

## 4.3: Variance

Recall the **expected value** of a real-valued random variable is the mean of the variable, and is a measure of the center of the distribution. Recall also that by taking the expected value of various transformations of the variable, we can measure other interesting characteristics of the distribution. In this section, we will study expected values that measure the spread of the distribution about the mean.

### Basic Theory

#### Definitions and Interpretations

As usual, we start with a random experiment modeled by a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . 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 random variable for the experiment, taking values in  $S \subseteq \mathbb{R}$ . Recall that  $\mathbb{E}(X)$ , the expected value (or mean) of  $X$  gives the center of the distribution of  $X$ .

The *variance* and *standard deviation* of  $X$  are defined by

1.  $\text{var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2)$
2.  $\text{sd}(X) = \sqrt{\text{var}(X)}$

Implicit in the definition is the assumption that the mean  $\mathbb{E}(X)$  exists, as a real number. If this is not the case, then  $\text{var}(X)$  (and hence also  $\text{sd}(X)$ ) are undefined. Even if  $\mathbb{E}(X)$  does exist as a real number, it's possible that  $\text{var}(X) = \infty$ . For the remainder of our discussion of the basic theory, we will assume that expected values that are mentioned exist as real numbers.

The variance and standard deviation of  $X$  are both measures of the spread of the distribution about the mean. Variance (as we will see) has nicer mathematical properties, but its physical unit is the square of that of  $X$ . Standard deviation, on the other hand, is not as nice mathematically, but has the advantage that its physical unit is the same as that of  $X$ . When the random variable  $X$  is understood, the standard deviation is often denoted by  $\sigma$ , so that the variance is  $\sigma^2$ .

Recall that the *second moment* of  $X$  about  $a \in \mathbb{R}$  is  $\mathbb{E}[(X - a)^2]$ . Thus, the variance is the second moment of  $X$  about the mean  $\mu = \mathbb{E}(X)$ , or equivalently, the *second central moment* of  $X$ . In general, the second moment of  $X$  about  $a \in \mathbb{R}$  can also be thought of as the *mean square error* if the constant  $a$  is used as an *estimate* of  $X$ . In addition, second moments have a nice interpretation in physics. If we think of the distribution of  $X$  as a mass distribution in  $\mathbb{R}$ , then the second moment of  $X$  about  $a \in \mathbb{R}$  is the *moment of inertia* of the mass distribution about  $a$ . This is a measure of the resistance of the mass distribution to any change in its rotational motion about  $a$ . In particular, the variance of  $X$  is the moment of inertia of the mass distribution about the center of mass  $\mu$ .



Figure 4.3.1: The moment of inertia about  $a$ .

The mean square error (or equivalently the moment of inertia) about  $a$  is minimized when  $a = \mu$ :

Let  $\text{mse}(a) = \mathbb{E}[(X - a)^2]$  for  $a \in \mathbb{R}$ . Then mse is minimized when  $a = \mu$ , and the minimum value is  $\sigma^2$ .

Proof

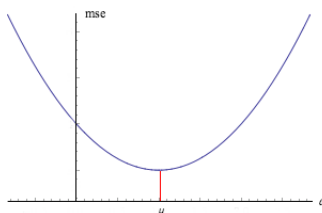


Figure 4.3.2:  $\text{mse}(a)$  is minimized when  $a = \mu$ .

The relationship between measures of center and measures of spread is studied in more detail in the advanced section on vector spaces of random variables.

#### Properties

The following exercises give some basic properties of variance, which in turn rely on basic properties of expected value. As usual, be sure to try the proofs yourself before reading the ones in the text. Our first results are computational formulas based on the change of variables formula for expected value

Let  $\mu = \mathbb{E}(X)$ .

1. If  $X$  has a discrete distribution with probability density function  $f$ , then  $\text{var}(X) = \sum_{x \in S} (x - \mu)^2 f(x)$ .
2. If  $X$  has a continuous distribution with probability density function  $f$ , then  $\text{var}(X) = \int_S (x - \mu)^2 f(x) dx$

Proof

1. This follows from the discrete version of the change of variables formula.
2. Similarly, this follows from the continuous version of the change of variables formula.

Our next result is a variance formula that is usually better than the definition for computational purposes.

$$\text{var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

Proof

Let  $\mu = \mathbb{E}(X)$ . Using the linearity of expected value we have

$$\text{var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - 2\mu \mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2 \quad (4.3.1)$$

Of course, by the change of variables formula,  $\mathbb{E}(X^2) = \sum_{x \in S} x^2 f(x)$  if  $X$  has a discrete distribution, and  $\mathbb{E}(X^2) = \int_S x^2 f(x) dx$  if  $X$  has a continuous distribution. In both cases,  $f$  is the probability density function of  $X$ .

Variance is always nonnegative, since it's the expected value of a nonnegative random variable. Moreover, any random variable that really is *random* (not a constant) will have strictly positive variance.

The nonnegative property.

1.  $\text{var}(X) \geq 0$
2.  $\text{var}(X) = 0$  if and only if  $\mathbb{P}(X = c) = 1$  for some constant  $c$  (and then of course,  $\mathbb{E}(X) = c$ ).

Proof

These results follow from the basic positive property of expected value. Let  $\mu = \mathbb{E}(X)$ . First  $(X - \mu)^2 \geq 0$  with probability 1 so  $\mathbb{E}[(X - \mu)^2] \geq 0$ . In addition,  $\mathbb{E}[(X - \mu)^2] = 0$  if and only if  $\mathbb{P}(X = \mu) = 1$ .

Our next result shows how the variance and standard deviation are changed by a linear transformation of the random variable. In particular, note that variance, unlike general expected value, is not a linear operation. This is not really surprising since the variance is the expected value of a nonlinear function of the variable:  $x \mapsto (x - \mu)^2$ .

If  $a, b \in \mathbb{R}$  then

1.  $\text{var}(a + bX) = b^2 \text{var}(X)$
2.  $\text{sd}(a + bX) = |b| \text{sd}(X)$

Proof

1. Let  $\mu = \mathbb{E}(X)$ . By linearity,  $\mathbb{E}(a + bX) = a + b\mu$ . Hence  $\text{var}(a + bX) = \mathbb{E}([(a + bX) - (a + b\mu)]^2) = \mathbb{E}[b^2(X - \mu)^2] = b^2 \text{var}(X)$ .
2. This result follows from (a) by taking square roots.

Recall that when  $b > 0$ , the linear transformation  $x \mapsto a + bx$  is called a *location-scale* transformation and often corresponds to a change of location and change of scale in the physical units. For example, the change from inches to centimeters in a measurement of length is a scale transformation, and the change from Fahrenheit to Celsius in a measurement of temperature is both a location and scale transformation. The previous result shows that when a location-scale transformation is applied to a random variable, the standard deviation does not depend on the location parameter, but is multiplied by the scale factor. There is a particularly important location-scale transformation.

Suppose that  $X$  is a random variable with mean  $\mu$  and variance  $\sigma^2$ . The random variable  $Z$  defined as follows is the *standard score* of  $X$ .

$$Z = \frac{X - \mu}{\sigma} \quad (4.3.2)$$

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

Proof

1. From the linearity of expected value,  $\mathbb{E}(Z) = \frac{1}{\sigma} [\mathbb{E}(X) - \mu] = 0$
2. From the [scaling property](#),  $\text{var}(Z) = \frac{1}{\sigma^2} \text{var}(X) = 1$ .

Since  $X$  and its mean and standard deviation all have the same physical units, the standard score  $Z$  is dimensionless. It measures the directed distance from  $\mathbb{E}(X)$  to  $X$  in terms of standard deviations.

Let  $Z$  denote the standard score of  $X$ , and suppose that  $Y = a + bX$  where  $a, b \in \mathbb{R}$  and  $b \neq 0$ .

1. If  $b > 0$ , the standard score of  $Y$  is  $Z$ .
2. If  $b < 0$ , the standard score of  $Y$  is  $-Z$ .

Proof

$E(Y) = a + bE(X)$  and  $sd(Y) = |b| sd(X)$ . Hence

$$\frac{Y - E(Y)}{sd(Y)} = \frac{b}{|b|} \frac{X - E(X)}{sd(X)} \quad (4.3.3)$$

As just noted, when  $b > 0$ , the variable  $Y = a + bX$  is a location-scale transformation and often corresponds to a change of physical units. Since the standard score is dimensionless, it's reasonable that the standard scores of  $X$  and  $Y$  are the same. Here is another standardized measure of dispersion:

Suppose that  $X$  is a random variable with  $E(X) \neq 0$ . The *coefficient of variation* is the ratio of the standard deviation to the mean:

$$cv(X) = \frac{sd(X)}{E(X)} \quad (4.3.4)$$

The coefficient of variation is also dimensionless, and is sometimes used to compare variability for random variables with different means. We will learn how to compute the variance of the sum of two random variables in the section on covariance.

### Chebyshev's Inequality

*Chebyshev's inequality* (named after Pafnuty Chebyshev) gives an upper bound on the probability that a random variable will be more than a specified distance from its mean. This is often useful in applied problems where the distribution is unknown, but the mean and variance are known (at least approximately). In the following two results, suppose that  $X$  is a real-valued random variable with mean  $\mu = E(X) \in \mathbb{R}$  and standard deviation  $\sigma = sd(X) \in (0, \infty)$ .

#### Chebyshev's inequality 1.

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}, \quad t > 0 \quad (4.3.5)$$

Proof

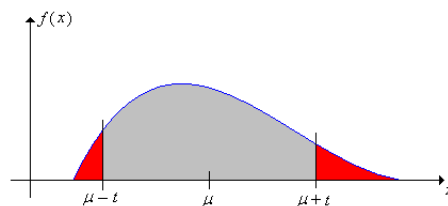


Figure 4.3.3: Chebyshev's inequality

Here's an alternate version, with the distance in terms of standard deviation.

#### Chebyshev's inequality 2.

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad k > 0 \quad (4.3.6)$$

Proof

Let  $t = k\sigma$  in the first version of [Chebyshev's inequality](#).

The usefulness of the Chebyshev inequality comes from the fact that it holds for any distribution (assuming only that the mean and variance exist). The tradeoff is that for many specific distributions, the Chebyshev bound is rather crude. Note in particular that the [first inequality](#) is useless when  $t \leq \sigma$ , and the [second inequality](#) is useless when  $k \leq 1$ , since 1 is an upper bound for the probability of any event. On the other hand, it's easy to construct a distribution for which [Chebyshev's inequality](#) is sharp for a specified value of  $t \in (0, \infty)$ . Such a distribution is given in an [exercise below](#).

### Examples and Applications

As always, be sure to try the problems yourself before looking at the solutions and answers.

## Indicator Variables

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

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

Proof

1. We proved this in the section on basic properties, although the result is so simple that we can do it again:  $\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p$ .
2. Note that  $X^2 = X$  since  $X$  only takes values 0 and 1. Hence  $\mathbb{E}(X^2) = p$  and therefore  $\text{var}(X) = p - p^2 = p(1 - p)$ .

The graph of  $\text{var}(X)$  as a function of  $p$  is a parabola, opening downward, with roots at 0 and 1. Thus the minimum value of  $\text{var}(X)$  is 0, and occurs when  $p = 0$  and  $p = 1$  (when  $X$  is deterministic, of course). The maximum value is  $\frac{1}{4}$  and occurs when  $p = \frac{1}{2}$ .

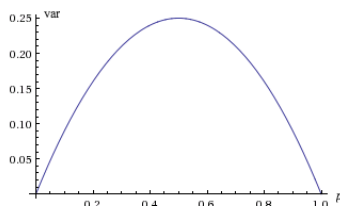


Figure 4.3.4: The variance of an indicator variable as a function of  $p$ .

## Uniform Distributions

Discrete uniform distributions are widely used in combinatorial probability, and model a point chosen *at random* from a finite set. The mean and variance have simple forms for the discrete uniform distribution on a set of evenly spaced points (sometimes referred to as a *discrete interval*):

Suppose that  $X$  has the discrete uniform distribution on  $\{a, a + h, \dots, a + (n - 1)h\}$  where  $a \in \mathbb{R}$ ,  $h \in (0, \infty)$ , and  $n \in \mathbb{N}_+$ . Let  $b = a + (n - 1)h$ , the right endpoint. Then

1.  $\mathbb{E}(X) = \frac{1}{2}(a + b)$ .
2.  $\text{var}(X) = \frac{1}{12}(b - a)(b - a + 2h)$ .

Proof

1. We proved this in the section on basic properties. Here it is again, using the formula for the sum of the first  $n - 1$  positive integers:

$$\mathbb{E}(X) = \frac{1}{n} \sum_{i=0}^{n-1} (a + ih) = \frac{1}{n} \left( na + h \frac{(n-1)n}{2} \right) = a + \frac{(n-1)h}{2} = \frac{a+b}{2} \quad (4.3.7)$$

2. Note that

$$\mathbb{E}(X^2) = \frac{1}{n} \sum_{i=0}^{n-1} (a + ih)^2 = \frac{1}{n-1} \sum_{i=0}^{n-1} (a^2 + 2ahi + h^2 i^2) \quad (4.3.8)$$

Using the formulas for the sum of the first  $n - 1$  positive integers, and the sum of the squares of the first  $n - 1$  positive integers, we have

$$\mathbb{E}(X^2) = \frac{1}{n} \left[ na^2 + 2ah \frac{(n-1)n}{2} + h^2 \frac{(n-1)n(2n-1)}{6} \right] \quad (4.3.9)$$

Using [computational formula](#) and simplifying gives the result.

Note that mean is simply the average of the endpoints, while the variance depends only on difference between the endpoints and the step size.

Open the [special distribution simulator](#), and select the discrete uniform distribution. Vary the parameters and note the location and size of the mean  $\pm$  standard deviation bar in relation to the probability density function. For selected values of the parameters, run the simulation 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

Next, 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 the continuous uniform distribution on the interval  $[a, b]$  where  $a, b \in \mathbb{R}$  with  $a < b$ . Then

1.  $\mathbb{E}(X) = \frac{1}{2}(a + b)$
2.  $\text{var}(X) = \frac{1}{12}(b - a)^2$

Proof

1.  $\mathbb{E}(X) = \int_a^b x \frac{1}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$
2.  $\mathbb{E}(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)}$ . The variance result then follows from (a), the [computational formula](#) and simple algebra.

Note that the mean is the midpoint of the interval and the variance depends only on the length of the interval. Compare this with the results in the discrete case.

Open the special distribution simulator, and select the continuous uniform distribution. This is the uniform distribution the interval  $[a, a + w]$ . Vary the parameters and note the location and size of the mean  $\pm$  standard deviation bar in relation to the probability density function. For selected values of the parameters, run the simulation 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

## 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. In the following problems, you will compute the mean and variance for each of the various types of dice. Be sure to compare the results.

A standard, fair die is thrown and the score  $X$  is recorded. Sketch the graph of the probability density function and compute each of the following:

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

Answer

1.  $\frac{7}{2}$
2.  $\frac{35}{12}$

An ace-six flat die is thrown and the score  $X$  is recorded. Sketch the graph of the probability density function and compute each of the following:

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

Answer

1.  $\frac{7}{2}$
2.  $\frac{15}{4}$

A two-five flat die is thrown and the score  $X$  is recorded. Sketch the graph of the probability density function and compute each of the following:

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

Answer

1.  $\frac{7}{2}$
2.  $\frac{11}{4}$

A three-four flat die is thrown and the score  $X$  is recorded. Sketch the graph of the probability density function and compute each of the following:

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

Answer

1.  $\frac{7}{2}$
2.  $\frac{9}{4}$

In the dice experiment, select one die. For each of the following cases, note the location and size of the mean  $\pm$  standard deviation bar in relation to the probability density function. Run the experiment 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

1. Fair die
2. Ace-six flat die
3. Two-five flat die
4. Three-four flat die

### The Poisson Distribution

Recall that the *Poisson distribution* is a discrete distribution on  $\mathbb{N}$  with probability density function  $f$  given by

$$f(n) = e^{-a} \frac{a^n}{n!}, \quad n \in \mathbb{N} \quad (4.3.10)$$

where  $a \in (0, \infty)$  is a parameter. The Poisson distribution is named after Simeon Poisson and is widely used to model the number of “random points” in a region of time or space; the parameter  $a$  is proportional to the size of the region. The Poisson distribution is studied in detail in the chapter on the Poisson Process.

Suppose that  $N$  has the Poisson distribution with parameter  $a$ . Then

1.  $\mathbb{E}(N) = a$
2.  $\text{var}(N) = a$

Proof

1. We did this computation in the previous section. Here it is again:

$$\mathbb{E}(N) = \sum_{n=0}^{\infty} n e^{-a} \frac{a^n}{n!} = e^{-a} \sum_{n=1}^{\infty} \frac{a^n}{(n-1)!} = e^{-a} a \sum_{n=1}^{\infty} \frac{a^{n-1}}{(n-1)!} = e^{-a} a e^a = a. \quad (4.3.11)$$

2. First we compute the *second factorial moment*:

$$\mathbb{E}[N(N-1)] = \sum_{n=1}^{\infty} n(n-1) e^{-a} \frac{a^n}{n!} = \sum_{n=2}^{\infty} e^{-a} \frac{a^n}{(n-2)!} = e^{-a} a^2 \sum_{n=2}^{\infty} \frac{a^{n-2}}{(n-2)!} = a^2 e^{-a} e^a = a^2 \quad (4.3.12)$$

Hence,  $E(N^2) = \mathbb{E}[N(N-1)] + \mathbb{E}(N) = a^2 + a$  and so  $\text{var}(N) = (a^2 + a) - a^2 = a$ .

Thus, the parameter of the Poisson distribution is both the mean and the variance of the distribution.

In the Poisson experiment, the parameter is  $a = rt$ . Vary the parameter and note the size and location of the mean  $\pm$  standard deviation bar in relation to the probability density function. For selected values of the parameter, run the experiment 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

### The Geometric Distribution

Recall that *Bernoulli trials* are independent trials each with two outcomes, which in the language of reliability, are called *success* and *failure*. The probability of success on each trial is  $p \in [0, 1]$ . A separate chapter on Bernoulli Trials explores this random process in more detail. It is named for Jacob Bernoulli. If  $p \in (0, 1]$ , the trial number  $N$  of the first success has the geometric distribution on  $\mathbb{N}_+$  with success parameter  $p$ . The probability density function  $f$  of  $N$  is given by

$$f(n) = p(1-p)^{n-1}, \quad n \in \mathbb{N}_+ \quad (4.3.13)$$

Suppose that  $N$  has the geometric distribution on  $\mathbb{N}_+$  with success parameter  $p \in (0, 1]$ . Then

1.  $\mathbb{E}(N) = \frac{1}{p}$
2.  $\text{var}(N) = \frac{1-p}{p^2}$

Proof

1. We proved this in the section on basic properties. Here it is again:

$$\mathbb{E}(N) = \sum_{n=1}^{\infty} n p (1-p)^{n-1} = -p \frac{d}{dp} \sum_{n=0}^{\infty} (1-p)^n = -p \frac{d}{dp} \frac{1}{p} = \frac{1}{p^2} = \frac{1}{p} \quad (4.3.14)$$

2. First we compute the *second factorial moment*:

$$\mathbb{E}[N(N-1)] = \sum_{n=2}^{\infty} n(n-1) (1-p)^{n-1} p = p(1-p) \frac{d^2}{dp^2} \sum_{n=0}^{\infty} (1-p)^n = p(1-p) \frac{d^2}{dp^2} \frac{1}{p} = p(1-p) \frac{2}{p^3} = \frac{2(1-p)}{p^2} \quad (4.3.15)$$

Hence  $\mathbb{E}(N^2) = \mathbb{E}[N(N-1)] + \mathbb{E}(N) = 2/p^2 - 1/p$  and hence  $\text{var}(X) = 2/p^2 - 1/p - 1/p^2 = 1/p^2 - 1/p$ .

Note that the variance is 0 when  $p = 1$ , not surprising since  $X$  is deterministic in this case.

In the negative binomial experiment, set  $k = 1$  to get the geometric distribution. Vary  $p$  with the scroll bar and note the size and location of the mean  $\pm$  standard deviation bar in relation to the probability density function. For selected values of  $p$ , run the experiment 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

Suppose that  $N$  has the geometric distribution with parameter  $p = \frac{3}{4}$ . Compute the true value and the Chebyshev bound for the probability that  $N$  is at least 2 standard deviations away from the mean.

Answer

1.  $\frac{1}{16}$
2.  $\frac{1}{4}$

## The Exponential Distribution

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

$$f(t) = re^{-rt}, \quad t \in [0, \infty) \quad (4.3.16)$$

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  $T$  has the exponential distribution with rate parameter  $r$ . Then

1.  $\mathbb{E}(T) = \frac{1}{r}$ .
2.  $\text{var}(T) = \frac{1}{r^2}$ .

Proof

1. We proved this in the section on basic properties. Here it is again, using integration by parts:

$$\mathbb{E}(T) = \int_0^\infty t r e^{-rt} dt = -t e^{-rt} \Big|_0^\infty + \int_0^\infty e^{-rt} dt = 0 - \frac{1}{r} e^{-rt} \Big|_0^\infty = \frac{1}{r} \quad (4.3.17)$$

2. Integrating by parts again and using (a), we have

$$\mathbb{E}(T^2) = \int_0^\infty t^2 r e^{-rt} dt = -t^2 e^{-rt} \Big|_0^\infty + \int_0^\infty 2t e^{-rt} dt = 0 + \frac{2}{r^2} \quad (4.3.18)$$

$$\text{Hence } \text{var}(T) = \frac{2}{r^2} - \frac{1}{r^2} = \frac{1}{r^2}$$

Thus, for the exponential distribution, the mean and standard deviation are the same.

In the gamma experiment, set  $k = 1$  to get the exponential distribution. Vary  $r$  with the scroll bar and note the size and location of the mean  $\pm$  standard deviation bar in relation to the probability density function. For selected values of  $r$ , run the experiment 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

Suppose that  $X$  has the exponential distribution with rate parameter  $r > 0$ . Compute the true value and the Chebyshev bound for the probability that  $X$  is at least  $k$  standard deviations away from the mean.

Answer

1.  $e^{-(k+1)}$
2.  $\frac{1}{k^2}$

## The Pareto Distribution

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

$$f(x) = \frac{a}{x^{a+1}}, \quad x \in [1, \infty) \quad (4.3.19)$$

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  $a$ . Then

1.  $\mathbb{E}(X) = \infty$  if  $0 < a \leq 1$  and  $\mathbb{E}(X) = \frac{a}{a-1}$  if  $1 < a < \infty$

2.  $\text{var}(X)$  is undefined if  $0 < a \leq 1$ ,  $\text{var}(X) = \infty$  if  $1 < a \leq 2$ , and  $\text{var}(X) = \frac{a}{(a-1)^2(a-2)}$  if  $2 < a < \infty$

Proof

1. We proved this in the section on basic properties. Here it is again:

$$\mathbb{E}(X) = \int_1^\infty x \frac{a}{x^{a+1}} dx = \int_1^\infty \frac{a}{x^a} dx = \frac{a}{-a+1} x^{-a+1} \Big|_1^\infty = \begin{cases} \infty, & 0 < a < 1 \\ \frac{a}{a-1}, & a > 1 \end{cases} \quad (4.3.20)$$

When  $a = 1$ ,  $\mathbb{E}(X) = \int_1^\infty \frac{1}{x} dx = \ln x \Big|_1^\infty = \infty$

2. If  $0 < a \leq 1$  then  $\mathbb{E}(X) = \infty$  and so  $\text{var}(X)$  is undefined. On the other hand,

$$\mathbb{E}(X^2) = \int_1^\infty x^2 \frac{a}{x^{a+1}} dx = \int_1^\infty \frac{a}{x^{a-1}} dx = ax^{-a+2} \Big|_1^\infty = \begin{cases} \infty, & 0 < a < 2 \\ \frac{a}{a-2}, & a > 2 \end{cases} \quad (4.3.21)$$

When  $a = 2$ ,  $\mathbb{E}(X^2) = \int_1^\infty \frac{2}{x} dx = \infty$ . Hence  $\text{var}(X) = \infty$  if  $1 < a \leq 2$  and  $\text{var}(X) = \frac{a}{a-2} - \left(\frac{a}{a-1}\right)^2$  if  $a > 2$ .

In the special distribution simulator, select the Pareto distribution. Vary  $a$  with the scroll bar and note the size and location of the mean  $\pm$  standard deviation bar. For each of the following values of  $a$ , run the experiment 1000 times and note the behavior of the empirical mean and standard deviation.

1.  $a = 1$
2.  $a = 2$
3.  $a = 3$

### The Normal Distribution

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

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R} \quad (4.3.22)$$

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.  $\text{var}(Z) = 1$

Proof

1. We proved this in the section on basic properties. Here it is again:

$$\mathbb{E}(Z) = \int_{-\infty}^\infty z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \Big|_{-\infty}^\infty = 0 - 0 \quad (4.3.23)$$

2. From (a),  $\text{var}(Z) = \mathbb{E}(Z^2) = \int_{-\infty}^\infty z^2 \phi(z) dz$ . Integrate by parts with  $u = z$  and  $dv = z\phi(z) dz$ . Thus,  $du = dz$  and  $v = -\phi(z)$ . Hence

$$\text{var}(Z) = -z\phi(z) \Big|_{-\infty}^\infty + \int_{-\infty}^\infty \phi(z) dz = 0 + 1 \quad (4.3.24)$$

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R} \quad (4.3.25)$$

Moreover, if  $Z$  has the standard normal distribution, then  $X = \mu + \sigma Z$  has the normal distribution with location parameter  $\mu$  and scale parameter  $\sigma$ . As the notation suggests, the location parameter is the mean of the distribution and the scale parameter is the standard deviation.

Suppose that  $X$  has the normal distribution with location parameter  $\mu$  and scale parameter  $\sigma$ . Then

1.  $\mathbb{E}(X) = \mu$
2.  $\text{var}(X) = \sigma^2$

Proof

We could use the probability density function, of course, but it's much better to use the representation of  $X$  in terms of the standard normal variable  $Z$ , and use properties of expected value and variance.



1.  $\mathbb{E}(X) = \mu + \sigma\mathbb{E}(Z) = \mu + 0 = \mu$
2.  $\text{var}(X) = \sigma^2\text{var}(Z) = \sigma^2 \cdot 1 = \sigma^2$  .

So to summarize, if  $X$  has a normal distribution, then its standard score  $Z$  has the standard normal distribution.

In the special distribution simulator, select the normal distribution. Vary the parameters and note the shape and location of the mean  $\pm$  standard deviation bar in relation to the probability density function. For selected parameter values, run the experiment 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

## Beta Distributions

The distributions in this subsection belong to the family of *beta distributions*, which are 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 a beta distribution with probability density function  $f$ . In each case below, graph  $f$  below and compute the mean and variance.

1.  $f(x) = 6x(1-x)$  for  $x \in [0, 1]$
2.  $f(x) = 12x^2(1-x)$  for  $x \in [0, 1]$
3.  $f(x) = 12x(1-x)^2$  for  $x \in [0, 1]$

Answer

1.  $\mathbb{E}(X) = \frac{1}{2}$ ,  $\text{var}(X) = \frac{1}{20}$
2.  $\mathbb{E}(X) = \frac{3}{5}$ ,  $\text{var}(X) = \frac{1}{25}$
3.  $\mathbb{E}(X) = \frac{2}{6}$ ,  $\text{var}(X) = \frac{1}{25}$

In the special distribution simulator, select the beta distribution. The parameter values below give the distributions in the previous exercise. In each case, note the location and size of the mean  $\pm$  standard deviation bar. Run the experiment 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

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

Suppose that a sphere has a random radius  $R$  with probability density function  $f$  given by  $f(r) = 12r^2(1-r)$  for  $r \in [0, 1]$ . Find the mean and standard deviation of each of the following:

1. The circumference  $C = 2\pi R$
2. The surface area  $A = 4\pi R^2$
3. The volume  $V = \frac{4}{3}\pi R^3$

Answer

1.  $\frac{6}{5}\pi$ ,  $\frac{2}{5}\pi$
2.  $\frac{8}{5}\pi$ ,  $\frac{2}{5}\sqrt{\frac{38}{7}}\pi$
3.  $\frac{8}{21}\pi$ ,  $\frac{8}{3}\sqrt{\frac{19}{1470}}\pi$

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.  $\text{var}(X)$

Answer

1.  $\frac{1}{2}$
2.  $\frac{1}{8}$

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 location and size of the mean  $\pm$  standard deviation bar in relation to the probability density function. Run the simulation 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation.

Suppose that the grades on a test are described by the random variable  $Y = 100X$  where  $X$  has the beta distribution with probability density function  $f$  given by  $f(x) = 12x(1-x)^2$  for  $x \in [0, 1]$ . The grades are generally low, so the teacher decides to “curve” the grades using the transformation  $Z = 10\sqrt{Y} = 100\sqrt{X}$ . Find the mean and standard deviation of each of the following variables:

1.  $X$
2.  $Y$
3.  $Z$

Answer

1.  $\mathbb{E}(X) = \frac{2}{5}$ ,  $\text{sd}(X) = \frac{1}{5}$
2.  $\mathbb{E}(Y) = 40$ ,  $\text{sd}(Y) = 20$
3.  $\mathbb{E}(Z) = 60.95$ ,  $\text{sd}(Z) = 16.88$

### Exercises on Basic Properties

Suppose that  $X$  is a real-valued random variable with  $\mathbb{E}(X) = 5$  and  $\text{var}(X) = 4$ . Find each of the following:

1.  $\text{var}(3X - 2)$
2.  $\mathbb{E}(X^2)$

Answer

1. 36
2. 29

Suppose that  $X$  is a real-valued random variable with  $\mathbb{E}(X) = 2$  and  $\mathbb{E}[X(X-1)] = 8$ . Find each of the following:

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

Answer

1. 10
2. 6

The expected value  $\mathbb{E}[X(X-1)]$  is an example of a *factorial moment*.

Suppose that  $X_1$  and  $X_2$  are independent, real-valued random variables with  $\mathbb{E}(X_i) = \mu_i$  and  $\text{var}(X_i) = \sigma_i^2$  for  $i \in \{1, 2\}$ . Then

1.  $\mathbb{E}(X_1 X_2) = \mu_1 \mu_2$
2.  $\text{var}(X_1 X_2) = \sigma_1^2 \sigma_2^2 + \sigma_1^2 \mu_2^2 + \sigma_2^2 \mu_1^2$

Proof

1. This is an important, basic result that was proved in the section on basic properties.
2. Since  $X_1^2$  and  $X_2^2$  are also independent, we have  $\mathbb{E}(X_1^2 X_2^2) = \mathbb{E}(X_1^2) \mathbb{E}(X_2^2) = (\sigma_1^2 + \mu_1^2)(\sigma_2^2 + \mu_2^2)$ . The result then follows from the [computational formula](#) and algebra.

Marilyn Vos Savant has an IQ of 228. Assuming that the distribution of IQ scores has mean 100 and standard deviation 15, find Marilyn's standard score.

Answer

$$z = 8.53$$

Fix  $t \in (0, \infty)$ . Suppose that  $X$  is the discrete random variable with probability density function defined by  $\mathbb{P}(X = t) = \mathbb{P}(X = -t) = p$ ,  $\mathbb{P}(X = 0) = 1 - 2p$ , where  $p \in (0, \frac{1}{2})$ . Then equality holds in [Chebyshev's inequality](#) at  $t$ .

Proof

Note that  $\mathbb{E}(X) = 0$  and  $\text{var}(X) = \mathbb{E}(X^2) = 2pt^2$ . So  $\mathbb{P}(|X| \geq t) = 2p$  and  $\sigma^2/t^2 = 2p$ .

This page titled [4.3: Variance](#) 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.