

3 Continuous Random Variables

3.1 Introduction

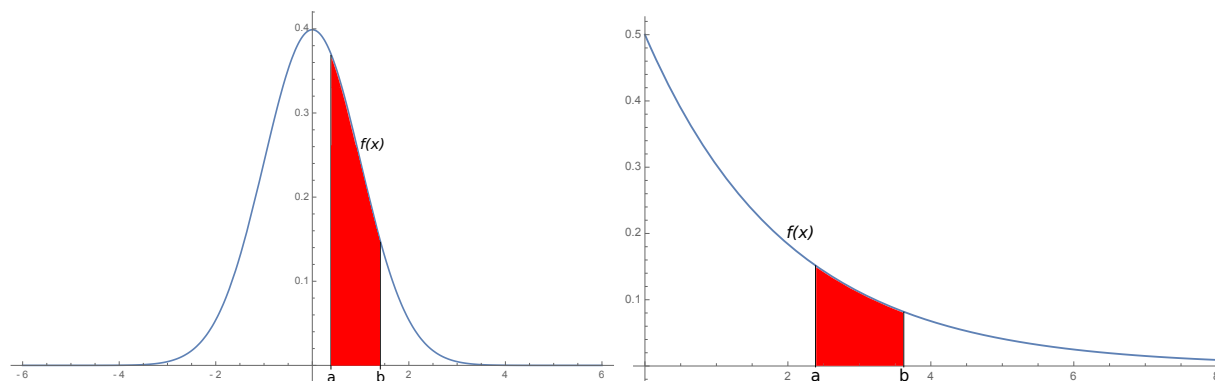
Many quantities of interest in the real world are not discrete in nature. Examples include the following:

1. The amount of rainfall in one day
2. The weight of an adult chimpanzee
3. The functional lifetime of an O-ring in a jet engine

A random variable which can take on any value within a range is called a *continuous random variable*. These are fundamentally different from discrete random variables in the following way. Recall that to specify a discrete random variable, all we have to do is assign probabilities between 0 and 1 to every possible output of the random variable in such a way that all the probabilities add up to 1. This is not possible for a random variable which can take values in an interval on the real number line. We must, therefore, use a different technique to describe a continuous random variable. Here we will use calculus for (essentially) the first time in the course.

3.2 Probability Density Functions

For a continuous random variable, rather than talk about the probability that a random variable equals a particular value, we talk about the probability that a random value falls within a particular range. (In fact, as we will see, the probability that a continuous random variable equals a specific value is 0.) A continuous random variable is described by a *probability density function (pdf)*. Here are examples of pdfs:



A pdf is a nonnegative function $f(x)$ where the total area under the curve is 1 (this is analogous to the discrete case where the probabilities of all the outputs sum to 1). The probability that a random variable falls into an interval $[a, b]$ is the area under the density curve between a and b . This is illustrated in red in the above pdfs. Since we are talking about areas under curves, we need to use calculus. In particular, we need to integrate! Here is the formal definition of a pdf:

Probability Density Function (pdf)

The function $f(y)$ is a *probability density function* (pdf) for a continuous random variable Y if $f(y) \geq 0$ for all y , and

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

The probability that Y falls into the interval $[a, b]$ is the area under the density curve between a and b , i.e.

$$\mathbb{P}(a \leq Y \leq b) = \int_a^b f(y)dy$$

Before we continue, let's mention one way in which continuous and discrete distributions are very different. For a continuous distribution, the probability of any single point is 0. We can see that from the density function since for a continuous random variable Y with density y ,

$$\mathbb{P}(Y = a) = \int_a^a f(y)dy = 0$$

Thus all the following probabilities are the same:

$$\mathbb{P}(a \leq Y \leq b) = \mathbb{P}(a \leq Y < b) = \mathbb{P}(a < Y \leq b) = \mathbb{P}(a < Y < b) = \int_a^b f(y)dy$$

In other words, it does not matter which inequality sign ($<$ or \leq) we use for the endpoints of the interval since the probability of hitting the endpoints is 0. This is not at all the case for the discrete case, where individual points have positive probability. If X is the number of flips of heads in 20 tosses of a fair coin, then $\mathbb{P}(8 \leq X \leq 12)$ and $\mathbb{P}(8 < X \leq 12)$ are very different!

Example. Consider the function

$$f(y) = \begin{cases} cy^2 & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the value of c for which $f(y)$ is a valid density function.

First note that $f(y)$ is nonnegative, so there is nothing to worry about there. Next we need to make sure that the density function integrates to 1.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(y)dy \\ &= \int_0^2 cy^2 dy && \text{since } f(y) \text{ is 0 outside } [0, 2] \\ &= c \frac{y^3}{3} \Big|_0^2 \\ &= \frac{8}{3}c \end{aligned}$$

Thus, choosing $c = 3/8$, we have a valid density function.

2. What is $P(1 \leq Y \leq 2)$?

Integrating the density function from 1 to 2, we get:

$$\begin{aligned}\mathbb{P}(1 \leq Y \leq 2) &= \int_1^2 f(y)dy \\ &= \int_1^2 \frac{3}{8}y^2 dy \\ &= \left. \frac{3}{8} \frac{y^3}{3} \right|_1^2 \\ &= \frac{7}{8}\end{aligned}$$

3.3 Cumulative Distribution Functions

Another way to describe a continuous random variable is with its *cumulative distribution function* (cdf)¹.

Cumulative Distribution Function (cdf)

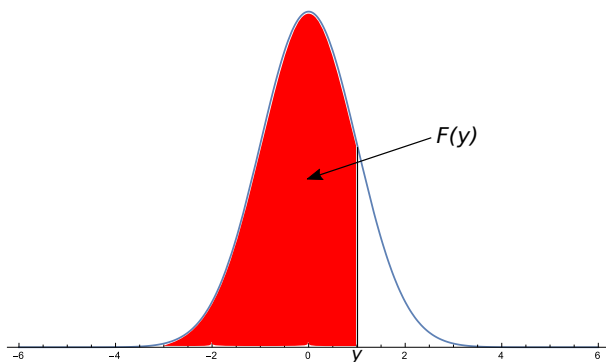
Let Y be a random variable. The *cumulative distribution function* cdf of Y , denoted $F(y)$, is defined by

$$F(y) = \mathbb{P}(Y \leq y)$$

If Y has density function $f(y)$, then

$$F(y) = \int_{-\infty}^y f(y)dy$$

The cdf gives the probability that our random variable Y is less than or equal to a certain value. For a continuous random variable, $F(y)$ can be visualized graphically as the area under the density curve to the left of y .



¹Sometimes you will see this called simply a *distribution function*. I will always use the term cdf.

Note that traditionally the cdf is written with an uppercase F , while the density is written with a lowercase f . The cdf is defined for discrete as well as continuous random variables, although we will never use it in the discrete case². The cdf has the following properties.

Properties of cdfs

Let Y be a random variable with cdf $F(y)$. Then:

1. $F(y)$ is a nondecreasing function.
2. $\lim_{y \rightarrow -\infty} F(y) = 0$
3. $\lim_{y \rightarrow \infty} F(y) = 1$
4. For a continuous random variable Y , the cdf $F(y)$ is a continuous function.

For a continuous random variable, the cdf and pdf are related via the fundamental theorem of calculus.

Relationship between cdf and pdf

Let Y be a continuous random variable with cdf $F(y)$ and density $f(y)$. Then we have the following relationships:

1.

$$F(y) = \int_{-\infty}^y f(y) dy$$

2.

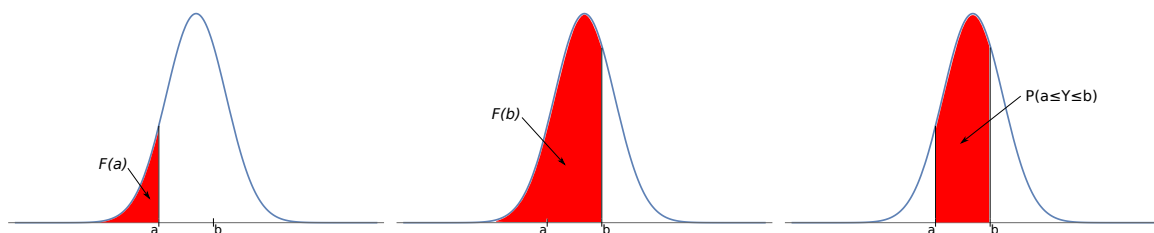
$$f(y) = \frac{DF(y)}{dy} = F'(y)$$

3.

$$\mathbb{P}(a \leq Y \leq b) = \int_a^b f(y) dy = F(b) - F(a)$$

We can illustrate the third relationship above, $\mathbb{P}(a \leq Y \leq b) = F(b) - F(a)$, using the graphs below. You can see that subtracting the first area from the second area yields the third area.

²The cdf for a discrete random variable is always a step function, since the cdf only increases on the finite or countable set of points which have positive probabilities.



You may ask why we bother at all with cdfs since it seems easier to work with densities. There are three good reasons to consider CDFs. From a mathematical standpoint, the cdf is a more fundamental object than the pdf; every continuous random variable has a cdf, while there are continuous random variables which have no density³. From a practical standpoint, the normal distribution, perhaps the most important distribution is all of probability, has a density which is unwieldy, thus we will use the cdf to compute probabilities involving that distribution. Finally, the cdf is used to define the *median* and *quartiles* of a probability distribution, which are important descriptors of a probability distribution.

3.4 Median and Quartiles

The *median* is the “middle” of a probability distribution. It is the value which separates the upper half of the distribution from the lower half of the distribution. The median is more robust to outlier values than the mean, and so in some cases may be a better descriptor of a typical outcome than the mean. Mathematically, if Y is a random variable, then we can define the median of Y as the value m such that $\mathbb{P}(Y \leq m) = \mathbb{P}(Y \geq m) = 1/2$. For a continuous random variable Y with cdf $F(y)$, we can define the median using the cdf as the value m for which $F(m) = 1/2$. Similarly we can define the 1st and 3rd quartiles (the median is sometimes called the 2nd quartile).

Median and quartiles

Let Y be a continuous random variable with cdf $F(y)$. Then we define the *median* m , *first quartile* Q_1 , and *third quartile* Q_3 by the following relationships:

$$F(Q_1) = \mathbb{P}(Y \leq Q_1) = 1/4$$

$$F(m) = \mathbb{P}(Y \leq m) = 1/2$$

$$F(Q_3) = \mathbb{P}(Y \leq Q_3) = 3/4$$

Let's revisit our example from the previous section.

Example. Let Y be a continuous random variable with density $f(y)$ defined by

$$f(y) = \begin{cases} \frac{3}{8}y^2 & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

³All continuous random variables we will encounter in this course will have a density

Find the median of Y .

Let m be the median of Y . Since $F(m) = 1/2$, where $F(y)$ is the cdf of Y , we need to first find the cdf. For $y < 0$, $F(y) = 0$ and for $y > 2$, $F(y) = 1$ (do you see why this is the case?) For $0 \leq y \leq 2$, which is the region we care about, we integrate the density to get the cdf.

$$\begin{aligned} F(y) &= \int_0^y \frac{3}{8} t^2 dt \\ &= \left. \frac{3}{8} \frac{t^3}{3} \right|_0^y \\ &= \frac{y^3}{8} \end{aligned}$$

To find the median, we solve $F(m) = 1/2$ for m .

$$\begin{aligned} \frac{m^3}{8} &= \frac{1}{2} \\ m^3 &= 4 \\ m &= 2^{2/3} \approx 1.59 \end{aligned}$$

3.5 Expectation and Variance

Just as with discrete random variables, we can talk about the expected value and variance of a continuous random variable. They have the exact same interpretations as in the discrete case. Expectation and variance work almost exactly the same way in the continuous case as in the discrete case. In fact, we can use the exact same formulas, if we make two key changes:

1. The probability mass function $p(y)$ is replaced by the probability density function $f(y)$.
2. Summation is replaced by integration.

Making these changes, we have the following definitions for the expected value of a continuous random variable.

Expected value of a continuous random variable

Let Y be a continuous random variable with density $f(y)$. Then we define the expected value by

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

If $g(y)$ is a real-valued function, then the expected value of $G(Y)$ is given by

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) f(y) dy$$

The variance of a continuous random variable is defined the same way as in the discrete case.

Variance of a continuous random variable

Let Y be a continuous random variable with density $f(y)$, and let $\mu = \mathbb{E}(Y)$. Then the variance of Y is defined by

$$\text{Var}(Y) = \mathbb{E}[(Y - \mu)^2]$$

This is usually computed using the Magic Variance Formula, which holds for continuous random variables as well:

$$\text{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2$$

Let's compute the expected value and variance of the continuous random variable we used in the section on probability densities.

Example. Let Y be the continuous random variable defined by the pdf

$$f(y) = \begin{cases} \frac{3}{8}y^2 & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected value and variance of Y ?

Using the formula for expected value,

$$\begin{aligned} \mathbb{E}(Y) &= \int_{-\infty}^{\infty} yf(y)dy \\ &= \int_0^2 y \frac{3}{8}y^2 dy \\ &= \frac{3}{8} \int_0^2 y^3 dy \\ &= \frac{3}{8} \frac{y^4}{4} \Big|_0^2 \\ &= \frac{3}{8} \frac{16}{4} = 1.5 \end{aligned}$$

For the variance, we will compute $\mathbb{E}(Y^2)$ and use the Magic Variance Formula.

$$\begin{aligned}\mathbb{E}(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y) dy \\ &= \int_0^2 y^2 \frac{3}{8} y^2 dy \\ &= \frac{3}{8} \int_0^2 y^4 dy \\ &= \frac{3}{8} \frac{y^5}{5} \Big|_0^2 \\ &= \frac{3}{8} \frac{32}{5} = 2.4\end{aligned}$$

Thus by the Magic Variance Formula we have:

$$\text{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = 2.4 - (1.5)^2 = 0.15$$

We will now look at three specific continuous distributions.

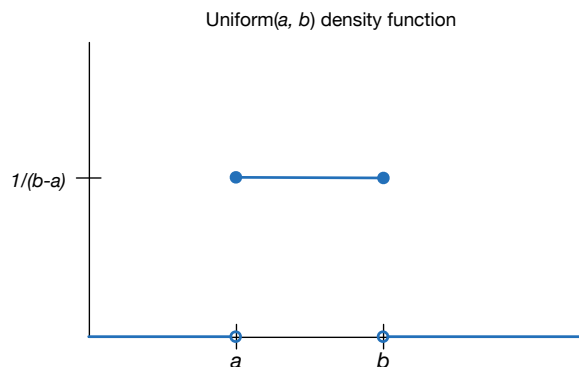
3.6 Continuous Uniform Distribution

The *continuous uniform distribution*, which we shall generally just call the *uniform distribution*, describes the probability distribution on a finite interval $[a, b]$ which has the property that all subintervals of equal length are equally probable. The uniform distribution must be specified on a finite interval and is not defined for interval of infinite length. Here are some examples where we can use the uniform distribution to model a problem.

1. A RIPTA bus is scheduled to arrive at tunnel on Thayer St. at 8:00 am, but experience has shown that its arrival times vary between 8:00 am and 8:15 am. We could model this as a uniform distribution on the time interval $[0, 15]$, representing the number of minutes the bus is behind schedule.
2. Strokkur, one of the most famous geysers in Iceland, erupts approximately once every 10 minutes⁴. For a given 10-minute interval, we can model the probability that Strokkur will erupt by a uniform distribution on the interval $[0, 10]$

A uniform distribution is specified in terms of parameters a and b , which are the endpoints of the interval $[a, b]$ on which the uniform distribution is defined. What is the density function for the uniform distribution? Look at the picture below:

⁴Its neighbor Geysir, from which we get the word “geyser”, hardly ever erupts these days.



The uniform density is a horizontal line between a and b and is 0 otherwise. (Does this make sense?) What is the height of the horizontal line? Since the integral of a probability density must integrate to 1, and since the uniform density is nothing more than a box, the area of the box must be 1. For that to be the case, the height of the box must be $1/(b - a)$. This is summarized below.

Uniform random variable

A continuous random variable Y has a *uniform distribution* on the interval $[a, b]$ if the pdf of Y is given by:

$$f(y) = \begin{cases} \frac{1}{b-a} & a \leq y \leq b \\ 0 & \text{otherwise} \end{cases}$$

Y a *uniform random variable*, which we can write as $Y \sim \text{Uniform}(a, b)$.

Let's do an example.

Example. A circle has a radius which is uniformly distributed on the interval $[0, 1]$. What is the expected value of the area of the circle?

Let $Y \sim \text{Uniform}(0, 1)$. Then Y has density

$$f(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The area of a circle of radius y is given by $g(y) = \pi y^2$. Then the expected value of the area

is:

$$\begin{aligned}
\mathbb{E}[g(Y)] &= \int_{-\infty}^{\infty} g(y)f(y)dy \\
&= \int_0^1 \pi y^2 dy \\
&= \pi \frac{y^3}{3} \Big|_0^1 \\
&= \frac{\pi}{3}
\end{aligned}$$

As with every other distribution, we are interested in the mean and the variance of the uniform distribution. These are given below.

Properties of the uniform distribution

Let Y have the *uniform distribution* on the interval $[a, b]$. Then

$$\begin{aligned}
\mathbb{E}(Y) &= \frac{a+b}{2} \\
Var(Y) &= \frac{(b-a)^2}{12}
\end{aligned}$$

It makes sense that the mean of the uniform distribution is halfway between the endpoints. To verify this, for $Y \sim \text{Uniform}(a, b)$ with density $f(y)$ as given above:

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} yf(y)dy \\
&= \int_a^b y \frac{1}{b-a} dy \\
&= \frac{1}{b-a} \frac{y^2}{2} \Big|_a^b \\
&= \frac{b^2 - a^2}{2(b-a)} \\
&= \frac{(b+a)(b-a)}{2(b-a)} \\
&= \frac{a+b}{2}
\end{aligned}$$

The variance can be found similarly by computing $\mathbb{E}(Y^2)$ and using the Magic Variance Formula.

3.7 Normal Distribution

The most important and most useful continuous probability distribution is the normal distribution, also known as the Gaussian distribution or the “bell curve” for its familiar bell-like

shape. Many observations from nature are approximately normally distributed, and we will see later that in the appropriate limit, just about everything has a normal distribution. In particular, we will see that for large enough n , we can approximate a binomial random variable with a normal distribution; this is especially useful since computations with the normal distribution are often easier than dealing with the pesky factorials in the binomial pmf.

Examples of the normal distribution from the natural (or artificial) world include:

1. Weights of chimpanzees (and just about any other species) follow the normal distribution.
2. Errors in scientific measurement (due to imperfect instrument and imperfect scientists) are normally distributed.
3. Scores on the SAT (and other standardized tests) are normally distributed (in this case by design).

We define a normal distribution by its density function.

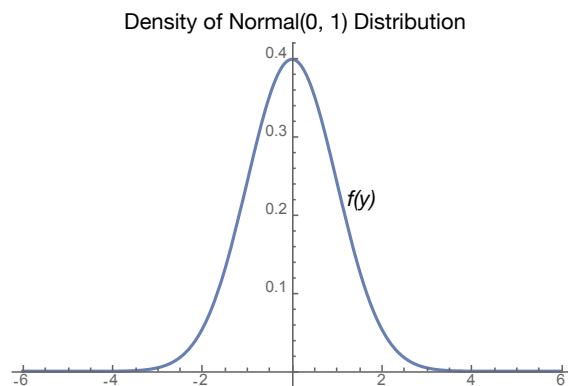
Normal distribution

A continuous random variable Y has a *normal distribution* with parameters μ and σ if its density $f(y)$ is given by

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Y is a *normal random variable*, denoted $Y \sim \text{Normal}(\mu, \sigma)$. Sometimes σ^2 is given as the parameter instead of σ .

A graph of the pdf for the normal distribution with parameters $\mu = 0$ and $\sigma = 1$ is shown below.



The parameters of the normal distribution, μ and σ , are the mean and standard deviation of the normal distribution.

Properties of the normal distribution

Let Y be a normal random variable with parameters μ and σ . Then

$$\begin{aligned}\mathbb{E}(Y) &= \mu \\ \text{Var}(Y) &= \sigma^2\end{aligned}$$

σ is the standard deviation of Y .

The fact that this is a valid pdf, i.e. that it integrates to 1, is a standard exercise in multivariable calculus⁵. The normal distribution with $\mu = 0$ and $\sigma = 1$ is so prevalent that it is called the *standard normal distribution*. It is represented by the letter Z .

Standard normal distribution

The *standard normal distribution* is a normal distribution with parameters $\mu = 0$ and $\sigma = 1$ and is designated by the letter Z . The pdf of Z is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Let Z be a standard normal random variable, and suppose we wish to calculate $\mathbb{P}(-1 \leq Z \leq 1)$. Using the density of the standard normal:

$$\mathbb{P}(-1 \leq Z \leq 1) = \int_{-1}^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Unfortunately, there is no nice antiderivative for the integrand, so we cannot compute the integral using the fundamental theorem of calculus. One option is to use numerical integration techniques, which can be done using software packages such as Matlab or Mathematica. Another option is to use tables for the cdf of the standard normal distribution. Although this is perhaps a bit “old-school”, it is important you know how to use these tables. There are many versions of the Z -table. The one I will provide (and the one which is on the course website) is the actual cdf for Z , i.e. it gives values for $F(z) = \mathbb{P}(Z \leq z)$. Since the standard normal is symmetric about 0, some tables only provide values on one side of the mean, since the others can be computed using symmetry. The table in the textbook by Wackerly et al (2008), for example, provides $\mathbb{P}(Z \geq z)$ for $z \geq 0$.

How do we compute this using a Z -table? Letting $F(z)$ be the cdf for the standard normal distribution, recall that for any continuous probability distribution we have:

$$\mathbb{P}(a \leq Z \leq b) = F(b) - F(a)$$

⁵The usual method is to square the integral, combine the integrands into a double integral, and change from cartesian to polar coordinates.

Then in this case, $\mathbb{P}(-1 \leq Z \leq 1) = F(1) - F(-1)$. Looking at the Z -table⁶, we see that $F(1) = 0.8413$ and $F(-1) = 0.1587$. Subtracting, we obtain $\mathbb{P}(-1 \leq Z \leq 1) = 0.6826$.

Let's do an example so we can practice using a Z -table.

Example. Let Z be a standard normal random variable, and let $F(z)$ be its cdf.

1. Find $\mathbb{P}(Z > 2)$.

Since $\mathbb{P}(Z > 2) = 1 - \mathbb{P}(Z \leq 2) = 1 - F(2)$, we just need to find $F(2)$. Looking this up in the Z -table, we see that $F(2) = 0.9772$, thus $\mathbb{P}(Z > 2) = 1 - 0.9772 = 0.0228$.

2. Find $\mathbb{P}(-2 \leq Z \leq 2)$

$\mathbb{P}(-2 \leq Z \leq 2) = F(2) - F(-2)$. From the Z -table, $F(2) = 0.9772$ and $F(-2) = 0.0228$. Thus we have $\mathbb{P}(-2 \leq Z \leq 2) = 0.9772 - 0.0228 = 0.9544$.

3. Find $\mathbb{P}(0 \leq Z \leq 1.73)$.

$\mathbb{P}(0 \leq Z \leq 1.73) = F(1.73) - F(0)$. From the Z -table, $F(1.73) = 0.9582$ (go to the row labeled 1.7, then across to the column corresponding to 0.03). We could use the table for $F(0)$, but since the distribution is symmetric about 0, we know $F(0) = 0.5$. Subtracting, we get $\mathbb{P}(0 \leq Z \leq 1.73) = 0.9582 - 0.5 = 0.4582$.

Before we go on, let us comment on two of the probabilities we already calculated. Recall the standard deviation of the standard normal random variable is 1. Then $\mathbb{P}(-1 \leq Z \leq 1)$ is the probability of falling within one standard deviation of the mean and $\mathbb{P}(-2 \leq Z \leq 2)$ is the probability of falling within two standard deviations of the mean. These are useful numbers to remember, since they are good guidelines for interpreting the normal distribution.

Guidelines for normal probabilities

Let Y be a random variable with a normal distribution. Then:

1. The probability of falling within 1 standard deviation of the mean is about 0.68
2. The probability of falling within 2 standard deviations of the mean is about 0.95
3. The probability of falling within 3 standard deviations of the mean is about 0.997

This is known as the *68-95-99.7 rule*

What do we do when we don't have a normal random variable. We can transform *any* random variable $Y \sim \text{Normal}(\mu, \sigma)$ into a standard normal random variable Z using the following formula:

$$Z = \frac{Y - \mu}{\sigma}$$

⁶The value of $F(z)$ is denoted by z in our table. The column on the left gives us the first decimal place and tells us what row to use. Then we read across the row to match the second decimal place

In other words, we subtract the mean and divide by the standard deviation.

Example. A machine produces ball bearings which diameters which are normally distributed with mean 3.0005 cm and standard deviation 0.0010 cm. Specifications require ball bearing diameters which lie in the interval 3.0000 ± 0.0020 cm. What fraction of the total production meets those specifications.

Let Y be the diameter of a ball bearing produced by the machine. Then $Y \sim \text{Normal}(3.0005, 0.0010)$. We want the probability that Y is within the required range, i.e. $\mathbb{P}(2.9980 \leq Y \leq 3.0020)$. To do this, we convert this to an interval involving the standard normal random variable Z .

$$\begin{aligned} y = 2.9980 & & z = \frac{2.9980 - 3.0005}{0.0010} = -2.5 \\ y = 3.0020 & & z = \frac{3.0020 - 3.0005}{0.0010} = 1.5 \end{aligned}$$

So $\mathbb{P}(2.9980 \leq Y \leq 3.0020) = \mathbb{P}(-2.5 \leq Z \leq 1.5)$. Looking at the Z -table, we find that $F(-2.5) = 0.0062$ and $F(1.5) = 0.9332$. Thus:

$$\mathbb{P}(2.9980 \leq Y \leq 3.0020) = \mathbb{P}(-2.5 \leq Z \leq 1.5) = F(1.5) - F(-2.5) = 0.9332 - 0.0062 = 0.9270$$

So approximately 92.7% of the ball bearings meet the required specifications.

3.8 Exponential Distribution

The final continuous distribution we will discuss is the exponential distribution. The exponential distribution belongs to the family of gamma distributions, and is perhaps the most useful member of that family; it is the only gamma distribution we will consider in this class.

The exponential distribution is used to model the length of time between events which occur independently and at a constant average rate. For example, it could be used the model the length of time between customer arrivals at a restaurant or phone calls at a call center. If this reminds you of a Poisson distribution, that is good! If we have a sequence of events which occur independently from each other and at a constant average rate, then:

1. The Poisson distribution is a discrete distribution which measures the number of events which occur in a fixed span of time.
2. The exponential distribution is a continuous distribution which measures the amount of time between two subsequent events.

Thus, the Poisson distribution and the exponential distribution ask different questions about the same problem.

The exponential distribution is also used to model the lifetime of electronic and mechanical components. Recall that we used the geometric distribution in a similar way; we did an example where the number of hours until a computer crashed was modeled as a geometric random variable. If we want to model the lifetime as a continuous random variable instead

of a discrete one, then we use the exponential distribution. The exponential distribution has another similarity to the geometric distribution; as we shall see, it is also memoryless. In fact, the geometric and exponential distributions are the only two probability distributions with this property.

Let's define the exponential distribution, then use it to model some problems.

Exponential distribution

A continuous random variable Y has an exponential distribution with parameter $\lambda > 0$ if its density function is:

$$f(y) = \begin{cases} \lambda e^{-\lambda y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

Y is an exponential random variable with parameter λ , which is denoted $Y \sim \text{Exponential}(\lambda)$.

You will sometimes see the exponential density written as $(1/\beta)e^{-y/\beta}$. I prefer the above version, since the parameter λ is the same as that in the Poisson distribution, i.e. the average rate at which the events occur.

As always, we first verify that the exponential distribution is a valid probability density.

$$\begin{aligned} \int_{-\infty}^{\infty} f(y) dy &= \int_0^{\infty} \lambda e^{-\lambda y} dy \\ &= \lambda \lim_{t \rightarrow \infty} \int_0^t e^{-\lambda y} dy \\ &= \lambda \left(-\frac{1}{\lambda} \right) \lim_{t \rightarrow \infty} e^{-\lambda y} \Big|_0^t \\ &= - \left(\lim_{t \rightarrow \infty} e^{-\lambda t} - 1 \right) \\ &= 1 \end{aligned}$$

since the limit of the exponential in the second-to-last line above is 0.

Next, we find the expected value and variance of the exponential distribution.

Properties of the exponential distribution

Let Y be an exponential random variable with parameter $\lambda > 0$. Then

$$\begin{aligned}\mathbb{E}(Y) &= \frac{1}{\lambda} \\ \text{Var}(Y) &= \frac{1}{\lambda^2}\end{aligned}$$

For the expected value, if $Y \sim \text{Exponential}(\lambda)$ has density $f(y)$, then

$$\begin{aligned}\mathbb{E}(Y) &= \int_{-\infty}^{\infty} y f(y) dy \\ &= \int_0^{\infty} \lambda y e^{-\lambda y} dy \\ &= \lambda \lim_{t \rightarrow \infty} \int_0^t \lambda y e^{-\lambda y} dy \\ &= \lambda \lim_{t \rightarrow \infty} \left(-\frac{1}{\lambda} y e^{-\lambda y} \Big|_0^t + \frac{1}{\lambda} \int_0^t e^{-\lambda y} dy \right) && \text{integration by parts} \\ &= -\lim_{t \rightarrow \infty} t e^{-\lambda t} + 0 + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda y} dy \\ &= \frac{1}{\lambda}\end{aligned}$$

where in the second-to-last line the limit is 0 (exponentials grow faster than any power) and the integral is 1 (since it's the integral of the exponential density). Similarly, using the Magic Variance Formula and integrating by parts twice, we can prove the variance of an exponential random variable.

Example. While procrastinating studying for APMA 1650, you decide to create a website for your cat. Suppose your cat's website gets an average of 5 unique visits per hour. What is the probability that your cat will go more than 30 minutes without a visitor?

We can model the number of visits to your cat's website per hour as a Poisson distribution with parameter $\lambda = 5$ and the time between visits as an exponential distribution with the same parameter $\lambda = 5$. In this case, we are interested in the time between visits. Let $X \sim \text{Exponential}(5)$. Then X has density $f(x) = 5e^{-5x}$. Then since we are working in units of hours, the probability we want is $\mathbb{P}(X \geq 1/2)$.

$$\begin{aligned}\mathbb{P}(X \geq 1/2) &= \int_{1/2}^{\infty} 5e^{-5x} dx \\ &= -\frac{5}{5} e^{-5x} \Big|_{1/2}^{\infty} \\ &= e^{-5/2} \approx 0.082\end{aligned}$$

Just as the geometric distribution is memoryless, the exponential distribution is memoryless. If we think of this in terms of the lifetime of, say, a light bulb, this means that the probability of the bulb burning out in the next 30 minutes is the same whether we just turned the light bulb on or whether it has been on for 5 days. To prove the memoryless property of the exponential distribution, it will be useful to have an expression for the cdf of the exponential distribution.

Cumulative distribution function (cdf) for the exponential distribution

Let Y be an exponential random variable with parameter $\lambda > 0$. Then its cdf $F(y)$ is given by:

$$F(y) = \begin{cases} 1 - e^{-\lambda y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

To see this, we just integrate the density function. Let $Y \sim \text{Exponential}(\lambda)$, and let $f(y)$ be its density. For $Y \leq 0$, the cdf $F(y) = 0$. (Why is this the case?) For $y > 0$, we use the definition of the cdf to get:

$$\begin{aligned} F(y) &= \mathbb{P}(Y \leq y) \\ &= \int_{-\infty}^y f(t) dt \\ &= \int_{0,y} \lambda e^{-\lambda t} dt \\ &= -e^{-\lambda t} \Big|_0^y \\ &= 1 - e^{-\lambda y} \end{aligned}$$

We can use this to show the memoryless property of the exponential distribution.

Memoryless property of the exponential distribution

Let Y be a geometric random variable with parameter λ . Then for all a and b

$$\mathbb{P}(Y > a + b | Y > a) = \mathbb{P}(Y > b)$$

To see this, we first use the exponential cdf $F(y)$ to see that

$$\mathbb{P}(Y > y) = 1 - \mathbb{P}(Y \leq y) = 1 - F(y) = 1 - (1 - e^{-\lambda y}) = e^{-\lambda y}$$

Using this and the definition of conditional probability:

$$\begin{aligned}
\mathbb{P}(Y > a + b | Y > a) &= \frac{\mathbb{P}(Y > a + b \cap Y > a)}{\mathbb{P}(Y > a)} \\
&= \frac{\mathbb{P}(Y > a + b)}{\mathbb{P}(Y > a)} \\
&= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\
&= e^{-\lambda b} \\
&= \mathbb{P}(Y > b)
\end{aligned}$$

We can also use the cdf of the exponential distribution to find the median. To do this we solve $F(m) = 1/2$, where $F(y)$ is the exponential cdf.

$$\begin{aligned}
F(m) &= \frac{1}{2} \\
1 - e^{-\lambda m} &= \frac{1}{2} \\
e^{-\lambda m} &= \frac{1}{2} \\
-\lambda m &= \log\left(\frac{1}{2}\right) = -\log 2 \\
m &= \frac{\log 2}{\lambda}
\end{aligned}$$

The median of the exponential distribution is thus a little smaller than the mean (by a factor of $\log 2 \approx 0.693$).

3.9 Bounds on Probabilities

Sometimes what we are interested in is the probability of an outlier event, i.e. the probability that a random variable deviates from the mean by more than a certain amount. Of course, if we know the exact probability distribution, this is not hard to calculate. However, even if we don't know the probability distribution, we can still obtain upper bounds on the probability of outlier events. In this section we will look at two inequalities which allow us to do exactly this. Markov's Inequality can be used if only the mean is known, but it gives us the weakest bound. Chebyshev's Inequality is a better bound, but required knowledge of both the mean and the variance. The moral of the story is: the more information we have about a probability distribution, the better estimates we can get.

3.9.1 Markov's Inequality

We first look at Markov's Inequality, which gives the probability that a nonnegative random variable exceeds a certain threshold in terms of the expected value of the random variable. To use Markov's Inequality, we only need to know the expected value.

Markov's Inequality

Let Y be a nonnegative random variable with expected value $E(Y)$. If $a > 0$, then

$$\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}(Y)}{a}$$

To see that this is true, let I_a be the indicator random variable for the event $(Y \geq a)$, i.e. define I_a by

$$I_a = \begin{cases} 1 & Y \geq a \\ 0 & \text{otherwise, i.e. if } Y < a \end{cases}$$

Next note that $aI_a \leq Y$. To see this is true, if $Y < a$, then $aI_a = 0 \leq Y$ since Y is nonnegative (this is why we require Y to be nonnegative). If $Y \geq a$, then $aI_a = a \leq Y$. Thus we have:

$$\begin{aligned} \mathbb{E}(aI_a) &\leq E(Y) \\ a[1 \cdot \mathbb{P}(Y \geq a) + 0 \cdot \mathbb{P}(Y < a)] &\leq E(Y) \\ a\mathbb{P}(Y \geq a) &\leq E(Y) \\ \mathbb{P}(Y \geq a) &\leq \frac{E(Y)}{a} \end{aligned}$$

where in the second line we used the definition of the expected value of a discrete random variable.

Example. Suppose we randomly select an article from a journal where the mean article length is known to be 1000 words. Find an upper bound on the probability that an article exceeds 1400 words in length.

Let Y be the length of an article in this journal. We know nothing about Y except that its mean is 1000. Certainly Y is nonnegative, thus we can find an upper bound for $\mathbb{P}(Y \geq 1400)$ using Markov's inequality:

$$\mathbb{P}(Y \geq 1400) \leq \frac{E(y)}{1400} = \frac{1000}{1400} \approx 0.71$$

This is not a great upper bound, but it's better than nothing!

3.9.2 Chebyshev's Inequality

The second inequality we will look at is Chebyshev's Inequality, which gives the probability that a random variable deviates from its mean by more than a certain amount. This does not require a nonnegative random variable, but it does require knowledge of both the mean and the variance of a random variable.

Chebyshev's Inequality

Let Y be a nonnegative random variable with expected value $\mathbb{E}(Y)$ and variance $Var(Y)$. If $a > 0$, then

$$\mathbb{P}(|Y - E(Y)| \geq a) \leq \frac{Var(Y)}{a^2}$$

Note the absolute value sign inside the probability for Chebyshev's Inequality. This means that Chebyshev's Inequality gives a bound for the probability of deviating from the mean *in either direction*. Contrast this to Markov's Inequality, which is a bound on the probability of *exceeding* a certain value.

To see this is true, take the nonnegative random variable $(Y - \mathbb{E}(Y))^2$ and substitute it into Markov's Inequality along with the constant a^2 . This gives us:

$$\mathbb{P}([Y - \mathbb{E}(Y)]^2 \geq a^2) \leq \frac{E[(Y - E(Y))^2]}{a^2} = \frac{Var(Y)}{a^2}$$

where the last equality uses the definition of variance. Then since $|x| \geq a$ if and only if $x^2 \geq a^2$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}([X - \mathbb{E}(X)]^2 \geq a^2) \leq \frac{Var(Y)}{a^2}$$

which is Chebyshev's Inequality.

Sometimes we like to measure deviation from the mean in terms of “numbers of standard deviations”. For the normal distribution, this is encapsulated in the 68-95-99.7 rule. We can state Chebyshev's Inequality in these terms if we like. Let Y be a random variable with mean μ and variance σ^2 . Recall that the standard deviation is the square root of the variance, so the standard deviation of Y is σ . Then Chebyshev's Inequality states that the probability of deviating at least k standard deviations from the mean is bounded by:

$$\mathbb{P}(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

To see this is true, just take $a = k\sigma$ in Chebyshev's Inequality above.

Example. Suppose we randomly select an article from a journal article length is distributed with a mean 1000 words and a standard deviation of 150 words. Find an upper bound on the probability that an article is outside of the range 600 - 1400 words?

Let Y be the length of an article in this journal. Here we know both the mean and the standard deviation of Y , we can use Chebyshev's Inequality. We are looking for the probability that Y deviates from the mean by at least 400. Recalling that the variance is the square of the standard deviation:

$$\mathbb{P}(Y \geq 1400 \cup Y \leq 600) = \mathbb{P}(|Y - 1000| \geq 400) \leq \frac{Var(Y)}{400^2} = \frac{150^2}{400^2} \approx 0.14$$

Let's compare this to the bound we got in the previous example using Markov's Inequality. Since $(Y \geq 1400) \subset (Y \geq 1400 \cup Y \leq 600)$, using the properties of probability:

$$\mathbb{P}(Y \geq 1400) \leq \mathbb{P}(Y \geq 1400 \cup Y \leq 600) \leq 0.14$$

This is a much better bound than we got from Markov's Inequality. If we have reason to suspect that the distribution of Y is symmetric about the mean, we can divide $\mathbb{P}(Y \geq 1400 \cup Y \leq 600)$ by 2 to get:

$$\mathbb{P}(Y \geq 1400) \leq \frac{\mathbb{P}(Y \geq 1400 \cup Y \leq 600)}{2} \leq \frac{0.14}{2} = 0.07$$

which is even better! It is important that we can only do this if we have (or suspect we have) a symmetric probability distribution. This is not true in general.

Just for comparison purposes, let's see how much better this bound is if we know the exact distribution of Y . Suppose Y is normally distributed with the same parameters, i.e. $Y \sim \text{Normal}(1000, 150)$. Then, standardizing to the standard normal random variable Z with cdf $F(y)$:

$$\begin{aligned} \mathbb{P}(600 \leq Y \leq 1400) &= \mathbb{P}\left(\frac{600 - 1000}{150} \leq Z \leq \frac{1400 - 1000}{150}\right) \\ &= \mathbb{P}(-2.67 \leq Z \leq 2.67) \\ &= F(2.67) - F(-2.67) \\ &= 0.9962 - 0.0038 = 0.9924 \end{aligned}$$

Thus we have:

$$\mathbb{P}(Y \geq 1400 \cup Y \leq 600) = 1 - \mathbb{P}(600 \leq Y \leq 1400) \leq 0.0076$$

Although Chebyshev's Inequality gives a decent bound on the probability of outliers, there is no substitute for knowing the actual probability distribution!