

1.9 Applications of Logarithms

The properties of logarithms are useful for a variety of applications. In this section, we discuss using a logarithm to transform data. We will see that data following an exponential model look linear in a semi-log transformation; data following a power law model look linear in a log-log transformation. We also consider an application to probability in relation to log-likelihood.

1.9.1 Logarithmic Transformations

Sometimes we look at data that are at many different scales. On a standard number line, we think of the numbers 1, 10, 100, and 1000 as spread very far apart. However, if we consider looking at a usual number line that will include all of these values, the relative space between 1 and 10 is very small and both seem very close to zero.

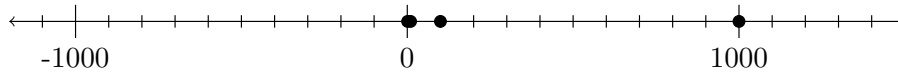


Figure 1.9.1 A number line showing the numbers 1, 10, 100, and 1000.

In a similar way, we might normally think of the numbers 0.1, 0.01, 0.001, and 0.0001 as all very close to 0. However, each value is a different order of magnitude exactly like the values 1, 10, 100, and 1000. The common logarithm is the logarithm for base ten ($b = 10$). Consequently, the common logarithm of these numbers would give us equally spaced integers.

$$\begin{aligned}
 0.0001 = 10^{-4} &\Leftrightarrow \log 0.0001 = -4 \\
 0.001 = 10^{-3} &\Leftrightarrow \log 0.001 = -3 \\
 0.01 = 10^{-2} &\Leftrightarrow \log 0.01 = -2 \\
 0.1 = 10^{-1} &\Leftrightarrow \log 0.1 = -1 \\
 1 = 10^0 &\Leftrightarrow \log 1 = 0 \\
 10 = 10^1 &\Leftrightarrow \log 10 = 1 \\
 100 = 10^2 &\Leftrightarrow \log 100 = 2 \\
 1000 = 10^3 &\Leftrightarrow \log 1000 = 3
 \end{aligned}$$

Because the logarithm spaces values apart according to the order of magnitude, we can often use logarithms to visualize data that occur at multiple orders of magnitude. Quality plotting tools allow us to plot the data but use the logarithm for their position in the graph. This is called using a logarithmic scale. We can choose to use a logarithmic scale for one or both axes.

Example 1.9.2 Brain size is strongly correlated with overall body mass in mammals. However, mammals cover a wide range of different sizes. The graph of brain size versus body mass for 96 species is shown in [Figure 1.9.3](#), based on data from *The Statistical Sleuth* by Ramsey and Schafer (2013). Because the elephant is so large relative to many other species, its data point requires a wide window in the figure. The majority of species, however, are much smaller and form a crowded cluster of points near the origin.

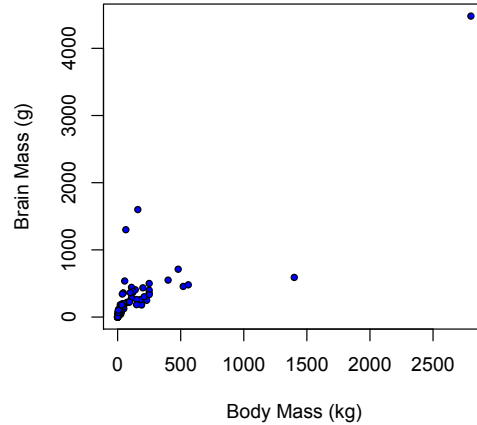


Figure 1.9.3 Plot of body mass (kg) and brain size (g) for 96 species of mammals.

This suggests replotting the data using a logarithmic scale. The same data is shown with logarithmic scales for both variables in the [Figure 1.9.4](#). Using a logarithmic scale on both axes is called a **log-log plot**. The transformed data spreads the points out more uniformly across the figure. In addition, the log-log plot suggests that the transformed data is approximately linear.

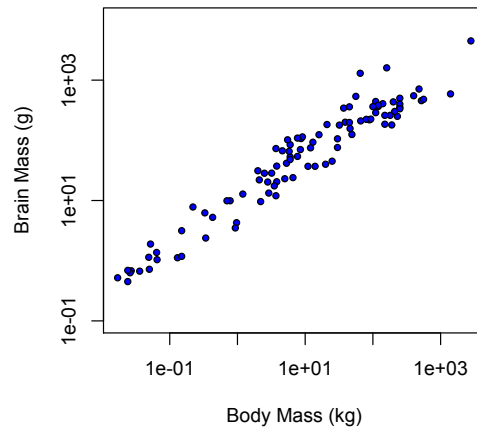


Figure 1.9.4 Log-log plot of body mass (kg) and brain size (g) for 96 species of mammals.

□

The previous example illustrated a dataset where transformed data look linear. Let us work out what that relation must be like.

Suppose we have raw data with variables (x, y) and we transform the data with logarithms. This creates two new variables, $u = \log(x)$ and $v = \log(y)$. The log-log plot is a figure showing data (u, v) but with the axes showing the original values on a logarithmic scale. If the transformed data are linear, there must be a model

$$v = au + b.$$

We now substitute our original variables and solve for y . We collect terms

in the logarithm.

$$\begin{aligned}\log(y) &= a \log(x) + b \\ &= \log(x^a) + \log(10^b) \\ &= \log(10^b \cdot x^a)\end{aligned}$$

Thus, we find $y = 10^b \cdot x^a$, which is a power law model. We summarize our result as a theorem for future reference.

Theorem 1.9.5 *Data (x, y) such that the transformed data $(\log(x), \log(y))$ (a log-log plot) has a linear relation will satisfy a power law relation.*

Another common transformation is a semi-log plot. This occurs when only the dependent variable is transformed. In other words, only the y -axis is transformed to a logarithmic scale. What relationship does this reveal?

Suppose we have raw data with variables (x, y) and we only transform y with a logarithm,

$$v = \log(y).$$

The semi-log plot is a figure showing data (x, v) but with the axes showing the original values on a logarithmic scale. If the transformed data are linear, there must be a model

$$v = ax + b.$$

We now substitute our original variables and solve for y .

$$\log(y) = ax + b.$$

To solve for y , we use the inverse operation to the logarithm, which is an exponential.

$$y = 10^{ax+b}.$$

Using the properties of exponents, we can rewrite this

$$y = 10^{ax} \cdot 10^b = 10^b \cdot (10^a)^x.$$

Thus, we find $y = AB^x$, with $A = 10^b$ and $B = 10^a$, which is an exponential model.

Theorem 1.9.6 *Data (x, y) such that the transformed data $(x, \log(y))$ (a semi-log plot) has a linear relation will satisfy an exponential relation.*

We can use the log-transformations to find the power law and exponential relations for actual data. If we know that (x, y) satisfies a power law for given data, then we know $(\log(x), \log(y))$ satisfies a linear model. We can calculate the slope and intercept of the transformed linear model and then solve for y . If we know that (x, y) satisfies an exponential model for given data, then we can find the equation of a line for $(x, \log(y))$ and then solve for y .

Example 1.9.7 Find the power law for (x, y) that includes data $(x, y) = (2, 5)$ and $(4, 8)$.

Solution. Power law data is linear under a log-log transformation, $u = \log(x)$ and $v = \log(y)$. The transformed points are $(u, v) = (\log x, \log y) = (\log(2), \log(5))$ and $(u, v) = (\log x, \log y) = (\log(4), \log(8))$. The slope is calculated and simplified using properties of logarithms,

$$\begin{aligned}m &= \frac{\Delta v}{\Delta u} = \frac{\log(8) - \log(5)}{\log(4) - \log(2)} \\ &= \frac{\log(8/5)}{\log(4/2)}\end{aligned}$$

$$= \frac{\log(8/5)}{\log(2)}$$

Using the point-slope form of a line, we have

$$\begin{aligned} v - \log(5) &= m(u - \log(2)) \\ &= \frac{\log(8/5)}{\log(2)}(u - \log(2)) \end{aligned}$$

Now we substitute back the original variables with $u = \log(x)$ and $v = \log(y)$. Alternatively, we could have done the work above using $\log x$ and $\log y$ in place of u and v . Our final equation then would be written

$$\log(y) - \log(5) = \frac{\log(8/5)}{\log(2)}(\log(x) - \log(2)).$$

We now proceed to apply the rules for logarithms to simplify our expression.

Our goal is to have a logarithm on the left equaling a logarithm on the right. The first step is to use the quotient rule of logarithms:

$$\log(y/5) = \frac{\log(8/5)}{\log(2)} \log(x/2).$$

On the right, we have a product of values. A linear log-log plot corresponds to a power law, and we ultimately want our equation to have x raised to a power. Because x currently appears within the logarithm, we will use the power rule for logarithms. The logarithm with x , $\log(x/2)$, is multiplied by an expression. That expression, which originally served as our slope, will become the power.

For simplicity in writing, we introduce a new symbol, $p = \frac{\log(8/5)}{\log(2)}$.

$$\begin{aligned} \log(y/5) &= p \cdot \log\left(\frac{x}{2}\right) \\ \log(y/5) &= \log\left(\frac{x}{2}\right)^p \end{aligned}$$

Now that we have the logarithm of an expression on the left and on the right, the expressions within the logarithms must be equal.

$$\begin{aligned} \frac{y}{5} &= \left(\frac{x}{2}\right)^p \\ y &= 5 \cdot \left(\frac{x}{2}\right)^p \end{aligned}$$

This equation is our model for the power law relation. \square

We now illustrate the similar process for a semi-log transformation. Exponentially related data will appear linear on a semi-log plot.

Example 1.9.8 Find the exponential law for (x, y) that includes data $(x, y) = (2, 4)$ and $(5, 10)$.

Solution. Exponential law data is linear under a semi-log transformation, $u = x$ and $v = \log(y)$. The transformed points are $(u, v) = (x, \log y) = (2, \log(4))$ and $(u, v) = (x, \log y) = (5, \log(10))$. The slope is calculated and simplified using properties of logarithms,

$$m = \frac{\Delta v}{\Delta u} = \frac{\log(10) - \log(4)}{5 - 2}$$

$$\begin{aligned}
 &= \frac{\log(10/4)}{3} \\
 &= \frac{\log(5/2)}{3}
 \end{aligned}$$

Using the point-slope form of a line, we can create our equation relating our variables.

$$\begin{aligned}
 v - \log(4) &= m(u - 2) \\
 \log y - \log 4 &= \frac{\log(5/2)}{3}(x - 2)
 \end{aligned}$$

We now seek to find an equivalent equation where a logarithm of an expression appears alone on each side of the equation. We first apply the quotient rule for logarithms on the left.

$$\begin{aligned}
 \log y - \log 4 &= \frac{\log(5/2)}{3}(x - 2) \\
 \log(y/4) &= \frac{\log(5/2)}{3}(x - 2)
 \end{aligned}$$

Because data that are linear in a semi-log plot have an exponential relation, we want to see how to get x into the exponent. On the right-hand side, we have a logarithm, $\log(5/2)$, that is multiplied by $(x - 2)$ and divided by 3. We can group those together as a single multiplication and then apply the power rule for logarithms.

$$\begin{aligned}
 \log(y/4) &= \frac{x - 2}{3} \cdot \log(5/2) \\
 \log(y/4) &= \log\left((5/2)^{(x-2)/3}\right)
 \end{aligned}$$

We can now eliminate the logarithm from both sides of the equation.

$$\begin{aligned}
 \frac{y}{4} &= (5/2)^{(x-2)/3} \\
 y &= 4 \cdot (5/2)^{(x-2)/3}
 \end{aligned}$$

□

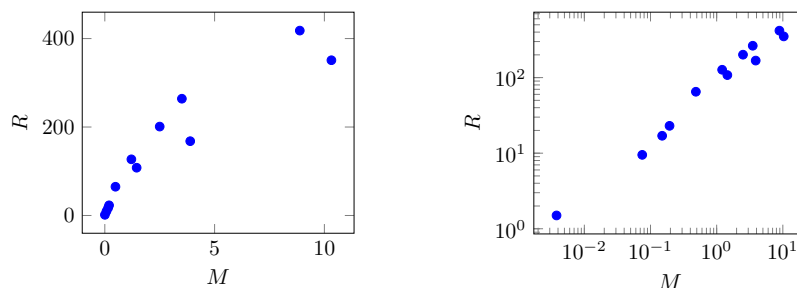
We end with an example of how this might relate to actual data.

Example 1.9.9 In 1967, Lasiewski and Dawson published an article in *The Condor* relating the body mass M (in kg) and resting metabolic rate R (in kcal/day) for birds. They tabulated the recorded body mass and metabolic rate for individual birds based on published studies. The following table includes twelve of these birds. Graph the data and determine if the data appear linear in a log-log plot. Then use a linear regression of transformed data to estimate the model

Table 1.9.10 Selected body mass and resting metabolic rate of birds, as tabulated in Lasiewski and Dawson (1967).

Bird	M (kg)	R (kcal/day)
Rufous hummingbird	0.0038	1.5
Common nighthawk	0.075	9.5
Common wood pigeon	0.150	17.0
Northern bobwhite	0.194	23.0
Wood duck	0.485	65
Pacific gull	1.21	127
Great horned owl	1.450	108
Wood stork	2.5	201
Brown pelican	3.51	264
Sandhill crane	3.89	168
Trumpeter swan	8.88	418
Andean condor	10.32	351

Solution. After entering the data into a spreadsheet or other graphing utility, we generate a scatterplot of the points (M, R) , as shown below on the left. You should see that the relationship of the data is increasing and concave down. If we modify both axes to use a logarithmic scale, as shown below on the right, we see that the transformed relationship looks reasonably linear. This suggests using a linear relation on the transformed coordinates.



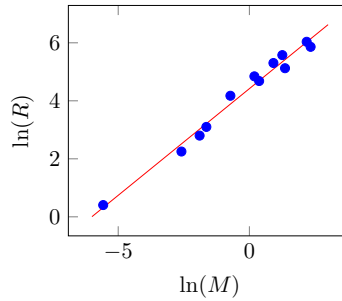
To form the linear model, we need to generate actual transformed values and not just a graph using logarithm scales. In the spreadsheet, we will create two new columns for $\ln(M)$ and for $\ln(R)$. For example, suppose the mass M of *Selasphorus rufus* appears in the spreadsheet in cell B2 and we want to generate the transformed variable $\ln(M)$ in cell D2. In cell D2, we would type the formula `=ln(B2)`. Copying this formula and pasting it into other cells will preserve the relative location. If you paste it into cell D3, you will discover it automatically changed the formula to `=ln(B3)`.

Once we have new columns $\ln(M)$ and $\ln(R)$, we can create a new scatterplot of data $(\ln(M), \ln(R))$ using *linear* scales. This new graph will have the same appearance as the original data using *logarithmic* scales, except that the axes show the logarithm of the data rather than the original data using logarithmic scales. With this new graph, we can find the linear trend line. The graph below shows the graph of the transformed data, along with a trend line using the formula

$$y = 0.7356x + 4.4192.$$

We change the variables to those plotted to give a transformed model

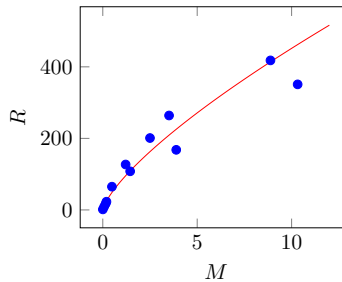
$$\ln(R) = 0.7356 \ln(M) + 4.4192.$$



We find the model of the relation for the original variables $M \mapsto R$ by solving for R and simplifying. To eliminate the logarithm, we apply the inverse operation of the natural exponential. Because our data are approximate, we can use decimal approximations for our formulas.

$$\begin{aligned}\ln(R) &= 0.7356 \ln(M) + 4.4192 \\ e^{\ln(R)} &= e^{0.7356 \ln(M) + 4.4192} \\ R &= e^{0.7356 \ln(M)} \cdot e^{4.4192} \\ &= e^{\ln(M^{0.7356})} \cdot 83.030 \\ &= 83.030 \cdot M^{0.7356}\end{aligned}$$

We see that the model is a power function, $R = 83.030M^{0.7356}$. The figure below shows the original data using linear axes along with this approximating model.



□

In practice, spreadsheets have built-in tools to accomplish the transformed calculations. When a spreadsheet application allows you to add a trend-line to a given plot, it often allows you to select a variety of different models. When it allows you to use an exponential model, the spreadsheet is internally using a semi-log transform, finding the linear trend-line, and then reporting back the resulting exponential model. Similarly, when a spreadsheet allows you to choose a power law model, the spreadsheet is internally finding the linear equation for a log-log transform and then reporting back the resulting power law model.

1.9.2 Enrichment: Log-Likelihood

Note 1.9.11 This section is included as an example of how logarithms play a more fundamental role in a more advanced sense than just transforming data. The content is optional. Subsequent sections do not rely on students having learned this material.

Suppose that we are performing an experiment that has a random outcome with two possibilities. We do not know in advance the probabilities associated with the two outcomes. For example, flipping a coin results in heads or tails.

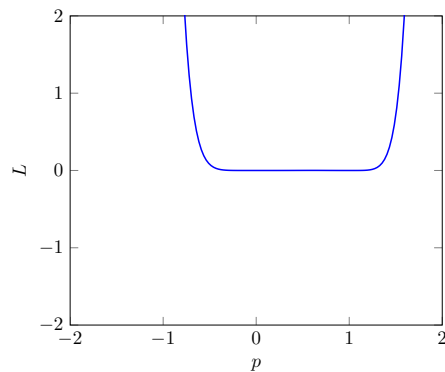
A fair coin has equal probabilities. A biased coin has unequal probabilities. If we suspected a coin was biased, we might want to determine the coin's true odds for heads versus tails. We would like to use repetition of an experiment in order to determine these probabilities.

In statistics, there is a method commonly used to estimate unknown parameters called the maximum likelihood principle. Each observation is assumed to have outcomes governed by a probability distribution characterized by certain model parameters. The likelihood L is the product of the probabilities densities associated with each observation. The maximum likelihood method adopts the parameter values that makes the likelihood as large as possible.

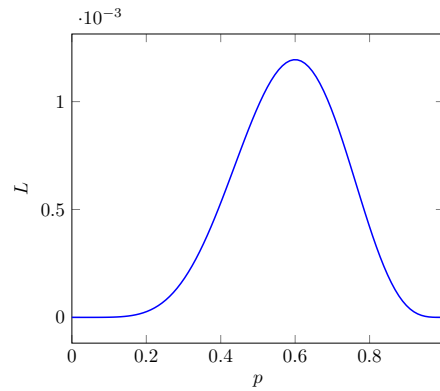
For our experiment with two different outcomes, the probability distribution is characterized by one parameter, p , which gives the probability of the first outcome (often called a success). The probability of the second outcome (often called a failure) will be $1 - p$ since probabilities must add to 1. Suppose that we repeated the experiment ten times and counted six successes and four failures. The likelihood is the product of the probabilities for these outcomes, using the expressions involving the parameter. The likelihood will be the product of six factors with p and four factors with $1 - p$. Writing these with powers, the likelihood is a function of p ,

$$L = p^6(1 - p)^4.$$

How will we maximize this value? Until we learn some calculus, we will need to find the maximum using a graph. The graph of this formula is shown below. (To make a graph, most graphing utilities require that you use the independent variable x in place of p .)



How do we interpret this graph? Because the parameter p is supposed to be a probability, we require $0 < p < 1$. But the graph doesn't seem to show a maximum there. This is because values of p outside the meaningful domain dominate the figure. If we redo the graph so that the domain only include $[0, 1]$, we get a better picture.



This graph has a maximum value at $p = 0.6$. We can also see why the earlier graph didn't show the maximum. The scale on the vertical axis for L for the restricted interval has an order of magnitude of 10^{-3} . If we had even more data than ten observations—and to estimate probabilities we need many more—this magnitude would be even smaller. Because of this effect that the likelihood shrinks in magnitude with more data, the likelihood value often drops below the smallest number a computer can represent. It would then be impossible to find the maximum likelihood parameter value.

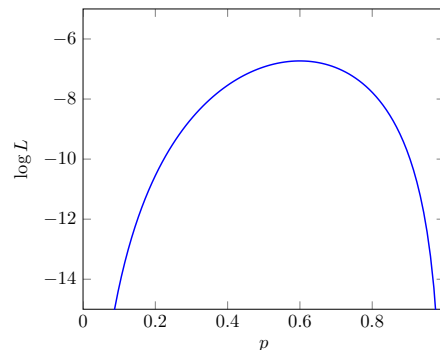
To avoid this issue, data scientists typically record the log-likelihood rather than the likelihood. Maximizing the log-likelihood will always give the same values as maximizing the likelihood itself. The log-likelihood is calculated as the natural logarithm of the likelihood,

$$\log L = \ln(L).$$

Because the logarithm of a product is equal to the sum of the logarithms of the factors, the log-likelihood is calculated by *adding* the logarithms of the probability densities corresponding to the observations. For our example,

$$\begin{aligned} \log L &= \ln(p^6 (1-p)^4) \\ &= \ln(p^6) + \ln((1-p)^4) \\ &= 6 \ln(p) + 4 \ln(1-p) \end{aligned}$$

A graph of the log-likelihood $\log L$ versus p is shown in the figure below. The maximum value again occurs at $p = 0.6$.



Example 1.9.12 An exponential time is a random time until some event occurs that is characterized by gaining no information by knowing how long has already passed without the event yet occurring. The time until a radioactive particle decays is an example of an exponential time. The mathematical model for the probability density of an exponential time t has a single parameter,

usually represented by the Greek letter lambda λ ,

$$f(t) = \lambda e^{-\lambda t}.$$

In a series of five experiments, the observed exponential times were recorded as $t_1 = 12.3$, $t_2 = 4.6$, $t_3 = 23.1$, $t_4 = 0.4$, and $t_5 = 10.5$. Calculate the log-likelihood for this collection of data, plot the log-likelihood, and determine the maximum likelihood value for the parameter λ .

Solution. The logarithm of the density is

$$\begin{aligned} \ln f(t) &= \ln(\lambda e^{-\lambda t}) \\ &= \ln(\lambda) + \ln(e^{-\lambda t}) \end{aligned}$$

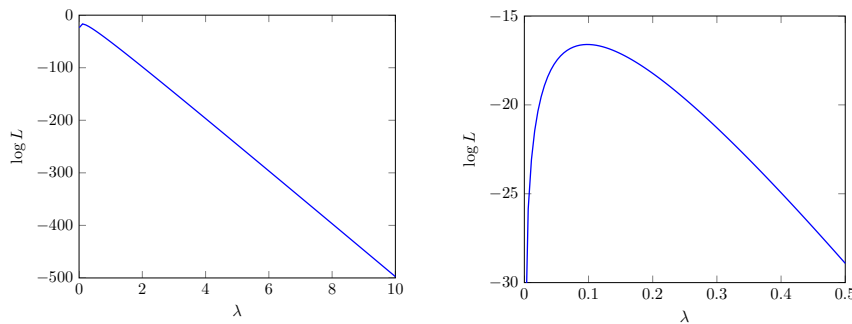
Because the natural logarithm and the exponential with the natural base e are inverses, we can simplify further to obtain

$$\ln f(t) = \ln(\lambda) - \lambda t.$$

The log-likelihood is the sum of the logarithms of the densities using the observed times. Each observation will result in adding $\ln(\lambda)$, so we obtain

$$\begin{aligned} \log L &= 5 \ln(\lambda) - \lambda(12.3 + 4.6 + 23.1 + 0.4 + 10.5) \\ &= 5 \ln(\lambda) - 50.9\lambda \end{aligned}$$

The parameter λ only needs to be a positive number. If we plot values $0 < \lambda < 10$ to explore where the maximum might be, we get the figure on the left. It shows the graph steadily decreasing, which means the maximum is close to zero. If we plot value $0 < \lambda < 0.5$, we get the figure on the right. The maximum value occurs at $\lambda = 0.098231$, which is our maximum likelihood estimate of the parameter.



□

1.9.3 Summary

- Transforming data with a logarithm allows us to view the distribution of data spread over a wide range of magnitudes.
- Data that appear linear in a log-log plot (both axes in logarithmic scale) follow a power law relation.
- Data that appear linear in a semi-log plot (only y -axis in logarithmic scale) follow an exponential relation.
- Estimating parameters for probability distributions is frequently based on maximum likelihood estimation. To avoid numerical underflow (expo-

nentially small magnitudes) of the likelihood, this is more common done using the log-likelihood.

1.9.4 Exercises

1. Suppose data for (t, M) appear linear in a semi-log plot. If the data include the points $(t, M) = (2, 5)$ and $(t, M) = (5, 2)$, find a linear model for the transformed data and use it to find the appropriate model for the original data.
2. Suppose data for (P, S) appear linear in a log-log plot. If the data include the points $(P, S) = (2, 5)$ and $(P, S) = (5, 2)$, find a linear model for the transformed data and use it to find the appropriate model for the original data.
3. A random experiment has two possible outcomes, high or low. The probability the result is high is represented by p , with $0 < p < 1$, and the probability the result is low is represented by $1 - p$. Twenty independent replicates of the experiment resulted in six highs and fourteen lows. Calculate the formula for the likelihood and use it to compute the log-likelihood. With a graph, estimate the maximum likelihood value for p .
4. An experiment results in randomly distributed exponential times. The probability density used in the likelihood has a single parameter λ ,

$$f(t) = \lambda e^{-\lambda t}.$$

Replicating the experiment six times results in measured times $t_1 = 0.826$, $t_2 = 0.293$, $t_3 = 0.218$, $t_4 = 0.024$, $t_5 = 0.561$, and $t_6 = 0.233$. Calculate the formula for the likelihood and use it to compute the log-likelihood. With a graph, estimate the maximum likelihood value for λ .

5. A manufacturer tracks quality control by testing random samples for proper performance. The number n of identified flaws is a random value that occurs with a probability

$$f(n) = a_n e^{-\lambda} \lambda^n,$$

where a_n does not depend on the model parameter λ . To find the maximum likelihood value for λ , the value of a_n does not matter. For five days of quality control tracking, the number of observed flaws were recorded. $n_1 = 4$, $n_2 = 2$, $n_3 = 5$, $n_4 = 4$, and $n_5 = 8$. Calculate the formula for the likelihood using $a_n = 1$ and use it to compute the log-likelihood. With a graph, estimate the maximum likelihood value for λ .