The likelihood ratio statistic

The second technique for constructing asymptotically exact confidence intervals is based on a quantity known as the *likelihood ratio statistic*. In order to motivate the development of this statistic, we begin with a simple coin flipping experiment that will highlight the importance of the likelihood ratio.

Let's say I have two biased coins. The probability of heads for coin 1 is $1/4$. The probability of heads for coin 2 is $3/4$. I select a coin at random and flip it four times, resulting in X heads. Regardless of the coin chosen, the possible values of X are $x = 0, 1, 2, 3, 4$. Table 3.9 gives the conditional probabilities associated with X for coin 1 and coin 2. These probabilities are the probability mass function values associated with the binomial(4, $1/4$) distribution for coin 1 and the binomial(4, $3/4$) distribution for coin 2. The bottom row of Table 3.9 contains the ratio of these two probabilities for each value of X. This ratio is known as the likelihood ratio. For example, if I select a coin at random, flip it four times and observe $x = 0$ heads, then the likelihood ratio tells me that it is 81 times more likely that the coin that was selected was coin 1. So this likelihood ratio is helpful in the effort to identify which coin was selected based on the number of heads that were observed.

	$x=0$	$x=1$	$x=2$	$x=3$	$x=4$
$P(X=x \mid \text{coin } 1)$	$81/256$	$27/64$	$27/128$	$3/64$	$1/256$
$P(X=x \mid \text{coin } 2)$	$1/256$	$3/64$	$27/128$	$27/64$	$81/256$
$\dfrac{P(X=x \mid \text{coin } 1)}{P(X=x \mid \text{coin } 2)}$	81	9	1	$1/9$	$1/81$

Table 3.9: Conditional distribution of X and likelihood ratio.

Table 3.9 contains the ratio of two probability mass functions. This ratio can be generalized to the ratio of likelihood functions, as defined next.

Definition 3.8 Let $X_1, X_2, \ldots, X_n$ be a random sample from a population whose probability distribution has a single unknown parameter θ and is described by $f(x)$. Let $\hat{\theta}$ be the maximum likelihood estimator of θ. The random variable

$$\Lambda = \frac{L(\theta)}{L\left(\hat{\theta}\right)}$$

is the *likelihood ratio statistic*.

The likelihood ratio statistic is typically introduced in conjunction with hypothesis testing (which is introduced in the next chapter), but it is defined here because it can also be used to construct asymptotically exact confidence intervals. Once again, the notation has been grossly oversimplified because the likelihood function is a function of the values from the random sample $X_1, X_2, \ldots, X_n$ and also θ. So it would be completely reasonable to write the definition of the likelihood ratio statistic as

$$\Lambda(X_1, X_2, \ldots, X_n, \theta) = \frac{L(X_1, X_2, \ldots, X_n, \theta)}{L\left(X_1, X_2, \ldots, X_n, \hat{\theta}\right)},$$

but compactness was chosen in the expression given in Definition 3.8. The likelihood ratio statistic

- is a random variable because the likelihood function and maximum likelihood estimator are functions of the data values, which are themselves random variables;

- is a random variable with positive support because the likelihood function is always positive;
- is less than or equal to one because $L(\theta)$ is maximized at the maximum likelihood estimator $\hat{\theta}$, so the numerator is always less than or equal to the denominator;
- has support on the interval $(0, 1]$ because of the previous two facts, so $0 < \Lambda \le 1$.

The probability distribution of the likelihood ratio statistic differs for each population probability distribution, so it alone is not of much use in constructing confidence intervals. However, as the following theorem shows, the opposite of twice the logarithm of the likelihood ratio statistic has a recognizable limiting probability density function—the chi-square distribution with one degree of freedom—which can be used to construct asymptotically exact confidence intervals.

Theorem 3.3 Let $X_1, X_2, \ldots, X_n$ be a random sample from a population whose probability distribution has a single unknown parameter θ and is described by $f(x)$. Let $\hat{\theta}$ be the maximum likelihood estimator of θ. Under certain regularity conditions concerning the population distribution,

$$-2\ln\Lambda = -2\ln\left(\frac{L(\theta)}{L\left(\hat{\theta}\right)}\right) \xrightarrow{D} \chi^2(1).$$

Proof An outline of the proof follows. For notational simplicity, let $l(\theta) = \ln L(\theta)$ for this proof only, so that

$$l'(\theta) = \frac{\partial \ln L(\theta)}{\partial\theta} \qquad \text{and} \qquad l''(\theta) = \frac{\partial^2 \ln L(\theta)}{\partial\theta^2}.$$

Expanding the log likelihood function $l(\theta)$ about $\hat{\theta}$ in a Taylor series gives

$$l(\theta) = l\left(\hat{\theta}\right) + l'\left(\hat{\theta}\right)\left(\theta - \hat{\theta}\right) + l''\left(\hat{\theta}\right)\frac{\left(\theta-\hat{\theta}\right)^2}{2!} + \cdots.$$

Recognizing that the derivative of the log likelihood function at $\hat{\theta}$ is zero (because $\hat{\theta}$ maximizes the log likelihood function) and ignoring the terms of order three and higher, the log likelihood function is approximately

$$l(\theta) \cong l\left(\hat{\theta}\right) + l''\left(\hat{\theta}\right)\frac{\left(\theta-\hat{\theta}\right)^2}{2}.$$

So $-2\ln\Lambda$ can be written as

$$\begin{aligned}
-2\ln\Lambda &= -2\ln\left(\frac{L(\theta)}{L\left(\hat{\theta}\right)}\right) \\
&= -2\left(\ln L(\theta) - \ln L\left(\hat{\theta}\right)\right) \\
&= -2\left(l(\theta) - l\left(\hat{\theta}\right)\right) \\
&\cong -2\left(l\left(\hat{\theta}\right) + l''\left(\hat{\theta}\right)\frac{\left(\theta-\hat{\theta}\right)^2}{2} - l\left(\hat{\theta}\right)\right) \\
&= -l''\left(\hat{\theta}\right)\left(\theta-\hat{\theta}\right)^2 \\
&= nI\left(\hat{\theta}\right)\left(\theta-\hat{\theta}\right)^2.
\end{aligned}$$

Since

$$\sqrt{nI(\theta)}\left(\hat{\theta}-\theta\right) \xrightarrow{D} N(0,1)$$

from Theorem 3.1, $I\left(\hat{\theta}\right) \xrightarrow{D} I(\theta)$, and the square of a standard normal random variable is a chi-square random variable with one degree of freedom,

$$-2\ln\Lambda \xrightarrow{D} \chi^2(1)$$

as desired. □

The regularity conditions in Theorem 3.3 are similar to those listed just after Theorem 2.4. Theorem 3.3 can be used to construct asymptotically exact confidence intervals for θ. Since

$$P\left(-2\ln\left(\frac{L(\theta)}{L\left(\hat{\theta}\right)}\right) < \chi^2_{1,\alpha}\right) = 1-\alpha,$$

an asymptotically exact two-sided confidence interval for θ is the set of all θ satisfying

$$-2\ln\left(\frac{L(\theta)}{L\left(\hat{\theta}\right)}\right) < \chi^2_{1,\alpha}.$$

Calculating the upper and lower bounds for a confidence interval for θ of this type typically requires the use of numerical methods, which is not ideal.

There is some nice geometry associated with this asymptotically exact confidence interval. The confidence interval can be rewritten as

$$\ln L(\theta) - \ln L\left(\hat{\theta}\right) > -\frac{\chi^2_{1,\alpha}}{2}$$

or

$$\ln L(\theta) > \ln L\left(\hat{\theta}\right) - \frac{\chi^2_{1,\alpha}}{2}.$$

So the asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for θ consists of all θ values associated with the log likelihood function being at a distance of less than $\chi^2_{1,\alpha}/2$ of its peak value at the maximum likelihood estimator. Figures will illustrate this geometry in each of the three examples that follow.

Example 3.28 Construct an asymptotically exact $100(1-\alpha)\%$ confidence interval for θ based on the likelihood ratio statistic for a random sample consisting of a single ($n=1$) observation X drawn from a population distribution with probability density function

$$f(x) = \theta x^{\theta-1} \qquad 0 < x < 1,$$

where θ is a positive unknown parameter. Apply this confidence interval procedure to the single observation $x = 0.63$ to give a 90% confidence interval for θ.

It might seem pointless to construct an asymptotically exact $100(1-\alpha)\%$ confidence interval for θ from just a single observation here for two reasons: (a) we already derived an *exact* confidence interval for θ in Example 3.1, and (b) an asymptotically exact confidence interval will probably not work well for $n=1$. The confidence interval is constructed here so as to see the mechanics and the geometry associated with the process. For a single observation, the likelihood function is the same as the probability density function

$$L(\theta) = f(x) = \theta x^{\theta - 1}.$$

The log likelihood function is

$$\ln L(\theta) = \ln \theta + (\theta - 1) \ln x.$$

The score is

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{1}{\theta} + \ln x.$$

Equating the score to zero gives the maximum likelihood estimate of θ as

$$\hat{\theta} = -\frac{1}{\ln x}.$$

The likelihood ratio statistic is

$$\Lambda = \frac{L(\theta)}{L\left(\hat{\theta}\right)} = \frac{\theta X^{\theta - 1}}{-\frac{1}{\ln X} X^{-1/\ln X - 1}} = -\theta \left(\ln X\right) \left(X^{\theta + 1/\ln X}\right).$$

Figure 3.34 shows that the likelihood ratio statistic Λ maps the support of X, which is $\mathcal{A} = \{x \,|\, 0 < x < 1\}$ to the support of Λ, which is $\mathcal{B} = \{\lambda \,|\, 0 < \lambda < 1\}$, although the transformation is not one-to-one. The transformation is plotted for $\theta = 0.5$, $\theta = 1$, and $\theta = 3$.

The asymptotically exact two-sided $100(1 - \alpha)\%$ confidence interval for θ consists of all θ satisfying

$$-2 \ln \left(\frac{L(\theta)}{L\left(\hat{\theta}\right)} \right) < \chi^2_{1, \alpha}.$$

In the case of the single data value $x = 0.63$ and $\alpha = 0.1$, the maximum likelihood estimate is

$$\hat{\theta} = -1/\ln x = -1/\ln 0.63 = 2.2.$$

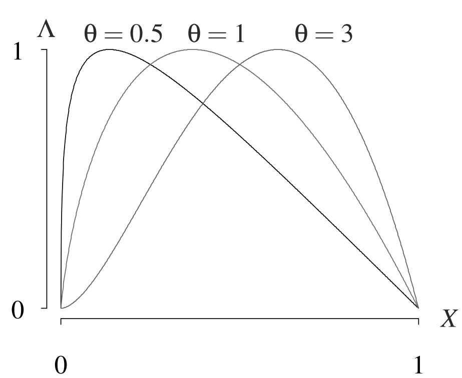

Figure 3.34: Likelihood ratio statistic Λ as a function of X.

Furthermore, the value of the log likelihood function at the maximum likelihood estimator is

$$\ln L\left(\hat{\theta}\right) = \ln L(2.2) = 0.23.$$

Next, find the 90th percentile of the chi-square distribution with one degrees of freedom via the R command `qchisq(0.9, 1)`, which gives

$$\chi^2_{1,0.1} = 2.7055.$$

After some algebra, the asymptotically exact 90% confidence interval consists of all values θ such that

$$\ln L(\theta) > \ln L\left(\hat{\theta}\right) - \frac{\chi^2_{1,0.1}}{2}$$

or

$$\ln L(\theta) > 0.23 - 1.35.$$

Figure 3.35 contains a plot of the log likelihood function, which is distinctly nonsymmetric because of the small sample size. The asymptotically exact 90% confidence interval corresponds to all of the values of the log likelihood function that are within 1.35 units of the peak value of the log likelihood function, which is indicated by a horizontal dashed line. The R function `uniroot` numerically solves for the confidence interval bounds in the code below. The variable `x` holds the data value, `theta` holds the maximum likelihood estimator, `logl` holds the value of the log likelihood function at the maximum likelihood estimator, and `crit` holds $\chi^2_{1,0.1}/2$.

```
x     = 0.63
theta = -1 / log(x)
logl  = log(theta) + (theta - 1) * log(x)
crit  = qchisq(0.9, 1) / 2
l = uniroot(function(th) log(th) + (th - 1) * log(x) - logl + crit,
            interval = c(0.1, theta))$root
u = uniroot(function(th) log(th) + (th - 1) * log(x) - logl + crit,
            interval = c(theta, 10))$root
```

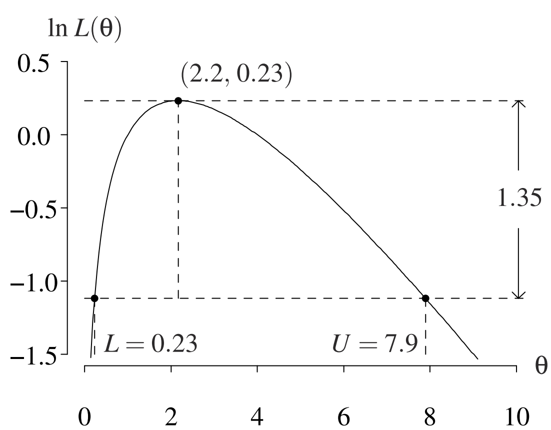

Figure 3.35: The log likelihood function near $\hat{\theta} = 2.2$ for the data set $x = 0.63$.

This code computes the asymptotically exact two-sided 90% confidence interval

$$0.23 < \theta < 7.9,$$

which is particularly nonsymmetric about the maximum likelihood estimator $\hat{\theta} = 2.2$. We are approximately 90% confident that θ lies between 0.23 and 7.9. The wide confidence interval is due to the fact that only a single observation X was collected. As a point of comparison, the exact two-sided 90% confidence interval from Example 3.1 was

$$0.11 < \theta < 6.5.$$

The asymptotically exact confidence interval developed here has approximate actual coverage for any finite value of n, which is always the case in practice. The next example considers a more appropriate application of the asymptotically exact confidence interval procedure, this time with $n = 200$.

Example 3.29 Construct an asymptotically exact $100(1-\alpha)\%$ confidence interval for λ based on the likelihood ratio statistic for a random sample $X_1, X_2, \ldots, X_n$ drawn from a Poisson(λ) population, where λ is a positive unknown parameter. Apply this confidence interval procedure to the horse kick data set from Example 2.9 to give a 95% confidence interval for λ.

From Example 2.9, the log likelihood function is

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \ln \lambda - \sum_{i=1}^{n} \ln(x_i!)$$

and the maximum likelihood estimate of λ is the sample mean

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

The asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for λ consists of all λ values satisfying

$$-2 \ln \left(\frac{L(\lambda)}{L(\hat{\lambda})}\right) < \chi^2_{1,\alpha}.$$

Recall that the horse kick data consisted of $n = 200$ corps-years of data, which is listed below. The $n = 200$ data values consist of 109 zeros, 65 ones, 22 twos, 3 threes, and 1

number of deaths per corps per year	0	1	2	3	4
number of observed values	109	65	22	3	1

four. The maximum likelihood estimate of λ is

$$\hat{\lambda} = \frac{109 \cdot 0 + 65 \cdot 1 + 22 \cdot 2 + 3 \cdot 3 + 1 \cdot 4}{200} = \frac{122}{200} = 0.61.$$

The geometry associated with the confidence interval is shown in Figure 3.36. Since the sample size ($n = 200$) is large, the log likelihood function is nearly symmetric. The the log likelihood function is maximized at the point

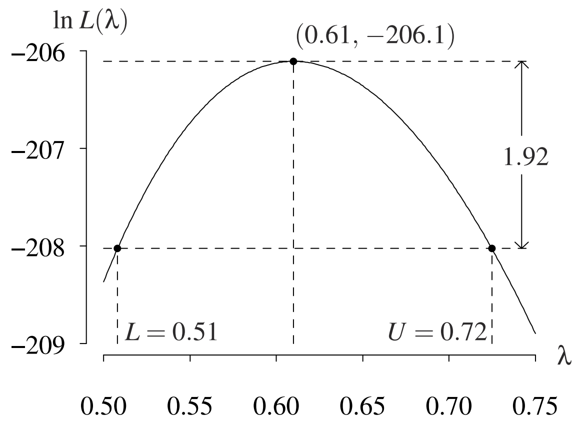

Figure 3.36: The log likelihood function near $\hat{\lambda} = 0.61$ for the horse kick data.

$$\left(\hat{\lambda}, \ln L(\hat{\lambda})\right) = (0.61, -206.1067).$$

The confidence interval bounds are found by drawing a horizontal line

$$\frac{\chi^2_{1,\alpha}}{2} = \frac{\chi^2_{1,0.05}}{2} = \frac{3.84}{2} = 1.92$$

units below the maximum point in Figure 3.36; the points where this horizontal line intersects the log likelihood function corresponds to the 95% confidence interval bounds. The R code below again uses the `uniroot` function to numerically solve for the lower and upper bound of the confidence interval.

```
x   = c(rep(0, 109), rep(1, 65), rep(2, 22), rep(3, 3), 4)
n   = length(x)
lam = mean(x)
t1  = sum(x)
t2  = sum(log(factorial(x)))
t3  = -n * lam + t1 * log(lam) - t2
t4  = qchisq(0.95, 1) / 2
l = uniroot(function(lambda) -n * lambda + t1 * log(lambda) - t2 -
            t3 + t4, lower = 0.1, upper = lam, tol = 1e-9)$root
u = uniroot(function(lambda) -n * lambda + t1 * log(lambda) - t2 -
            t3 + t4, lower = lam, upper = 0.9, tol = 1e-9)$root
```

The asymptotically exact two-sided 95% confidence interval, with bounds rounded to two significant digits, is

$$0.51 < \lambda < 0.72.$$

This approximate confidence interval can be compared to the approximate confidence intervals in Examples 3.12 and 3.26, which are

$$0.51 < \lambda < 0.73 \qquad \text{and} \qquad 0.50 < \lambda < 0.72.$$

All three of these confidence intervals are close to one another because the large sample size, $n = 200$, means that the asymptotic relationship on which two of them are based is quite accurate.

One of the problems associated with constructing asymptotically exact confidence intervals is that you cannot tell how well they perform for a finite sample size. And of course, the rate of convergence depends on the population distribution. In this light, a Monte Carlo simulation is conducted to see how close $-2\ln\Lambda$ comes to the chi-square distribution with one degree of freedom when samples of size $n = 200$ are drawn from a Poisson population with λ arbitrarily set to $\lambda = 3$. The following R code generates 100,000 random values of $-2\ln\Lambda$, which are stored in the vector `stat`.

```
lam  = 3
n    = 200
nrep = 100000
stat = numeric(nrep)
for (i in 1:nrep) {
  x = rpois(n, lam)
  lamhat = mean(x)
  t1  = sum(x)
  t2  = sum(log(factorial(x)))
  num = -n * lam + t1 * log(lam) - t2
  den = -n * lamhat + t1 * log(lamhat) - t2
  stat[i] = -2 * (num - den)
}
```

The empirical cumulative distribution function associated with these values, along with the cumulative distribution function of a chi-square random variable with one degree of freedom are plotted in Figure 3.37. The two cumulative distribution functions are nearly indistinguishable, so at least for a Poisson population with $\lambda = 3$ and a sample size of $n = 200$, using the likelihood ratio statistic will provide a confidence interval whose actual coverage is near the stated coverage.

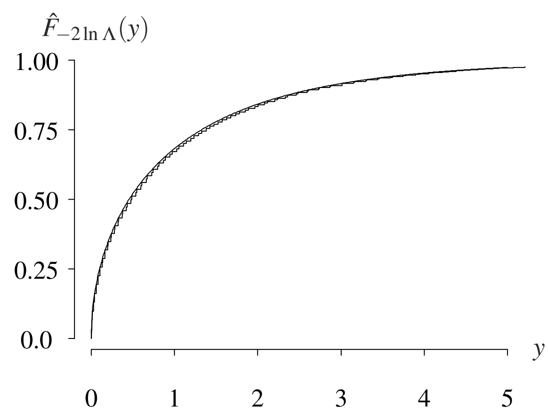

Figure 3.37: Cumulative distribution functions for $-2\ln\Lambda$ for a Poisson population and $\chi^2(1)$.