Extended Introductory MLE Example

John Fox
York SPIDA

2010

1 Preliminaries

The log-likelihood (from the lecture notes):

$$\log_e L(\pi) = x \log_e \pi + (n - x) \log_e (1 - \pi)$$

where x is the number of heads in n independent flips of the coin.

The derivative of the log-likelihood (also from the notes):

$$\begin{aligned}\frac{d \log_e L(\pi)}{d\pi} &= \frac{x}{\pi} + (n - x)\frac{1}{1 - \pi}(-1) \\ &= \frac{x}{\pi} - \frac{n - x}{1 - \pi}\end{aligned}$$

Setting the derivative to 0 and solving for π produces the MLE (again, from the notes),

$$\widehat{\pi} = \frac{x}{n}.$$

To get the Fisher information, we need the second derivative of the log likelihood, which is

$$\frac{d^2 \log_e L(\pi)}{d\pi^2} = -\frac{x}{\pi^2} - \frac{n - x}{(1 - \pi)^2}$$

Noting that the expected number of heads is $E(x) = n\pi$, the Fisher information is

$$\begin{aligned}\mathcal{I}(\pi) &= -E\left[\frac{d^2 \log_e L(\pi)}{d\pi^2}\right] \\ &= \frac{n\pi}{\pi^2} + \frac{n - n\pi}{(1 - \pi)^2} \\ &= \frac{n}{\pi(1 - \pi)}\end{aligned}$$

Then the asymptotic variance of the sample proportion $\widehat{\pi}$ is

$$\mathcal{V}(\widehat{\pi}) = \frac{1}{\mathcal{I}(\pi)} = \frac{\pi(1 - \pi)}{n}$$

and the estimated asymptotic standard error of the sample proportion is

$$\mathrm{SE}(\widehat{\pi}) = \sqrt{\frac{\widehat{\pi}(1 - \widehat{\pi})}{n}}$$

which is the familiar result for the standard error of a proportion.

2 Wald Test

Suppose that we want to test the hypothesis that the coin is fair, H_0: $\pi = .5$, and that our sample has $x = 7$ heads in $n = 10$ flips, so that $\widehat{\pi} = 7/10 = .7$. Then

$$\text{SE}(\widehat{\pi}) = \sqrt{\frac{.7(1-.3)}{10}} = 0.1449$$

and the Wald test statistic is

$$Z_0 = \frac{\widehat{\pi} - \pi_0}{\text{SE}(\widehat{\pi})} = \frac{.7 - .5}{0.1449} = 1.380$$

for which the two-sided p-value (from the standard-normal distribution) is $p = .167$.

3 Likelihood-Ratio Test

The likelihood is

$$L(\pi) = \pi^x (1-\pi)^{n-x}$$

and so the log-likelihood at the MLE and at the hypothesized value of π are, respectively,

$$\begin{aligned} \log_e L(\widehat{\pi}) &= \log_e[.7^7(1-.7)^3] = -6.1086 \\ \log_e L(\pi_0) &= \log_e[.5^7(1-.5)^3] = -6.9315 \end{aligned}$$

The likelihood-ratio (LR) test statistic is

$$\begin{aligned} G_0^2 &= 2[\log_e L(\widehat{\pi}) - \log_e L(\pi_0)] \\ &= 2(-6.1086 - -6.9315) = 1.646 \end{aligned}$$

From the χ^2 distribution with one degree of freedom, the p-value for this test statistic is $p = .199$.

Note: Converting the chisquare test statistic to a standard-normal test statistic produces $Z_0 = \sqrt{1.646} = 1.283$, which differs somewhat from the Wald test statistic.

4 Score Test

We need the score at the null value

$$\begin{aligned} S(\pi_0) &= \left. \frac{d \log_e L(\pi)}{d\pi} \right|_{\pi=\pi_0} = \frac{x}{\pi_0} - \frac{n-x}{1-\pi_0} \\ &= \frac{7}{.5} - \frac{10-7}{1-.5} = 8.0 \end{aligned}$$

and the Fisher information at the null value

$$\begin{aligned} \mathcal{I}(\pi_0) &\equiv \frac{n}{\pi_0(1-\pi_0)} \\ &= \frac{10}{.5(1-.5)} = 40.0 \end{aligned}$$

Then the score statistic is

$$S_0 = \frac{S(\pi_0)}{\sqrt{\mathcal{I}(\pi_0)}} = \frac{8.0}{\sqrt{40.0}} = 1.265$$

From the standard-normal distribution, the two-sided p-value is $p = .206$. In this case, therefore, the score statistic is closer to the likelihood-ratio statistic than the Wald statistic is.

5 An Exact Test

In this simple setting, an exact binomial test is available (as you likely learned in basic statistics). The distribution of the sample proportion $\widehat{\pi}$ and of the number of heads X is

$$\Pr(\widehat{\pi} = \frac{x}{n}) = \Pr(X = x) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

(being more careful now to distinguish the random variable X from a particular realized value x). The null distribution for the current example follows from setting $\pi = \pi_0 = .5$ and $n = 10$, and is given in the following table:

x	$\widehat{\pi} = \frac{x}{10}$	$\Pr(X = x)$
0	0.0	.000977
1	.1	.009766
2	.2	.043945
3	.3	.117188
4	.4	.205078
5	.5	.246094
6	.6	.205078
7	.7	.117188
8	.8	.043945
9	.9	.009766
10	1.0	.000977

Having obtained $x = 7$ heads (i.e., $\widehat{\pi} = .7$), the two-sided p-value for the hypothesis is $\Pr(X \leq 3 \text{ or } X \geq 7) = .3437$, quite different from the results produced by the three asymptotic tests. Of course, $n = 10$ is a very small sample.

6 Confidence Intervals

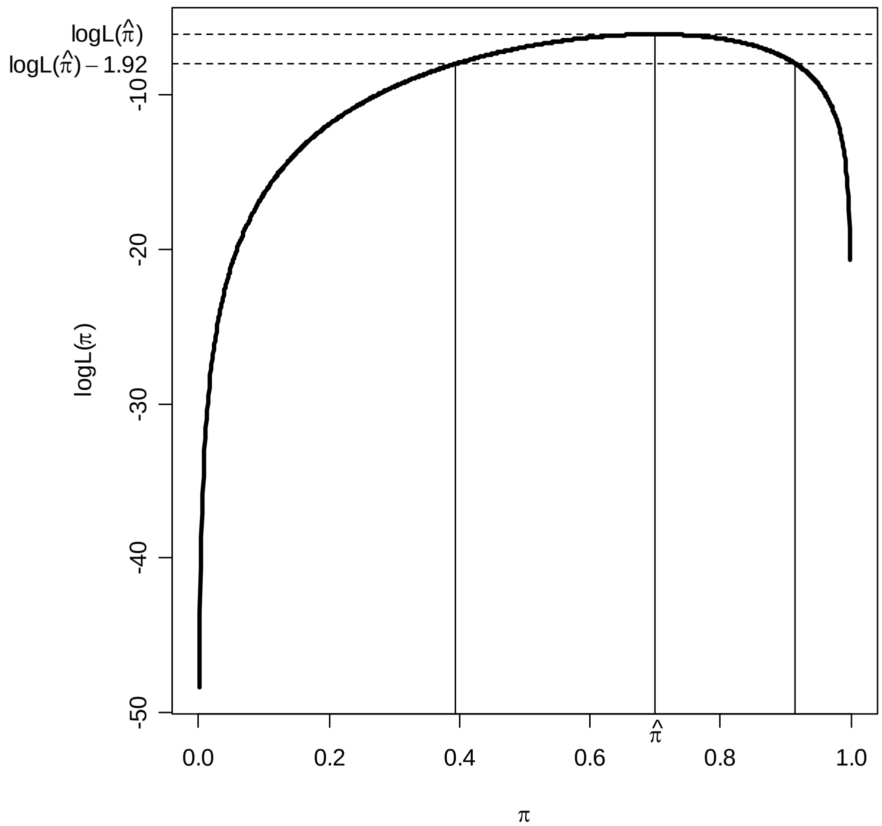
As mentioned in the lecture, the Wald and LR statistics can be inverted to produce confidence intervals. The Wald interval is particularly simple; e.g., for a 95-percent confidence interval:

$$\begin{aligned}\pi &= \widehat{\pi} \pm 1.96 \times \text{ SE}(\widehat{\pi}) \\ &= .7 \pm 1.96 \times 0.1449 \\ &= .7 \pm .284 \\ &= (.416, .984)\end{aligned}$$

The confidence interval based on the LR statistic includes all values of π that cannot be rejected when tested as hypotheses given the observed value of $\widehat{\pi}$. To be statistically significant at the .05 level, a chisquare statistic with one degree of freedom must be at least $1.96^2 = 3.84$. That is, the confidence interval includes all values of π for which

$$\begin{aligned}2[\log_e L(\widehat{\pi}) - \log_e L(\pi)] &\leq 3.84 \\ \log_e L(\widehat{\pi}) - \log_e L(\pi) &\leq 1.92\end{aligned}$$

This is illustrated, for the example, in the following graph:



The 95 percent confidence interval for π runs from .394 to .915.

With only $n = 10$ observations, neither the Wald interval nor the LR interval can be trusted.