STA 6384, Report 1.12

Carson Slater *Baylor University*

Problem: Reproduce the results in Agresti's Example 1.4.3, p. 15 using R.

Introduction

This document reproduces the analysis from Example 1.4.3 of the provided text. The example aims to estimate the proportion of vegetarians, π, based on a sample of $n = 25$ students from an honors class. In the sample, $x = 0$ students identified as vegetarian. The sample proportion is therefore $\hat{\pi} = 0/25 = 0$.

We will calculate three different 95% confidence intervals for π: the Wald interval, the score interval, and the likelihood-ratio interval.

---

1. The Wald Confidence Interval

The Wald confidence interval is calculated using the formula:

$$\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

For a 95% confidence level, $z_{0.025} \approx 1.96$. When $\hat{\pi} = 0$, the standard error term becomes 0, causing the interval to collapse to a single point.

```
## The 95% Wald confidence interval is: (0, 0)
```

As the text notes, the resulting interval of $(0, 0)$ is not very sensible, as it implies π is exactly 0 with 95% confidence.

---

2. The Score Confidence Interval (Wilson Interval)

The score interval is based on inverting the score test. It includes all values of π_0 that are not rejected by the test. The test statistic is:

$$z = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

We can calculate this interval in R using `prop.test()` with the `correct = FALSE` argument.

```
## The 95% Score (Wilson) confidence interval is: (0, 0.133)
```

This result, $(0.0, 0.133)$, matches the text.

Verifying with Score Tests

To illustrate how the interval is constructed, we can perform the score test for two values of π_0, as done in the text.

Test 1: $H_0 : \pi = 0.10$

```
## z-statistic: -1.67
## P-value: 0.0956
```

Since the P-value (0.0956) is greater than 0.05, we do not reject H_0. Thus, $\pi_0 = 0.10$ is included in the confidence interval.

Test 2: $H_0 : \pi = 0.20$

```
## z-statistic: -2.5
## P-value: 0.0124
```

Here, the P-value (0.0124) is less than 0.05, so we reject H_0. Thus, $\pi_0 = 0.20$ is not included in the interval.

---

3. The Likelihood-Ratio Confidence Interval

The likelihood-ratio interval is formed by all π_0 values for which the likelihood-ratio test statistic is less than or equal to the chi-squared critical value. The test statistic is:

$$-2[L(\pi_0) - L(\hat{\pi})] \leq \chi_1^2(0.05)$$

For $x = 0$, the log-likelihood function is $L(\pi) = n \log(1 - \pi)$. The maximum occurs at $\hat{\pi} = 0$, where $L(\hat{\pi}) = L(0) = 0$. The condition becomes:

$$-2[n \log(1 - \pi_0)] \leq \chi_1^2(0.05)$$

$$-50 \log(1 - \pi_0) \leq 3.841$$

The lower bound is 0. The upper bound is found by solving for equality.

$$\pi_0 = 1 - \exp\left(-\frac{\chi_1^2(0.05)}{50}\right)$$

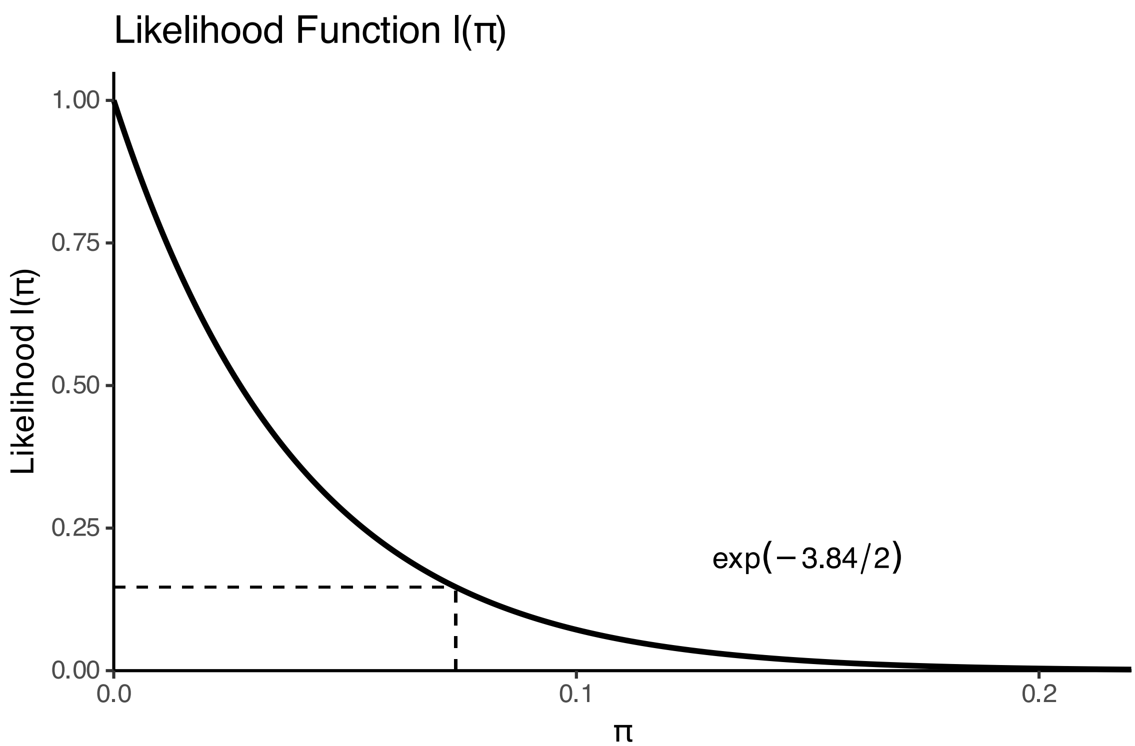
```
## The chi-squared critical value is: 3.841

## The 95% Likelihood-Ratio confidence interval is: (0.0, 0.074)
```

The result, $(0.0, 0.074)$, matches the text.

---

4. Graphical Representation

The text provides Figure 1.2, which plots the likelihood and log-likelihood functions. We can reproduce these plots.



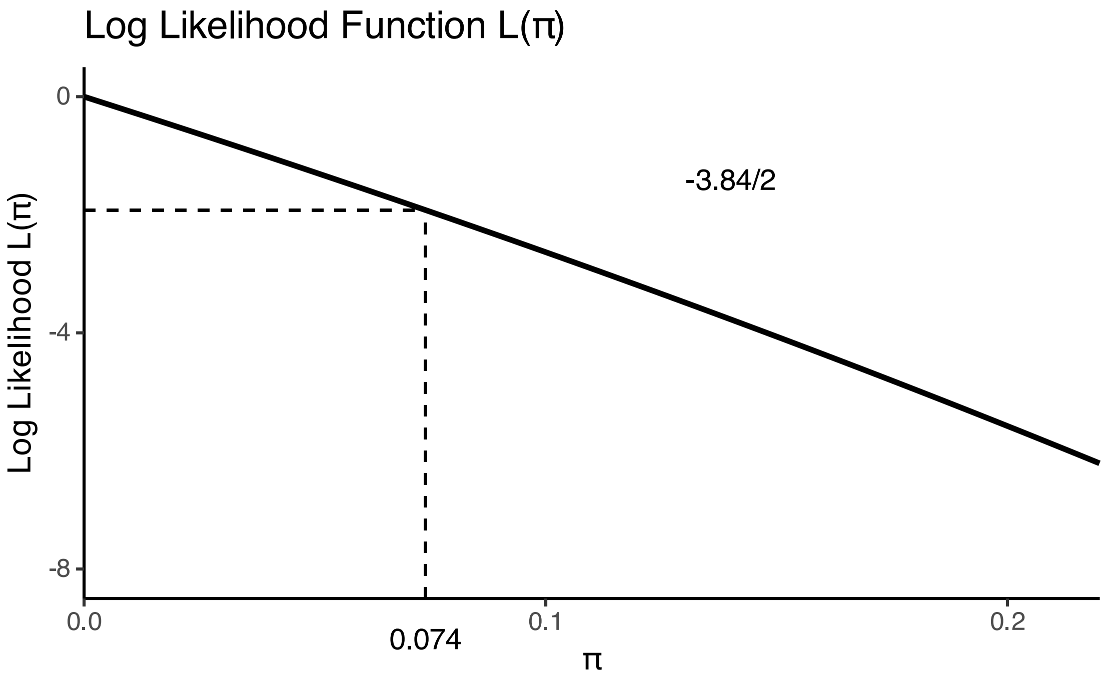


Figure 1.2 Reproduction: Binomial likelihood and log likelihood when y = 0 in n = 25 trials.

The generated plots closely resemble Figure 1.2 from the text, illustrating how the likelihood-ratio confidence interval is determined graphically. The upper bound of the interval (0.074) corresponds to the π value where the log-likelihood drops by $\chi^2_1(0.05)/2 = 1.92$ from its maximum value.
