STA 6384, Report 1.6

Carson Slater *Baylor University*

Problem: Work problem 1.7, p. 29 of Agresti. As a "part f", provide a Bayesian 95% credible set for π, using a uniform prior. Compare your answer to the confidence intervals obtained in the rest of the problem.

1.7 In a crossover trial comparing a new drug to a standard, π denotes the probability that the new one is judged better. It is desired to estimate π and test $H_0 : \pi = 0.50$ against $H_a : \pi \neq 0.50$. In 20 independent observations, the new drug is better each time.

a. Find and sketch the likelihood function. Is it close to the quadratic shape that large-sample normal approximations utilize?

Let Y be the number of times the new drug is judged better in n independent observations. The random variable Y follows a binomial distribution with parameters n and π, where π is the probability that the new drug is judged better. The probability mass function is given by:

$$P(Y = y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$$

The likelihood function, $L(\pi; y)$, is the probability of observing the data y as a function of the parameter π. In this problem, we have $n = 20$ trials, and the new drug was judged better in every trial, so $y = 20$.

Substituting $n = 20$ and $y = 20$ into the formula, we get:

$$L(\pi) = L(\pi; y = 20) = \binom{20}{20}\pi^{20}(1-\pi)^{20-20}$$

Since $\binom{20}{20} = 1$ and $(1-\pi)^0 = 1$, the likelihood function simplifies to:

$$L(\pi) = \pi^{20}, \quad \text{for } \pi \in [0, 1]$$

The function is $L(\pi) = \pi^{20}$.

- At $\pi = 0$, $L(0) = 0$.
- At $\pi = 1$, $L(1) = 1$.
- The function is monotonically increasing on the interval $[0, 1]$.
- The function is very flat for values of π close to 0 and rises very steeply as π approaches 1.

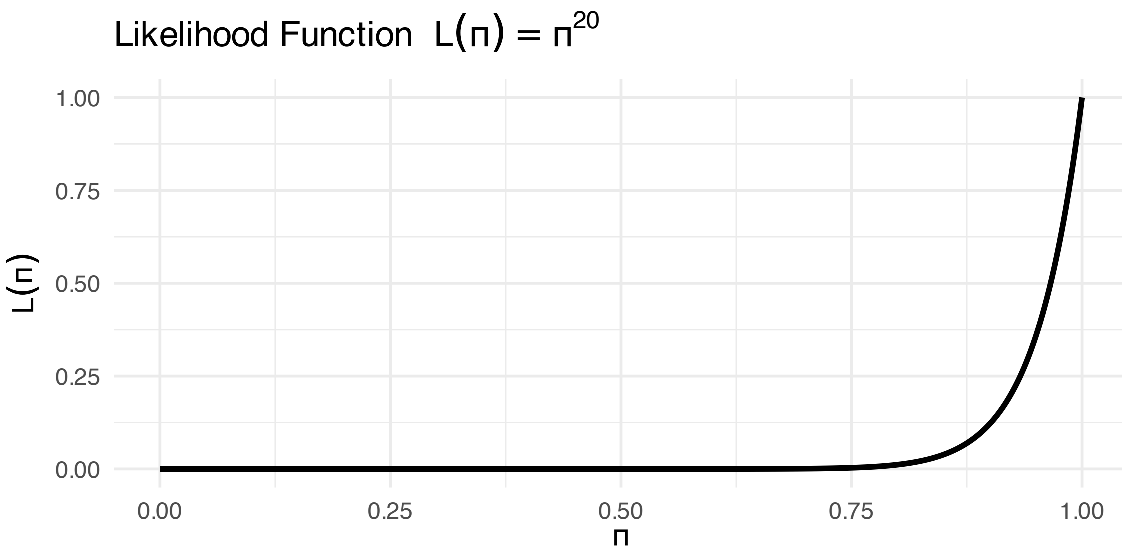
The maximum value of the likelihood function occurs at $\pi = 1$. A sketch of the function is shown below.

Large-sample normal approximations for inference (like the Wald test) rely on the log-likelihood function being approximately quadratic in the vicinity of the maximum likelihood estimate (MLE). This implies that the likelihood function itself should be approximately bell-shaped (symmetric) around the MLE.

In this case, the likelihood function $L(\pi) = \pi^{20}$ is **not close to a quadratic shape**.

- **Symmetry:** The function is highly skewed, not symmetric like a parabola.
- **Maximum Location:** The maximum of the function does not occur inside the parameter space $(0, 1)$ but at the boundary, $\hat{\pi} = 1$. A quadratic approximation is centered at the MLE, but here the MLE is at the edge of the domain, so a symmetric bell shape around it is not possible.

Because the MLE is on the boundary and the likelihood is extremely skewed, the quadratic approximation used for large-sample inference is very poor in this situation. [Gemini 2.5 pro-Assisted]



b. Give the ML estimate of π. Conduct a Wald test and construct a 95% Wald confidence interval for π. Are these sensible?

For a binomial experiment with y successes in n trials, the maximum likelihood estimate (MLE) of the probability of success π is given by:

$$\hat{\pi} = \frac{y}{n}$$

In this problem, we have $y = 20$ successes in $n = 20$ trials. Therefore, the ML estimate is:

$$\hat{\pi} = \frac{20}{20} = 1$$

This result is consistent with our finding in part (a), where the likelihood function $L(\pi) = \pi^{20}$ was maximized at $\pi = 1$.

The Wald test is used to test the hypothesis $H_0 : \pi = \pi_0$ against $H_a : \pi \neq \pi_0$. For this problem, $\pi_0 = 0.50$. The Wald test statistic is:

$$z = \frac{\hat{\pi} - \pi_0}{\text{SE}(\hat{\pi})}$$

The standard error of $\hat{\pi}$ is estimated using the formula:

$$\text{SE}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

Substituting our values $\hat{\pi} = 1$ and $n = 20$:

$$\text{SE}(\hat{\pi}) = \sqrt{\frac{1(1 - 1)}{20}} = \sqrt{\frac{0}{20}} = 0$$

Since the standard error is 0, the test statistic z becomes:

$$z = \frac{1 - 0.5}{0}$$

This expression is undefined due to division by zero. **Therefore, the Wald test cannot be conducted.** This breakdown occurs because the MLE, $\hat{\pi} = 1$, lies on the boundary of the parameter space, causing the estimated standard error to collapse to zero.

The formula for a $(1 - \alpha)\%$ Wald confidence interval for π is:

$$\hat{\pi} \pm z_{\alpha/2} \cdot \text{SE}(\hat{\pi})$$

For a 95% confidence interval, $\alpha = 0.05$, and $z_{\alpha/2} = z_{0.025} = 1.96$. Using our previously calculated values:

$$1 \pm 1.96 \cdot (0)$$

This gives a confidence interval of $[1, 1]$, or the single point $\{1\}$.

The results from the Wald procedure are **not sensible**.

- **Wald Test:** A statistical test that cannot be computed is not useful or sensible. It highlights the failure of the method under these specific data conditions.
- **Wald Confidence Interval:** The interval $[1, 1]$ suggests that we are 100% certain that the true value of π is exactly 1. While the evidence ($y = 20$ out of $n = 20$) is strong that π is very high, it is not definitive proof that it is exactly 1. For example, if the true probability were $\pi = 0.95$, the chance of observing 20 successes in a row is $0.95^{20} \approx 0.358$, which is not an impossible event. A sensible confidence interval should reflect this uncertainty by having a non-zero width.

The failure of both the Wald test and the confidence interval is a direct consequence of the issue identified in part (a): the likelihood function is not well-approximated by a quadratic function when the MLE is on the boundary of the parameter space. These methods, which rely on large-sample theory, are inappropriate for this dataset.

c. Conduct a score test, reporting the p-value. Construct a 95% score confidence interval. Interpret.

Unlike the Wald test which evaluates the standard error at the ML estimate ($\hat{\pi}$), the score test evaluates it at the null hypothesis value (π_0). This often leads to better performance in small samples or with extreme proportions.

We test the hypotheses $H_0 : \pi = 0.50$ against $H_a : \pi \neq 0.50$. The score test statistic is given by:

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

Here, $\hat{\pi} = 1$, $\pi_0 = 0.5$, and $n = 20$. The standard error is calculated under the null hypothesis:

$$\text{SE}_0 = \sqrt{\frac{0.50(1-0.50)}{20}} = \sqrt{\frac{0.25}{20}} = \sqrt{0.0125} \approx 0.1118$$

Now we can compute the test statistic:

$$z = \frac{1 - 0.50}{0.1118} \approx 4.472$$

The two-sided p-value is the probability of observing a result at least this extreme under the standard normal distribution:

$$p\text{-value} = 2 \times p(Z > 4.472)$$

This probability is extremely small ($\approx 7.74 \times 10^{-6}$).

The score confidence interval consists of all values of π_0 for which the null hypothesis $H_0 : \pi = \pi_0$ would not be rejected at the $\alpha = 0.05$ significance level. We find the endpoints of the interval by solving the equation $|z| = z_{\alpha/2}$ for π_0:

$$\left| \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \right| = 1.96$$

Squaring both sides gives the quadratic equation in terms of π_0:

$$(\hat{\pi} - \pi_0)^2 = \frac{(1.96)^2}{n}\pi_0(1-\pi_0)$$

Substituting $\hat{\pi} = 1$ and $n = 20$:

$$(1 - \pi_0)^2 = \frac{3.8416}{20}\pi_0(1-\pi_0)$$

$$(1 - \pi_0)^2 = 0.19208\pi_0(1-\pi_0)$$

Rearranging into the standard quadratic form $A\pi_0^2 + B\pi_0 + C = 0$:

$$(1 + 0.19208)\pi_0^2 - (2 + 0.19208)\pi_0 + 1 = 0$$

$$1.19208\pi_0^2 - 2.19208\pi_0 + 1 = 0$$

Solving this quadratic equation for π_0 gives the two roots that form the interval's endpoints:

$$\pi_0 = \frac{2.192 \pm \sqrt{(-2.19208)^2 - 4(1.19208)(1)}}{2(1.19208)}$$

$$\pi_0 = \frac{2.19208 \pm \sqrt{4.8052 - 4.76832}}{2.38416} = \frac{2.19208 \pm \sqrt{0.03688}}{2.38416} \approx \frac{2.19208 \pm 0.19204}{2.38416}$$

The two solutions are:

$$\pi_L = \frac{2.19208 - 0.19204}{2.38416} \approx 0.8389$$

$$\pi_U = \frac{2.19208 + 0.19204}{2.38416} \approx 1.0000$$

Thus, the 95% score confidence interval is approximately **(0.839, 1.0)**.

- **Test:** The extremely small p-value (< 0.00001) gives us overwhelming evidence to reject the null hypothesis. We can conclude that the probability of the new drug being judged better is not 0.50. The data strongly suggests the new drug is superior.
- **Confidence Interval:** We are 95% confident that the true probability π for the new drug being judged better is between 0.839 and 1.0. Unlike the Wald interval of $\{1\}$, the score interval has a non-zero width, which sensibly reflects that while π is very high, we cannot be completely certain it is exactly 1 based on a finite sample. This interval is far more reasonable and trustworthy than the Wald interval in this case.

d. Conduct a likelihood-ratio test and construct a likelihood-based 95% confidence interval. Interpret.

We test the hypotheses $H_0 : \pi = 0.50$ against $H_a : \pi \neq 0.50$. The likelihood-ratio test statistic is given by:

$$\lambda = -2 \ln \left(\frac{\sup_{\pi \in H_0} L(\pi)}{\sup_{\pi \in H_a} L(\pi)} \right) = -2 \ln \left(\frac{L(\pi_0)}{L(\hat{\pi})} \right)$$

Under H_0, the likelihood is evaluated at the specified value $\pi_0 = 0.5$:

$$L(\pi_0) = L(0.5) = (0.5)^{20}$$

The unrestricted maximum of the likelihood occurs at the MLE, $\hat{\pi} = 1$:

$$L(\hat{\pi}) = L(1) = 1^{20} = 1$$

Now, we compute the test statistic:

$$\lambda = -2 \ln \left(\frac{(0.5)^{20}}{1} \right) = -2 \ln((0.5)^{20}) = -2 \times 20 \times \ln(0.5)$$

$$\lambda = -40 \times (-0.69315) \approx 27.726$$

Under the null hypothesis, this test statistic follows a chi-squared distribution with degrees of freedom equal to the difference in the number of free parameters, which is $df = 1$.

$$\lambda \sim \chi^2_1$$

The p-value is the probability of observing a test statistic at least as extreme as 27.726:

$$p\text{-value} = P(\chi^2_1 \geq 27.726)$$

The critical value for a χ^2_1 distribution at $\alpha = 0.05$ is 3.841. Since our test statistic of 27.726 is far greater than this, the p-value is extremely small (1.4×10^{-7}).

The likelihood-based confidence interval is the set of all parameter values π_0 for which the LRT would not reject $H_0 : \pi = \pi_0$ at the $\alpha = 0.05$ level. This means we seek all π_0 that satisfy:

$$-2 \ln \left(\frac{L(\pi_0)}{L(\hat{\pi})} \right) \leq \chi^2_{1,0.05}$$

Substituting the likelihood functions and the critical value $\chi^2_{1,0.05} \approx 3.841$:

$$-2 \ln \left(\frac{\pi_0^{20}}{1} \right) \leq 3.841$$

$$-40 \ln(\pi_0) \leq 3.841$$

We divide by -40, which reverses the inequality sign:

$$\ln(\pi_0) \geq -\frac{3.841}{40}$$

$$\ln(\pi_0) \geq -0.096025$$

To find π_0, we exponentiate both sides:

$$\pi_0 \geq e^{-0.096025}$$

$$\pi_0 \geq 0.9084$$

The upper bound of the interval is the MLE, $\hat{\pi} = 1$. Therefore, the 95% likelihood-based confidence interval is approximately **[0.908, 1.0]**.

- **Test:** The extremely small p-value ($\approx 1.4 \times 10^{-7}$) provides very strong evidence to reject the null hypothesis. We conclude that the probability of the new drug being judged better is significantly different from 0.50. The data strongly indicate the new drug is superior.
- **Confidence Interval:** We are 95% confident that the true probability π that the new drug is judged better falls within the interval [0.908, 1.0]. This result is sensible. It reflects the strong evidence in the data by being concentrated at high values of π, but its non-zero width acknowledges sampling uncertainty. This interval is considered very reliable, often having better coverage properties than the Wald interval, especially in cases like this.

e. Construct an exact binomial test. Interpret.

The exact binomial test is appropriate here because the assumptions for large-sample tests (Wald, Score, LRT) are not met, as the ML estimate is on the boundary of the parameter space. This test uses the binomial distribution directly to calculate the p-value and is valid for any sample size.

We are testing the hypotheses:

$$H_0 : \pi = 0.50 \quad \text{against} \quad H_a : \pi \neq 0.50$$

The observed data is $y = 20$ successes in $n = 20$ trials. Under the null hypothesis, the number of successes Y follows a binomial distribution with $n = 20$ and $\pi = 0.5$.

$$Y \sim \text{Bin}(20, 0.5)$$

The p-value is the probability of observing a result as extreme or more extreme than $y = 20$. For a two-sided test, we consider extreme outcomes in both tails of the distribution. The outcome of 20 successes is the most extreme possible result. The equally extreme result in the other tail is 0 successes.

The p-value is therefore the sum of the probabilities of these two outcomes:

$$p\text{-value} = p(Y = 20) + p(Y = 0)$$

Using the binomial probability formula, $P(Y = k) = \binom{n}{k}\pi^k(1 - \pi)^{n-k}$:

$$P(Y = 20) = \binom{20}{20}(0.5)^{20}(1 - 0.5)^0 = 1 \cdot (0.5)^{20}$$

$$P(Y = 0) = \binom{20}{0}(0.5)^0(1 - 0.5)^{20} = 1 \cdot (0.5)^{20}$$

Summing these probabilities:

$$p\text{-value} = (0.5)^{20} + (0.5)^{20} = 2 \cdot (0.5)^{20} = 2 \cdot \left(\frac{1}{2}\right)^{20} = \frac{2}{1,048,576}$$

$$p\text{-value} \approx 0.000001907$$

The p-value is extremely small ($p \approx 1.9 \times 10^{-6}$), which is far below any conventional significance level (e.g., $\alpha = 0.05$ or $\alpha = 0.01$).

Therefore, we **reject the null hypothesis** H_0.

There is overwhelming statistical evidence to conclude that the probability of the new drug being judged better is not 0.50. The observed result of 20 successes in 20 trials is exceptionally unlikely to have occurred if the true probability was 50%. We can be very confident that the new drug is superior to the standard.

f. provide a Bayesian 95% credible set for π, using a uniform prior. Compare your answer to the confidence intervals obtained in the rest of the problem.

For this part, we perform a Bayesian analysis to find a 95% credible set for π. A credible set is the Bayesian analogue to a frequentist confidence interval.

A Bayesian analysis combines a prior belief about a parameter with the likelihood of the observed data to form an updated posterior belief.

- **Likelihood:** As established in part (a), the likelihood function for the data ($y = 20, n = 20$) is $L(\pi) = \pi^{20}$.
- **Prior Distribution:** You specified a uniform prior for π. This expresses a belief that all values of π between 0 and 1 are equally likely before observing the data. A uniform prior is a special case of the Beta distribution, specifically Beta$(\pi|\alpha = 1, \beta = 1)$.

$$p(\pi) = 1, \quad \text{for } \pi \in [0, 1]$$

- **Posterior Distribution:** The posterior distribution is proportional to the product of the likelihood and the prior:
$$p(\pi|y) \propto L(\pi|y) \times p(\pi)$$
$$p(\pi|y) \propto \pi^{20} \times 1 = \pi^{20}$$
Recognizing this as the kernel of a Beta distribution, we can use the property of conjugacy. For a Beta(α, β) prior and binomial likelihood, the posterior is Beta($y + \alpha, n - y + \beta$). With $y = 20, n = 20, \alpha = 1, \beta = 1$, the posterior distribution for π is:
$$\pi|y \sim \text{Beta}(20 + 1, 20 - 20 + 1) = \textbf{Beta(21, 1)}$$
The probability density function (pDF) for this posterior is $p(\pi|y) = 21\pi^{20}$ for $\pi \in [0, 1]$.

We need to find an interval $[L, U]$ such that $p(L < \pi < U|y) = 0.95$. We will construct an equal-tailed interval by finding the 2.5th and 97.5th percentiles of the posterior distribution. First, we find the cumulative distribution function (CDF), $F(x)$, for the Beta(21, 1) distribution:

$$F(x) = \int_0^x 21t^{20} dt = \left[t^{21}\right]_0^x = x^{21}$$

Now we find the lower (L) and upper (U) bounds:

- Lower bound (2.5th percentile): $L^{21} = 0.025 \implies L = (0.025)^{1/21} \approx 0.8389$
- Upper bound (97.5th percentile): $U^{21} = 0.975 \implies U = (0.975)^{1/21} \approx 0.9988$

Thus, the 95% Bayesian credible set for π is approximately **[0.839, 0.999]**.

Let's summarize the four intervals calculated for π:

- **Wald Interval (b):** [1.0, 1.0]
- **Score Interval (c):** [0.839, 1.0]
- **Likelihood-Ratio Interval (d):** [0.908, 1.0]
- **Bayesian Credible Set (f):** [0.839, 0.999]

Interpretation of Comparison:

- The **Wald interval** is clearly nonsensical. Its zero width is an artifact of the method's failure when the MLE is on the boundary.
- The **Bayesian credible set** and the **Score confidence interval** are remarkably similar. Their lower bounds are identical, and their upper bounds differ by only 0.001. This illustrates a known connection between Bayesian methods with non-informative priors and certain frequentist procedures.

- The **Likelihood-Ratio interval** gives a slightly different result, with a higher lower bound (0.908). This is not surprising, as it is based on a different asymptotic theory (χ^2 distribution of the log-likelihood ratio) than the Score test (normal distribution of the score function).
- All three sensible methods (Score, LRT, Bayesian) tell a consistent story: the true value of π is very high, but we cannot be certain it is exactly 1. They all provide a plausible range for π that reflects the sampling uncertainty, which is a much more reasonable conclusion than the one provided by the Wald interval.