Exact Confidence Intervals for Binomial Proportions

Dr. Jimmy A. Doi
Department of Statistics
Cal Poly San Luis Obispo

Email: jdoi@calpoly.edu
Web: www.calpoly.edu/~jdoi

March 25, 2004
Clinical Trial Dataset: Cancer

- Drug 1: The New Treatment
- Drug 2: The Standard Treatment

<table>
<thead>
<tr>
<th>Response</th>
<th>Drug 1</th>
<th>Drug 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>Failure</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

- Proportion of Success on Drug 1 = 60%
- Proportion of Success on Drug 2 = 32%
- Treatment Difference = 28%

Does this data suggest Drug 1 is significantly better than Drug 2? Confidence intervals are useful in addressing this question.
Outline

• Introduction
  – Confidence Interval for $\pi$
    * Asymptotic vs. Exact Methods

• Current Research
  – Confidence Interval for $\pi_1 - \pi_2$
    * Standard vs. Proposed Exact Methods

• Summary
Parameter of interest is any quantitative aspect of the population (true average $\mu$, true percentage $\pi$)

For this presentation, our focus will be on the population proportion $\pi$. 
Example: Illicit Drug Usage

Survey Question to Undergraduates

“Have you ever experimented with an illegal drug/substance for non-medicinal purposes?”

Parameter Space of $\pi$

\[ \pi \]

Population

\[ \pi \]

\[ \text{Population} \]

\[ \pi \]
Example: Illicit Drug Usage (cont.)

Given a sample from the population, let the sample proportion be denoted as $\hat{\pi}$.

Problem:
Is the estimate close to the true value ... 

... or relatively far away?
Confidence Intervals

Instead of a point estimator like $\hat{\pi}$, use an interval estimator $(l, u)$.

“Worth Going to War in Iraq?”

“Yes”: 55% ± 3% (Gallup Poll, 3/18/04)
Confidence Intervals: Definitions

- Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from the population.

- Generate intervals of the form

  $$[L(\mathbf{X}), U(\mathbf{X})]$$

  which capture the true parameter $\theta$ for a high percentage of all possible samples.

- These intervals are known as **confidence intervals**.

- Given the interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$, its
  
  - **coverage probability** is $P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.
  
  - **confidence coefficient** ($C$) is $\inf_\theta P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.
    (typical values of $C$ are 90%, 95%, and 99%)
Confidence Intervals: Comparison

- Consider two competing CI methods:
  \[[L_1(X), U_1(X)]\] versus \[[L_2(X), U_2(X)]\]

- To compare competing methods, one should examine coverage probability and interval length properties

Ex. Recall the illicit drug usage proportion \(\pi\):

Clearly, the interval \([0,1]\) contains the true parameter. This is a 100% confidence interval for \(\pi\), but it has no utility.
Asymptotic versus Exact Methods

Our parameter of interest, $\pi$, often stems from the proportion parameter of the binomial distribution:

- $X$ is binomial($n, \pi$), a discrete random variable
- Prob. Mass Function $= \text{bin}(x, n, \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$

**Problem:** Using $\text{bin}(x, n, \pi)$ to create a CI for $\pi$ can be an inconvenient and cumbersome chore

**Solution:** Use an asymptotic approximation approach (e.g. normal distribution)
Asymptotic Approximation for Binomial

\( X \) is binomial\((n, \pi)\)

Prob. Mass Function = \( \text{bin}(x, n, \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \)
Confidence Interval for $\pi$: Asymptotic vs. Exact

Asymptotic CI: $[L_1(X), U_1(X)]$, Wald
Given $X = x$, a popular asymptotic 95% CI for $\pi$ is ($\hat{\pi} = x/n$):

$$\hat{\pi} \pm 1.96 \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

Exact CI: $[L_2(X), U_2(X)]$, Clopper & Pearson
Given $X = x$, a well known exact 95% CI for $\pi$ has endpoints with values of $\pi$ that satisfy:

$$\sum_{k=x}^{n} \binom{n}{k} \pi^k (1 - \pi)^{n-k} = 0.025$$

and

$$\sum_{k=0}^{x} \binom{n}{k} \pi^k (1 - \pi)^{n-k} = 0.025.$$
Coverage Probability

- $X$ is binomial$(n, \pi)$
- For $x \in \{0, 1, 2, \ldots, n\}$, generate CI for each sample point:

<table>
<thead>
<tr>
<th>$X = x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>$\cdots$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CI(x)$</td>
<td>$(l_0, u_0)$</td>
<td>$(l_1, u_1)$</td>
<td>$(l_2, u_2)$</td>
<td>$\cdots$</td>
<td>$(l_n, u_n)$</td>
</tr>
</tbody>
</table>

- For each $\pi \in [0, 1]$, set $A = \{x : \pi \in CI(x)\}$
- coverage probability $= \sum_{x \in A} \text{bin}(x, n, \pi)$
Asymptotic CI Coverage Probability

\(X \sim \text{binomial}(n=15, \pi)\), Asymptotic CI Coverage

- Coverage Probability
- \(\pi\)
Asymptotic CI Coverage Probability

$X \sim \text{binomial}(n=15, \pi)$, Asymptotic CI Coverage

Coverage Probability

0.80 0.85 0.90 0.95 1.00

0.2 0.4 0.6 0.8

$\pi$
Exact CI Coverage Probability

X~binomial(n=15, π), Exact CI Coverage
Exact Coverage Probability Example

\[ X \sim \text{bin}(n = 5, \pi) \]

\[
\binom{n}{x} \pi^x (1 - \pi)^{n-x}
\]

\[ A = \{ x : \pi \in CI(x) \} \]

\[ CP = \sum_{x \in A} \binom{n}{x} \pi^x (1 - \pi)^{n-x} \]
Exact Coverage Probability Example

\[ X \sim \text{bin}(n = 5, \pi) \]

\[ \text{bin}(x, n, \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \]

\[ A = \{ x : \pi \in CI(x) \} \]

\[ \text{CP} = \sum_{x \in A} \text{bin}(x, n, \pi) \]

![Graph showing coverage probability for a binomial distribution with n=5 and \( \pi \). The graph includes intervals for different values of \( X \) and corresponding confidence intervals.](image)
Exact Coverage Probability Example

\[ X \sim \text{bin}(n = 5, \pi) \]

\[ \text{bin}(x, n, \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \]

\[ A = \{ x : \pi \in CI(x) \} \]

\[ CP = \sum_{x \in A} \text{bin}(x, n, \pi) \]
Asymptotic versus Exact Methods

Asymptotic:

- Traditional method based on asymptotic theory
- **Benefit**: Easy to implement, maintains nominal level for large sample size(s)
- **Drawback**: For small to moderate sample sizes, as found in many clinical trials, asymptotic test performs very poorly

Exact:

- Exact testing is based on the true underlying distribution(s)
- **Benefit**: No approximations used, guaranteed to maintain nominal level regardless of sample size(s)
- **Drawback**: Difficult to implement, time consuming, computationally intensive
Clinical Trial Background

- Clinical trial to compare a standard and new treatment (i.e. New Trt = Drug 1, Standard Trt = Drug 2)

\[ \pi_1 \text{ and } \pi_2 \text{ represent true response rates of the new and standard treatments, respectively} \]

- Assume a higher response rate indicates stronger efficacy of the drug

- Comparison of the two parameters \( \pi_1 \) and \( \pi_2 \) through their difference, \( \pi_1 - \pi_2 = \delta \)
### 2 × 2 Contingency Table

<table>
<thead>
<tr>
<th>Response</th>
<th>Population 1</th>
<th>Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>$X_1$</td>
<td>$X_2$</td>
</tr>
<tr>
<td>Failure</td>
<td>$n_1 - X_1$</td>
<td>$n_2 - X_2$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_1$</td>
<td>$n_2$</td>
</tr>
</tbody>
</table>

- 2 independent response variables $(X_1, X_2)$ distributed as binomial $(n_1, \pi_1)$ and binomial $(n_2, \pi_2)$, respectively
- Let $x_1$ and $x_2$ be observed values from the data which will be denoted as $\mathbf{x}$
- Let the sample space of $(X_1, X_2)$ be denoted by $\Omega_{\mathbf{x}} = \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\}$. 
Standard vs. Proposed Exact CI Methods

Goal: Construct a confidence interval for $\pi_1 - \pi_2$

Problem with existing method:
- The Standard (Unconditional) Exact Method
  - computationally intensive and time consuming algorithm
  - potentially leads to conservative outcomes

Research on alternative approach:
- The Proposed (Unconditional) Exact Method
  - computationally less intensive method via modified algorithm
  - less conservative outcomes
Comparison of Methods: Confidence Intervals

- Consider sample size ratios \((n_1 : n_2) \in \{1 : 1, 2 : 1, 3 : 1\}\)
  where \((n_1 + n_2) \in \{20, 60, 100\}\)

- Confidence interval comparisons
  - Coverage probability
  - Expected length
Comparison of Methods: Coverage Probability

Method for fixed $\pi_2$:

• Choose $\pi_2 \in [0, 1]$

• For $\pi_1 \in [0, 1]$, define $\delta = \pi_1 - \pi_2$

• Set $A = \{(x_1, x_2) : \delta \in CI(x_1, x_2)\}$

• coverage probability $= \sum\sum_{(x_1, x_2) \in A} \binom{x_1}{n_1, \pi_1} \binom{x_2}{n_2, \pi_2}$
Coverage Prob. Plots for $n_1 : n_2 = 1 : 1$, 95% CI, $\pi_2 = 25\%$

- $n_1=10$, $n_2=10$, $\pi_2=0.25$
- $n_1=30$, $n_2=30$, $\pi_2=0.25$
- $n_1=50$, $n_2=50$, $\pi_2=0.25$

- Standard Exact Method
- Proposed Exact Method
Coverage Prob. Plots for $n_1 : n_2 = 2 : 1$, 95% CI, $\pi_2 = 25\%$

- $n_1=13$, $n_2=7$, $\pi_2=0.25$
- $n_1=40$, $n_2=20$, $\pi_2=0.25$
- $n_1=66$, $n_2=34$, $\pi_2=0.25$

- (13,7)
- (40,20)
- (66,34)
Coverage Prob. Plots for $n_1 : n_2 = 3 : 1$, 95% CI, $\pi_2 = 25\%$
Comparison of Methods: Expected Length

Method for fixed $\pi_2$:

- Choose $\pi_2 \in [0, 1]$
- $\pi_1 \in [0, 1]$
- expected length =

\[
\sum_{(x_1, x_2) \in \Omega_X} \|CI(x_1, x_2)\| \binom{x_1}{n_1, \pi_1} \binom{x_2}{n_2, \pi_2}
\]
Expected Length Plots for $n_1 : n_2 = 1 : 1$, 95% CI, $\pi_2 = 25\%$

- $n_1=10$, $n_2=10$, $\pi_2=0.25$
- $n_1=30$, $n_2=30$, $\pi_2=0.25$
- $n_1=50$, $n_2=50$, $\pi_2=0.25$

Graphs illustrate the length of the confidence interval for different sample sizes and confidence levels. The plots compare the standard exact method (solid line) and the proposed exact method (dotted line).
**Expected Length Plots for $n_1: n_2 = 2:1$, 95% CI, $\pi_2 = 25\%$**

- **Exp. Length**
  - $n_1=13, n_2=7, \pi_2=0.25$
  - $n_1=40, n_2=20, \pi_2=0.25$
  - $n_1=66, n_2=34, \pi_2=0.25$

- **Graphs**
  - (13,7)
  - (40,20)
  - (66,34)

- **Legend**
  - Standard Exact Method
  - Proposed Exact Method

---

31
Expected Length Plots for $n_1 : n_2 = 3 : 1, 95\% CI, \pi_2 = 25\%$

(15,5) $n_1=15, n_2=5, \pi_2=0.25$

(45,15) $n_1=45, n_2=15, \pi_2=0.25$

(75,25) $n_1=75, n_2=25, \pi_2=0.25$

- Standard Exact Method
- Proposed Exact Method
Summary

- Confidence intervals for $\pi$:
  - Asymptotic methods (normality) are not appropriate for small to moderate sample sizes
  - Exact methods, though conservative, offer improved performances

- Confidence intervals for $\pi_1 - \pi_2$:
  - benefits of using the proposed method are most apparent when $n_1$ and $n_2$ are unbalanced
  - gains outweigh losses
Example: Illicit Drug Usage

Survey Question to Undergraduates

“Have you ever experimented with an illegal drug/substance for non-medicinal purposes?”

What is the true value across surveys (n=750)?
Example: Illicit Drug Usage

Survey Question to Undergraduates

“Have you ever experimented with an illegal drug/substance for non-medicinal purposes?”

What is the true value across surveys (n=750)?

0.571