STAT 226 Lecture 1 & 2

Yibi Huang

Outline

- Variable Types
- Review of Binomial Distributions
- Likelihood and Maximum Likelihood Method
- Tests for Binomial Proportions
- Confidence Intervals for Binomial Proportions

Variable Types

Regression methods are used to analyze data when the response variable is **numerical**.

- e.g., temperature, blood pressure, heights, speeds, income
- Covered in Stat 222 & 224

Methods in **categorical data analysis** are used when the response variable are **categorical**, e.g.,

- gender (male, female),
- political philosophy (liberal, moderate, conservative),
- region (metropolitan, urban, suburban, rural)
- Covered in Stat 226 & 227 (Don't take both STAT 226 and 227)

In either case, the explanatory variables can be numerical or categorical.

Nominal and Ordinal Categorical Variables

- **Nominal**: unordered categories, e.g.,
  - transport to work (car, bus, bicycle, walk, other)
  - favorite music (rock, hiphop, pop, classical, jazz, country, folk)
- **Ordinal**: ordered categories
  - patient condition (excellent, good, fair, poor)
  - government spending (too high, about right, too low)

We pay special attention to — **binary variables**: success or failure for which nominal-ordinal distinction is unimportant.

Review of Binomial Distributions

Binomial Distributions (Review)

If n Bernoulli trials are performed:

- only two possible outcomes for each trial (success, failure)
- $\pi = P(\text{success})$, $1 - \pi = P(\text{failure})$, for each trial,
- trials are independent
- Y = number of successes out of n trials

then we say Y has a **binomial distribution**, denoted as

$$Y \sim \text{Binomial}\,(n, \pi).$$

The probability function of Y is

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad y = 0, 1, \ldots, n.$$

where $\binom{n}{y} = \dfrac{n!}{y!\,(n-y)!}$ is the *binomial coefficient* and

$m! = m$ factorial $= m \times (m-1) \times (m-2) \times \cdots \times 1$ Note that $0! = 1$

Example: Are You Comfortable Getting a Covid Booster?

Response (Yes, No). Suppose $\pi = \text{Pr(Yes)} = 0.4$.

Let y = # answering Yes among $n = 3$ randomly selected people.

Example: Are You Comfortable Getting a Covid Booster?

Response (Yes, No). Suppose $\pi = \Pr(\text{Yes}) = 0.4$.

Let y = # answering Yes among $n = 3$ randomly selected people.

$$P(y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y} = \frac{3!}{y!(3-y)!}(0.4)^y(0.6)^{3-y}$$

$$P(0) = \frac{3!}{0!3!}(0.4)^0(0.6)^3 = (0.6)^3 = 0.216$$

$$P(1) = \frac{3!}{1!2!}(0.4)^1(0.6)^2 = 3(0.4)(0.6)^2 = 0.432$$

$$P(2) = \frac{3!}{2!1!}(0.4)^2(0.6)^1 = 3(0.4)^2(0.6) = 0.288$$

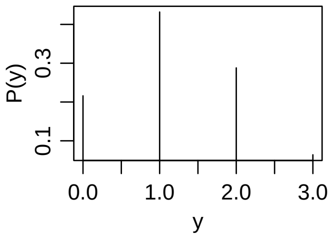
$$P(3) = \frac{3!}{3!0!}(0.4)^3(0.6)^0 = (0.4)^3 = 0.064$$

y	0	1	2	3	Total
$P(y)$	0.216	0.432	0.288	0.064	1

Binomial Probabilities in R

```
dbinom(x=0, size=3, p=0.4)
[1] 0.216
dbinom(0, 3, 0.4)
[1] 0.216
dbinom(1, 3, 0.4)
[1] 0.432
dbinom(x=0:3, size=3, p=0.4)
[1] 0.216 0.432 0.288 0.064
```

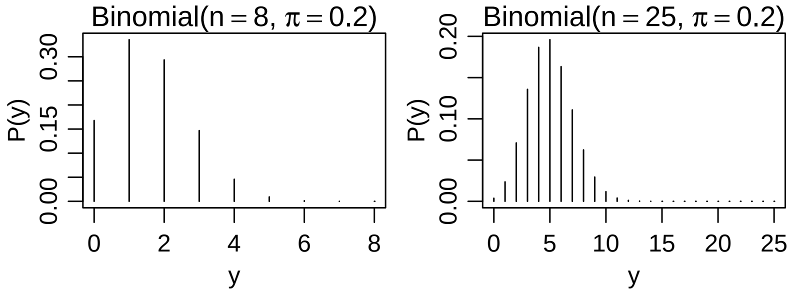
```
plot(0:3, dbinom(0:3, 3, .4), type = "h", xlab = "y", ylab = "P(y)")
```



Binomial Distribution Facts

If Y is a Binomial (n, π) random variable, then

- $\mathrm{E}(Y) = n\pi$
- SD $= \sigma(Y) = \sqrt{\mathrm{Var}(Y)} = \sqrt{n\pi(1-\pi)}$
- Binomial (n, π) can be approx. by Normal $(n\pi, n\pi(1-\pi))$ when n is large ($n\pi \geq 5$ and $n(1-\pi) \geq 5$).



Likelihood & Maximum Likelihood Estimation

A Probability Question

Let π be the proportion of US adults that are willing to get an Omicron booster.

A sample of 5 subjects are randomly selected. Let Y be the number of them that are willing to get an Omicron booster. What is $P(Y = 3)$?

Answer: Y is Binomial ($n = 5, \pi$) (Why?)

$$P(Y = y; \pi) = \frac{n!}{y!\,(n-y)!}\pi^y(1-\pi)^{n-y}$$

If π is known to be 0.3, then

$$P(Y = 3; \pi) = \frac{5!}{3!2!}(0.3)^3(0.7)^2 = 0.1323.$$

A Statistics Question

Of course, in practice we don't know π
and we collect data to estimate it.

How shall we choose a "good" estimator for π?

An *estimator* is a **formula** based on the data (a statistic) that we plan to use to estimate a parameter (π) after we collect the data.

Once the data are collected, we can calculate the **value** of the statistic: an *estimate* for π.

A Statistics Question

Suppose 8 of 20 randomly selected U.S. adults said they are willing to get an Omicron booster

What can we infer about the value of

$$\pi = \text{proportion of U.S. adults that are comfortable getting a booster?}$$

The chance to observe $Y = 8$ in a random sample of size $n = 20$ is

$$P(Y = 8; \pi) = \begin{cases} \binom{20}{8}(0.3)^8(0.7)^{12} \approx 0.1143 & \text{if } \pi = 0.3 \\ \binom{20}{8}(0.6)^8(0.4)^{12} \approx 0.0354 & \text{if } \pi = 0.6 \end{cases}$$

It appears that $\pi = 0.3$ is **more likely** to be π than $\pi = 0.6$, since the former gives a higher prob. to observe the outcome $y = 8$.

We say the *likelihood* of $\pi = 0.3$ is higher than that of $\pi = 0.6$.

Maximum Likelihood Estimate (MLE)

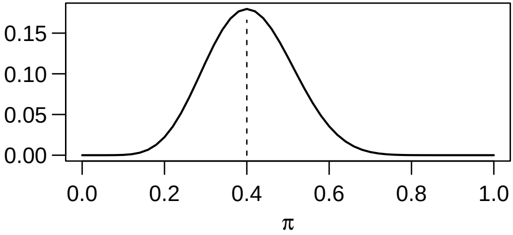
The *maximum likelihood estimate* (*MLE*) of a parameter (like π) is the value at which the likelihood function is maximized.

Example. If 8 of 20 randomly selected U.S. adults are comfortable getting the booster, the likelihood function

$$\ell(\pi \mid y = 8) = \binom{20}{8} \pi^8 (1 - \pi)^{12}$$

reaches its max at $\pi = 0.4$,

the MLE for π is $\widehat{\pi} = 0.4$ given the data $y = 8$.



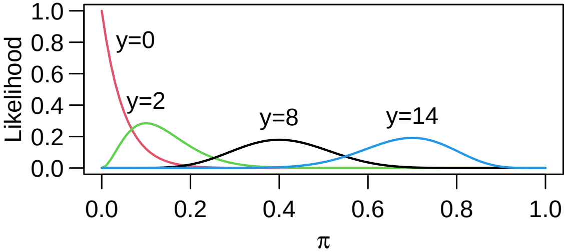
Maximum Likelihood Estimate (MLE)

The probability

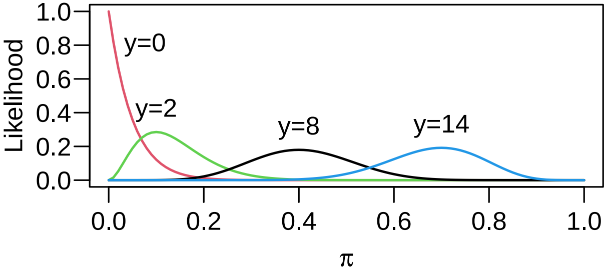
$$P(Y = y; \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} = \ell(\pi \mid y)$$

viewed as a function of π, is called the *likelihood function*,
(or just **likelihood**) of π, denoted as $\ell(\pi \mid y)$.

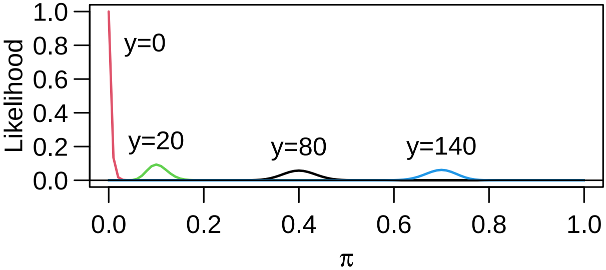
It measure the "plausibility" of a value being the true value of π.



Likelihood functions $\ell(\pi \mid y)$ at different values of y for $n = 20$.



Likelihood functions $\ell(\pi \mid y)$ for various values of y when $n = 20$.



Likelihood functions $\ell(\pi \mid y)$ at various values of y when $n = 200$.

Likelihood in General

In general, suppose the observed data $(Y_1, Y_2, \ldots, Y_n)$ have a joint probability distribution with some parameter(s) called θ

$$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = f(y_1, y_2, \ldots, y_n \mid \theta)$$

The *likelihood function* for the parameterθ is

$$\ell(\theta \mid \text{data}) = \ell(\theta \mid y_1, y_2, \ldots, y_n) = f(y_1, y_2, \ldots, y_n \mid \theta).$$

- Note the likelihood function regards the probability as a function of the parameter θ rather than as a function of the data $y_1, y_2, \ldots, y_n$.
- If

$$\ell(\theta_1 \mid y_1, \ldots, y_n) > \ell(\theta_2 \mid y_1, \ldots, y_n),$$

then θ_1 appears more plausible to be the true value of θ than θ_2 does, given the observed data $y_1, \ldots, y_n$.

Maximizing the Log-likelihood

Rather than maximizing the likelihood, it is often computationally easier to maximize its natural logarithm, called the *log-likelihood*,

$$\log \ell(\pi \mid y)$$

which results in the same answer since logarithm is strictly increasing,

$$x_1 > x_2 \iff \log(x_1) > \log(x_2).$$

So

$$\ell(\pi_1 \mid y) > \ell(\pi_2 \mid y) \iff \log \ell(\pi_1 \mid y) > \log \ell(\pi_2 \mid y).$$

Example (MLE for Binomial)

If the observed data $Y \sim$ Binomial (n, π) but π is unknown, the likelihood of π is

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

and the *log-likelihood* is

$$\log \ell(\pi \mid y) = \log \binom{n}{y} + y \log(\pi) + (n-y)\log(1-\pi).$$

From calculus, we know a function $f(x)$ reaches its max at $x = x_0$ if

$$\frac{d}{dx}f(x) = 0 \text{ at } x = x_0, \quad \text{and } \frac{d^2}{dx^2}f(x) < 0 \text{ at } x = x_0.$$

Example (MLE for Binomial)

$$\frac{d}{d\pi}\log\ell(\pi \mid y) = \frac{y}{\pi} - \frac{n-y}{1-\pi} = \frac{y-n\pi}{\pi(1-\pi).}$$

equals 0 when

$$\frac{y-n\pi}{\pi(1-\pi)} = 0$$

That is, when $y - n\pi = 0$.

Solving for π gives the ML estimator (MLE) $\boxed{\widehat{\pi} = \frac{y}{n}}$.

and $\frac{d^2}{d\pi^2}\log\ell(\pi \mid y) = -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2} < 0$ for any $0 < \pi < 1$

Thus, we know $\log\ell(\pi \mid y)$ reaches its max when $\pi = y/n$.

So MLE of π is $\widehat{\pi} = \frac{y}{n}$ = sample proportion of successes.

MLEs for Other Inference Problems

- If $Y_1, Y_2, \ldots, Y_n$ are i.i.d. $N(\mu, \sigma^2)$,

the MLE for μ is the **sample mean** $\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$.

- In simple linear regression,

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

When the errors ε_i are i.i.d. normal,
the usual **least squares estimates** for β_0 and β_1 are the MLEs.

i.i.d. = Independent and identically distributed
(same distribution each ε_i).

Hypothesis Tests of a Binomial Proportion

Hypothesis Tests of a Binomial Proportion

If the observed data $Y \sim$ Binomial (n, π), recall the MLE for π is

$$\hat{\pi} = Y/n.$$

Recall that since $Y \sim$ Binomial (n, π), the mean and standard deviation (SD) of Y are respectively,

$$\mathrm{E}[Y] = n\pi, \qquad \mathrm{SD}(Y) = \sqrt{n\pi(1-\pi)}.$$

The mean and SD of $\hat{\pi}$ are thus respectively

$$\mathrm{E}(\hat{\pi}) = \mathrm{E}\left(\frac{Y}{n}\right) = \frac{\mathrm{E}(Y)}{n} = \pi,$$

$$\mathrm{SD}(\hat{\pi}) = \mathrm{SD}\left(\frac{Y}{n}\right) = \frac{\mathrm{SD}(Y)}{n} = \sqrt{\frac{\pi(1-\pi)}{n}}.$$

By CLT, as n gets large, $\dfrac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}} \sim N(0, 1)$.

Hypothesis Tests for a Binomial Proportion

The textbook lists 3 different tests for testing

$$\text{H}_0\text{: } \pi = \pi_0 \text{ v.s. } \text{H}_a\text{: } \pi \neq \pi_0 \text{ (or 1-sided alternative.)}$$

- **Score Test** uses the *score statistic* $z_s = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}}$
- **Wald Test** uses the *Wald statistic* $z_w = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\hat{\pi}(1 - \hat{\pi})/n}}$
- **Likelihood Ratio Test**: we'll introduce shortly

As n gets large,

$$\text{both } z_s \text{ and } z_w \sim N(0, 1),$$
$$\text{both } z_s^2 \text{ and } z_w^2 \sim \chi^2_1.$$

based on which, P-value can be computed.

Example (Will You Get the COVID-19 Vaccine?)

Pew Research Institute surveyed 12,648 U.S. adults during Nov. 18-29, 2020 about their intention to be vaccinated for COVID-19. Among the 1264 respondents in the 18-29 age group, 695 said they would probably or definitely get the vaccine if it's available today.

- estimate of $\pi = \hat{\pi} = \dfrac{695}{1264} \approx 0.55$

Example (Will You Get the COVID-19 Vaccine?)

Pew Research Institute surveyed 12,648 U.S. adults during Nov. 18-29, 2020 about their intention to be vaccinated for COVID-19. Among the 1264 respondents in the 18-29 age group, 695 said they would probably or definitely get the vaccine if it's available today.

- estimate of $\pi = \hat{\pi} = \dfrac{695}{1264} \approx 0.55$

Want to test whether 60% of 18-29 year-olds in the U.S. would probably or definitely get the vaccine.

$$\text{H}_0\text{: } \pi = 0.6 \text{ v.s. } \text{H}_a\text{: } \pi \neq 0.6$$

- Score statistic $z_s = \dfrac{0.55 - 0.6}{\sqrt{0.6 \times 0.4/1264}} \approx -3.64$
- Wald statistic $z_w = \dfrac{0.55 - 0.6}{\sqrt{0.55 \times 0.45/1264}} \approx -3.58$

Note that the P-values computed using $N(0, 1)$ or χ_1^2 are identical.

P-value for the score test

```
2*pnorm(-3.64)
[1] 0.0002726
pchisq(3.64^2,df=1,lower.tail=F)
[1] 0.0002726
```

P-value for the Wald test

```
2*pnorm(-3.58)
[1] 0.0003436
pchisq(3.58^2,df=1,lower.tail=F)
[1] 0.0003436
```

See slides `L01_supp_chisq_table.pdf` for more details about chi-squared distributions.

Likelihood Ratio Test (LRT)

Recall the likelihood function for a binomial proportion π is

$$\ell(\pi|y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}.$$

To test H_0: $\pi = \pi_0$ v.s. H_a: $\pi \neq \pi_0$, let

- ℓ_0 be the max. likelihood under H_0, which is $\ell(\pi_0|y)$

Likelihood Ratio Test (LRT)

Recall the likelihood function for a binomial proportion π is

$$\ell(\pi|y) = \binom{n}{y}\pi^{y}(1-\pi)^{n-y}.$$

To test H_0: $\pi = \pi_0$ v.s. H_a: $\pi \neq \pi_0$, let

- ℓ_0 be the max. likelihood under H_0, which is $\ell(\pi_0|y)$
- ℓ_1 be the max. likelihood over all possible π, which is $\ell(\hat{\pi}|y)$ where $\hat{\pi} = y/n$ is the MLE of π.

Likelihood Ratio Test (LRT)

Recall the likelihood function for a binomial proportion π is

$$\ell(\pi|y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}.$$

To test H_0: $\pi = \pi_0$ v.s. H_a: $\pi \neq \pi_0$, let

- ℓ_0 be the max. likelihood under H_0, which is $\ell(\pi_0|y)$
- ℓ_1 be the max. likelihood over all possible π, which is $\ell(\hat{\pi}|y)$ where $\hat{\pi} = y/n$ is the MLE of π.

Observe that

- $\ell_0 \leq \ell_1$ always

Likelihood Ratio Test (LRT)

Recall the likelihood function for a binomial proportion π is

$$\ell(\pi|y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}.$$

To test H_0: $\pi = \pi_0$ v.s. H_a: $\pi \neq \pi_0$, let

- ℓ_0 be the max. likelihood under H_0, which is $\ell(\pi_0|y)$
- ℓ_1 be the max. likelihood over all possible π, which is $\ell(\hat{\pi}|y)$ where $\hat{\pi} = y/n$ is the MLE of π.

Observe that

- $\ell_0 \leq \ell_1$ always
- Under H_0, we expect $\hat{\pi} \approx \pi_0$ and hence $\ell_0 \approx \ell_1$.

Likelihood Ratio Test (LRT)

Recall the likelihood function for a binomial proportion π is

$$\ell(\pi|y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}.$$

To test H_0: $\pi = \pi_0$ v.s. H_a: $\pi \neq \pi_0$, let

- ℓ_0 be the max. likelihood under H_0, which is $\ell(\pi_0|y)$
- ℓ_1 be the max. likelihood over all possible π, which is $\ell(\hat{\pi}|y)$ where $\hat{\pi} = y/n$ is the MLE of π.

Observe that

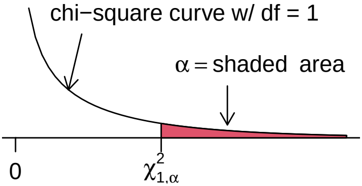
- $\ell_0 \leq \ell_1$ always
- Under H_0, we expect $\hat{\pi} \approx \pi_0$ and hence $\ell_0 \approx \ell_1$.
- $\ell_0 \ll \ell_1$ is a sign to reject H_0

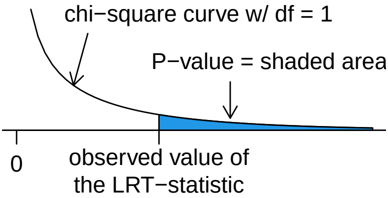
Likelihood Ratio Test Statistic (LRT Statistic)

The *likelihood-ratio test statistic* (LRT statistic) for testing H_0: $\pi = \pi_0$ v.s. H_a: $\pi \neq \pi_0$ equals

$$-2\log(\ell_0/\ell_1).$$

- Here $\log$ is the natural log
- LRT statistic $-2\log(\ell_0/\ell_1)$ is always nonnegative since $\ell_0 \leq \ell_1$
- When n is large, $-2\log(\ell_0/\ell_1) \sim \chi^2_1$.
  - Reject H_0 at level α if $-2\log(\ell_0/\ell_1) > \chi^2_{1,\alpha}$ = `qchisq(1-alpha, df=1)`
  - P-value $= P(\chi^2_1 >$ observed LRT statistic)





Likelihood Ratio Test Statistic for a Binomial Proportion

Recall the likelihood function for a binomial proportion π is

$$\ell(\pi|y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}.$$

Thus

$$\frac{\ell_0}{\ell_1} = \frac{\binom{n}{y}\pi_0^y(1-\pi_0)^{n-y}}{\binom{n}{y}(\frac{y}{n})^y(1-(\frac{y}{n}))^{n-y}} = \left(\frac{n\pi_0}{y}\right)^y \left(\frac{n\,(1-\pi_0)}{n-y}\right)^{n-y}$$

and hence the LRT statistic is

$$\begin{aligned} -2\log(\ell_0/\ell_1) &= 2y\log\left(\frac{y}{n\pi_0}\right) + 2(n-y)\log\left(\frac{n-y}{n\,(1-\pi_0)}\right) \\ &= 2\left\{O_{yes}\times\left[\log\left(\frac{O_{yes}}{E_{yes}}\right)\right] + O_{no}\times\left[\log\left(\frac{O_{no}}{E_{no}}\right)\right]\right\} \end{aligned}$$

where $O_{yes} = y$ and $O_{no} = n - y$ are the observed counts of yes & no, and $E_{yes} = n\pi_0$ and $E_{no} = n(1-\pi_0)$ are the expected counts of yes & no under H_0.

Example (COVID-19 , Cont'd)

Among the 1264 respondents in the 18-29 age group , 695 answered "yes", 569 answered "no", so

$$O_{yes} = y = 695, \quad O_{no} = n - y = 569.$$

Under H_0: $\pi = 0.6$, we expect 60% of the 1264 subjects to answer "yes" and 40% to answer "no." Don't round $n\pi_0$ and $n(1 - \pi_0)$ to integers.

$$E_{yes} = n\pi_0 = 1264 \times 0.6 = 758.4,$$

$$E_{no} = n(1 - \pi_0) = 1264 \times 0.4 = 505.6.$$

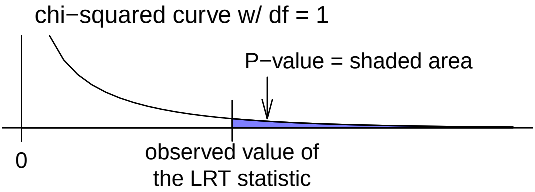
LRT statistic $= 2\left[695 \log\left(\frac{695}{758.4}\right) + 569 \log\left(\frac{569}{505.6}\right)\right] \approx 13.091$

which exceeds the critical value $\chi^2_{1,\alpha} = \chi^2_{1,0.05} = 3.84$ at $\alpha = 0.05$ and hence H_0 is rejected 5% level

```
qchisq(1-0.05, df=1)
[1] 3.841
```

P-value of LRT test of Porportions

Even though H_a is **two-sided**, the P-value remains to be the **upper tail** probability below, since a large deviation of $\widehat{\pi} = y/n$ from π_0 would lead to a large LRT statistic, no matter $\pi_0 > \widehat{\pi}$ or $\pi_0 < \widehat{\pi}$.



For the COVID-19 example, the P-value is $P(\chi_1^2 > 13.09)$, which is

```
pchisq(13.09, df=1, lower.tail=F)
[1] 0.0002969
```

Confidence Intervals for Binomial Proportions

Duality of Confidence Intervals and Significance Tests

For a 2-sided test of θ, the dual $100(1-\alpha)\%$ confidence interval (CI) for the parameter θ consists of all those θ^* values that a two-sided test of H$_0$: $\theta = \theta^*$ is not rejected at level α. E.g.,

- the dual 90% Wald CI for π is the collection of all π_0 such that a 2-sided Wald test of H$_0$: $\pi = \pi_0$ having a P-value $> 10\%$

Duality of Confidence Intervals and Significance Tests

For a 2-sided test of θ, the dual $100(1 - \alpha)\%$ confidence interval (CI) for the parameter θ consists of all those θ^* values that a two-sided test of H_0: $\theta = \theta^*$ is not rejected at level α. E.g.,

- the dual 90% Wald CI for π is the collection of all π_0 such that a 2-sided Wald test of H_0: $\pi = \pi_0$ having a P-value $> 10\%$
- the dual 95% score CI for π is the collection of all π_0 such that a 2-sided score test of H_0: $\pi = \pi_0$ having a P-value $> 5\%$

Duality of Confidence Intervals and Significance Tests

For a 2-sided test of θ, the dual $100(1-\alpha)\%$ confidence interval (CI) for the parameter θ consists of all those θ^* values that a two-sided test of H$_0$: $\theta = \theta^*$ is not rejected at level α. E.g.,

- the dual 90% Wald CI for π is the collection of all π_0 such that a 2-sided Wald test of H$_0$: $\pi = \pi_0$ having a P-value $> 10\%$
- the dual 95% score CI for π is the collection of all π_0 such that a 2-sided score test of H$_0$: $\pi = \pi_0$ having a P-value $> 5\%$

E.g., If the 2-sided P-value for testing H$_0$: $\pi = 0.2$ is 6%, then

- 0.2 is in the 95% CI

Duality of Confidence Intervals and Significance Tests

For a 2-sided test of θ, the dual $100(1-\alpha)\%$ confidence interval (CI) for the parameter θ consists of all those θ^* values that a two-sided test of H_0: $\theta = \theta^*$ is not rejected at level α. E.g.,

- the dual 90% Wald CI for π is the collection of all π_0 such that a 2-sided Wald test of H_0: $\pi = \pi_0$ having a P-value $> 10\%$
- the dual 95% score CI for π is the collection of all π_0 such that a 2-sided score test of H_0: $\pi = \pi_0$ having a P-value $> 5\%$

E.g., If the 2-sided P-value for testing H_0: $\pi = 0.2$ is 6%, then

- 0.2 is in the 95% CI
    - The corresponding α for a 95% CI is 5%. As p-value $= 6\% > \alpha = 5\%$, H_0: $\pi = 0.2$ is not rejected so 0.2 in the 95% CI.

Duality of Confidence Intervals and Significance Tests

For a 2-sided test of θ, the dual $100(1-\alpha)\%$ confidence interval (CI) for the parameter θ consists of all those θ^* values that a two-sided test of H_0: $\theta = \theta^*$ is not rejected at level α. E.g.,

- the dual 90% Wald CI for π is the collection of all π_0 such that a 2-sided Wald test of H_0: $\pi = \pi_0$ having a P-value $> 10\%$
- the dual 95% score CI for π is the collection of all π_0 such that a 2-sided score test of H_0: $\pi = \pi_0$ having a P-value $> 5\%$

E.g., If the 2-sided P-value for testing H_0: $\pi = 0.2$ is 6%, then

- 0.2 is in the 95% CI
  - The corresponding α for a 95% CI is 5%. As p-value $= 6\% > \alpha = 5\%$, H_0: $\pi = 0.2$ is not rejected so 0.2 in the 95% CI.
- but 0.2 is NOT in the 90% CI

Duality of Confidence Intervals and Significance Tests

For a 2-sided test of θ, the dual $100(1-\alpha)\%$ confidence interval (CI) for the parameter θ consists of all those θ^* values that a two-sided test of H_0: $\theta = \theta^*$ is not rejected at level α. E.g.,

- the dual 90% Wald CI for π is the collection of all π_0 such that a 2-sided Wald test of H_0: $\pi = \pi_0$ having a P-value $> 10\%$
- the dual 95% score CI for π is the collection of all π_0 such that a 2-sided score test of H_0: $\pi = \pi_0$ having a P-value $> 5\%$

E.g., If the 2-sided P-value for testing H_0: $\pi = 0.2$ is 6%, then

- 0.2 is in the 95% CI
    - The corresponding α for a 95% CI is 5%. As p-value $= 6\% > \alpha = 5\%$, H_0: $\pi = 0.2$ is not rejected so 0.2 in the 95% CI.
- but 0.2 is NOT in the 90% CI
    - The corresponding α for a 90% CI is 10%. As p-value $= 6\% < \alpha = 10\%$, H_0: $\pi = 0.2$ is rejected so 0.2 NOT in the 90% CI.

Wald Confidence Intervals (Wald CIs)

For a Wald test, H_0: $\pi = \pi^*$ is not rejected at level α if

$$\left| \frac{\hat{\pi} - \pi^*}{\sqrt{\hat{\pi}(1 - \hat{\pi})/n}} \right| < z_{\alpha/2},$$

so a $100(1 - \alpha)\%$ Wald CI is

$$\left(\hat{\pi} - z_{\alpha/2} \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}},\ \ \hat{\pi} + z_{\alpha/2} \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} \right).$$

where

confidence level $100(1 - \alpha)\%$	90%	95%	99%
$z_{\alpha/2}$	1.645	1.96	2.576

- Introduced in STAT 220 and 234

Drawbacks:

- Wald CI for π collapses whenever $\hat{\pi} = 0$ or 1.
- Actual coverage prob. for Wald CI is usually much less than $100(1 - \alpha)\%$ if π close to 0 or 1, unless n is quite large.

Score Confidence Intervals (Score CIs)

For a Score test, H_0 $\pi = \pi^*$ is not rejected at level α if

$$\left|\frac{\hat{\pi} - \pi^*}{\sqrt{\pi^*(1-\pi^*)/n}}\right| < z_{\alpha/2}.$$

A $100(1-\alpha)\%$ score confidence interval consists of those π^* satisfying the inequality above.

Example. If $\hat{\pi} = 0$, the 95% score CI consists of those π^* satisfying

$$\left|\frac{0 - \pi^*}{\sqrt{\pi^*(1-\pi^*)/n}}\right| < 1.96.$$

After a few steps of algebra, we can show such π^*'s are those satisfying $0 < \pi^* < \frac{1.96^2}{n+1.96^2}$. The 95% score CI for π when $\hat{\pi} = 0$ is thus

$$\left(0, \frac{1.96^2}{n + 1.96^2}\right),$$

which is NOT collapsing!

Score CI (Cont'd)

The end points of the score CI can be shown to be

$$\frac{(y + z^2/2) \pm z_{\alpha/2}\sqrt{n\hat{\pi}(1-\hat{\pi}) + z^2/4}}{n + z^2} \quad \text{where} \quad z = z_{\alpha/2}.$$

- midpoint of the score CI, $\dfrac{\hat{\pi} + z^2/2n}{1 + z^2/n}$, is between $\hat{\pi}$ and 0.5.
- better than the Wald CI, that the actual coverage probabilities are closer to the nominal levels.

Agresti-Coull Confidence Intervals

Recall the midpoint for a $100(1-\alpha)\%$ score CI is

$$\tilde{\pi} = \frac{y + z^2/2}{n + z^2}, \quad \text{where } z = z_{\alpha/2},$$

which looks as if we add $z^2/2$ more successes and $z^2/2$ more failures to the data before we estimate π.

This inspires the Agresti-Coull $100(1-\alpha)\%$ confidence interval:

$$\tilde{\pi} \pm z\sqrt{\frac{\tilde{\pi}(1-\tilde{\pi})}{n+z^2}} \quad \text{where } \tilde{\pi} = \frac{y + z^2/2}{n + z^2} \quad \text{and } z = z_{\alpha/2}.$$

which is essentially a Wald-type interval after adding $z^2/2$ more successes and $z^2/2$ more failures to the data, where $z = z_{\alpha/2}$.

95% "Plus-Four" Confidence Intervals

At 95% level, $z_{\alpha/2} = z_{0.025} = 1.96$, the midpoint of the Agresti-Coull CI is

$$\frac{y + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2} = \frac{y + 1.96^2/2}{n + 1.96^2} \approx \frac{y+2}{n+4}.$$

Hence some approximate the 95% Agresti-Coull correction to the Wald CI by **adding 2 successes and 2 failures** before computing $\hat{\pi}$ and then compute the Wald CI:

$$\hat{\pi}^* \pm 1.96\sqrt{\frac{\hat{\pi}^*(1-\hat{\pi}^*)}{n+4}}, \quad \text{where } \hat{\pi}^* = \frac{y+2}{n+4}.$$

- This is so called the "Plus-Four" confidence interval

95% "Plus-Four" Confidence Intervals

At 95% level, $z_{\alpha/2} = z_{0.025} = 1.96$, the midpoint of the Agresti-Coull CI is

$$\frac{y + z^2_{\alpha/2}/2}{n + z^2_{\alpha/2}} = \frac{y + 1.96^2/2}{n + 1.96^2} \approx \frac{y+2}{n+4}.$$

Hence some approximate the 95% Agresti-Coull correction to the Wald CI by **adding 2 successes and 2 failures** before computing $\hat{\pi}$ and then compute the Wald CI:

$$\hat{\pi}^* \pm 1.96\sqrt{\frac{\hat{\pi}^*(1-\hat{\pi}^*)}{n+4}}, \quad \text{where } \hat{\pi}^* = \frac{y+2}{n+4}.$$

- This is so called the "Plus-Four" confidence interval
- Note the "Plus-Four" CI is for 95% confidence level only

95% "Plus-Four" Confidence Intervals

At 95% level, $z_{\alpha/2} = z_{0.025} = 1.96$, the midpoint of the Agresti-Coull CI is

$$\frac{y + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2} = \frac{y + 1.96^2/2}{n + 1.96^2} \approx \frac{y+2}{n+4}.$$

Hence some approximate the 95% Agresti-Coull correction to the Wald CI by **adding 2 successes and 2 failures** before computing $\hat{\pi}$ and then compute the Wald CI:

$$\hat{\pi}^* \pm 1.96\sqrt{\frac{\hat{\pi}^*(1-\hat{\pi}^*)}{n+4}}, \quad \text{where } \hat{\pi}^* = \frac{y+2}{n+4}.$$

- This is so called the "Plus-Four" confidence interval
- Note the "Plus-Four" CI is for 95% confidence level only
- At 90% level, $z_{\alpha/2} = z_{0.05} = 1.645$, Agresti-Coull CI would add $z_{\alpha/2}^2/2 = 1.645^2/2 \approx 1.35$ more successes and 1.35 more failures.

Likelihood Ratio Confidence Intervals (LR CIs)

A LR test will not reject H_0: $\pi = \pi^*$ at level α if

$$-2\log(\ell_0/\ell_1) = -2\log\left(\frac{\ell(\pi^*|y)}{\ell(\hat{\pi}\,|\,y)}\right) < \chi^2_{1,\alpha}.$$

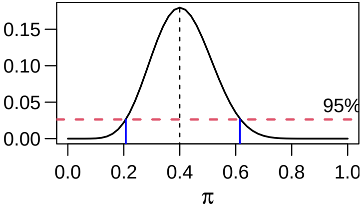
A $100(1-\alpha)\%$ likelihood ratio CI consists of those π^* with likelihood

$$\ell(\pi^*|y) > e^{-\chi^2_{1,\alpha}/2}\ell(\hat{\pi}|y)$$

E.g., the 95% LR CI contains those π^* with likelihood above $e^{-\chi^2_{1,0.05}/2} = e^{-3.84/2} \approx 0.0147$ multiple of the max. likelihood.

- No close form expression for end points of a LR CI
- Can use software to find the end points numerically

Likelihood $\ell(\pi|y)$ for $n = 20$, $y = 8$.



Likelihood Ratio Confidence Intervals Do Not Collapse at 0

Recall the LRT statistic for testing H_0: $\pi = \pi_0$ against H_a: $\pi \neq \pi_0$ is

$$-2\log(\ell_0/\ell_1) = 2y\log\left(\frac{y}{n\pi_0}\right) + 2(n-y)\log\left(\frac{n-y}{n(1-\pi_0)}\right)$$

and the H_0: $\pi = \pi_0$ is rejected if $-2\log(\ell_0/\ell_1) > \chi^2_{1,\alpha}$. Hence the $100(1-\alpha)\%$ LR confidence interval consists of those π_0 satisfying

$$2y\log\left(\frac{y}{n\pi_0}\right) + 2(n-y)\log\left(\frac{n-y}{n(1-\pi_0)}\right) \leq \chi^2_{1,\alpha}$$

In particular, when $y = 0$, the 95% LR CI consists of those π_0 satisfying

$$-2n\log(1-\pi_0) < \chi^2_{1,0.05} = 3.84.$$

That is, $(0, 1 - e^{-3.84/(2n)})$, which is NOT collapsing, either!

Example (Political Party Affiliation)

A survey about the political party affiliation of residents in a town found 4 of 400 in the sample to be Independents.

Want a 95% CI for π = proportion of Independents in the town.

- estimate of $\pi = 4/400 = 0.01$
- Wald CI: $0.01 \pm 1.96\sqrt{\dfrac{0.01 \times (1-0.01)}{400}} \approx (0.00025, 0.01975)$.
- 95% Score CI contains those π^* satisfying

$$\frac{0.01 - \pi^*}{\sqrt{\pi^*(1-\pi^*)/400}} < 1.96$$

  which is the interval $(0.0039, 0.0254)$.
- 95% Agresti-Coull CI: adding $z^2/2 = z^2_{0.05}/2 = 1.96^2/2 \approx 1.92$. The estimate of π is $(4 + 1.92)/(400 + 3.84) \approx 0.01466$

$$0.01466 \pm 1.96\sqrt{\frac{0.01466 \times (1 - 0.01466)}{403.84}} \approx (0.00294, 0.02638).$$

R Function “prop.test()” for Score Test and CI

The R function `prop.test()` performs the **score test** and produces the **score CI**.

- It test H_0: $\pi = 0.5$ vs H_a: $\pi \neq 0.5$ by default
- Uses continuity correction by default.

```
prop.test(4,400)


    1-sample proportions test with continuity correction


data:  4 out of 400, null probability 0.5
X-squared = 382, df = 1, p-value <2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.003208 0.027187
sample estimates:
   p
0.01
```

R Function “prop.test()” for Score Test and CI

To perform a score test of H_0: $\pi = 0.02$ vs H_a: $\pi \neq 0.02$ **without** the continuity correction . . .

```
prop.test(4,400, p=0.02, correct=F)


    1-sample proportions test without continuity correction

data:  4 out of 400, null probability 0.02
X-squared = 2, df = 1, p-value = 0.2
alternative hypothesis: true p is not equal to 0.02
95 percent confidence interval:
 0.003895 0.025427
sample estimates:
   p
0.01
```

The 95% CI matches the score CI computed earlier.

R function for Other CIs of Binomial Proportions

The function `binom.confint()` in the package `binom` can produce confidence intervals for several methods.

You need to first install the `binom` package **just once, ever**.

To check if the `binom` package has installed on your computer,

```
library(binom)
```

If you get an error message,

```
# Error in library(binom) : there is no package called 'binom'
```

that means the `binom` library is not installed. You can run the following command to install the `binom` library.

If `FALSE`, you can install the library using the command below

```
install.packages("binom")
```

Now one can use `binom.confint()` to find the CIs.

```
# Wald CI
binom.confint(4, 400, conf.level = 0.95, method = "asymptotic")
      method x   n mean      lower   upper
1 asymptotic 4 400 0.01 0.0002493 0.01975

# Score CI, also called ``Wilson"
binom.confint(4, 400, conf.level = 0.95, method = "wilson")
  method x   n mean    lower   upper
1 wilson 4 400 0.01 0.003895 0.02543

# Agresti-Coull CI
binom.confint(4, 400, conf.level = 0.95, method = "ac")
         method x   n mean    lower   upper
1 agresti-coull 4 400 0.01 0.002939 0.02638

# Likelihood-Ratio Test CI
binom.confint(4, 400, conf.level = 0.95, method = "lrt")
  method x   n mean    lower   upper
1    lrt 4 400 0.01 0.003136 0.02308
```

Example (Political Party Affiliation) LR CI

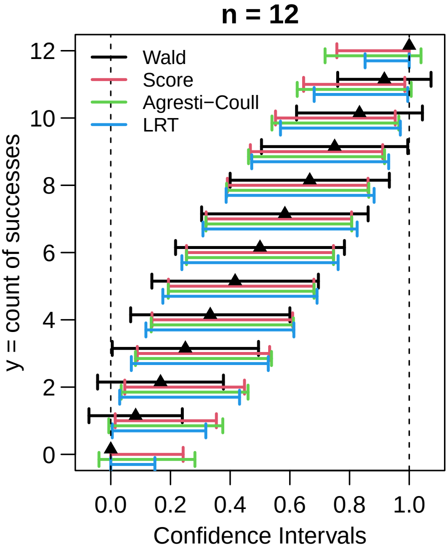
Recall the 95% LR confidence interval consists of those π_0 satisfying

$$2y\log\left(\frac{y}{n\pi_0}\right) + 2(n-y)\log\left(\frac{n-y}{n\left(1-\pi_0\right)}\right) \leq \chi^2_{1,0.05} = 3.8415$$

To verify the LRT confidence interval $(0.003135542, 0.02307655)$ given by `binom.confint()`, let's plug the end points in to the LRT test statistic above and see if we obtain 3.84146

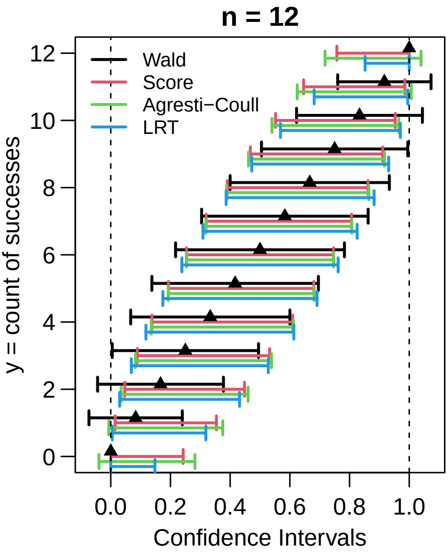
```
y = 4
n = 400
pi0 = c(0.003135542, 0.02307655)
2*y*log(y/n/pi0) + 2*(n-y)*log((n-y)/n/(1-pi0))
[1] 3.806 3.841
pi0 = c(0.003115255, 0.02307735)
2*y*log(y/n/pi0) + 2*(n-y)*log((n-y)/n/(1-pi0))
[1] 3.841 3.841
```

Comparison of Wald, Score, Agresti-Coull, and LRT CIs



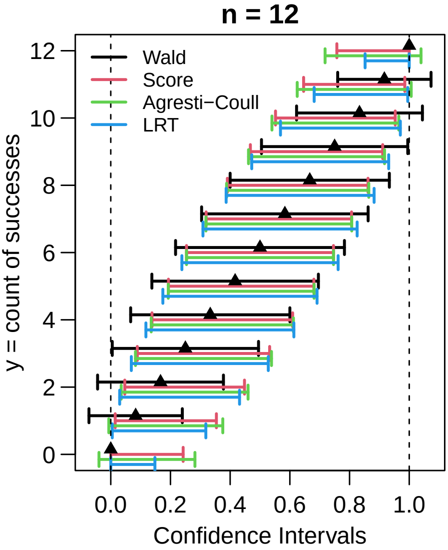
- End points of Score, Agresti-Coull, and LRT CIs are generally closer to 0.5 than those for the Wald CIs

Comparison of Wald, Score, Agresti-Coull, and LRT CIs



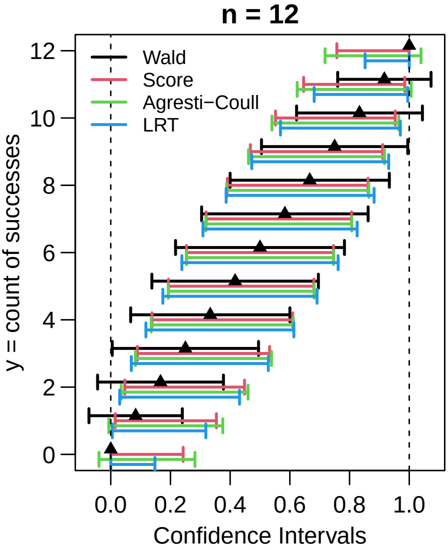
- End points of Score, Agresti-Coull, and LRT CIs are generally closer to 0.5 than those for the Wald CIs
- End points of Wald and Agresti-Coull CIs may fall outside of $[0, 1]$, while those of Score and LRT CIs always fall between 0 and 1

Comparison of Wald, Score, Agresti-Coull, and LRT CIs



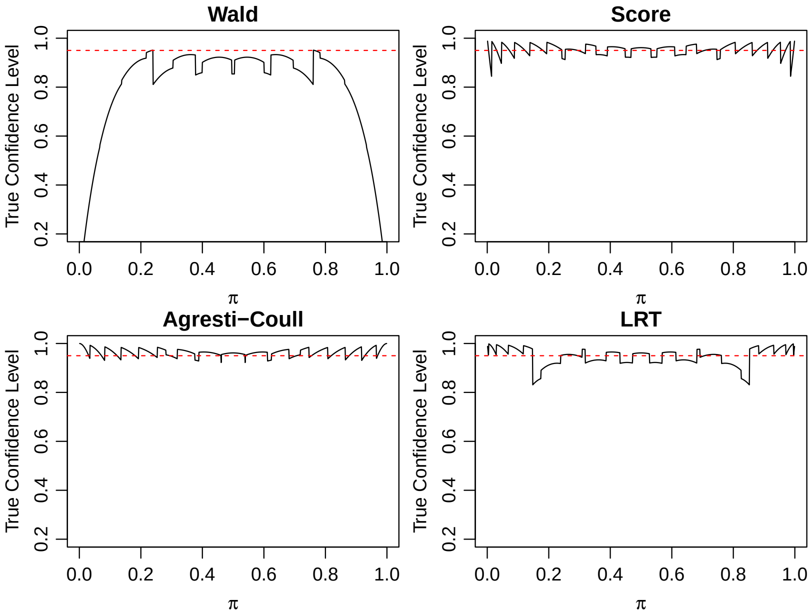
- End points of Score, Agresti-Coull, and LRT CIs are generally closer to 0.5 than those for the Wald CIs
- End points of Wald and Agresti-Coull CIs may fall outside of $[0, 1]$, while those of Score and LRT CIs always fall between 0 and 1
- Agresti-Coull CIs always contain the Score CIs

Comparison of Wald, Score, Agresti-Coull, and LRT CIs

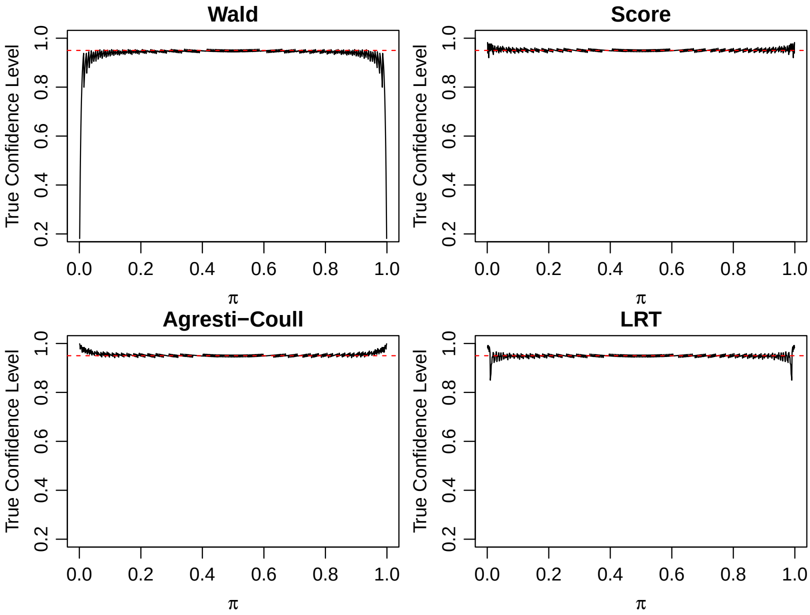


- End points of Score, Agresti-Coull, and LRT CIs are generally closer to 0.5 than those for the Wald CIs
- End points of Wald and Agresti-Coull CIs may fall outside of $[0, 1]$, while those of Score and LRT CIs always fall between 0 and 1
- Agresti-Coull CIs always contain the Score CIs
- Score CIs are narrower than Wald CIs unless y/n is close to 0 or 1.

True Confidence Levels for Various Types of CIs When $n = 12$



True Coverage Probabilities for Various CIs When $n = 200$



True Confidence Levels of Various CIs

- How are true confidence levels computed? Why do the curves look jumpy? See HW2.

True Confidence Levels of Various CIs

- How are true confidence levels computed? Why do the curves look jumpy? See HW2.
- Wald CIs tend to be farthest below the 0.95 level. In fact, the true level can be as low as 0 when π is close to 0 or 1

True Confidence Levels of Various CIs

- How are true confidence levels computed? Why do the curves look jumpy? See HW2.
- Wald CIs tend to be farthest below the 0.95 level. In fact, the true level can be as low as 0 when π is close to 0 or 1
- Score CIs are closer to the 0.95 level, though it may fall below 0.95 when π is close to 0 or 1

True Confidence Levels of Various CIs

- How are true confidence levels computed? Why do the curves look jumpy? See HW2.
- Wald CIs tend to be farthest below the 0.95 level. In fact, the true level can be as low as 0 when π is close to 0 or 1
- Score CIs are closer to the 0.95 level, though it may fall below 0.95 when π is close to 0 or 1
- Agresti-Coull CIs are usually conservative (true level are above 0.95) especially when π close to 0 or 1.

True Confidence Levels of Various CIs

- How are true confidence levels computed? Why do the curves look jumpy? See HW2.
- Wald CIs tend to be farthest below the 0.95 level. In fact, the true level can be as low as 0 when π is close to 0 or 1
- Score CIs are closer to the 0.95 level, though it may fall below 0.95 when π is close to 0 or 1
- Agresti-Coull CIs are usually conservative (true level are above 0.95) especially when π close to 0 or 1.
- LRT CIs are better than Wald but generally not as good as Score or Agresti-Coull CIs

True Confidence Levels of Various CIs

- How are true confidence levels computed? Why do the curves look jumpy? See HW2.
- Wald CIs tend to be farthest below the 0.95 level. In fact, the true level can be as low as 0 when π is close to 0 or 1
- Score CIs are closer to the 0.95 level, though it may fall below 0.95 when π is close to 0 or 1
- Agresti-Coull CIs are usually conservative (true level are above 0.95) especially when π close to 0 or 1.
- LRT CIs are better than Wald but generally not as good as Score or Agresti-Coull CIs
- When n gets larger, all 4 types of intervals become closer to the 0.95 level, though Wald CIs remain poor when π is close to 0 or 1

How To Compute the True Confidence Levels? (1)

Consider the true confidence level the 95% Wald CI when $n = 12$ and $\pi = 0.1$, i.e., the probability that the 95% Wald confidence interval (Wald CI) below

$$\left(\hat{\pi} - 1.96\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}},\ \ \hat{\pi} + 1.96\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right) \quad \text{where } \hat{\pi} = y/n$$

contains $\pi = 0.1$ when $y \sim \text{Binomial}(n = 12, \pi = 0.1)$.

If y has a Binomial$(n = 12, \pi = 0.1)$ distribution, the possible values of y are the integers $0, 1, 2, \ldots, 12$.

We can calculate the corresponding Wald CI for each possible value of y on the next page.

See also: https://yibi-huang.shinyapps.io/shiny/

```r
n = 12
y = 0:n
p = y/n
CI.lower = p - 1.96*sqrt(p*(1-p)/n)
CI.upper = p + 1.96*sqrt(p*(1-p)/n)
data.frame(y, CI.lower, CI.upper)
    y CI.lower CI.upper
1   0  0.00000   0.0000
2   1 -0.07305   0.2397
3   2 -0.04420   0.3775
4   3  0.00500   0.4950
5   4  0.06661   0.6001
6   5  0.13772   0.6956
7   6  0.21710   0.7829
8   7  0.30439   0.8623
9   8  0.39994   0.9334
10  9  0.50500   0.9950
11 10  0.62247   1.0442
12 11  0.76029   1.0730
13 12  1.00000   1.0000
```

Which of the Wald intervals contain $\pi = 0.1$?

```r
n = 12
y = 0:n
p = y/n
CI.lower = p - 1.96*sqrt(p*(1-p)/n)
CI.upper = p + 1.96*sqrt(p*(1-p)/n)
data.frame(y, CI.lower, CI.upper)
    y CI.lower CI.upper
1   0  0.00000   0.0000
2   1 -0.07305   0.2397
3   2 -0.04420   0.3775
4   3  0.00500   0.4950
5   4  0.06661   0.6001
6   5  0.13772   0.6956
7   6  0.21710   0.7829
8   7  0.30439   0.8623
9   8  0.39994   0.9334
10  9  0.50500   0.9950
11 10  0.62247   1.0442
12 11  0.76029   1.0730
13 12  1.00000   1.0000
```

Which of the Wald intervals contain $\pi = 0.1$?

Only the CIs for $y = 1, 2, 3, 4$.

When $y \sim$ Binomial$(n = 12, \pi = 0.1)$,

$$P(\text{95\% Wald CI contains } \pi = 0.1)$$

$$= P(y = 1) + P(y = 2) + P(y = 3) + P(y = 4)$$

$$= \binom{12}{1}(0.1)^1(0.9)^{11} + \binom{12}{2}(0.1)^2(0.9)^{10} + \binom{12}{3}(0.1)^3(0.9)^9 + \binom{12}{4}(0.1)^4(0.9)^8.$$

The four Binomial probabilities above can be found using

```
dbinom(1:4, size = 12, p=0.1)
[1] 0.37657 0.23013 0.08523 0.02131
```

and hence their total is

```
sum(dbinom(1:4, size = 12, p=0.1))
[1] 0.7132
```

The true confidence level of a 95% Wald CI is just 71%, far below the nominal 95% level.