Chapter 1 Introduction

Yibi Huang
Department of Statistics
University of Chicago

Outline

- Variable types
- Review of binomial and multinomial distributions
- Likelihood and maximum likelihood method
- Inference for a binomial proportion (Wald, score and likelihood ratio tests and confidence intervals)
- Sample sample inference

Variable Types

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

- e.g., temperature, blood pressure, heights, speeds, income
- Stat 22200, Stat 22400

Methods in *categorical data analysis* are used when the response variable takes **categorical** (or **qualitative**) values

- e.g.,
  - gender (male, female),
  - political philosophy (liberal, moderate, conservative),
  - region (metropolitan, urban, suburban, rural)
- Stat 22600

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

Two Types of 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 and Multinomial Distributions

Binomial Distributions (Review)

If n Bernoulli trials are performed:

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

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

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

The probability function of Y is

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

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

$$m! = \text{“}m \text{ factorial”} = m \times (m-1) \times (m-2) \times \cdots \times 1.$$

Note that $0! = 1$.

Example

Vote (Dem, Rep). Suppose $\pi = Pr(Dem) = 0.4$.

Sample $n = 3$ voters, let y = number of Dem votes among them.

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

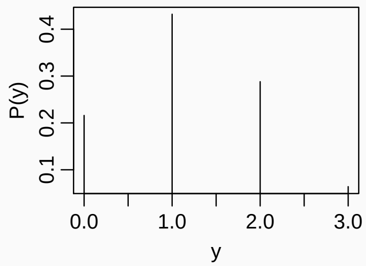
$$P(0) = \frac{3!}{0!3!}(0.4)^0(0.6)^3 = (0.6)^3 = 0.216$$

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

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

R Codes

```
> 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(0:3, 3, 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)")
```

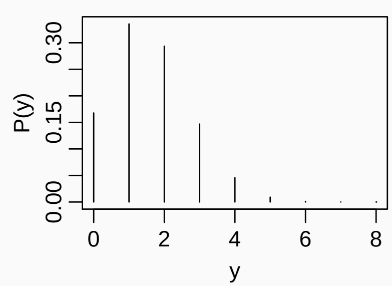


Facts About the Binomial Distribution

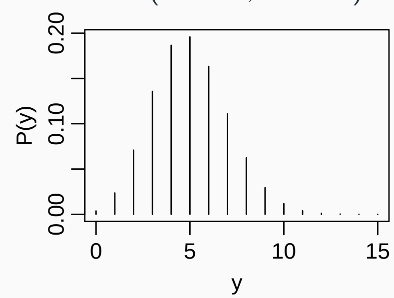
If Y is a binomial (n, π) random variable, then

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

binomial$(n = 8, \pi = 0.2)$



binomial$(n = 25, \pi = 0.2)$



Multinomial Distribution — Generalization of Binomial

If n trials are performed:

- in each trial there are $c > 2$ possible outcomes (categories)
- $\pi_i = \mathrm{P}(\text{category } i)$, for each trial, $\sum_{i=1}^{c} \pi_i = 1$
- trials are **independent**
- Y_i = number of trials fall in category i out of n trials

then the joint distribution of $(Y_1, Y_2, \ldots, Y_c)$ has a **multinomial distribution**, with probability function

$$\mathrm{P}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_c = y_c) = \frac{n!}{y_1!\, y_2! \cdots y_c!} \pi_1^{y_1} \pi_2^{y_2} \cdots \pi_c^{y_c}$$

where $0 \le y_i \le n$ for all i and $\sum_i y_i = n$.

Example

Suppose proportions of individuals with genotypes AA, Aa, and aa in a large population are

$$(\pi_{AA}, \pi_{Aa}, \pi_{aa}) = (0.25, 0.5, 0.25).$$

Randomly sample $n = 5$ individuals from the population.

The chance of getting 2 AA's, 2 Aa's, and 1 aa is

$$\begin{aligned}\mathrm{P}(Y_{AA} = 2, Y_{Aa} = 2, Y_{aa} = 1) &= \frac{5!}{2!\,2!\,1!}(0.25)^2(0.5)^2(0.25)^1 \\ &= \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1)(2 \cdot 1)(1)}(0.25)^2(0.5)^2(0.25)^1 \approx 0.117\end{aligned}$$

and the chance of getting no AA, 3 Aa's, and 2 aa's is

$$\begin{aligned}\mathrm{P}(Y_{AA} = 0, Y_{Aa} = 3, Y_{aa} = 2) &= \frac{5!}{0!\,3!\,2!}(0.25)^0(0.5)^3(0.25)^2 \\ &= \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(1)(3 \cdot 2 \cdot 1)(2 \cdot 1)}(0.25)^0(0.5)^3(0.25)^2 \approx 0.078\end{aligned}$$

Facts About the Multinomial Distribution

If $(Y_1, Y_2, \ldots, Y_c)$ has a multinomial distribution with n trials and category probabilities $(\pi_1, \pi_2, \cdots, \pi_c)$, then

- $\mathrm{E}(Y_i) = n\pi_i$ for $i = 1, 2, \ldots, c$
- $\sigma(Y_i) = \sqrt{\mathrm{Var}(Y_i)} = \sqrt{n\pi_i(1-\pi_i)}$,
- $\mathrm{Cov}(Y_i, Y_j) = -n\pi_i\pi_j$

Likelihood and Maximum Likelihood Estimation

A Probability Question

A push pin is tossed $n = 5$ times. Let Y be the number of times the push pin lands on its head. What is $\mathrm{P}(Y = 3)$?

Answer. As the tosses are indep., Y is binomial $(n = 5, \pi)$

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

where $\pi = \mathrm{P}$(push pin lands on its head in a toss).

If π is known to be 0.4, then

$$\mathrm{P}(Y = 3; \pi) = \frac{5!}{3!2!}(0.4)^3(0.6)^2 = 0.2304.$$

A Statistics Question

Suppose a push pin is observed to land on its head $Y = 8$ times in $n = 20$ tosses. Can we infer about the value of

$$\pi = \mathrm{P}(\text{push pin lands on its head in a toss})?$$

The chance to observe $Y = 8$ in $n = 20$ tosses is

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

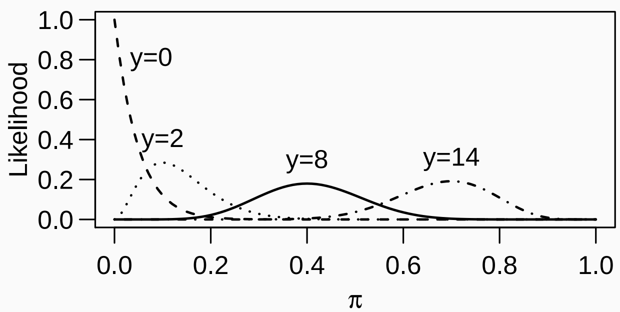
It appears that $\pi = 0.3$ is more likely than $\pi = 0.6$, since the former gives a higher prob. to observe the outcome $Y = 8$.

The probability

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

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

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



Curves for the likelihood $\ell(\pi|y)$ at different values of y for $n = 20$.

Likelihood

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

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

The *likelihood function* for the parameterθ is

$$\ell(\theta) = \ell(\theta|y_1, y_2, \ldots, y_n) = f(y_1, y_2, \ldots, y_n|\theta).$$

- 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

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

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

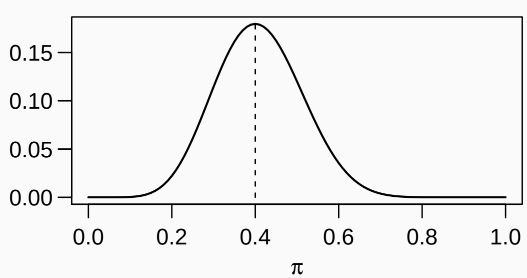
Maximum Likelihood Estimate (MLE)

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

Example. If a push pin lands on head $Y = 8$ times in $n = 20$ tosses, the likelihood function

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

reach its maximum at $\pi = 0.4$, the MLE for π is $\widehat{\pi} = 0.4$ given the data $Y = 8$.



Maximizing the Log-likelihood

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

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

which is equivalent since *logarithm* is *strictly increasing*,

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

So

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

Example (MLE for Binomial)

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

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

and the *log-likelihood* is

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

From calculus, we know a function $f(x)$ reaches its max at $x = x_0$ if $\frac{d}{dx}f(x) = 0$ at $x = x_0$ and $\frac{d^2}{dx^2}f(x) < 0$ at $x = x_0$. As

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

equals 0 when $\pi = y/n$ and $\frac{d^2}{d\pi^2}\log \ell(\pi|y) = -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2} < 0$ is always true, we know $\log \ell(\pi|y)$ reaches its max when $\pi = y/n$. So the MLE of π is y/n.

More Facts about MLEs

- If $Y_1, Y_2, \ldots, Y_n$ are i.i.d. $N(\mu, \sigma^2)$, the MLE of μ is the **sample mean** $\sum_{i=1}^n Y_i/n$.
- In ordinary linear regression,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

when the noise ε_i are i.i.d. normal, the usual **least squares estimates** for $\beta_0, \beta_1, \ldots, \beta_p$ are MLEs.

Large Sample Optimality of MLEs

MLEs are not always the best estimators but they have a number of good properties.

MLEs are

- asymptotically unbiased — the bias of MLE approaches 0 as the sample size n gets large,
- asymptotically efficient — no other estimates have smaller limiting variance than the MLE as n gets large
- asymptotically normal — the large sample distribution of the MLE is approx. normal.

All the above are true under most circumstances, though sometimes the sample size required can be quite large.

Computation Issues of MLEs. In many cases, the MLEs can not be solve directly (no analytical expression exists), and numerical tools are needed to compute the values of the MLEs.

Inference for a Binomial Proportion

Inference for 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,

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

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

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

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

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

Significance Test for a Binomial Proportion

The text lists 3 different tests for testing

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

- **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 will explain later

As n gets large,

$$\text{both } z_s \text{ and } z_w \sim N(0,1),$$
$$\text{both } z_s^2 \text{ and } z_w^2 \sim \chi_1^2.$$

based on which, P-value can be computed.

Example (U.S. in Another World War)

When 2004 General Social Survey asked subjects "*do you expect the U.S. to fight in another world war within the next 10 years?*" 460 of 828 subjects answered "yes". Want to test if $\pi = 0.5$ where π is the proportion of the population that would answered"yes".

- estimate of $\pi = \hat{\pi} = 460/828 \approx 0.556$
- Score statistic $z_s = \dfrac{0.556 - 0.5}{\sqrt{0.5 \times 0.5/828}} = 3.22,$
- Wald statistic $z_w = \dfrac{0.556 - 0.5}{\sqrt{0.556 \times 0.444/828}} \approx 3.24,$

Example (U.S. in Another World War)

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

```
> 2*pnorm(3.22,lower.tail=F)        #P-value for score test
[1] 0.001281906
> pchisq(3.22^2,df=1,lower.tail=F)
[1] 0.001281906

> 2*pnorm(3.24,lower.tail=F)        #P-value for Wald test
[1] 0.001195297
> pchisq(3.24^2,df=1,lower.tail=F)
[1] 0.001195297
```

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 true
- 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

$$H_0\colon \pi = \pi_0 \text{ v.s. } H_a\colon \pi \neq \pi_0$$

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 $-2\log(\ell_0/\ell_1) > \chi^2_{1,\alpha}$
  - P-value $= \mathrm{P}(\chi^2_1 >$ 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

$$\begin{aligned}-2\log(\ell_0/\ell_1) &= 2y\log\left(\frac{y}{n\pi_0}\right) + 2(n-y)\log\left(\frac{n-y}{n(1-\pi_0)}\right)\\ &= 2\sum_{i=\text{yes,no}} \text{Observed}_i \times \left[\log\left(\frac{\text{Observed}_i}{\text{Fitted}_i}\right)\right]\end{aligned}$$

where $\text{Observed}_{\text{yes}} = y$ and $\text{Observed}_{\text{no}} = n-y$ are the observed counts, and $\text{Fitted}_{\text{yes}} = n\pi_0$ and $\text{Fitted}_{\text{no}} = n(1-\pi_0)$ are the fitted counts under H_0.

Example (U.S. in Another World War, Cont'd)

In the survey, 460 answered "yes", 368 answered "no," so

$$\text{Observed}_{\text{yes}} = 460, \qquad \text{Observed}_{\text{no}} = 368.$$

Under H_0: $\pi = 0.5$, we expected half of the 828 subjects, to answer "yes" and half to answer "no,"

$$\text{Fitted}_{\text{yes}} = 828 \times 0.5 = 414, \qquad \text{Fitted}_{\text{no}} = 828 - 414 = 414.$$

Thus the LRT statistic is

$$2\left[460\log\left(\frac{460}{414}\right) + 368\log\left(\frac{368}{414}\right)\right] \approx 10.24$$

which exceeds the critical value $\chi^2_{1,0.05}$ at level 0.05

```
> qchisq(0.05, df=1, lower.tail=F)
[1] 3.841459
```

so H_0 is rejected.

The P-value is $\mathrm{P}(\chi^2_1 > 10.24)$, which is

Duality of Confidence Intervals and Significance Tests

For a 2-sided significance test of θ, the dual $100(1-\alpha)\%$ confidence interval for the parameter θ consisted 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 two-sided Wald test of H_0: $\pi = \pi_0$ having P-value > 10%
- the dual 95% score CI for π is the collection of all π_0 such that a two-sided score test of H_0: $\pi = \pi_0$ having 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
- but 0.2 is 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 confidence interval 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	90%	95%	99%
$z_{\alpha/2}$	1.645	1.96	2.58

- learned in Stat220 and Stat234

Drawbacks:

- Wald CI for π collapses if $\hat{\pi} = 0$ 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 $0 < \pi^* < \frac{1.96^2}{n+1.96^2}$. Thus the 95% score CI for π when $\hat{\pi} = 0$ is

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

which is NOT collapsing!

Score CI (Cont'd)

In Problem 1.18 in the textbook, the end points of the score CI are shown to be

$$\frac{(n\hat{\pi} + z^2/2) \pm z_{\alpha/2}\sqrt{n\hat{\pi}(1-\hat{\pi}) + z^2/4}}{n + z^2}$$

where $z = z_{\alpha/2}$.

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

Agresti-Coull Confidence Intervals

Recall the midpoint for a 95% score CI is

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

This inspires Agresti-Coull correction to the Wald CI that we **add 2 successes and 2 failures** before computing $\hat{\pi}$ and then compute the Wald CI:

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

- simpler formula than score CIs
- also perform reasonably well

Likelihood Ratio Confidence Intervals (LR CIs)

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

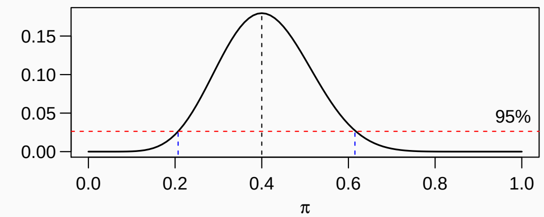
$$-2\log(\ell_0/\ell_1) = -2\log(\ell(\pi^*|y)/\ell(\hat{\pi}|y)) < \chi^2_{1,\alpha}.$$

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 that is at least $e^{-\chi^2_{1,0.05}/2} = e^{-3.84/2} \approx 0.0147$ multiple of the max. likelihood.

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



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

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 $\pi = 4/400 \approx 0.01$
- Wald CI: $0.01 \pm 1.96\sqrt{\frac{0.01\times(1-0.01)}{400}} \approx (0.00025, 0.01975)$.
- Agresti-Coull CI: estimate of π is $(4+2)/(400+4) \approx 0.0149$

$$0.0149 \pm 1.96\sqrt{\frac{0.0149\times(1-0.0149)}{404}} \approx (0.00306, 0.02665).$$

- 95% Score CI contains those π^* satisfying

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

which is the interval (0.00390.0254).

R Functions for Tests and CIs for Binomial Proportions

`prop.test()` performs the **score test** and computes the **score CI**.

- Default test is for H_0: $\pi = 0.5$ vs H_a: $\pi \neq 0.5$
- 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 = 382.2, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.003208199 0.027187351
sample estimates:
   p
0.01
```

If want a score test of H_0: $\pi = 0.02$ vs H_a: $\pi \neq 0.02$ **without** 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 = 2.0408, df = 1, p-value = 0.1531
alternative hypothesis: true p is not equal to 0.02
95 percent confidence interval:
 0.003895484 0.025426565
sample estimates:
   p
0.01
```

The 95% CI is the same as the score CI we computed before.

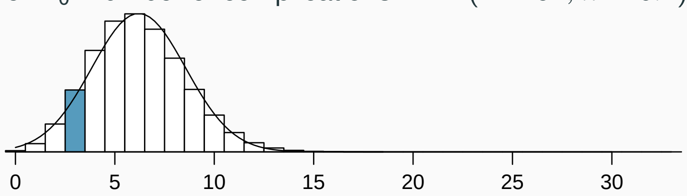
Small Sample Binomial Inference

Example: Medical Consultants for Organ Donors

- People providing an organ for donation sometimes seek the help of a special "medical consultant" These consultants assist the patient in all aspects of the surgery, with the goal of reducing the possibility of complications during the medical procedure and recovery.
- One consultant tried to attract patients by noting the average complication rate for liver donor surgeries in the US is about 10%, but her clients have only had 3 complications in the 62 liver donor surgeries she has facilitated.
- Is this strong evidence that her work meaningfully contributes to reducing complications (and therefore she should be hired!)?

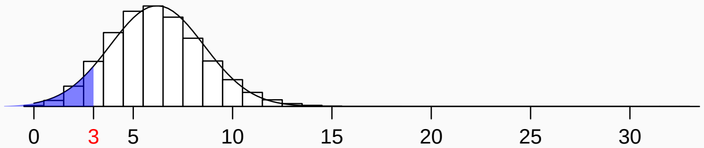
Example: Medical Consultants for Organ Donors (Cont'd)

- H_0: $\pi = 0.1$ vs. H_a: $\pi < 0.1$
- estimate of π is $3/62 \approx 0.048$
- Wald, score, likelihood ratio tests are based on *large samples*: only appropriate when *numbers of successes and failures are both at least 10* (or 15), but there were only 3 successes (having complications) in this example
- For small sample, one can use the exact distribution of the data — binomial, instead of its normal approximation.
- Under H_0: number of complications $\sim Bin(n = 62, \pi = 0.1)$

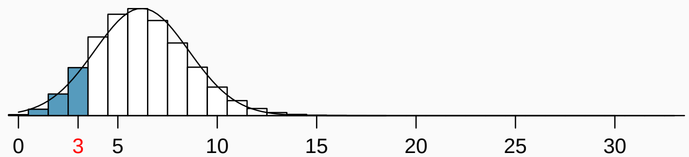


Exact Binomial Tests

For conventional large sample tests based on normal approximation, the lower one sided P-value is the area under the normal curve below 3



For exact binomial tests, the lower one-sided P-value is the area under the probability histogram below 3.



Exact Binomial Tests

Let X = number of complications among 62 liver donars $\sim Bin(n = 62, \pi = 0.1)$ under H_0.

$$P(X = k) = \binom{62}{k}(0.1)^k(0.9)^{62-k}$$

The lower one-sided P-value for the exact binomial test for the consultant's claim is

$$\begin{aligned} P\text{-value} = P(X \le 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{62}{0}(0.1)^0(0.9)^{62} + \binom{62}{1}(0.1)^1(0.9)^{61} \\ &\quad + \binom{62}{2}(0.1)^2(0.9)^{60} + \binom{62}{3}(0.1)^3(0.9)^{59} \\ &= 0.0015 + 0.0100 + 0.0340 + 0.0755 \\ &= 0.1210 \end{aligned}$$

The evidence is not strong enough to support the consultant's claim.

Exact Binomial Tests in R

The R function to do exact binomial test is `binom.test()`.

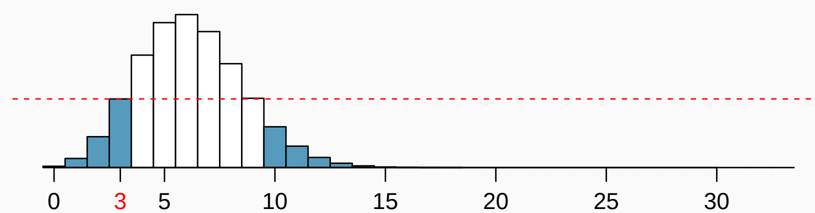
```
> binom.test(3, 62, p=0.1, alternative="less")

Exact binomial test

data:  3 and 62
number of successes = 3, number of trials = 62, p-value = 0.121
alternative hypothesis: true probability of success is less than 0.1
95 percent confidence interval:
 0.0000000 0.1203362
sample estimates:
probability of success
             0.0483871
```

The p-value given by R is 0.121, which agrees with our calculation.

Two-Sided Exact Binomial Tests

For testing H_0: $\pi = \pi_0$, suppose the observed binomial count is k_{obs}.

- P-value $= P(X \le k_{obs}) = \sum_{k \le k_{obs}} \binom{n}{k}\pi_0^k(1-\pi_0)^{n-k}$ for a lower one-sided alternative H_a: $\pi < \pi_0$
- P-value $= P(X \ge k_{obs}) = \sum_{k \ge k_{obs}} \binom{n}{k}\pi_0^k(1-\pi_0)^{n-k}$ for a upper one-sided alternative H_a: $\pi > \pi_0$
- If the alternative is two-sided H_a: $\pi \ne \pi_0$, the P-value is the sum of all the $P(X = k)$ such that $P(X = k) \le P(X = k_{obs})$

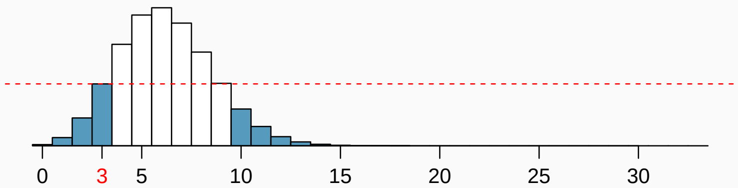


Example: Medical Consultants for Organ Donors (Cont'd)

In this example, the observed count k_{obs} is 3.

As $P(X = 9) > P(X = 3)$ and $P(X = k) < P(X = 3)$ for all $k \geq 10$, the two-sided P-value is

$$P(X \leq 3) + P(X \geq 10) \approx 0.0872 + 0.1210 = 0.2082$$



Note that the two-sided P-value for an exact binomial test may not be twice of the one-sided P-value since a binomial distribution may not be symmetric

Two-Sided Exact Binomial Tests in R

```
> binom.test(3, 62, p=0.1, alternative="two.sided")

Exact binomial test

data:  3 and 62
number of successes = 3, number of trials = 62, p-value = 0.2081
alternative hypothesis: true probability of success is not equal to 0.1
95 percent confidence interval:
 0.01009195 0.13496195
sample estimates:
probability of success
             0.0483871
```

The p-value given by R 0.2081 agrees with our calculation.

Exact Binomial Confidence Intervals

- Just like Wald, score, or LRT confidence intervals, one can invert the two-sided exact binomial test to construct confidence intervals for π.
- The $100(1-\alpha)\%$ exact binomial confidence interval for π is the collection of those π_0 such that the two-sided P-value for testing H_0: $\pi = \pi_0$ using the exact binomial test is at least α.
- The computation of the exact binomial confidence interval is tedious to do by hand, but easy for a computer.
- For the medical consultant example, the 95% exact confidence interval for π is (0.0101, 0.1350) from the R output in the previous slide.