Categorical Data Analysis
Lecture 2 – Review

Graduate School of Advanced Science and Engineering
Rei Monden

2024.12.5

Contents of this lecture

Today's contents

- Types of variables
- Binomial and multinomial distributions (review)
- Statistical inference for a proportion
- Significance tests and confidence intervals about a binomial parameter (Wald, score, likelihood ratio)

Types of variables

Variables

Type of variable	**Scale**	**Features**
Qualitative (categorical)	Nominal	Represents only categorical differences (e.g., gender, occupation).
	Ordinal	Nominal variables whose categories can be ordered (e.g., disagree $\rightarrow$ neutral $\rightarrow$ agree), but there is no ‘distance’ between categories.
Quantitative (continuous)	Interval	The order and distance between values is meaningful, but there is no origin (e.g., temperature).
	Ratio	There is a meaningful origin (i.e., value 0). E.g., length or weight.

In this course, among the *qualitative* variables, we will consider separately the case of binary variables (with two categories) and the remaining (with ≥ 3 categories).

Response and explanatory variables

- Response variable = Dependent variable.
  Usually written as $\mathbf{Y}$ or $\mathbf{y}$.

- Explanatory variable = Independent variable.
  Usually written as $\mathbf{X}$ or $\mathbf{x}$.

Example:
Is happiness related to annual income and savings?

- Response variable: Happiness.
- Explanatory variables: Annual income, savings.

In this lecture, we will focus on studying models for qualitative response variables, using qualitative and/or quantitative explanatory variables.

Variables

Classical regression analysis is a statistical framework that allows analyzing quantitative response variables.

- E.g., temperature, blood pressure, height, speed, income, etc.

Categorical data analysis encompasses a family of statistical methods whose response variable is qualitative (categorical).

- E.g., gender (male/female), prefecture, disease stage (cancer stage), etc.

And, as mentioned before, explanatory variables can be either numerical or categorical in both cases.

Exercise 2-1

Answer to the following exercises in Rmarkdown and generate a PDF file from it.
Report all the exercises from this lecture in one file and submit it via Teams("第２回課題").

1. In the following examples, identify the natural response variable and explanatory variables.
   a. Attitude toward gun control (favor, oppose), gender (female, male), mother's education (high school, college).
   b. Heart disease (yes, no), blood pressure, cholesterol level.
   c. Race (white, nonwhite), religion (Catholic, Jewish,Muslim, Protestant, none), vote for president (Democrat, Republican, Green), annual income.

–> Continue to the next slide...

Exercise 2-1

2. Which scale of measurement is most appropriate for the following variables-nominal or ordinal?
    a. UK political party preference (Labour, Liberal Democrat, Conservative, other).
    b. Highest education degree obtained (none, high school, bachelor's, master's, doctorate).
    c. Patient condition (good, fair, serious, critical).
    d. Hospital location (London, Boston, Madison, Rochester, Toronto).
    e. Favorite beverage (beer, juice, milk, soft drink, wine, other).
    f. Rating of a movie with 1 to 5 stars (hated it, didn't like it, liked it, really liked it, loved it).

Binomial and multinomial distributions (review)

Binomial distribution

Consider a set of n Bernoulli trials, verifying the following conditions:

- Each trial's outcome is binary (success/failure).
- The probability of success (say, π) and the probability of failure $(1 - \pi)$ are constant across trials.
- Trials are independent from each other.

Under these conditions, the total number of ''successes'' (say, Y) out of the n trials is said to follow a binomial distribution with parameters n and π:

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

Binomial distribution

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

The probability mass function of Y is

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \text{ for } y = 0, 1, \dots, n,$$

where

- $\binom{n}{y} = \frac{n!}{y!(n-y)!}$ is the binomial coefficient
- $m! = m \times (m-1) \times (m-2) \times \cdots \times 1$ is the factorial of m.

Note: $0! = 1$.

Binomial distribution – Example

The probability of winning a certain lottery is 0.4.
Three randomly selected people draw this lottery.

Denote

$$Y = \text{number of wins among the 3 people.}$$

Assuming that the assumptions of Bernoulli trials are met, then

$$Y \sim \text{binomial}(n = 3, \pi = .4).$$

The probability mass function is

$$\begin{aligned} P(Y = y) &= \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y} \\ &= \frac{3!}{y!(3-y)!}(0.4)^y(0.6)^{3-y}, \text{ for } y = 0, 1, 2, 3. \end{aligned}$$

Binomial distribution – Example

$$\boxed{P(Y=y)=\frac{3!}{y!(3-y)!}(0.4)^y(0.6)^{3-y}}$$

So, for instance,

- $P(0)=\frac{3!}{0!3!}(0.4)^0(0.6)^3=(0.6)^3=0.216.$
- $P(1)=\frac{3!}{1!2!}(0.4)^1(0.6)^2=3(0.4)(0.6)^2=0.432.$

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

R Code

```r
> dbinom(x    = 0,    # number of successes
+        size = 3,    # number of trials
+        p    = 0.4)  # prob. success of each trial
```

```
[1] 0.216
```

```r
> dbinom(0, 3, 0.4)   # unnamed, but *ordered*, arguments
```

```
[1] 0.216
```

```r
> dbinom(1, 3, 0.4)
```

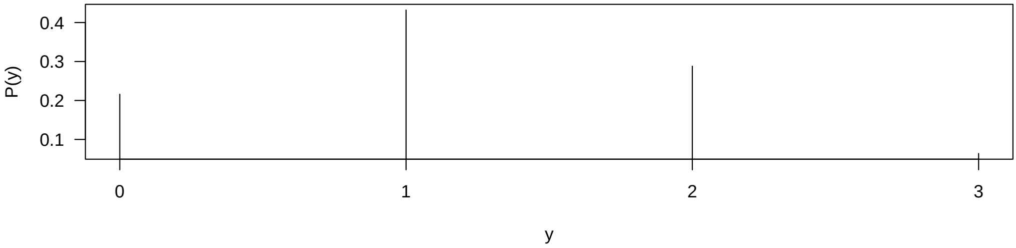
```
[1] 0.432
```

```r
> dbinom(0:3, 3, 0.4) # vectorized computation
```

```
[1] 0.216 0.432 0.288 0.064
```

R Code

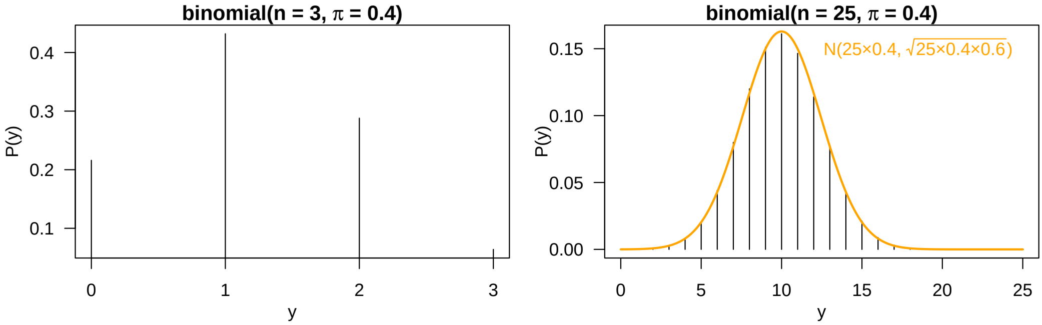
```
> plot(x    = 0:3,
+      y    = dbinom(0:3, 3, .4),
+      type = "h",       # vertical lines
+      xlab = "y",
+      ylab = "P(y)",
+      las  = 1,         # y-axis labels position
+      xaxt = "n")       # remove default x-axis tick marks
> axis(side = 1,         # bottom side, i.e., x-axis
+      at   = 0:3)       # position of tick marks
```



Binomial distribution – Properties

If $Y \sim \text{binomial}(n, \pi)$ then:

- $E(Y) = n\pi$
- $\sigma(Y) = \sqrt{\text{Var}(Y)} = \sqrt{n\pi(1-\pi)}$
- When n is large enough (say, $n\pi > 5$ and $n(1-\pi) > 5$), Binomial(n, π) is approximately Normal$(\mu = n\pi, \sigma = \sqrt{n\pi(1-\pi)})$.



Exercise 2-2

Each of 100 multiple-choice questions on an exam has four possible answers, but only one correct response. For each question, a student randomly selects one response as the answer.

a. Specify the probability distribution of the student's number of correct answers on the exam.
b. Based on the mean and standard deviation of that distribution, would it be surprising if the student made at least 50 correct responses? Explain your reasoning.

Multinomial distribution: General form of the binomial distribution

Consider a set of n trials, verifying the following conditions:

- There are $c > 2$ possible outcome categories in each trial.
- $\pi_i = \text{P(category i)}$ for each trial, and $\sum_{i=1}^{c} \pi_i = 1$.
- Trials are independent from each other.

Under these conditions, and letting Y_i denote the total number of trials for which category i was observed, then we say that the joint distribution of $(Y_1, Y_2, \dots, Y_c)$ follows a multinomial distribution:

$$(Y_1, Y_2, \dots, Y_c) \sim \text{multinomial}(n; \pi_1, \dots, \pi_c).$$

Multinomial distribution: General form of the binomial distribution

$$\boxed{(Y_1, Y_2, \ldots, Y_c) \sim \text{multinomial}(n; \pi_1, \ldots, \pi_c)}$$

The probability mass function is

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

where $y_i = 0, 1, \ldots, n$ for all $i = 1, \ldots, c$ and $\sum_i y_i = n$.

Note: The binomial distribution is the special case when $c = 2$.

Multinomial distribution – Example

Let's assume that Japanese blood types are distributed as follows:
Type A: 40%; Type B: 20%; Type O: 30%; Type AB: 10%.
Therefore, $(\pi_A, \pi_B, \pi_O, \pi_{AB}) = (0.4, 0.2, 0.3, 0.1)$.
Furthermore, five Japanese people are randomly selected.

Denote

$$Y_t = \text{ number of Japanese with blood type } t,$$

where $t =$ A, B, O, or AB.

Assuming that the trials conditions are met, then

$$(Y_{\mathsf{A}}, Y_{\mathsf{B}}, Y_{\mathsf{O}}, Y_{\mathsf{AB}}) \sim \text{multinomial}(5; 0.4, 0.2, 0.3, 0.1).$$

Multinomial distribution – Example

The probability mass function is

$$P(Y_{\mathsf{A}} = y_{\mathsf{A}}, Y_{\mathsf{B}} = y_{\mathsf{B}}, Y_{\mathsf{O}} = y_{\mathsf{O}}, Y_{\mathsf{AB}} =_{\mathsf{AB}}) = \frac{5!}{y_{\mathsf{A}}!y_{\mathsf{B}}!y_{\mathsf{O}}!y_{\mathsf{AB}}!}(0.4)^{y_{\mathsf{A}}}(0.2)^{y_{\mathsf{B}}}(0.3)^{y_{\mathsf{O}}}(0.1)^{y_{\mathsf{AB}}},$$

such that $y_{\mathsf{A}} + y_{\mathsf{B}} + y_{\mathsf{O}} + y_{\mathsf{AB}} = 5$.

Example:
What is the probability of observing 2 type A, 2 type B, 1 type O, and 0 type AB?

$$\begin{aligned} P(Y_A = 2, Y_B = 2, Y_O = 1, Y_A B = 0) &= \frac{5!}{2!2!1!0!}(0.4)^2(0.2)^2(0.3)^1(0.1)^0 \\ &= 30(0.4)^2(0.2)^2(0.3)^1(0.1)^0 \\ &= 0.058 \end{aligned}$$

Multinomial distribution – Properties

If $(Y_1, Y_2, \ldots, Y_c) \sim \text{multinomial}(n; \pi_1, \ldots, \pi_c)$ then, for $i = 1, \ldots, c$:

- $E(Y_i) = n\pi_i$
- $\sigma(Y_i) = \sqrt{\text{Var}(Y_i)} = \sqrt{n\pi_i(1 - \pi_i)}$
- $Cov(Y_i, Y_j) = -n\pi_i\pi_j$, for $i \neq j$.

Statistical inference for a proportion

Inference for proportions

The parameters of a binomial or multinomial distribution (i.e., proportions) are typically unknown.
Sample data is needed to estimate these parameters.

Commonly, one uses the maximum likelihood estimation (MLE) method to perform the estimation.

Likelihood function and maximum likelihood estimation

Statistical modeling first requires that we choose a probability distribution for the response variable.
(E.g., the binomial.)

Such a distribution is meant to capture (to *model*) reality as closely as possible.
(E.g., are the assumptions from Bernoulli trials minimally met?)

We then look for the unknown parameters of this probability distribution.
The main idea is that the distribution depends on the unknown parameters.
We need to find the best values!

Our job is to find the parameter values that make our observed data look as likely as possible.

Likelihood

Example:
Suppose the number successes from $n = 10$ Bernoulli trials is $y = 0$.

Under the binomial model we have that $Y \sim$ binomial$(10, \pi)$.
Note how the data are fixed and known (i.e., 0 successes in 10 trials).
Parameter π, although also considered to be fixed in the population, is however unknown.
We would like to find its *fixed* value in the population!

The probability of observing 0 successes in 10 trials is

$$P(0) = \frac{10!}{0!10!}\pi^0(1-\pi)^{10} = (1-\pi)^{10}.$$

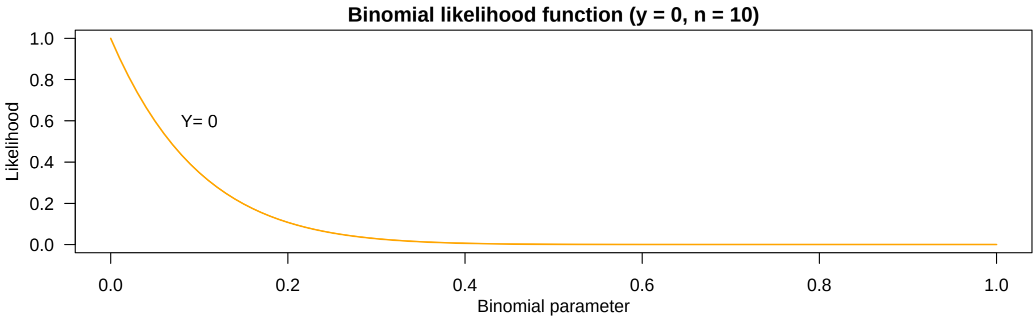
In this example, the likelihood function is given by

$$l(\pi) = (1-\pi)^{10}, \text{ for } 0 \leq \pi \leq 1.$$

Likelihood

$$\boxed{l(\pi) = (1-\pi)^{10}, 0 \leq \pi \leq 1}$$

We can evaluate the likelihood function at any possible value for π in $[0, 1]$.
The plot below shows the likelihood function evaluated in the $[0, 1]$ interval.



Likelihood

In the general situation, when the observed data $\{Y_1, Y_2, \dots, Y_n\}$ follow a joint probability distribution with parameter θ

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

then the likelihood function is

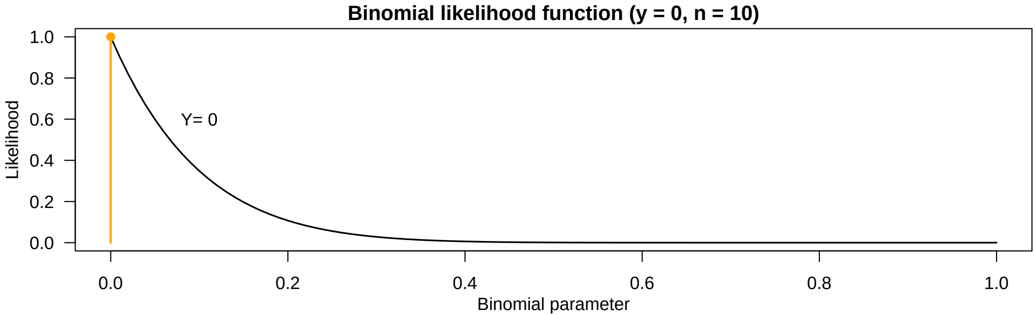
$$l(\theta) = l(\theta | y_1, y_2, \dots, y_n) = f(y_1, y_2, \dots, y_n | \theta).$$

Note. Parameter θ can be a *vector* of parameters
(e.g., $\theta = (\pi_1, \pi_2, \dots, \pi_c)$ for the multinomial distribution).

In the likelihood function, the probability distribution is regarded as a function of θ, for fixed and known observed data $y_1, y_2, \dots, y_n$.

Maximum likelihood estimation

Maximum likelihood estimation consists of finding the parameter value θ that maximizes the likelihood function.
That value, say $\hat{\theta}$, is called the *maximum likelihood estimate* of parameter θ.

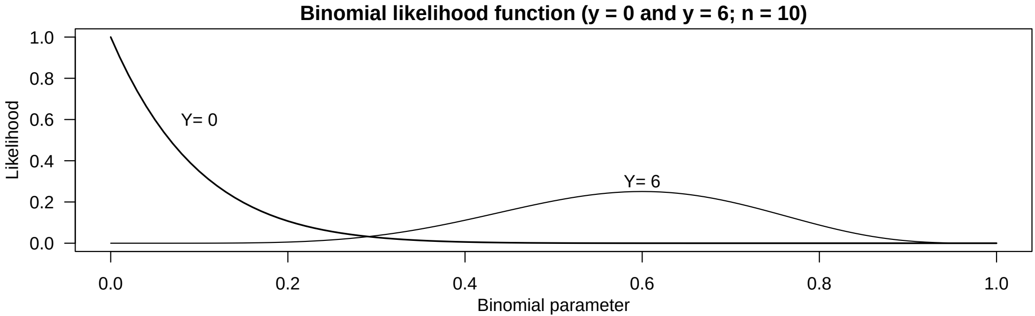


Since the likelihood function is maximum at $\pi = 0$, the maximum likelihood estimate of π is $\hat{\pi} = 0$, when there are $y = 0$ successes in a total of $n = 10$ trials.

Maximum likelihood estimate for binomial distribution

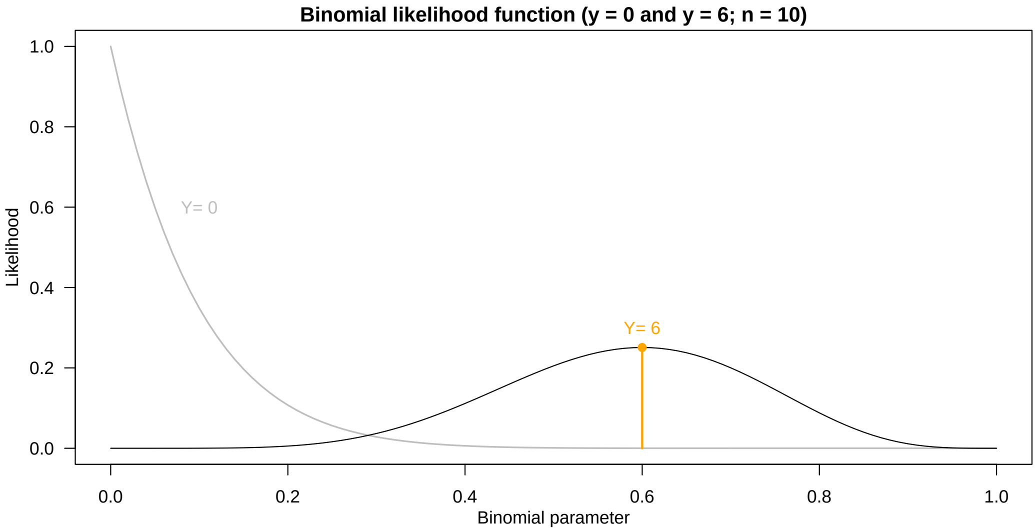
In general, the maximum likelihood estimate of π for y successes in n trials is

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



Example:
What is the maximum likelihood estimate $\hat{\pi}$ when $n = 10$ and $y = 6$?

Maximum likelihood estimate for binomial distribution



Maximizing the *log*-likelihood

Instead of maximizing the *likelihood* function, it is simpler (in terms of the math and the algorithms involved) to maximize the log-likelihood function instead:

$$\log l(\theta) = \log l(\theta | y_1, y_2, \ldots, y_n),$$

where log() is the *natural* logarithm.

In the notation of the binomial distribution example, we maximize

$$\log l(\pi | y).$$

This works because the log() function is a *monotonic increasing* function:

$$x_1 > x_2 \Longleftrightarrow \log(x_1) > \log(x_2),$$

and therefore

$$l(\pi_1 | y) > l(\pi_2 | y) \Longleftrightarrow \log l(\pi_1 | y) > \log l(\pi_2 | y).$$

Maximum likelihood estimate for binomial distribution

Given $Y \sim \text{binomial}(n, \pi)$ where n is known and π is unknown, the likelihood function is

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

and the corresponding log-likelihood function is

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

From calculus, we know that x_0 is a maximum of (at least twice differentiable function) $f(x)$ if

- $\frac{d}{dx} f(x) = 0$

and

- $\frac{d^2}{dx^2} f(x) < 0.$

Maximum likelihood estimate for binomial distribution

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

First derivative of $\log l(\pi|y)$*:*

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

Solving $\frac{d}{d\pi} \log l(\pi|y) = 0$ implies that $\pi = \frac{y}{n}$.

This is our current (and only) candidate for maximum.

Maximum likelihood estimate for binomial distribution

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

Second derivative of $\log l(\pi|y)$*:*

$$\frac{d^2}{d\pi^2} \log l(\pi|y) = -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2}.$$

Observe that $\frac{d^2}{d\pi^2} \log l(\pi|y) < 0$ for any value π in $[0, 1]$.

Conclusion:
The maximum likelihood estimate of π is $\hat{\pi} = \frac{y}{n}$.

Maximum likelihood estimate – Some notes

Although the maximum likelihood estimator is not always the best, it does have some desirable asymptotic properties:

- Asymptotically unbiased:
  As the sample size n increases, the bias of the maximum likelihood estimate approaches 0.
- Asymptotically efficient:
  There is no other estimator whose variance is smaller than that of the maximum likelihood estimator as the sample size n increases.
- Asymptotically normal:
  The sampling distribution of a maximum likelihood estimator approaches normality as the sample size n increases.

In some, but not all, cases, the sample size required to satisfy the above properties may be quite large.

Exercise 2-3

In a particular city, a proportion π of the population supports an increase in the minimum wage. For a random sample of size 2, let Y equal the number of persons who support an increase.

a. Assuming $\pi = 0.50$, specify the probabilities for the possible values y of Y, and find the distribution's mean and standard deviation.
b. Suppose you observe $y = 1$ and do not know the value of π. Find and sketch the likelihood function, and explain why the ML estimate is $\hat{\pi} = 0.50$.

Significance tests and confidence intervals about a binomial parameter

Significance tests and CIs about a binomial parameter

We learned today that, for $Y \sim \text{binomial}(n, \pi)$, we have that

- The MLE is $\hat{\pi} = \frac{Y}{n}$.
- $E(Y) = n\pi$.
- $\text{var}(Y) = n\pi(1-\pi)$.

As a consequence, we also have that

- $E(\hat{\pi}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E\left(Y\right) = \pi$.
- $\text{var}(\hat{\pi}) = \text{var}\left(\frac{Y}{n}\right) = \frac{1}{n^2}\text{var}(Y) = \frac{\pi(1-\pi)}{n}$.

Finally, from the Central Limit Theorem, as n increases we have that

$$\frac{\hat{\pi} - E(\hat{\pi})}{\sqrt{\text{var}(\hat{\pi})}} = \frac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}} \sim \mathcal{N}(0, 1).$$

This formula is the base for several significance tests about π.

Significance tests about a binomial parameter

There are three types of test statistics to perform the following test:

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

- Score test:

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

- Wald test:

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

- Likelihood ratio test: We will see it in a moment.

Significance tests about a binomial parameter

As the sample size n increases,

- The sampling distributions of both z_S and z_W approximate $\mathcal{N}(0, 1)$.
- The sampling distributions of both z_S^2 and z_W^2 approximate χ_1^2.

We use these approximations to the sampling distributions to perform inference (i.e., compute p-values and confidence intervals).

Confidence intervals about a binomial parameter

The confidence intervals are derived from the sampling distributions in the customary fashion:

- $100(1-\alpha)\%$ score confidence interval:

$$\hat{\pi} \pm z_{\alpha/2}\sqrt{\pi_0(1-\pi_0)/n}$$

- $100(1-\alpha)\%$ Wald confidence interval:

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

- Likelihood ratio test: (We will not see this today.)

Significance tests about a binomial parameter – Example

In 2022, we asked the question:

Do you think Japan's employment rate will decline in the next 10 years?

460 out of 828 respondents answered 'yes'.

Let π denote the proportion of people in the population who would answer 'yes'.
Hypotheses to test:

$$\mathsf{H}_0 : \pi = 0.5$$
$$\mathsf{H}_a : \pi \neq 0.5$$

- MLE of π: $\hat{\pi} = 460/828 = 0.556$.
- Score statistic: $z_S = \frac{0.556-0.5}{\sqrt{0.5(1-0.5)/828}} = 3.20$.
- Wald statistic: $z_W = \frac{0.556-0.5}{\sqrt{0.556(1-0.556)/828}} = 3.22$.

Significance tests about a binomial parameter – Example

Note that p-values using $\mathcal{N}(0,1)$ or χ^2_1 are the same.

```
> 2 * pnorm(3.20, lower.tail = FALSE)          # score test's p-value
```

```
[1] 0.001374276
```

```
> pchisq(3.20^2, df = 1, lower.tail = FALSE) # score test's p-value
```

```
[1] 0.001374276
```

```
> 2 * pnorm(3.22, lower.tail = FALSE)          # Wald test's p-value
```

```
[1] 0.001281906
```

```
> pchisq(3.22^2, df = 1, lower.tail = FALSE) # Wald test's p-value
```

```
[1] 0.001281906
```

Conclusion:
Using either the score or Wald's test, we reject H_0 (at, say, $\alpha = 5\%$ significance level).

Confidence intervals about a binomial parameter – Example

95% score confidence interval:

```
> n        <- 828                             # sample size
> pi.hat   <- 460 / 828                       # MLE
> pi.0     <- .5                              # pi under H0
> z.crit   <- qnorm(.025, lower.tail = FALSE) # z_(alpha/2)
> SE.score <- sqrt(pi.0 * (1 - pi.0) / n)     # score's SE (i.e., denominator of test statistic)
> c(pi.hat - z.crit * SE.score, pi.hat + z.crit * SE.score)
```

```
[1] 0.5214988 0.5896123
```

95% Wald confidence interval:

```
> n        <- 828                             # sample size
> pi.hat   <- 460 / 828                       # MLE
> z.crit   <- qnorm(.025, lower.tail = FALSE) # z_(alpha/2)
> SE.wald  <- sqrt(pi.hat * (1 - pi.hat) / n) # Wald's SE (i.e., denominator of test statistic)
> c(pi.hat - z.crit * SE.wald, pi.hat + z.crit * SE.wald)
```

```
[1] 0.5217097 0.5894014
```

Likelihood ratio test (LRT)

Recall the likelihood function for the binomial proportion π:

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

In order to test $\mathrm{H}_0 : \pi = \pi_0$ *versus* $\mathrm{H}_a : \pi \neq \pi_0$, define:

- $\ell_0 = \ell(\pi_0|y)$:
  The maximum value of the likelihood function under H_0.
- $\ell_1 = \ell(\hat{\pi}|y)$:
  The maximum value of the likelihood function for all possible π values.
  By definition, this is the likelihood function evaluated at the MLE $\hat{\pi}$.

Likelihood ratio test (LRT)

Several properties:

- $\ell_0 \leq \ell_1$, by definition.
- We will reject H_0 if $\ell_0 \ll \ell_1$.

LRT statistic

The likelihood ratio test statistic for testing

$$\begin{aligned} \mathsf{H}_0 &: \pi = \pi_0 \\ \mathsf{H}_a &: \pi \neq \pi_0 \end{aligned}$$

is

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

- Since $\ell_0 \leq \ell_1$, the LRT statistic is always nonnegative.
- For large n, $2\log(\ell_1/\ell_0) \sim \chi^2_1$ if H_0 is true.
  - p-value $= \mathsf{P}(\chi^2_1 > 2\log(\ell_1/\ell_0))$.
  - We reject H_0 if $p < \alpha$, at significance level α.

Significance test for binomial proportions – LRT

Recall the likelihood function for the binomial proportion π:

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

We have that

$$\begin{aligned}\frac{\ell_1}{\ell_0} &= \frac{\binom{n}{y}\left(\frac{y}{n}\right)^y\left(1-\frac{y}{n}\right)^{n-y}}{\binom{n}{y}\pi_0^y(1-\pi_0)^{n-y}} \\ &= \left(\frac{y}{n\pi_0}\right)^y\left(\frac{n-y}{n(1-\pi_0)}\right)^{n-y}.\end{aligned}$$

Significance test for binomial proportions – LRT

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

From this, the LRT statistic is

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

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

Significance test for binomial proportions – LRT example

Let's recover the running example where we had 460 people answering 'yes' and $828 - 460 = 368$ answering 'no' to a survey on Japan's employment rate:

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

Under $H_0 : \pi = 0.5$, half of the 828 people are expected to answer 'yes' and the other half are expected to answer 'no':

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

Therefore, the LRT statistic is

$$2\log(\ell_1/\ell_0) = 2\left[460 \times \log\left(\frac{460}{414}\right) + 368 \times \log\left(\frac{368}{414}\right)\right] = 10.24.$$

Significance test for binomial proportions – LRT example

Critical value $\chi^2_{1,0.05}$ at significance level $\alpha = 0.05$:

```
> qchisq(0.05, df = 1, lower.tail = FALSE)
```

```
[1] 3.841459
```

Since the test statistic (10.24) is larger than the critical value (3.84), we decide to reject H_0.

The p-value is given by $P(\chi^2_1 > 10.24)$ and is equal to:

```
pchisq(10.24, df = 1, lower.tail = FALSE)
```

```
## [1] 0.001374276
```

Score vs Wald vs likelihood ratio test

- For the usual regression model with normal residuals, all tests coincide.
- For other models like the ones we will learn in this course, they get close to each other as the sample size n increases.
- For small to moderate sample sizes:
  Avoid Wald's test (use the score or LR test instead).

Exercise 2-4

When the 2010 General Social Survey asked subjects in the US whether they would be willing to accept cuts in their standard of living to protect the environment, 486 of 1374 subjects said *yes*.

a. Estimate the population proportion who would say *yes*. Construct and interpret a 99% confidence interval for this proportion.
b. Conduct a significance test to determine whether a majority or minority of the population would say *yes*. Use significance level $\alpha = 5\%$. Report and interpret the P-value.

Exercise 2-5

If Y is a random variable and c is a positive constant, then the standard deviation of the probability distribution of cY equals $c\sigma(Y)$. Suppose Y is a binomial variate and let $\pi = Y/n$.

a. Based on the binomial standard deviation for Y, show that $\sigma(\pi) = \sqrt{\pi(1-\pi)/n}$.
b. Explain why it is easier to estimate π precisely when it is near 0 or 1 than when it is near 0.50.

In the next lecture

We are going to skip 2.1.2, 2.1.4, 2.3.5, 2.3.6, 2.4.6, 2.5-2.7.