Inference on Proportion

Demetris Athienitis

Framework

Parameters are often estimated using *maximum likelihood* (ML) That is, finding value of the parameters (of interest) that maximize the *likelihood function* or equivalently the log of the likelihood function.

Definition (Likelihood function)

The probability of the observed data, expressed as a function of the parameter is called a likelihood function.

Definition (MLE)

The maximum likelihood estimator (MLE) is defined to be the parameter value for which the likelihood function is maximized.

Example (for illustration)

Consider a widget that either works (success) or does not work (failure). Hence, if each attempt with the widget is identical and independent, the number of successes follows a $\text{Bin}(n, \pi)$.

Out of 10 attempts, 7 yielded a success. Which π value is most likely to yield this outcome?

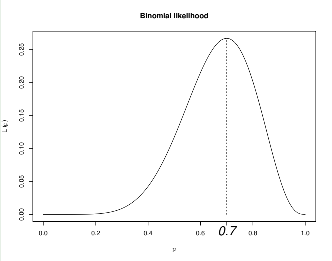
$\text{Bin}(10, ?)$	$P(Y = 7)$
$\pi = 0.5$	0.1172
$\pi = 0.6$	0.2150
$\pi = 0.7$	0.2668
$\pi = 0.8$	0.2013

Maximized at $\pi = 0.7$. Makes sense!

Example (continued)

Let's look slightly more rigorously. The likelihood function is

$$L(\pi|y,n) := \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}, \qquad y=7,\ n=10,\ \pi\in[0,1]$$



Maximized at $\hat{\pi} = y/n = 7/10$.

Finding the MLE

1. Write down the likelihood as a function of π
2. Take the log of the likelihood function (This isn't necessary but frequently makes the calculus easier)
3. Take the derivative with respect to π
4. Set the derivative equal to zero and solve for π. This is the MLE!
5. Check that the second derivative is negative to ensure we found a maximum

MLE Key facts

- If $y_1, y_2, \ldots, y_n$ are i.i.d. from a normal distribution, then

$$L(\mu, \sigma^2 | \mathbf{y}) = \prod_{i=1}^{n} f(\mu, \sigma^2 | y_i)$$

  where $f(\cdot)$ is the p.d.f. MLEs are $\hat{\mu} = \bar{y}$ and $\hat{\sigma}^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$
- In ordinary linear regression, with Y being normal, the least squares estimators of the regression coefficients are also the MLEs
- For large sample size n, MLEs are optimal (no other estimator has smaller mean squared error: variance plus squared bias). This is true in fairly broad generality
- For large n, the sampling distribution of the MLE is approximately normal. Again, this is true in fairly broad generality

MLE Key facts

- Recall that $\hat{\pi}$ is *unbiased* with $E(\hat{\pi}) = \pi$ and *consistent* with $V(\hat{\pi}) \underset{n\to\infty}{\longrightarrow} 0$. MLEs are generally consistent
- $\hat{\pi}$ is a sample mean for 0-1 data, so by the Central Limit Theorem, the sampling distribution is approximately normal for large n. Again, this is generally true for MLEs

Framework

First we introduce two “classical” methods for introduction but then will refer to newer methods that are now considered “standard”.

Start with the usual hypothesis test

$$\mathrm{H}_0 : \pi = \pi_0 \qquad \text{vs} \qquad \mathrm{H}_a : \pi \neq \pi_0$$

and $p = \hat{\pi}$.

Wald

Under the null,

$$TS = \frac{p - \pi_0}{\sqrt{p(1-p)/n}} \overset{\text{approx.}}{\sim} N(0,1)$$

We fail to reject the null when

$$\left| \frac{p - \pi_0}{\sqrt{p(1-p)/n}} \right| < z_{1-\alpha/2}$$

Solving for π_0 we obtain the $100(1-\alpha)\%$ CI

$$p \mp z_{1-\alpha/2}\sqrt{p(1-p)/n}$$

Remark

When $p = 0$ or 1, the CI collapses to (0,0) or (1,1).

Score/Wilson

Fully adopting the null hypothesis, π_0 is used in the standard error, so that under the null,

$$TS = \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \overset{\text{approx.}}{\sim} N(0,1)$$

We fail to reject the null when

$$\left| \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \right| < z_{1-\alpha/2}$$

Solving for π_0 requires the use of the quadratic formula and is a bit more complex and generally we let software solve for us.

Review and Other methods

- **Agresti-Coull**, has become the new norm and works well for small sample sizes
- Wald, Score and Agresti-Coull are approximate methods since the approximate the distribution using the normal. They are equivalent when sample size is very large
- **Clopper-Pearson** a.k.a. "exact" which is recommended when n is small seeing how it is "exact"

Software

In R

Use `binom.confint{binom}`

```
binom.confint(y, n, conf.level = 0.95, methods = ``all'')
```

with `methods =`

- `asymptotic` for Wald
- `wilson` or `prop.test` for Score/Wilson
- `ac` for Agresti Coull
- `exact` for Clopper-Pearson
- `all` for all methods (which includes one not listed here)

Example

An experiment yielded 5 successes out of 17 trials.

```
> library(binom)
> binom.confint(x=5,n=17,conf.level=0.95,method="all")
          method x  n      mean     lower     upper
1  agresti-coull 5 17 0.2941176 0.1298740 0.5342570
2     asymptotic 5 17 0.2941176 0.0775217 0.5107136
5          exact 5 17 0.2941176 0.1031355 0.5595827
10     prop.test 5 17 0.2941176 0.1137660 0.5595199
11        wilson 5 17 0.2941176 0.1327999 0.5313311
```

These 2-sided CIs are equivalent to their corresponding 2-sided hypothesis test. For example, to test $H_0 : \pi = 0.5 \quad \text{vs} \quad H_a : \pi \neq 0.5$

- $\pi = 0.5$ is a plausible value since it is in all the CIs
- Fail to reject H_0; p-value greater than 5%, since we used 95% levels

We learned

- Estimating proportion of successes via MLE
- Inference on proportion using approximate and exact methods