Lecture 15
Interval Estimation

Chao Song

College of Ecology

Lanzhou University

November 11, 2024

Interval estimation

We have used the **pivotal method** for constructing confidence intervals. However, it is sometimes impossible to find a pivotal quantity. Under these situations, we need other techniques to find confidence intervals.

- Asymptotic distribution of statistics;
- Distribution free confidence intervals;
- Resampling based confidence intervals.

Confidence interval of maximum likelihood estimate

Recall that the maximum likelihood estimate $\hat{\theta}$ for a parameter θ asymptotically has a **normal** distribution

$$\hat{\theta} \sim N(\theta, I(\theta)^{-1})$$

where $I(\theta)$ is the Fisher information defined as

$$\begin{aligned} I(\theta) &= E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right] \\ &= -E\left(\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right) = -nE\left(\frac{\partial^2 \ln f(x|\theta)}{\partial \theta^2}\right) \end{aligned}$$

Confidence interval of maximum likelihood estimate

The asymptotic properties of maximum likelihood estimates provide a generally applicable approach to deriving confidence interval.

$$P(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sqrt{I(\theta)^{-1}}} < z_{\alpha/2}) \approx 1 - \alpha$$

Thus, the $100(1 - \alpha)\%$ confidence interval for θ is

$$\hat{\theta} \pm \frac{z_{\alpha/2}}{\sqrt{I(\theta)}}$$

Because θ is unknown, $I(\theta)$ is approximated by the observed Fisher information $I(\hat{\theta})$, i.e. Fisher information evaluated at the maximum likelihood estimate $\hat{\theta}$.

Confidence interval of maximum likelihood estimate

Previously, we can analytically derive the distribution of sample mean based on a sample from a normal distribution to construct its confidence interval. Now, we consider using the maximum likelihood framework to do so.

For a sample $X_1, X_2, \ldots, X_n$, the log-likelihood is

$$\ln L(\mu, \sigma^2) = \sum_{i=1}^{n} \Big(\ln(\frac{1}{\sigma\sqrt{2\pi}}) - \frac{(X_i - \mu)^2}{2\sigma^2} \Big)$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma^2}$$

$$\frac{\partial^2 \ln L(\mu, \sigma^2)}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$I(\mu) = -E\Big(\frac{\partial^2 \ln L(\mu, \sigma^2)}{\partial \mu^2}\Big) = \frac{n}{\sigma^2}$$

Confidence interval of maximum likelihood estimate

From previous lectures, we know that the maximum likelihood estimate for μ is $\overline{X}$. The confidence interval can thus be approximated by using observed Fisher information for deriving the variance.

$$\overline{X} \pm z_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{n}}$$

where $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$ is the variance of MLE evaluated at MLE $\overline{X}$.

Confidence interval of maximum likelihood estimate

Recall the problem of calculating confidence interval for $\mu_X - \mu_Y$ when we cannot assume that $\sigma_X = \sigma_Y$. We can use the maximum likelihood approach to deriving its confidence interval. Recall that

$$(X - Y) \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Thus, using the result of confidence interval for the mean of a normal distribution, the confidence interval is

$$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{s_X^2/n + s_Y^2/m}$$

Note here that the variance of the MLE should be its variance evaluated at the MLE. Thus $s_X^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$ and $s_Y^2 = \sum_{i=1}^{m}(X_i - \overline{X})^2$

Confidence interval of maximum likelihood estimate

Comments on confidence interval of the difference of means:

- Depending on whether we can assume equal variance of X and Y, the confidence interval for $\mu_X - \mu_Y$ differs. It is thus critical to assess whether the equal variance assumption is valid or not.
- The confidence interval based on equal variance performs very poorly when variances are not actually equal and sample size of X and Y differ substantially.
- if the sample variance differs a lot between X and Y and their respective sample size are vastly different, it is safer and more robust to construct confidence interval assuming unequal variance.

Confidence interval for proportions

Let X be the number of success in n independent Bernoulli trials. How do we construct confidence interval for the success probability p?

The log-likelihood is

$$\begin{aligned}
\ln L(p) &= \ln\left(\mathbf{C}_n^k p^X (1-p)^{n-X}\right) \\
&= \ln \mathbf{C}_n^X + x \ln p + (n-x)\ln(1-p) \\
\frac{d \ln L(p)}{dp} &= \frac{x}{p} - \frac{n-x}{1-p} \\
\frac{d^2 \ln L(p)}{dp^2} &= -\frac{x}{p^2} - \frac{n-x}{(1-p)^2} \\
I(p) &= -E\left(\frac{d^2 \ln L(p)}{dp^2}\right) = \frac{n}{p} + \frac{n}{1-p} = \frac{n}{p(1-p)}
\end{aligned}$$

Confidence interval for proportions

The maximum likelihood estimate $\hat{p} = X/n$ is obtained by

$$\frac{d \ln L(p)}{dp} = \frac{x}{p} - \frac{n-x}{1-p} = 0$$

Using the asymptotic properties of maximum likelihood estimate

$$P\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2} \right) \approx 1 - \alpha$$

Using observed Fisher information, the $100(1-\alpha)\%$ confidence interval is

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

Confidence interval for proportions

We use the asymptotic properties of MLE to construct confidence interval. When n is large and p is not too small, the coverage probability is approximately correct. But if n is not sufficiently large or if p is fairly close to 0 or 1, improvements are needed.

The **Wilson score method** directly solve the inequality

$$z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}$$

to obtain the confidence interval

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

Confidence interval for proportions

Agresti and Coull (1998) suggested that we use $\tilde{p} = (X+2)/(n+4)$ as an estimator for p when n is small or if X is close to 0 or n. The confidence interval is

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$$

If we form 95% confidence interval, $z_{\alpha/2} = 1.96 \approx 2$. The 95% confidence interval using the Wilson score method, it is centered at

$$\frac{\hat{p} + z^2_{\alpha/2}/(2n)}{1 + z^2_{\alpha/2}/n} = \frac{X + z^2_{\alpha/2}/2}{n + z^2_{\alpha/2}} \approx \frac{X+2}{n+4}$$

Thus it is roughly consistent with the Agresti and Coull method.

Distribution-free confidence intervals

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be a random sample sorted from the smallest to the largest. We call $X_{(j)}$ the jth order statistics of the random sample. For the 100p% percentile of the distribution m, we have

$$P(X_{(i)} < m < X_{(j)}) = \sum_{k=i}^{j-1} \mathbf{C}_n^k p^k (1-p)^{n-k} = 1 - \alpha$$

This approach only uses the order statistics to construct confidence intervals. Little is assumed about the underlying distribution, except that the distribution is continuous. Thus, these confidence intervals are called **distribution-free confidence intervals**.

Distribution-free confidence intervals

Example: Suppose we have a sample $X_1 < X_2 < X_3 < X_4 < X_5$. One confidence interval of the median m is

$$P(X_1 < m < X_5) = \sum_{k=1}^{4} \mathbf{C_5^k} (\frac{1}{2})^k (\frac{1}{2})^{5-k} = 0.9375$$

$$P(X_2 < m < X_4) = \sum_{k=2}^{3} \mathbf{C_5^k} (\frac{1}{2})^k (\frac{1}{2})^{5-k} = 0.625$$

The interval (X_1, X_n) tends to get wider as n increases, thus we are not "pinning down" m very well. However, if we used the interval (X_2, X_{n-1}) or (X_3, X_{n-2}), we would obtain shorter intervals, but also smaller confidence coefficient.

Distribution-free confidence intervals

As you can see, confidence interval based on order statistic has a prominent shortcoming: we can calculate the confidence coefficient of an interval, but we cannot construct an interval with a pre-specified confidence coefficient.

This approach, therefore, is not widely used in practice. Only use it if there are not other available approach to calculate confidence interval.

Resampling based confidence intervals

Suppose that we need to find the distribution of some statistic, but we do not know its sampling distribution. We observed the values of $X_1, X_2, \ldots, X_n$. The empirical distribution found by placing weight $1/n$ on each X_i is a best estimate of that distribution. A resampling based confidence interval can be constructed using the following steps:

- Sample from $X_1, X_2, \ldots, X_n$ with replacement;
- Calculate the statistic of interest from the sample drawn;
- Repeat the above procedures many times to obtain a empirical distribution of the statistic;
- Obtain the confidence interval of the statistic from its empirical distribution

This approach is also referred to as **bootstrapping.** It allows us to substitute computation for theory for statistical inference.

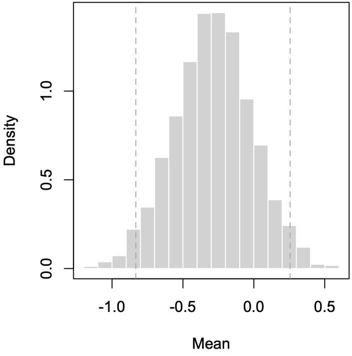
Resampling based confidence interval

Example: We have a random sample of size 10:

$0.17, -0.27, -1.70, 0.89, -0.14, 0.88, -0.87, 0.25, -1.65, -0.45.$

Use bootstrapping to find the 95% confidence interval of the mean μ.

Using 5000 iterations of resampling and the percentile method, we obtain the 95% confidence interval as $(-0.832, 0.256)$.



Resampling based confidence interval

A few comments about resampling based confidence interval

- After obtaining the empirical distribution of the statistic, there are alternative methods in addition to the percentile method. The resulting CI can differ depending on which method you choose.
- Resampling approach is effective when the original sample size is big. After all, the method relies on using the empirical distribution of data to approximate the underlying distribution.
- Iterations of resampling should be sufficiently large to obtain reliable empirical distribution of the statistic. If computation is not too time consuming, it is better to have large number of iterations just to be safe.