Interval Estimation

Yanrong Yang

Research School of Finance, Actuarial Studies and Statistics
The Australian National University

April 26, 2017

Content

1. Definition of Interval Estimation
2. Examples to Illustrate How to Find Interval Estimators
3. Summary

Interval Estimator

Let $Y_1, \ldots, Y_n$ be a sample from a population whose probability distribution depends on an unknown parameter θ.

Definition (Interval Estimator or Confidence Interval)

Suppose that we can find two functions of $Y_1, \ldots, Y_n$ which are denoted by $L = g(Y_1, \ldots, Y_n)$, $U = h(Y_1, \ldots, Y_n)$, such that $P(L \leq \theta \leq U) = 1 - \alpha$. Then $[L, U]$ is a $100(1-\alpha)\%$ confidence interval (CI) for θ.

1. L: the lower bound (LB);
2. U: the upper bound (UB);
3. $1 - \alpha$: the coverage coefficient.

Example 4

Question: Suppose that 6.2 is a number chosen randomly between 0 and c. Find an 80% confidence interval for c.

Analysis:

1. $Y_1 \sim U(0, c)$, $n = 1$.
2. A pivotal quantity $X = Y_1/c \sim U(0, 1)$.
3. $0.8 = P(0.1 \leq X \leq 0.9) = P(0.1 \leq Y_1/c \leq 0.9) = P(10Y_1/9 \leq c \leq 10Y_1)$.
4. $[10Y_1/9, 10Y_1]$ is an 80% CI for c.

Remark

According to the definition of CI, $1 - \alpha = 0.8$ is the coverage coefficient; $L = g(Y_1) = 10Y_1/9$ is the lower bound; and $U = h(Y_1) = 10Y_1$ is the upper bound.

Different Styles of CIs

There are three kinds of CIs, including central CI, upper range CI and lower range CI.

Definition

$I = [L, U] = [g(Y_1, \ldots, Y_n), h(Y_1, \ldots, Y_n)]$ be a $100(1-\alpha)\%$ confidence interval for θ.

1. Upper range CI: $[L, \infty)$, i.e. $P(\theta \geq L) = 1 - \alpha$.
2. Lower range CI: $(-\infty, U]$, i.e. $P(\theta \leq U) = 1 - \alpha$.
3. Central CI: $[L, U]$ such that $P(\theta < L) = P(\theta > U) = \alpha/2$.

Example 5

Question: Find an upper and a lower range CI for c in Example 4.
Analysis:

1. $0.8 = P(X \leq 0.8) = P(Y_1/c \leq 0.8) = P(5Y_1/4 \leq c)$.
2. An upper range 80% CI for c is $[5Y_1/4, \infty)$.
3. $0.2 = P(X \geq 0.2) = P(Y_1/c \geq 0.2) = P(c \leq 5Y_1)$.
4. A lower range 80% CI for c is $(-\infty, 5Y_1]$.

Remark

In Example 4, $Y_1 = 6.2$. Then the upper range and lower range CIs are $[7.75, \infty)$ and $(-\infty, 31]$ respectively. For the lower range CI, since c should be non-negative and no less than $Y_1 = 6.2$, a refined lower range CI is $[6.2, 31]$.

Example 6

Question: Suppose that 1.2, 3.9 and 2.4 are a random sample from a normal distribution with variance 7. Find a 95% confidence interval for the normal mean.

Analysis:

1. $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$, $n = 3$, $\sigma^2 = 7$.
2. A pivotal quantity $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.
3. $1 - \alpha = P(-z_{\alpha/2} < Z < z_{\alpha/2}) =$ $P\left(-z_{\alpha/2}\sigma/\sqrt{n} < \bar{Y} - \mu < z_{\alpha/2}\sigma/\sqrt{n}\right) =$ $P\left(\bar{Y} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{Y} + z_{\alpha/2}\sigma/\sqrt{n}\right)$.
4. A $1 - \alpha$ CI is $(\bar{Y} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{Y} + z_{\alpha/2}\sigma/\sqrt{n})$.

Remark

In this Example, $1 - \alpha = 0.95$, $z_{\alpha/2} = z_{0.025} = 1.96$, $n = 3$, $\bar{y} = 2.5$, $\sigma^2 = 7$.

Example 7

Question: Suppose that 1.2, 3.9 and 2.4 are a random sample from a normal distribution with unknown variance. Find a 95% confidence interval for the normal mean.

Analysis:

1. $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$, $n = 3$.
2. A pivotal quantity $T = \frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$.
3. $1-\alpha = P(-t_{\alpha/2} < T < t_{\alpha/2}) = P(\bar{Y} - t_{\alpha/2}S/\sqrt{n} < \mu < \bar{Y} + t_{\alpha/2}S/\sqrt{n})$.
4. A $1-\alpha$ CI for μ is $(\bar{Y} - t_{\alpha/2}(n-1)S/\sqrt{n}, \bar{Y} + t_{\alpha/2}(n-1)S/\sqrt{n})$.

Remark

$t_{\alpha/2}(n-1) = t_{0.025}(2) = 4.303$.

Example 8

Question: 200 people were randomly sampled from the population of Australia, and their heights measured. The sample mean was 1.673 and the sample standard deviation was 0.310. Find a 95% confidence interval for the average height of all Australians.

Analysis:

1. Y_i is the ith height, $i = 1, \ldots, n$.
2. $n = 200$, $\bar{Y} = 1.673$, $S = 0.310$.
3. A pivotal quantity $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1)$.
4. $Z = \frac{\bar{Y}-\mu}{S/\sqrt{n}} = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \cdot \frac{S}{\sigma} \dot{\sim} N(0,1)$ is a practical pivotal quantity.
5. A $1-\alpha$ CI for μ is $(\bar{Y} - z_{\alpha/2}S/\sqrt{n}, \bar{Y} + z_{\alpha/2}S/\sqrt{n})$.

Example 9

Question: Suppose that we toss a bent coin 100 times and get 72 heads. Find a 95% confidence interval for the probability of a head.

Analysis:

1. Y: the number of heads out of the $n = 100$ tosses;
   p: the probability of a head on a single toss.
2. $Y \sim Bin(n, p)$ and its value is $y = 72$.
3. $Y = \sum_{i=1}^{n} Y_i$, $Y_i \sim Bern(p)$ and then by CLT, $Y \dot{\sim} N(Np, np(1-p))$.
4. A pivotal quantity $Z_1 = \frac{Y-np}{\sqrt{np(1-p)}} \dot{\sim} N(0,1)$.
5. $1-\alpha = P\left(-z_{\alpha/2} < \frac{Y-np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right) =$
   $P\left(\left[\frac{Y-np}{\sqrt{np(1-p)}}\right]^2 < z_{\alpha/2}^2\right) =$
   $P\left(p^2(1+z_{\alpha/2}^2/n) - p(2Y/n + z_{\alpha/2}^2/n) + Y^2/n^2 < 0\right) = P(a < p < b)$,

Example 9 continuing

where a and b are the roots of the quadratic equality $p^2(1 + z^2_{\alpha/2}/n) - p\left(2\hat{p} + z^2/n\right) + \hat{p}^2$. So the $1 - \alpha$ CI for p is

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

Another Approach for Example 9

1. Another pivotal quantity
$$Z_2 = \frac{Y-np}{\sqrt{n\hat{p}(1-\hat{p})}} = \frac{Y-np}{\sqrt{np(1-p)}} \cdot \frac{\sqrt{np(1-p)}}{\sqrt{n\hat{p}(1-\hat{p})}} \dot{\sim} N(0,1).$$
2. $1-\alpha = P\left(-z_{\alpha/2} < \frac{Y-np}{\sqrt{n\hat{p}(1-\hat{p})}} < z_{\alpha/2}\right) =$
$$P\left(\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$
3. A $1-\alpha$ CI for p is $\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$.

CI for the difference between two population means

Consider two samples $X_1, \ldots, X_n \sim i.i.d.(\mu_X, \sigma_X^2)$ and $Y_1, \ldots, Y_m \sim i.i.d.(\mu_Y, \sigma_Y^2)$. $(X_1, \ldots, X_n) \perp (Y_1, \ldots, Y_m)$. The goal is to construct a $1-\alpha$ CI for $\mu_X - \mu_Y$.

As n and m are large, approximate CI can be found by

1. σ_X^2 and σ_Y^2 are known: pivotal quantity $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \dot{\sim} N(0,1)$.
2. σ_X^2 and σ_Y^2 are unknown: $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}} \dot{\sim} N(0,1)$.

As n and m are fixed, moreover, X_i and Y_i are normal,

1. σ_X^2 and σ_Y^2 are known: pivotal quantity $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \sim N(0,1)$.
2. $\sigma_X^2 = \sigma_Y^2$ are unknown: $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim t(n+m-2)$.

Some Technique

Why the statistic $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \dot{\sim} N(0,1)$?

1. By CLT, we have $\frac{\bar{X}-\mu_X}{\sqrt{\sigma_X^2/n}} \xrightarrow{d} N(0,1)$ and $\frac{\bar{Y}-\mu_Y}{\sqrt{\sigma_Y^2/m}} \xrightarrow{d} N(0,1)$.
2. $\frac{\bar{X}-\mu_X}{\sqrt{\sigma_Y^2/m}} = \frac{\sqrt{\sigma_X^2/n}}{\sqrt{\sigma_Y^2/m}} \cdot \frac{\bar{X}-\mu_X}{\sqrt{\sigma_X^2/n}} \xrightarrow{d} N\left(0, \frac{\sigma_X^2}{\sigma_Y^2} \cdot \lim \frac{m}{n}\right)$.
3. By $X \perp Y$, $\frac{\bar{X}-\mu_X}{\sqrt{\sigma_Y^2/m}} - \frac{\bar{Y}-\mu_Y}{\sqrt{\sigma_Y^2/m}} \xrightarrow{d} N\left(0, \frac{\sigma_X^2}{\sigma_Y^2} \cdot \lim \frac{m}{n} + 1\right)$

Example 10

Question: You have a bent $1 coin and a bent $2 coin. You toss the $1 coin 200 times and get 108 heads. You toss the $2 coin 300 times and get 141 heads. Find a 90% CI for the difference between the probability of a head on the $1 coin and the probability of a head on the $2 coin.

Analysis:

1. $X \sim Bin(n,p)$, $Y \sim Bin(m,q)$, $X \perp Y$.
2. A pivotal quantity is $\frac{\hat{p}-\hat{q}-(p-q)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{\hat{q}(1-\hat{q})}{m}}} \dot{\sim} N(0,1)$.
3. $1-\alpha = P\left(-z_{\alpha/2} < \frac{\hat{p}-\hat{q}-(p-q)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{\hat{q}(1-\hat{q})}{m}}} < z_{\alpha/2}\right) = P\left(a < p-q < b\right)$ with $a = \hat{p}-\hat{q}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{\hat{q}(1-\hat{q})}{m}}$ and $b = \hat{p}-\hat{q}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{\hat{q}(1-\hat{q})}{m}}$.

Example 11

Question: Suppose that $Y_1, \ldots, Y_n \sim i.i.d.N(\mu, \sigma^2)$. Find a $1-\alpha$ CI for σ^2.

Analysis:

1. A pivotal quantity is $Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
2. $1-\alpha = P\left(\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi_{\alpha/2}(n-1)\right) = P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$.
3. A $1-\alpha$ CI for σ^2 is $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$.

Summary

1. Distinguish point estimation and interval estimation;
2. how to find an interval estimation or confidence interval;
3. exact CI and Approximate CI.