Statistics

Interval Estimation

Shiu-Sheng Chen

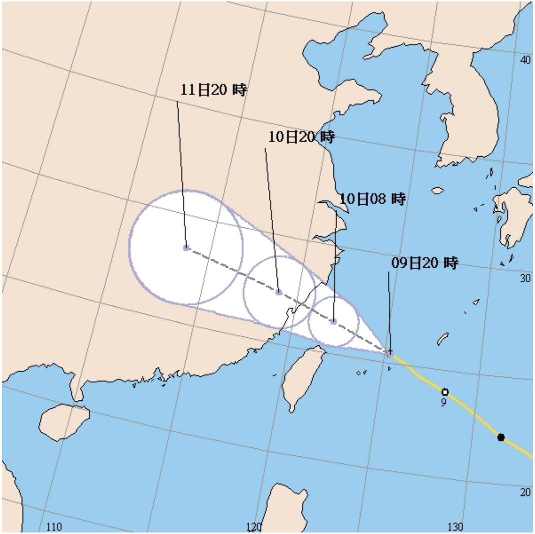
Department of Economics
National Taiwan University

Fall 2019

Section 1

Interval Estimation

Typhoon Track Forecast



Interval Estimator

- An interval estimator is a random interval (or range) in which we can be confident about that it contains the true unknown population parameter.

Interval Estimator

Definition (Interval Estimator)

Let $\{X_i\}_{i=1}^n \sim^{i.i.d.} f(x, \theta)$, and $L(X_1, \ldots, X_n) < U(X_1, \ldots, X_n)$ be two statistics such that

$$P\left(L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n)\right) = 1 - \alpha$$

The random interval $[L(X_1, \ldots, X_n), U(X_1, \ldots, X_n)]$ is called an $100(1-\alpha)\%$ interval estimator of the parameter θ.

- $1 - \alpha$ is called the coverage probability or simply, the coverage
- By construction, we would like to find a random interval such that it can contain the true parameter with probability $1 - \alpha$

Interval Estimator

Coverage Probability

$$P\left(L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n)\right) = 1 - \alpha$$

- In general, $1 - \alpha = 0.90$, 0.95 or 0.99
- Fixed parameter: θ
- Random variables: $L(X_1, \ldots, X_n)$ and $U(X_1, \ldots, X_n)$
  - $L(X_1, \ldots, X_n)$ and $U(X_1, \ldots, X_n)$ can be either linear or nonlinear.

Example 1

- Let

$$\{X_i\} \sim^{i.i.d.} N(\mu, \sigma^2)$$

and assume that σ is known for the sake of simplicity

- We would like to construct a 95% interval estimator of μ, which is of the form:

$$[\bar{X}_n - c, \ \ \bar{X}_n + c]$$

That is,

$$L(X_1, X_2, \ldots, X_n) = \bar{X}_n - c,$$

$$U(X_1, X_2, \ldots, X_n) = \bar{X}_n + c.$$

- How to find c? $\Rightarrow$ It depends on the coverage probability

Example 1

- Let $1-\alpha = 0.95$, we want to find

$$P(\bar{X}_n - c \leq \mu \leq \bar{X}_n + c) = 0.95$$

- Clearly,

$$\{X_i\} \sim^{i.i.d.} N(\mu, \sigma^2) \;\Rightarrow\; \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \;\Rightarrow\; \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$P\left(-Z_{0.025} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{0.025}\right) = 0.95$$

Example 1

- Rearrange

$$P\left(\bar{X}_n - Z_{0.025}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + Z_{0.025}\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- Since $Z_{0.025} = 1.96$,

$$c = 1.96\frac{\sigma}{\sqrt{n}}$$

Example 1

- Hence,

$$P\left(\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- The random interval

$$\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}},\ \ \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

is called a 95% interval estimator of μ.

  - Suppose that $\sigma = 15$, $n = 100$, then the interval estimator is

$$\left[\bar{X}_n - 2.94,\ \ \bar{X}_n + 2.94\right]$$

  - Moreover, if $\bar{X}_n = 170$, the interval estimate is

$$\left[167.06,\ \ 172.94\right]$$

Interval Estimator vs. Interval Estimate

- Interval Estimator

$$P\Big(\mu \in [\bar{X}_n - 2.94,\ \ \bar{X}_n + 2.94]\Big) = 0.95$$

- Interval Estimate (e.g., $\bar{X}_n = 170$)

$$P\Big(\mu \in [167.06,\ \ 172.94]\Big) = 0 \text{ or } 1$$

Section 2

Pivotal Quantity and Interval Estimator

Pivotal Quantity

Definition (Pivotal Quantity)

Let $\{X_i\}_{i=1}^n \sim^{i.i.d.} f(x,\theta)$, and $\varphi(\theta, X_1, X_2, \ldots, X_n)$ be a random variable whose distribution does not depend on θ. Then $\varphi(\cdot)$ is called a *Pivotal Quantity* or simply a *Pivotal*.

- For example, suppose that $\{X_i\}_{i=1}^n \sim^{i.i.d.} N(\mu, \sigma^2)$, and σ known. Then it is clear that

$$\varphi(\mu, \sigma, X_1, X_2, \ldots, X_n) = \left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \right) \sim N(0,1)$$

  That is, its distribution does not depend on μ and σ

Construct Interval Estimators via a Pivotal

The Procedure

- Determine $1-\alpha = 0.90$, 0.95 or 0.99
- Construct a pivotal quantity:

$$\varphi(\theta, X_1, X_2, \ldots, X_n)$$

- Use the sampling distribution of $\varphi(\cdot)$ to find l and u such that
  - $P(l \leq \varphi(\theta, X_1, X_2, \ldots, X_n) \leq u) = 1-\alpha$
  - $P(\varphi(\theta, X_1, X_2, \ldots, X_n) \leq l) = \frac{\alpha}{2}$
  - $P(\varphi(\theta, X_1, X_2, \ldots, X_n) \geq u) = \frac{\alpha}{2}$
- For instance, if the sampling distribution of φ is $N(0,1)$, then

$$l = -Z_{\frac{\alpha}{2}}, \quad u = Z_{\frac{\alpha}{2}}$$

Construct Interval Estimators via a Pivotal

The Procedure

- Rearrange

$$P(l \leq \varphi(\theta, X_1, X_2, \ldots, X_n) \leq u) = 1 - \alpha$$

into

$$P\Big(L(X_1, X_2, \ldots, X_n) \leq \theta \leq U(X_1, X_2, \ldots, X_n)\Big) = 1 - \alpha$$

- Hence, the interval estimator of θ is

$$\Big[L(X_1, X_2, \ldots, X_n),\ \ U(X_1, X_2, \ldots, X_n)\Big]$$

Construct Interval Estimators via a Pivotal

- Now we return to Example 1 and use the pivotal to construct a 95% IE for μ:

$$\{X_i\}_{i=1}^n \sim^{i.i.d.} N(\mu, \sigma^2), \quad \sigma^2 \text{ known}$$

- Construct the pivotal:

$$\varphi = \left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \right) \sim N(0,1)$$

- Hence, $l = -Z_{0.025} = -1.96$, $u = Z_{0.025} = 1.96$, and

$$P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96 \right) = 0.95$$

Construct Interval Estimators via a Pivotal

- Rearrange

$$P\left(-1.96 \le \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \le 1.96\right) = 0.95$$

into

$$P\left(\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- That is, a 95% IE for μ is

$$\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Example 2: Interval Estimator for μ with unknown σ^2

- Given

$$\{X_i\}_{i=1}^{n} \sim^{i.i.d.} N(\mu, \sigma^2), \quad \sigma^2 \text{ unknown}$$

  construct a $100(1-\alpha)\%$ interval estimator of μ

- Construct the pivotal:

$$\varphi = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}}$$

  - Note that we replace σ with S_n
  - What is the sampling distribution of φ?

Example 2: Interval Estimator for μ with unknown σ^2

- We have shown that

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1), \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

  and $\bar{X}_n \perp S_n^2$

- Hence,

$$\varphi = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \sim t(n-1)$$

- That is,

$$P\left(-t_{\frac{\alpha}{2}}(n-1) \le \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \le t_{\frac{\alpha}{2}}(n-1)\right)$$

Example 2: Interval Estimator for μ with unknown σ^2

- Hence, a $100(1-\alpha)\%$ interval estimator of μ is

$$\left[\bar{X}_n - t_{\frac{\alpha}{2}}(n-1)\frac{S_n}{\sqrt{n}},\quad \bar{X}_n + t_{\frac{\alpha}{2}}(n-1)\frac{S_n}{\sqrt{n}}\right]$$

Example 3: Approximate Interval Estimator for μ with unknown σ^2

- Given

$$\{X_i\}_{i=1}^n \sim^{i.i.d.} (\mu, \sigma^2), \quad \sigma^2 \text{ unknown}$$

construct an approximate $100(1-\alpha)\%$ interval estimator of μ

- Construct the pivotal:

$$\varphi = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{S_n}\right)$$

  - What is the sampling distribution of φ? We do not know.

Example 3: Approximate Interval Estimator for μ with unknown σ^2

- The asymptotic theory saves us.
- By CLT, CMT and Slutsky's theorem

$$\varphi = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$$

- A $100(1-\alpha)\%$ approximate interval estimator of μ is

$$\left[\bar{X}_n - Z_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}, \quad \bar{X}_n + Z_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}\right]$$

Example 4: Approximate Interval Estimator for Population Proportion p

- Given $\{X_i\} \sim^{i.i.d.}$ Bernoulli(p), we would like to find an approximate IE for p
  - Recall that $\hat{p} = \bar{X}_n$, $Var(\hat{p}) = Var(\bar{X}_n) = \frac{p(1-p)}{n}$
- By CLT, CMT and Slutsky's theorem

$$\varphi = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{d} N(0, 1)$$

- A $100(1-\alpha)\%$ approximate interval estimator of μ is

$$\left[\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Confidence Intervals