4. INTERVAL ESTIMATION

Interval estimation is more informative than point estimation, and is very important in practice.

Confidence Interval is the most commonly used interval estimation. More general type is *Confidence Set* which may consist of several intervals.

4.1 What is an interval estimator

Example 1. Let us start with a simple example. A random sample $X_1, \cdots, X_n$ are drawn from $N(\mu, 1)$. Then

$$\sqrt{n}(\bar{X} - \mu) \sim N(0, 1).$$

Hence

$$P(-1.96 \le \sqrt{n}(\bar{X} - \mu) \le 1.96) = 0.95,$$

or

$$P(\bar{X} - 1.96/\sqrt{n} < \mu < \bar{X} + 1.96/\sqrt{n}) = 0.95.$$

So a 95% confidence interval for μ is

$$(\bar{X} - 1.96/\sqrt{n},\ \bar{X} + 1.96/\sqrt{n}).$$

Suppose $n = 4$, $\bar{X} = 2.25$. Then a 95% C.I. is $(2.25 - 0.98, 2.25 + 0.98) = (1.27, 3.23)$.

Question: what is $P(1.27 < \mu < 3.23)$? — Note μ is a unknown constant!

Answer: $(1.27, 3.23)$ is one realisation of the **r**andom interval $(\bar{X} - 0.98,\ \bar{X} + 0.98)$ which covers μ with probability 0.95.

Definition. Suppose the joint distribution of $\mathbf{X} = (X_1, \cdots, X_n)$ depends some parameter θ.

If $\theta_l(\mathbf{X})$ and $\theta_u(\mathbf{X})$ are two statistics for which

$$P\{\theta_l(\mathbf{X}) < \theta < \theta_u(\mathbf{X})\} = 1 - \alpha,$$

$(\theta_l(\mathbf{X}),\ \theta_u(\mathbf{X}))$ is called a $100(1-\alpha)\%$ *confidence interval* for θ.

Remark. (i) $1 - \alpha$ is called the confidence level, which is usually set at 0.95, 0.99 or 0.999. Naturally for given α, we shall search for the interval with the shortest length $\theta_u(\mathbf{X}) - \theta_l(\mathbf{X})$, which gives the **most accurate** estimation.

(ii) We may have $\theta_l(\mathbf{X}) = -\infty$ or $\theta_u(\mathbf{X}) = \infty$, giving a one-sided interval.

(iii) In general we may use non-interval type set $S(\mathbf{X}) \subset \Theta$ as the estimator for θ, i.e.

$$P\{\theta \in S(\mathbf{X})\} = 1 - \alpha.$$

We call $S(\mathbf{X})$ a $100(1-\alpha)\%$ **c**onfidence set.

4.2 Method of pivotal functions

Set-up: Observations – $\mathbf{X}$, Parameter of interest – θ (possible other parameters).

Definition. A function of $\mathbf{X}$ and θ alone is a *pivotal function* if its distribution does not depend on any unknown parameters (i.e. can be calculated numerically).

Examples of pivotal functions: Let $X_1, \cdots, X_n$ be a random sample.

1. for exponential distribution with mean θ, $\theta^{-1}\sum_j X_j$ is a pivotal,
2. for $N(\mu, 1)$, $\bar{X} - \mu$ is a pivotal,
3. for $N(\mu, \sigma^2)$, $\sqrt{n}(\bar{X} - \mu)/S$ is a pivotal.

Construction of C.I. based on a pivotal:

Step 1: Find a pivotal function $T = T(\mathbf{X}, \theta)$, and identify its distribution

Step 2: Use the distribution to identify a range of values t_1, t_2 such that

$$P(t_1 < T < t_2) = 1 - \alpha.$$

Step 3: Manipulate the inequalities

$$T > t_1 \quad \text{and} \quad T < t_2$$

to find a set of values for θ. The values included in this set form a $100(1-\alpha)\%$ C.I. for θ.

Example 2. (*Student's t-interval*)
$(X_1, \cdots, X_n)$ is a random sample from $N(\mu, \sigma^2)$. A pivotal function is

$$T = \sqrt{n}(\bar{X} - \mu)/S \sim t_{n-1},$$

where $S^2 = \frac{1}{n-1}\sum_j (X_j - \bar{X})^2$.

For $n = 10$, we have $P(-2.26 < T < 2.26) = 0.95$. Note that $T > -2.26$ is equivalent to

$$\mu < \bar{X} + 2.26S/\sqrt{n}.$$

(Similar for $T < 2.26$.) A 95% C.I. for μ is $(\bar{X} - 2.26S/\sqrt{n},\ \bar{X} + 2.26S/\sqrt{n})$.

Note. (i) In general it is not easy to identify a pivotal function.

(ii) For location parameter μ, $\bar{X} - \mu$ is **likely** to be pivotal.

(iii) The MLE is asymptotically pivotal:

$$\widehat{\theta} \sim N(\theta, \mathcal{I}^{-1}(\theta)).$$

Hence, $\sqrt{\mathcal{I}(\theta)}(\widehat{\theta} - \theta) \sim N(0, 1)$ approximately for large n.

4.3 Inverting Tests

One easy way to get a confidence set is to invert a test; see Example 1.

Theorem. Suppose that we have a size α test for the Null Hypothesis $H_0: \ \theta = \theta_0$. For each $\theta_0 \in \Theta$, suppose the set $A(\theta_0)$ is the collection of $\mathbf{x}$ for which H_0 is *not rejected*. Then the set

$$S(\mathbf{x}) = \{\theta : \mathbf{x} \in A(\theta), \theta \in \Theta\}$$

is a family of confidence sets for θ at confidence level $1 - \alpha$.

The proof of this is obvious, because when θ is the true parameter value

$$P_\theta[\theta \in S(\mathbf{X})] = P_\theta[\mathbf{X} \in A(\theta)] = 1 - \alpha.$$

Example 3. Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from an exponential distribution with mean $1/\lambda$. The UMPT with size 5% for

$$H_0: \ \lambda = \lambda_0 \quad \text{against} \quad H_1: \ \lambda > \lambda_0$$

is to reject H_0 when $\sum X_i < \frac{1}{2\lambda_0}\chi^2_{(2n),0.05}$, where $\chi^2_{(2n),0.05}$ is the lower 5% point of the distribution χ^2_{2n}. The acceptance region here is $\sum X_i > \frac{1}{2\lambda_0}\chi^2_{(2n),0.05}$, which on inversion gives

$$\left(\frac{\chi^2_{(2n),0.05}}{2\sum X_i},\ \infty\right)$$

as the confidence set for λ at the 95% level of confidence.

4.4 Bootstrap confidence intervals

Let $X_1, \cdots, X_n$ be sample from unknown distribution of F. We are interested in constructing confidence intervals for some characteristic $\theta = \theta(F)$ of distribution F.

We adopt the so-called *nonparametric bootstrap method* now. It draws a bootstrap sample

$$X_1^*, \cdots, X_n^*$$

independently from the uniform distribution over n discrete points $\{X_1, \cdots, X_n\}$.

Now in principle the distribution of any statistic based on $(X_1^*, \cdots, X_n^*)$ is known, *conditionally on* $(X_1, \cdots, X_n)$.

4.4.1 Bootstrap percentiles

Let $\widehat{\theta} = T(X_1, \cdots, X_n)$ be an estimator for θ. Define

$$\widehat{\theta}^* = T(X_1^*, \cdots, X_n^*).$$

Let l_α^* and u_α^* be, respectively, the lower and upper α-point of the distribution $\widehat{\theta}^*$, i.e.

$$P\{\widehat{\theta}^* \le l_\alpha^* | X_1, \cdots, X_n\} = \alpha,$$

$$P\{\widehat{\theta}^* > u_\alpha^* | X_1, \cdots, X_n\} = \alpha.$$

Then **the** $(1-\alpha)$**100%-th bootstrap interval** for θ is defined as

$$(l^*_{\alpha/2},\ u^*_{\alpha/2}].$$

In practice we draw B sets of bootstrap samples for some large B, resulting in B bootstrap estimates $\widehat{\theta}_1^*, \cdots, \widehat{\theta}_B^*$. Then $l^*_{\alpha/2}$ and $u^*_{\alpha/2}$ are, respectively, the $[B\alpha/2]$-th smallest and the $[B\alpha/2]$-th largest values among $\widehat{\theta}_1^*, \cdots, \widehat{\theta}_B^*$.

Note. Both $l^*_{\alpha/2}$ and $u^*_{\alpha/2}$ are statistics depending on $\{X_1, \cdots, X_n\}$.

Intuitively, we **m**ust start with a *good* estimator $\widehat{\theta}$. In fact, if $\widehat{\theta} = T(X_1, \cdots, X_n)$ fulfils some conditions, it holds that as $n \to \infty$,

$$P(l^*_{\alpha/2} < \theta \le u^*_{\alpha/2}) \to 1 - \alpha.$$

4.4.2 Bootstrap-t

Suppose we have a *studentised 'pivot'*

$$V = (\widehat{\theta} - \theta)/\widehat{\sigma},$$

where $\widehat{\theta}$ is an estimator for θ, and $\widehat{\sigma}^2$ is an estimator for $\mathrm{Var}(\widehat{\theta})$. Define its bootstrap version

$$V^* = (\widehat{\theta}^* - \widehat{\theta})/\widehat{\sigma}^*.$$

Let

$$\begin{aligned} \alpha/2 &= P\{V^* \le l^*_{\alpha/2} | X_1, \cdots, X_n\} \\ &= P\{V^* > u^*_{\alpha/2} | X_1, \cdots, X_n\}. \end{aligned}$$

Then the $100(1-\alpha)\%$-th bootstrap-t interval for θ is

$$[\widehat{\theta} - \widehat{\sigma} u^*_{\alpha/2},\ \widehat{\theta} - \widehat{\sigma} l^*_{\alpha/2}).$$

Remark. The estimator $\widehat{\sigma}^2$ may be obtained by bootstrap, then $\widehat{\sigma}^{2*}$ will be calculated via a **n**ested bootstrap.

Ideally we require that the distribution, or at least the asymptotic distribution, of V does not depend on anything unknown. This will make the approximation

$$P\big\{\widehat{\theta} - \widehat{\sigma} u^*_{\alpha/2} \le \theta < \widehat{\theta} - \widehat{\sigma} l^*_{\alpha/2}\big\} \approx \alpha$$

more accurate.

We give a heuristic argument for the above assertion.

If V is an asymptotic pivot in the sense that

$$P(V \le x) = \Phi(x) + \frac{1}{\sqrt{n}} g(x, \theta) + O(\frac{1}{n}), \tag{5}$$

similarly in the bootstrap world we have

$$\begin{aligned} P(V^* \le x | X_1, \cdots, X_n) &= \Phi(x) + \frac{1}{\sqrt{n}} g(x, \widehat{\theta}) + O_p(\frac{1}{n}) \\ &= \Phi(x) + \frac{1}{\sqrt{n}} g(x, \theta) + O_p(\frac{1}{n}). \end{aligned} \tag{6}$$

Hence

$$l_\alpha = l^*_\alpha + O_p(\frac{1}{n}).$$

In contrast if $\Phi(x)$ depends on θ, then $\Phi(x)$ should be replaced by $\Phi(x, \theta)$ in (5), and by $\Phi(x, \widehat{\theta})$ in (6). Now

$$l_\alpha = l^*_\alpha + O_p(\frac{1}{\sqrt{n}}).$$