Statistical Inference (MA862)

Lecture Slides

Topic 3: Interval Estimation

Interval Estimation

- Our aim is to find an interval in $\Theta \subseteq \mathbb{R}$ such that the interval covers the unknown parameter with a specified high probability.
- Note that for a RV X and two real constants $a > 0$, $b > 0$,

$$P\left(a < X < b\right) = P\left(X < b < bX/a\right).$$

  Though these two probabilities are same, there is a basic difference in RHS and LHS statements. For the LHS, we are taking about probability that a random quantity X belongs to a fixed interval (a, b). For the RHS, we are taking about probability that a random interval $(X, bX/a)$ contains a fixed point b.
- For example, let $X \sim U(0, 1)$, $a = 0.5$, and $b = 1$. In this case $P\left(X < 1 < 2X\right) = P(0.5 < X < 1) = 0.5$.

Interval Estimation

- In interval estimation we will try to find a random interval based on the RS so that it contains the fixed unknown parameter with a pre-specified probability.
- Interval estimation is quite useful in practice. For example one may interested to find a upper limit of mean of toxic level of some drug or food.

Interval Estimation

Definition 3.1: An interval estimate of a real valued parameter θ is any pair of functions $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$ of random sample only (do not involve any unknown parameters) that satisfy $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ for all $\boldsymbol{x}$ in the support of RS. The random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is called an interval estimator of θ.

Remark 3.1:

- Though in the definition the closed interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is considered, the interval may be closed, open or semi-open.
- If $L(\boldsymbol{x}) = -\infty$, then $U(\boldsymbol{x})$ provides an upper limit and the corresponding interval estimator is called lower interval estimator.
- If $U(\boldsymbol{x}) = \infty$, then $L(\boldsymbol{x})$ provides a lower limit and the corresponding interval estimator is called upper interval estimator.

Example

Example 3.1: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, 1)$. Conciser $L_1(\boldsymbol{x}) = x_1 - 1$, $U_1(\boldsymbol{x}) = x_1 + 1$, $L_2(\boldsymbol{x}) = \overline{x} - 1$, and $U_2(\boldsymbol{x}) = \overline{x} + 1$. Then both $[L_1(\boldsymbol{X}), U_1(\boldsymbol{X})]$ and $[L_2(\boldsymbol{X}), U_2(\boldsymbol{X})]$ are interval estimator of μ.
Which one should we use? Note that here the lengths of both intervals are same, hence one should use that which has more probability that the random interval includes μ.

$$
\begin{aligned}
P\left(X_1 - 1 \leq \mu \leq X_1 + 1\right) &= P\left(-1 \leq X_1 - \mu \leq 1\right) = 2\Phi(1) - 1, \\
P\left(\overline{X} - 1 \leq \mu \leq \overline{X} + 1\right) &= P\left(-\sqrt{n} \leq \sqrt{n}\left(\overline{X} - \mu\right) \leq \sqrt{n}\right) \\
&= 2\Phi(\sqrt{n}) - 1.
\end{aligned}
$$

Now as $\Phi(\cdot)$ is an increasing function, we should prefer $[L_2(\boldsymbol{X}), U_2(\boldsymbol{X})]$ over $[L_1(\boldsymbol{X}), U_1(\boldsymbol{X})]$.

Remarks

Remark 3.2:

- In the previous example we talk about maximum probability when the length of the intervals are fixed. Similarly in some other situations, we can consider minimum length for fixed probability. Though we will not pursue these concepts further.
- Clearly we are loosing precision in interval estimation compared to point estimation. Do we have any gain? Consider the previous example. A reasonable point estimator of μ is $\overline{X}$. However, $P\left(\overline{X} = \mu\right) = 0$ as $\overline{X}$ is a CRV. On the other hand $P\left(\overline{X} - 1 \leq \mu \leq \overline{X} + 1\right) > 0$. Hence in interval estimation we have some confidence which we gain by reducing precision.

Confidence Interval (CI)

Definition 3.2: Let $\alpha \in (0, 1)$. An interval estimator $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is said to be a confidence interval of level $1 - \alpha$ (or a $100(1 - \alpha)\%$ confidence interval) if $P_\theta\left(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})\right) \geq 1 - \alpha$ for all $\theta \in \Theta$.

Remark 3.3:

- Typical values of α are 0.1, 0.05, 0.01.
- If we are able to have large number of realization of a RS, and calculate a $100(1 - \alpha)\%$ CI for each realization, then about $100(1 - \alpha)\%$ times the true parameter will be inside the calculated CI. This is the physical mean of a CI and is called relative frequency interpretation of CI.

Method of Finding CI

Definition 3.3: A random variable $T = T(\boldsymbol{X}; \theta)$ is called a pivot (or a pivotal quantity) if the distribution of T does not involve any unknown parameters.

Example 3.2: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ and μ and σ both are unknown. Then $\overline{X} - \mu$ is not a pivot as $\overline{X} - \mu \sim N(0, \sigma^2/n)$. However, $\frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu\right) \sim N(0, 1)$ and $\frac{\sqrt{n}}{S}\left(\overline{X} - \mu\right) \sim t_{n-1}$ and hence are pivot.

Example 3.3: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Exp(\lambda)$. Then $2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$ (why?) and hence is a pivot.

Remark 3.4: Pivot is a function of random sample and unknown parameters, but its' distribution is independent of all unknown parameters. Hence pivot is not a statistic.

Method of Finding CI

- Let T be a pivot.
- Find two real numbers a and b such that

$$P\left(a \leq T(\boldsymbol{X};\theta) \leq b\right) \geq 1-\alpha.$$

  Note that a and b are independent of all unknown parameters as the distribution of T does not involve any unknown parameter.
- A $100(1-\alpha)\%$ CI for θ is

$$C(\boldsymbol{x}) = \left\{\theta \in \Theta : a \leq T(\boldsymbol{x};\theta) \leq b\right\}.$$

  Note that $C(\boldsymbol{x})$ does not involve any unknown parameters as a and b are independent of all unknown parameters.

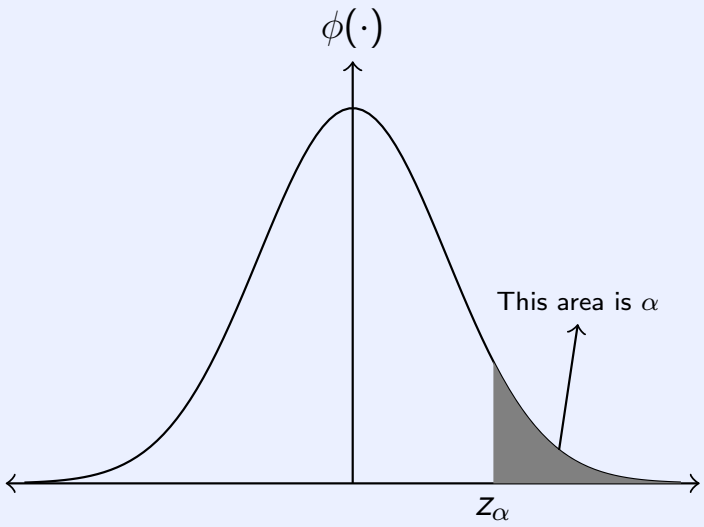
Remark 3.5: If $T(\boldsymbol{x};\theta)$ is monotone in $\theta \in \Theta$ for each $\boldsymbol{x}$, then $C(\boldsymbol{x})$ is an interval. Otherwise it will be a general set.

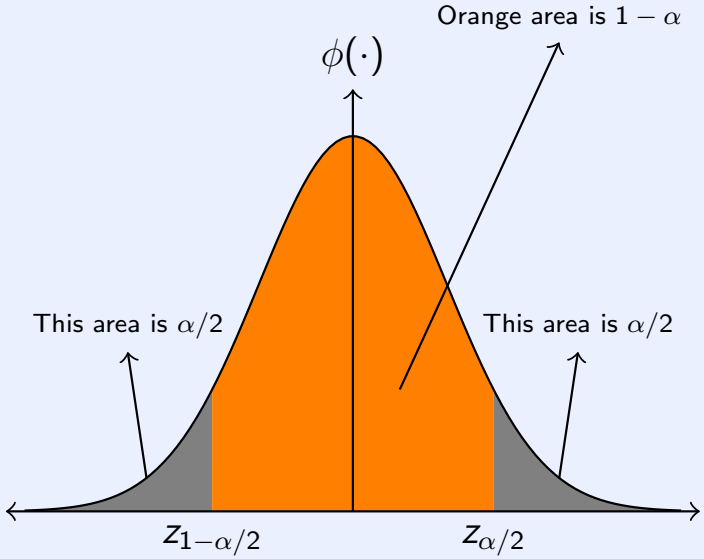
Examples

Example 3.4: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ is unknown and $\sigma > 0$ is known. A $100(1-\alpha)\%$ symmetric CI for μ is

$$C(\boldsymbol{X}) = \left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right],$$

where z_α is a real number such that $P(X > z_\alpha) = \alpha$ for $X \sim N(0, 1)$.





Remarks

Remark 3.6:

- We can have infinite number of choices for a and b. For example $a = z_{1-\alpha_1}$ and $b = z_{\alpha_2}$, where $\alpha_1 > 0$, $\alpha_2 > 0$, and $\alpha_1 + \alpha_2 = \alpha$. Of course they will not be symmetric CI.
- The CI in the previous example is called symmetric CI.
- This is the minimum length CI based on $\overline{X}$.
- z_α is called upper α-point of standard normal distribution.
- Note that in the previous example, $\overline{X}$ is a sufficient statistic for μ and the CI is constructed using sufficient statistic. Whenever possible, we try to find pivot that involves sufficient statistic.

Examples

Example 3.5: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ is known and $\sigma > 0$ is unknown. A $100(1-\alpha)\%$ symmetric CI for σ^2 is

$$c(\boldsymbol{X}) = \left[\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_{n,1-\alpha/2}} \right],$$

where $\chi^2_{n,\alpha}$ is called upper α-point of χ^2-distribution with degrees of freedom n and is a real number such that $P\left(X > \chi^2_{n,\,\alpha}\right) = \alpha$ for a random variable $X \sim \chi^2_n$.

Examples

Example 3.6: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are unknown. We are interested in CI of μ. A $100(1-\alpha)\%$ symmetric CI for μ is

$$c(\boldsymbol{X}) = \left[\overline{X} - \frac{S}{\sqrt{n}} t_{n-1,\alpha/2},\ \overline{X} + \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}\right],$$

where $t_{n,\alpha}$ is called upper α-point of t-distribution with degrees of freedom n.

Example 3.7: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are unknown. A $100(1-\alpha)\%$ symmetric CI for σ^2 is

$$c(\boldsymbol{X}) = \left[\frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\chi^2_{n-1,\alpha/2}},\ \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\chi^2_{n-1,1-\alpha/2}}\right].$$