Lezione 10 - Interval Estimation

Let's assume a random sample $X_1, \dots, X_n \overset{iid}{\sim} f(-,\theta)$, where θ is an unknown parameter, and let us assume the observed sample $x_1, \dots, x_n \in \mathbb{R}^n$

Our goal is to estimate the scalar parameter, θ, with a pair of estimators in order to identify a real interval that is an estimate of possible values of θ.

We have so far only introduced the estimator T, which even in best case scenario where $\underbrace{E(T) = \theta, \forall\theta}_{\Rightarrow \text{ unbiased}}$, if T is absolutely continuous is such that

$$P(T = \theta) = 0 \quad \forall\theta$$

Therefore instead of only using one estimator, we will use a interval estimator which estimates an interval from the sample, within which θ possibly lies.

Replacing the random sample with observed sample, we will have a real interval that is an interval estimate of θ, that is an interval within which we expect the find the true value of the parameter θ.

Confidence Intervals

Let $X_1, \dots, X_n$ be a random sample with density function F_θ and $\theta \in \Theta \subset \mathbb{R}^k$

We define a two-sided confidence interval (CI) for θ with confidence level $1-\alpha$ as a random interval determined by two statistics, $L(X_1, \dots, X_n)$ and $U(X_1, \dots, X_n)$ such that:

$$P_\theta\Big(L(X_1, \dots, X_n) < \theta < U(X_1, \dots, X_n)\Big) = 1 - \alpha$$

Where α is a small number, e.g 1%, 2,5%, 5% or 10%.

The interval estimate will be (l, u) where l, u are the observed values: $l = L(x_1, \dots, x_n)$ and $u = U(x_1, \dots, x_n)$

Confidence Intervals for Gaussian population with KNOWN variance

Fixing a confidence level α (e.g 5%, 10%, 1%)

Let our random sample be $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$, where σ_0^2 is known but μ is not.

We define a <u>two-sided confidence interval</u> for μ with confidence $1-\alpha$ as:

$$\left(\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}},\ \bar{x} + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$$

Where $\bar{x}_n$ is the observed value of $\bar{X}_n$.

This interval has been derived from the pivot (a statistic whose distribution does NOT depend on the parameter of interest):

$$\frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$$

Where:

$$P\Big(\underbrace{-z_{1-\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} < z_{1-\alpha/2}}_{*}\Big) = 1-\alpha$$

We therefore solve the inequality $*$ to obtain of interval.

The length of the confidence interval $\left(\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}},\ \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$ is: $L = 2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}}$

The storer the interval the more precise the estimate, therefore if α is fixed:

$$2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}} \longrightarrow 0 \text{ if } n \longrightarrow +\infty$$

If n is fixed, but the confidence level $(1-\alpha)$ increase ($\Rightarrow \alpha$ decreases) then:

$$1 - \frac{\alpha}{2} \longrightarrow 1 \Leftrightarrow z_{1-\frac{\alpha}{2}} \longrightarrow \infty$$

Hence the length increases, consequently the CI is <u>less precise</u>.

It is therefore necessary to balance confidence level and length.

Note: If $1-\alpha \uparrow$, the precision decreases, whereas if $n \uparrow$, the precision increases.