Statistics

2023 Lectures
Part 9 - Interval Estimation

Institute of Economic Studies
Faculty of Social Sciences
Charles University

Confidence intervals

- So far an estimator was used to produce a single number, hopefully close to unknown parameter.
- Can we find an interval estimator that will cover the unknown parameter with a certain probability?

Let $X_1, \ldots, X_n$ be random sample from the distribution with $f(x, \theta)$ and $\alpha \in (0, 1)$ a fixed number. Typically $\alpha = 0.05$ or 0.01.

Definition 46: A pair of statistics $L(X)$ and $U(X)$ is a $(1-\alpha)-$level confidence interval for θ if for all $\theta \in \Theta$

$$P_\theta(L(X) \leq \theta \leq U(X)) = 1 - \alpha.$$

A statistics L is a $(1-\alpha)$-level lower confidence bound if for all $\theta \in \Theta$

$$P_\theta(L(X) \leq \theta) = 1 - \alpha,$$

a statistics U is a $(1-\alpha)$-level upper confidence bound if for all $\theta \in \Theta$

$$P_\theta(U(X) \geq \theta) = 1 - \alpha.$$

Terminology, ambiguity and notation

- The interval $(L(X), U(X))$ varies from sample to sample and is therefore random and the probability that it covers θ (the "true value" of the parameter) is $1-\alpha$.
- For a specific sample x, $(L(x), U(x))$ may and may not cover θ and all randomness is gone. But since we do not know which is the case, we use the word "confidence".
- To stress out the difference between $(L(X), U(X))$ and $(L(x), U(x))$, we call them the probability and sample confidence interval, respectively.

Notation:

- if $Z \sim N(0,1)$ then z_α denotes $(1-\alpha)$-(lower) quantile of Z, i.e., $P(Z \leq z_\alpha) = 1-\alpha$ and $z_{1-\alpha} = -z_\alpha$, $P(|Z| \leq z_{\frac{\alpha}{2}}) = 1-\alpha$.
- $t_{\alpha,\nu}$ denotes $(1-\alpha)$-(lower) quantile of $T \sim t_\nu$.
- similarly for χ^2, F distributions

Example

Example 90: Let $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ where σ^2 is known. Then

$$\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

is $(1-\alpha)$- level confidence interval of θ.

$$P\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \theta < \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right)$$

which equals $1-\alpha$ from the definition of quantiles and since $\frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

Confidence intervals based on CLTs

- For large samples we can use CLT to derive approximate confidence intervals.

Theorem 60: (without proof) If T is an efficient estimator of θ based on a random sample from distribution with $f(x, \theta)$ then the random variable

$$\sqrt{nI(\theta)}(T - \theta)$$

has asymptotically standard normal distribution.

Theorem 61: (without proof) Let $\hat{\theta}_n$ be the MLE estimator of θ in a problem for which MLE satisfies assumptions of the Theorem 60. Then for large n

$$\left(\hat{\theta}_n - \frac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\theta}_n)}}, \hat{\theta}_n + \frac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\theta}_n)}}\right)$$

is an approximate $(1 - \alpha)$-confidence interval for parameter θ.

Example: Bernoulli distribution

Example 91:

Let $X_1, \ldots, X_n$ be random sample from Bernoulli distribution with p unknown

MLE of p is $\hat{p} = \bar{X}$ and $I(p) = \frac{1}{p(1-p)}$

$Var(\hat{p}) = \frac{1}{n^2} Var S_n = \frac{p(1-p)}{n} = \frac{1}{nI(p)}$

Then we get

$$\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

based on approximation

$$-z_{\frac{\alpha}{2}} < \sqrt{nI(\hat{p})}(\hat{p} - p) < z_{\frac{\alpha}{2}}.$$

Example: $N(\theta, \sigma^2)$, σ^2 unknown

Example 92:

Set $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ and is independent of $\bar{X}$.

Then

$$T = \frac{\bar{X} - \theta}{S}\sqrt{n} \sim t_{n-1}.$$

Hence, if $t_{\frac{\alpha}{2},n-1}$ denotes $(1 - \frac{\alpha}{2})$-quantile of t_{n-1},

$$\left(\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right)$$

is a $(1 - \alpha)$-confidence interval for θ.

Example: $N(\mu, \theta)$, μ known and unknown

Example 93:
First, assume μ is known. Then

$$U = \frac{\sum (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

and

$$P\left(\chi^2_{1-\frac{\alpha}{2},n} < \frac{\sum (X_i - \mu)^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2},n}\right) = 1 - \alpha.$$

If μ is unknown then we replace U by $V = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.

- if μ is known, the former interval is usually shorter (but not always!)
- it is possible to cut the tails of χ^2 in many different ways; for large degrees of freedom, however, the shortest intervals are close to $\frac{\alpha}{2}$.