Stat 2911 Lecture Notes

Class 35, 2017

Uri Keich

Deriving CIs for a location and for a scale parameter (with normal examples), Approximate CIs based on parametric bootstrap

Confidence Intervals

Def. An interval valued function

$$I_\alpha^{(n)} : \mathbb{R}^n \to \{(a,b) \in \mathbb{R}^2 : a < b\}$$

defines a $100(1-\alpha)\%$ confidence interval for θ if for $\forall \theta$ and $X_1, \dots, X_n$ ind. F_θ-distributed RVs

$$P_\theta\left(\theta \in I_\alpha^{(n)}(X_1, \dots, X_n)\right) \geq 1-\alpha.$$

In this case, given a sample $x_1, \dots, x_n$, $I_\alpha(x_1, \dots, x_n)$ is a $100(1-\alpha)\%$ CI for θ.

Constructing CIs

Given a sample $\underline{x} = x_1, \dots, x_n$ from F_θ let $\hat{\theta}(\underline{x})$ be the corresponding estimate of θ.

$\hat{\Theta} = \hat{\theta}(X_1, \dots, X_n)$ is the estimator RV when X_i are iid F_θ RVs.

Example 1) $X_i \sim N(\theta, 1)$, $\hat{\theta}(\underline{x}) = \bar{x}$, $\hat{\Theta} = \bar{X} \sim N(\theta, 1/n)$

Assume that

$$h(\theta; \hat{\theta}) = P_\theta(\hat{\Theta} \geq \hat{\theta}) = P_\theta(\bar{X} \geq \bar{x})$$

is cont. and strictly $\uparrow$ in θ so we can solve for θ_L:

$$h(\theta_L; \hat{\theta}(\underline{x})) = P_{\theta_L}(\hat{\Theta} \geq \hat{\theta}(\underline{x})) = \alpha/2.$$

Similarly, assuming that

$$g(\theta;\hat{\theta}) = P_\theta(\hat{\Theta} \leq \hat{\theta}) = P_\theta(\bar{X} \leq \bar{x})$$

is cont. and strictly $\downarrow$ in θ we can solve for θ_R:

$$g(\theta_R;\hat{\theta}(\underline{x})) = P_{\theta_R}(\hat{\Theta} \leq \hat{\theta}(\underline{x})) = \alpha/2 .$$

<u>Claim.</u> (θ_L, θ_R) is a $100(1-\alpha)\%$ CI for θ.

$\Rightarrow$ Given a sample $x_1,\ldots,x_n$ from the $N(\theta,1)$ dist.
$\left(\bar{x} - \frac{z_{1-\alpha/2}}{\sqrt{n}}, \bar{x} + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)$ is a $100(1-\alpha)\%$ CI for θ.

In general, to find θ_R we need to solve the equation

$$P_{\theta_R}(\hat{\Theta} \leq \hat{\theta}(\underline{x})) = \alpha/2 .$$

That is, find θ_R s.t. if $X_i \sim F_{\theta_R}$ then the observed $\hat{\theta}(\underline{x})$ is the $\alpha/2$ quantile of $\hat{\Theta} = \hat{\theta}(\underline{X})$.

In our $N(\theta,1)$ example we were able to do this based on the observation that in this case

$$P_\theta(\hat{\Theta} \leq \hat{\theta}) = P_\theta(\underbrace{\hat{\Theta} - \theta}_{\text{a RV whose dist. is invariant of } \theta\text{; } N(0,1/n)} \leq \hat{\theta} - \theta)$$

Indeed, in such cases, where the dist. of $\hat{\Theta} - \theta$ is invariant of θ and free of unknown parameters (typically when θ is a location parameter) the problem simplifies as follows.

The equation

$$\alpha/2 = g(\theta_R; \hat{\theta} = \hat{\theta}(\underset{\sim}{x})) = P_{\theta_R}(\underbrace{\hat{\Theta} - \theta_R}_{Y} \le \underbrace{\hat{\theta} - \theta_R}_{y})$$

implies $\hat{\theta} - \theta_R$ $(=y)$ is the $\alpha/2$ quantile of the dist. of $\hat{\Theta} - \theta_R$ (Y), where $X_i \sim F_{\theta_R}$.

But by the assumed invariance, $\hat{\theta} - \theta_R$ is the $\alpha/2$ quantile of the "universal" dist. of $\hat{\Theta} - \theta$ where $X_i \sim F_\theta$ for <u>any</u> θ.

Let F be this universal dist. ($N(0, 1/n)$ in our example)

By universal here we mean free of unknown parameters.

Assuming, as in our $N(\theta,1)$ example, F^{-1} exists we have

$$\hat{\theta} - \theta_R = F^{-1}(\alpha/2)$$

$$\Rightarrow \quad \theta_R = \hat{\theta} - F^{-1}(\alpha/2)$$

If this dist. of $\hat{\Theta} - \theta$ is also symmetric about 0 then

$$\theta_R = \hat{\theta} + F^{-1}(1 - \alpha/2).$$

In our $N(\theta,1)$ example $\hat{\Theta} - \theta = \bar{X} - \theta \sim N(0, 1/n)$ so

$$(\hat{\Theta} - \theta) = 1/\sqrt{n} \cdot Z \quad \text{where } Z \sim N(0,1)$$

(q-q plots)

$$\Rightarrow \quad F^{-1}(\alpha/2) = 1/\sqrt{n}\, \Phi^{-1}(\alpha/2) = z_{\alpha/2}/\sqrt{n}$$

$$\Rightarrow \quad \theta_R = \hat{\theta} - z_{\alpha/2}/\sqrt{n}$$

$$= \hat{\theta} + z_{1-\alpha/2}/\sqrt{n}$$

The same kind of procedure applies to θ^L, which is computed by solving

$$\alpha/2 = h(\theta_L; \hat{\theta}(\underline{x})) = P_{\theta_L}(\hat{\Theta} \geq \hat{\theta}(\underline{x})) = P_{\theta_L}(\hat{\Theta} - \theta_L \geq \hat{\theta} - \theta_L).$$

Again, assuming the dist. of $\hat{\Theta} - \theta$ is, as in our example, universal with an invertible cont. CDF F:

$$\alpha/2 = 1 - F(\hat{\theta} - \theta_L)$$

$$\Rightarrow \theta_L = \hat{\theta} - F^{-1}(1 - \alpha/2).$$

Again, in our $N(\theta, 1)$ example we have

$$F^{-1}(1-\alpha/2) = 1/\sqrt{n} \cdot \Phi^{-1}(1 - \alpha/2) = z_{1-\alpha/2}/\sqrt{n}$$

$$\Rightarrow \theta_L = \hat{\theta} - z_{1-\alpha/2}/\sqrt{n}.$$

The same idea works even when the $\alpha/2$ quantile of $F = F_{\hat{\Theta}-\theta}$ cannot be found explicitly.

For example, we can use parametric bootstrap to construct approximate CIs. Specifically, the idea is to approximate $F^{-1}(\alpha/2)$ and $F^{-1}(1-\alpha/2)$.

We do this by drawing bootstrap samples $\underline{x}^{*1}, \ldots, \underline{x}^{*N}$, where each $\underline{x}^{*i} = (x_1^{*i}, \ldots, x_n^{*i})$ is a sample of size n from F_θ, with $\theta = \hat{\theta}(\underline{x})$ ($N(\bar{x}, 1)$ in our example).

We then compute $\hat{\theta}^{*i} = \hat{\theta}(\underline{X}^{*i}) \qquad i = 1, \ldots, N$

$= \bar{X}^{*i} = \frac{1}{n}\sum_{j=1}^{n} X_j^{*i}$ in $N(\theta,1)$

Then, examining the empirical dist. of $\hat{\theta}^{*i} - \hat{\theta}(\underline{X})$ we find its β quantile, q_β^* for $\beta = \alpha/2$ and $\beta = 1-\alpha/2$.

Finally, we approximate θ_R and θ_L by

$$\tilde{\theta}_R = \hat{\theta}(\underline{X}) - q^*_{\alpha/2} = \bar{X} - q^*_{\alpha/2} \text{ in our case}$$

$$\tilde{\theta}^L = \hat{\theta}(\underline{X}) - q^*_{1-\alpha/2},$$

$(\tilde{\theta}_L, \tilde{\theta}_R)$ is an <u>approximate</u> $100(1-\alpha)\%$ CI for θ.

In parametric bootstrap we sample using $\theta = \hat{\theta}(\underline{X})$.
If $F_{\hat{\theta}-\theta}$ is indeed invariant of θ it shouldn't matter which θ we use in sampling.
If the dist. of $\hat{\theta}-\theta$ varies with θ then there is a question about the validity of this procedure, though if it only "slowly varies" with θ then this approach is reasonable (try multiple values of θ).

More examples.

2) $X_i \sim N(\theta, \sigma^2)$ where $\sigma^2 > 0$ is known.

$\hat{\theta} = \bar{x}$ as before and now $\hat{\Theta} = \bar{X} \sim N(\theta, \sigma^2/n)$.

Solve: $\quad \alpha/2 = h(\theta_L, \hat{\theta}) = P_{\theta_L}(\hat{\Theta} \geq \hat{\theta})$

Again, the dist. of $\hat{\Theta} - \theta \sim N(0, \sigma^2/n)$ is universal (free of unknown parameters), so

$$\theta_L = \hat{\theta} - F^{-1}(1 - \alpha/2).$$

Since $\hat{\Theta} - \theta = \sigma/\sqrt{n} \cdot Z$ where $Z \sim N(0,1)$,

$$F^{-1}(1 - \alpha/2) = \sigma/\sqrt{n}\, \Phi^{-1}(1 - \alpha/2) = z_{1-\alpha/2} \cdot \sigma/\sqrt{n}.$$

$$\Rightarrow \theta_L = \hat{\theta} - z_{1-\alpha/2} \cdot \sigma/\sqrt{n}.$$

Similarly,

$$\theta_R = \hat{\theta} - z_{\alpha/2} \cdot \sigma/\sqrt{n} = \hat{\theta} + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}.$$

$\Rightarrow \left(\hat{\theta} - z_{1-\alpha/2} \cdot \sigma/\sqrt{n},\ \hat{\theta} + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\right)$ is a $100(1-\alpha)\%$ CI for θ.

How does this CI change with σ? With n?

3) $X_i \sim N(\mu, \theta^2)$ where $\mu \in \mathbb{R}$ is unknown but we are only interested in estimating $\theta = \sigma_{X_i} = \sigma$. (scale)

Our estimator is

$$\hat{\Theta} = \sqrt{\frac{1}{n-1} S^2} = \sqrt{\frac{1}{n-1} \underbrace{\sum_1^n (x_i - \bar{x})^2}_{S^2}} .$$

Need to solve

$$\alpha/2 = g(\theta_R; \hat{\theta}) = P_{\theta_R}(\hat{\Theta} \le \hat{\theta}),$$

only now the dist. of $\hat{\Theta} - \theta$ is not invariant of θ.

However, the dist. of $\hat{\Theta}/\theta$ is invariant of θ (θ is the scale):

Let $Z_i = \frac{X_i - \mu}{\sigma}$, then Z_i are iid $N(0,1)$ and

$$S^2/\sigma^2 = \sum_1^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

$$= \sum_1^n \left(\frac{X_i - \mu}{\sigma} + \frac{\mu - \bar{X}}{\sigma}\right)^2$$

$$= \sum_1^n (Z_i - \bar{Z})^2$$

Clearly, the dist. of the RHS in invariant of μ and σ^2; it is in fact χ^2_{n-1}.

It follows that the dist. of

$$\hat{\Theta}/\theta = \sqrt{S^2/(n-1)} \, \frac{1}{\sigma} = \frac{1}{\sqrt{n-1}} \sqrt{S^2/\sigma^2} \sim \frac{1}{\sqrt{n-1}} \sqrt{\chi^2_{n-1}}$$

is invariant of both unknown parameters.

More generally, assume that the dist. of $\hat{\Theta}/\theta$ is universal (this will typically be the case when θ is a scale parameter), and let F be its CDF.

Then if F is invertible then we can readily solve

$$\alpha/2 = P_{\theta_R}(\hat{\Theta} \le \hat{\theta}) = P_{\theta_R}\left(\hat{\Theta}/\theta_R \le \hat{\theta}/\theta_R\right)$$
$$= F(\hat{\theta}/\theta_R)$$

$$\Rightarrow \quad \theta_R = \hat{\theta}/F^{-1}(\alpha/2)$$

Similarly,

$$\alpha/2 = P_{\theta_L}(\hat{\Theta} \ge \theta) = P_{\theta_L}\left(\hat{\Theta}/\theta_L \ge \hat{\theta}/\theta_L\right)$$
$$= 1 - F(\hat{\theta}/\theta_L) \quad \text{(if } F \text{ is cont.)}$$

$$\Rightarrow \quad \theta_L = \hat{\theta}/F^{-1}(1-\alpha/2)$$

Going back to our $N(\mu,\sigma^2)$ example we saw that

$$\hat{\Theta}/\theta \sim \frac{1}{\sqrt{n-1}}\sqrt{\chi^2_{n-1}}$$

It follows that

$$F^{-1}(p) = \frac{1}{\sqrt{n-1}}\sqrt{F^{-1}_{\chi^2_{n-1}}(p)}$$

Thus, if we denote $F^{-1}_{\chi^2_{n-1}}(p) = \chi^{n-1}_p$ then

$$\theta_R = \hat{\theta} \Big/ \sqrt{\frac{\chi^{n-1}_{\alpha/2}}{n-1}}$$

$$= \frac{\sqrt{S^2/(n-1)}}{\sqrt{\frac{\chi^{n-1}_{\alpha/2}}{n-1}}} = \sqrt{S^2/\chi^{n-1}_{\alpha/2}}$$

$$\theta_L = \hat{\theta} \Big/ \sqrt{\frac{\chi^{n-1}_{1-\alpha/2}}{n-1}} = \sqrt{S^2/\chi^{n-1}_{1-\alpha/2}}$$

$\Rightarrow \left(\sqrt{S^2/\chi^{n-1}_{1-\alpha/2}}, \sqrt{S^2/\chi^{n-1}_{\alpha/2}}\right)$ is a $100(1-\alpha)\%$ CI for σ.

ex. $\left(S^2/\chi^{n-1}_{1-\alpha/2}, S^2/\chi^{n-1}_{\alpha/2}\right)$ is a $100(1-\alpha)\%$ CI for σ^2.

These CIs are sensitive to deviations from the normal dist.

Exercise: What if μ is known? Hint: $\hat{\theta} = \frac{1}{n}\sum_1^n (X_i - \mu)^2$

Again, we can use parametric bootstrap when estimating such a scale parameter θ and the dist. of $\hat{\Theta}/\theta$ is invariant of θ (or slowly varying with θ):

Generate bootstrap samples $\underline{x}^{*1}, \ldots, \underline{x}^{*N}$, where each $\underline{x}^{*i} = (x_1^{*i}, \ldots, x_n^{*i})$ is a sample of size n from F_θ, with $\theta = \hat{\Theta}(\underline{x})$ ($N(0, \hat{\sigma}^2)$ or $N(\hat{\mu}, \hat{\sigma}^2)$ in our example).

We then compute $\hat{\Theta}^{*i} = \hat{\Theta}(\underline{x}^{*i}) \quad i = 1, \ldots, N$

$$= \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} \left(x_j^{*i} - \overline{x}^{*i}\right)^2} \text{ (in our example)}$$

Let q_β^* be the β quantile of the empirical dist. of $\hat{\Theta}^{*i}/\hat{\Theta}(\underline{x})$. Then

$$\left(\hat{\Theta}/q^*_{1-\alpha/2}, \; \hat{\Theta}/q^*_{\alpha/2}\right)$$

is an approximate $100(1-\alpha)\%$ CI for θ.