

CHAPTER 5

Confidence Regions

We quantify the difference between an estimator T and θ giving an interval estimate

$$[L(x) \quad R(x)]$$

Confidence interval

$$X \sim p_\theta(x) \quad \theta \in \mathbb{R}$$

a confidence interval for θ is
a map

$$x \mapsto [L(x), R(x)]$$

such that

$$P_\theta [L(x) \leq \theta \leq R(x)] \geq c$$

where c is said the confidence level
(usually 0.90, 0.95, 0.99)

-
- Thus a CI is a stochastic interval that has a high probability of containing θ .
 - The important point is that, given ANY θ , we have a fixed coverage probability of θ .
 - After the data are observed the CI is just a non-stochastic interval

Example Let $X \sim N(\theta, 1)$

then

$$[X - 1.96, X + 1.96]$$

is a CI for θ at level 0.95

$$\begin{aligned}
 \text{Proof: } & P_\theta (X - 1.96 \leq \theta \leq X + 1.96) = \\
 & = P_\theta (-1.96 \leq \theta - X \leq 1.96) = \\
 & = P_\theta (1.96 \geq X - \theta \geq -1.96) \\
 & = P_\theta (-1.96 \leq X - \theta \leq 1.96)
 \end{aligned}$$

$$\rightarrow P_{\theta} (-1.96 \leq Z \leq 1.96)$$

with $Z \sim N(0,1)$

$$= \Phi(1.96) - \Phi(-1.96) = 0.95$$

If we sample $X \sim N(\theta, 1)$
 we know the the interval
 $[X - 1.96, X + 1.96]$ contains θ
 95% of the time in the long run.

Suppose that we get $X = 10$.

Then the realized CI is

$$10 \pm 1.96 = \begin{cases} 11.96 \\ 8.04 \end{cases}$$

But we can't interpret as :
 we have the 95% chance of
 having $8.04 \leq \theta \leq 11.96$.

As θ is FIXED writing

$$P_{\theta} (8.04 \leq \theta \leq 11.96).$$

does not have any sense

Confidence region at level C for θ

is a stochastic subset G_x of Θ
such that

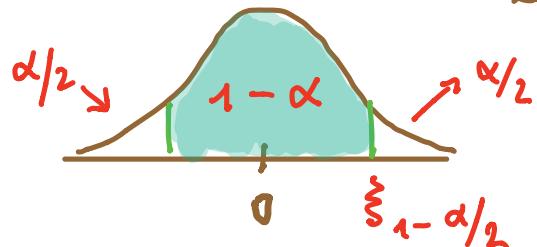
$$P_\theta(G_x \ni \theta) \geq C$$

for all $\theta \in \Theta$.

EXAMPLE 5.2 $(X_1, \dots, X_n) \stackrel{iid}{\sim} N(\mu, \sigma^2)$
with σ^2 known

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

let $C = 1 - \alpha$



$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$$



$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right) = 1 - \alpha$$

5.3 Pivots

$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is said to be a pivot.

because $P_\mu \left(a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b \right)$

is known and does not depend on μ .

Note that the pivot is a function of the data \underline{X} and of the unknown parameter, $\theta = \mu$.

The confidence interval is obtained by inverting the pivot

Example 5.4 $(X_1, \dots, X_n) \stackrel{iid}{\sim} N(\mu, \sigma^2)$
with σ^2 unknown.

Here $T = \frac{\bar{X} - \mu}{S_x/\sqrt{n}}$

is a pivot because

$$P_{\mu, \sigma} \left(a \leq \frac{\bar{X} - \mu}{S_x / \sqrt{n}} \leq b \right) = P(a \leq t_{m-1, 1-\alpha/2} \leq b)$$

is known and does not depend on the parameter $\theta = (\mu, \sigma^2)$.

Using this we find a confidence interval

$$P_{\mu, \sigma} \left(-t_{m-1, 1-\alpha/2} \leq \frac{\bar{X} - \mu}{S_x / \sqrt{n}} \leq t_{m-1, 1-\alpha/2} \right) = 1 - \alpha$$

So we get the CI :

$$\left[\bar{X} - \frac{S_x}{\sqrt{n}} t_{m-1, 1-\alpha/2}, \bar{X} + \frac{S_x}{\sqrt{n}} t_{m-1, 1-\alpha/2} \right]$$

- the interval is wider than the interval

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \xi_{1-\alpha/2}$$

- the length of the CI is random.
- the difference between the intervals disappears for $n \rightarrow \infty$.

- Non symmetric CI can be constructed from

$$\left[\bar{X} - \frac{S_x}{\sqrt{n}} t_{m-1, 1-\alpha}; \bar{X} - \frac{S_x}{\sqrt{n}} t_{m-1, \beta} \right]$$



the shortest CI is obtained for $\beta = \gamma = \alpha/2$.

Example 5.5 CI for θ - in Uniform

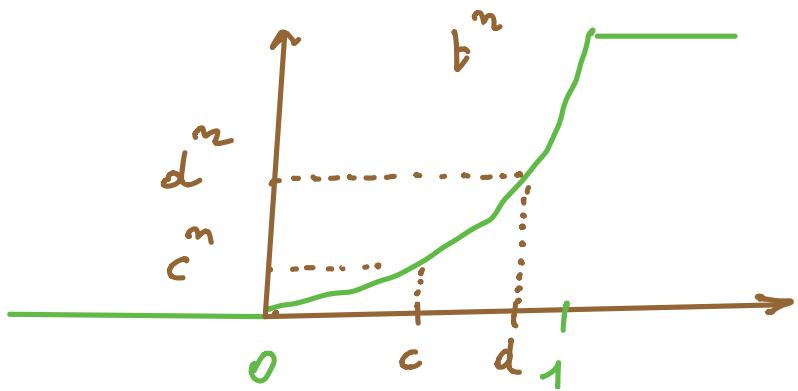
$$\underline{X} \stackrel{\text{iid}}{\sim} U(0, \theta) \Rightarrow \left(\frac{X_1}{\theta}, \dots, \frac{X_n}{\theta} \right) \stackrel{\text{iid}}{\sim} U(0, 1)$$

Every function of \uparrow is a pivot
So also

$$T = \frac{X_{(n)}}{\theta}$$

is a pivot, and we know

$$P_\theta \left(\frac{X_{(n)}}{\theta} \leq t \right) = t^n \quad 0 \leq t \leq 1$$



$$P_\theta \left(c \leq \frac{X_{(n)}}{\theta} \leq d \right) = 1 - \alpha$$

$$d^n - c^n = 1 - \alpha \quad \text{for instance}$$

$c = d^{1/n}$, $d = 1$ is a $(1-\alpha)$ -C.I.

//

Near-Pivots

Pivots do not exist always.

So we can use approximate pivots

Example 5.2 Binomial

$$X_n \sim \text{Bin}(n, p) \Rightarrow T = \frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{L} N(0, 1)$$

Thus T is a pivot asymptotically

$$P_p \left(-\xi_{1-\alpha/2} < \frac{X - np}{\sqrt{np(1-p)}} < \xi_{1-\alpha/2} \right) \approx 1-\alpha$$

Solve $(X - np)^2 \leq \xi_{1-\alpha/2}^2 [np(1-p)]$

and get L and R .

→ Notebook 5

We can also use the fact that by the LLN

$$\frac{X}{n} \xrightarrow{P} p \quad n \rightarrow \infty$$

and by Slutsky lemma

$$\frac{X - np}{\sqrt{n \frac{X}{n} \left(1 - \frac{X}{n}\right)}} \xrightarrow{L} N(0,1)$$

Recall from Probability ↴

Slutsky lemma 5.15

If S_n and T_n are sequences of r.v. such that for $n \rightarrow \infty$

$$S_n \xrightarrow{P} \sigma \quad \text{and} \quad T_n \xrightarrow{L} T$$

► $S_n + T_n \xrightarrow{L} \sigma + T$

► $T_n / S_n \xrightarrow{L} T/\sigma \quad \text{if } \sigma \neq 0$

So we can use this second approximated point to get the CI

$$\frac{x}{n} \pm \xi_{1-\alpha/2} \frac{1}{\sqrt{n}} \sqrt{\frac{x}{n} \left(1 - \frac{x}{n}\right)}$$

Example in a poll suppose that

$X = \# \text{ successes for a candidate}$
with a sample of size n .

Let $x = 90$ $n = 200$

a 95% CI for p = Probability of success.

is

$$0.45 \pm 1.96 \sqrt{\frac{0.45 \times 0.55}{200}}$$
$$0.45 \pm 0.0689$$

$$45\% \pm 6.89\%$$

Note that the size of the population has no role in determining the length of the CI.

- Newspapers promise a 2% deviation from the truth
- With $n=1500$ and $\hat{p} = 1/2$ we get a deviation of about

$$1.96 \sqrt{\frac{0.5 \times 0.5}{1500}} = 0.025$$

i.e. 2.5%.

Estimated standard error

Note that the MLE estimate of p is $\hat{P}_{ML} = \frac{x}{n}$

The standard error is

$$\sigma\left(\frac{x}{n}\right) = \sqrt{\frac{p(1-p)}{n}}$$

The MLE of the standard error is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{1}{\sqrt{n}} \sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)}$$

Therefore the approximate CI is

$$\hat{P}_{ML} \pm \xi_{1-\alpha/2} \cdot \hat{se}_{ML}$$

5.4 MLE as Near-Pivots.

Under certain conditions the Maximum likelihood estimators are asymptotically normal

We start from the simplest case

$$X \sim p_{\theta}(x)$$
$$\theta \in \Theta \subset \mathbb{R}.$$

with a single parameter and a single observation.

We assume that the log-likelihood

$$\theta \mapsto l_{\theta}(x) = \log p_{\theta}(x)$$

is differentiable for every x and also the derivative is continuous.

Two definitions

► The score function of the model

is $\ell'_\theta(x) = \frac{d}{d\theta} \ell_\theta(x)$

► The Fisher information for θ in X is the number

$$i_\theta = \text{var}_\theta \ell'_\theta(x)$$

Example $X \sim \text{Bin}(n, \theta)$

$$\begin{aligned} \ell_\theta(x) &= \log \left[\binom{n}{x} \theta^x (1-\theta)^{n-x} \right] \\ &= \text{const.} + x \log \theta + (n-x) \log(1-\theta) \end{aligned}$$

$$\ell'_\theta(x) = \frac{x}{\theta} - \frac{n-x}{1-\theta} \quad (0 < \theta < 1)$$

$$\ell'_\theta(x) = \frac{x}{\theta} - \frac{n-x}{1-\theta} \quad X \sim \text{Bin}(n, \theta)$$

Now
1) $E_\theta(\ell'_\theta(x)) = E_\theta\left(\frac{x}{\theta} - \frac{n-x}{1-\theta}\right)$

$$= \frac{m\theta}{\theta} - \frac{m-m\theta}{1-\theta}$$

$$= m - m = \boxed{0}.$$

Thus, the expected value of the score is 0.

$$\begin{aligned} 2) \text{Var}_\theta(\ell'_\theta(X)) &= E_\theta \left[\left(\frac{x}{\theta} - \frac{m-x}{1-\theta} \right)^2 \right] \\ &= E_\theta \left(\frac{x-m\theta}{\theta(1-\theta)} \right)^2 = \frac{1}{\theta^2(1-\theta)^2} E_\theta[(x-m\theta)^2] \\ &= \frac{m\theta(1-\theta)}{\theta^2(1-\theta)^2} = \boxed{\frac{m}{\theta(1-\theta)}} \end{aligned}$$

Therefore the Fisher information for θ in X is $\frac{m}{\theta(1-\theta)}$
and is proportional to m .

Now if we have a sample

$$(X_1, \dots, X_n) \stackrel{iid}{\sim} p_\theta(x)$$

► The log-likelihood is $\sum_{i=1}^n \ell_\theta(x_i)$

► The score is $\sum_{i=1}^n \ell'_\theta(x_i)$

► The Fisher information is

$$\text{var}_\theta \left(\sum_{i=1}^n \ell'_\theta(x_i) \right) = n i_\theta$$

► The MLE is the solution of the likelihood equation

$$\sum_{i=1}^n \ell'_\theta(x_i) = 0$$

unless the likelihood takes its maximum on the boundary

► We assume that the model is identifiable :

if $\theta \neq \theta'$ then $p_\theta(x) \neq p_{\theta'}(x)$

with positive probability.

► Assume that Θ is compact and convex

Theorem 5.9 -

Under certain regularity conditions listed below if $\hat{\theta}_n$ is the MLE of θ and θ is an interior point of Θ ,

$$\text{under } \mathcal{S}, \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(\theta, i_{\theta}^{-1})$$

The regularity conditions define a large subclass of models.

1) $|l'_{\theta}(x)| \leq L(x) \quad \forall \theta$

where $L(\cdot)$ is a function such that $E_{\theta}(L^2(X_i)) < \infty$

2) The function $\theta \mapsto i_{\theta}$ is continuous and positive.

If n is large, The theorem implies that under \mathcal{S}

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, i_{\theta}^{-1})$$

$$rmle(\theta - \bar{\theta}) \approx N(0, 1)$$

is a near pivot.

We can deduce an approximate CI of level $(1-\alpha)$ in this way :

The standard error of $\hat{\theta}_n$ is

$$se = \frac{1}{\sqrt{n i_0'}}$$

The estimated standard error is

$$\hat{se} = \frac{1}{\sqrt{n \hat{i}_0'}}$$

where \hat{i}_0' is an estimator of i_0' .

And the CI is

$$\hat{\theta} \pm z_{1-\alpha/2} \frac{1}{\sqrt{n \hat{i}_0'}}$$

Estimators of i_θ

► The plug-in estimator is

► The observed information

$$\widehat{i}_\theta = - \frac{1}{n} \ell_{\widehat{\theta}}''(x_i)$$

where $\ell_{\theta}''(x) = \frac{d^2}{d\theta^2} \log p_\theta(x)$.

NOTE

- the observed information is a random variable.
- the Fisher information is a number

→ Notebook 5

Example - Binomial $X \sim \text{Bin}(n, \theta)$

$$i_\theta = \frac{n}{\theta(1-\theta)}$$

Plug-in estimator. With $\hat{\theta} = \frac{x}{n}$,

$$i_{\hat{\theta}} = \frac{n}{\frac{x}{n} \left(1 - \frac{x}{n}\right)} = \boxed{\frac{n^3}{x(n-x)}}$$

Observed information

$$\begin{aligned} \ell''_\theta(x) &= \frac{d}{d\theta} \left(\frac{x}{\theta} - \frac{n-x}{1-\theta} \right) \\ &= -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \end{aligned}$$

$$- \ell''_{\hat{\theta}}(x) = \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}$$

Then substitute $\hat{\theta}$ to θ

$$\hat{i}_\theta = \frac{x}{(x/n)^2} + \frac{n-x}{(1-x/n)^2} = \boxed{\frac{n^3}{x(n-x)}} .$$

CURVATURE

The observed information gives the curvature of the log-likelihood

Why? Use a Taylor approximation

1) Use the normalized log-lik :

$$\ell(\theta) - \ell(\hat{\theta}) = \log \frac{L(\theta)}{L(\hat{\theta})}$$

2) Expand $\ell(\theta)$ at the 2nd order

$$\ell(\theta) \simeq \ell(\hat{\theta}) + \ell'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2} \ell''(\hat{\theta})(\theta - \hat{\theta})^2$$

Quadratic approximation

$$\ell(\theta) - \ell(\hat{\theta}) \simeq -\frac{1}{2} \widehat{\chi}_{\theta} (\theta - \hat{\theta})^2$$