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

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

a confidence interval for θ is a map

$$X \longmapsto [L(X), R(X)]$$

such that

$$P_\theta\left(L(X) \leq \theta \leq R(X) \right) \geq C$$

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

Proof: $P_\theta(X - 1.96 \leq \theta \leq X + 1.96) =$

$= P_\theta(-1.96 \leq \theta - X \leq 1.96) =$

$= P_\theta(1.96 \geq X - \theta \geq -1.96)$

$= P_\theta(-1.96 \leq X - \theta \leq 1.96)$

$$= P_\theta(-1.96 \le Z \le 1.96)$$

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 $8.04 \le \theta \le 11.96$.

As θ is FIXED writing
$$P_\theta(8.04 \le \theta \le 11.96).$$
does not have any sense

Confidence <u>region</u> at level c for θ

is a stochastic subset G_X of Θ such that

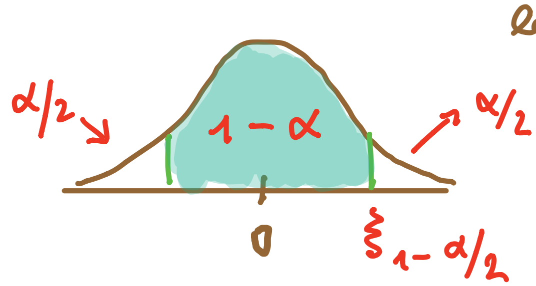
$$P_\theta (G_X \ni \theta) \geq c$$

for all $\theta \in \Theta$.

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

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

let $c = 1-\alpha$



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

$$\Updownarrow$$

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

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

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

is a pivot because

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

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

Using this we find a confidence interval

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

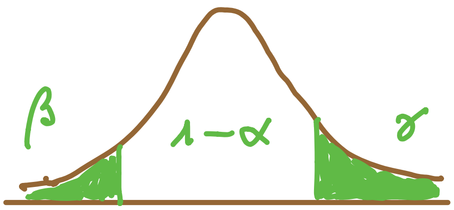
So we get the CI:

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

- the interval is wider than the interval
$$\bar{X} \pm \frac{\sigma}{\sqrt{n}}\, \xi_{1-\alpha/2}$$
- the lenght of the CI is random.
- the difference between the intervals disappears for $n \to \infty$.

- Non symmetric CI can be constructed from

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



with $\beta + \gamma = \alpha$.

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

Example 5.5 CI for θ - in Uniform

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

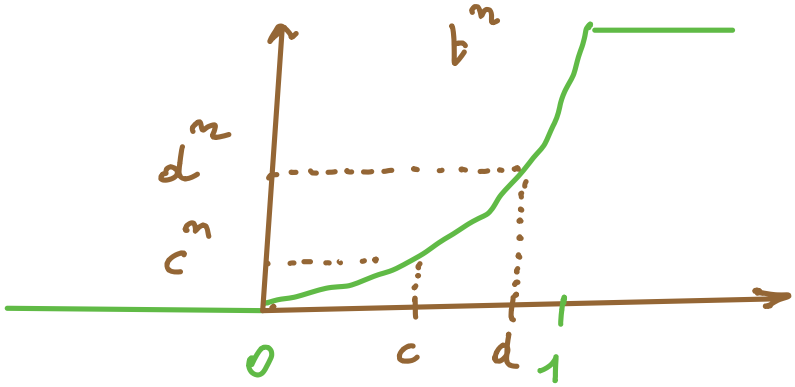
Every function of $\left(\frac{X_1}{\theta}, \dots, \frac{X_n}{\theta}\right)$ is a pivot

So also

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

is a pivot, and we know

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



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

$d^n - c^n = 1-\alpha$ for instance.

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

Near-Pivots

Pivots do not exist always.
So we can use approximate pivots

Example 5.2 Binomial

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

Thus T is a pivot asymptotically

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

Solve $\quad (X - np)^2 \leq \xi^2_{1-\alpha/2}\,[np(1-p)]$

and get L and R.

$\rightarrow$ Notebook 5

We can also use the fact that by the LLN

$$\frac{X}{n} \xrightarrow{P} p \qquad n \to \infty$$

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 $n \to \infty$

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

- $S_n + T_n \xrightarrow{L} \sigma + T$
- $T_s / S_n \xrightarrow{L} T/\sigma$ if $\sigma \neq 0$

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

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

Example in a poll suppose that X = # successes for a candidate with a sample of size n.

Let $X = 90$ $\qquad n = 200$

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

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

5.4 MLE as Near-Pivots.

Under certain conditions the Maximum likelihood estimators are <u>asymptotically normal</u>

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

$$\theta \longmapsto \ell_\theta(x) = \log p_\theta(x)$$

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
$$i_\theta = \operatorname{var}_\theta \ell'_\theta(X)$$

Example $\quad X \sim Bin(n,\theta)$

$$\ell_\theta(x) = \log\left[\binom{n}{x}\theta^x(1-\theta)^{n-x}\right]$$

$$= \text{const.} + x\log\theta + (n-x)\log(1-\theta)$$

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

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

Now

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

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

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

Thus, the expected value of the score is 0.

2) $$\mathrm{var}_\theta(\ell'_\theta(X)) = E_\theta\left[\left(\frac{X}{\theta} - \frac{n-X}{1-\theta}\right)^2\right]$$

$$= E_\theta\left(\frac{X - n\theta}{\theta(1-\theta)}\right)^2 = \frac{1}{\theta^2(1-\theta)^2} E_\theta\left[(X - n\theta)^2\right]$$

$$= \frac{n\theta(1-\theta)}{\theta^2(1-\theta)^2} = \boxed{\frac{n}{\theta(1-\theta)}}$$

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

Now if we have a sample

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

- 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

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

- 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,

under θ, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(\theta, i_\theta^{-1})$

The regularity conditions define a large subclass of models.

1) $|\ell_\theta'(x)| \leq L(x) \quad \forall \theta$

where $L(\cdot)$ is a function such that $E_\theta(L^2(X_1)) < \infty$

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

If n is large, The theorem implies that under θ

$$\sqrt{n i_\theta}(\hat{\theta} - \theta) \approx N(0,1)$$

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

$$se = \frac{1}{\sqrt{n\, i_\theta}}$$

The estimated standard error is

$$\hat{se} = \frac{1}{\sqrt{n\, \hat{i}_\theta}}$$

where $\hat{i}_\theta$ is an estimator of i_θ.

And the CI is

$$\boxed{\hat{\theta} \pm \xi_{1-\alpha/2} \frac{1}{\sqrt{n\, \hat{i}_\theta}}}$$

$\boxed{\text{Estimators of } i_\theta}$

- The plug-in estimator $i_{\hat\theta}$
- The observed information

$$\widehat{i_\theta} = -\frac{1}{n} \ell''_{\hat\theta}(x_i)$$

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

$\boxed{\text{NOTE}}$
- the observed information is a random variable.
- the Fisher information is a number

→ Notebook 5

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

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

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

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

$$= -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$$

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

Then substitute $\hat{\theta}$ to θ

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

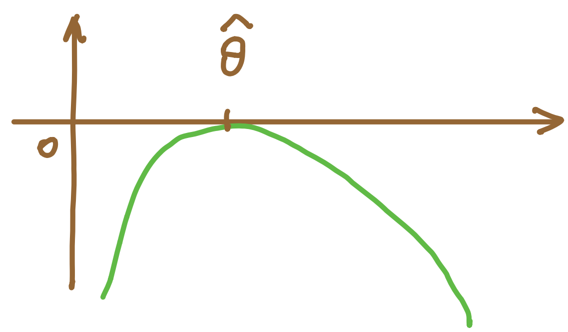
CURVATURE

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

$$\boxed{\ell(\theta) - \ell(\hat{\theta}) \simeq -\frac{1}{2}\,\hat{\imath}_\theta\,(\theta - \hat{\theta})^2}$$