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 region at level c for θ

is a stochastic subset G_X of Θ such that

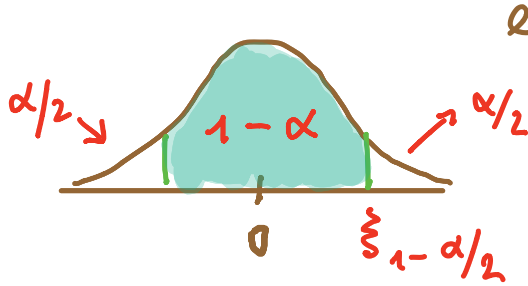
$$P_\theta\left(G_X \ni \theta \right) \geq c$$

for all $\theta \in \Theta$.

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

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

let $c = 1-\alpha$



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

$$\Updownarrow$$

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

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

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

is a pivot because

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

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

Using this we find a confidence interval

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

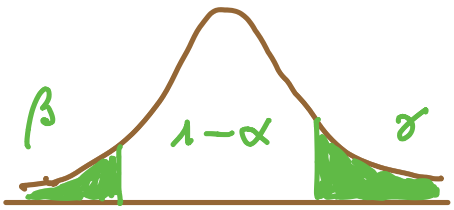
So we get the CI:

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

- 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 $\beta = \gamma = \alpha/2$.

<u>Example 5.5</u> CI for θ - in Uniform

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

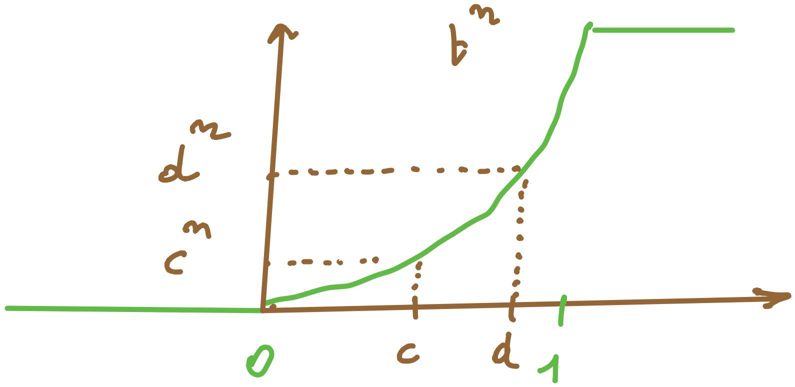
Every function of ⟶ is a pivot

So also

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

is a pivot, and we know

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



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

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

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

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

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

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

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

<u>Example</u> 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 $X \sim Bin(n, \theta)$

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

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

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

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

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

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

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

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

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

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

The <u>regularity conditions</u> 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 $\vartheta \mapsto i_\vartheta$ 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 <u>plug-in estimator</u> $i_{\hat{\theta}}$
- The <u>observed information</u>

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

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

→ <u>Notebook 5</u>

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

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

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

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

Then substitute $\hat\theta$ to θ

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

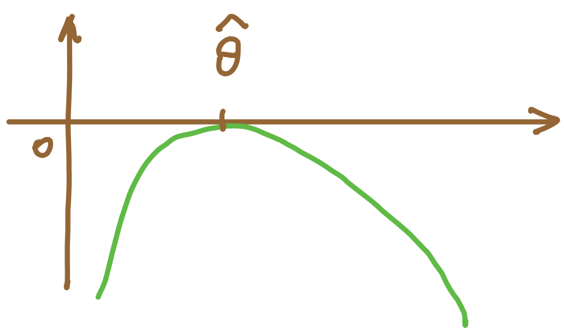
Interpretation of the information

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

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

$\longrightarrow$ Notebook 5

LEMMA 5.10

Under certain regularity conditions

- $E_\theta \, \ell_\theta'(X_1) = 0$
- $E_\theta \, \ell_\theta''(X_1) = -i_\theta$

Proof Differentiate the identity.

$$\int p_\theta(x)\,dx = 1 \qquad \forall \theta$$

(for discrete distribution, replace $\int$ by Σ)

$$\frac{d}{d\theta}\int p_\theta(x)\,dx = 0$$

$$\Updownarrow \quad \text{under certain regularity conditions}$$

$$\int p_\theta'(x)\,dx = 0$$

$$\Updownarrow \quad \text{as } \ell_\theta'(x) = \frac{p_\theta'(x)}{p_\theta(x)}$$

$$\int \ell_\theta'(x)\,p_\theta(x)\,dx = 0$$

$$\Updownarrow$$

$$\boxed{E_\theta \, \ell_\theta'(X_1) = 0}$$

Differentiate twice the identity

$$\frac{d^2}{d\theta^2} \int p_\theta(x)\, dx = 0$$

$$\Updownarrow$$

$$\int p''_\theta(x)\, dx = 0$$

As $$\ell''_\theta(x) = \frac{d}{d\theta}\left[\frac{p'_\theta(x)}{p_\theta(x)}\right]$$

$$= \frac{p''_\theta(x)\, p_\theta(x) - p'_\theta(x)^2}{p_\theta(x)^2}$$

$$= \frac{p''_\theta(x)}{p_\theta(x)} - \left(\frac{p'_\theta(x)}{p_\theta(x)}\right)^2$$

$$\ell''_\theta(x) = \frac{p''_\theta(x)}{p_\theta(x)} - \ell'_\theta(x)^2$$

we have also

$$\ell''_\theta(x)\, p_\theta(x) = p''_\theta(x) - \ell'_\theta(x)^2\, p_\theta(x)$$

$$\int p_\theta''(x)\,dx = 0$$

$$\Updownarrow$$

$$\int \ell_\theta''(x)\, p_\theta(x)\,dx + \int \ell_\theta'(x)^2\, p_\theta(x)\,dx = 0$$

$$\Updownarrow$$

$$E_\theta\, \ell_\theta''(X_1) + E_\theta\left[\ell_\theta'(X_1)^2\right] = 0$$

$$\Updownarrow$$

$$\boxed{E_\theta\, \ell_\theta''(X_1) = -\mathrm{var}_\theta\left(\ell_\theta'(X_1)\right) = -i_\theta}$$

//

Example 5.11 Poisson

$$\ell_\theta'(x) = \frac{d}{d\theta} \log\left(\frac{e^{-\theta}\,\theta^x}{x!}\right) = \frac{x}{\theta} - 1$$

$$i_\theta = \mathrm{var}_\theta\left(\frac{X}{\theta} - 1\right) = \frac{1}{\theta^2}\,\theta = \frac{1}{\theta}.$$

$$i_\theta = E_\theta\left(-\ell_\theta''(X_1)\right) = -E_\theta\left(-\frac{X_1}{\theta^2}\right) = \frac{1}{\theta}$$

Estimate i_θ with

- $i_{\hat\theta} = \frac{1}{\hat\theta} = \frac{1}{\bar X}$, or

- $\hat{i}_\theta = \frac{-1}{n} \sum_{i=1}^{n} \ell''_{\hat\theta}(X_i) = \frac{1}{n} \sum \frac{X_i}{\hat\theta^2}$

$$= \frac{1}{n \bar X^2} \cdot n \bar X = \frac{1}{\bar X}.$$

Confidence interval:

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

Example 5.13 Exponential

$$\ell'_\lambda(x) = \frac{d}{d\theta} \log(\lambda e^{-\lambda x}) = \frac{1}{\lambda} - x \qquad \hat\lambda = \frac{1}{\bar x}$$

$$i_\lambda = \mathrm{var}_\lambda\left(\frac{1}{\lambda} - X_1\right) = \mathrm{var}_\lambda(X_1) = \frac{1}{\lambda^2}.$$

$$i_{\hat\lambda} = \frac{1}{(1/\bar x^2)} = \bar x^2. \quad = \hat{i}_{\hat\lambda}$$

Confidence interval

$$\frac{1}{\bar{X}} \pm z_{1-\alpha/2} \frac{1}{\sqrt{n}\,\bar{X}}$$

Proof of Theorem 5.9

Notation: θ = true parameter
ϑ = generic parameter different from θ.

$$M_n(\vartheta) = \frac{1}{n} \sum_{i=1}^{n} \ell_\vartheta(X_i)$$

The proof is structured as follows

a) Proof that the MLE $\hat{\theta}_n$

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{for } n \to \infty$$

i.e. that the MLE is consistent.

b) Proof that $\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{L} \frac{1}{i_\theta} N(0, i_\theta)$

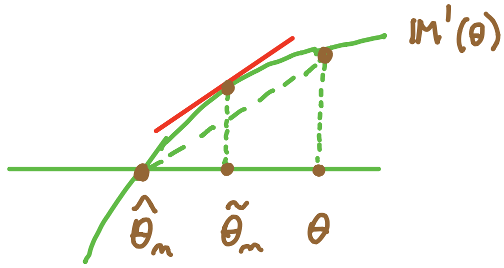
We skip the proof of a)

Proof of b)

$\hat{\theta}_n$ satisfies the likelihood eqns. that can be written as

$$\mathbb{M}_n'(\hat{\theta}_n) = 0$$

By the mean value theorem there is a point $\tilde{\theta}_n$ between θ and $\hat{\theta}_n$ s.t.



$$\frac{\mathbb{M}_n'(\theta) - \mathbb{M}_n'(\hat{\theta}_n)}{\theta - \hat{\theta}_n} = \mathbb{M}_n''(\tilde{\theta}_n)$$

Thus, we can write

$$\sqrt{n}\,(\hat{\theta}_n - \theta) = -\sqrt{n}\,\frac{\mathbb{M}_n'(\theta)}{\mathbb{M}_n''(\tilde{\theta}_n)}$$

$$= - \frac{\sqrt{n}\ \frac{1}{n}\sum_1^n \ell'_\theta(X_i)}{\frac{1}{n}\sum_1^n \ell''_{\tilde{\theta}_n}(X_i)} = - \frac{\sqrt{n}\ \frac{1}{n}\sum_1^n Y_i}{\frac{1}{n}\sum_1^n U_i}$$

Apply the central limit theorem to the numerator.

Let $Y_i = \ell'_\theta(X_i)$ $\quad E_\theta(Y_i) = 0$

$Y_i \sim$ iid $\quad \mathrm{var}_\theta(Y_i) = i_\theta$.

The numerator is

$$\sqrt{n}\left(\frac{1}{n}\sum Y_i - E(Y_1)\right) \xrightarrow{L} N(0, i_\theta)$$

The denominator is

$$\frac{1}{n}\sum_1^n \ell''_{\tilde{\theta}_n}(X_i) = \frac{1}{n}\sum_1^n U_i \qquad (*)$$

- $U_1, \ldots, U_n$ are dependent and the LLN cannot be applied as such.

- But $\tilde{\theta}_n \xrightarrow{P} \theta$ and it can be proved that $\frac{1}{n}\sum_1^n U_i$ behaves like $\frac{1}{n}\sum_1^n \ell''_\theta(X_i)$ that satisfies the LLN and

$$\boxed{\frac{1}{n}\sum_1^n \ell''_\theta(x_i) \xrightarrow{P} E_\theta \ell''_\theta(x_1) = -i_\theta}$$

given this we can conclude that

$$\frac{1}{n}\sum_1^n U_i \xrightarrow{P} -i_\theta .$$

and by Slutsky theorem

$$\sqrt{n}\,(\hat{\theta}_n - \theta) = \frac{A_n}{B_n}$$

where $A_n \xrightarrow{L} N(0, i_\theta)$

and $B_n \xrightarrow{P} -(-i_\theta) = i_\theta$

Thus

$$\boxed{\frac{A_n}{B_n} \xrightarrow{L} \frac{1}{i_\theta} \cdot N(0, i_\theta) = N(0, i_\theta^{-1})}$$

The only thing that has still to be proved is that:

$$\frac{1}{n}\sum_{1}^{n}\left(\ell''_{\tilde{\theta}_n}(X_i)-\ell''_{\theta}(X_i)\right)\xrightarrow{P}0$$

skip this

This requires a further assumption of the

- existence of $\ell'''_{\vartheta}(x)$ and
- $|\ell'''_{\vartheta}(x)|\leq K(x) \quad \vartheta\in$ (an interval around θ).

with $K(\cdot)$ satisfying $E_\theta K(X_1)<\infty$.

We apply the mean value theorem to the $\ell''_{\vartheta}(x)$.

$$\left|\frac{\ell''_{\vartheta}(x)-\ell''_{\theta}(x)}{\vartheta-\theta}\right|=\left|\ell'''_{\tilde{\vartheta}}(x)\right|\leq K(x)$$

Thus, for all $\delta>0$

$$P_\theta\left\{\left|\frac{1}{n}\sum_{1}^{n}\left[\ell''_{\tilde{\theta}_n}(X_i)-\ell''_{\theta}(X_i)\right]\right|>\delta,\right.$$

$$|\tilde{\theta}_n - \theta| \leq \varepsilon)$$

$$\leq P_\theta\left(\frac{1}{n}\sum_1^n K(X_i)\,|\tilde{\theta}_n - \theta| > \delta\right) \qquad (**)$$

By the LLN $\frac{1}{n}\sum_1^n K(X_i) \to E_\theta K(X_1) < \infty$

thus, there exist a constant M s.t.

$$P_\theta\left(\frac{1}{n}\sum_1^n K(X_i) < M\right) \to 1$$

From $(**)$ $\quad P_\theta\left(|\hat{\theta}_n - \theta| > \frac{\delta}{M}\right) \to 0$

and thus also

$$\frac{1}{n}\sum_1^n \ell''_{\tilde{\theta}_n}(X_i) \xrightarrow{P} \frac{1}{n}\sum_1^n \ell''_\theta(X_i) \quad //$$

5.4.2* Multidimensional parameters

Let $\underset{\sim}{\theta}$ $k \times 1$ $\quad k > 1$.

The score $= \dot{\ell}_\theta(x) = \nabla_\theta \, \ell_\theta(x)$

$$= \left(\frac{\partial}{\partial \theta_1} \ell_\theta(x), \ldots, \frac{\partial}{\partial \theta_k} \ell_\theta(x) \right)$$

The Fisher information

$$\underset{\sim}{i}_\theta = \left[i_\theta(r,t) \right] \in \mathbb{R}^{k \times k} \quad \text{where}$$

$$i_\theta(r,t) = \operatorname{cov}_\theta \left(\frac{\partial}{\partial \theta_r} \ell_\theta(X_1), \frac{\partial}{\partial \theta_t} \ell_\theta(X_1) \right)$$

Theorem 5.9 is still valid but

$$\sqrt{n} \left(\underline{\hat{\theta}}_n - \underline{\theta} \right) \xrightarrow{L} MN \left(\underline{0}, \underset{\sim}{i}_\theta^{-1} \right)$$

That is:

near pivot: $\left(n \, i_\theta \right)^{1/2} \left(\underline{\hat{\theta}}_n - \underline{\theta} \right) \xrightarrow{L} \underline{Z}$

where $\quad \underline{Z} = (Z_1, \ldots, Z_k) \sim \text{iid} \; N(0,1)$

The matrix $i_\theta^{1/2} = A$ is such that

$$A^T A = i_\theta .$$

Note that

$$(\hat{\underline{\theta}} - \underline{\theta})^T n\, i_\theta\, (\hat{\underline{\theta}} - \underline{\theta}) \xrightarrow{L} \underline{Z}^T \underline{Z} \sim \chi^2_k$$

With an estimator $\hat{i}_\theta$ of i_θ we define a confidence region of level $1-\alpha$ with

$$\boxed{\left\{ \theta : (\hat{\underline{\theta}} - \underline{\theta})^T n\, \hat{i}_\theta\, (\hat{\underline{\theta}} - \underline{\theta}) \leq \chi^2_{k, 1-\alpha} \right\}}$$

that is an ellipsoid. Because the Fisher information matrix i_θ is positive definite.

Example — trinomial distribution

Let $(Y_1, Y_2, Y_3)^T$ be counts obtained after sampling from a population with 3 categories.

Specifically, pneumoconiosis in a sample of miners (after 20 years of exposition) is found

Severe	Present	Absent	Tot
3	6	34	43

Let $\underset{\sim}{X}_i = (X_{i1}, X_{i2}, X_{i3})$ $i = 1, \dots, n$ be an indicator of the category.

$$\boxed{(X_{i1}, X_{i2}, X_{i3}) \sim \text{Mult}\left(1, \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}\right)}$$

We have a random sample $(\underline{X}_1, \underline{X}_2, \dots \underline{X}_n)$ from this distribution

The sum

$$\sum_{i=1}^{n} \underline{X}_i = \underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} .$$

are the observed counts of workers having severe, present, absent pneumoconiosis.

$$\boxed{\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \text{Mult}\left(n, \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}\right)}$$

We want a confidence region for (p_1, p_2, p_3).

The density is

$$P(\underline{x}_1, \dots, \underline{x}_n) = \prod_{i=1}^{n} \left(p_1^{x_{i1}} p_2^{x_{i2}} p_3^{x_{i3}}\right)$$

$$= p_1^{\sum x_{i1}} p_2^{\sum x_{i2}} p_3^{x_{i3}}$$

The log-likelihood is

$$\ell(\underline{p}) = \sum_{i=1}^{n} x_{i1} \log p_1 + \sum_{i=1}^{n} x_{i2} \log p_2 + \sum_{i=1}^{n} x_{i3} \log p_3$$

$$= Y_1 \log p_1 + Y_2 \log p_2 + Y_3 \log p_3$$

$$= Y_1 \log p_1 + Y_2 \log p_2 + Y_3 \log(1 - p_2 - p_3)$$

The score function for a single observation

$$\frac{\partial \ell}{\partial p_1} = \frac{X_{i,1}}{p_1} - \frac{X_{i,3}}{p_3}$$

$$\frac{\partial \ell}{\partial p_1} = \frac{X_{i,2}}{p_2} - \frac{X_{i,3}}{p_3}$$

Fisher information matrix for a single observation

$$i_p = \begin{bmatrix} \frac{1}{p_1} + \frac{1}{p_3} & \frac{1}{p_3} \\ \frac{1}{p_3} & \frac{1}{p_2} + \frac{1}{p_3} \end{bmatrix}$$

Proof First note that for the $\mathrm{Mult}\left(1, \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}\right)$ the covariance matrix is

$$\begin{pmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 \\ \cdot & p_2(1-p_2) & -p_2p_3 \\ \cdot & \cdot & p_3(1-p_3) \end{pmatrix}$$

Thus,

$$\mathrm{var}_p\left(\frac{X_{i,1}}{p_1} - \frac{X_{i,3}}{p_3}\right) = \frac{p_1(1-p_1)}{p_1^2} + \frac{p_3(1-p_3)}{p_3^2} + 2$$

$$= \frac{1}{p_1} + \frac{1}{p_3}.$$

$$\mathrm{cov}_p\left(\frac{X_{i1}}{p_1} - \frac{X_{i,3}}{p_3}, \frac{X_{i2}}{p_2} - \frac{X_{i3}}{p_3}\right) =$$

$$= \frac{1}{p_1p_2}\mathrm{cov}_p(X_{i1}, X_{i2}) - \frac{1}{p_1p_3}\mathrm{cov}_p(X_{i1}, X_{i3}) -$$

$$- \frac{1}{p_3p_2}\mathrm{cov}_p(X_{i3}, X_{i2}) + \frac{1}{p_3^2}\cdot \mathrm{var}_p(X_{i3})$$

$$= \frac{-p_1p_2}{p_1p_2} - \left(\frac{-p_1p_3}{p_1p_3}\right) - \left(\frac{-p_2p_3}{p_2p_3}\right) + \frac{1-p_3}{p_3} = 1 + \frac{1-p_3}{p_3}$$

$$= \frac{1}{p_3}. \qquad //$$

Inverse of the Fisher inf. matrix:

$$\underline{i}_p^{-1} = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 \\ & p_2(1-p_2) \end{bmatrix}$$

The estimator of i_p can be found using the ML Estimates

$$\hat{\underset{\sim}{p}} = \left(\frac{y_1}{n}, \frac{y_2}{n}\right)^T$$

The approximate elliptical confidence region is

$$\underline{p} = (p_1, p_2)^T : (\hat{\underline{p}} - \underline{p})\, n\, \underline{i}_p\, (\hat{\underline{p}} - \underline{p}) \leq \chi^2_2$$

→ Notebook. 5

$\boxed{\text{Delta method}}$

Sometimes we are interested in large sample confidence intervals for a function $g(\theta)$.

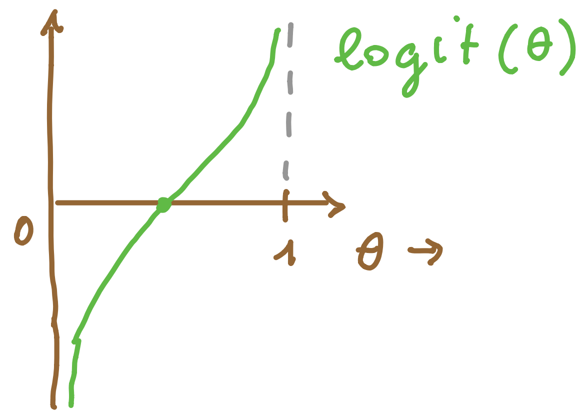
Example $X \sim Bin(n, \theta)$; $\hat{\theta} = X/n$.

$$i_\theta = n/\theta(1-\theta) \qquad \hat{se} = \frac{1}{\sqrt{n\, i_{\hat{\theta}}}} = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

95% CI $\boxed{\hat{\theta} \pm 1.96 \sqrt{\hat{\theta}(1-\hat{\theta})/n}}$

$$\frac{\theta}{1-\theta} = \text{odds} = p(\text{lose})/p(\text{win})$$

$$\text{log-odds} = \log \frac{\theta}{1-\theta} = \text{logit}(\theta)$$



$$g(\theta) = \log \frac{\theta}{1-\theta}$$

95% CI for $g(\theta)$?

Delta method

If $g : \Theta \to \mathbb{R}$ is differentiable

$$\sqrt{n}\,\left(g(\hat{\theta}) - g(\theta)\right) \xrightarrow{L} N\left(0, g'(\theta)^2 i_\theta^{-1}\right)$$

<u>Proof</u>: It can be proved that

$\sqrt{n}\,\left(g(\hat{\theta}) - g(\theta)\right)$ has the same limit distribution as the first-order Taylor expansion

$\sqrt{n}\, g'(\theta)\,(\hat{\theta} - \theta)$ at θ.

<u>Detail</u> :

$$g(\hat{\theta}_n) \approx g(\theta) + g'(\theta)(\hat{\theta}_n - \theta)$$

$$\sqrt{n}\,\left(g(\hat{\theta}_n) - g(\theta)\right) \approx \sqrt{n}\, g'(\theta)(\hat{\theta}_n - \theta)$$

Then we know that

$$\sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow{L} U \sim N(0, i_\theta^{-1})$$

Thus $\sqrt{n}\; g'(\theta)\,(\hat{\theta} - \theta) \xrightarrow{L} g'(\theta)\, U$

and $g'(\theta)\, U \sim N(0, g'(\theta)^2 i_\theta^{-1})$ //

... Example.

$$\phi = g(\theta) = \log \frac{\theta}{1-\theta}$$

$$g'(\theta) = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}.$$

$$g'(\theta)^2\, i_\theta^{-1} = \frac{1}{\theta^2(1-\theta)^2}\,\theta(1-\theta) = \frac{1}{\theta(1-\theta)}$$

$$\sqrt{n}\,(\hat{\phi}_n - \phi) \xrightarrow{L} N\left(0, \frac{1}{\theta(1-\theta)}\right)$$

$$\hat{se} = \frac{1}{\sqrt{n\,\hat{\theta}(1-\hat{\theta})}}$$

$$\boxed{\hat{\phi} \pm 1.96 \frac{1}{\sqrt{n\hat{\theta}(1-\hat{\theta})}}.}$$

$$\boxed{\text{Generalization of the delta method}}$$

If $\underset{\sim}{\theta} \in \Theta$ and $\underset{\sim}{\theta}$ is k-dimensional

If $g(\underset{\sim}{\theta})$ is a transformation

$g : \Theta \to \mathbb{R}$

with gradient $\nabla_\theta g(\underline{\theta}) = g'(\underline{\theta})_{(1\times k)}$

If $\underset{k\times k}{\underline{i}_\theta}$ is the Fisher information matrix

Then

$$\sqrt{n}\left(g(\hat{\underline{\theta}}) - g(\underline{\theta})\right) \xrightarrow{L} MN(\underline{0}, g'(\underline{\theta})\, i_{\underline{\theta}}\, g'(\theta)^T)$$

5.5 - Confidence regions and tests

Let $(X_1, \dots, X_n) \sim$ i.i.d $N(\mu, \sigma^2)$

1) The t test does not reject

$$H_0 : \mu = \mu_0$$

at level α when

$$-t_{n-1, 1-\alpha/2} \leq \frac{\bar{X} - \mu_0}{S_X/\sqrt{n}} \leq t_{n-1, 1-\alpha/2}$$

2) The conf. int. with confidence level $1-\alpha$ for μ is

$$\bar{X} \pm S_X/\sqrt{n} \cdot t_{n-1, 1-\alpha/2} .$$

A general result holds

1) The set of all values θ_0 that are NOT rejected in testing $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$ is a confidence interval for θ of level $1-\alpha$.

2) Given a confidence interval G_x for θ of level $1-\alpha$ the critical region

$$\{ \underline{x} : \quad G_x \not\ni \theta_0 \}$$

gives a test of level α for the null $H_0: \theta = \theta_0$.

Proof. — 1) Given θ_0, let $K(\theta_0)$ a critical region of level α for $H_0: \theta = \theta_0$. Then consider

$$G_X = \{ \theta_0 : \underline{X} \notin K(\theta_0) \}$$

As the test has level α then

$$P_{\theta^0}(\underline{X} \in K(\theta_0)) \leq \alpha \text{ for all } \theta_0$$

so that

$$P_{\theta_0}(\underline{X} \in G_X) \geq 1 - \alpha$$

2) Given G_X, a confidence region of level $1-\alpha$,

$$P_{\theta^0}(G_X \ni \theta_0) \geq 1 - \alpha$$

to test $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$ we consider the critical region

$$K(\theta_0) = \{ \underline{X} : G_X \not\ni \theta_0 \}$$

The test has level α because

$$P_{\theta_0}\left(\underline{X} \in K(\theta_0)\right) =$$

$$= P_{\theta_0}\left(G_X \not\ni \theta_0 \right) \leq \alpha \qquad \square$$

Likelihood Ratio Regions

We apply the procedure of deducing confidence regions from tests to the LRT.

Given $H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

The LRT rejects the null for

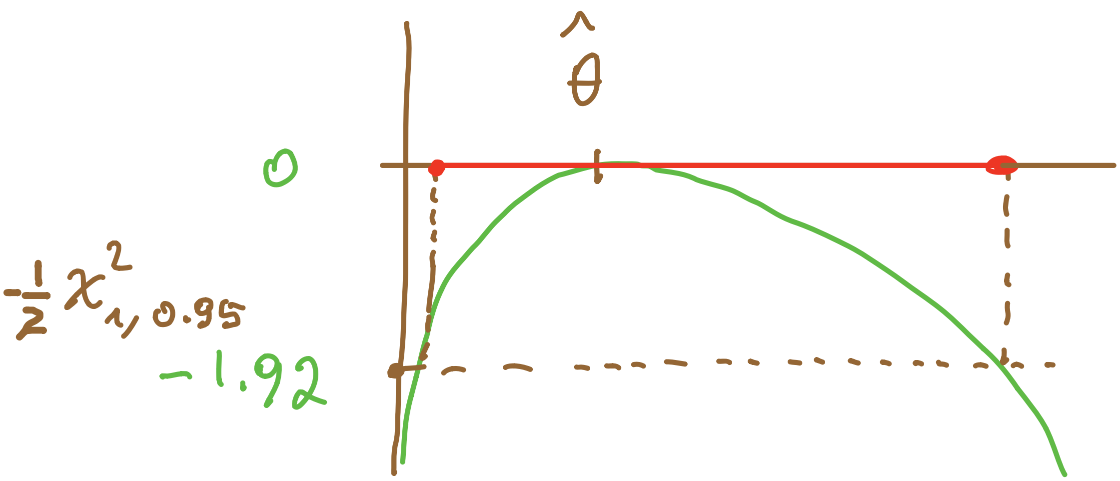
$$2 \log \frac{p_{\hat{\theta}}(x)}{p_{\theta_0}(x)} \geq \chi^2_{1,1-\alpha}$$

and "accepts" for

$$2 \log p_{\hat{\theta}} / p_{\theta_0} < \chi^2_{1,1-\alpha}$$

Therefore we can define a confidence region

$$\left\{ \theta_0 : \log p_{\theta_0}(x) - \log p_{\hat{\theta}}(x) \geq -\frac{1}{2}\chi^2_{1,1-\alpha} \right\}$$



$= \ell(\theta_0) - \ell(\hat{\theta})$

Note: the form of the region depends on the likelihood and may be

- asymmetric
- a union of disconnecte intervals.

Example 5.23

Let $(X_1, \ldots, X_n) \sim \mathrm{Exp}(\lambda)$ iid

$$\ell(\lambda) = n \log \lambda - \lambda \sum X_i.$$

$\rightarrow$ Notebook 5

5.7 Bayesian "confidence" regions

The logical choice is a region with probability $1-\alpha$ under the posterior. $\rightarrow$ CREDIBLE REGION

(a) symmetric, or

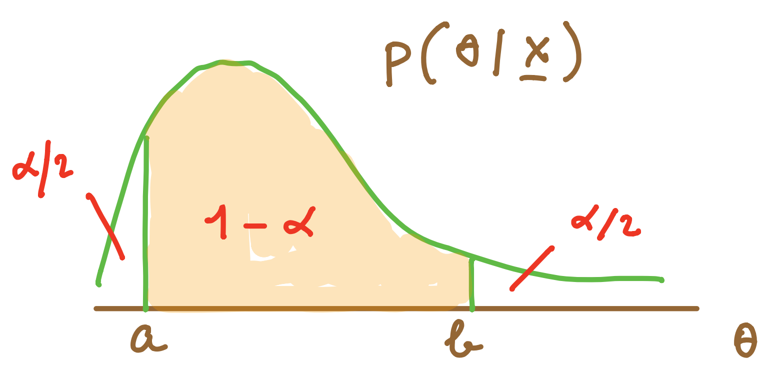
(b) smallest possible.

→ A Bayesian region is <u>not a confidence region.</u>

→ $1-\alpha$ refers to the probability that Θ will be inside the found region

→ For a Bayesian this is more important than having long-run guarantees.

▶ Example of (a)



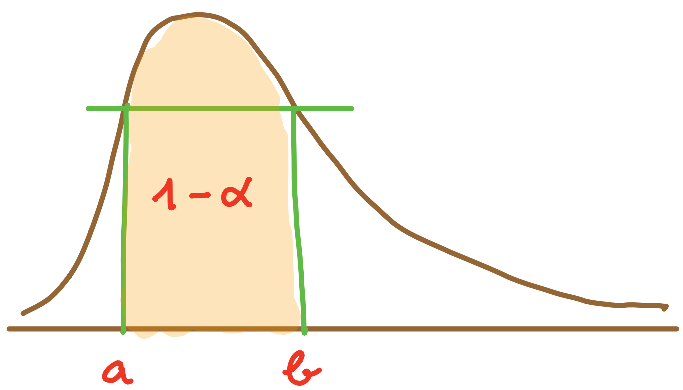
▶ Example of (b). Highest Posterior density. HPD

Is a region R in Θ such that

(i) $P(\theta \in R \mid \underline{x}) = 1-\alpha$

(ii) for $\theta' \in R$ and $\theta'' \notin R$

$$p(\theta' / \underline{x}) \geq p(\theta'' \mid \underline{x}).$$



Examples $\rightarrow$ Notebook.5

Example: $X \sim \text{Binomial}(10, \theta)$

Observe $X = 2$ successes

For $\theta \sim \text{Beta}(1,1)$

$$\theta \mid x \sim \text{Beta}(1+2, 1+8)$$

95% credible interval (0.06, 0.52)

After seeing the data we believe That $\theta \in (0.06, 0.52)$ with prob. 0.95.

95% HPD region $(0.0405, 0.4837)$ is narrower. See → Notebook 5

Notes:

HPD may not have symmetric Tails

HPD may not be an interval.

Large-sample properties of Bayesian procedures

- Bayesian ESTIMATORS are asymptotically normally distributed
- Differences with ML ESTIMATORS disappear as $n \to \infty$.

Bernstein- Von Mises

- Let $\underset{\sim}{X} = (X_1, \dots, X_n) \sim$ iid $P_\theta(x)$
- Let Θ be the r.v. with prior distribution
- Under the same assumptions of Theorem 5.9
- Fisher information is invertible and continuous

- ML estimator $\hat{\theta}_n$ consistent

Then

for every prior distribution continuous and with strictly positive density on Θ then

$$\left\| P(\Theta_n \in B \mid X_1, \dots, X_n) - N\left(\hat{\theta}_n, \frac{1}{n} i_\theta^{-1}\right)(B) \right\|_{TV} \xrightarrow[P_\theta]{} 0 \text{ as } n \to \infty$$

Here

- $\| P - Q \|_{TV} = \sup_{B \in \text{Boreliani}} | P(B) - Q(B) |$

- $N\left(\hat{\theta}_n, \frac{1}{n} i_\theta^{-1}\right)(B)$ is the probability that a normal distributed variable centred in $\hat{\theta}_n$ and with variance i_θ^{-1}/n appartenga a B.