

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

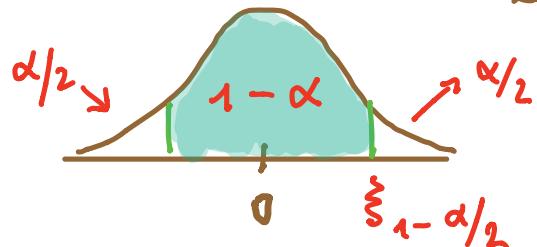
$$P_\theta(G_x \ni \theta) \geq C$$

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

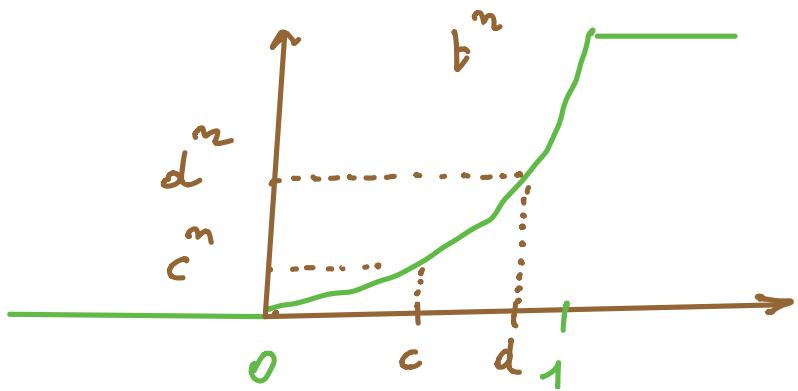
$$\underline{X} \stackrel{\text{iid}}{\sim} U(0, \theta) \Rightarrow \left(\frac{X_1}{\theta}, \dots, \frac{X_n}{\theta} \right) \stackrel{\text{iid}}{\sim} U(0, 1)$$

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 $n \rightarrow \infty$

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

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

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

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 $x = 90 \quad n = 200$

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

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

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 θ

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

is a mean 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

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

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

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

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

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

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

→ Notebook 5

LEMMA 5.10

Under certain regularity conditions

$$\triangleright E_{\theta} \ell_{\theta}'(x_i) = 0$$

$$\triangleright E_{\theta} \ell_{\theta}''(x_i) = -i_{\theta}$$

Proof Differentiate the identity.

$$\int P_{\theta}(x) dx = 1 \quad \forall \theta$$

(for discrete distribution, replace \int by Σ)

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

\Updownarrow | under certain regularity conditions |

$$\int P_{\theta}'(x) dx = 0$$

$$\Downarrow \quad \text{as } e_{\theta}'(x) = \frac{P_{\theta}'(x)}{P_{\theta}(x)}$$

$$\int e_{\theta}'(x) P_{\theta}(x) dx = 0$$

\Updownarrow

$$E_{\theta} e_{\theta}'(x_i) = 0$$

Differentiate Twice the identity

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



$$\int P_\theta''(x) dx = 0$$

$$\begin{aligned} \text{As } \ell_\theta''(x) &= \frac{d}{d\theta} \left[\frac{P_\theta'(x)}{P_\theta(x)} \right] \\ &= \frac{P_\theta''(x) P_\theta(x) - P_\theta'(x)^2}{P_\theta(x)^2} \end{aligned}$$

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

$$\ell_\theta''(x) = \frac{P_\theta''(x)}{P_\theta(x)} - \ell_\theta'(x)^2$$

We have also

$$\ell_\theta''(x) P_\theta(x) = P_\theta''(x) - \ell_\theta'(x)^2 P_\theta(x)$$

$$\begin{aligned}
 \int p_{\theta}''(x) dx &= 0 \\
 \Updownarrow \\
 \int e_{\theta}''(x) p_{\theta}(x) dx + \int e_{\theta}'(x)^2 p_{\theta}(x) dx &= 0 \\
 \Updownarrow \\
 E_{\theta} e_{\theta}''(x_1) + E_{\theta} [e_{\theta}'(x_1)^2] &= 0 \\
 \Updownarrow \\
 \boxed{E_{\theta} e_{\theta}''(x_1) = -\text{var}_{\theta}(e_{\theta}'(x_1)) = -i_{\theta}}
 \end{aligned}$$

Example 5.11 Poisson

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

$$i_{\theta} = \text{var}_{\theta}\left(\frac{x}{\theta} - 1\right) = \frac{1}{\theta^2} \theta = \frac{1}{\theta} .$$

$$i_{\theta} = E_{\theta} \left(-e_{\theta}''(x_1) \right) = -E_{\theta} \left(-\frac{x_1}{\theta^2} \right) = \frac{1}{\theta}$$

Estimate i_{θ} with

$$\blacktriangleright \hat{i}_{\theta} = \frac{1}{\hat{\theta}} = \frac{1}{\bar{x}}, \text{ or}$$

$$\begin{aligned}\blacktriangleright \hat{i}_{\theta} &= \frac{1}{n} \sum_{i=1}^n e''_{\theta}(x_i) = \frac{1}{n} \sum \frac{x_i}{\hat{\theta}^2} \\ &= \frac{1}{n \bar{x}^2} \cdot n \bar{x} = \frac{1}{\bar{x}}.\end{aligned}$$

Confidence interval:

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

Example 5.13 Exponential

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

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

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

Confidence interval

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

Proof of Theorem 5.9

Notation : $\theta = \text{true parameter}$
 $\vartheta = \text{generic parameter different from } \theta.$

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

The proof is structured as follows

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

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

i.e. that the MLE is consistent.

b) Proof that $\sqrt{n} (\hat{\theta}_m - \theta) \xrightarrow{d} \frac{1}{i_0} N(0, i_0)$

We skip the proof of a)

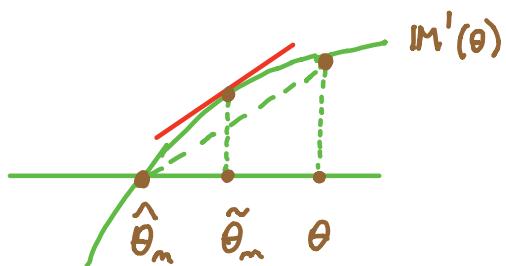
Proof of b)

$\hat{\theta}_m$ satisfies the likelihood eqns.

that can be written as

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

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



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

Thus, we can write

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

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

Apply the central limit theorem
to the numerator.

Let $Y_i = e'_\theta(x_i)$ $E_\theta(Y_i) = 0$
 $Y_i \sim \text{iid}$ $\text{var}_\theta(Y_i) = i_\theta$.

The numerator is _____

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

The denominator is

$$\frac{1}{n} \sum_1^n e''_{\tilde{\theta}_n}(x_i) = \frac{1}{n} \sum_1^n U_i \quad (*)$$

- U_1, \dots, U_n are dependent
and the LLN cannot be applied
as much.

► But $\tilde{\theta}_n \xrightarrow{P} \theta$ and we will prove

$\frac{1}{n} \sum_{i=1}^n U_i$ behaves like $\frac{1}{n} \sum_{i=1}^n l''_\theta(x_i)$

That satisfies the LLN and

$$\frac{1}{n} \sum_{i=1}^n l''_\theta(x_i) \xrightarrow{P} E_\theta l''_\theta(x_i) = -i_\theta$$

Given this we can conclude that

$$\frac{1}{n} \sum_{i=1}^n U_i \xrightarrow{P} -i_\theta.$$

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

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

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

$$\frac{1}{m} \sum_i^m (\ell_{\hat{\theta}_m}''(x_i) - \ell_{\theta}''(x_i)) \xrightarrow{P} 0$$

This requires a further assumption of the

- existence of $\ell_{\vartheta}'''(x)$ and
- $|\ell_{\vartheta}'''(x)| \leq K(x) \quad \vartheta \in \underset{\theta}{\text{---}}$

with $K(\cdot)$ satisfying $E_{\theta} K(x_i) < \infty$.

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

$$\left| \frac{\ell_{\vartheta}'''(x) - \ell_{\theta}'''(x)}{\vartheta - \theta} \right| = |\ell_{\vartheta}'''(z)| \leq K(x)$$

Thus, for all $\delta > 0$

$$P_{\theta} \left\{ \left| \frac{1}{m} \sum_i^m [\ell_{\hat{\theta}_m}''(x_i) - \ell_{\theta}''(x_i)] \right| > \delta \right\}$$

$$\leq P_\theta \left(\frac{1}{n} \sum_1^n K(x_i) \mid \tilde{\theta}_n - \theta \mid > \delta \right) \quad (**)$$

By the LLN $\frac{1}{n} \sum_1^n K(x_i) \rightarrow E_\theta K(x_1) < \infty$

thus, there exist a constant M s.t.

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

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

and thus also

$$\frac{1}{n} \sum_1^n \ell_{\tilde{\theta}_n}''(x_i) \xrightarrow{P} \frac{1}{n} \sum_1^n \ell_\theta''(x_i) \quad // \uparrow$$

5.4.2 * Multidimensional parameters

Let $\underline{\theta} \in \mathbb{R}^{k \times 1}$ $k > 1$.

The score = $\dot{l}_{\underline{\theta}}(\underline{x}) = \nabla_{\underline{\theta}} l_{\underline{\theta}}(\underline{x})$

$$= \left(\frac{\partial}{\partial \theta_1} l_{\underline{\theta}}(\underline{x}), \dots, \frac{\partial}{\partial \theta_k} l_{\underline{\theta}}(\underline{x}) \right)$$

The Fisher information

$$\dot{i}_{\underline{\theta}} = [i_{\underline{\theta},(r,t)}] \in \mathbb{R}^{k \times k} \text{ where}$$

$$i_{\underline{\theta},(r,t)} = \text{cov}_{\underline{\theta}} \left(\frac{\partial}{\partial \theta_r} l_{\underline{\theta}}(\underline{x}_i), \frac{\partial}{\partial \theta_t} l_{\underline{\theta}}(\underline{x}_i) \right)$$

Theorem 5.9 is still valid but

$$\sqrt{n} (\hat{\underline{\theta}}_m - \underline{\theta}) \xrightarrow{L} MN(\underline{0}, \dot{i}_{\underline{\theta}}^{-1})$$

That is:

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

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

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

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

Note that

$$(\hat{\theta} - \underline{\theta})^T n i_{\theta} (\hat{\theta} - \underline{\theta}) \stackrel{L}{\rightarrow} Z^T Z \sim \chi_k^2$$

With an estimator \hat{i}_{θ} of i_{θ}

we define a confidence region
of level $1-\alpha$ with

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

that is an ellipsoid. Because
the Fisher information matrix i_{θ}
is positive definite.

Example 5.17 Multinomial.

Suppose that we have (Y_1, Y_2, Y_3, Y_4) multinomially distributed with probabilities

$$\underline{\underline{P}} = (P_1, P_2, P_3, P_4)$$

and size n . This describes the number of people with blood phenotypes A, B, AB, O.

We take a sample of $n = 200$ people obtaining

A	B	AB	O	Tot
89	18	12	81	200

We want a confidence interval for $\underline{\underline{P}}$.

log-likelihood -

$$L(\underline{\underline{P}}) = \log \left(\begin{smallmatrix} 200 \\ 89 \ 18 \ 12 \ 81 \end{smallmatrix} \right) + \sum_{i=1}^n y_i \log P_i.$$

$$= \text{const.} + \sum_{i=1}^4 \gamma_i \log P_i + \gamma_4 \log(1 - P_1 - P_2 - P_3)$$

The MLE is (see Exercise 3.15)

$$\left(\frac{\gamma_1}{n}, \frac{\gamma_2}{n}, \frac{\gamma_3}{n}, \frac{\gamma_4}{n} \right) = \left(\frac{89}{200}, \frac{18}{200}, \frac{12}{200}, \frac{81}{200} \right)$$

The score function is

$$\frac{\partial \ell}{\partial P_r} = \frac{Y_r}{P_r} - \frac{Y_4}{P_4} \quad r = 1, 2, 3$$

n i.p. elements.

$$\text{var}_P \left(\frac{Y_1}{P_1} - \frac{Y_4}{P_4} \right) = \frac{n(1-P_1)}{P_1} + \frac{n(1-P_4)}{P_4} - \frac{1}{n} (-n P_1 P_4)$$

$$\begin{aligned}
 &= \frac{m q_1}{P_1} + \frac{m q_4}{P_4} + 2m \\
 &= \frac{m q_j P_4 + m P_j q_4 + 2m P_j P_4}{P_j P_4}
 \end{aligned}$$

$$= \frac{m (P_4 + P_j)}{P_4 P_j} = \boxed{m \left(\frac{1}{P_j} + \frac{1}{P_4} \right)}$$

$$\text{cov}_P \left[\frac{Y_1}{P_1} - \frac{Y_4}{P_4}, \frac{Y_2}{P_2} - \frac{Y_4}{P_4} \right] =$$

$$\begin{aligned}
 &= \text{cov}_P \left[\frac{Y_1}{P_1}, \frac{Y_2}{P_2} \right] - \text{cov}_P \left[\frac{Y_1}{P_1}, \frac{Y_4}{P_4} \right] \\
 &\quad - \text{cov}_P \left[\frac{Y_4}{P_4}, \frac{Y_2}{P_2} \right] + \text{cov} \left[\frac{Y_4}{P_4}, \frac{Y_4}{P_4} \right]
 \end{aligned}$$

$$\begin{aligned}
 &= -\frac{m \cancel{P_1 P_2}}{\cancel{P_1 P_2}} - \frac{-m \cancel{P_1 P_4}}{\cancel{P_1 P_4}} - \frac{-m \cancel{P_4 P_2}}{\cancel{P_4 P_2}} + \frac{m q_4}{P_4} \\
 &= -\cancel{m} + \cancel{m} + m + \frac{m (1 - P_4)}{P_4}
 \end{aligned}$$

$$\begin{aligned}
 &= n \left[1 + \frac{1 - p_4}{p_4} \right] = n \left[\frac{p_4 + 1 - p_4}{p_4} \right] \\
 &= \boxed{\frac{n}{p_4}}
 \end{aligned}$$

Thus

$$i_p = \begin{bmatrix} \frac{1}{p_1} + \frac{1}{p_4} & \frac{1}{p_4} & \frac{1}{p_4} \\ . & \frac{1}{p_2} + \frac{1}{p_4} & \frac{1}{p_4} \\ . & . & \frac{1}{p_3} + \frac{1}{p_4} \end{bmatrix}$$

The estimator of i_p can be found using $\hat{P}_r = \frac{y_r}{n}$ $r = 1, 2, 3$

$$\hat{i}_p = \begin{bmatrix} P_1(1-P_1) & -P_1P_2 & -P_1P_3 \\ . & P_2(1-P_2) & -P_2P_3 \\ . & . & P_3(1-P_3) \end{bmatrix}$$

• 30 - 131]

Delta method

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

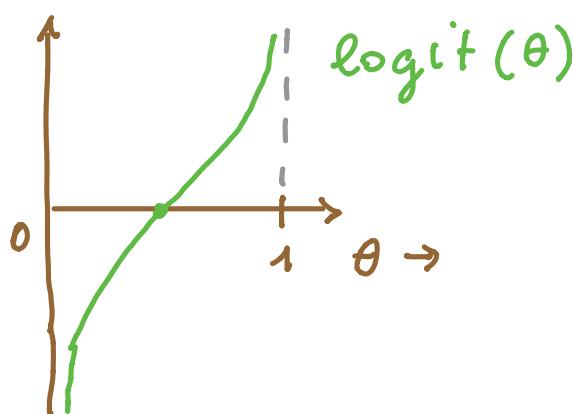
Example $X \sim \text{Bin}(n, \theta)$; $\hat{\theta} = \bar{x}/n$.

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

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

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

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



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

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

→ Delta method

If $g : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable

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

Proof: It can be proved that

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

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

Detail :

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

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

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.

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

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

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

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

$$\text{Se} = \frac{1}{\sqrt{n \hat{\theta}(1-\hat{\theta})}}$$

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

.