CentraleSupélec

Statistics and Learning

Lecturers (alphabetic order):

Julien Bect, Gilles Faÿ, Ziad Kobeissi, Laurent Le Brusquet,
Vincent Lescarret, Arshak Minasyan, Arthur Tenenhaus† & Xujia Zhu

† Course coordinator

Lecture 3/9

Asymptotic distributions and confidence intervals

Course objectives

- Take the asymptotic approach one step further, introducing asymptotic distributions.
- Understand confidence intervals and learn how to construct them (using asymptotic arguments if necessary)

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

Recap: Mathematical framework

In this section:

- We consider a statistical model

$$\left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \left\{\mathbb{P}^{X}_{\theta},\, \theta \in \Theta\right\}\right),$$

  assumed (most of the time) to be parametric ($\Theta \subset \mathbb{R}^p$).

- $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, defined on a common $(\Omega, \mathscr{F}, \mathbb{P}_\theta)$.

- We want to estimate a "quantity of interest":
  - either θ itself (we assume in this case that $\Theta \subset \mathbb{R}^p$),
  - or, more generally, $\eta = g(\theta) \in \mathbb{R}^q$.

Recap: Mathematical framework

In this section:

- We consider a statistical model

$$\left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \left\{\mathbb{P}_\theta^X, \theta \in \Theta\right\}\right),$$

  assumed (most of the time) to be parametric ($\Theta \subset \mathbb{R}^p$).

- $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, defined on a common $(\Omega, \mathscr{F}, \mathbb{P}_\theta)$.

- We want to estimate a "quantity of interest":
  - either θ itself (we assume in this case that $\Theta \subset \mathbb{R}^p$),
  - or, more generally, $\eta = g(\theta) \in \mathbb{R}^q$.

Recap: Mathematical framework

In this section:

- We consider a statistical model

$$\left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \left\{\mathbb{P}_\theta^{X}, \theta \in \Theta\right\}\right),$$

  assumed (most of the time) to be parametric ($\Theta \subset \mathbb{R}^p$).

- $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, defined on a common $(\Omega, \mathscr{F}, \mathbb{P}_\theta)$.

- We want to estimate a "quantity of interest":
  - either θ itself (we assume in this case that $\Theta \subset \mathbb{R}^p$),
  - or, more generally, $\eta = g(\theta) \in \mathbb{R}^q$.

Lecture outline

Lecture outline

Probability refresher: the Central Limit Theorem (CLT)

Theorem

Let

- a sequence $(X_n)_{n\in\mathbb{N}^*}$ of IID random vectors taking values in $\mathbb{R}^d$, with finite second order moments.
- $\mu = \mathbb{E}(X_1)$ and $\Sigma = \mathrm{var}(X_1) \in \mathbb{R}^{d\times d}$.

Then :
$$\boxed{\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,\Sigma),}$$

with $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^n X_i$ the sample mean.

$\Rightarrow$ The sample mean $\bar{X}_n$ is said to be

- an asymptotically Gaussian estimator of μ def: asympt. normality
- with convergence rate $\frac{1}{\sqrt{n}}$. def: convergence rate

Probability refresher: the Central Limit Theorem (CLT)

Theorem

Let

- a sequence $(X_n)_{n\in\mathbb{N}^*}$ of IID random vectors taking values in $\mathbb{R}^d$, with finite second order moments.
- $\mu = \mathbb{E}(X_1)$ and $\Sigma = \mathrm{var}(X_1) \in \mathbb{R}^{d\times d}$.

Then : $$\boxed{\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,\Sigma),}$$

with $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ the sample mean.

$\Rightarrow$ The sample mean $\bar{X}_n$ is said to be

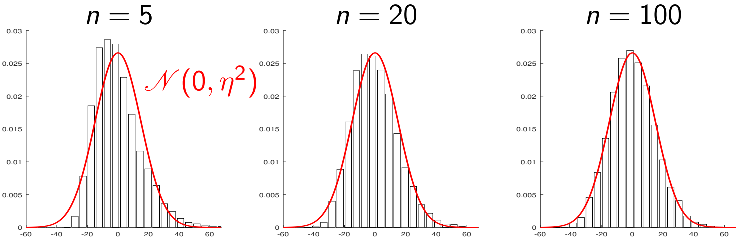
- an asymptotically Gaussian estimator of μ def: asympt. normality
- with convergence rate $\frac{1}{\sqrt{n}}$. def: convergence rate

Example: component reliability

Recall that

- $X_i \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, $\theta > 0$, and $\quad \eta = \mathbb{E}_\theta(X_1) = \frac{1}{\theta}$.
- $\hat{\eta}_n = \bar{X}_n$ is obtained by ML and the method of moments.

➠ Direct application of the CLT: $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta^2\right)$.



Histograms of $\sqrt{n}\left(\bar{X}_n - \eta\right)$ obtained from 10000 realizations of $\underline{X}_n$

Convergence rate

Let $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$ be a consistent estimator of $\eta = g(\theta)$.

Definition

If there exists a sequence $(a_n)_{n \in \mathbb{N}^*}$ of positive numbers such that:

- $\lim\limits_{n \to \infty} a_n = \infty$,
- $a_n \left(\hat{\eta}_n - \eta\right) \xrightarrow[n \to \infty]{\mathrm{d}} Z$,
- where Z is a non-degenerate* random variable (or vector),

then $\hat{\eta}_n$ converges to η at the rate $\frac{1}{a_n}$.

* We say that Z is degenerate if:

- scalar case: $\exists c \in \mathbb{R}, \;\; Z = c$ a.s.;
- vector case: $\exists a \in \mathbb{R}^q \setminus \{0\}, \exists c \in \mathbb{R}, \;\; \sum_{j=1}^{q} a_j Z^{(j)} = c$ a.s.

Remark. If Z has a second order moment, it can be shown that:

Z is non-degenerate iff its covariance matrix is invertible. Proof

Convergence rate

Let $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$ be a consistent estimator of $\eta = g(\theta)$.

Definition

If there exists a sequence $(a_n)_{n\in\mathbb{N}^*}$ of positive numbers such that:

- $\lim_{n\to\infty} a_n = \infty$,
- $a_n\left(\hat{\eta}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} Z$,
- where Z is a non-degenerate* random variable (or vector),

then $\hat{\eta}_n$ converges to η at the rate $\frac{1}{a_n}$.

* We say that Z is degenerate if:

- scalar case: $\exists c \in \mathbb{R},\ \ Z = c\ \text{ a.s.}$;
- vector case: $\exists a \in \mathbb{R}^q \setminus \{0\}, \exists c \in \mathbb{R},\ \ \sum_{j=1}^q a_j Z^{(j)} = c\ \text{ a.s.}$

Remark. If Z has a second order moment, it can be shown that:

Z is non-degenerate iff its covariance matrix is invertible.

Proof

Asymptotic normality

Let $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$ be a consistant estimator of $\eta = g(\theta)$.

Definition

If there exists

- a sequence $(a_n)_{n\in\mathbb{N}^*}$ of positive numbers s.t. $\lim_{n\to\infty} a_n = \infty$,
- a symmetric positive-definite matrix $\Sigma(\theta)$,

such that

$$a_n\left(\hat{\eta}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \Sigma(\theta)\right), \tag{1}$$

then we say that $\hat{\eta}_n$ is asymptotically normal.

Vocabulary. $\Sigma(\theta)$ is called the asymptotic covariance matrix (asymptotic variance, in the scalar case).

Note: it can be proved that (1) with $a_n \to +\infty$ implies (weak) consistency.

Convergence in distribution to a constant

Asymptotic normality

Let $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$ be a consistant estimator of $\eta = g(\theta)$.

Definition

If there exists

- a sequence $(a_n)_{n\in\mathbb{N}^*}$ of positive numbers s.t. $\lim\limits_{n\to\infty} a_n = \infty$,
- a symmetric positive-definite matrix $\Sigma(\theta)$,

such that

$$a_n\left(\hat{\eta}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \Sigma(\theta)\right), \qquad (1)$$

then we say that $\hat{\eta}_n$ is asymptotically normal.

Vocabulary. $\Sigma(\theta)$ is called the asymptotic covariance matrix (asymptotic variance, in the scalar case).

Note: it can be proved that (1) with $a_n \to +\infty$ implies (weak) consistency.

Convergence in distribution to a constant

Asymptotic normality

Let $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$ be a consistant estimator of $\eta = g(\theta)$.

Definition

If there exists

- a sequence $(a_n)_{n\in\mathbb{N}^*}$ of positive numbers s.t. $\lim_{n\to\infty} a_n = \infty$,
- a symmetric positive-definite matrix $\Sigma(\theta)$,

such that

$$a_n\left(\hat{\eta}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}\left(0, \Sigma(\theta)\right), \tag{1}$$

then we say that $\hat{\eta}_n$ is asymptotically normal.

Vocabulary. $\Sigma(\theta)$ is called the asymptotic covariance matrix (asymptotic variance, in the scalar case).

Note: it can be proved that (1) with $a_n \to +\infty$ implies (weak) consistency.

Convergence in distribution to a constant

Lecture outline

The continuous mapping theorem

Theorem (Mann-Wald)

Let

- $h : \mathbb{R}^d \to \mathbb{R}^q$ a measurable function,
- Y a random vector, taking values in $\mathbb{R}^d$,

such that

$$\boxed{h \text{ is continuous at the point } Y\text{, almost surely.}}$$

Then, for any sequence $(Y_n)_{n\in\mathbb{N}^*}$ of RV with values in $\mathbb{R}^d$,

$$\begin{array}{llll} \text{(i)} & Y_n \xrightarrow{\text{as}} Y & \Rightarrow & h(Y_n) \xrightarrow{\text{as}} h(Y), \\ \text{(ii)} & Y_n \xrightarrow{\mathbb{P}} Y & \Rightarrow & h(Y_n) \xrightarrow{\mathbb{P}} h(Y), \\ \text{(iii)} & Y_n \xrightarrow{\text{d}} Y & \Rightarrow & h(Y_n) \xrightarrow{\text{d}} h(Y). \end{array}$$

Proof: see CIP for the case where h is continuous. General case: admit.

Example: component reliability (cont'd)

Recall that

- $X_i \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, $\theta > 0$, and $\eta = \mathbb{E}_\theta(X_1) = \frac{1}{\theta}$.
- $\hat{\eta}_n = \bar{X}_n$ is obtained by ML and the method of moments.

Law of large numbers (strong and in L^2):

$$\hat{\eta}_n = \bar{X}_n \xrightarrow{\text{as}, L^2} \eta.$$

By the continuous mapping theorem:

$$\hat{\theta}_n = \frac{1}{\hat{\eta}_n} \xrightarrow{\text{as}} \frac{1}{\eta} = \theta,$$

therefore $\hat{\theta}_n$ is strongly consistent for the estimation of θ.

Remark: it can be shown that $\hat{\theta}_n$ is also consistent the L^2 sense.

Example: component reliability (cont'd)

Recall that

- $X_i \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, $\theta > 0$, and $\eta = \mathbb{E}_\theta(X_1) = \frac{1}{\theta}$.
- $\hat{\eta}_n = \bar{X}_n$ is obtained by ML and the method of moments.

Law of large numbers (strong and in L^2):

$$\hat{\eta}_n = \bar{X}_n \xrightarrow{\text{as}, L^2} \eta.$$

By the continuous mapping theorem:

$$\hat{\theta}_n = \frac{1}{\hat{\eta}_n} \xrightarrow{\text{as}} \frac{1}{\eta} = \theta,$$

therefore $\hat{\theta}_n$ is strongly consistent for the estimation of θ.

Remark: it can be shown that $\hat{\theta}_n$ is also consistent the L^2 sense.

Slutsky's theorem

Theorem

Let

- $(X_n)_{n\in\mathbb{N}^*}$ a sequence of random vectors that converges in distribution to a RV X:

$$X_n \xrightarrow[n\to\infty]{\mathrm{d}} X,$$

- $(Y_n)_{n\in\mathbb{N}^*}$ a sequence of random vectors that converges in distribution (therefore in probability) to a constant c:

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} c,$$

Then

$$(X_n, Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} (X, c).$$

Remark: $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} c$ implies $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ (constant limit).

Slutsky's theorem

Theorem

Let

- $(X_n)_{n\in\mathbb{N}^*}$ a sequence of random vectors that converges in distribution to a RV X:

$$X_n \xrightarrow[n\to\infty]{\mathrm{d}} X,$$

- $(Y_n)_{n\in\mathbb{N}^*}$ a sequence of random vectors that converges in distribution (therefore in probability) to a constant c:

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} c,$$

Then

$$(X_n, Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} (X, c).$$

Remark: $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} c$ implies $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ (constant limit).

Example: component reliability (cont'd)

Recall that (CLT) $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta^2\right)$.

Since $\bar{X}_n \xrightarrow[n\to\infty]{\text{as}} \eta$ (constant), we have by Slutsky's theorem:

$$\left(\sqrt{n}\left(\bar{X}_n - \eta\right), \bar{X}_n\right) \xrightarrow[n\to\infty]{\text{d}} (Z, \eta) \quad \text{with } Z \sim \mathcal{N}\left(0, \eta^2\right).$$

Therefore, by the continuous mapping theorem,

$$\sqrt{n}\,\frac{\left(\bar{X}_n - \eta\right)}{\bar{X}_n} \xrightarrow[n\to\infty]{\text{d}} \frac{Z}{\eta} \sim \mathcal{N}\left(0, 1\right),$$

since $(z, y) \mapsto \frac{z}{y}$ is continuous at any point where $y \neq 0$.

Remark. This result will be used to construct an asymptotic CI. exercise 4

Example: component reliability (cont'd)

Recall that (CLT) $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right)$.

Since $\bar{X}_n \xrightarrow[n\to\infty]{\text{as}} \eta$ (constant), we have by Slutsky's theorem:

$$\left(\sqrt{n}\left(\bar{X}_n - \eta\right), \bar{X}_n\right) \xrightarrow[n\to\infty]{\mathrm{d}} (Z, \eta) \quad \text{with } Z \sim \mathcal{N}\left(0, \eta^2\right).$$

Therefore, by the continuous mapping theorem,

$$\sqrt{n}\,\frac{\left(\bar{X}_n - \eta\right)}{\bar{X}_n} \xrightarrow[n\to\infty]{\mathrm{d}} \frac{Z}{\eta} \sim \mathcal{N}(0,1),$$

since $(z, y) \mapsto \frac{z}{y}$ is continuous at any point where $y \neq 0$.

Remark. This result will be used to construct an asymptotic CI. exercise 4

Example: component reliability (cont'd)

Recall that (CLT) $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta^2\right)$.

Since $\bar{X}_n \xrightarrow[n\to\infty]{\text{as}} \eta$ (constant), we have by Slutsky's theorem:

$$\left(\sqrt{n}\left(\bar{X}_n - \eta\right), \bar{X}_n\right) \xrightarrow[n\to\infty]{\text{d}} (Z, \eta) \quad \text{with } Z \sim \mathcal{N}\left(0, \eta^2\right).$$

Therefore, by the continuous mapping theorem,

$$\sqrt{n}\,\frac{\left(\bar{X}_n - \eta\right)}{\bar{X}_n} \xrightarrow[n\to\infty]{\text{d}} \frac{Z}{\eta} \sim \mathcal{N}\left(0, 1\right),$$

since $(z, y) \mapsto \frac{z}{y}$ is continuous at any point where $y \neq 0$.

Remark. This result will be used to construct an asymptotic CI. exercise 4

Linearization method ("delta method")

"Delta theorem" (**scalar** case)

Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of RV with values in $\mathbb{R}$, s.t.

$$\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} Z,$$

with Y a random variable, taking values in $\mathbb{R}$, and $m \in \mathbb{R}$.

Then, for any $h : \mathbb{R} \to \mathbb{R}$ that is differentiable at m,

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} h'(m)\,Z,$$

Intuition: $h(y) - h(m) \;\approx\; h'(m)\,(y-m)$.

Proof

Linearization method ("delta method")

"Delta theorem" (**scalar** case)

Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of RV with values in $\mathbb{R}$, s.t.

$$\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} Z,$$

with Y a random variable, taking values in $\mathbb{R}$, and $m \in \mathbb{R}$.

Then, for any $h : \mathbb{R} \to \mathbb{R}$ that is differentiable at m,

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} h'(m)\,Z,$$

Intuition: $h(y) - h(m) \approx h'(m)\,(y - m)$.

Proof

Linearization method (“delta method”)

“Delta theorem” (**scalar** case)

Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of RV with values in $\mathbb{R}$, s.t.

$$\sqrt{n}\,(Y_n - m) \;\xrightarrow[n\to\infty]{\mathrm{d}}\; Z,$$

with Y a random variable, taking values in $\mathbb{R}$, and $m \in \mathbb{R}$.

Then, for any $h : \mathbb{R} \to \mathbb{R}$ that is differentiable at m,

$$\sqrt{n}\,(h(Y_n) - h(m)) \;\xrightarrow[n\to\infty]{\mathrm{d}}\; h'(m)\,Z,$$

Intuition: $h(y) - h(m) \;\approx\; h'(m)\,(y - m)$.

Proof

Linearization method ("delta method")

"Delta theorem" (**vectorial** case)

Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of RV with values in $\mathbb{R}^d$, s.t.

$$\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} Z,$$

with Y a RV taking values in $\mathbb{R}^d$ and $m \in \mathbb{R}^d$.

Then, for any $h : \mathbb{R}^d \to \mathbb{R}^q$ that is differentiable at m,

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} (Dh)(m)\,Z,$$

with $(Dh)(m)$ the Jcobian matrix of h at m:

$$(Dh)(m) = \Big((\partial_j h_i)(m)\Big)_{1\le i\le q,\, 1\le j\le d}.$$

Linearization method ("delta method")

"Delta theorem" (**vectorial** case)

Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of RV with values in $\mathbb{R}^d$, s.t.

$$\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} Z,$$

with Y a RV taking values in $\mathbb{R}^d$ and $m \in \mathbb{R}^d$.

Then, for any $h : \mathbb{R}^d \to \mathbb{R}^q$ that is differentiable at m,

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} (Dh)(m)\,Z,$$

with $(Dh)(m)$ the Jcobian matrix of h at m:

$$(Dh)(m) = \Big((\partial_j h_i)(m)\Big)_{1\le i\le q,\ 1\le j\le d}.$$

Delta theorem in the Gaussian case

Scalar case.
If $\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,\, \sigma^2)$, then

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0,\, (h'(m))^2\sigma^2\right).$$

Vector case

If $\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,\, \Sigma)$, then

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0,\, (Dh)(m)\,\Sigma\,(Dh)(m)^\top\right).$$

Delta theorem in the Gaussian case

Scalar case.

If $\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}(0,\, \sigma^2)$, then

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}\left(0,\; (h'(m))^2\sigma^2\right).$$

Vector case

If $\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}(0,\, \Sigma)$, then

$$\sqrt{n}\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}\left(0,\; (Dh)(m)\,\Sigma\,(Dh)(m)^\top\right).$$

Example: component reliability (cont'd)

We already saw that:

- $\hat{\theta}_n = 1/\bar{X}_n$ is a consistent estimator of θ,
- $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta^2\right)$, where $\eta = \frac{1}{\theta}$.

Using the delta method with $h(\eta) = \frac{1}{\eta}$

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0,\, \eta^2\left(h'(\eta)\right)^2\right),$$

$$h'(\eta) = -\frac{1}{\eta^2} \quad \Longrightarrow \quad \sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \theta^2\right).$$

Conclusion: $\hat{\theta}_n$ is asymptotically Gaussian,

and its convergence rate is $\frac{1}{\sqrt{n}}$.

Example: component reliability (cont'd)

We already saw that:

- $\hat{\theta}_n = 1/\bar{X}_n$ is a consistent estimator of θ,
- $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta^2\right)$, where $\eta = \frac{1}{\theta}$.

Using the delta method with $h(\eta) = \frac{1}{\eta}$

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0,\, \eta^2\left(h'(\eta)\right)^2\right),$$

$$h'(\eta) = -\frac{1}{\eta^2} \quad\Longrightarrow\quad \sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \theta^2\right).$$

Conclusion: $\hat{\theta}_n$ is asymptotically Gaussian,

and its convergence rate is $\frac{1}{\sqrt{n}}$.

Example: component reliability (cont'd)

We already saw that:

- $\hat{\theta}_n = 1/\bar{X}_n$ is a consistent estimator of θ,
- $\sqrt{n}\left(\bar{X}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right)$, where $\eta = \frac{1}{\theta}$.

Using the delta method with $h(\eta) = \frac{1}{\eta}$

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0,\ \eta^2\left(h'(\eta)\right)^2\right),$$

$$h'(\eta) = -\frac{1}{\eta^2} \quad \Longrightarrow \quad \sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \theta^2\right).$$

Conclusion: $\hat{\theta}_n$ is asymptotically Gaussian,

and its convergence rate is $\frac{1}{\sqrt{n}}$.

Asymptotic comparison of (scalar) estimators (1/2)

With **asymptotic variances.**
Exemple of use with "component reliability" for $\eta = \mathbb{E}_\theta(X_1)$.

1) For $\hat{\eta}^{(1)} = \bar{X}_n$, we have (CLT): $\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right)$.

2) For $\hat{\eta}^{(2)} = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$ (see lecture #1) ?

- Since $\mathbb{E}\left(X_1^2\right) = 2\eta^2$ and $\mathbb{E}\left(X_1^4\right) = 24\eta^4$, we have (CLT):

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - 2\eta^2\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, 20\,\eta^4\right).$$

- Hence, using the delta method with $h(z) = \sqrt{\frac{1}{2}z}$,

$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{5}{4}\eta^2\right).$$

Asymptotic comparison of (scalar) estimators (1/2)

With **asymptotic variances.**
Exemple of use with "component reliability" for $\eta = \mathbb{E}_\theta(X_1)$.

1) For $\hat{\eta}^{(1)} = \bar{X}_n$, we have (CLT): $\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta^2\right)$.

2) For $\hat{\eta}^{(2)} = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$ (see lecture #1) ?

- Since $\mathbb{E}\left(X_1^2\right) = 2\eta^2$ and $\mathbb{E}\left(X_1^4\right) = 24\eta^4$, we have (CLT):

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\eta^2\Big) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, 20\,\eta^4\right).$$

- Hence, using the delta method with $h(z) = \sqrt{\frac{1}{2}z}$,

$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \frac{5}{4}\eta^2\right).$$

Asymptotic comparison of (scalar) estimators (2/2)

In summary:

$$\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right),$$
$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{5}{4}\eta^2\right).$$

We observe that

- the two estimators are asymptotically normal,
- have the same convergence rate,
- but the asymptotic variance of $\hat{\eta}^{(1)}$ is smaller.

➠ We say that $\hat{\eta}^{(1)}$ is asymptotically preferable to $\hat{\eta}^{(2)}$.

Asymptotic comparison of (scalar) estimators (2/2)

In summary:

$$\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right),$$
$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{5}{4}\eta^2\right).$$

We observe that

- the two estimators are asymptotically normal,
- have the same convergence rate,
- but the asymptotic variance of $\hat{\eta}^{(1)}$ is smaller.

➠ We say that $\hat{\eta}^{(1)}$ is asymptotically preferable to $\hat{\eta}^{(2)}$.

Asymptotic comparison of (scalar) estimators (2/2)

In summary:

$$\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right),$$
$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{5}{4}\eta^2\right).$$

We observe that

- the two estimators are asymptotically normal,
- have the same convergence rate,
- but the asymptotic variance of $\hat{\eta}^{(1)}$ is smaller.

➠ We say that $\hat{\eta}^{(1)}$ is asymptotically preferable to $\hat{\eta}^{(2)}$.

Asymptotic comparison of (scalar) estimators (2/2)

In summary:

$$\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta^2\right),$$
$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{5}{4}\eta^2\right).$$

We observe that

- the two estimators are asymptotically normal,
- have the same convergence rate,
- but the asymptotic variance of $\hat{\eta}^{(1)}$ is smaller.

➠ We say that $\hat{\eta}^{(1)}$ is asymptotically preferable to $\hat{\eta}^{(2)}$.

Asymptotic comparison of (scalar) estimators (2/2)

In summary:

$$\sqrt{n}\left(\hat{\eta}^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}\left(0, \eta^2\right),$$
$$\sqrt{n}\left(\hat{\eta}^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathscr{N}\left(0, \frac{5}{4}\eta^2\right).$$

We observe that

- the two estimators are asymptotically normal,
- have the same convergence rate,
- but the asymptotic variance of $\hat{\eta}^{(1)}$ is smaller.

➠ We say that $\hat{\eta}^{(1)}$ is asymptotically preferable to $\hat{\eta}^{(2)}$.

Lecture outline

Asymptotic efficiency

Recall the Cramér-Rao lower bound (scalar parameter)

$\forall \hat{\theta}$ regular UE of θ, $\forall \theta \in \Theta$,

$$R_\theta\left(\hat{\theta}\right) = \mathrm{var}_\theta\left(\hat{\theta}\right) \ \geq \ \frac{1}{n}\, I_1^{-1}(\theta),$$

with $I_1(\theta) = \mathrm{var}_\theta\left(S_\theta(X_1)\right)$.

➠ When equality holds for all θ, the estimator is called efficient.

Asymptotic efficiency

Definition. An estimator is called asymptotically efficient if

- it is asymptotically normal at the rate $\frac{1}{\sqrt{n}}$,
- with asymptotic variance $I_1^{-1}(\theta)$.

Remark: this definition is valid for the vector-valued case as well, replacing the variance by the covariance matrix

Asymptotic efficiency

Recall the Cramér-Rao lower bound (scalar parameter)

$\forall \hat{\theta}$ regular UE of θ, $\forall \theta \in \Theta$,

$$R_\theta\left(\hat{\theta}\right) = \mathrm{var}_\theta\left(\hat{\theta}\right) \ \geq \ \frac{1}{n}\, I_1^{-1}(\theta),$$

with $I_1(\theta) = \mathrm{var}_\theta\left(S_\theta(X_1)\right)$.

➠ When equality holds for all θ, the estimator is called efficient.

Asymptotic efficiency

Definition. An estimator is called asymptotically efficient if

- it is asymptotically normal at the rate $\frac{1}{\sqrt{n}}$,
- with asymptotic variance $I_1^{-1}(\theta)$.

Remark: this definition is valid for the vector-valued case as well, replacing the variance by the covariance matrix

Asymptotic efficiency of the MLE

Context: $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$ and, $\forall \theta \in \Theta$, P_θ admits a pdf f_θ.

Definition: regular model

The statistical model is called regular if

- conditions C_0–C_2 are verified (def. given in lecture 2)
- The conditions C_3 & C_4 are verified Conditions C_3 & C_4
- $\forall \theta \in \Theta$, the Fisher information matrix $I_1(\theta)$ is positive definite.

Theorem

If the statistical model is regular and if the MLE $\hat{\theta}_n$ is consistent, then it is asymptotically efficient :

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, I_1^{-1}(\theta)\right).$$

Counter-example with a non-regular model

Asymptotic efficiency of the MLE

Context: $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$ and, $\forall \theta \in \Theta$, P_θ admits a pdf f_θ.

Definition: regular model

The statistical model is called regular if

- conditions C_0–C_2 are verified (def. given in lecture 2)
- The conditions C_3 & C_4 are verified Conditions C_3 & C_4
- $\forall \theta \in \Theta$, the Fisher information matrix $I_1(\theta)$ is positive definite.

Theorem

If the statistical model is regular and if the MLE $\hat{\theta}_n$ is consistent, then it is asymptotically efficient :

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, I_1^{-1}(\theta)\right).$$

Counter-example with a non-regular model

Fisher information in regular models

Reminder. The Fisher information brought by $\underline{X}$ is the matrix

$$I(\theta) = \mathrm{var}_\theta(S_\theta) = \mathbb{E}_\theta\left(S_\theta\, S_\theta^\top\right).$$

Proposition: another expression for the FIM

In a regular model, we have

$$I(\theta) = -\,\mathbb{E}_\theta\left(\nabla_\theta\left(S_\theta^\top\right)\right), \qquad (\star)$$

In other words : $\forall \theta \in \Theta$, $\forall j \leq p$, $\forall k \leq p$,

$$(I(\theta))_{j,k} = -\,\mathbb{E}_\theta\left(\frac{\partial}{\partial \theta_j} S_\theta^{(k)}\right) = -\,\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k} \ln f_\theta(\underline{X})\right).$$

Remark: actually, if C_0–C_3 hold, then C_4 and $(\star)$ are equivalent.

Fisher information in regular models

Reminder. The Fisher information brought by $\underline{X}$ is the matrix

$$I(\theta) = \mathrm{var}_\theta(S_\theta) = \mathbb{E}_\theta\left(S_\theta\, S_\theta^\top\right).$$

Proposition: another expression for the FIM

In a regular model, we have

$$I(\theta) = -\,\mathbb{E}_\theta\left(\nabla_\theta\left(S_\theta^\top\right)\right), \qquad (\star)$$

In other words : $\forall\theta \in \Theta$, $\forall j \leq p$, $\forall k \leq p$,

$$\left(I(\theta)\right)_{j,k} = -\,\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta_j} S_\theta^{(k)}\right) = -\,\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ln f_\theta(\underline{X})\right).$$

Remark: actually, if C_0–C_3 hold, then C_4 and $(\star)$ are equivalent.

Example: component reliability (cont'd)

Question: is $\hat{\theta}_n = 1/\bar{X}_n$ asymptotically efficient?

We have already computed the score: $S_\theta(X_1) = \frac{1}{\theta} - X_1$.

Computation of Fisher's information (two approaches):

Comput. of $\mathbb{E}_\theta\left(S_\theta(X_1)^2\right)$	Comput. of $-\mathbb{E}_\theta\left(\frac{\partial S_\theta}{\partial \theta}(X_1)\right)$
$I_1(\theta) = \mathrm{var}_\theta(X_1) = \eta^2 = \frac{1}{\theta^2}$	$I_1(\theta) = -\mathbb{E}_\theta\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$

Conclusion: since $\sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \theta^2\right)$,

$$\hat{\theta}_n = \frac{1}{\bar{X}_n} \text{ is asymptotically efficient.}$$

➠ We recover the conclusions of the theorem (C_0–C_4 hold indeed).

Example: component reliability (cont'd)

Question: is $\hat{\theta}_n = 1/\bar{X}_n$ asymptotically efficient?

We have already computed the score: $S_\theta(X_1) = \frac{1}{\theta} - X_1$.

Computation of Fisher's information (two approaches):

Comput. of $\mathbb{E}_\theta\left(S_\theta(X_1)^2\right)$	Comput. of $-\mathbb{E}_\theta\left(\frac{\partial S_\theta}{\partial \theta}(X_1)\right)$
$I_1(\theta) = \text{var}_\theta(X_1) = \eta^2 = \frac{1}{\theta^2}$	$I_1(\theta) = -\mathbb{E}_\theta\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$

Conclusion: since $\sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \theta^2\right)$,

$$\hat{\theta}_n = \frac{1}{\bar{X}_n} \text{ is asymptotically efficient.}$$

➠ We recover the conclusions of the theorem (C_0–C_4 hold indeed).

Example: component reliability (cont'd)

Question: is $\hat{\theta}_n = 1/\bar{X}_n$ asymptotically efficient?

We have already computed the score: $S_\theta(X_1) = \frac{1}{\theta} - X_1$.

Computation of Fisher's information (two approaches):

Comput. of $\mathbb{E}_\theta\left(S_\theta(X_1)^2\right)$	Comput. of $-\mathbb{E}_\theta\left(\frac{\partial S_\theta}{\partial \theta}(X_1)\right)$
$I_1(\theta) = \mathrm{var}_\theta(X_1) = \eta^2 = \frac{1}{\theta^2}$	$I_1(\theta) = -\mathbb{E}_\theta\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$

Conclusion: since $\sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \theta^2\right)$,

$$\hat{\theta}_n = \frac{1}{\bar{X}_n} \text{ is asymptotically efficient.}$$

➠ We recover the conclusions of the theorem (C_0–C_4 hold indeed).

Lecture outline

Lecture outline

Motivation

Problem

A point estimator necessarily makes some estimation error.
How can we "report" this error?

Two approaches:

- provide, in addition to the estimated value,
  - the distribution of the estimator $\hat{\eta}$, exact or approximate,
  - or at least some "measure of dispersion" (e.g., its standard deviation);
- give, instead of a point estimation $\hat{\eta}$,

$$\boxed{\text{a confidence interval for } \eta.}$$

Motivation

Problem

A point estimator necessarily makes some estimation error.
How can we "report" this error?

Two approaches:

- provide, in addition to the estimated value,
  - the distribution of the estimator $\hat{\eta}$, exact or approximate,
  - or at least some "measure of dispersion" (e.g., its standard deviation);
- give, instead of a point estimation $\hat{\eta}$,

$$\boxed{\text{a confidence interval for } \eta.}$$

Motivation

Problem

A point estimator necessarily makes some estimation error.
How can we "report" this error?

Two approaches:

- provide, in addition to the estimated value,
  - the distribution of the estimator $\hat{\eta}$, exact or approximate,
  - or at least some "measure of dispersion" (e.g., its standard deviation);
- give, instead of a point estimation $\hat{\eta}$,

  $\boxed{\text{a confidence interval for } \eta.}$

Coverage probability

Reminder. $\eta = g(\theta)$.

Let

- $\mathcal{P}(N)$ the power set (set of all subsets) of $N = g(\Theta)$.
- a statistic $C(\underline{X})$ with values in $\mathcal{P}(N)$.

Goal. Having $\eta \in C(\underline{X})$ with high probability.

Definition

For $\theta \in \Theta$, the coverage probability of $C(\underline{X})$ is defined as:

$$\mathbb{P}_\theta\left(\eta \in C\left(\underline{X}\right)\right)$$

⚠ In general, the coverage probability depends on the underlyign distribution, i.e., on θ.

Coverage probability

Reminder. $\eta = g(\theta)$.

Let

- $\mathcal{P}(N)$ the power set (set of all subsets) of $N = g(\Theta)$.
- a statistic $C(\underline{X})$ with values in $\mathcal{P}(N)$.

Goal. Having $\eta \in C(\underline{X})$ with high probability.

Definition

For $\theta \in \Theta$, the coverage probability of $C(\underline{X})$ is defined as:

$$\mathbb{P}_\theta \left(\eta \in C(\underline{X})\right)$$

⚠ In general, the coverage probability depends on the underlyign distribution, i.e., on θ.

Coverage probability

Reminder. $\eta = g(\theta)$.

Let

- $\mathcal{P}(N)$ the power set (set of all subsets) of $N = g(\Theta)$.
- a statistic $C(\underline{X})$ with values in $\mathcal{P}(N)$.

Goal. Having $\eta \in C(\underline{X})$ with high probability.

Definition

For $\theta \in \Theta$, the coverage probability of $C(\underline{X})$ is defined as:

$$\mathbb{P}_\theta\left(\eta \in C\left(\underline{X}\right)\right)$$

⚠ In general, the coverage probability depends on the underlyign distribution, i.e., on θ.

Confidence regions and confidence intervals

We aim to control the coverage probability.

Let $\alpha \in]0, 1[$.

Definition: confidence region with **level** $1 - \alpha$

A confidence region with level (at least) $1 - \alpha$ for η is a statistics $C(\underline{X})$ taking values in $\mathcal{P}(N)$, such that:

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta\left(\eta \in C\left(\underline{X}\right)\right) \geq 1 - \alpha.$$

We say that $C(\underline{X})$ is a confidence region with level *exactly* $1 - \alpha$ if

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta\left(\eta \in C\left(\underline{X}\right)\right) = 1 - \alpha.$$

(Some authors also write: of "size" $1 - \alpha$.)

Scalar case. If $C(\underline{X})$ is an interval, it is called a confidence interval.

Confidence regions and confidence intervals

We aim to control the coverage probability.

Let $\alpha \in]0,1[$.

Definition: confidence region with **level** $1-\alpha$

A confidence region with level (at least) $1-\alpha$ for η is a statistics $C\left(\underline{X}\right)$ taking values in $\mathcal{P}(N)$, such that:

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta\left(\eta \in C\left(\underline{X}\right)\right) \geq 1-\alpha.$$

We say that $C\left(\underline{X}\right)$ is a confidence region with level *exactly* $1-\alpha$ if

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta\left(\eta \in C\left(\underline{X}\right)\right) = 1-\alpha.$$

(Some authors also write: of "size" $1-\alpha$.)

Scalar case. If $C(\underline{X})$ is an interval, it is called a confidence interval.

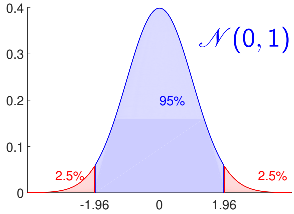
Example: $\mathcal{N}(\mu, \sigma_0^2)$ n-sample, with known σ_0^2

Since $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma_0^2}{n}\right)$, $T = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \sim \mathcal{N}(0,1)$, for $\alpha = 5\%$:

$$\mathbb{P}_\mu\left(\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \in [-1.96,\ 1.96]\right) \approx 1 - \alpha = 95\%,$$

where 1.96 is the quantile of order 97.5% of the distribution $\mathcal{N}(0,1)$.

def.: quantile



We "pivot" to obtain a CI with level exactly 95% :

$$\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \in [-1.96,\ 1.96] \quad \Leftrightarrow \quad \mu \in C(\underline{X}) = \left[\bar{X} - 1.96\,\frac{\sigma_0}{\sqrt{n}},\ \bar{X} + 1.96\,\frac{\sigma_0}{\sqrt{n}}\right].$$

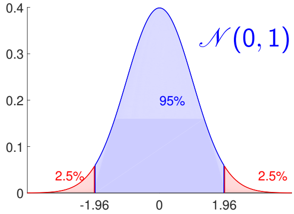
Example: $\mathcal{N}(\mu, \sigma_0^2)$ n-sample, with known σ_0^2

Since $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma_0^2}{n}\right)$, $T = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \sim \mathcal{N}(0,1)$, for $\alpha = 5\%$:

$$\mathbb{P}_\mu\left(\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \in [-1.96,\ 1.96]\right) \approx 1 - \alpha = 95\%,$$

where 1.96 is the quantile of order 97.5% of the distribution $\mathcal{N}(0,1)$.

def.: quantile

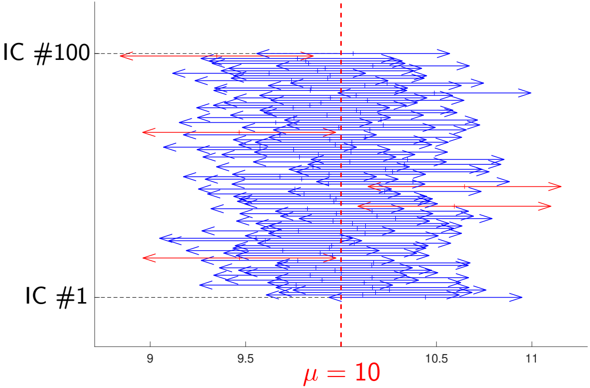


We "pivot" to obtain a CI with level exactly 95% :

$$\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \in [-1.96,\ 1.96] \quad \Leftrightarrow \quad \mu \in C(\underline{X}) = \left[\bar{X} - 1.96\,\frac{\sigma_0}{\sqrt{n}},\ \bar{X} + 1.96\,\frac{\sigma_0}{\sqrt{n}}\right].$$

Interpretation: simulations

We simulate 100 realizations with $\mu = 10$ and $\sigma_0 = 1$.



In red: realizations where the IC does not contain $\mu = 10$.

➠ The proportion of cases where the CI does not contain μ is (approx.) α.

Lecture outline

Pivotal functions

The method can be formalized using pivotal functions.

Definitions

A function

$$T : \underline{\mathcal{X}} \times N \;\to\; \mathbb{R}$$

is called pivotal if the distribution of the RV $T = T(\underline{X}, \eta)$ does not depend on θ. We say that the distribution of $T(\underline{X}, \eta)$ is free from the parameter.

Back to the **example**: $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma_0^2)$ with known σ_0.

Then $T = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma_0}$ is pivotal since

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma_0} \;\sim\; \mathcal{N}(0,1).$$

Remark: we can also choose $T = \sqrt{n}\,(\bar{X}_n - \mu) \sim \mathcal{N}(0, \sigma_0^2)$.

Pivotal functions

The method can be formalized using pivotal functions.

Definitions

A function

$$T : \underline{\mathcal{X}} \times N \;\to\; \mathbb{R}$$

is called pivotal if the distribution of the RV $T = T(\underline{X}, \eta)$ does not depend on θ. We say that the distribution of $T(\underline{X}, \eta)$ is free from the parameter.

Back to the **example**: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma_0^2)$ with known σ_0.

Then $T = \sqrt{n}\, \frac{\bar{X}_n - \mu}{\sigma_0}$ is pivotal since

$$\sqrt{n}\, \frac{\bar{X}_n - \mu}{\sigma_0} \;\sim\; \mathcal{N}(0, 1).$$

Remark: we can also choose $T = \sqrt{n}\, \left(\bar{X}_n - \mu\right) \sim \mathcal{N}(0, \sigma_0^2)$.

Probability refresher: quantiles

Definition: quantile of order r

Let $F(x)$ be the cdf of a probability distribution on $\mathbb{R}$.

For $0 < r < 1$, the quantile of order r of the distribution is defined as:

$$q_r = \inf\{x \in \mathbb{R},\ F(x) \geq r\} = \min\{x \in \mathbb{R},\ F(x) \geq r\}.$$

Properties:

- If F is continuous, then $F(q_r) = r$.
- If, in addition, F is strictly increasing, then $q_r = F^{-1}(r)$.

Probability refresher: quantiles

Definition: quantile of order r

Let $F(x)$ be the cdf of a probability distribution on $\mathbb{R}$.

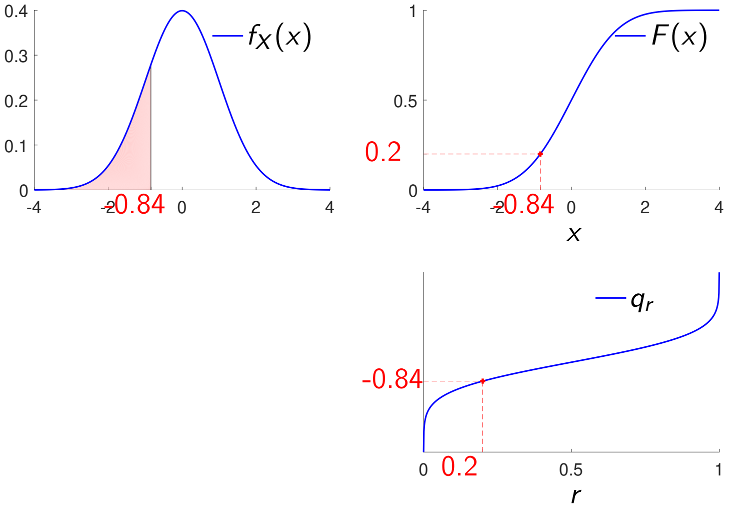
For $0 < r < 1$, the quantile of order r of the distribution is defined as:

$$q_r = \inf\left\{x \in \mathbb{R},\ F(x) \geq r\right\} = \min\left\{x \in \mathbb{R},\ F(x) \geq r\right\}.$$

Properties:

- If F is continuous, then $F(q_r) = r$.
- If, in addition, F is strictly increasing, then $q_r = F^{-1}(r)$.

Quantile function of the $\mathcal{N}(0,1)$ distribution



How to use pivotal functions

Let $T\left(\underline{X}, \eta\right)$ be a pivotal function and $\alpha \in \left]0, 1\right[$.

Proposition

Assume that the cdf F of $T\left(\underline{X}, \eta\right)$ is continuous and strictly increasing, and denote by $q_r = F^{-1}(r)$ the quantile of order r.

Then, for all $\gamma \in [0, \alpha]$:

$$\begin{aligned} C^{\gamma}\left(\underline{X}\right) &= \left\{\eta \in N \text{ such that } q_{\gamma} \leq T\left(\underline{X}, \eta\right) \leq q_{\gamma+1-\alpha}\right\} \\ &= T^{-1}\left(\underline{X}, \left[q_{\gamma}, q_{\gamma+1-\alpha}\right]\right) \end{aligned}$$

is a confidence interval for η with level exactly $1 - \alpha$.

Proof.
$$\begin{aligned} \mathbb{P}_{\theta}\left(g(\theta) \in C^{\gamma}\left(\underline{X}\right)\right) &= \mathbb{P}_{\theta}\left(q_{\gamma} \leq T\left(\underline{X}, \eta\right) \leq q_{\gamma+1-\alpha}\right) \\ &= F\left(q_{\gamma+1-\alpha}\right) - F\left(q_{\gamma}\right) = 1 - \alpha \end{aligned}$$
$\square$

How to use pivotal functions

Let $T(\underline{X}, \eta)$ be a pivotal function and $\alpha \in]0, 1[$.

Proposition

Assume that the cdf F of $T(\underline{X}, \eta)$ is continuous and strictly increasing, and denote by $q_r = F^{-1}(r)$ the quantile of order r.

Then, for all $\gamma \in [0, \alpha]$:

$$\begin{aligned} C^{\gamma}(\underline{X}) &= \{\eta \in N \text{ such that } q_{\gamma} \leq T(\underline{X}, \eta) \leq q_{\gamma+1-\alpha}\} \\ &= T^{-1}(\underline{X}, [q_{\gamma}, q_{\gamma+1-\alpha}]) \end{aligned}$$

is a confidence interval for η with level exactly $1 - \alpha$.

Proof.

$$\begin{aligned} \mathbb{P}_{\theta}(g(\theta) \in C^{\gamma}(\underline{X})) &= \mathbb{P}_{\theta}(q_{\gamma} \leq T(\underline{X}, \eta) \leq q_{\gamma+1-\alpha}) \\ &= F(q_{\gamma+1-\alpha}) - F(q_{\gamma}) = 1 - \alpha \end{aligned}$$

$\square$

Example: $\mathscr{N}(\mu, \sigma_0^2)$ n-sample, with known σ_0^2

Consider once more the pivotal function

$$T(\underline{X}, \mu) = \sqrt{n}\, \frac{(\bar{X} - \mu)}{\sigma_0} \;\sim\; \mathscr{N}(0, 1).$$

For all $\gamma \leq \alpha$, we obtain a CI with level (exactly) $1 - \alpha$:

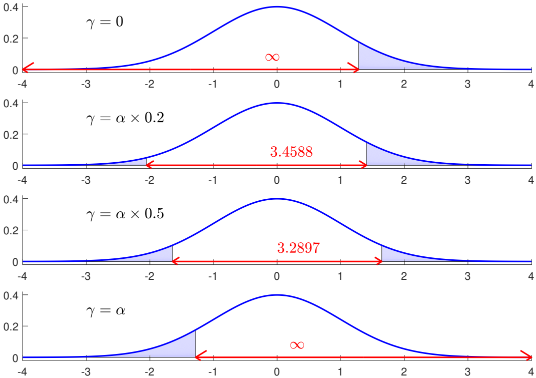
$$C^\gamma = \left[\bar{X} - \frac{\sigma_0}{\sqrt{n}}\, q_{1-\alpha+\gamma}, \quad \bar{X} - \frac{\sigma_0}{\sqrt{n}}\, q_\gamma \right],$$

with q_r the quantile of order r of the $\mathscr{N}(0, 1)$ distribution.

For instance, with $\gamma = \frac{\alpha}{2}$ and $\alpha = 0.05$:

$$\begin{aligned} -q_{1-\alpha+\gamma} &= -q_{0.975} \approx -1.96 \\ -q_\gamma &= -q_{0.025} \approx +1.96 \end{aligned}$$

How to choose γ ?



Density of the $\mathcal{N}(0,1)$ distribution and corresponding quantiles for $\alpha = 0.1$ and several values of γ (in red: $q_{\gamma+1-\alpha} - q_\gamma$).

Usual criterion: value s.t. the CI has minimal length (here $\gamma = \dfrac{\alpha}{2}$).

Example: component reliability (cont'd)

It can be proved that:

$$T\left(\underline{X}, \eta\right) = \frac{\bar{X}}{\eta} \sim \Gamma\left(n, n\right).$$

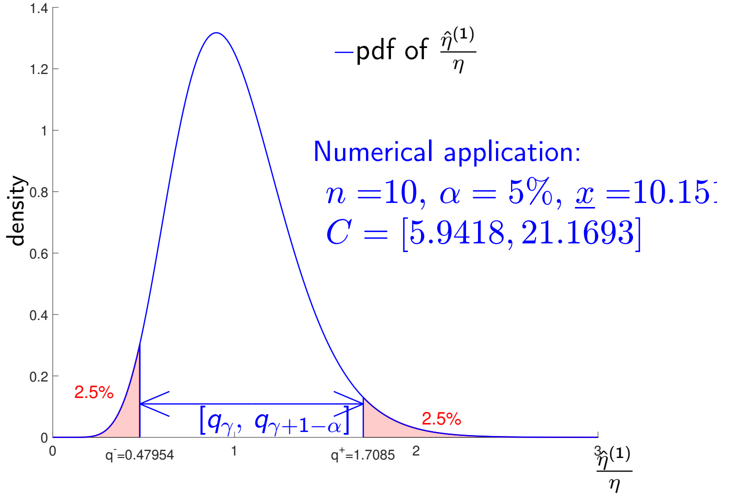
Thus, a CI with level exactly $1-\alpha$ is :

$$C^{\gamma} = \left[\frac{\bar{X}}{q_{\gamma+1-\alpha}}, \frac{\bar{X}}{q_{\gamma}}\right],$$

with q_r the quantile of order r of the $\Gamma(n, n)$ distribution.

Choice of γ: we can take $\gamma = \frac{\alpha}{2}$ for simplicity, or search numerically for the value γ such that the length $1/q_{\gamma} - 1/q_{1+\gamma-\alpha}$ is minimal.

Example: component reliability (cont'd)



Probability density function of the pivotal distribution $\Gamma(n, n)$ and corresponding quantiles for $\alpha = 0.05$ and $\gamma = \frac{\alpha}{2}$.

Lecture outline

Motivation and goal

Problem

It is sometimes (often) difficult to find a pivotal function.

Solution: use once again an asymptotic approach.

- Intervals with "approximate guarantees" will be obtained.
- Computation becomes easier using the previously introduced tools

(CLT, Slutsky, delta method. . .).

⚠ Any analysis carried out in an asymptotic setting is

approximate when n is finite.

➠ The results can be poor for small n. . .

Motivation and goal

Problem

It is sometimes (often) difficult to find a pivotal function.

Solution: use once again an asymptotic approach.

- Intervals with “approximate guarantees” will be obtained.
- Computation becomes easier using the previously introduced tools

  (CLT, Slutsky, delta method...).

⚠ Any analysis carried out in an asymptotic setting is approximate when n is finite.

➠ The results can be poor for small n...

Motivation and goal

Problem

It is sometimes (often) difficult to find a pivotal function.

Solution: use once again an asymptotic approach.

- Intervals with "approximate guarantees" will be obtained.
- Computation becomes easier using the previously introduced tools

  (CLT, Slutsky, delta method...).

⚠ Any analysis carried out in an asymptotic setting is

approximate when n is finite.

➠ The results can be poor for small n...

Asymptotic confidence regions (intervals)

We set $\underline{X}_n = (X_1, \ldots, X_n)$. Recall that $\eta = g(\theta)$ and $N = g(\Theta)$.

Definition: asymptotic confidence region

An asymptotic confidence region with level (at least) $1 - \alpha$ is a statistic $C_n(\underline{X}_n)$, with values in $\mathcal{P}(N)$, such that

$$\forall \theta \in \Theta, \quad \lim_{n \to \infty} \mathbb{P}_\theta \left(g(\theta) \in C_n \left(\underline{X}_n \right) \right) \geq 1 - \alpha.$$

(variant: "exactly" if equality holds for all θ.)

Recall that for an "exact" CR with level (at least) $1 - \alpha$,

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta \left(g(\theta) \in C_n \left(\underline{X}_n \right) \right) \geq 1 - \alpha$$

(here, "exact" means "non asymptotic").

Asymptotic confidence regions (intervals)

How to use **asymptotic pivotal functions**.

Their use is illustrated in:

- the parameter of a Rayleigh distribution ➠ exercise 3

  This is an exercise mixing **definitions** and questions.

- the component reliability example ➠ exercise 4

  It can be proved that

$$C_n = \left[\left(1 - \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right) \bar{X}_n, \left(1 + \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right) \bar{X}_n \right]$$

  is an asymptotic CI with level $1 - \alpha$ for η where q_r the quantile of order r of the $\mathcal{N}(0, 1)$ distribution.

⚠ *The design of asymptotic CI is part of the course (and exam).*

Asymptotic confidence regions (intervals)

How to use **asymptotic pivotal functions**.

Their use is illustrated in:

- the parameter of a Rayleigh distribution exercise 3

  This is an exercise mixing **definitions** and questions.

- the component reliability example exercise 4

  It can be proved that

$$C_n = \left[\left(1 - \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right) \bar{X}_n, \left(1 + \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right) \bar{X}_n \right]$$

  is an asymptotic CI with level $1-\alpha$ for η where q_r the quantile of order r of the $\mathcal{N}(0,1)$ distribution.

⚠ *The design of asymptotic CI is part of the course (and exam).*

Asymptotic confidence regions (intervals)

How to use **asymptotic pivotal functions**.

Their use is illustrated in:

- the parameter of a Rayleigh distribution [exercise 3]

  This is an exercise mixing **definitions** and questions.

- the component reliability example [exercise 4]

  It can be proved that

$$C_n = \left[\left(1 - \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}\right) \bar{X}_n,\ \left(1 + \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}\right) \bar{X}_n \right]$$

  is an asymptotic CI with level $1-\alpha$ for η where q_r the quantile of order r of the $\mathcal{N}(0,1)$ distribution.

⚠ *The design of asymptotic CI is part of the course (and exam).*

Summary and preview

We have seen and will practice in PC 3:

- the tools to establish the convergence in distribution and the convergence rate of a sequence of estimators,
- The use of the (asymptotic) distribution of a sequence of estimators to construct confidence intervals or regions.

We will cover in Lecture 4:

- decision-making through statistical hypothesis testing,
- the construction of such a test,
- the risks associated with this decision.

Summary and preview

We have seen and will practice in PC 3:

- the tools to establish the convergence in distribution and the convergence rate of a sequence of estimators,
- The use of the (asymptotic) distribution of a sequence of estimators to construct confidence intervals or regions.

We will cover in Lecture 4:

- decision-making through statistical hypothesis testing,
- the construction of such a test,
- the risks associated with this decision.

Lecture outline

Lecture outline

Exercise 1 (Estimation of the probability of an event) solution

Let $(X_n)_{n\geq 1}$ be a sequence of IID RV with values in $(\mathcal{X}, \mathscr{A})$.

For a given $A \in \mathscr{A}$, we estimate $\eta = \mathbb{P}(X_1 \in A)$ by:

$$\hat{\eta}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in A}.$$

Question

Study the asymptotic behaviour of $\hat{\eta}_n^{(1)}$.

Exercise 2 (Asymptotic distribution)

solution

Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, with $\theta > 0$.

Let η denote the probability of exceeding a given threshold $x_0 > 0$:

$$\eta = \mathbb{P}_\theta(X \geq x_0) = \exp(-\theta x_0) .$$

Questions

1. Study the asymptotic behaviour of the sample mean $\bar{X}_n$.
2. Propose an estimator $\hat{\eta}_n^{(1)}$ as a function of $\bar{X}_n$, using the substitution method.
3. Study the asymptotic behaviour of $\hat{\eta}_n^{(1)}$.
4. Let $\hat{\eta}_n^{(2)} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \geq x_0}$. Is one of the estimators asymptotically preferable to the other?

Exercise 3 (Rayleigh distribution: asymptotic CI)

solution

This is a long exercise about the concept **of asymptotic confidence interval**.

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathscr{R}\left(\sigma^2\right)$, with $\sigma^2 > 0$.

Rayleigh distribution

Questions ❶-❸ detail how to obtain asymptotic IC using **asymptotic pivotales functions.**

Questions ❹-❺ show how to compute **coverage probability** in the context of asymptotic confidence intervals.

Exercise 3 (Rayleigh distribution: asymptotic CI)

solution

Definition

A (sequence of) function(s)

$$T_n : \mathcal{X}^n \times N \;\to\; \mathbb{R}$$

is an asymptotic pivotal function if the limit distribution of $T_n(\underline{X}_n, \eta)$ does not depend on θ :

$$T_n(\underline{X}_n, \eta) \xrightarrow[n \to \infty]{\mathrm{d}} T_\infty.$$

where T_∞ is a RV whose distribution is free of θ.

Definition given with the lesson notations

➠ For the exercise, $\eta = \theta = \sigma^2$.

Exercise 3 (Rayleigh distribution: asymptotic CI)

solution

How to use asymptotic pivotal functions:

- exactly as we used the non-asymptotic ones !
- The obtained intervals are asymptotic confidence intervals.

Questions

1. Give the asymptotic distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. Using the asymptotic distribution of $\bar{X}_n$, propose an asymptotic pivotal functions,
3. Give a confidence interval for η with level exactly $1-\alpha$.

Exercise 3 (Rayleigh distribution: asymptotic CI)

solution

Reminder : the **coverage probability** of a CI is its **true** level.

Computing the coverage probability of $C_n(\underline{X}_n)$ requires the use of **Cumulative distribution function (CDF) of** T_n.

Here, T_n depends on $\bar{X}_n$ whose distribution is not a standard one.
➠ The CDF can however be numerically computed..

Questions

4. Show that $\dfrac{1}{\sigma}\sum_{i=1}^{n} X_i \sim \mathcal{SR}(n,1)$
   where $F^{(n)}$ is the CDF of the $\mathcal{SR}(n,1)$ distribution.
   Sum of Rayleigh distributions.
5. Give the coverage probability of $C_n(\underline{X}_n)$ as a function of $F^{(n)}$.

Exercise 4 (Asymptotic CI for the Component reliability application)

solution

Reminder about "Component reliability"

➠ $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$ et $\eta = \dfrac{1}{\theta}$

Questions

1. Show that
$$T_n = \sqrt{n}\,\frac{(\bar{X}_n - \eta)}{\bar{X}_n} \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}(0, 1).$$
is an asymtotically pivotal function (voir exercise 3 for a definition of this term).
2. Use this pivotal function to design an asymptotic CI with level $1 - \alpha$.
3. Determine the cov. prob. of the obtained asymptotic CI.

back to slide 38

Lecture outline

Solution of exercise 1

back to questions

Using the CLT with $Y_i = \mathbb{1}_{X_i \in A} \overset{\text{iid}}{\sim} \mathrm{Ber}(\eta)$:

$$\sqrt{n}\left(\hat{\eta}_n - \eta\right) \xrightarrow[n \to \infty]{\text{d}} \mathcal{N}\left(0, \eta(1-\eta)\right).$$

Concl.: if $0 < \eta < 1$, then $\hat{\eta}_n$ is asymptotically Gaussian, with

- convergence rate: $\frac{1}{\sqrt{n}}$,
- asymptotic variance: $\eta(1-\eta)$.

Solution of exercise 1

back to questions

Using the CLT with $Y_i = \mathbb{1}_{X_i \in A} \overset{\text{iid}}{\sim} \mathrm{Ber}(\eta)$:

$$\sqrt{n}\left(\hat{\eta}_n - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta(1-\eta)\right).$$

Concl.: if $0 < \eta < 1$, then $\hat{\eta}_n$ is asymptotically Gaussian, with

- convergence rate: $\frac{1}{\sqrt{n}}$,
- asymptotic variance: $\eta(1-\eta)$.

Using the CLT with $Y_i = \mathbb{1}_{X_i \in A} \overset{\text{iid}}{\sim} \mathrm{Ber}(\eta)$:

$$\sqrt{n}\left(\hat{\eta}_n - \eta\right) \xrightarrow[n\to\infty]{\text{d}} \mathcal{N}\left(0, \eta(1-\eta)\right).$$

Concl.: if $0 < \eta < 1$, then $\hat{\eta}_n$ is asymptotically Gaussian, with

- convergence rate: $\frac{1}{\sqrt{n}}$,
- asymptotic variance: $\eta(1-\eta)$.

back to questions

❶ Using CLT:

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\theta}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)$$

❷ $\eta = \exp\left(-\frac{x_0}{\frac{1}{\theta}}\right) = h\left(\frac{1}{\theta}\right)$

with $h : u \mapsto \exp\left(-\frac{x_0}{u}\right)$ continuous on $\mathbb{R}_+^*$.

Applying method of moments with $\bar{X}_n$ estimator of $\frac{1}{\theta}$:

$$\hat{\eta}_n^{(1)} = h\left(\bar{X}_n\right) = \exp\left(-\frac{x_0}{\bar{X}_n}\right)$$

❸ h is differentiable on $\mathbb{R}_+^*$ with $h'(u) = \frac{x_0}{u^2} \exp\left(-\frac{x_0}{u}\right)$.
Using the Delta-theorem in the Gaussian cae, we get:

$$\sqrt{n}\left(h\left(\bar{X}_n\right) - h\left(\frac{1}{\theta}\right)\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, h'\left(\frac{1}{\theta}\right)^2 \frac{1}{\theta^2}\right)$$

Let:

$$\sqrt{n}\left(\hat{\eta}_n^{(1)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \left(x_0\theta \exp\left(-\theta x_0\right)\right)^2\right)$$

The asymptotic variance of $\hat{\eta}_n^{(1)}$ is $\sigma_1^2(\theta) = \left(x_0\theta \exp\left(-\theta x_0\right)\right)^2$.

Solution of exercise 2

back to questions

❹ $\hat{\eta}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ with $Z_i = \mathbb{1}_{X_i \geq x_0} \implies \begin{cases} Z_1, \ldots, Z_n \text{ IID} \\ Z_1 \sim \mathrm{Ber}(\eta) \end{cases}$

Using the result of exercise 1:

Exercise 1

$$\sqrt{n}\left(\hat{\eta}_n^{(2)} - \eta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta(1-\eta)\right)$$

with $\eta = \exp\left(-\theta x_0\right)$, we obtain the asymptotic variance:

$$\sigma_2^2(\theta) = \exp\left(-\theta x_0\right)\left(1 - \exp\left(-\theta x_0\right)\right).$$

Solution of exercise 2

back to questions

Let $\Delta(\theta) = \sigma_2^2(\theta) - \sigma_1^2(\theta)$.

$$\begin{aligned}\Delta(\theta) &= \exp\left(-\theta x_0\right)\left(1 - \exp\left(-\theta x_0\right) - x_0^2\theta^2 \exp\left(-\theta x_0\right)\right) \\ &= \exp\left(-\theta x_0\right)\varphi(\theta x_0)\end{aligned}$$

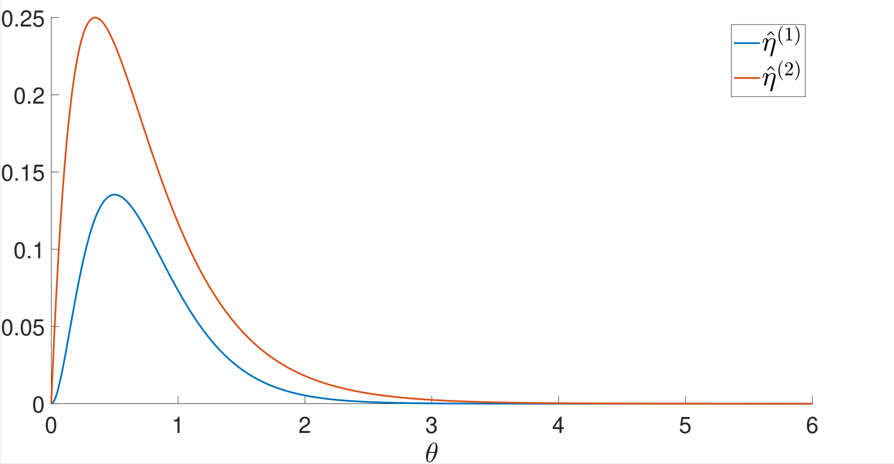
with $\varphi(u) = 1 - \exp(-u)(1 + u^2)$.

Analyzing the sign of the derivative of φ leads to $\varphi > 0$ on $\mathbb{R}_+$.

$\hat{\eta}_n^{(1)}$ is asymptotically preferable to $\hat{\eta}_n^{(2)}$.

Solution of exercise 2

back to questions



Plots of bothasymptotic variances for $x_0 = 2.0$.

❶ Direct application of the CLT:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \sigma\sqrt{\frac{\pi}{2}}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma^2\left(2-\frac{\pi}{2}\right)\right).$$

$$\sqrt{n}\left(\frac{\frac{\bar{X}_n}{\sigma} - \sqrt{\frac{\pi}{2}}}{\sqrt{2-\frac{\pi}{2}}}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1).$$

❷ Thus:

$T_n = \sqrt{n}\left(\frac{\frac{\bar{X}_n}{\sigma} - \sqrt{\frac{\pi}{2}}}{\sqrt{2-\frac{\pi}{2}}}\right)$ is an asymptotic pivotal function.

back to questions

❸ Since $T_n \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$, we have with asymptotic probability $1-\alpha$ that:

$$-q_{1-\frac{\alpha}{2}} \leq \sqrt{n}\left(\frac{\frac{\bar{X}_n}{\sigma} - \sqrt{\frac{\pi}{2}}}{\sqrt{2-\frac{\pi}{2}}}\right) \leq q_{1-\frac{\alpha}{2}}$$

with q_r the quantile of order r of the $\mathcal{N}(0,1)$ distribution.

It comes the asymptotic confidence intervals with level $1-\alpha$:

$$\sqrt{\frac{2}{\pi}}\bar{X}_n \frac{1}{1+\frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\frac{4}{\pi}-1}} \leq \sigma \leq \sqrt{\frac{2}{\pi}}\bar{X}_n \frac{1}{1-\frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\frac{4}{\pi}-1}}$$

back to questions

The asymptotic CI can be simplified with a Taylor approximation:

Asymptotic CI with level (exactly) $1-\alpha$ for σ

$$C_n = \sqrt{\frac{2}{\pi}}\bar{X}_n \left[1 - \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\frac{4}{\pi} - 1},\ 1 + \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\frac{4}{\pi} - 1}\right]$$

Solution of exercise 3

back to questions

❹ As $X_i \sim \mathcal{R}(\sigma^2)$. σ is a scale parameter : $\dfrac{X_i}{\sigma} \sim \mathcal{R}(1)$.

As X_i are IID, it comes $\dfrac{1}{\sigma}\sum_{i=1}^{n} X_i \sim \mathcal{SR}(n, 1)$.

❺ Coverage probability of $I_n\left(\underline{X}_n\right)$

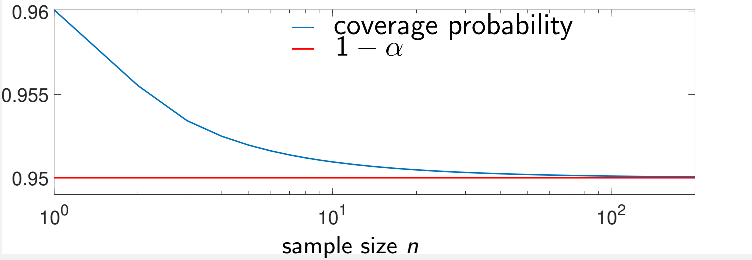
$$\begin{aligned}
\mathbb{P}_\theta\left(\sigma \in C_n\left(\underline{X}_n\right)\right) &= \mathbb{P}_\theta\left(a_n \leq \frac{1}{\sigma}\sum_{i=1}^{n} X_i \leq b_n\right) \\
&= F^{(n)}(b_n) - F^{(n)}(a_n)
\end{aligned}$$

$$\text{with } \left\{\begin{aligned}
a_n &= n\sqrt{\frac{\pi}{2}} - \sqrt{n}\sqrt{2 - \frac{\pi}{2}}\, q_{1-\frac{\alpha}{2}} \\
b_n &= n\sqrt{\frac{\pi}{2}} + \sqrt{n}\sqrt{2 - \frac{\pi}{2}}\, q_{1-\frac{\alpha}{2}}
\end{aligned}\right.$$

Remark. Here the coverage probability does not depend on θ. This is a special case because σ is a scale parameter.

Solution of exercise 3

back to questions



Coverage probability of the asymptotic CI $C_n(\underline{X}_n)$ with $\alpha = 5\%$.

Remark. Observe that we have indeed a confidence interval with asymptotic level (exactly) $1-\alpha$:

$$\forall \theta, \quad \lim_{n \to \infty} \mathbb{P}_\theta\left(\sigma \in C_n\left(\underline{X}_n\right)\right) = 1 - \alpha.$$

Solution of exercise 4

❶ It has already been shown (CLT, Slutski, Mann-Wald) that

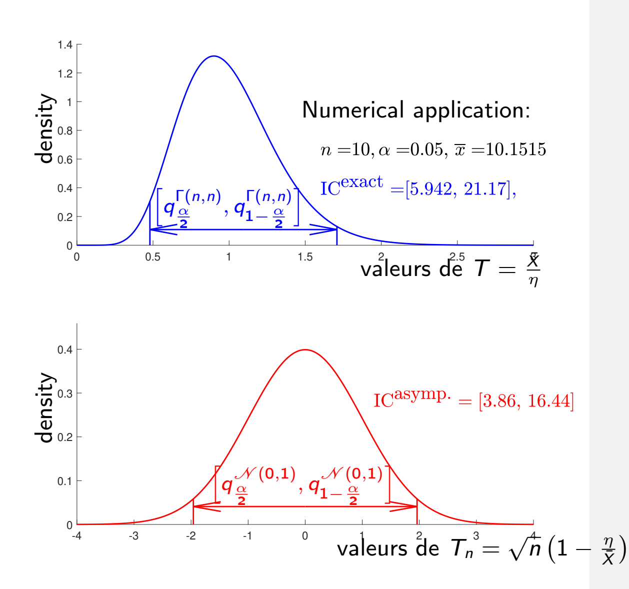
$$T_n\left(\underline{X}_n, \eta\right) = \sqrt{n}\,\frac{\left(\bar{X}_n - \eta\right)}{\bar{X}_n} \xrightarrow[n\to\infty]{\text{d}} \sim \mathcal{N}\left(0,1\right),$$

➠ Thus, T_n is an asymptotic pivotal function.

❷ Asymptotic CI with level (exactly) $1-\alpha$ for η :

$$C_n = \left[\left(1 - \frac{1}{\sqrt{n}}\,q_{1-\frac{\alpha}{2}}\right)\bar{X}_n,\ \left(1 + \frac{1}{\sqrt{n}}\,q_{1-\frac{\alpha}{2}}\right)\bar{X}_n\right]$$

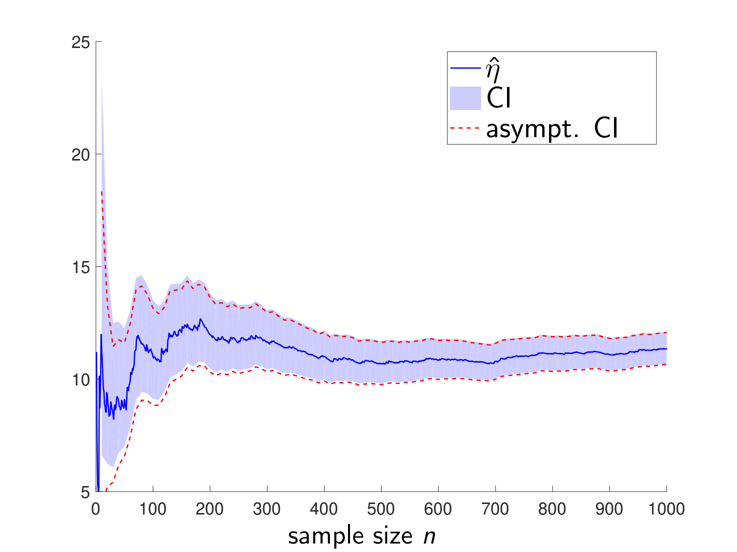
with q_r the quantile of order r of the $\mathcal{N}(0,1)$ distribution.



⚠ Do not confuse intervals on pivotal functions $\left[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}\right]$ and confidence interval for η.

Solution of exercise 4

back to questions



Comparison of exact and asymptotic CIs, as a function of n

Solution of exercise 4

back to questions

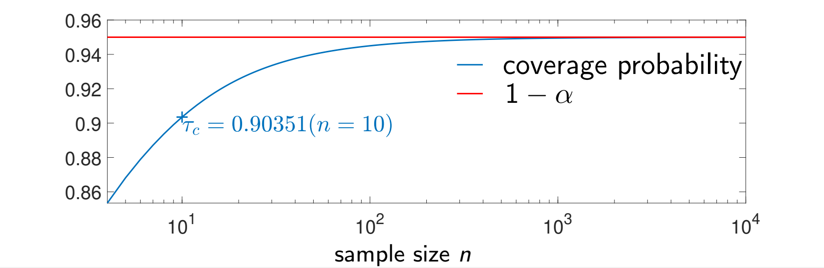
❸ Coverage probability of $I_n\left(\underline{X}_n\right)$

$$\begin{aligned}
\mathbb{P}_\theta\left(\eta \in C_n\left(\underline{X}_n\right)\right) &= \mathbb{P}_\theta\left(\eta \in \left[\left(1 - \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}\right), \left(1 + \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}\right)\right] \bar{X}_n\right) \\
&= \mathbb{P}_\theta\left(\frac{1}{1 + \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}} \leq \frac{\bar{X}_n}{\eta} \leq \frac{1}{1 - \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}}\right)
\end{aligned}$$

Since (reminder) $\dfrac{\bar{X}_n}{\eta} \sim \Gamma\left(n, n\right)$, it comes:

$$\mathbb{P}_\theta\left(\eta \in C_n\left(\underline{X}_n\right)\right) = F^{\Gamma(n,n)}\left(\frac{1}{1 - \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}}\right) - F^{\Gamma(n,n)}\left(\frac{1}{1 + \frac{1}{\sqrt{n}}\, q_{1-\frac{\alpha}{2}}}\right)$$

with $F^{\Gamma(n,n)}$ the cdf of the $\Gamma(n, n)$ distribution.



Coverage probability of the asympt. CI with level 95%

Remarks.

- The property $\forall\theta,\ \lim_{n\to\infty} \tau^c_{n,\theta}\left(C_n(\underline{X}_n)\right) \geq 1-\alpha$ is verified.
- Usually the coverage probability depends on θ. It is not the case here because η is a scale parameter.

Lecture outline

Proof

As Z has a moment of order 2, we can define:

- its mean $\mu = \mathbb{E}(Z)$,
- its covariance matrix $\Sigma_Z = \mathbb{E}((Z-\mu)(Z-\mu)^\top)$.

We start by noting that if it exists $a \in \mathbb{R}^q \setminus \{0\}$ and $c \in \mathbb{R}$ s.t. $a^\top Z = c$; a.s., then $c = a^\top \mu$.

An intermediate result

Let V be a **positive** scalar random variable. We have :

$$\mathbb{E}(V) = 0 \iff V = 0 \text{ a.s.} \qquad (*)$$

Proof (cont'd)

Let $a \in \mathbb{R}^q \setminus \{0\}$ et $c \in \mathbb{R}$.

$$
\begin{aligned}
a^\top Z = c \text{ a.s.} \quad & \iff \quad a^\top (Z-\mu) = 0 \text{ a.s.} \\
& \iff \quad a^\top (Z-\mu)(Z-\mu)^\top a = 0 \text{ a.s.} \\
& \iff \quad \mathbb{E}(a^\top (Z-\mu)(Z-\mu)^\top a) = 0 \text{ (utilisant } (*)) \\
& \iff \quad a^\top \Sigma_Z a = 0
\end{aligned}
$$

As the matrix Σ_Z is positive-definite, $a^\top \Sigma_Z a = 0$ (with $a \neq 0$) is equivalent to $a \in \mathrm{Ker}(\Sigma_Z)$.
Thus,

$$
\begin{aligned}
Z \text{ dégénérée} \quad & \iff \quad \exists a \neq 0 \text{ t.q. } a^\top Z = c \text{ a.s.} \\
& \iff \quad \exists a \neq 0 \in \mathrm{Ker}(\Sigma_Z) \\
& \iff \quad \Sigma_Z \text{ non inversible}
\end{aligned}
$$

$\square$

back to slide 7

Relation between convergence in distribution and in proba.

We already know that convergence in probability implies convergence in distribution. Let $(Y_n)_{n\in\mathbb{N}^*}$ be a sequence of RV with values in $\mathbb{R}^d$.

Proposition

If $Y_n \xrightarrow{\mathrm{d}} c$, with $c \in \mathbb{R}^d$ a constant, then $Y_n \xrightarrow{\mathbb{P}} c$.

Corollary

If there exists $c \in \mathbb{R}^d$,

- a RV Z with values in $\mathbb{R}^d$,
- a sequence $(a_n)_{n\in\mathbb{N}^*}$ of real numbers such that $\lim\limits_{n\to\infty} a_n = \infty$,

such that

$$a_n\,(Y_n - c) \xrightarrow[n\to\infty]{\mathrm{d}} Z$$

then

$$Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c.$$

Proof (exercise): use above proposition and Slutsky's theorem (see below). □

back to slide 8

Proof "Delta-theorem" (scalar case)

Consider the function ψ defined by :

$$\psi(y) = \begin{cases} \dfrac{h(y) - h(m)}{y - m} & \text{si } y \neq m, \\ h'(m) & \text{si } y = m; \end{cases}$$

ψ is continuous at m because h est differentiable at m. Since $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} m$,

$$\psi(Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} \psi(m) = h'(m),$$

and thus (Slutsky)

$$\left(\sqrt{n}(Y_n - m), \psi(Y_n)\right) \xrightarrow[n\to\infty]{\mathrm{d}} \left(Z, h'(m)\right).$$

Finally, we have

$$\sqrt{n}\,(h(Y_n) - h(m)) = \sqrt{n}\,(Y_n - m)\,\psi(Y_n) \xrightarrow[n\to\infty]{\mathrm{d}} h'(m)\,Z. \quad \square$$

back to slide 13

Regular models: regularity conditions C_3 and C_4

Reminder: C_0, C_1 and C_2 were defined in Lecture #2.

Regularity condition C_3

$\theta \mapsto f_\theta(\underline{x})$ is twice continuously differentiable for ν-almost all $\underline{x}$.

Regularity condition C_4

At any point $\theta \in \Theta$, we have

$$\int_{\mathcal{S}} \nabla_\theta \nabla_\theta^\top f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = \nabla_\theta \int_{\mathcal{S}} \nabla_\theta^\top f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}).$$

In other words: $\forall \theta \in \Theta$, $\forall k \leq p$, $\forall j \leq p$,

$$\int_{\mathcal{S}} \frac{\partial^2 f_\theta(\underline{x})}{\partial \theta_k \partial \theta_j}\, \nu(\mathrm{d}\underline{x}) = \frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} \frac{\partial f_\theta(\underline{x})}{\partial \theta_j}\, \nu(\mathrm{d}\underline{x}).$$

back to slide 20

Example: an MLE that is not asymptotically Gaussian

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{U}_{[0,\theta]}$, with $\theta > 0$ unknown.

⚠ This model is not regular (why?).

It can be proved that (cf. PC 1, exercise 1.2)

- $\hat{\theta}_n = \max_{i \leq n} X_i$ is the MLE of θ, and
- $n\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\text{d}} -Z$ with $Z \sim \mathcal{E}\left(\lambda = \dfrac{1}{\theta}\right)$.

In this particular case

- the MLE is not asymptotically Gaussian;
- the convergence rate is $\frac{1}{n}$: faster than $\frac{1}{\sqrt{n}}$.

back to slide 20

The Rayleigh $\mathscr{R}\left(\sigma^2\right)$ distribution

$X \sim \mathscr{R}\left(\sigma^2\right)$ if X admits the pdf

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathbb{1}_{\mathbb{R}^+}(x).$$

Moments

- mean : $\mathbb{E}_\sigma(X) = \sigma\sqrt{\frac{\pi}{2}}$
- variance : $\mathrm{var}_\sigma(X) = \sigma^2\left(2 - \frac{\pi}{2}\right)$

Property

if $X \sim \mathscr{R}(\sigma^2)$ then $Y = X^2 \sim \mathcal{E}\left(\frac{1}{2\sigma^2}\right)$.

Sum of Rayleigh distributions

We define (for the exercise) the following distribution:

$$\text{If } (X_1, \ldots, X_n) \overset{\text{iid}}{\sim} \mathcal{R}(\sigma^2), \text{ then } Z = \sum^{n} X_i \sim \mathcal{SR}(n, \sigma^2).$$