

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

1/69

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)

2/69

Lecture outline

- 1 – Convergence rate and asymptotic distribution
- 2 – Confidence regions and confidence intervals
- 3 – Standard exercises (with solutions)
- 4 – Appendices

3/69

Recap: Mathematical framework

In this section:

- ▶ We consider a **statistical model**

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

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

- ▶ $X_1, X_2, \dots \stackrel{\text{iid}}{\sim} \mathbb{P}_{\theta}$, defined on a common $(\Omega, \mathcal{F}, \mathbb{P}_{\theta})$.
- ▶ 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, $\eta = g(\theta) \in \mathbb{R}^q$.

4/69

Lecture outline

1 – Convergence rate and asymptotic distribution

1.1 – Definitions and examples

1.2 – Theoretical tools

1.3 – Asymptotic efficiency

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

Lecture outline

1 – Convergence rate and asymptotic distribution

1.1 – Definitions and examples

1.2 – Theoretical tools

1.3 – Asymptotic efficiency

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

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 $\Sigma = \text{var}(X_1) \in \mathbb{R}^{d \times d}$.

Then :

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \Sigma),$$

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

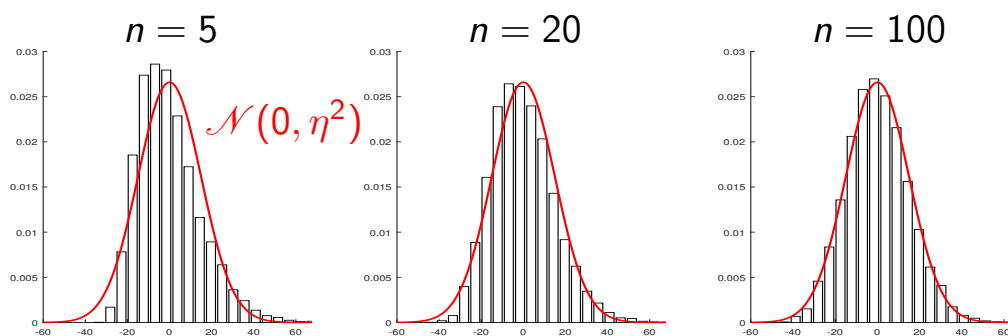
5/69

Example: component reliability

Recall that

- ▶ $X_i \stackrel{\text{iid}}{\sim} \mathcal{E}(\theta)$, $\theta > 0$, 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.

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



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

6/69

Convergence rate

Let $\hat{\eta}_n = \hat{\eta}_n(X_1, \dots, X_n)$ 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:

- ▶ $\lim_{n \rightarrow \infty} a_n = \infty$,
 - ▶ $a_n (\hat{\eta}_n - \eta) \xrightarrow[n \rightarrow \infty]{d} Z$,
 - ▶ 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

7/69

Asymptotic normality

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

Definition

If there exists

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

such that

$$a_n (\hat{\eta}_n - \eta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \Sigma(\theta)), \quad (1)$$

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 $a_n \rightarrow +\infty$ implies (weak) **consistency**.

Convergence in distribution to a constant

8/69

Lecture outline

1 – Convergence rate and asymptotic distribution

1.1 – Definitions and examples

1.2 – Theoretical tools

1.3 – Asymptotic efficiency

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

The continuous mapping theorem

Theorem (Mann-Wald)

Let

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

such that

h is continuous at the point Y , almost surely.

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

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

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

Example: component reliability (cont'd)

Recall that

- ▶ $X_i \stackrel{\text{iid}}{\sim} \mathcal{E}(\theta)$, $\theta > 0$, 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.

10/69

Slutsky's theorem

Theorem

Let

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

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

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

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

Then

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

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

11/69

Example: component reliability (cont'd)

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

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

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

Therefore, by the continuous mapping theorem,

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

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

12/69

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.

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

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

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

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

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

 Proof

13/69

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.

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

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

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

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

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

$$(Dh)(m) = \left((\partial_j h_i)(m) \right)_{1 \leq i \leq q, 1 \leq j \leq d}.$$

14/69

Delta theorem in the Gaussian case

Scalar case.

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

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

Vector case

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

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

15/69

Example: component reliability (cont'd)

We already saw that:

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

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

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

Conclusion: $\hat{\theta}_n$ is **asymptotically Gaussian**,
and its **convergence rate** is $\frac{1}{\sqrt{n}}$.

16/69

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): $\sqrt{n}(\hat{\eta}^{(1)} - \eta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \eta^2)$.

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

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

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

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

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

17/69

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

In summary:

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

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 $\hat{\eta}^{(2)}$.

18/69

Lecture outline

1 – Convergence rate and asymptotic distribution

1.1 – Definitions and examples

1.2 – Theoretical tools

1.3 – Asymptotic efficiency

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

Asymptotic efficiency

Recall the Cramér-Rao lower bound (scalar parameter)

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

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

with $I_1(\theta) = \text{var}_{\theta}(S_{\theta}(X_1))$.

⇒ 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

19/69

Asymptotic efficiency of the MLE

Context: $X_1, X_2, \dots \stackrel{\text{iid}}{\sim} P_{\theta}$ 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** :

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

Fisher information in regular models

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

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

Proposition: another expression for the FIM

In a regular model, we have

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

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.

21/69

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 $\mathbb{E}_\theta (S_\theta(X_1)^2)$

$$I_1(\theta) = \text{var}_\theta(X_1) = \eta^2 = \frac{1}{\theta^2}$$

Comput. of $-\mathbb{E}_\theta \left(\frac{\partial S_\theta}{\partial \theta}(X_1) \right)$

$$I_1(\theta) = -\mathbb{E}_\theta \left(-\frac{1}{\theta^2} \right) = \frac{1}{\theta^2}$$

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

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

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

22/69

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

2.1 – Definition and example

2.2 – Exact confidence intervals

2.3 – Asymptotic confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

2.1 – Definition and example

2.2 – Exact confidence intervals

2.3 – Asymptotic confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

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 $\hat{\eta}$,

a **confidence interval** for η .

23/69

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:

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

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

24/69

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:

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

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

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

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

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

25/69

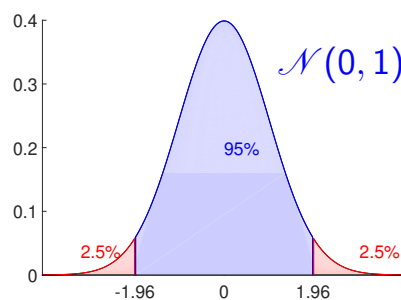
Example: $\mathcal{N}(\mu, \sigma_0^2)$ n -sample, with known σ_0^2

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



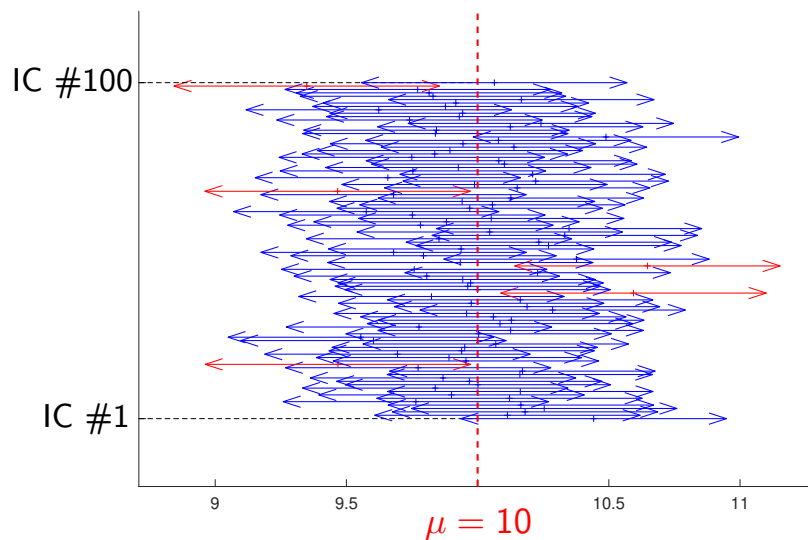
We “pivot” to obtain a **CI with level exactly 95%** :

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

26/69

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 .

27/69

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

2.1 – Definition and example

2.2 – Exact confidence intervals

2.3 – Asymptotic confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

Pivotal functions

The method can be formalized using **pivotal functions**.

Definitions

A function

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

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**: $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma_0^2)$ 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 $T = \sqrt{n} (\bar{X}_n - \mu) \sim \mathcal{N}(0, \sigma_0^2)$.

28/69

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:

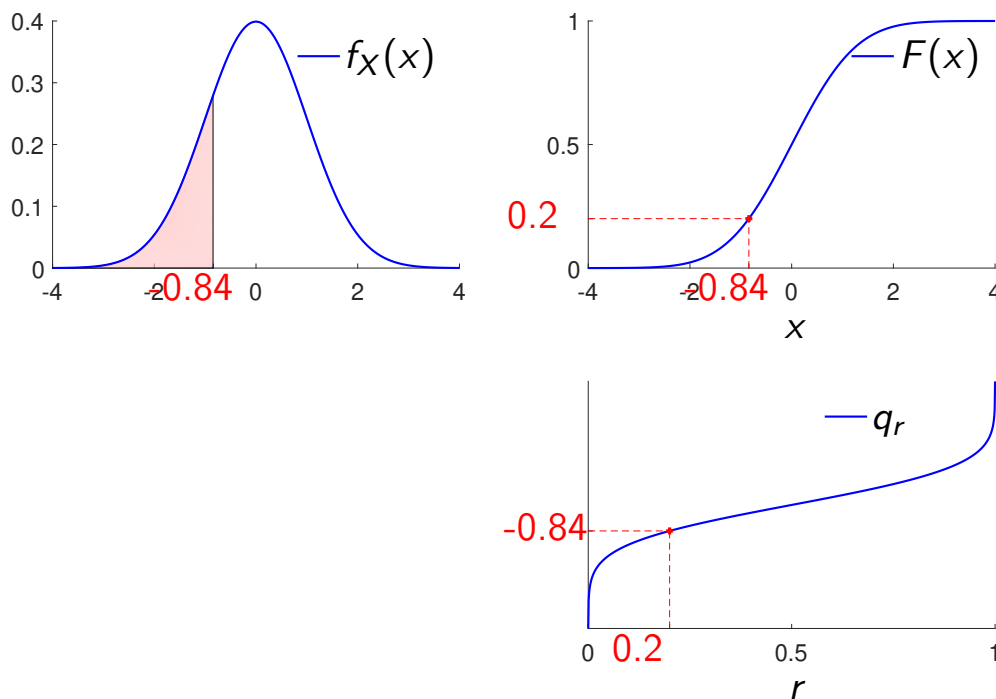
$$q_r = \inf \{x \in \mathbb{R}, F(x) \geq r\} = \min \{x \in \mathbb{R}, F(x) \geq r\}.$$

Properties:

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

29/69

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



30/69

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

□

31/69

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

Consider once more the pivotal function

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

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 $\mathcal{N}(0, 1)$ distribution.

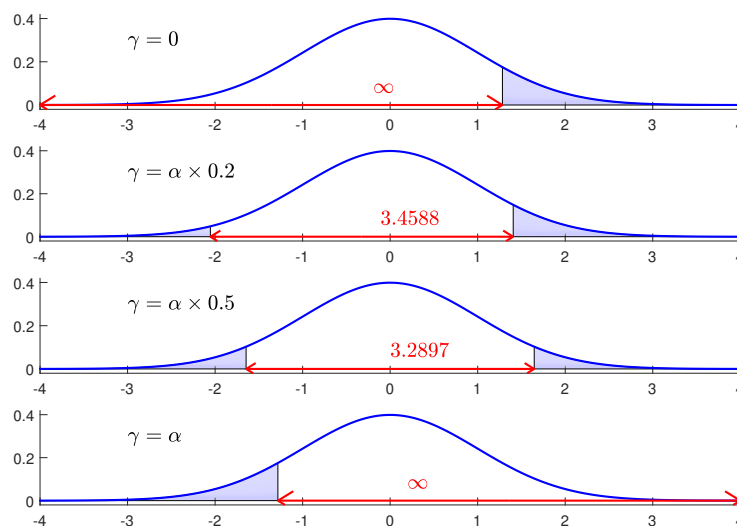
For instance, with $\gamma = \frac{\alpha}{2}$ and $\alpha = 0.05$:

$$-q_{1-\alpha+\gamma} = -q_{0.975} \approx -1.96$$

$$-q_\gamma = -q_{0.025} \approx +1.96$$

32/69

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 $\gamma = \frac{\alpha}{2}$).

33/69

Example: component reliability (cont'd)

It can be proved that:

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

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

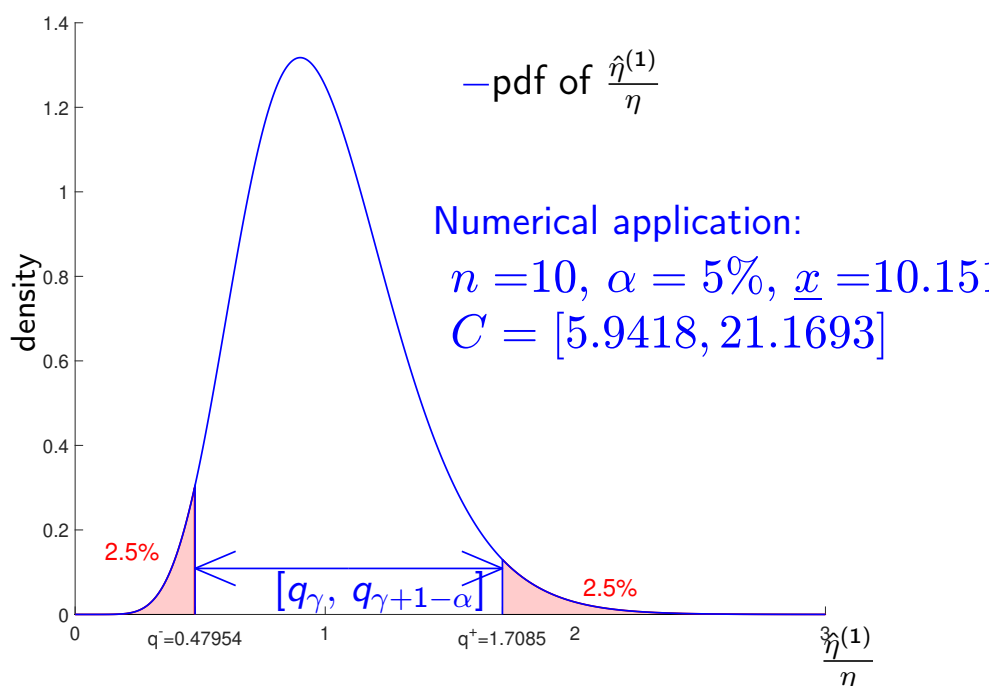
$$C^\gamma = \left[\frac{\bar{X}}{q_{\gamma+1-\alpha}}, \frac{\bar{X}}{q_\gamma} \right],$$

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.

34/69

Example: component reliability (cont'd)



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

35/69

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

2.1 – Definition and example

2.2 – Exact confidence intervals

2.3 – Asymptotic confidence intervals

3 – Standard exercises (with solutions)

4 – Appendices

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 $\underline{X}_n = (X_1, \dots, X_n)$. 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

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

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

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

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

(here, “exact” means “non asymptotic”).

37/69

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).

38/69

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.

39/69

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

3.1 – Questions

3.2 – Solutions

4 – Appendices

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

3.1 – Questions

3.2 – Solutions

4 – Appendices

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 $(\mathcal{X}, \mathcal{A})$.

For a given $A \in \mathcal{A}$, 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 $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{E}(\theta)$, 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?

41/69

Exercise 3 (Rayleigh distribution: asymptotic CI)

 solution

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

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{R}(\sigma^2)$, with $\sigma^2 > 0$.

 Rayleigh distribution

Questions ①-③ detail how to obtain asymptotic IC using **asymptotic pivotale functions**.

Questions ④-⑤ show how to compute **coverage probability** in the context of asymptotic confidence intervals.

42/69

Exercise 3 (Rayleigh distribution: asymptotic CI)

 solution

Definition

A (sequence of) function(s)

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

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

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

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

Definition given with the lesson notations

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

43/69

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 $1 - \alpha$.

44/69

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 $\frac{1}{\sigma} \sum_{i=1}^n X_i \sim \mathcal{SR}(n, 1)$

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 $F^{(n)}$.

45/69

Exercise 4 (Asymptotic CI for the Component reliability application)

▶ solution

Reminder about “Component reliability”

▶ $(X_1, \dots, X_n) \stackrel{\text{iid}}{\sim} \mathcal{E}(\theta)$ et $\eta = \frac{1}{\theta}$

Questions

- 1 Show that

$$T_n = \sqrt{n} \frac{(\bar{X}_n - \eta)}{\bar{X}_n} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

is an asymptotically 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

46/69

Lecture outline

1 – Convergence rate and asymptotic distribution

2 – Confidence regions and confidence intervals

3 – Standard exercises (with solutions)

3.1 – Questions

3.2 – Solutions

4 – Appendices

Solution of exercise 1

[back to questions](#)

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

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

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

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

❶ Using CLT:

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

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 $\frac{1}{\theta}$:

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

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 case, we get:

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

Let:

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

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

$$\textcircled{4} \hat{\eta}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n Z_i \text{ with } Z_i = \mathbb{1}_{X_i \geq x_0} \implies \begin{cases} Z_1, \dots, Z_n \text{ IID} \\ Z_1 \sim \text{Ber}(\eta) \end{cases}$$

Using the result of exercise 1:

[Exercise 1](#)

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

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

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

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

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

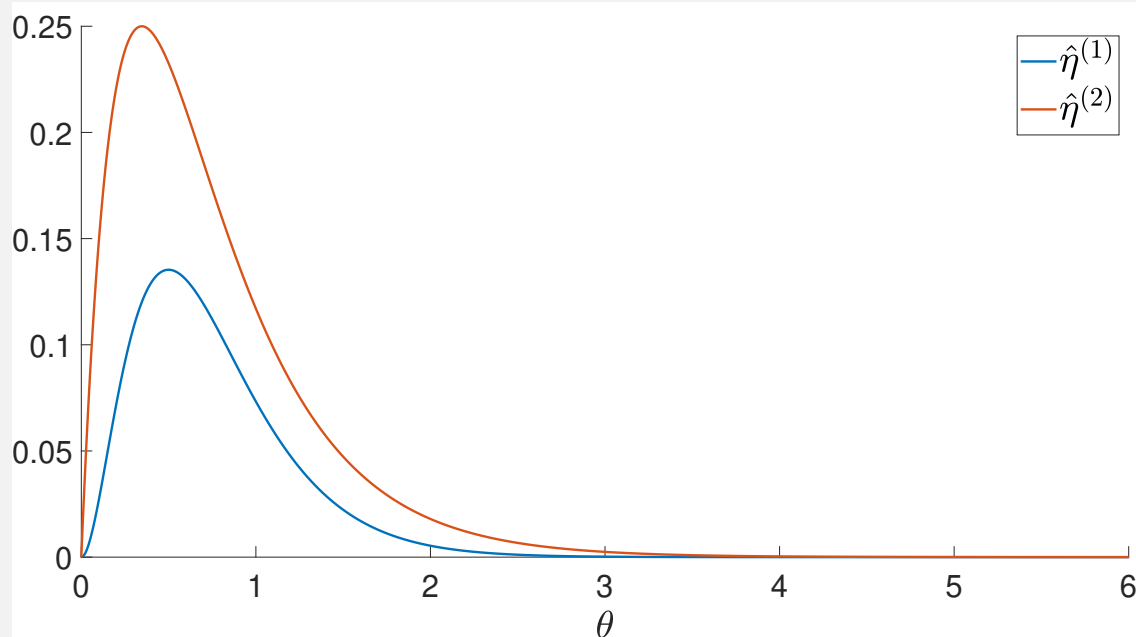
with $\varphi(u) = 1 - \exp(-u)(1 + u^2)$.

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 both asymptotic variances for $x_0 = 2.0$.

52/69

Solution of exercise 3

[back to questions](#)

❶ Direct application of the CLT:

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

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

❷ Thus:

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

53/69

Solution of exercise 3

[back to questions](#)

③ Since $T_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)$, 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 $1 - \alpha$:

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

54/69

Solution of exercise 3

[back to questions](#)

The asymptotic CI can be simplified with a Taylor approximation:

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

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

55/69

Solution of exercise 3

[back to questions](#)

④ As $X_i \sim \mathcal{R}(\sigma^2)$. σ is a scale parameter : $\frac{X_i}{\sigma} \sim \mathcal{R}(1)$.

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

⑤ Coverage probability of $I_n(\underline{X}_n)$

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

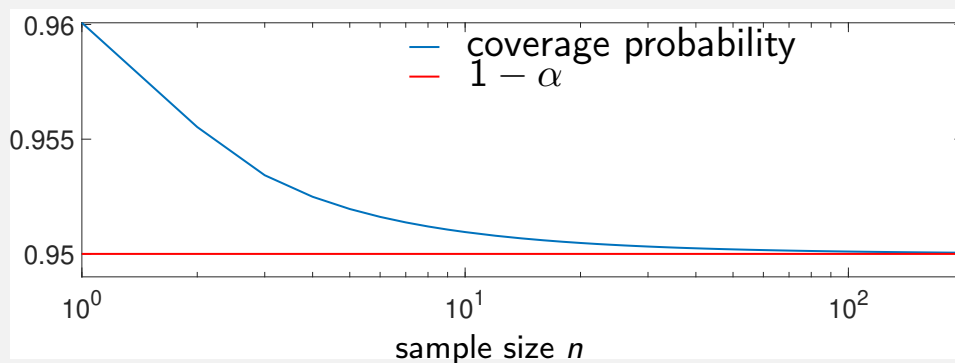
$$\text{with } \begin{cases} a_n &= n\sqrt{\frac{\pi}{2}} - \sqrt{n}\sqrt{2 - \frac{\pi}{2}}q_{1-\frac{\alpha}{2}} \\ b_n &= n\sqrt{\frac{\pi}{2}} + \sqrt{n}\sqrt{2 - \frac{\pi}{2}}q_{1-\frac{\alpha}{2}} \end{cases}$$

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

56/69

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) $1 - \alpha$:

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

57/69

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

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

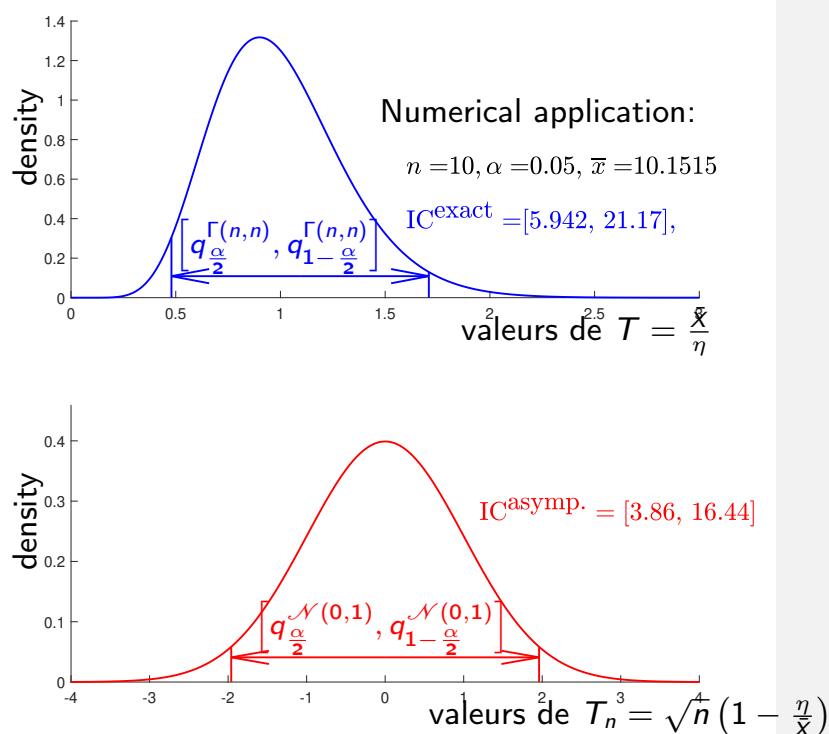
➡ Thus, T_n is an asymptotic pivotal function.

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.

58/69

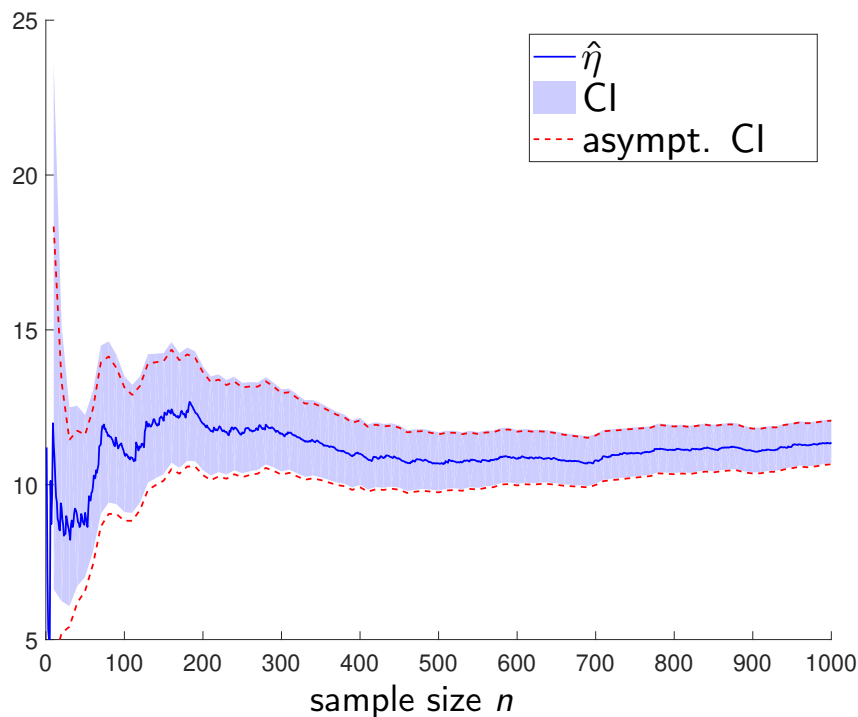


⚠ Do not confuse intervals on pivotal functions $[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$ and confidence interval for η .

59/69

Solution of exercise 4

[back to questions](#)



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

60/69

Solution of exercise 4

[back to questions](#)

③ Coverage probability of $I_n(\underline{X}_n)$

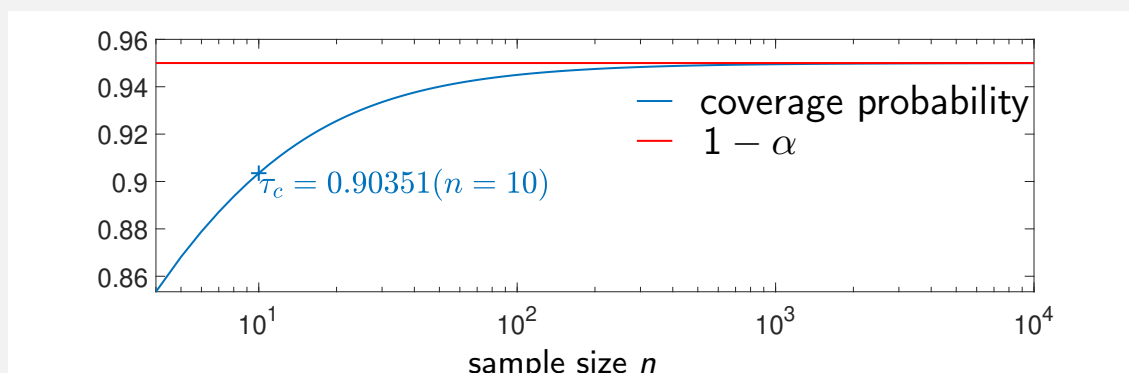
$$\begin{aligned} \mathbb{P}_\theta(\eta \in C_n(\underline{X}_n)) &= \mathbb{P}_\theta\left(\eta \in \left[\left(1 - \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right), \left(1 + \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right)\right] \bar{X}_n\right) \\ &= \mathbb{P}_\theta\left(\frac{1}{1 + \frac{1}{\sqrt{n}}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

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

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

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

61/69



Coverage probability of the asympt. CI with level 95%

Remarks.

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

62/69

Lecture outline

- 1 – Convergence rate and asymptotic distribution
- 2 – Confidence regions and confidence intervals
- 3 – Standard exercises (with solutions)
- 4 – Appendices

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 :

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

63/69

Proof (cont'd)

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

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

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

Thus,

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

□

back to slide ??

64/69

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 $Y_n \xrightarrow{d} c$, 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 $\lim_{n \rightarrow \infty} a_n = \infty$,

such that

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

then

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

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

[back to slide ??](#)

65/69

Proof "Delta-theorem" (scalar case)

Consider the function ψ defined by :

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

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

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

and thus (Slutsky)

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

Finally, we have

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

[back to slide ??](#)

66/69

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

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

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


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

back to slide 20

67/69

Example: an MLE that is not asymptotically Gaussian

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{U}_{[0, \theta]}$, 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
- ▶ $n(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} -Z$ with $Z \sim \mathcal{E}\left(\lambda = \frac{1}{\theta}\right)$.

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 ??

68/69

The Rayleigh $\mathcal{R}(\sigma^2)$ distribution

$X \sim \mathcal{R}(\sigma^2)$ 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 : $\text{var}_\sigma(X) = \sigma^2 \left(2 - \frac{\pi}{2}\right)$

Property

if $X \sim \mathcal{R}(\sigma^2)$ 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:

If $(X_1, \dots, X_n) \stackrel{\text{iid}}{\sim} \mathcal{R}(\sigma^2)$, then $Z = \sum_{i=1}^n X_i \sim \mathcal{SR}(n, \sigma^2)$.