CentraleSupélec

Statistics and Learning

Julien Bect

(julien.bect@centralesupelec.fr)

Teaching : CentraleSupélec / dept. of Statistics and Signal Processing

Research : Laboratory of Signals and Systems (L2S)

J. Bect & L. Le Brusquet — 1A — Statistics and Learning

Lecture 3/9

Asymptotic distributions and confidence regions intervals

In this lecture you will learn how to...

- Take the asymptotic approach one step further, introducing asymptotic distributions.
- Learn what confidence intervals are and show how to construct them (using, again, asymptotic arguments if needed)

Lecture outline

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

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 :

- $\lim_{n \to \infty} a_n = \infty$,
- $a_n (\hat{\eta}_n - \eta) \xrightarrow[n \to \infty]{d} Z$,
- 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 : $\exists a_1, \ldots, a_q, c \in \mathbb{R},\ \ \sum_{j=1}^{q} a_j Z^{(j)} = c$ a.s. ;

Exercise. Let Z be a random vector with finite second order moments.
➠ Prove that Z is non-degenerate iff its covariance matrix is invertible.

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. $\lim\limits_{n\to\infty} a_n = \infty$,
- a symmetric positive-definite matrix $\Sigma(\theta)$,

such that

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

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 consistency.

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 rel numbers such that $\lim_{n\to\infty} a_n = \infty$,

such that

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

then

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

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

Probability refresher : the Central Limit Theorem (CLT)

The CLT gives us the convergence rate of the empirical mean.

Theorem

Let

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

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

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

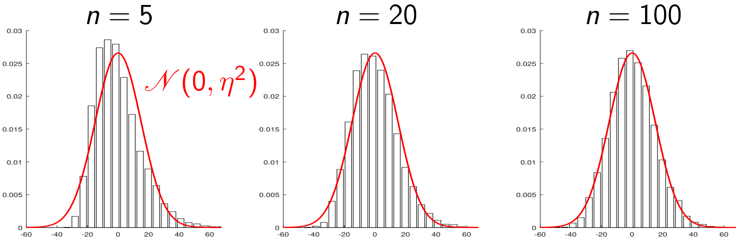
Convergence rate of the sample mean.
$\bar{X}_n$ converges to $\mathbb{E}(X_1)$ at the rate $\frac{1}{\sqrt{n}}$.

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



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

Another exemple : indicator function

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

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

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

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

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

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

Lecture outline

Slutsky's theorem

Theorem

Let

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

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

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

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

Then

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

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

The continuous mapping theorem

Theorem (Mann-Wald)

Let

- $h : \mathbb{R}^d \to \mathbb{R}^q$ a continuous 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 $\mathbb{R}^d$,

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

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.

Exercise : prove that $\hat{\theta}_n$ is also consistent the L^2 sense.

Example : component reliability (cont'd)

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

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

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

Therefore, by the continuous mapping theorem,

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

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

Linearization method ("delta method")

Theorem ("delta theorem")

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

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

Y a random vector, 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,

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

where $(Dh)(m)$ is the Jacobian matrix of h at m :

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

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

Special cases

Gaussian case

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

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

Scalar case

If $d = q = 1$ and $\sqrt{n}\,(Y_n - m) \xrightarrow[n\to\infty]{d} Z$, then

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

Exercise. If $h'(m) = 0$, and if h is twice differentiable at m, show that

$$n\,(h(Y_n) - h(m)) \xrightarrow[n\to\infty]{d} \frac{1}{2}\,h''(m)\,Z^2.$$

Proof (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 $Y_n \xrightarrow[n\to\infty]{d} m$,

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

and thus (Slutsky)

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

Finally, we have

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

Example : component reliability (cont'd)

Application : comparing estimators of $\eta = \mathbb{E}_\theta(X_1)$.

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

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

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

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

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

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

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

(Actually, it can be proved that $\hat\eta^{(1)}$ is efficient ; see comput. of the FIM below).

Asymptotic comparison of (scalar) estimators

Let $\hat{\eta}_n$ and $\tilde{\eta}_n$ be two estimators of $\eta = g(\theta) \in \mathbb{R}$,

- asymptotically Gaussian.
- with asymptotic variances $\sigma^2(\theta)$ and $\tilde{\sigma}^2(\theta)$.

Definition : asymptotically preferable

If

- the two estimators have the same convergence rate,
- $\sigma^2(\theta) \leq \tilde{\sigma}^2(\theta) \quad \forall \theta \in \Theta$,

then we say that

$\hat{\eta}_n$ is asymptotically preferable to $\tilde{\eta}_n$

("strictly" if $\exists \theta \in \Theta$ such that $\sigma^2(\theta) < \tilde{\sigma}^2(\theta)$).

Note : comparing vector-valued estimators $\Rightarrow$ compare matrices...

Example : component reliability (cont'd)

We already saw that

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

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

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

hence, since $h'(\eta) = -\frac{1}{\eta^2}$,

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

➠ the estimator $\hat{\theta}_n$ is asymptotically Gaussian.

Lecture outline

Asymptotic efficiency

Recall the Cramér-Rao lower bound (scalar parameter)

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

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

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

➠ 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 with rate $\frac{1}{\sqrt{n}}$,
- is its asymptotic variance is such that $\Sigma(\theta) = I_1^{-1}(\theta)$.

Remark : this definition is valid for the vector-valued case as well.

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_1–C_4 hold, (C_3 and C_4 defined below)
- $\forall \theta \in \Theta$, 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 :

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

Counter-example in a non-regular model

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. TD1, exercise 1.3)

- $\hat{\theta}_n = \max_{i \leq n} X_i$ is the MLE of θ, and
- $n\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{d} -Z$ 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 $\frac{1}{\sqrt{n}}$.

Partial solution of exercise 1.3 (TD 1)

By maximization of the likelihood, we get (see TD1) :

$$\hat{\theta}_n = \max_{i=1\dots n} X_i.$$

We have, for $0 \leq t \leq \theta$:

$$\mathbb{P}_\theta\left(\hat{\theta}_n \leq t\right) = \left(\frac{t}{\theta}\right)^n,$$

thus, for all $u \leq 0$:

$$\mathbb{P}_\theta\left(n\left(\hat{\theta}_n - \theta\right) \leq u\right) = \begin{cases} \left(1 + \dfrac{u}{n\theta}\right)^n & \text{if } u \geq -n\theta, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, for all $u \leq 0$,

$$\mathbb{P}_\theta\left(n\left(\hat{\theta}_n - \theta\right) \leq u\right) \xrightarrow[n\to\infty]{} \exp\left(\frac{u}{\theta}\right).$$

$\square$

Regularity conditions : C_3 and C_4

We assume that C_1 and C_2 hold.

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

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

Consequence of $\mathrm{C}_3 + \mathrm{C}_4$: Fisher information matrix

We assume that C_1 and C_2 hold.

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

$$I_{\underline{X}}(\theta) = \mathbb{V}_\theta(S_\theta(\underline{X})) = \mathbb{E}_\theta\left(S_\theta(\underline{X})\, S_\theta(\underline{X})^\top\right).$$

Proposition : another expression for the FIM

If conditions C_1–C_4 hold, then

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

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

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

Remark : actually, if C_1–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 $-\mathbb{E}_\theta\left(\frac{\partial S_\theta}{\partial\theta}(X_1)\right)$
$I_1(\theta) = \mathbb{V}_\theta(X_1) = \eta^2 = \frac{1}{\theta^2}$	$I_1(\theta) = -\mathbb{E}_\theta\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$

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

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

➠ We recover the conclusions of the theorem (C_1–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 $\hat{\eta}$,

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

Confidence regions and confidence intervals

Recall that $\eta = g(\theta)$. We denote by $\mathcal{P}(N)$ the subsets of $N = g(\Theta)$.

Definition : confidence region

Let $\alpha \in]0,1[$. A confidence region with level (at least) $1-\alpha$ for η is a statistics $I_\alpha(\underline{X})$ taking values in $\mathcal{P}(N)$, such that :

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

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

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

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

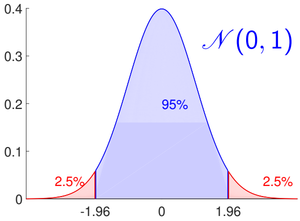
Scalar case : if $I_\alpha(\underline{X})$ is an interval, it is called a confidence interval.

Example : $\mathcal{N}(\mu, 1)$ n-sample, with known σ_0^2

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

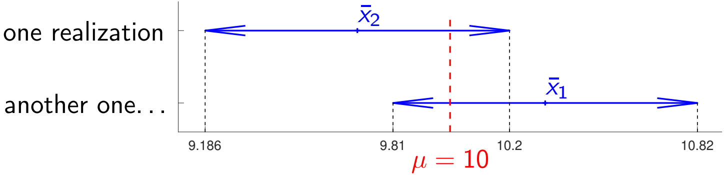
$$\mathbb{P}_\mu\left(\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma_0} \in \left[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}\right]\right) = 1-\alpha,$$

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



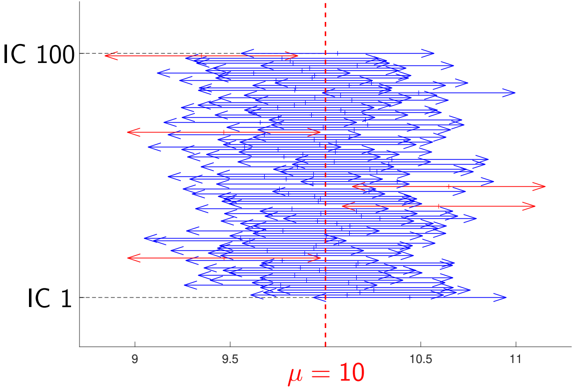
CI with level exactly $1-\alpha$:

$$\left[\bar{X} - 1.96\,\frac{\sigma_0}{\sqrt{n}},\ \bar{X} + 1.96\,\frac{\sigma_0}{\sqrt{n}}\right]$$



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 doe snot contain μ is (approx.) α.

Lecture outline

Pivotal functions

The method can be formalized using pivotal functions.

Definitions

A function

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

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** : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma_0^2)$ 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 $\mathbb{R}$.

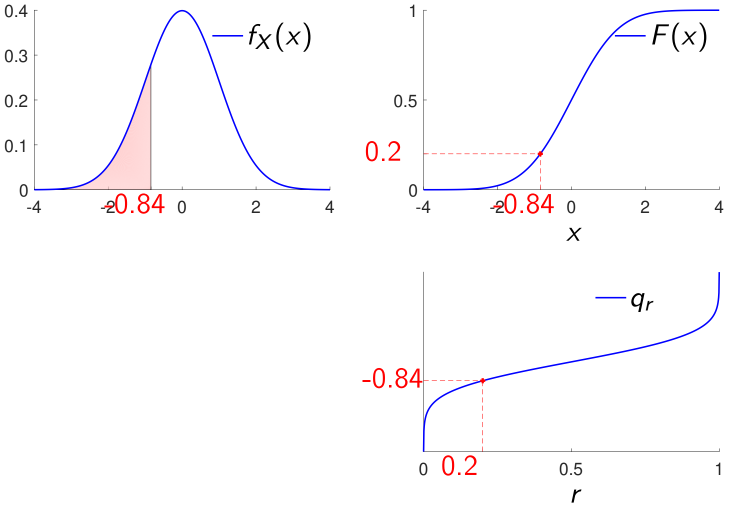
For $0 < r < 1$, the quantile of order r of the distribution is defined as :

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

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

Proposition

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

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

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

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

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

Example : $\mathcal{N}(\mu, 1)$ 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) $1 - \alpha$:

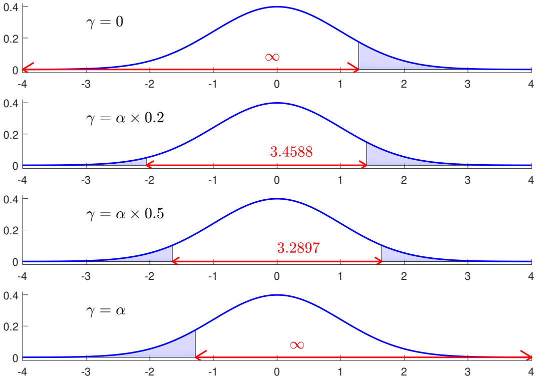
$$I_\alpha^\gamma = \left[\bar{X} + \frac{\sigma_0}{\sqrt{n}}\, q_\gamma, \quad \bar{X} + \frac{\sigma_0}{\sqrt{n}}\, q_{1-\alpha+\gamma}\right],$$

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

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

$$q_\gamma \approx -1.96 \quad \text{and} \quad q_{1-\alpha+\gamma} \approx 1.96.$$

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

Example : component reliability (cont'd)

It can be proved that :

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

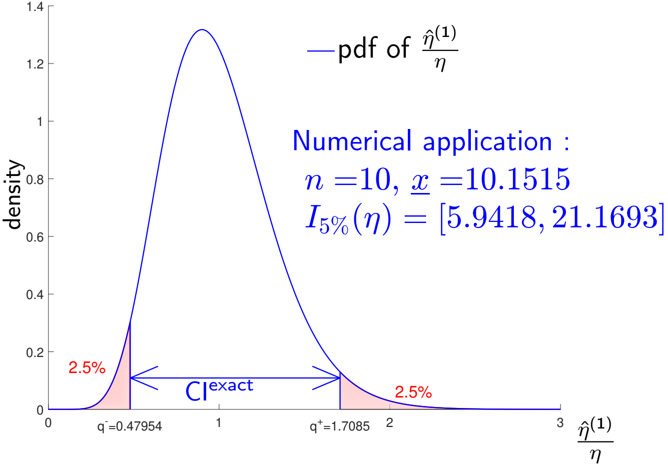
Thus, a CI with level exactly $1 - \alpha$ is :

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

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

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

Another exemple : the Rayleigh distribution

Definition : Rayleigh distribution with parameter σ^2

$$X \sim \mathscr{R}\left(\sigma^2\right) \quad \text{if (déf.)} \quad f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right),\ x \geq 0.$$

Pivotal function. Finding a pivotal function in this case requires the knowledge of some "fine" properties of the Rayleigh distribution...

It can be proved that : if $X \sim \mathscr{R}(\sigma^2)$ then $Y = X^2 \sim \mathcal{E}\left(\frac{1}{2\sigma^2}\right)$;

➠ thus $T\left(\underline{X}, \sigma^2\right) = \frac{1}{n\sigma^2} \sum_{i=1}^{n} X_i^2 \sim \Gamma\left(n, \frac{1}{2}\right)$ is pivotal for σ^2.

Hence a CI with level exactly $1 - \alpha$:

$$I_\alpha^{\gamma=\frac{\alpha}{2}} = \left[\frac{1}{nq_{1-\frac{\alpha}{2}}} \sum_{i=1}^{n} X_i^2,\ \frac{1}{nq_{\frac{\alpha}{2}}} \sum_{i=1}^{n} X_i^2\right].$$

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.
- Comput. become easier with the tools that we already have (CLT, Slutsky, delta method...).

⚠ Any analysis carried out in an asymptotic setting is

approximate when n is finite.

➠ The results can be bad 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 $I_{n,\alpha}(\underline{X}_n)$, with values in $\mathcal{P}(N)$, such that

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

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

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

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

(here, « exact » means « non asymptotic »).

Asymptotic pivotal function

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 $T_n\left(\underline{X}_n, \eta\right)$ does not depend on θ :

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

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

How to use asymptotic pivotal functions :

➠ exactly as we used the non-asymptotic ones !

Example : component reliability (cont'd)

We already saw that (Slutsky + continuity theorem)

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

➠ Asymptotic pivotal function :

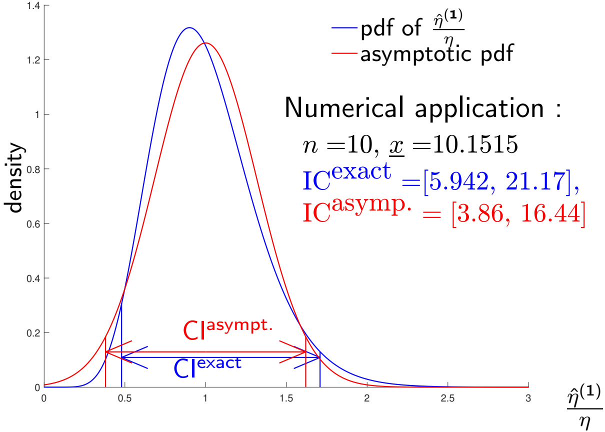
$$T_n\left(\underline{X}_n, \eta\right) = \sqrt{n}\,\frac{\bar{X} - \eta}{\bar{X}}.$$

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

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

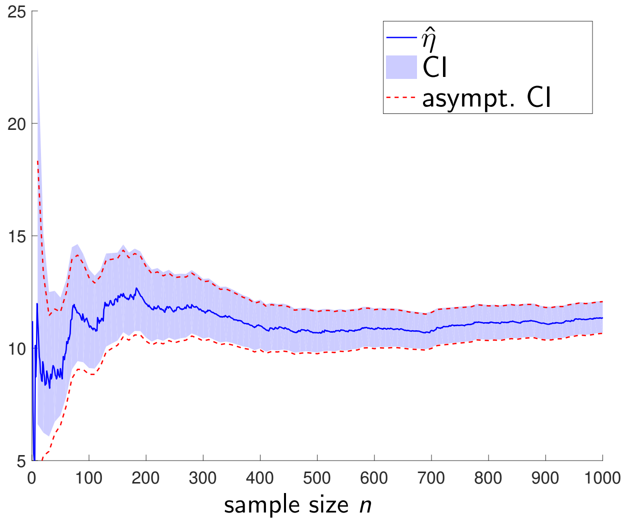
with q_r the quantile of order r of the $\mathcal{N}(0,1)$ distribution.

Example : component reliability (cont'd)



Comparison of the pdfs of pivotal functions (exact and asymptotic)

Example : component reliability (cont'd)



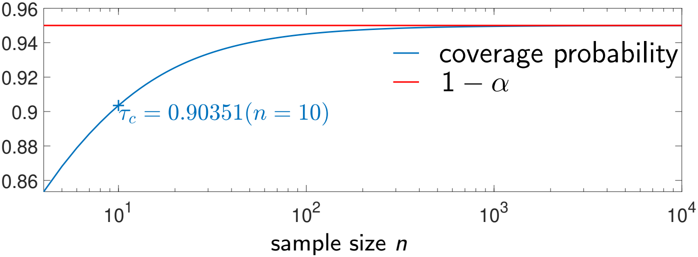
Comparison of exact and asymptotic CIs, as a function of n

Coverage probability of a confidence interval

Definition

For $\theta \in \Theta$, the coverage probability of $I_\alpha\left(\underline{X}_n\right)$ is defined by

$$\tau^c_{n,\theta}\left(I_\alpha\left(\underline{X}_n\right)\right) = \mathbb{P}_\theta\left(\eta \in I_\alpha\left(\underline{X}_n\right)\right)$$



Ex. « component reliability » : $\tau^c_{n,\theta}$ for the asympt. CI with level 95%

Remark. If $I_\alpha\left(\underline{X}_n\right)$ is an asympt. CI with level $1-\alpha$, then :

$$\forall \theta, \lim_{n\to\infty} \tau^c_\theta\left(I_\alpha(\underline{X}_n)\right) \geq 1-\alpha.$$