Théorie de l'estimation

Université Hassiba Benbouali de Chlef

Définition d'un estimateur

Définition 1.1

Un estimateur d'une grandeur θ est une statistique $\widehat{\theta}_n$ à valeurs dans l'ensemble des valeurs possibles de θ. Une estimation de θ est une réalisation de l'estimateur $\widehat{\theta}_n$.

Exemple

Le nombre moyen de guérisons est un estimateur "naturel" de la probabilité p :

$$\widehat{p}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Estimateur

On ne dispose que de la valeur de l'estimateur prise en les observations : $\widehat{\theta}_n = \varphi(x_1, \ldots, x_n)$

Exemples

- Efficacité d'un médicament : $\widehat{p}_n := \frac{1}{n}\sum_{i=1}^{n} x_i = 0.81$
- Durées de vie d'ampoules : $\widehat{\lambda}_n := \frac{n}{\sum_{i=1}^{n} x_i} = 0.033$

- On souhaite que cette estimation soit proche du paramètre inconnu
- Quelle confiance avoir en cette estimation ?

Pour le savoir, on étudie les propriétés théoriques de l'estimateur

- Propriétés à n fixé : biais, variance, risque quadratique
- Propriétés asymptotiques ($n \to +\infty$) : convergence, vitesse

Qualité d'un estimateur

Définition 1.2

On appelle biais de l'estimateur $\widehat{\theta}_n$ la fonction $\mathbb{B}_n(\theta) = \mathbb{E}[\widehat{\theta}_n] - \theta$.

- Un estimateur $\widehat{\theta}_n$ de θ est sans biais si et seulement si $\mathbb{B}_n(\theta) = 0$.
- $\widehat{\theta}_n$ est un estimateur asymptotiquement sans biais pour le paramètre θ, si pour tout $\theta \in \Theta$

$$\lim_{n \to +\infty} \mathbb{B}_n(\theta) = 0$$

Qualité d'un estimateur (suite)

Exemples

Qualité d'un estimateur (suite)

- $\bar{X}_n$ est un estimateur sans biais de la moyenne de la population m car $\mathbb{E}[\bar{X}_n] = m$.
- S_n^2 est un estimateur biaisé de la variance de la population σ^2, son biais est égale à :

$$\mathbb{B}_n(\sigma^2) = \mathbb{E}[S_n^2] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}.$$

- S_n^2 est un estimateur asymptotiquement sans biais pour σ^2,

$$\lim_{n \to +\infty} \mathbb{B}_n(\sigma^2) = \lim_{n \to +\infty} \frac{-\sigma^2}{n} = 0.$$

Qualité d'un estimateur (suite)

Définition 1.3

Le risque quadratique ou erreur quadratique moyenne est :
$EQM(\widehat{\theta}_n) = \mathbb{E}\left[(\widehat{\theta}_n - \theta)^2\right]$

Définition 1.4

L'estimateur $\widehat{\theta}_n$ converge en moyenne quadratique vers θ si et seulement si son erreur quadratique moyenne tend vers 0 quand n tend vers l'infini :

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{MQ} \theta \Leftrightarrow \lim_{n\to+\infty} \mathbb{E}\left[(\widehat{\theta}_n - \theta)^2\right] = 0.$$

Qualité d'un estimateur (suite)

L'erreur quadratique moyenne s'écrit :

$$
\begin{aligned}
EQM(\widehat{\theta}_n) &= \mathbb{E}\left[(\widehat{\theta}_n - \theta)^2\right] = \mathbb{E}\left[\left(\widehat{\theta}_n - \mathbb{E}[\widehat{\theta}_n] + \mathbb{E}[\widehat{\theta}_n] - \theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\widehat{\theta}_n - \mathbb{E}[\widehat{\theta}_n]\right)^2\right] + 2\mathbb{E}\left[\left(\widehat{\theta}_n - \mathbb{E}[\widehat{\theta}_n]\right)\right]\mathbb{E}\left[\left(\mathbb{E}[\widehat{\theta}_n] - \theta\right)\right] \\
&\quad + \mathbb{E}\left[\left(\mathbb{E}[\widehat{\theta}_n] - \theta\right)^2\right] \\
&= \mathrm{Var}[\widehat{\theta}_n] + \left(\mathbb{E}[\widehat{\theta}_n] - \theta\right)^2 \\
&= \text{Variance de l'estimateur} \; + \; \text{carré de son biais}
\end{aligned}
$$

Qualité d'un estimateur (suite)

Remarque

Si $\widehat{\theta}_n$ est un estimateur sans biais, $EQM(\widehat{\theta}_n) = \mathrm{Var}[\widehat{\theta}_n]$.

La moyenne empirique

La moyenne empirique $\bar{X}_n$ est un estimateur sans biais et convergent en moyenne quadratique de $\mathbb{E}[X]$.

Qualité d'un estimateur (suite)

Définition 1.5

Soient T_1 et T_2 deux estimateurs de θ, si $EQM(T_1) \leq EQM(T_2)$, $\forall\theta \in \Theta$, on dit alors que T_1 est préférable à T_2, si l'inégalité est stricte, on dira que T_1 est strictement préférable à T_2.

Le meilleur estimateur possible de θ est un estimateur sans biais et de variance minimale (**ESBVM**).

Théorème de Rao-Blackwell

Théorème 1.6

S'il existe une statistique exhaustive T et un estimateur sans biais $\widehat{\theta}_n$ de θ , alors $Z = \mathbb{E}[\widehat{\theta}_n|T]$ est un estimateur sans biais de θ , de variance inférieure à celle de $\widehat{\theta}_n$.

Théorème de Lehmann-Scheffé

Théorème 1.7

Si $\widehat{\theta}_n$ est un estimateur sans biais de θ et T est une statistique exhaustive et complète, alors $Z = \mathbb{E}[\widehat{\theta}_n|T]$ est l'unique estimateur sans biais de θ, de variance minimale parmi tous les estimateurs sans biais de θ.

Corollaire

Pour trouver un estimateur optimal, il suffit de trouver un estimateur sans biais fonction d'une statistique exhaustive et complète.

Modèle de Bernoulli

La moyenne empirique $\widehat{p}_n = \bar{X}_n$ fonction de la statistique exhaustive et complète $\sum_{i=1}^{n} X_i$ c'est l'**ESBVM** de p.

Exemple

Information de Fisher

Définition 1.8

Pour $\theta \in \mathbb{R}$, si la loi des observations vérifie les conditions de régularité, on appelle quantité d'information (de Fisher) sur θ apportée par l'échantillon $X_1, \ldots, X_n$, la quantité :

$$\mathcal{I}_n(\theta) = \mathrm{Var}\left[\frac{\partial \ln \mathcal{L}(\theta, X_1, \ldots, X_n)}{\partial \theta}\right]$$

Information de Fisher (suite)

On peut montrer que $\mathbb{E}\left[\frac{\partial \ln \mathcal{L}(\theta, X_1, \ldots, X_n)}{\partial \theta}\right] = 0$. Par conséquent, la quantité d'information peut aussi s'écrire sous la forme :

$$\mathcal{I}_n(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}(\theta, X_1, \ldots, X_n)}{\partial \theta}\right)^2\right]$$

On montre que l'on a également :

$$\mathcal{I}_n(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}(\theta, X_1, \ldots, X_n)}{\partial \theta^2}\right]$$

Cette écriture peut s'avérer pratique pour les calculs.

Information de Fisher (suite)

Échantillon de loi Bernoulli

$$\begin{aligned}
\mathcal{I}_n(p) &= \mathrm{Var}\left[\frac{\partial \ln \mathcal{L}(p, X_1, \ldots, X_n)}{\partial p}\right] = \mathrm{Var}\left[\frac{\displaystyle\sum_{i=1}^{n} X_i - np}{p(1-p)}\right] \\
&= \frac{\mathrm{Var}\left[\displaystyle\sum_{i=1}^{n} X_i\right]}{p^2(1-p)^2} = \frac{np(1-p)}{p^2(1-p)^2} = \frac{n}{p(1-p)}.
\end{aligned}$$

Inégalité de Fréchet-Darmois-Cramer-Rao (FDCR)

Si la loi des observations vérifie les conditions de régularité, alors pour tout estimateur $\widehat{\theta}_n$ de θ, on a :

$$\operatorname{Var}[\widehat{\theta}_n] \geq \frac{\left[\dfrac{\partial}{\partial\theta}\mathbb{E}[\widehat{\theta}_n]\right]^2}{\mathcal{I}_n(\theta)}.$$

Remarque

- Si $\widehat{\theta}_n$ est un estimateur sans biais de θ, alors $\mathbb{E}[\widehat{\theta}_n] = \theta$, donc $\operatorname{Var}[\widehat{\theta}_n] \geq \dfrac{1}{\mathcal{I}_n(\theta)}$.
- La quantité $\dfrac{1}{\mathcal{I}_n(\theta)}$ est appelée la borne de Cramer-Rao.

Estimateur efficace

- Un estimateur efficace est un estimateur pour lequel l'inégalité FDCR est une égalité.
- Si un estimateur sans biais est efficace, alors il est forcément de variance minimale et sa variance est égale à la borne de Cramer-Rao.

Exemple

La moyenne empirique $\widehat{p}_n = \bar{X}_n$ est un estimateur efficace. En effet,

$$\mathrm{Var}[\bar{X}_n] = \frac{\mathrm{Var}[X]}{n} = \frac{p(1-p)}{n} = \frac{1}{\mathcal{I}_n(p)}$$

Méthodes d'estimation

- Méthode des moments
- Maximum de vraisemblance

Méthode des moments

Soit $\varphi = (\varphi_1, \dots, \varphi_k)$ une application de $\mathcal{X}$ dans $\mathbb{R}^k$, telle que l'application :

$$\begin{aligned} \Phi : \quad \Theta &\longrightarrow \mathbb{R}^k \\ \theta &\longmapsto \mathbb{E}[\varphi(x)] = \Phi(\theta). \end{aligned}$$

soit injective.
On définit l'estimateur $\widehat{\theta}_n$ comme solution dans Θ (quand elle existe) de l'équation :

$$\Phi(\theta) = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i).$$

Méthode des moments (suite)

Souvent lorsque $\mathcal{X} \subset \mathbb{R}$, la fonction qu'on prend est $\varphi_i(x) = x^i$ et Φ correspond donc au i$^{\text{ème}}$ moment de la variable X. On résout le système d'équation suivant :

$$\begin{cases} \mathbb{E}[X] & = & \frac{1}{n}\sum_{i=1}^{n} X_i \\ \mathbb{E}[X^2] & = & \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \vdots & & \vdots \\ \mathbb{E}[X^k] & = & \frac{1}{n}\sum_{i=1}^{n} X_i^k \end{cases}$$

Le modèle de Bernoulli

L'estimation par la méthode des moments consiste a résoudre l'équation :

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i \Leftrightarrow p = \bar{X}_n,$$

alors

$$\widehat{p}_n = \bar{X}_n.$$

Loi exponentielle

On a

$$\mathbb{E}\left[X\right] = \frac{1}{\lambda},$$

l'estimateur par moment associé s'écrit

$$\widehat{\lambda}_n = \frac{1}{\overline{X}_n}.$$

Le modèle Uniforme $\mathcal{U}[0, \theta]$

On résout l'équation :

$$\begin{aligned} \mathbb{E}[X] &= \bar{X}_n \\ \Rightarrow \frac{\theta}{2} &= \bar{X}_n \\ \Rightarrow \widehat{\theta}_n &= 2\bar{X}_n \end{aligned}$$

Le modèle Normal $\mathcal{N}(m, \sigma^2)$

On résout le système d'équations :

$$\begin{cases} \mathbb{E}[X] = \bar{X}_n \\ \mathbb{E}[X^2] = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases} \Rightarrow \begin{cases} m = \bar{X}_n \\ m^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

on obtient :

$$\begin{cases} \hat{m} = \bar{X}_n \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 \end{cases}$$

Estimateur du maximum de vraisemblance

Soit $X_1, \ldots, X_n$ i.i.d. de loi $\mathbb{P}_\theta$, $\{\mathbb{P}_\theta, \theta \in \Theta\}$ dominée, $\Theta \subset \mathbb{R}^d$, $\theta \mapsto \mathcal{L}(\theta, X_1, \ldots, X_n)$ vraisemblance associée.

Définition 2.1

On appelle estimateur du maximum de vraisemblance tout estimateur $\widehat{\theta}_n$ satisfaisant

$$\mathcal{L}(\widehat{\theta}_n, X_1, \ldots, X_n) = \sup_{\theta \in \Theta} \mathcal{L}(\theta, X_1, \ldots, X_n).$$

Exemple introductif

On considère n tirages successifs à pile ou face, avec ses n variables de Bernoulli prenant la valeur 1 avec la probabilité p quand face sort et la valeur 0 avec la probabilité $1-p$ quand c'est pile. La valeur du paramètre p nous est inconnue.
Par exemple pour $n = 10$, on obtient comme résultat

$$0,\ 1,\ 1,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1$$

On écrit la vraisemblance :

Exemple introductif (suite)

$$
\begin{aligned}
\mathcal{L}(p, x_1, \ldots, x_n) &= \prod_{i=1}^{n} \mathbb{P}(X = x_i) \\
&= p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i} \\
&= p^6 (1-p)^4.
\end{aligned}
$$

Voici quelques valeurs numériques :

Exemple introductif (suite)

p	0.3	0.4	0.5	0.6	0.7	0.8
$p^6(1-p)^4$	0.00018	0.00053	0.00097	0.0012	0.00095	0.00036

On s'aperçoit que la quantité $p^6(1-p)^4$ admet un maximum au voisinage de $p = 0.6$, voir Figure 1

Exemple introductif (suite)

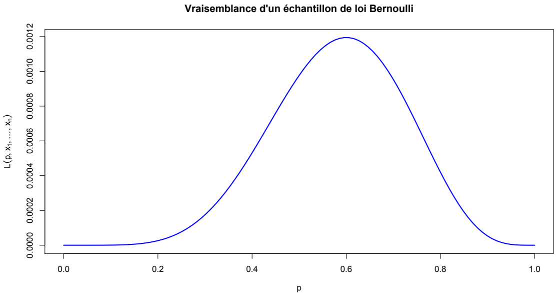

FIGURE – Vraisemblance pour un échantillon de taille 10 de la loi de Bernoulli, pour $p = 0.6$.

Remarques

- Log-vraisemblance :

$$\begin{aligned}\theta \mapsto \ell_n(\theta, X_1, \dots, X_n) &= n^{-1} \log \mathcal{L}(\theta, X_1, \dots, X_n) \\ &= n^{-1} \sum_{i=1}^{n} \log f_\theta(X_i).\end{aligned}$$

  Bien défini si $f_\theta(\cdot) > 0$ μ-pp.

  Max. vraisemblance = max. log-vraisemblance.

- L'estimateur du maximum de vraisemblance ne dépend pas du choix de la mesure dominante μ.
- Racine de l'équation de vraisemblance : tout estimateur $\widehat{\theta}_n$ vérifiant

$$\nabla_\theta \ell_n(\widehat{\theta}_n, X_1, \dots, X_n) = 0.$$

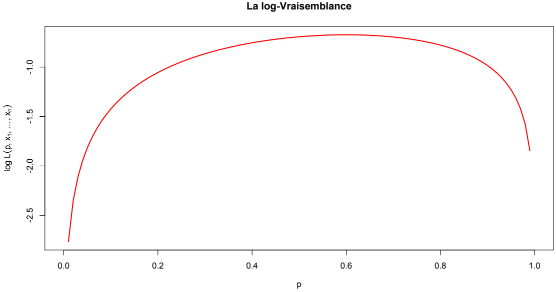

FIGURE – La log-vraisemblance pour un échantillon de taille 10 de la loi de Bernoulli, pour $p = 0.6$.

Exemple : modèle normal

L'expérience statistique est engendrée par un n-échantillon de loi $\mathcal{N}(m, \sigma^2)$, le paramètre est $\theta = (m, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+ \setminus \{0\}$.

- Vraisemblance

$$\mathcal{L}((m, \sigma^2), X_1, \ldots, X_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\tfrac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - m)^2 \right).$$

- Log-vraisemblance

$$\log \mathcal{L}\left((m, \sigma^2), X_1, \ldots, X_n\right) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - m)^2.$$

Exemple : modèle normal

Equation(s) de vraisemblance

$$\begin{cases} \partial_m \ell_n\left((m,\sigma^2), X_1, \ldots, X_n\right) & = & \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - m) \\ \\ \partial_{\sigma^2} \ell_n\left((m,\sigma^2), X_1, \ldots, X_n\right) & = & -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(X_i - m)^2. \end{cases}$$

Solution de ces équations (pour $n \geq 2$) :

$$\boxed{\widehat{\theta}_n = \left(\overline{X}_n, \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right)}$$

et on vérifie que $\widehat{\theta}_n = \widehat{\theta}_n$.

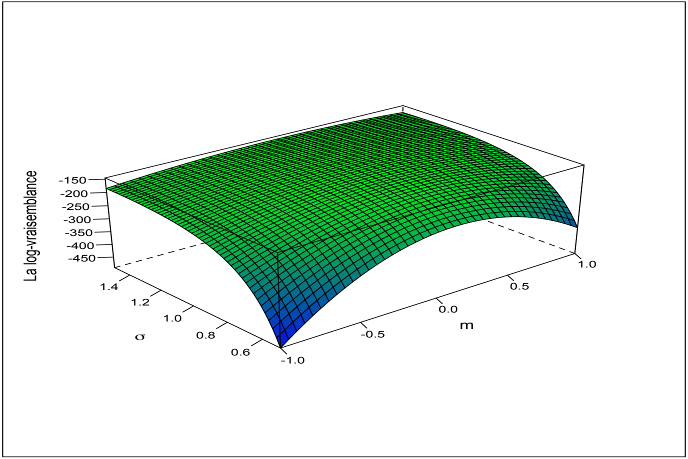

FIGURE – La log-Vraisemblance pour un échantillon de taille 100 de la loi Normale $\mathcal{N}(0, 1)$.

Exemple : modèle de Poisson

- Vraisemblance

$$\mathcal{L}(\theta, X_1, \ldots, X_n) = \frac{1}{\prod_{i=1}^n X_i!} e^{-n\theta} \theta^{\sum_{i=1}^n X_i}.$$

- Log-vraisemblance

$$\ell_n(\theta, X_1, \ldots, X_n) = c(X_1, \ldots, X_n) - n\theta + \sum_{i=1}^n X_i \log \theta.$$

- Equation de vraisemblance

$$-n + \sum_{i=1}^n X_i \frac{1}{\theta} = 0, \quad \text{soit} \quad \boxed{\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}_n}$$

Exemple : modèle de Laplace

L'expérience statistique est engendrée par un n-échantillon de loi de Laplace de paramètre $\theta \in \Theta = \mathbb{R}$. La densité par rapport à la mesure de Lebesgue :

$$f_\theta(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\theta|}{\sigma}\right),$$

où $\sigma > 0$ est connu.

- Vraisemblance

$$\mathcal{L}(\theta, X_1, \ldots, X_n) = (2\sigma)^{-n} \exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n} |X_i - \theta|\right)$$

- Log-vraisemblance

$$\ell_n(\theta, X_1, \ldots, X_n) = -n\log(2\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n} |X_i - \theta|.$$

Exemple : modèle de Laplace

Maximiser $\mathcal{L}_n(\theta, X_1, \dots, X_n)$ revient à minimiser la fonction $\theta \mapsto \sum_{i=1}^n \left| X_i - \theta \right|$, dérivable presque partout de dérivée constante par morceaux. Equation de vraisemblance :

$$\sum_{i=1}^n \mathrm{sign}(X_i - \theta) = 0.$$

Soit $X_{(1)} \leq \ldots \leq X_{(n)}$ la statistique d'ordre.

- n pair : $\widehat{\theta}_n$ n'est pas unique ; tout point de l'intervalle $\left[X_{(\frac{n}{2})}, X_{(\frac{n}{2}+1)} \right]$ est un EMV.
- n impair : $\widehat{\theta}_n = X_{\left(\frac{n+1}{2}\right)}$, l'EMV est unique.
- pour tout n, la médiane empirique est un EMV.

Exemple : modèle de Cauchy

- Vraisemblance

$$\mathcal{L}(\theta, X_1, \ldots, X_n) = \pi^{-n} \prod_{i=1}^{n} \frac{1}{1+(X_i-\theta)^2}.$$

- Log-vraisemblance

$$\ell_n(\theta, X_1, \ldots, X_n) = -n \log \pi - \sum_{i=1}^{n} \log \left(1+(X_i-\theta)^2\right)$$

- Equation de vraisemblance

$$\boxed{\sum_{i=1}^{n} \frac{X_i-\theta}{1+(X_i-\theta)^2} = 0}$$

pas de solution explicite et admet en général plusieurs solutions.

Exemple : modèle de Cauchy (suite)

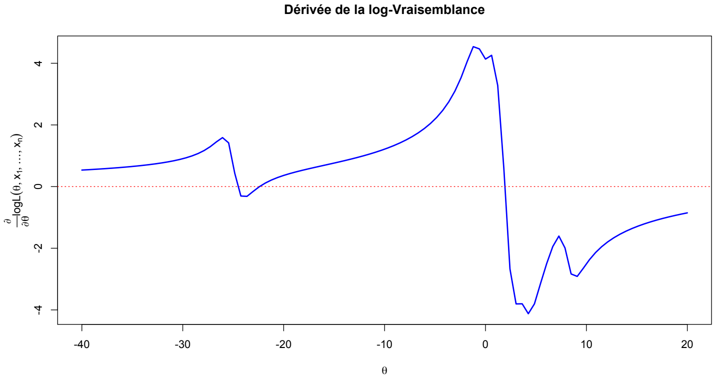

FIGURE – Dérivée de log-vraisemblance de la loi de Cauchy

Exemple : modèle uniforme

- On a

$$\begin{aligned}\mathcal{L}(\theta, X_1, \ldots, X_n) &= \theta^{-n} \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(X_i) \\ &= \theta^{-n} \mathbb{1}_{\{\max\limits_{1 \leq i \leq n} X_i \leq \theta\}}\end{aligned}$$

- La fonction de vraisemblance n'est pas régulière.
- L'estimateur du maximum de vraisemblance est $\widehat{\theta}_n = \max\limits_{1 \leq i \leq n} X_i$.

Exemple : modèle uniforme (suite)

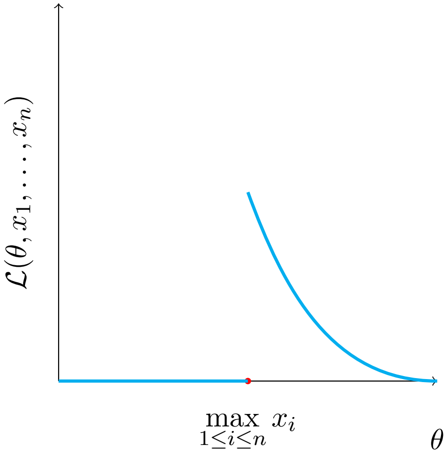

FIGURE – La vraisemblance pour la loi uniforme $\mathcal{U}[0, \theta]$.

Estimation des paramètres de la loi Gamma

- Soit $(X_1, X_2, \ldots, X_n)$ n observations i.i.d. de loi Gamma($\theta = (\alpha, \beta) \in \Theta = (\mathbb{R}_+^* \times \mathbb{R}_+^*)$)

$$f_\theta(x) = \Gamma(\alpha)^{-1} \beta^\alpha x^{\alpha-1} \mathrm{e}^{-\beta x}.$$

- log-vraisemblance

$$\log \mathcal{L}((\alpha, \beta)), X_1, \ldots, X_n) = -n \log \Gamma(\alpha) + n\alpha \log(\beta) + (\alpha - 1) \sum_{i=1}^{n} \log$$

- Le maximum ne se calcule pas explicitement.
- La minimisation par rapport à β pour α fixée est explicite : $\hat{\beta}_n(\alpha) = \alpha / \bar{X}_n$. L'estimateur du MV est obtenu en maximisant par rapport à α la fonction $\alpha \mapsto \log \mathcal{L}((\alpha, \hat{\beta}_n(\alpha)), X_1, \ldots, X_n)$.

Estimation par intervalle de confiance

Définition 3.1

Un intervalle de confiance de seuil (ou niveau de signification) $\alpha \in [0, 1]$ pour un paramètre θ, est un intervalle aléatoire $\mathbf{IC}$ tel que $\mathbb{P}(\mathbf{IC} \ni \theta) = 1 - \alpha$.

- Les bornes de l'intervalle de confiance $\mathbf{IC}$ dépendent de l'échantillon, elles sont donc aléatoires.
- Par abus de langage, on note souvent $\mathbb{P}(\mathbf{IC} \in \theta) = 1 - \alpha$.

Construction

Pour construire un intervalle de confiance, on utilise une variable aléatoire dont on connaît la distribution de probabilité.

Définition 3.2

Une fonction pivotale pour le paramètre θ est une fonction des observations $(X_1, \ldots, X_n)$ et du paramètre θ dont la loi ne dépend pas du paramètre θ.

On recherche dans la suite des fonctions pivotales particulières adaptées aux cas étudiés.

On envisage deux cas :

- La variable aléatoire mesurée est normale et le nombre de réalisations est quelconque,
- La variable aléatoire mesurée n'est pas normale et le nombre de réalisations est important. On parlera d'intervalle de confiance asymptotique.

Échantillons gaussiens

Théorème 3.3

Si $(X_1, \ldots, X_n)$ *est un échantillon de la loi* $\mathcal{N}(m, \sigma^2)$*, alors :*

- $\sqrt{\frac{n}{\sigma^2}}\left(\bar{X}_n - m\right)$ *suit la loi normale* $\mathcal{N}(0,1)$*.*
- $\sqrt{\frac{n-1}{S_n^2}}\left(\bar{X}_n - m\right)$ *suit la loi de Student* $\mathcal{T}_{(n-1)}$*.*
- $\frac{nS_n^2}{\sigma^2}$ *suit la loi du chi-deux* $\chi^2_{(n-1)}$*.*

Intervalle de confiance pour la moyenne d'un échantillon gaussien

Quand la variance est connue, après centrage et réduction de la moyenne empirique, d'après le Théorème 3.3, on obtient :

$$\mathbb{P}\left[\sqrt{\frac{n}{\sigma^2}}\left(\bar{X}_n - m\right) \in \left[-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right]\right] = 1-\alpha$$

où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$.

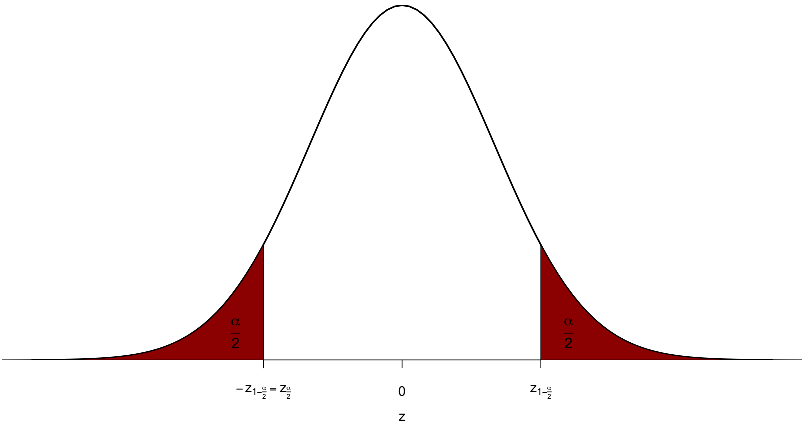

FIGURE – Quantile de la loi normale $\mathcal{N}(0,1)$.

Or :

$$\begin{aligned} & \sqrt{\frac{n}{\sigma^2}}\left(\bar{X}_n - m\right) \in \left[-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right] \\ \Leftrightarrow \quad & \bar{X}_n - m \in \left[-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}, z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right] \\ \Leftrightarrow \quad & m \in \left[\bar{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]. \end{aligned}$$

L'intervalle de confiance bilatéral symétrique pour l'espérance d'une loi normale s'écrit donc au niveau $1-\alpha$ sous la forme suivante :

$$\mathbf{IC}(m) = \left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

$\bar{x}_n$ est la réalisation de $\bar{X}_n$ sur l'échantillon.

Une application numérique

On prend $n = 100$ et $\sigma = 1$. $\bar{x}_{100} = 0.174$.

- Pour $\alpha = 0.1$, $z_{1-\frac{\alpha}{2}} = 1.65$

$$\begin{aligned}\mathbf{IC}(m) &= \left[0.174 - 1.65\frac{1}{\sqrt{100}}, 0.174 + 1.65\frac{1}{\sqrt{100}}\right] \\ &= [0.009, 0.339]\end{aligned}$$

- Pour $\alpha = 0.05$, $z_{1-\frac{\alpha}{2}} = 1.96$

$$\begin{aligned}\mathbf{IC}(m) &= \left[0.174 - 1.96\frac{1}{\sqrt{100}}, 0.174 + 1.96\frac{1}{\sqrt{100}}\right] \\ &= [-0.022, 0.37]\end{aligned}$$

Remarque

- Remarquons que si α augmente (ou que si n augmente), l'amplitude de l'intervalle de confiance diminue.
- Pour $\alpha = 0.05$, si on répète 100 fois l'expérience, on aura 100 intervalles de confiance différents. En moyenne, θ sera dans 95 de ces intervalles.

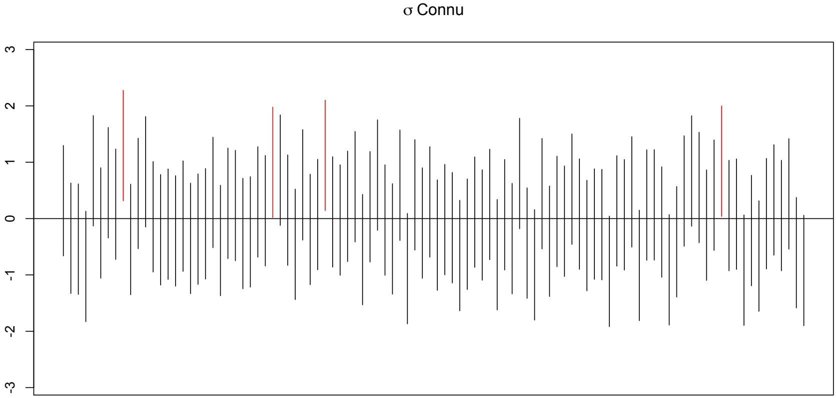

Figure – Intervalles de confiances

Quand la variance est inconnue, d'après le Théorème 3.3, $\sqrt{\dfrac{n-1}{S_n^2}}\left(\bar{X}_n - m\right)$ suit la loi de Student $\mathcal{T}_{(n-1)}$. On obtient :

$$\mathbb{P}\left[\sqrt{\frac{n-1}{S_n^2}}\left(\bar{X}_n - m\right) \in \left[-t_{n-1,1-\frac{\alpha}{2}}, t_{n-1,1-\frac{\alpha}{2}}\right]\right] = 1-\alpha$$

où $t_{n-1,1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student $\mathcal{T}_{(n-1)}$. L'intervalle de confiance bilatéral symétrique pour l'espérance d'une loi normale s'écrit donc au niveau $1-\alpha$ sous la forme suivante :

$$\mathbf{IC}(m) = \left[\bar{x}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{s_n}{\sqrt{n}}\right]$$

$\bar{x}_n$ et s_n sont les réalisations de $\bar{X}_n$ et S_n sur l'échantillon.

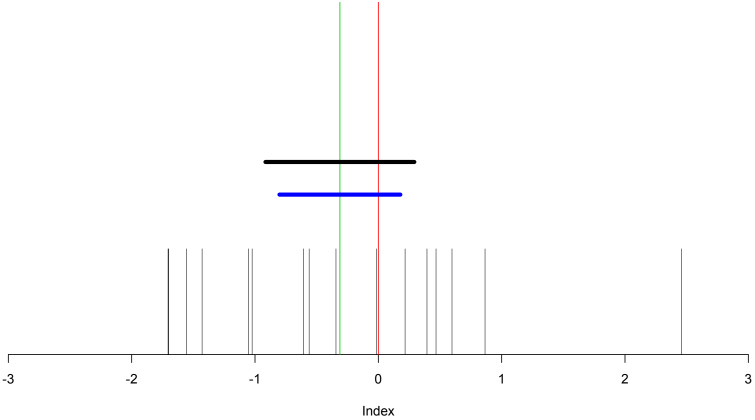

FIGURE – Intervalles de confiance basés sur la loi de Student (noir) et la loi normale (bleu).

Intervalle de confiance pour variance d'un échantillon gaussien

On considère la variance empirique $S_n^2 := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$ comme fonction pivotale pour σ^2. D'après le Théorème 3.3, $\frac{nS_n^2}{\sigma^2}$ suit la loi du chi-deux $\chi^2_{(n-1)}$. On a donc

$$\mathbb{P}\left[\frac{nS_n^2}{\sigma^2} \in \left[\chi^2_{(n-1),\frac{\alpha}{2}}, \chi^2_{(n-1),1-\frac{\alpha}{2}}\right]\right] = 1-\alpha$$

où $\chi^2_{(n-1),\frac{\alpha}{2}}$ est le quantile d'ordre $\frac{\alpha}{2}$ de la loi du chi-deux $\chi^2_{(n-1)}$ et $\chi^2_{(n-1),1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi du chi-deux $\chi^2_{(n-1)}$.

Quand l'espérance est inconnue, l'intervalle de confiance bilatéral pour la variance d'une loi normale s'écrit donc au niveau $1-\alpha$ sous la forme suivante :

$$\boxed{\mathbf{IC}(\sigma^2) = \left[\frac{ns_n^2}{\chi^2_{(n-1),1-\frac{\alpha}{2}}}, \frac{ns_n^2}{\chi^2_{(n-1),\frac{\alpha}{2}}}\right]}$$

s_n^2 est la réalisation de S_n^2 sur l'échantillon.

Intervalles de confiance asymptotiques

Définition 3.4

Soit $\alpha \in [0,1]$. On appelle intervalle de confiance asymptotique pour θ tout intervalle de la forme $[L_n, U_n]$, où L_n et U_n sont des fonctions telles que :

$$\lim_{n \to +\infty} \mathbb{P}\left([L_n, U_n] \ni \theta\right) \to 1 - \alpha.$$

Intervalle de confiance pour une proportion

Soit $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{B}(p)$. On cherche un intervalle de confiance asymptotique pour p. Un estimateur de p est $\widehat{p}_n = \bar{X}_n$. Par le TCL :

$$\sqrt{n}(\widehat{p}_n - p) \xrightarrow{d} \mathcal{N}\left(0, p(1-p)\right).$$

Un intervalle de confiance asymptotique de niveau $1-\alpha$ est donc donné par

$$\mathbf{IC}(p) = \left[\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sqrt{p(1-p)}}{\sqrt{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

Le problème est que cet intervalle de confiance dépend de p qui est inconnu.

Intervalle de confiance pour une proportion (suite)

On a d'après la loi des grands nombres

$$\bar{X}_n \xrightarrow{P} p.$$

Par continuité, on a

$$\sqrt{\bar{X}_n(1-\bar{X}_n)} \xrightarrow{P} \sqrt{p(1-p)},$$

et donc

$$\frac{\sqrt{p(1-p)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{P} 1.$$

On obtient donc d'après Slutsky

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \times \frac{\sqrt{p(1-p)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

Intervalle de confiance pour une proportion (suite)

Un intervalle de confiance asymptotique de niveau $1-\alpha$ est donné par

$$\boxed{\mathbf{IC}(p) = \left[\widehat{p}_n - z_{1-\frac{\alpha}{2}} \frac{\sqrt{\widehat{p}_n(1-\widehat{p}_n)}}{\sqrt{n}}, \widehat{p}_n + z_{1-\frac{\alpha}{2}} \frac{\sqrt{\widehat{p}_n(1-\widehat{p}_n)}}{\sqrt{n}}\right]}$$