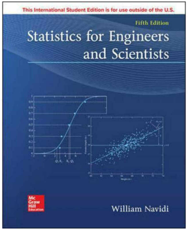
Introduction aux statistiques pour l'Ingénieur

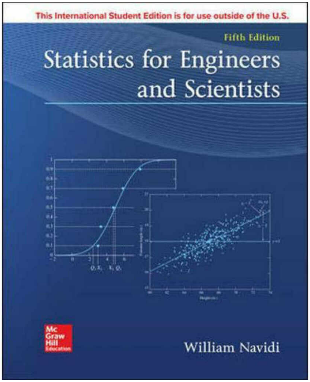
Stéphane Canu
asi.insa-rouen.fr/enseignants/~scanu
scanu@insa-rouen.fr



ITI 3, INSA Rouen Normandie, mars 2023

Lecture road map

1. Rappels
2. Technicités mathématiques
3. Information de Fisher
4. Borne de Cramer Rao

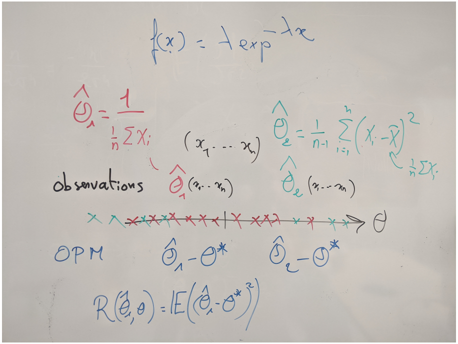


https://moodle.insa-rouen.fr/course/view.php?id=93

Comment mesurer la qualité d'un estimateur ?

On défini son risque

$$R_{\widehat{\theta}}(\theta) = \mathbb{E}\left((\widehat{\theta}(X_1, \dots, X_n) - \theta)^2\right)$$



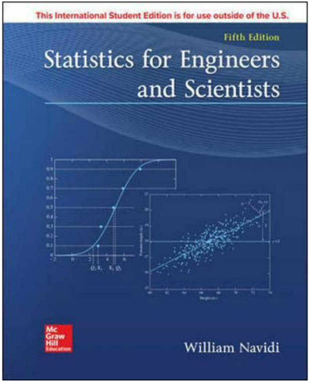
$$R_{\widehat{\theta}}(\theta) = \underbrace{\mathbb{E}\left(\left(\widehat{\theta} - \mathbb{E}(\widehat{\theta})\right)^2\right)}_{\textit{Variance}} + \underbrace{\left(\mathbb{E}(\widehat{\theta}) - \theta\right)^2}_{\textit{Biais}^2}$$

une stratégie pour minimiser le risque R

Trouver un estimateur sans biais de variance minimale

Lecture road map

1. Rappels
2. Technicités mathématiques
3. Information de Fisher
4. Borne de Cramer Rao



https://moodle.insa-rouen.fr/course/view.php?id=93

Technicités (1) : Intégrale et dérivées

Si

Alors

$$\frac{\partial}{\partial y}\int_{D(x)} \phi(x,y)\, dx \;=\; \int_{D(x)} \frac{\partial \phi(x,y)}{\partial y}\, dx$$

Exemple : prenons $x = (X_1, \ldots, X_n)$, $y = \theta$ et $\phi(x,y)$

$$\mathbb{E}\left(\frac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}\right) = 0$$

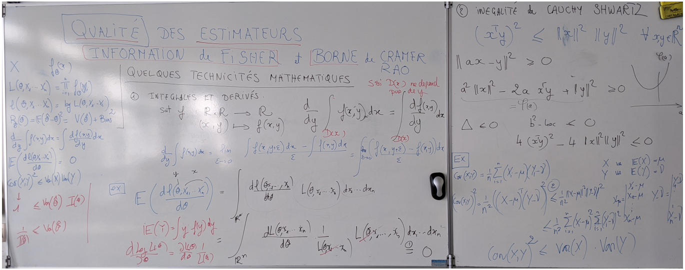
Technicités (2) : inégalité de Cauchy Schwartz

$\forall (x, y) \in \mathbb{R}^n$

$$(x^\top y)^2 \leq \|x\|^2 \|y\|^2$$

Exemple :

$$cov(X, Y)^2 \leq Var(X) Var(Y)$$



Information de Fisher : définition

Soit X une variable aléatoire parente de densité ou probabilité $f_\theta(X)$ Avec une log vraisemblance

$$\ell(\theta, X_1, \ldots, X_n) = \log L(\theta, X_1, \ldots, X_n) = \log \prod_{i=1}^{n} f_\theta(X_i)$$

L'information de Fisher est

$$I_n(\theta) = Var\Big(\frac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}\Big)$$

Pour $\theta \in \mathbb{R}^p$, L'information de Fisher est la matrice $p \times p$

$$I_n(\theta) = \mathbb{E}\big(\nabla \ell(\theta, X_1, \ldots, X_n) \nabla \ell(\theta, X_1, \ldots, X_n)^\top\big)$$

Information de Fisher : moyen de calcul

$$I_n(\theta) = Var(\ell'(\theta, X_1, \ldots, X_n))$$

Or

$$\mathbb{E}\left(\frac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}\right) = 0$$

et donc aussi

$$\frac{\partial}{\partial \theta}\mathbb{E}\left(\frac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}\right) = 0$$

D'où :

L'information de Fisher est aussi

$$I_n(\theta) = -\mathbb{E}(\ell''(\theta, X_1, \ldots, X_n)) = -\mathbb{E}\left(\frac{\partial^2 \ell(\theta, X_1, \ldots, X_n)}{\partial \theta^2}\right)$$

Information de Fisher : Recette et exemple

$$
\begin{aligned}
f(x) &= \lambda \exp^{-\lambda x} \\
L(\lambda, X_1, \ldots, X_n) &= \lambda^n \prod_{i=1}^{n} \exp^{-\lambda X_i} \\
\ell(\lambda, X_1, \ldots, X_n) &= n \log \lambda - \lambda \sum_{i=1}^{n} X_i \\
\frac{\partial \ell(\lambda, X_1, \ldots, X_n)}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} X_i \\
\frac{\partial^2 \ell(\lambda, X_1, \ldots, X_n)}{\partial \lambda^2} &= -\frac{n}{\lambda^2}
\end{aligned}
$$

$$
I_n(\lambda) = \frac{n}{\lambda^2}
$$

Matrice (d'information) de Fisher : Exemple

Estimation jointe des paramètres d'une loi normale ($\theta = (\mu, \sigma^2)$)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x-\mu)^2}{2\sigma^2}$$

$$\ell(\theta, X_1, \ldots, X_n) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi$$

$$\nabla \ell(\theta, X_1, \ldots, X_n) = \begin{pmatrix} -\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) \\ \frac{1}{\sigma^4} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{2\sigma^2} \end{pmatrix}$$

$$I(\mu, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

Théorème de Cramer Rao : énoncé

si

- le support de la distribution ne dépend pas de θ
- la log vraissemblance est différentiable et
- $\ldots$
- $\widehat{\theta}$ est un estimateur sans biais de θ

alors

$$BCR(\theta) = \frac{1}{I_n(\theta)} \leq \mathrm{Var}\big(\widehat{\theta}(X_1, \ldots, X_n)\big)$$

Théorème de Cramer Rao : preuve

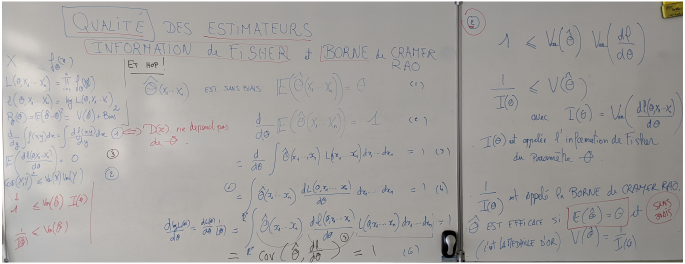
$$\mathbb{E}(\widehat{\theta}) = \theta \quad \Rightarrow \quad \frac{\partial \mathbb{E}(\widehat{\theta})}{\partial \theta} = 1$$

$$\int \widehat{\theta}(x_1, \dots, x_n) \frac{\partial L(x_1, \dots, x_n)}{\partial \theta}\, dx_1, \dots, dx_n = 1$$

$$\text{cov}\left(\widehat{\theta}(X_1, \dots, X_n), \frac{\partial \ell(X_1, \dots, X_n)}{\partial \theta}\right) = 1 \quad \text{car } \mathbb{E}\left(\frac{\partial \ell(\theta, X_1, \dots, X_n)}{\partial \theta}\right) = 0$$

or d'après Cauchy Schwartz

$$1 \leq \text{Var}\big(\widehat{\theta}(X_1, \dots, X_n)\big)\ \text{Var}\left(\frac{\partial \ell(X_1, \dots, X_n)}{\partial \theta}\right) = \text{Var}\big(\widehat{\theta}(X_1, \dots, X_n)\big) I(\theta)$$



Théorème de Cramer Rao : Exemple

Estimation de l'espérance d'une loi normale avec σ^2 connu ($\theta = \mu$)

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x-\mu)^2}{2\sigma^2}$

$$\ell(\mu, X_1, \ldots, X_n) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi$$

$$\frac{\partial \ell(\mu, X_1, \ldots, X_n)}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)$$

or $Var(X_i - \mu) = \sigma^2$ et donc

$$Var\left(\frac{\partial \ell(\mu, X_1, \ldots, X_n)}{\partial \mu}\right) = \frac{n}{\sigma^2}$$

L'estimateur $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est sans biais et efficace.

Théorème de Cramer Rao : Exemple

Estimation de la variance d'une loi normale μ connu ($\theta = \sigma^2$)

$$\frac{\partial \ell(\sigma^2, X_1, \ldots, X_n)}{\partial \sigma^2} = \frac{1}{2\sigma^4} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{2\sigma^2}$$

or, si X_i suit une loi normale, $\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}$ suit une loi du chi 2 à n degrés de liberté, et donc $Var(\sum_{i=1}^{n}(X_i - \mu)^2) = 2n\sigma^4$ d'où

$$Var\left(\frac{\partial \ell(\sigma^2, X_1, \ldots, X_n)}{\partial \sigma^2}\right) = \frac{n}{2\sigma^4}$$

L'estimateur $\widehat{\theta} = S_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n}(X_i - \bar{X})^2$ est bien sans biais mais pas efficace car :

$$BCR(\sigma^2) = \frac{2\sigma^4}{n} \leq Var(S_{n-1}) = \frac{2\sigma^4}{n-1}$$

minimum variance unbiased estimator (MVUE)

Si

$$\frac{\partial \ell(X_1, \ldots, X_n)}{\partial \theta} = A(n, \theta)\big(\widehat{\theta}(X_1, \ldots, X_n) - \theta\big)$$

Alors $\widehat{\theta}(X_1, \ldots, X_n)$ est un estimateur sans biais de risque minimal (UMUE)

Preuve : puisque $\mathbb{E}\left(\frac{\partial \ell(X_1,\ldots,X_n)}{\partial \theta}\right) = 0$,

$$\mathbb{E}\big(\widehat{\theta}(X_1, \ldots, X_n) - \theta\big) = 0$$

D'autre part, on a $I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ell(X_1,\ldots,X_n)}{\partial \theta^2}\right) = A(n, \theta)$ et $Var\left(\frac{\partial \ell(X_1,\ldots,X_n)}{\partial \theta}\right) = A(n, \theta)^2 Var\big(\widehat{\theta}(X_1, \ldots, X_n)\big)$ d'où :

$$Var\big(\widehat{\theta}(X_1, \ldots, X_n)\big) = \frac{1}{A(n, \theta)} = \frac{1}{I(\theta)}$$

Estimateur d'une fonction de θ

estimateur sans biais d'une fonction de θ : $\mathbb{E}(\widehat{u}(\theta)) = u(\theta)$

$$\mathbb{E}(\widehat{u}(\theta)) = u(\theta) \quad \Rightarrow \quad \frac{\partial \mathbb{E}(\widehat{u}(\theta))}{\partial \theta} = u'(\theta)$$

Dans ce cas :

$$BCR(\theta) = \frac{u'(\theta)}{I_n(\theta)} \leq \mathrm{Var}\big(\widehat{u}(\theta(X_1, \ldots, X_n))\big)$$

Conclusion - fiche recette

1. Vraisemblance : $L(\theta, X_1, \dots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$
2. Log vraisemblance : $\ell(\theta, X_1, \dots, X_n) = log\ L = \sum_{i=1}^{n} log\ f_\theta(X_i)$
3. Calcul de la dérivée de la log vraisemblance $\frac{\partial \ell(\theta, X_1, \dots, X_n)}{\partial \theta}$

S'il existe un estimateur $\widehat{\theta}(X_1, \dots, X_n)$ de θ tel que

$$\frac{\partial \ell(X_1, \dots, X_n)}{\partial \theta} = A(n, \theta)\big(\widehat{\theta}(X_1, \dots, X_n) - \theta\big)$$

Alors $\widehat{\theta}(X_1, \dots, X_n)$ est un estimateur sans biais de risque minimal (UMUE)
S'il existe une fonction u de $\mathbb{R}$ sur $\mathbb{R}$

$$\frac{\partial \ell(X_1, \dots, X_n)}{\partial \theta} = A(n, \theta)\big(\widehat{u}(X_1, \dots, X_n) - u(\theta)\big)$$

Alors $\widehat{u}(\theta)$ est un estimateur sans biais de risque minimal de $u(\theta)$

Exemple (1) : la loi de Bernouilli

$$
\begin{aligned}
f(x) &= p^x(1-p)^{1-x} \\
L(\lambda, X_1, \ldots, X_n) &= \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} \\
\ell(\lambda, X_1, \ldots, X_n) &= log p \sum_{i=1}^{n} X_i + log(1-p)(n - \sum_{i=1}^{n} X_i) \\
\frac{\partial \ell(\lambda, X_1, \ldots, X_n)}{\partial \lambda} &= \frac{\sum_{i=1}^{n} X_i}{p} - \frac{n - \sum_{i=1}^{n} X_i}{1-p} \\
&= \frac{n}{p(1-p)}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - p\Big)
\end{aligned}
$$

$$
\widehat{\theta}(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad A(n,p) = \frac{n}{p(1-p)}\,, \quad Var(\widehat{\theta}) = \frac{p(1-p)}{n}
$$

Exemple (2) : la loi exponentielle

$$\begin{aligned}
f(x) &= \lambda \exp^{-\lambda x} \\
L(\lambda, X_1, \ldots, X_n) &= \lambda^n \prod_{i=1}^{n} \exp^{-\lambda X_i} \\
\ell(\lambda, X_1, \ldots, X_n) &= n \log \lambda - \lambda \sum_{i=1}^{n} X_i \\
\frac{\partial \ell(\lambda, X_1, \ldots, X_n)}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} X_i \\
&= -n\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{\lambda}\Big)
\end{aligned}$$

$$u(\lambda) = \frac{1}{\lambda}, \quad \widehat{u}(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$Var(\widehat{u}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = -\frac{u'(\lambda)}{A(n,\lambda)} = \frac{1}{n\lambda^2},$$

Borne de Cramer Rao des estimateurs biaisés

Soit $\widehat{\theta}$ un estimateur de biais B avec donc :

$$B(\theta) = \mathbb{E}(\widehat{\theta}) - \theta$$

$$\mathbb{E}(\widehat{\theta}) = \theta + B(\theta) \quad \Rightarrow \quad \frac{\partial \mathbb{E}(\widehat{\theta})}{\partial \theta} = 1 + B'(\theta)$$

Dans ce cas :

$$BCR(\theta) = \frac{1 + B'(\theta)}{I_n(\theta)} \leq \mathrm{Var}\big(\widehat{\theta}(X_1, \ldots, X_n)\big)$$

Estimateur et vraisemblance

Le modèle : X de loi parente de distribution $f_\theta(x)$
Échantillon i.i.d.

$$(X_1, \ldots, X_n)$$

Vraisemblance :

$$L(\theta, X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$$

Log vraisemblance :

$$\ell(\theta, X_1, \ldots, X_n) = log\, L(\theta, X_1, \ldots, X_n) = \sum_{i=1}^{n} log\, f_\theta(X_i)$$

Estimateur du max de vraisemblance

$$\widehat{\theta}_{MV} = \arg\max_\theta \ell(\theta, X_1, \ldots, X_n) \quad \Leftrightarrow \quad \left.\frac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}\right|_{\theta = \widehat{\theta}_{MV}} = 0$$

Propriété des estimateurs max de vraisemblance

Si

- la probabilité ou de la densité dépend d'un paramètre θ
- le support de la probabilité ou de la densité ne dépend pas de θ
- si I existe, est inversible et quelques autres conditions plus techniques

Asymptotiquement sans biais

$$\widehat{\theta}_{MV} \underset{n\to\infty}{\longrightarrow} \theta^\star$$

Asymptotiquement efficace

$$Var(\widehat{\theta}_{MV}) \underset{n\to\infty}{\longrightarrow} I_n^{-1}(\theta^\star)$$

Asymptotiquement normal

$$\sqrt{n}(\widehat{\theta}_{MV} - \theta^\star) \underset{n\to\infty}{\longrightarrow} \mathcal{N}\big(0, I_1^{-1}(\theta^\star)\big)$$

note : $I_n(\theta^\star) = nI_1(\theta^\star)$