CentraleSupélec

Statistique et apprentissage

Chargés de cours (ordre alphabétique) :

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

[†] Coordinateur du cours

Cours 2/9

Estimation ponctuelle

Objectifs du cours 2

- Savoir quantifier les performances d'un estimateur
- Savoir comparer des estimateurs
- Introduire l'approche asymptotique

Plan du cours

Plan du cours

Rappel : cadre mathématique

Données

- Formellement, un élément $\underline{x}$ d'un ensemble $\underline{\mathcal{X}}$.
- ex : $\underline{\mathcal{X}} = \mathbb{R}^n$, $\mathbb{R}^{n \times d}$, $\{\text{mots}\}$, un espace fonctionnel, etc.

Des données aux VA

- Point de vue a priori : avant réalisation de l'expérience.
- Modélisation : VA $\underline{X}$ à valeur dans $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$,
- **mais** la loi de $\underline{X}$ est inconnue.

Modélisation statistique

- On suppose $\underline{X}$ définie sur $(\Omega, \mathscr{F}, \mathbb{P})$, avec $\mathbb{P} \in \mathscr{P}$.
- $\mathscr{P}$: un ensemble de mesure de proba possibles sur $(\Omega, \mathscr{F})$
- Formellement, $\mathscr{M} = \left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \mathscr{P}^{\underline{X}}\right)$, avec $\mathscr{P}^{\underline{X}} = \{\mathbb{P}^{\underline{X}}, \mathbb{P} \in \mathscr{P}\}$.

Réalisation canonique : $\Omega = \underline{\mathcal{X}}$, $\mathscr{F} = \underline{\mathscr{A}}$, $\underline{X} = \mathrm{Id}_{\underline{\mathcal{X}}}$ et $\mathscr{P} = \mathscr{P}^{\underline{X}}$.

Rappel : cadre mathématique

Données

- Formellement, un élément $\underline{x}$ d'un ensemble $\underline{\mathcal{X}}$.
- ex : $\underline{\mathcal{X}} = \mathbb{R}^n$, $\mathbb{R}^{n \times d}$, $\{\text{mots}\}$, un espace fonctionnel, etc.

Des données aux VA

- Point de vue a priori : avant réalisation de l'expérience.
- Modélisation : VA $\underline{X}$ à valeur dans $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$,
- **mais** la loi de $\underline{X}$ est inconnue.

Modélisation statistique

- On suppose $\underline{X}$ définie sur $(\Omega, \mathscr{F}, \mathbb{P})$, avec $\mathbb{P} \in \mathscr{P}$.
- $\mathscr{P}$: un ensemble de mesure de proba possibles sur $(\Omega, \mathscr{F})$
- Formellement, $\mathscr{M} = \left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \mathscr{P}^{\underline{X}}\right)$, avec $\mathscr{P}^{\underline{X}} = \{\mathbb{P}^{\underline{X}}, \mathbb{P} \in \mathscr{P}\}$.

Réalisation canonique : $\Omega = \underline{\mathcal{X}}$, $\mathscr{F} = \underline{\mathscr{A}}$, $\underline{X} = \mathrm{Id}_{\underline{\mathcal{X}}}$ et $\mathscr{P} = \mathscr{P}^{\underline{X}}$.

Rappel : cadre mathématique

Données

- Formellement, un élément $\underline{x}$ d'un ensemble $\underline{\mathcal{X}}$.
- ex : $\underline{\mathcal{X}} = \mathbb{R}^n$, $\mathbb{R}^{n \times d}$, $\{\text{mots}\}$, un espace fonctionnel, etc.

Des données aux VA

- Point de vue a priori : avant réalisation de l'expérience.
- Modélisation : VA $\underline{X}$ à valeur dans $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$,
- **mais** la loi de $\underline{X}$ est inconnue.

Modélisation statistique

- On suppose $\underline{X}$ définie sur $(\Omega, \mathscr{F}, \mathbb{P})$, avec $\mathbb{P} \in \mathscr{P}$.
- $\mathscr{P}$: un ensemble de mesure de proba possibles sur $(\Omega, \mathscr{F})$
- Formellement, $\mathscr{M} = \left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \mathscr{P}^{\underline{X}}\right)$, avec $\mathscr{P}^{\underline{X}} = \{\mathbb{P}^{\underline{X}}, \, \mathbb{P} \in \mathscr{P}\}$.

Réalisation canonique : $\Omega = \underline{\mathcal{X}}$, $\mathscr{F} = \underline{\mathscr{A}}$, $\underline{X} = \mathrm{Id}_{\underline{\mathcal{X}}}$ et $\mathscr{P} = \mathscr{P}^{\underline{X}}$.

Rappel : cadre mathématique (suite)

Important

Comme $\mathbb{P} \in \mathscr{P}$ est inconnue, on doit concevoir des procédures statistiques qui « fonctionnent bien » (en un sens à préciser) pour **toutes** les lois $\mathbb{P} \in \mathscr{P}$.

Famille paramétrée de probabilités

- Usuellement, on écrit $\mathscr{P} = \{\mathbb{P}_\theta,\ \theta \in \Theta\}$.
- θ : paramètre inconnu (scalaire, vectoriel, fonctionnel...)
- Dans la suite, on suppose un modèle paramétrique : $\Theta \subset \mathbb{R}^p$.

Cas important : n-échantillon (iid) d-varié (→ tableau $n \times d$)

- $\underline{\mathcal{X}} = \mathcal{X}^n$, avec $\mathcal{X} \subset \mathbb{R}^d$, munis de leurs tribus boréliennes,
- $\underline{X} = (X_1, \ldots, X_n)$ avec $X_i \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, et donc $\mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$.

Rappel : cadre mathématique (suite)

Important

Comme $\mathbb{P} \in \mathscr{P}$ est inconnue, on doit concevoir des procédures statistiques qui « fonctionnent bien » (en un sens à préciser) pour **toutes** les lois $\mathbb{P} \in \mathscr{P}$.

Famille paramétrée de probabilités

- Usuellement, on écrit $\mathscr{P} = \{\mathbb{P}_\theta,\ \theta \in \Theta\}$.
- θ : paramètre inconnu (scalaire, vectoriel, fonctionnel...)
- Dans la suite, on suppose un modèle paramétrique : $\Theta \subset \mathbb{R}^p$.

Cas important : n-échantillon (iid) d-varié (→ tableau $n \times d$)

- $\underline{\mathcal{X}} = \mathcal{X}^n$, avec $\mathcal{X} \subset \mathbb{R}^d$, munis de leurs tribus boréliennes,
- $\underline{X} = (X_1, \ldots, X_n)$ avec $X_i \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, et donc $\mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$.

Rappel : cadre mathématique (suite)

Important

Comme $\mathbb{P} \in \mathscr{P}$ est inconnue, on doit concevoir des procédures statistiques qui « fonctionnent bien » (en un sens à préciser) pour **toutes** les lois $\mathbb{P} \in \mathscr{P}$.

Famille paramétrée de probabilités

- Usuellement, on écrit $\mathscr{P} = \{\mathbb{P}_\theta,\ \theta \in \Theta\}$.
- θ : paramètre inconnu (scalaire, vectoriel, fonctionnel. . .)
- Dans la suite, on suppose un modèle paramétrique : $\Theta \subset \mathbb{R}^p$.

Cas important : n-échantillon (iid) d-varié (→ tableau $n \times d$)

- $\underline{\mathcal{X}} = \mathcal{X}^n$, avec $\mathcal{X} \subset \mathbb{R}^d$, munis de leurs tribus boréliennes,
- $\underline{X} = (X_1, \ldots, X_n)$ avec $X_i \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, et donc $\mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$.

Estimation (ponctuelle)

Paramètre à estimer

- On s'intéresse à un paramètre $\eta = g(\theta)$, où $g : \Theta \mapsto \mathbb{R}$ ou $\mathbb{R}^q$.
- Sa valeur est inconnue, puisque θ est inconnu.

Définition informelle : estimation

Deviner (inférer) la valeur de η à partir d'une réalisation $\underline{x}$ de $\underline{X}$.

Définition : estimateur

On appelle estimateur toute statistique $\hat{\eta} = \varphi(\underline{X})$ à valeurs dans l'ensemble $N = g(\Theta)$ des valeurs possibles de η.

Remarque : le mot « estimateur » peut désigner soit la VA $\hat{\eta}$ soit la fonction φ. En pratique on confond les deux et on notera (abusivement) $\hat{\eta} = \hat{\eta}(\underline{X})$.

Estimation (ponctuelle)

Paramètre à estimer

- On s'intéresse à un paramètre $\eta = g(\theta)$, où $g : \Theta \mapsto \mathbb{R}$ ou $\mathbb{R}^q$.
- Sa valeur est inconnue, puisque θ est inconnu.

Définition informelle : estimation

Deviner (inférer) la valeur de η à partir d'une réalisation $\underline{x}$ de $\underline{X}$.

Définition : estimateur

On appelle estimateur toute statistique $\hat{\eta} = \varphi(\underline{X})$ à valeurs dans l'ensemble $N = g(\Theta)$ des valeurs possibles de η.

Remarque : le mot « estimateur » peut désigner soit la VA $\hat{\eta}$ soit la fonction φ. En pratique on confond les deux et on notera (abusivement) $\hat{\eta} = \hat{\eta}(\underline{X})$.

Exemple 1

n-échantillon iid gaussien : $\underline{X} = (X_1, \ldots X_n)$ avec

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$,
- $\theta = \left(\mu, \sigma^2\right)$,
- $\Theta = \mathbb{R} \times]0; +\infty[$.

Dans cet exemple on supposera qu'on veut estimer la moyenne μ ;

- ici $\eta = \mu$ et $g : \theta = \left(\mu, \sigma^2\right) \mapsto \mu$,
- σ^2 est inconnu aussi (paramètre de nuisance).

Exemple 1 (suite)

Quelques estimateurs possibles...

- $\hat{\mu}_1 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ (méthode des moments / EMV),
- $\hat{\mu}_2 = \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé,
- $\hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n$,
- $\hat{\mu}_4 = \bar{X}_n + c$ pour un $c \neq 0$ fixé,
- $\hat{\mu}_5 = \mathrm{med}(X_1, \ldots, X_n)$,
- ...

Questions

- L'un de ces estimateurs est-il « meilleur » que les autres ?
- Peut-on trouver un estimateur « optimal » ?
- En quel sens ?

Exemple 1 (suite)

Quelques estimateurs possibles...

- $\hat{\mu}_1 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ (méthode des moments / EMV),
- $\hat{\mu}_2 = \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé,
- $\hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n$,
- $\hat{\mu}_4 = \bar{X}_n + c$ pour un $c \neq 0$ fixé,
- $\hat{\mu}_5 = \mathrm{med}(X_1, \ldots, X_n)$,
- ...

Questions

- L'un de ces estimateurs est-il « meilleur » que les autres ?
- Peut-on trouver un estimateur « optimal » ?
- En quel sens ?

Autres exemples

Exemple 1'

- Même modèle statistique que l'exemple 1, mais
- $g(\theta) = \sigma^2$.
- Dans ce cas, μ est vu comme un paramètre de nuisance.

Exemple 1''

- Toujours le même modèle statistique, mais
- $g(\theta) = \theta = (\mu, \sigma^2)$.
- Ici le paramètre à estimer est vectoriel.

Autres exemples

Exemple 1'

- Même modèle statistique que l'exemple 1, mais
- $g(\theta) = \sigma^2$.
- Dans ce cas, μ est vu comme un paramètre de nuisance.

Exemple 1"

- Toujours le même modèle statistique, mais
- $g(\theta) = \theta = (\mu, \sigma^2)$.
- Ici le paramètre à estimer est vectoriel.

Autres exemples (suite)

Exemple 2

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, i.e., $f_\theta(x) = \theta\, e^{-\theta x}\, \mathbb{1}_{x \geq 0}$,
- $\Theta =]0, +\infty[$,
- $g(\theta) = \mathbb{E}_\theta(X_1) = 1/\theta$.

Exemple 2'

- Même modèle statistique, mais
- $g(\theta) = \mathbb{P}_\theta\left(X_1 > x_0\right) = e^{-\theta x_0}$ pour un $x_0 > 0$ donné.

Exemple 3 (hors programme)

- statistique non-paramétrique

complément

Autres exemples (suite)

Exemple 2

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, i.e., $f_\theta(x) = \theta\, e^{-\theta x}\, \mathbb{1}_{x \geq 0}$,
- $\Theta =]0, +\infty[$,
- $g(\theta) = \mathbb{E}_\theta(X_1) = 1/\theta$.

Exemple 2'

- Même modèle statistique, mais
- $g(\theta) = \mathbb{P}_\theta\left(X_1 > x_0\right) = e^{-\theta x_0}$ pour un $x_0 > 0$ donné.

Exemple 3 (hors programme)

- statistique non-paramétrique

complément

Autres exemples (suite)

Exemple 2

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, i.e., $f_\theta(x) = \theta\, e^{-\theta x}\, \mathbb{1}_{x \geq 0}$,
- $\Theta =]0, +\infty[$,
- $g(\theta) = \mathbb{E}_\theta(X_1) = 1/\theta$.

Exemple 2'

- Même modèle statistique, mais
- $g(\theta) = \mathbb{P}_\theta\left(X_1 > x_0\right) = e^{-\theta x_0}$ pour un $x_0 > 0$ donné.

Exemple 3 (hors programme)

- statistique non-paramétrique

complément

Plan du cours

Notion générale de risque

Objectif

Quantifier les performances d'un estimateur

On se donne une fonction de perte $L : N \times N \to \mathbb{R}$.

- Rappel : $N = g(\Theta)$ est l'ensemble des valeurs possibles de η.
- Interprétation : on perd $L(\eta, \eta')$ si l'on estime η' alors que la valeur vraie est η.

Risque

Étant donné une fonction de perte L, on définit le risque $R_\theta(\hat{\eta})$ de l'estimateur $\hat{\eta}$, pour la valeur $\theta \in \Theta$ du paramètre, par

$$R_\theta\left(\hat{\eta}\right) = \mathbb{E}_\theta\left(L\left(g(\theta), \hat{\eta}\right)\right).$$

Notion générale de risque

Objectif

Quantifier les performances d'un estimateur

On se donne une fonction de perte $L : N \times N \to \mathbb{R}$.

- Rappel : $N = g(\Theta)$ est l'ensemble des valeurs possibles de η.
- Interprétation : on perd $L(\eta, \eta')$ si l'on estime η' alors que la valeur vraie est η.

Risque

Étant donné une fonction de perte L, on définit le risque $R_\theta(\hat{\eta})$ de l'estimateur $\hat{\eta}$, pour la valeur $\theta \in \Theta$ du paramètre, par

$$R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left(L\left(g(\theta), \hat{\eta}\right)\right).$$

Notion générale de risque

Objectif

Quantifier les performances d'un estimateur

On se donne une fonction de perte $L : N \times N \to \mathbb{R}$.

- Rappel : $N = g(\Theta)$ est l'ensemble des valeurs possibles de η.
- Interprétation : on perd $L(\eta, \eta')$ si l'on estime η' alors que la valeur vraie est η.

Risque

Étant donné une fonction de perte L, on définit le risque $R_\theta(\hat{\eta})$ de l'estimateur $\hat{\eta}$, pour la valeur $\theta \in \Theta$ du paramètre, par

$$R_\theta\left(\hat{\eta}\right) = \mathbb{E}_\theta\left(L\left(g(\theta), \hat{\eta}\right)\right).$$

Risque quadratique

Risque quadratique

On appelle risque quadratique le risque associé à la fonction de perte

$$L(\eta, \eta') = \|\eta - \eta'\|^2,$$

c'est-à-dire :

$$R_\theta\left(\hat{\eta}\right) = \mathbb{E}_\theta\left(\|g(\theta) - \hat{\eta}\|^2\right).$$

Remarques

- Aussi appelée « erreur quadratique moyenne » (EQM).
- C'est le risque le plus couramment utilisé
  (pour des raison de simplicité, comme on va le voir) ;
- dans la suite de ce cours on ne considérera que ce risque.

Risque quadratique

Risque quadratique

On appelle risque quadratique le risque associé à la fonction de perte

$$L(\eta, \eta') = \|\eta - \eta'\|^2,$$

c'est-à-dire :

$$R_\theta\left(\hat{\eta}\right) = \mathbb{E}_\theta\left(\|g(\theta) - \hat{\eta}\|^2\right).$$

Remarques

- Aussi appelée « erreur quadratique moyenne » (EQM).
- C'est le risque le plus couramment utilisé
  (pour des raison de simplicité, comme on va le voir) ;
- dans la suite de ce cours on ne considérera que ce risque.

Exemple 1 (rappel)

n-échantillon iid gaussien : $\underline{X} = (X_1, \dots X_n)$ avec

- $X_1, X_2, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$,
- $\theta = \left(\mu, \sigma^2\right)$,
- $\Theta = \mathbb{R} \times]0; +\infty[$.

Dans cet exemple on supposera qu'on veut estimer la moyenne μ ;

- ici $\eta = \mu$ et $g : \theta = \left(\mu, \sigma^2\right) \mapsto \mu$,
- σ^2 est inconnu aussi (paramètre de nuisance).

Exemple 1 : risque de l'estimateur $\hat{\mu}_1$

Considérons l'estimateur

$$\hat{\mu}_1 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Pour tout $\theta = (\mu, \sigma^2) \in \Theta$, on a le résultat suivant :

Risque quadratique de la moyenne empirique

$$R_\theta(\hat{\mu}_1) = \mathbb{E}_\theta\left((\hat{\mu}_1 - \mu)^2\right) = \frac{\sigma^2}{n}.$$

Remarque : le résultat reste valable dès que les X_i admettent des moments d'ordre 2

(on n'utilise pas le caractère gaussien)

Exemple 1 : risque de l'estimateur $\hat{\mu}_1$

Considérons l'estimateur

$$\hat{\mu}_1 = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Pour tout $\theta = \left(\mu, \sigma^2\right) \in \Theta$, on a le résultat suivant :

Risque quadratique de la moyenne empirique

$$R_\theta\left(\hat{\mu}_1\right) = \mathbb{E}_\theta\left(\left(\hat{\mu}_1 - \mu\right)^2\right) = \frac{\sigma^2}{n}.$$

Remarque : le résultat reste valable dès que les X_i admettent des moments d'ordre 2

(on n'utilise pas le caractère gaussien)

Exemple 1 : risque de l'estimateur $\hat{\mu}_1$ (calcul)

On observe que

$$\mathbb{E}_\theta(\hat{\mu}_1) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i) = \mu.$$

Donc

$$\begin{aligned} R_\theta\left(\hat{\mu}_1\right) = \operatorname{var}_\theta\left(\hat{\mu}_1\right) &= \frac{1}{n^2}\operatorname{var}_\theta\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}_\theta\left(X_i\right) = \frac{\sigma^2}{n} \end{aligned}$$

Exemple 1 : risque de l'estimateur $\hat{\mu}_1$ (calcul)

On observe que

$$\mathbb{E}_\theta(\hat{\mu}_1) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta(X_i) = \mu.$$

Donc

$$\begin{aligned} R_\theta\left(\hat{\mu}_1\right) = \mathrm{var}_\theta\left(\hat{\mu}_1\right) &= \frac{1}{n^2}\,\mathrm{var}_\theta\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}_\theta\left(X_i\right) = \frac{\sigma^2}{n} \end{aligned}$$

$\square$

Biais d'un estimateur

Soit $\hat{\eta}$ un estimateur de $\eta = g(\theta)$ admettant un moment d'ordre 1, $\forall \theta \in \Theta$.

Définitions : biais / estimateur sans biais

On définit le biais de l'estimateur $\hat{\eta}$ au point $\theta \in \Theta$ par

$$\mathrm{b}_\theta(\hat{\eta}) = \mathbb{E}_\theta(\hat{\eta}) - g(\theta).$$

On dit que $\hat{\eta}$ est un estimateur sans biais (ESB) si

$$\mathrm{b}_\theta(\hat{\eta}) = 0, \qquad \forall \theta \in \Theta.$$

Exemple 1

- On a déjà vu que $\hat{\mu}_1 = \bar{X}_n$ est un ESB de μ.
- Plus généralement, on note que $\hat{\mu} = \alpha + \beta \bar{X}_n$ est un ESB de μ $\iff \alpha = 0$ et $\beta = 1$.

Biais d'un estimateur

Soit $\hat{\eta}$ un estimateur de $\eta = g(\theta)$ admettant un moment d'ordre 1, $\forall \theta \in \Theta$.

Définitions : biais / estimateur sans biais

On définit le biais de l'estimateur $\hat{\eta}$ au point $\theta \in \Theta$ par

$$\mathrm{b}_\theta(\hat{\eta}) = \mathbb{E}_\theta(\hat{\eta}) - g(\theta).$$

On dit que $\hat{\eta}$ est un estimateur sans biais (ESB) si

$$\mathrm{b}_\theta(\hat{\eta}) = 0, \qquad \forall \theta \in \Theta.$$

Exemple 1

- On a déjà vu que $\hat{\mu}_1 = \bar{X}_n$ est un ESB de μ.
- Plus généralement, on note que $\hat{\mu} = \alpha + \beta \bar{X}_n$ est un ESB de μ $\iff \alpha = 0$ et $\beta = 1$.

Décomposition biais-variance

Rappel : on considère toujours le risque quadratique.

Soit $\hat{\eta}$ un estimateur de $\eta = g(\theta)$, admettant un moment d'ordre 2, $\forall \theta \in \Theta$.

Proposition : Décomposition biais-variance (cas scalaire)

Si le paramètre à estimer est scalaire ($\eta \in \mathbb{R}$), on a :

$$R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left((\hat{\eta} - g(\theta))^2\right) = \mathrm{var}_\theta(\hat{\eta}) + \mathrm{b}_\theta(\hat{\eta})^2.$$

exercice 1

Remarque : en sommant sur le composantes, on généralise au cas vectoriel :

$$R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left(\|\hat{\eta} - g(\theta)\|^2\right) = \mathrm{tr}\left(\mathrm{var}_\theta(\hat{\eta})\right) + \|\mathrm{b}_\theta(\hat{\eta})\|^2,$$

où $\mathrm{var}_\theta(\hat{\eta})$ est la matrice de covariance de $\hat{\eta}$.

Décomposition biais-variance

Rappel : on considère toujours le risque quadratique.

Soit $\hat{\eta}$ un estimateur de $\eta = g(\theta)$, admettant un moment d'ordre 2, $\forall \theta \in \Theta$.

Proposition : Décomposition biais-variance (cas scalaire)

Si le paramètre à estimer est scalaire ($\eta \in \mathbb{R}$), on a :

$$R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left((\hat{\eta} - g(\theta))^2\right) = \mathrm{var}_\theta(\hat{\eta}) + \mathrm{b}_\theta(\hat{\eta})^2.$$

exercice 1

Remarque : en sommant sur le composantes, on généralise au cas vectoriel :

$$R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left(\|\hat{\eta} - g(\theta)\|^2\right) = \mathrm{tr}\left(\mathrm{var}_\theta(\hat{\eta})\right) + \|\mathrm{b}_\theta(\hat{\eta})\|^2,$$

où $\mathrm{var}_\theta(\hat{\eta})$ est la matrice de covariance de $\hat{\eta}$.

Exemple 1 : risque de quelques estimateurs

$$\hat{\mu}_1 = \bar{X}_n \qquad R_\theta(\hat{\mu}_1) = \frac{\sigma^2}{n} + 0^2$$

$$\hat{\mu}_2 = \mu_0 \qquad R_\theta(\hat{\mu}_2) = 0^2 + (\mu - \mu_0)^2$$

$$\hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n \qquad R_\theta(\hat{\mu}_3) = \frac{1}{4}\frac{\sigma^2}{n} + \frac{1}{4}(\mu - \mu_0)^2$$

$$\hat{\mu}_4 = \bar{X}_n + c \qquad R_\theta(\hat{\mu}_4) = \frac{\sigma^2}{n} + c^2$$

$$\hat{\mu}_5 = \mathrm{med}(X_1, \ldots, X_n) \qquad R_\theta(\hat{\mu}_5) \approx 1.57\frac{\sigma^2}{n} + 0^2 \quad (n \to +\infty)$$

Exercice : Calculer $R_\theta(\hat{\mu}_j)$, $2 \leq j \leq 4$ exercice 1

Remarque : seul le résultat pour $\hat{\mu}_5$ utilise le caractère gaussien.

Estimateurs admissibles

Définition : relation d'ordre sur l'ensemble des estimateurs

On dira que $\hat{\eta}'$ est préférable (au sens large) à $\hat{\eta}$ si

- $\forall \theta \in \Theta,\ R_\theta(\hat{\eta}') \leq R_\theta(\hat{\eta})$,

On dira qu'il est strictement préférable à $\hat{\eta}$ si, de plus,

- $\exists \theta \in \Theta,\ R_\theta(\hat{\eta}') < R_\theta(\hat{\eta})$.

Remarques

- La relation « préférable à » est un ordre partiel sur les fonctions de risque.
- Il n'existe pas en général un estimateur optimal, càd un estimateur préférable à tous les autres (sauf à restreindre la classe d'estimateurs considérés).

Admissibilité

On dit que $\hat{\eta}$ est un estimateur admissible s'il n'existe pas un estimateur $\hat{\eta}'$ qui lui est strictement préférable.

Estimateurs admissibles

Définition : relation d'ordre sur l'ensemble des estimateurs

On dira que $\hat{\eta}'$ est préférable (au sens large) à $\hat{\eta}$ si

- $\forall \theta \in \Theta,\ R_\theta(\hat{\eta}') \leq R_\theta(\hat{\eta})$,

On dira qu'il est strictement préférable à $\hat{\eta}$ si, de plus,

- $\exists \theta \in \Theta,\ R_\theta(\hat{\eta}') < R_\theta(\hat{\eta})$.

Remarques

- La relation « préférable à » est un ordre partiel sur les fonctions de risque.
- Il n'existe pas en général un estimateur optimal, càd un estimateur préférable à tous les autres (sauf à restreindre la classe d'estimateurs considérés).

Admissibilité

On dit que $\hat{\eta}$ est un estimateur admissible s'il n'existe pas un estimateur $\hat{\eta}'$ qui lui est strictement préférable.

Exemple 1 (suite)

$$\hat{\mu}_1 = \bar{X}_n \qquad R_\theta(\hat{\mu}_1) = \frac{\sigma^2}{n} + 0^2$$

$$\hat{\mu}_2 = \mu_0 \qquad R_\theta(\hat{\mu}_2) = 0^2 + (\mu - \mu_0)^2$$

$$\hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n \qquad R_\theta(\hat{\mu}_3) = \frac{1}{4}\frac{\sigma^2}{n} + \frac{1}{4}(\mu - \mu_0)^2$$

$$\hat{\mu}_4 = \bar{X}_n + c \qquad R_\theta(\hat{\mu}_4) = \frac{\sigma^2}{n} + c^2$$

- $\hat{\mu}_1$ est strictement préférable à $\hat{\mu}_4$, donc $\hat{\mu}_4$ n'est pas admissible.
- On peut montrer
  - que $\hat{\mu}_1$, $\hat{\mu}_2$, et $\hat{\mu}_3$ sont deux à deux incomparables, exercice 1
  - mais qu'ils sont tous les trois admissibles (preuve hors-programme)

Exemple 1 (suite)

$$\hat{\mu}_1 = \bar{X}_n \qquad R_\theta(\hat{\mu}_1) = \frac{\sigma^2}{n} + 0^2$$

$$\hat{\mu}_2 = \mu_0 \qquad R_\theta(\hat{\mu}_2) = 0^2 + (\mu - \mu_0)^2$$

$$\hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n \qquad R_\theta(\hat{\mu}_3) = \frac{1}{4}\frac{\sigma^2}{n} + \frac{1}{4}(\mu - \mu_0)^2$$

$$\hat{\mu}_4 = \bar{X}_n + c \qquad R_\theta(\hat{\mu}_4) = \frac{\sigma^2}{n} + c^2$$

- $\hat{\mu}_1$ est strictement préférable à $\hat{\mu}_4$, donc $\hat{\mu}_4$ n'est pas admissible.
- On peut montrer
  - que $\hat{\mu}_1$, $\hat{\mu}_2$, et $\hat{\mu}_3$ sont deux à deux incomparables, exercice 1
  - mais qu'ils sont tous les trois admissibles (preuve hors-programme)

Exemple 1 (suite)

$$\hat{\mu}_1 = \bar{X}_n \qquad R_\theta(\hat{\mu}_1) = \frac{\sigma^2}{n} + 0^2$$

$$\hat{\mu}_2 = \mu_0 \qquad R_\theta(\hat{\mu}_2) = 0^2 + (\mu - \mu_0)^2$$

$$\hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n \qquad R_\theta(\hat{\mu}_3) = \frac{1}{4}\frac{\sigma^2}{n} + \frac{1}{4}(\mu - \mu_0)^2$$

$$\hat{\mu}_4 = \bar{X}_n + c \qquad R_\theta(\hat{\mu}_4) = \frac{\sigma^2}{n} + c^2$$

- $\hat{\mu}_1$ est strictement préférable à $\hat{\mu}_4$, donc $\hat{\mu}_4$ n'est pas admissible.
- On peut montrer
  - que $\hat{\mu}_1$, $\hat{\mu}_2$, et $\hat{\mu}_3$ sont deux à deux incomparables, exercice 1
  - mais qu'ils sont tous les trois admissibles (preuve hors-programme)

Plan du cours

Contexte et motivation

On se place dans la classe des estimateurs sans biais de $g(\theta)$,

➠ pour un ESB, $R_\theta(\hat{\eta}) = \mathrm{var}_\theta(\hat{\eta})$.

Objectif de cette partie : montrer qu'il existe une borne de la forme

$$\mathrm{var}_\theta(\hat{\eta}) \;\geq\; v_{\min}(\theta), \qquad \forall \theta \in \Theta,$$

valable pour (presque) **tous** les ESB de $g(\theta)$.

Utilité d'une telle borne ?

1. Établir qu'un certain niveau de précision n'est pas atteignable par un estimateur sans biais.
2. Établir qu'un ESB donné est optimal (c.-à-d. minimise, dans la classe des ESB, le risque $R_\theta(\hat{\eta}), \forall \theta \in \Theta$).

Contexte et motivation

On se place dans la classe des estimateurs sans biais de $g(\theta)$,

➠ pour un ESB, $R_\theta(\hat{\eta}) = \mathrm{var}_\theta(\hat{\eta})$.

Objectif de cette partie : montrer qu'il existe une borne de la forme

$$\mathrm{var}_\theta(\hat{\eta}) \;\geq\; v_{\min}(\theta), \qquad \forall \theta \in \Theta,$$

valable pour (presque) **tous** les ESB de $g(\theta)$.

Utilité d'une telle borne ?

1. Établir qu'un certain niveau de précision n'est pas atteignable par un estimateur sans biais.
2. Établir qu'un ESB donné est optimal (c.-à-d. minimise, dans la classe des ESB, le risque $R_\theta(\hat{\eta}), \forall \theta \in \Theta$).

Contexte et motivation

On se place dans la classe des estimateurs sans biais de $g(\theta)$,

➠ pour un ESB, $R_\theta(\hat{\eta}) = \mathsf{var}_\theta(\hat{\eta})$.

Objectif de cette partie : montrer qu'il existe une borne de la forme

$$\mathsf{var}_\theta(\hat{\eta}) \;\geq\; v_{\mathsf{min}}(\theta), \qquad \forall\theta \in \Theta,$$

valable pour (presque) **tous** les ESB de $g(\theta)$.

Utilité d'une telle borne ?

1. Établir qu'un certain niveau de précision n'est pas atteignable par un estimateur sans biais.
2. Établir qu'un ESB donné est optimal (c.-à-d. minimise, dans la classe des ESB, le risque $R_\theta(\hat{\eta}), \forall\theta \in \Theta$).

Condition de régularité C_0 et C_1

Condition de régularité C_0

Modèle dominé : il existe une mesure (σ-finie) ν sur $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$, et une famille (f_θ) de densité de probabilité par rapport à ν, tq

$$\forall A \in \underline{\mathscr{A}}, \quad \mathbb{P}_\theta\left(\underline{X} \in A\right) = \int_A f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}).$$

Condition de régularité C_1

Les densités f_θ ont même support : $\exists \mathcal{S} \in \underline{\mathscr{A}}$,

$$\forall \theta \in \Theta, \quad \mathbb{1}_{f_\theta > 0} = \mathbb{1}_{\mathcal{S}} \quad \nu\text{-pp.}$$

- En conséquence, on peut supposer que $f_\theta(x) > 0 \ \Leftrightarrow \ x \in \mathcal{S}$.

Condition de régularité C_0 et C_1

Condition de régularité C_0

Modèle dominé : il existe une mesure (σ-finie) ν sur $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$, et une famille (f_θ) de densité de probabilité par rapport à ν, tq

$$\forall A \in \underline{\mathscr{A}}, \quad \mathbb{P}_\theta\left(\underline{X} \in A\right) = \int_A f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}).$$

Condition de régularité C_1

Les densités f_θ ont même support : $\exists \mathcal{S} \in \underline{\mathscr{A}}$,

$$\forall \theta \in \Theta, \quad \mathbb{1}_{f_\theta > 0} = \mathbb{1}_{\mathcal{S}} \quad \nu\text{-pp.}$$

- En conséquence, on peut supposer que $f_\theta(x) > 0 \;\Leftrightarrow\; x \in \mathcal{S}$.

Condition de régularité C_1 : exemples / contre-exemple

Considérons un n-échantillon IID, univarié :

$$\underline{X} \sim f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i)$$

(avec un abus de notation usuel pour la notation des densités).

Remarque : si C_1 est vraie pour $n = 1$ avec $\mathcal{S} = \mathcal{S}_1$,
alors elle est vraie aussi pour tout $n \geq 2$ avec $\mathcal{S} = \mathcal{S}_1^n$.

Quelques exemples. . .

1. $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma^2 > 0$: C_1 est vraie avec $\mathcal{S}_1 = \mathbb{R}$,
2. $\mathcal{E}(\theta)$: C_1 est vraie avec $\mathcal{S}_1 = [0, +\infty)$.
3. $\mathcal{U}_{[0,\theta]}$: C_1 n'est pas vraie !

Condition de régularité C_1 : exemples / contre-exemple

Considérons un n-échantillon IID, univarié :

$$\underline{X} \sim f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i)$$

(avec un abus de notation usuel pour la notation des densités).

Remarque : si C_1 est vraie pour $n = 1$ avec $\mathcal{S} = \mathcal{S}_1$,
alors elle est vraie aussi pour tout $n \geq 2$ avec $\mathcal{S} = \mathcal{S}_1^n$.

Quelques exemples...

1. $\mathscr{N}(\mu, \sigma^2)$ avec $\sigma^2 > 0$: C_1 est vraie avec $\mathcal{S}_1 = \mathbb{R}$,
2. $\mathcal{E}(\theta)$: C_1 est vraie avec $\mathcal{S}_1 = [0, +\infty)$.
3. $\mathcal{U}_{[0,\theta]}$: C_1 n'est pas vraie !

Une autre condition de régularité

On suppose que C_0 et C_1 sont vraies.

Condition de régularité C_2

- i. Θ est un ouvert de $\mathbb{R}^p$,
- ii. $\theta \mapsto f_\theta(\underline{x})$ est différentiable ν-presque pour tout $\underline{x}$,
- iii. et, en tout point $\theta \in \Theta$, on a

$$\int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = \nabla_\theta \int_{\mathcal{S}} f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = 0.$$

Autrement dit : $\forall \theta \in \Theta$, $\forall k \leq p$,

$$\int_{\mathcal{S}} \frac{\partial f_\theta(\underline{x})}{\partial \theta_k}\, \nu(\mathrm{d}\underline{x}) = \frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = 0.$$

Une autre condition de régularité

On suppose que C_0 et C_1 sont vraies.

Condition de régularité C_2

1. Θ est un ouvert de $\mathbb{R}^p$,
2. $\theta \mapsto f_\theta(\underline{x})$ est différentiable ν-presque pour tout $\underline{x}$,
3. et, en tout point $\theta \in \Theta$, on a

$$\int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = \nabla_\theta \int_{\mathcal{S}} f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = 0.$$

Autrement dit : $\forall \theta \in \Theta$, $\forall k \leq p$,

$$\int_{\mathcal{S}} \frac{\partial f_\theta(\underline{x})}{\partial \theta_k}\, \nu(\mathrm{d}\underline{x}) = \frac{\partial}{\partial \theta_k} \int_{\mathcal{S}} f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = 0.$$

Le vecteur score

Définition / propriété : score

Supposons C_0, C_1, C_2-i et C_2-ii vérifiées, et posons $\forall \underline{x} \in \mathcal{S}$

$$S_\theta(\underline{x}) = \nabla_\theta \left(\ln f_\theta(\underline{x}) \right) = \begin{pmatrix} \frac{\partial \ln f_\theta(\underline{x})}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ln f_\theta(\underline{x})}{\partial \theta_p} \end{pmatrix}.$$

Alors

i. On appelle score le vecteur aléatoire $S_\theta = S_\theta(\underline{X})$.
ii. C_2-iii $\Leftrightarrow \forall \theta \in \Theta$, le score S_θ est centré sous $\mathbb{P}_\theta$.

Remarques :

- Bien défini puisque $\underline{X} \in \mathcal{S}$ $\mathbb{P}_\theta$-ps, $\forall \theta \in \Theta$.
- L'estimateur du maximum de vraisemblance annule le score (rappel : on suppose $\Theta \subset \mathbb{R}^p$ ouvert).

Le vecteur score (démonstration)

Remarquons que

$$\nabla_\theta\left(\ln f_\theta\right) = \frac{1}{f_\theta}\,\nabla_\theta f_\theta,$$

et donc, pour tout $\theta \in \Theta$,

$$\begin{aligned}
\mathbb{E}_\theta\left(S_\theta\right) &= \int_{\mathcal{S}} S_\theta(\underline{x})\, f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x}) \\
&= \int_{\mathcal{S}} \frac{1}{f_\theta(\underline{x})}\,\nabla_\theta f_\theta(\underline{x})\, f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x}) \\
&= \int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x}).
\end{aligned}$$

Ainsi,

$$\mathbb{E}_\theta\left(S_\theta\right) = 0 \quad \Leftrightarrow \quad \int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x}) = 0 \quad (\mathrm{C}_2\text{-iii}). \quad \square$$

Exemple 2

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$ avec $\theta \in \Theta =]0, +\infty[$.

On calcule la vraisemblance, pour $x_1, \ldots, x_n \geq 0$:

$$\mathcal{L}(\theta; \underline{x}) = f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \theta^n \, e^{-\theta \sum x_i},$$

puis la log-vraisemblance :

$$\ln \mathcal{L}(\theta; \underline{x}) = \ln f_\theta(\underline{x}) = n \ln \theta - \theta \sum x_i,$$

et enfin le score :

$$S_\theta(\underline{X}) = \sum_{i=1}^{n} S_\theta(X_i) = n \left(\frac{1}{\theta} - \bar{X}_n \right).$$

Exemple 2

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$ avec $\theta \in \Theta =]0, +\infty[$.

On calcule la vraisemblance, pour $x_1, \ldots, x_n \geq 0$:

$$\mathcal{L}(\theta; \underline{x}) = f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \theta^n \, e^{-\theta \sum x_i},$$

puis la log-vraisemblance :

$$\ln \mathcal{L}(\theta; \underline{x}) = \ln f_\theta(\underline{x}) = n \ln \theta - \theta \sum x_i,$$

et enfin le score :

$$S_\theta(\underline{X}) = \sum_{i=1}^{n} S_\theta(X_i) = n \left(\frac{1}{\theta} - \bar{X}_n \right).$$

Exemple 2

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$ avec $\theta \in \Theta =]0, +\infty[$.

On calcule la vraisemblance, pour $x_1, \ldots, x_n \geq 0$:

$$\mathcal{L}(\theta; \underline{x}) = f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \theta^n \, e^{-\theta \sum x_i},$$

puis la log-vraisemblance :

$$\ln \mathcal{L}(\theta; \underline{x}) = \ln f_\theta(\underline{x}) = n \ln \theta - \theta \sum x_i,$$

et enfin le score :

$$S_\theta(\underline{X}) = \sum_{i=1}^{n} S_\theta(X_i) = n \left(\frac{1}{\theta} - \bar{X}_n \right).$$

Remarque sur la condition C_2-iii

Rappel de C_2-iii : $\forall\theta \in \Theta$,

$$\int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = \nabla_\theta \int_{\mathcal{S}} f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = 0,$$

ou, de manière équivalente : $\mathbb{E}_\theta(S_\theta) = 0$.

Deux approches pour vérifier cette condition :

1. Calculer explicitement $\mathbb{E}_\theta(S_\theta) = \int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})$.
2. Utiliser une condition de domination : mq $\forall\theta_0 \in \Theta$, $\exists\mathscr{V} \subset \Theta$ voisinage de θ_0 et $g : \underline{\mathcal{X}} \to \mathbb{R}$ une fonction ν-intégrable tq

$$\forall\theta \in \mathscr{V},\ \forall\underline{x} \in \mathcal{S},\ \forall k \leq p, \quad \left|\frac{\partial f_\theta(\underline{x})}{\partial\theta_k}\right| \leq g(\underline{x}).$$

Remarque sur la condition C_2-iii

Rappel de C_2-iii : $\forall \theta \in \Theta$,

$$\int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = \nabla_\theta \int_{\mathcal{S}} f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = 0,$$

ou, de manière équivalente : $\mathbb{E}_\theta(S_\theta) = 0$.

Deux approches pour vérifier cette condition :

1. Calculer explicitement $\mathbb{E}_\theta(S_\theta) = \int_{\mathcal{S}} \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})$.
2. Utiliser une condition de domination : mq $\forall \theta_0 \in \Theta$, $\exists \mathscr{V} \subset \Theta$ voisinage de θ_0 et $g : \underline{\mathcal{X}} \to \mathbb{R}$ une fonction ν-intégrable tq

$$\forall \theta \in \mathscr{V},\ \forall \underline{x} \in \mathcal{S},\ \forall k \leq p, \quad \left| \frac{\partial f_\theta(\underline{x})}{\partial \theta_k} \right| \leq g(\underline{x}).$$

Soit un modèle statistique vérifiant C_0-C_2,
et soit $\hat{\eta}$ un estimateur de $\eta = g(\theta) \in \mathbb{R}$.

Définition : estimateur régulier

On dit que $\hat{\eta}$ est régulier si

1. $\mathbb{E}_\theta\left(\hat{\eta}^2\right) < +\infty$, $\forall \theta \in \Theta$,
2. $\theta \mapsto \mathbb{E}_\theta\left(\hat{\eta}\right)$ est différentiable, avec

$$\nabla_\theta \mathbb{E}_\theta\left(\hat{\eta}\right) = \int_{\mathcal{S}} \hat{\eta}(\underline{x})\, \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}), \qquad \forall \theta \in \Theta.$$

Remarque : si $\hat{\eta}$ est un ESB régulier de $g(\theta)$, alors

$$(\nabla g)(\theta) = \int_{\mathcal{S}} \hat{\eta}(\underline{x})\, \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}), \qquad \forall \theta \in \Theta.$$

Théorème / définition : inégalité de Cramér-Rao

Soit un modèle statistique vérifiant C_0-C_2, dans lequel le score S_θ admet des moments d'ordre deux pour tout $\theta \in \Theta$.

Soit $\mathrm{var}_\theta(S_\theta)$ la matrice de covariance du score, que l'on suppose inversible pour tout $\theta \in \Theta$.

Soit $\hat{\eta}$ est un ESB régulier de $g(\theta)$. Alors, $\forall \theta \in \Theta$,

$$R_\theta(\hat{\eta}) = \mathrm{var}_\theta(\hat{\eta}) \geq \nabla g(\theta)^\top \mathrm{var}_\theta(S_\theta)^{-1} \nabla g(\theta).$$

Un ESB est dit efficace s'il atteint cette borne pour tout θ.

Démonstration

Inégalité de Cramér-Rao 2/2

Théorème / définition : inégalité de Cramér-Rao

Soit un modèle statistique vérifiant C_0-C_2, dans lequel le score S_θ admet des moments d'ordre deux pour tout $\theta \in \Theta$.

Soit $\mathrm{var}_\theta(S_\theta)$ la matrice de covariance du score, que l'on suppose inversible pour tout $\theta \in \Theta$.

Soit $\hat{\eta}$ est un ESB régulier de $g(\theta)$. Alors, $\forall \theta \in \Theta$,

$$R_\theta\left(\hat{\eta}\right) = \mathrm{var}_\theta\left(\hat{\eta}\right) \;\geq\; \nabla g(\theta)^\top \, \mathrm{var}_\theta\left(S_\theta\right)^{-1} \nabla g(\theta).$$

Un ESB est dit efficace s'il atteint cette borne pour tout θ.

Démonstration

Information de Fisher

On suppose toujours les conditions C_0–C_2.

Définition : information de Fisher

On appelle information de Fisher apportée par $\underline{X}$ la matrice $p \times p$

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

qui apparaît dans la borne de Cramér-Rao.

Proposition

Soit $I_n(\theta)$ l'information de Fisher dans un n-échantillon IID. Alors

$$I_n(\theta) = n\, I_1(\theta).$$

L'inégalité de CR s'écrit alors : $\mathrm{var}_\theta(\hat{\eta}) \geq \frac{1}{n} \nabla g(\theta)^\top I_1(\theta)^{-1} \nabla g(\theta)$.

Information de Fisher

On suppose toujours les conditions C_0–C_2.

Définition : information de Fisher

On appelle information de Fisher apportée par $\underline{X}$ la matrice $p \times p$

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

qui apparaît dans la borne de Cramér-Rao.

Proposition

Soit $I_n(\theta)$ l'information de Fisher dans un n-échantillon IID. Alors

$$I_n(\theta) = n\, I_1(\theta).$$

L'inégalité de CR s'écrit alors : $\mathrm{var}_\theta(\hat{\eta}) \geq \frac{1}{n} \nabla g(\theta)^\top I_1(\theta)^{-1} \nabla g(\theta)$.

Démonstration

On remarque que le score est additif dans un échantillon IID :

$$\begin{aligned} S_\theta &= \nabla_\theta \left(\ln f_\theta(\underline{x}) \right) \\ &= \nabla_\theta \left[\ln \left(\prod_{i=1}^{n} f_\theta^{X_1}(X_i) \right) \right] = \sum_{i=1}^{n} \underbrace{\nabla_\theta \left(\ln f_\theta^{X_1}(X_i) \right)}_{Z_i} . \end{aligned}$$

On a donc

$$\mathrm{var}_\theta\left(S_\theta\right) = \sum_{i=1}^{n} \mathrm{var}_\theta\left(Z_i\right) = n\ \mathrm{var}_\theta\left(Z_1\right) = n\, I_1(\theta)$$

puisque $Z_1, \ldots, Z_n$ sont IID, et distribuées comme le score dans un échantillon de taille 1. □

Démonstration

On remarque que le score est additif dans un échantillon IID :

$$\begin{aligned} S_\theta &= \nabla_\theta \left(\ln f_\theta(\underline{x})\right) \\ &= \nabla_\theta \left[\ln \left(\prod_{i=1}^{n} f_\theta^{X_1}(X_i)\right)\right] = \sum_{i=1}^{n} \underbrace{\nabla_\theta \left(\ln f_\theta^{X_1}(X_i)\right)}_{Z_i}. \end{aligned}$$

On a donc

$$\mathrm{var}_\theta\left(S_\theta\right) = \sum_{i=1}^{n} \mathrm{var}_\theta\left(Z_i\right) = n\ \mathrm{var}_\theta\left(Z_1\right) = n\, I_1(\theta)$$

puisque $Z_1, \ldots, Z_n$ sont IID, et distribuées comme le score dans un échantillon de taille 1. $\square$

Exemple 1 : estimation de μ

Rappels : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ et $\theta = (\mu, \sigma^2)$

- $\hat{\mu}_n = \bar{X}_n$ est l'EMV de μ,
- $\hat{\mu}_n$ est sans biais et $R_\theta(\hat{\mu}_n) = \text{var}_\theta(\hat{\mu}_n) = \frac{\sigma^2}{n}$.

Un peu de calcul (voir TD 2) permet de montrer que l'information de Fisher dans ce modèle vaut

$$I_n(\theta) = n \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Inégalité de Cramér-Rao avec $g(\theta) = \mu$: $\forall \hat{\mu}'_n$ ESB de μ,

$$R_\theta(\hat{\mu}'_n) = \text{var}_\theta(\hat{\mu}'_n) \geq \frac{\sigma^2}{n},$$

donc $\hat{\mu}_n = \bar{X}_n$ est efficace.

Exemple 1 : estimation de μ

Rappels : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ et $\theta = (\mu, \sigma^2)$

- $\hat{\mu}_n = \bar{X}_n$ est l'EMV de μ,
- $\hat{\mu}_n$ est sans biais et $R_\theta(\hat{\mu}_n) = \text{var}_\theta(\hat{\mu}_n) = \frac{\sigma^2}{n}$.

Un peu de calcul (voir TD 2) permet de montrer que l'information de Fisher dans ce modèle vaut

$$I_n(\theta) = n \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Inégalité de Cramér-Rao avec $g(\theta) = \mu$: $\forall \hat{\mu}'_n$ ESB de μ,

$$R_\theta(\hat{\mu}'_n) = \text{var}_\theta(\hat{\mu}'_n) \geq \frac{\sigma^2}{n},$$

donc $\hat{\mu}_n = \bar{X}_n$ est efficace.

Exemple 1' : estimation de σ^2

Même modèle statistique, mais on veut estimer $g(\theta) = \sigma^2$.

Un peu de calcul (voir TD 2) permet de montrer que

- l'EMV $S_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$ est biaisé ;
- $\hat{\sigma}_n^2 = (S_n')^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$ est un ESB de σ^2 et sa variance vaut :

$$\mathrm{var}_\theta\left(\hat{\sigma}_n^2\right) = \frac{2\,\sigma^4}{n-1}.$$

Conclusion : $\hat{\sigma}_n^2$ n'est pas un estimateur efficace, puisque

$$\mathrm{var}_\theta\left(\hat{\sigma}_n^2\right) > \frac{2\,\sigma^4}{n}.$$

(Attention terminologie trompeuse ! On peut montrer en utilisant le théorème de Lehmann-Scheffé que $\hat{\sigma}_n^2$ est un ESB de *variance minimale* pour ce problème, donc optimal pour le risque quadratique dans la classe des ESB.)

Exemple 1' : estimation de σ^2

Même modèle statistique, mais on veut estimer $g(\theta) = \sigma^2$.

Un peu de calcul (voir TD 2) permet de montrer que

- l'EMV $S_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$ est biaisé ;
- $\hat{\sigma}_n^2 = (S_n')^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$ est un ESB de σ^2 et sa variance vaut :

$$\mathrm{var}_\theta\left(\hat{\sigma}_n^2\right) = \frac{2\,\sigma^4}{n-1}.$$

Conclusion : $\hat{\sigma}_n^2$ n'est pas un estimateur efficace, puisque

$$\mathrm{var}_\theta\left(\hat{\sigma}_n^2\right) > \frac{2\,\sigma^4}{n}.$$

(Attention terminologie trompeuse ! On peut montrer en utilisant le théorème de Lehmann-Scheffé que $\hat{\sigma}_n^{\mathbf{2}}$ est un ESB de *variance minimale* pour ce problème, donc optimal pour le risque quadratique dans la classe des ESB.)

Plan du cours

Motivation / notations

Problème

Il est parfois (souvent !) difficile d'établir les propriétés *exactes* des procédures statistiques.

(estimateur ponctuel, mais aussi IC, tests, etc. (voir plus loin))

Approche(s) asymptotique(s) $\rightarrow$ propriétés approchées

- $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, définis sur un même $(\Omega, \mathscr{F}, \mathbb{P}_\theta)$
- Suite d'estimateurs : $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$
- Propriétés des estimateurs lorsque $n \to \infty$?

Remarque : on a ici non plus un mais une *suite* $(\mathscr{M}_n)_{n \geq 1}$ de modèles stat.,

$$\mathscr{M}_n = \left(\mathcal{X}^n, \mathscr{A}^{\otimes n}, \left\{\mathrm{P}_\theta^{\otimes n}, \, \theta \in \Theta\right\}\right),$$

que l'on réalise sur un même espace $(\Omega, \mathscr{F})$ sous-jacent.

Motivation / notations

Problème

Il est parfois (souvent !) difficile d'établir les propriétés *exactes* des procédures statistiques.

(estimateur ponctuel, mais aussi IC, tests, etc. (voir plus loin))

Approche(s) asymptotique(s) $\to$ propriétés approchées

- $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, définis sur un même $(\Omega, \mathscr{F}, \mathbb{P}_\theta)$
- Suite d'estimateurs : $\hat{\eta}_n = \hat{\eta}_n(X_1, \ldots, X_n)$
- Propriétés des estimateurs lorsque $n \to \infty$?

Remarque : on a ici non plus un mais une *suite* $(\mathscr{M}_n)_{n\geq 1}$ de modèles stat.,

$$\mathscr{M}_n = \left(\mathcal{X}^n, \mathscr{A}^{\otimes n}, \left\{\mathrm{P}_\theta^{\otimes n},\, \theta \in \Theta\right\}\right),$$

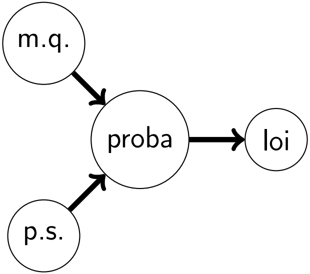
que l'on réalise sur un même espace $(\Omega, \mathscr{F})$ sous-jacent.

Rappel de proba : modes de convergence

Principaux modes de convergence utiles en stats :

- convergence presque sûre,
- convergence dans L^2 (en moyenne quadratique),
- convergence en probabilité,
- convergence en loi.

Implications entre certains modes de convergence : Compléments



Consistance

Soit $(\hat{\eta}_n)$ une suite d'estimateurs de $\eta = g(\theta)$.

Consistance (faible)

On dit que $\hat{\eta}_n$ est estimateur consistant de $\eta = g(\theta)$ si, $\forall \theta \in \Theta$,

$$\hat{\eta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta} g(\theta).$$

(notez l'abus de langage !)

Consistances forte et en moyenne quadratique

On dit que l'estimateur $\hat{\eta}_n$ est fortement consistant (resp. consistant en moyenne quadratique) si, $\forall \theta \in \Theta$,

$$\hat{\eta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta-\text{p.s.}} g(\theta) \qquad \left(\text{resp.,}\quad \hat{\eta}_n \xrightarrow[n\to\infty]{L^2(\mathbb{P}_\theta)} g(\theta)\right).$$

Remarque : on dit parfois « convergent » au lieu de « consistant ».

Consistance

Soit $(\hat{\eta}_n)$ une suite d'estimateurs de $\eta = g(\theta)$.

Consistance (faible)

On dit que $\hat{\eta}_n$ est estimateur consistant de $\eta = g(\theta)$ si, $\forall \theta \in \Theta$,

$$\hat{\eta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta} g(\theta).$$

(notez l'abus de langage !)

Consistances forte et en moyenne quadratique

On dit que l'estimateur $\hat{\eta}_n$ est fortement consistant (resp. consistant en moyenne quadratique) si, $\forall \theta \in \Theta$,

$$\hat{\eta}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta-\text{p.s.}} g(\theta) \qquad \left(\text{resp.,} \quad \hat{\eta}_n \xrightarrow[n\to\infty]{L^2(\mathbb{P}_\theta)} g(\theta)\right).$$

Remarque : on dit parfois « convergent » au lien de « consistant ».

Rappel de proba : loi des grands nombres

Soit $(X_k)_{k \geq 1}$ une suite de VA réelles ou vectorielles.

Loi forte des grands nombres

Si les X_k sont IID et admettent des moments d'ordre 1, alors

$$\bar{X}_n \xrightarrow[n \to \infty]{\text{p.s.}} \mathbb{E}(X_1).$$

Loi des grands nombres dans L^2

Si les X_k sont IID et admettent des moment d'ordre 2, alors

$$\bar{X}_n \xrightarrow[n \to \infty]{L^2} \mathbb{E}(X_1).$$

Preuve (cas scalaire) : $\mathbb{E}\left(\left(\bar{X}_n - \mathbb{E}(X_1)\right)^2\right) = \mathrm{var}_\theta(\bar{X}_n) = \frac{1}{n}\,\mathrm{var}_\theta(X_1) \to 0.$ $\square$

Rappel de proba : loi des grands nombres

Soit $(X_k)_{k\geq 1}$ une suite de VA réelles ou vectorielles.

Loi forte des grands nombres

Si les X_k sont IID et admettent des moments d'ordre 1, alors

$$\bar{X}_n \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}(X_1).$$

Loi des grands nombres dans L^2

Si les X_k sont IID et admettent des moment d'ordre 2, alors

$$\bar{X}_n \xrightarrow[n\to\infty]{L^2} \mathbb{E}(X_1).$$

Preuve (cas scalaire) : $\mathbb{E}\left(\left(\bar{X}_n - \mathbb{E}(X_1)\right)^2\right) = \mathrm{var}_\theta(\bar{X}_n) = \frac{1}{n}\,\mathrm{var}_\theta(X_1) \to 0.$ $\square$

Consistance : exemples

A) n-échantillon IID, dont la loi admet un moment d'ordre 1.

- i.e., $\mathbb{E}_\theta(\|X_1\|) < +\infty$, pour tout $\theta \in \Theta$.
- $\bar{X}_n$ est un estimateur fortement consistant de $\eta = \mathbb{E}_\theta(X_1)$.
- On ne peut rien dire sur le risque quadratique sans hypothèse supplémentaire (il peut même être infini).

B) n-échantillon IID, dont la loi admet un moment d'ordre 2.

- i.e., $\mathbb{E}_\theta(\|X_1\|^2) < +\infty$, pour tout $\theta \in \Theta$.
- $\bar{X}_n$ est un estimateur fortement consistant et consistant en moyenne quadratique de $\eta = \mathbb{E}_\theta(X_1)$.

Consistance : exemples

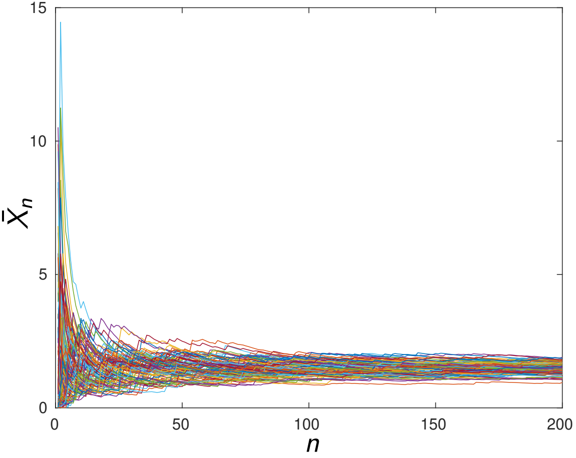
A) n-échantillon IID, dont la loi admet un moment d'ordre 1.

- i.e., $\mathbb{E}_\theta(\|X_1\|) < +\infty$, pour tout $\theta \in \Theta$.
- $\bar{X}_n$ est un estimateur fortement consistant de $\eta = \mathbb{E}_\theta(X_1)$.
- On ne peut rien dire sur le risque quadratique sans hypothèse supplémentaire (il peut même être infini).

B) n-échantillon IID, dont la loi admet un moment d'ordre 2.

- i.e., $\mathbb{E}_\theta(\|X_1\|^2) < +\infty$, pour tout $\theta \in \Theta$.
- $\bar{X}_n$ est un estimateur fortement consistant et consistant en moyenne quadratique de $\eta = \mathbb{E}_\theta(X_1)$.

Consistance : exemples (suite)



Convergence de $\bar{X}_n$ vers la vraie moyenne

(pour un n-échantillon de loi Gamma, ici de moyenne $\mu = 1.5$)

Consistance : exemples (suite)

C) n-échantillon IID (lois quelconques)

- Soit $A \in \mathscr{A}$ et $\eta = g(\theta) = \mathbb{P}_\theta(X_1 \in A)$.
- Fréquence : $\hat{\eta}_n = \frac{1}{n} \operatorname{card}\{i \leq n \mid X_i \in A\}$
- $\hat{\eta}_n$ est un estimateur fortement consistant et consistant en moyenne quadratique de η.

Application : histogrammes

exercice 3

D) EMV d'un n-échantillon IID uniforme (voir TD 1)

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{U}_{[0,\theta]}$
- On estime $\eta = \theta$ par $\hat{\eta}_n = \max_{i \leq n} X_i$.
- $\hat{\eta}_n$ est consistant, fortement et en mq.

E) Estimateur du maximum de vraisemblance

complément

Consistance : exemples (suite)

C) n-échantillon IID (lois quelconques)

- Soit $A \in \mathscr{A}$ et $\eta = g(\theta) = \mathbb{P}_\theta(X_1 \in A)$.
- Fréquence : $\hat{\eta}_n = \frac{1}{n} \operatorname{card}\{i \leq n \mid X_i \in A\}$
- $\hat{\eta}_n$ est un estimateur fortement consistant et consistant en moyenne quadratique de η.

Application : histogrammes

exercice 3

D) EMV d'un n-échantillon IID uniforme (voir TD 1)

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{U}_{[0,\theta]}$
- On estime $\eta = \theta$ par $\hat{\eta}_n = \max_{i \leq n} X_i$.
- $\hat{\eta}_n$ est consistant, fortement et en mq.

E) Estimateur du maximum de vraisemblance

complément

Consistance : exemples (suite)

C) n-échantillon IID (lois quelconques)

- Soit $A \in \mathscr{A}$ et $\eta = g(\theta) = \mathbb{P}_\theta\,(X_1 \in A)$.
- Fréquence : $\hat{\eta}_n = \frac{1}{n}$ card $\{i \leq n \mid X_i \in A\}$
- $\hat{\eta}_n$ est un estimateur fortement consistant et consistant en moyenne quadratique de η.

Application : histogrammes

exercice 3

D) EMV d'un n-échantillon IID uniforme (voir TD 1)

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{U}_{[0,\theta]}$
- On estime $\eta = \theta$ par $\hat{\eta}_n = \max_{i \leq n} X_i$.
- $\hat{\eta}_n$ est consistant, fortement et en mq.

E) Estimateur du maximum de vraisemblance

complément

Conclusion du cours et transition vers la séquence suivante

Nous avons vu et développerons en TD 2 :

- la quantification de la performance d'un estimateur via le calcul d'un risque,
- la comparaison d'estimateurs entre eux et une notion d'optimalité,
- l'étude asymptotique des estimateurs.

Nous aborderons dans la séquence 3 :

- la notion de vitesse de convergence d'un estimateur,
- la définition et la construction d'intervalle ou régions de confiance.

Conclusion du cours et transition vers la séquence suivante

Nous avons vu et développerons en TD 2 :

- la quantification de la performance d'un estimateur via le calcul d'un risque,
- la comparaison d'estimateurs entre eux et une notion d'optimalité,
- l'étude asymptotique des estimateurs.

Nous aborderons dans la séquence 3 :

- la notion de vitesse de convergence d'un estimateur,
- la définition et la construction d'intervalle ou régions de confiance.

Plan du cours

Plan du cours

Exercice 1 (risque quadratique)

corrigé

Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_*^+$.
On souhaite estimer $g(\theta) = \mu$. On s'intéresse aux estimateurs

$$\hat{\mu}_1 = \bar{X}_n, \qquad \hat{\mu}_2 = \mu_0, \qquad \hat{\mu}_3 = \frac{1}{2}\mu_0 + \frac{1}{2}\bar{X}_n, \qquad \hat{\mu}_4 = \bar{X}_n + c,$$

où μ_0 et c sont des réels fixés.

Questions

1. Montrer la formule de décomposition biais-variance dans le cas scalaire. retour au slide 17
2. Calculer le risque quadratique de chacun des estimateurs.
3. Montrer que $\hat{\mu}_2$ et $\hat{\mu}_3$ ne sont pas comparables.
4. Montrer que $\hat{\mu}_4$ n'est pas admissible. retour au slide 20

Exercice 2 (efficacité d'un estimateur)

corrigé

Soient $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(\theta)$ avec $\theta \in \Theta =]0,1[$.

On rappelle que (cf. exercices du cours 1) :

- la log-vraisemblance du n échantillon s'écrit

$$\ln \mathcal{L}(\theta; \underline{x}) = \ln f_\theta(\underline{x}) = n \ln(1-\theta) - \ln\left(\frac{\theta}{1-\theta}\right) \sum_{i=1}^{n} x_i,$$

- l'EMV s'écrit $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Questions

1. Vérifier que le modèle satisfait aux hypothèses de l'inégalité de Cramér-Rao, et calculer la borne de Cramér-Rao.
2. L'EMV $\hat{\theta}_n$ est-il efficace ?

Exercice 3 (consistance de l'histogramme)

corrigé

Considérons :

retour au slide 41

- un n-échantillon $X_1, \ldots, X_n$, avec X_i à valeurs dans $]a, b] \subset \mathbb{R}$,
- une partition de $]a, b]$ en K classes adjacentes $A_k =]a_{k-1}, a_k]$, pour $k \in \{1, \ldots, K\}$, avec $a_0 = a$, $a_K = b$,
- le vecteur $\eta \in \mathbb{R}^K$ avec $\eta^{(k)} = \mathrm{P}(X_1 \in A_k)$.

Histogramme

Représentation graphique de la répartition empirique d'une variable aléatoire à l'aide de rectangles, dont les bases sont les intervalles A_k et dont les surfaces sont proportionnelles aux fréquences $\hat{\eta}_n^{(k)}$ des classes :

$$\hat{\eta}_n^{(k)} = \frac{1}{n} \operatorname{card}\{i \leq n \mid X_i \in A_k\}, \qquad 1 \leq k \leq K.$$

Question. Montrer que $\hat{\eta}_n = \left(\hat{\eta}_n^{(1)}, \ldots, \hat{\eta}_n^{(K)}\right)$ est un estimateur de η fortement consistant et consistant en moyenne quadratique.

Exercice 4 (consistance en moyenne quadratique)

corrigé

Soit $\hat{\eta}_n$ un estimateur d'un paramètre scalaire $\eta = g(\theta) \in \mathbb{R}$, indicé par la taille n de l'échantillon observé.

Question

Montrer que $\hat{\eta}_n$ est consistant en moyenne quadratique si, et seulement si, les conditions suivantes sont vérifiées pour tout $\theta \in \Theta$:

i. $\mathrm{b}_\theta\left(\hat{\eta}_n\right) \to 0$,
ii. $\mathrm{var}_\theta\left(\hat{\eta}_n\right) \to 0$.

Plan du cours

Corrigé de l'exercice 1

retour au sujet

❶ Décomposition biais-variance

$$R_\theta(\hat\eta) = \mathbb{E}_\theta\left((\hat\eta - g(\theta))^2\right) = \mathrm{var}_\theta(\hat\eta) + \mathrm{b}_\theta(\hat\eta)^2.$$

Démonstration

$$\begin{aligned}
R_\theta(\hat\eta) &= \mathbb{E}_\theta\left((\hat\eta - g(\theta))^2\right) \\
&= \mathbb{E}_\theta\left((\hat\eta - \mathbb{E}_\theta(\hat\eta) + \mathrm{b}_\theta(\hat\eta))^2\right) \\
&= \underbrace{\mathbb{E}_\theta\left((\hat\eta - \mathbb{E}_\theta(\hat\eta))^2\right)}_{\mathrm{var}_\theta(\hat\eta)} + \mathrm{b}_\theta(\hat\eta)^2 + 2\,\underbrace{\mathbb{E}_\theta\left(\hat\eta - \mathbb{E}_\theta(\hat\eta)\right)}_{=0}\,\mathrm{b}_\theta(\hat\eta) \\
&= \mathrm{var}_\theta(\hat\eta) + \mathrm{b}_\theta(\hat\eta)^2.
\end{aligned}$$

□

Corrigé de l'exercice 1

retour au sujet

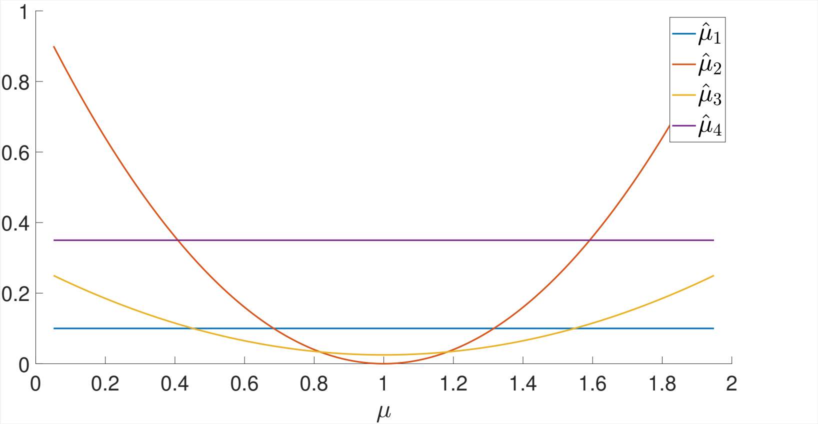
❷ On calcule le biais et la variance de l'estimateur, puis on conclut par la décomposition biais-variance.

	espérance	biais	variance	risque quadratique
$\bar{X}_n$	μ	0	$\frac{\sigma^2}{n}$	$\frac{\sigma^2}{n}$
μ_0	μ_0	$\mu_0 - \mu$	0	$(\mu_0 - \mu)^2$
$\frac{1}{2}\,(\mu_0 + \bar{X}_n)$	$\frac{1}{2}\,(\mu_0 + \mu)$	$\frac{1}{2}\,(\mu_0 - \mu)$	$\frac{1}{4}\,\frac{\sigma^2}{n}$	$\frac{1}{4}\,\frac{\sigma^2}{n} + \frac{1}{4}\,(\mu_0 - \mu)^2$
$\bar{X}_n + c$	$\mu + c$	c	$\frac{\sigma^2}{n}$	$\frac{\sigma^2}{n} + c^2$

Rappel : $\mathrm{var}_\theta(\alpha X + \beta) = \alpha^2\,\mathrm{var}_\theta(X)$.

Corrigé de l'exercice 1

retour au sujet



Tracé des 4 risques pour $\sigma^2 = 1$, $n = 10$, $\mu_0 = 1$ et $c = 0.5$.

❸ Calculons le risque en deux points bien choisis.

Pour $\theta = (\mu_0, 1)$ on a

$$R_\theta(\hat{\mu}_2) = 0, \qquad R_\theta(\hat{\mu}_3) = \frac{1}{4n}, \qquad \text{donc } R_\theta(\hat{\mu}_2) < R_\theta(\hat{\mu}_3).$$

Pour $\theta = \left(\mu_0 + \frac{1}{\sqrt{n}}, 1\right)$ on a

$$R_\theta(\hat{\mu}_2) = \frac{1}{n}, \qquad R_\theta(\hat{\mu}_3) = \frac{1}{2n}, \qquad \text{donc } R_\theta(\hat{\mu}_2) > R_\theta(\hat{\mu}_3).$$

Donc les estimateurs $\hat{\mu}_2$ et $\hat{\mu}_3$ ne sont pas comparables.

retour au sujet

❹ On a :

$$\begin{cases} R_\theta\left(\hat{\mu}_4\right) & = & \dfrac{\sigma^2}{n} + c^2 \\ \\ R_\theta\left(\hat{\mu}_1\right) & = & \dfrac{\sigma^2}{n} \end{cases}$$

Donc, $\forall \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_*^+$, $R_\theta\left(\hat{\mu}_4\right) > R_\theta\left(\hat{\mu}_1\right)$

On en déduit que $\hat{\mu}_4$ n'est pas admissible.

 retour au sujet

❶ On doit vérifier que le modèle vérifie les conditions de régularité C_1 et C_2, et que l'information de Fisher ne s'annule pas.

✏ C_1 : comme $\Theta =]0,1[$, les densités

$$f_\theta(\underline{x}) \;=\; \theta^{\sum_{i=1}^n x_i}\,(1-\theta)^{n-\sum_{i=1}^n x_i}$$

ont toutes pour support $\mathcal{S} = \{0,1\}^n$.

✏ C_2 : $\Theta =]0,1[$ est un ouvert de $\mathbb{R}$, $\theta \mapsto f_\theta(\underline{x})$ est dérivable sur Θ pour tout $\underline{x}$, et le score

$$S_\theta(\underline{X}) = \frac{\partial(\ln f_\theta)}{\partial\theta}(X_i) = \frac{n}{\theta(1-\theta)}\,\left(\bar{X}_n - \theta\right)$$

Corrigé de l'exercice 2

retour au sujet

est centré : $\mathbb{E}_\theta\left(S_\theta(\underline{X})\right) = \dfrac{n}{\theta(1-\theta)}\left(\mathbb{E}_\theta(\bar{X}_n) - \theta\right) = 0.$

✏ Enfin, on vérifie que l'information de Fisher ne s'annule pas :

$$I(\theta) = \mathrm{var}_\theta\left(S_\theta(\underline{X})\right) = \left(\frac{n}{\theta(1-\theta)}\right)^2 \mathrm{var}_\theta(\bar{X}_n) = \frac{n}{\theta(1-\theta)} > 0.$$

✏ La borne de Cramér-Rao pour θ vaut

$$I(\theta)^{-1} = \frac{1}{n}\theta(1-\theta).$$

Corrigé de l'exercice 2

retour au sujet

❷ L'estimateur $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est sans biais :

$$\mathbb{E}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta(X_1) = \theta,$$

et sa variance vaut

$$\mathrm{var}(\hat{\theta}) = \frac{1}{n}\mathrm{var}(X_1) = \frac{\theta(1-\theta)}{n} = I(\theta)^{-1}.$$

Il est donc efficace. □

Remarque : on peut vérifier facilement que $\hat{\theta}_n$ est un estimateur régulier (voir définition slide 29), puisque

- ⓐ la densité f_θ est dérivable par rapport à θ,
- ⓑ les intégrales se réduisent à des sommes finies sur $\{0,1\}^n$.

retour au sujet

❶ Consistance forte

Rappel : $\hat{\eta}_n \xrightarrow{\text{ps}} \eta$ ssi $\hat{\eta}_n^{(k)} \xrightarrow{\text{ps}} \eta^{(k)}$, $\forall k$.

Pour tout $k \in \{1, \ldots, K\}$, on a :

$$\hat{\eta}_n^{(k)} = \frac{1}{n} \operatorname{card}\{i \leq n \mid X_i \in A_k\} = \frac{1}{n} \sum_{i=1}^{n} Z_i \quad \text{avec } Z_i = 1_{A_k}(X_i).$$

La loi forte des grands nombres appliquée à $(Z_i)_{i \geq 1}$ donne alors :

$$\hat{\eta}_n^{(k)} = \frac{1}{n} \sum_{i=1}^{n} Z_i \xrightarrow[n \to \infty]{\text{p.s.}} \mathbb{E}(Z_1) = \eta^{(k)}.$$

puisque $Z_1, Z_2, \ldots \overset{\text{iid}}{\sim} \mathrm{Ber}\left(\eta^{(k)}\right)$.

retour au sujet

❷ Consistance en moyenne quadratique

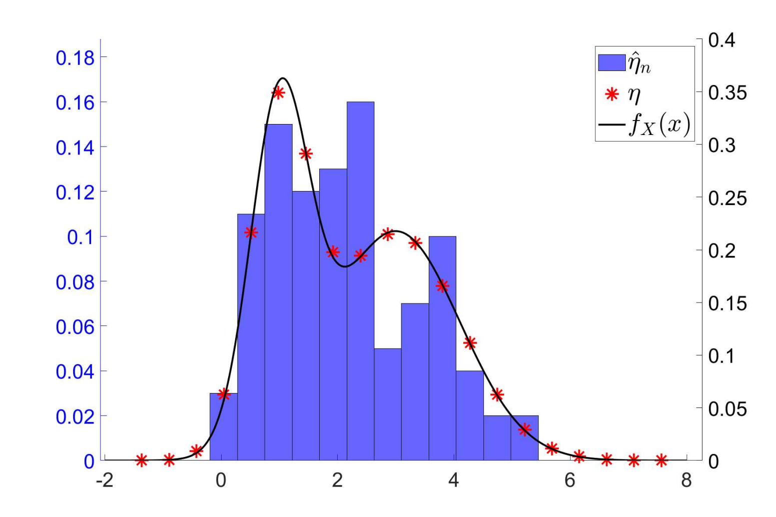
$$\mathbb{E}\left(\|\hat{\eta}_n - \eta\|^2\right) = \sum_{k=1}^{K} \mathbb{E}\left(\left(\hat{\eta}_n^{(k)} - \eta^{(k)}\right)^2\right)$$

À k fixé : $\hat{\eta}_n^{(k)} = \overline{Z}_n$ avec $Z_i \sim \mathrm{Ber}(\eta^{(k)})$ de variance finie.

Appliquant la loi des grands nombres dans L^2 :

$$\overline{Z}_n = \hat{\eta}_n^{(k)} \xrightarrow[n\to\infty]{L^2} \eta^{(k)}, \quad \text{i.e.} \quad \mathbb{E}\left(\left(\hat{\eta}_n^{(k)} - \eta^{(k)}\right)^2\right) \xrightarrow[n\to\infty]{} 0$$

$$\text{Ainsi } \mathbb{E}\left(\|\hat{\eta}_n - \eta\|^2\right) \xrightarrow[n\to\infty]{} 0 \text{ et } \quad \hat{\eta}_n \xrightarrow[n\to\infty]{L^2} \eta$$

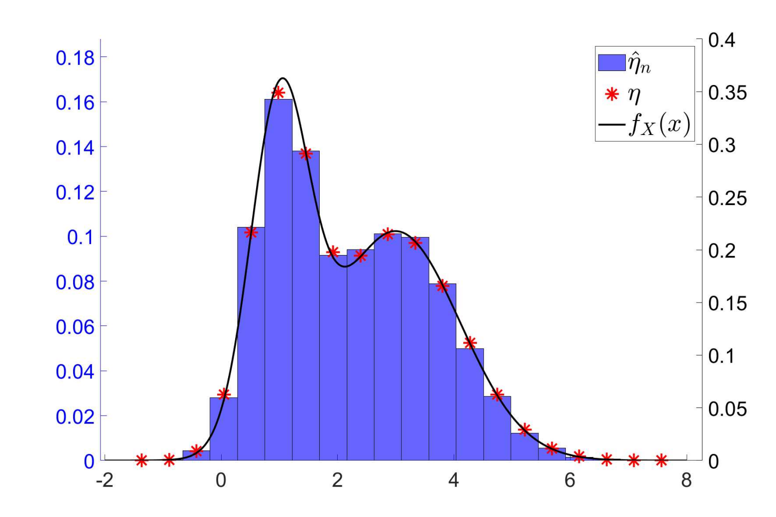


Une réalisation d'histogramme calculé pour $N = 100$ et $K = 20$.

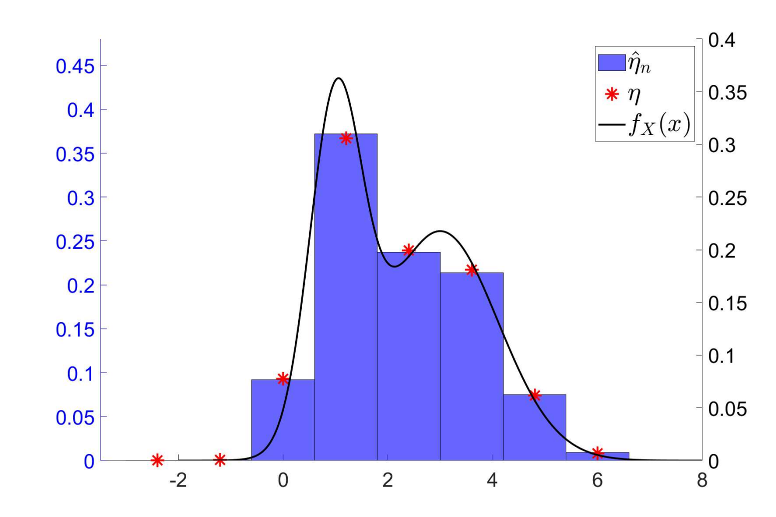
Remarque. La loi utilisée dans l'exemple admet une densité $f_X(x)$.

Corrigé de l'exercice 3

retour au sujet



Une réalisation d'histogramme calculé pour $N = 10000$ *et* $K = 20$.



Une réalisation d'histogramme calculé pour $N = 10000$ *et* $K = 8$.

Corrigé de l'exercice 4

retour au sujet

Considérons la décomposition biais-variance du risque quadratique :

$$\mathbb{E}_\theta\left((\hat{\eta} - g(\theta))^2\right) = \mathrm{var}_\theta(\hat{\eta}) + \mathrm{b}_\theta(\hat{\eta})^2.$$

Les deux termes de la somme sont positifs, donc

$$\mathbb{E}_\theta\left((\hat{\eta} - g(\theta))^2\right) \to 0 \quad \Leftrightarrow \quad \begin{cases} \mathrm{var}_\theta(\hat{\eta}) \to 0, \\ \mathrm{b}_\theta(\hat{\eta}) \to 0. \end{cases}$$

Ceci démontre l'équivalence annoncée. $\square$

Plan du cours

Plan du cours

Exemples hors-programme (statistique non-paramétrique)

Exemple 3

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}$,
- $\theta = \mathrm{P}$, la loi inconnue,
- $\Theta = \{\text{lois de proba sur } (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$,
- $g(\theta) = F$: fonction de répartition des X_i.

Exemple 4

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}_\theta$,
- P_θ : loi de densité de probabilité $\theta(x)$
- $\Theta = \left\{\text{ddp sur } \mathbb{R}, \text{ de classe } \mathscr{C}^2, \text{ avec } \int \theta''(x)^2 \,\mathrm{d}x < +\infty\right\}$
- $g(\theta) = \theta$.

retour au slide 10

Démonstration de l'inégalité de Cramér-Rao

Remarque préliminaire : puisque $\hat{\eta}$ est un ESB régulier de $g(\theta)$, g est nécessairement dérivable.

Fixons $\theta \in \Theta$ et posons $c = \mathrm{cov}_\theta\left(S_\theta,\, \hat{\eta}\right) \in \mathbb{R}^p$. Alors, $\forall a \in \mathbb{R}^p$,

$$\mathrm{var}_\theta\left(\hat{\eta} - a^\top S_\theta\right) = \mathrm{var}_\theta\left(\hat{\eta}\right) - 2a^\top c + a^\top \mathrm{var}_\theta\left(S_\theta\right) a \;\geq\; 0.$$

En particulier, pour $a = \mathrm{var}_\theta\left(S_\theta\right)^{-1} c \in \mathbb{R}^p$, on obtient :

$$\mathrm{var}_\theta\left(\hat{\eta}\right) - c^\top \mathrm{var}_\theta\left(S_\theta\right)^{-1} c \geq 0.$$

Pour conclure, puisque S_θ est centré et $\hat{\eta}$ un ESB régulier,

$$\begin{aligned} c = \mathbb{E}_\theta\left(\hat{\eta} S_\theta\right) &= \int_{\mathcal{S}} \hat{\eta}(\underline{x}) \cdot \frac{1}{f_\theta(\underline{x})} \nabla_\theta f_\theta(\underline{x}) \cdot f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) \\ &= \int_{\mathcal{S}} \hat{\eta}(\underline{x})\, \nabla_\theta f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x}) = \nabla_\theta \mathbb{E}_\theta\left(\hat{\eta}\right) = \nabla g(\theta). \qquad \square \end{aligned}$$

retour au slide 30

Rappel de proba : modes de convergence

- convergence presque sûre :

$$T_n \xrightarrow{\text{ps}} T \qquad \text{si} \qquad \mathbb{P}\left(T_n \to T\right) = 1$$

- convergence dans L^2 (en moyenne quadratique) :

$$T_n \xrightarrow{L^2} T \qquad \text{si} \qquad \mathbb{E}\left(\|T_n - T\|^2\right) \to 0$$

$$\text{ssi} \qquad \forall j \leq p, \quad T_n^{(j)} \xrightarrow{L^2} T^{(j)}$$

- convergence en probabilité :

$$T_n \xrightarrow{\mathrm{P}} T \qquad \text{si} \qquad \forall \varepsilon > 0, \quad \mathbb{P}\left(\|T_n - T\| \geq \varepsilon\right) \to 0$$

- convergence en loi :

$$T_n \xrightarrow{\text{loi}} T \qquad \text{si} \qquad \forall \varphi, \quad \mathbb{E}\left(\varphi(T_n)\right) \to \mathbb{E}\left(\varphi(T)\right),$$

avec $\varphi : \mathbb{R}^p \to \mathbb{R}$ continue et bornée.

retour au slide 36

Consistance de l'EMV

L'EMV minimise le critère

$$\gamma_n(\theta) = -\frac{1}{n} \ln f_\theta(\underline{X}) = -\frac{1}{n} \sum_{k=1}^{n} \ln f_\theta(X_i).$$

Fixons $\theta_\star \in \Theta$ et supposons $X_i \overset{\text{iid}}{\sim} f_{\theta_\star}$. Alors, $\forall \theta \in \Theta$,

$$\gamma_n(\theta) - \gamma_n(\theta_\star) = \frac{1}{n} \sum_{k=1}^{n} \ln \frac{f_{\theta_\star}(X_i)}{f_\theta(X_i)} \xrightarrow[n\to+\infty]{\text{ps}} \int_{\mathcal{S}_1} \ln \frac{f_{\theta_\star}(x)}{f_\theta(x)} \, f_{\theta_\star}(x)\, \nu_1(\mathrm{d}x).$$

(en supposant que $Z_i = \frac{f_{\theta_\star}(X_i)}{f_\theta(X_i)}$ admet un moment d'ordre 1).

Définition / propriété : divergence de Kullback-Leibler

$$D_{\mathrm{KL}}\left(f_{\theta_\star} || f_\theta\right) = \int_{\mathcal{S}_1} \ln \frac{f_{\theta_\star}(x)}{f_\theta(x)} \, f_{\theta_\star}(x)\, \nu_1(\mathrm{d}x) \geq 0$$

Consistance de l'EMV (suite)

Posons $\Delta_n(\theta_\star, \theta) = \frac{1}{n}\sum_{k=1}^{n} \ln \frac{f_{\theta_\star}(X_i)}{f_\theta(X_i)}$ et $\Delta(\theta_\star, \theta) = D_{\mathrm{KL}}\left(f_{\theta_\star} || f_\theta\right)$.

On a $\Delta_n(\theta_\star, \theta) \xrightarrow[n\to+\infty]{\mathbb{P}_{\theta_\star}-\mathrm{ps}} \Delta(\theta_\star, \theta)$ pour tout θ, et $\Delta(\theta_\star, \theta_\star) = 0$.

Théorème : consistance de l'EMV

Supposons que, pour tout $\theta_\star \in \Theta$,

1. $\sup_{\theta\in\Theta} |\Delta_n(\theta_\star, \theta) - \Delta(\theta_\star, \theta)| \xrightarrow[n\to+\infty]{\mathbb{P}_{\theta_\star}} 0$
2. et, pour tout $\epsilon > 0$,

$$\inf_{\theta\in\Theta,\, \|\theta-\theta_\star\|\geq\epsilon} \Delta(\theta_\star, \theta) > 0.$$

Alors l'EMV est consistant.

retour au slide 41