Statistique Mathématique

Salim Lardjane

Université Bretagne Sud

Partie IV
Estimateurs Uniformément Sans Biais de Variance Minimale

Statistiques exhaustives

Soit $X_1, \ldots, X_n$ un échantillon i.i.d associée à une v.a. X de loi $\mathcal{B}(\theta)$, c'est-à-dire que :

$$\mathbb{P}(X_i = x) = \theta^x (1-\theta)^{1-x} \quad (x = 0, 1).$$

Soit Y une *statistique*, c'est-à-dire une variable aléatoire fonction des observations $X_1, \ldots, X_n$.

Plus précisément, soit :

$$Y = X_1 + \cdots + X_n.$$

Y est le nombre de succès observés sur n essais de Bernoulli, la probabilité de succès d'un essai étant θ.

Statistiques exhaustives

Théorème. La loi conditionnelle de $X_1, \ldots, X_n$ sachant Y ne dépend pas de θ.

On dit que la statistique Y est *exhaustive* pour θ.

Preuve. Par définition, nous avons :

$$\begin{aligned} &\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n | Y = y) \\ &= \frac{\mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n, Y = y)}{\mathbb{P}_\theta(Y = y)}. \end{aligned}$$

Or la probabilité au numérateur vaut 0 à moins que l'on ait $y = x_1 + \cdots + x_n$, auquel cas :

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n | Y = y) = \frac{\theta^y (1-\theta)^{n-y}}{C_n^y \theta^y (1-\theta)^{n-y}} = \frac{1}{C_n^y}.$$

Statistiques exhaustives

Par exemple, si l'on sait qu'on obtient 3 succès en 5 lancers, la probabilité que le résultat soit $SESSE$ est $1/C_5^3 = 1/10$, où on note S pour succès et E pour échec.

Statistiques exhaustives

Une conséquence du résultat précédent est que *pour prendre une décision statistique*, on peut *se baser uniquement sur* $X_1 + \cdots + X_n$ et ignorer le détail des observations.

En effet, supposons que le statisticien A observe $X_1, \ldots, X_n$ et qu'il prenne une décision sur cette base, par exemple qu'il estime θ par $T(X_1, \ldots, X_n)$.

Supposons également que la statisticienne B observe uniquement $X_1 + \cdots + X_n$ et tire $X'_1, \cdots, X'_n$ selon la loi conditionnelle de $X_1, \ldots, X_n$ sachant Y, c'est-à-dire que :

$$\mathbb{P}(X'_1 = x_1, \ldots, X'_n = x_n | Y = y) = \frac{1}{C_n^y}.$$

Ce tirage est possible car la loi conditionnelle *ne dépend pas du paramètre inconnu* θ.

Nous avons alors le résultat suivant.

Statistiques exhaustives

Théorème. $(X'_1, \ldots, X'_n)$ *et* $(X_1, \ldots, X_n)$ *sont de même loi*.

Preuve. Etant données les valeurs $x_1, \ldots, x_n$, soit $y = x_1 + \cdots + x_n$. Si

$$X'_1 = x_1, \ldots, X'_n = x_n$$

alors $Y = X_1 + \cdots + X_n = y$ et le tirage effectué par B produit $X'_1 = x_1, \ldots, X'_n = x_n$ avec la probabilité :

$$\begin{aligned}
\mathbb{P}_\theta(X'_1 = x_1, \ldots, X'_n = x_n) &= \mathbb{P}_\theta(Y = y) \cdot \\
&\quad \cdot \mathbb{P}_\theta(X'_1 = x_1, \ldots, X'_n = x_n | Y = y) \\
&= C_n^y \theta^y (1-\theta)^{n-y} \frac{1}{C_n^y} \\
&= \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n).
\end{aligned}$$

Par conséquent, les estimateurs $T(X_1, \ldots, X_n)$ et $T(X'_1, \ldots, X'_n)$ *sont de même loi*. On peut donc estimer indifférement θ à l'aide de l'un ou l'autre, *avec les mêmes propriétés*.

Théorème de Factorisation

Soit X une variable ou un vecteur aléatoire de densité $f_\theta(x)$ et soit $Y = u(X)$ une statistique.

Le plus souvent $X = (X_1, \ldots, X_n)$, un échantillon i.i.d. d'observations, et Y est un *estimateur* de θ.

Théorème (de factorisation). $Y = u(X)$ *est une statistique exhaustive pour* θ *si et seulement si la densité de* X *peut être factorisée sous la forme :*

$$f_\theta(x) = g(\theta, u(x))h(x).$$

Dans le cas d'un échantillon de Bernoulli, vu précédemment, on a :

$$f_\theta(x_1, \ldots, x_n) = \theta^y(1-\theta)^{n-y}$$

$$u(x) = y = \sum_{i=1}^{n} x_i$$

$$h(x) \equiv 1.$$

Théorème de Factorisation

Preuve du théorème. On donne la preuve dans le cas discret uniquement.

Si $Y = u(X)$ est une statistique exhaustive pour θ, alors :

$$\begin{aligned}\mathbb{P}_\theta(X = x) &= \mathbb{P}_\theta(X = x, Y = u(x)) \\ &= \mathbb{P}_\theta(Y = u(x))\mathbb{P}(X = x|Y = u(x)) \\ &= g(\theta, u(x))h(x).\end{aligned}$$

Réciproquement, si $f_\theta(x) = g(\theta, u(x))h(x)$, alors :

$$\mathbb{P}_\theta(X = x|Y = y) = \frac{\mathbb{P}_\theta(X = x, Y = y)}{\mathbb{P}_\theta(Y = y)}$$

et le numérateur vaut 0 sauf si $y = u(x)$, auquel cas l'expression précédente devient :

$$\frac{\mathbb{P}_\theta(X = x)}{\mathbb{P}_\theta(Y = y)} = \frac{g(\theta, u(x))h(x)}{\sum_{z:u(z)=y} g(\theta, u(z))h(z)}.$$

Théorème de Factorisation

Les termes en g au numérateur et au dénominateur sont égaux à $g(\theta, y)$ et peuvent donc être éliminés.

Ainsi, on obtient :

$$\mathbb{P}_\theta(X = x | Y = y) = \frac{h(x)}{\sum_{z:u(z)=y} h(z)},$$

expression qui ne dépend pas de θ, d'où l'exhaustivité de Y pour θ.

Exemple

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de loi $\mathcal{N}(\mu, \sigma^2)$.

On a donc :

$$
\begin{aligned}
f_\theta(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]
\end{aligned}
$$

Posons $\theta = (\mu, \sigma^2)$ et soient

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Exemple

Alors, on peut écrire :

$$x_i - \bar{x} = x_i - \mu - (\bar{x} - \mu)$$

et

$$\begin{aligned} s^2 &= \frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu)\right. \\ &\left. + n(\bar{x}-\mu)^2\right]. \end{aligned}$$

Or $\sum_{i=1}^{n}(x_i - \mu) = n(\bar{x} - \mu)$, d'où

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 - (\bar{x}-\mu)^2$$

et la densité conjointe s'écrit :

$$f_\theta(x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2} e^{-ns^2/2\sigma^2} e^{-n(\bar{x}-\mu)^2/2\sigma^2}$$

Exemple

Si μ et σ^2 sont toutes deux inconnues, alors $\theta = (\mu, \sigma^2)$ et $(\bar{X}, S^2)$ est exhaustive pour θ, avec :

$$h(x) \equiv 1.$$

Si σ^2 est connue et μ inconnue, $\theta = \mu$ et $\bar{X}$ est exhaustive pour θ. C'est une conséquence du théorème de factorisation avec

$$h(x) = (2\pi\sigma^2)^{-n/2} e^{-ns^2/2\sigma^2}.$$

Si μ est connue et σ^2 est inconnue, alors $\theta = \sigma^2$ et $\sum_{i=1}^{n}(X_i - \mu)^2$ est exhaustive pour θ. C'est une conséquence du théorème de factorisation avec

$$h(x) \equiv 1$$

dans l'expression initiale de la densité conjointe.

Théorème de Rao-Blackwell

Afin de mieux comprendre le cheminement conduisant au théorème de Rao-Blackwell, commençons par considérer une expérience en deux étapes :

> Etape 1. On observe une v.a.r. X de densité $\frac{1}{2}x^2e^{-x}$ $(x > 0)$
>
> Etape 2. Si $X = x$ on tire Y selon la loi uniforme sur $]0, x[$.

On souhaite déterminer $\mathbb{E}(Y)$.

Théorème de Rao-Blackwell

Méthode 1 (densité conjointe). Nous avons :

$$\begin{aligned} f(x,y) &= f_X(x)f_Y(y|x) \\ &= \frac{1}{2}x^2e^{-x}\cdot\frac{1}{x} \\ &= \frac{1}{2}x\,e^{-x} \quad (0<y<x) \end{aligned}$$

En toute généralité, on a :

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy.$$

Dans notre cas, on a $g(x,y)=y$, d'où :

$$\begin{aligned} \mathbb{E}(Y) &= \int_0^{\infty}\left(\int_0^x \frac{y}{2}dy\right)x\,e^{-x}dx \\ &= \int_0^{\infty}\frac{x^3}{4}e^{-x}dx = \frac{\Gamma(4)}{4} \\ &= \frac{3!}{4} = \frac{3}{2} \end{aligned}$$

Théorème de Rao-Blackwell

Méthode 2. Cette méthode fonctionne bien lorsque l'espérance conditionnelle est facile à calculer. Dans notre cas :

$$\mathbb{E}(Y|X=x) = \frac{x}{2}$$

puisque $Y|X=x \sim \mathcal{U}_{]0,x[}$.

Par conséquent :

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}[\mathbb{E}(Y|X)] \quad \text{(Lemme 1 ci-après)} \\
&= \mathbb{E}(X/2) \quad \text{(Remarque ci-après)} \\
&= \int_0^\infty \frac{x}{2} \cdot \frac{1}{2}\, x^2\, e^{-x}\, dx \\
&= \frac{3}{2}.
\end{aligned}$$

Théorème de Rao-Blackwell

Remarque. Si on obtient, par exemple, que $\mathbb{E}[Y|X=x] = x^2 + 3x + 4$, alors on peut écrire

$$\mathbb{E}[Y|X] = X^2 + 3X + 4.$$

Ainsi, $\mathbb{E}(Y|X)$ est une fonction $g(X)$ de la v.a. X.

Lorsque $X = x$, on note $g(x) = \mathbb{E}[Y|X = x]$.

Théorème de Rao-Blackwell

Lemme 1. *Soient X_1, X_2 deux variables aléatoires réelles, alors*

$$\mathbb{E}[\mathbb{E}(X_2|X_1)] = \mathbb{E}(X_2).$$

Preuve. Soit $g(X_1) = \mathbb{E}(X_2|X_1)$. Alors, on a :

$$\begin{aligned}
\mathbb{E}[g(X_1)] &= \int_{-\infty}^{+\infty} g(x_1) f_1(x_1) dx_1 \\
&= \int_{-\infty}^{+\infty} \mathbb{E}[X_2|X_1 = x_1] f_1(x_1) dx_1 \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_2 f_{2|1}(x_2|x_1) f_1(x_1) dx_1 dx_2 \\
&= \int_{-\infty}^{+\infty} x_2 \left(\int_{-\infty}^{+\infty} f(x_1, x_2) dx_1 \right) dx_2 \\
&= \int_{-\infty}^{+\infty} x_2 f_2(x_2) dx_2 \\
&= \mathbb{E}(X_2).
\end{aligned}$$

.

Théorème de Rao-Blackwell

Lemme 2. *Soient X_1, X_2 deux variables aléatoires réelles et supposons $\mu_1 = \mathbb{E}(X_1)$ et $\mu_2 = \mathbb{E}(X_2)$ bien définies, alors :*

$$\mathbb{E}\left[(X_2 - \mathbb{E}(X_2|X_1))(\mathbb{E}[X_2|X_1] - \mu_2)\right] = 0.$$

Preuve. L'espérance s'écrit :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x_2 - \mathbb{E}(X_2|X_1 = x_1)] \cdot [\mathbb{E}(X_2|X_1 = x_1) - \mu_2] \cdot f_1(x_1) f_{2|1}(x_2|x_1) dx_1 dx_2$$

$$= \int_{-\infty}^{+\infty} f_1(x_1)[\mathbb{E}(X_2|X_1 = x_1) - \mu_2] \cdot \cdot \left(\int_{-\infty}^{+\infty}[x_2 - \mathbb{E}(X_2|X_1 = x_1)] f_{2|1}(x_2|x_1) dx_2\right) dx_1$$

Or l'intégrale entre parenthèse vaut :

$$\mathbb{E}(X_2|X_1 = x_1) - \mathbb{E}(X_2|X_1 = x_1) = 0,$$

d'où le résultat.

Théorème de Rao-Blackwell

Lemme 3. *Soient* X_1, X_2 *deux variables aléatoires réelles, de variances bien définies, alors :*

$$V(X_2) \geq V(\mathbb{E}[X_2|X_1]).$$

Preuve. Nous pouvons écrire :

$$\begin{aligned}
V(X_2) &= \mathbb{E}[(X_2 - \mu_2)^2] \\
&= \mathbb{E}([(X_2 - \mathbb{E}(X_2|X_1)) \\
&\quad +(\mathbb{E}(X_2|X_1) - \mu_2)]^2) \\
&= \mathbb{E}[(X_2 - \mathbb{E}(X_2|X_1))^2] \\
&\quad +\mathbb{E}[(\mathbb{E}(X_2|X_1) - \mu_2)^2] \quad \text{(Lemme 2)} \\
&\geq \mathbb{E}[(\mathbb{E}(X_2|X_1) - \mu_2)^2]
\end{aligned}$$

car les deux termes sont positifs.

Mais en vertu du Lemme 1, on a

$$\mathbb{E}[\mathbb{E}(X_2|X_1)] = \mu_2$$

Ainsi,

$$\mathbb{E}[(\mathbb{E}(X_2|X_1) - \mu_2)^2] = V(\mathbb{E}[X_2|X_1])$$

d'où le résultat.

Théorème de Rao-Blackwell

Lemme 4. *Soient* X_1, X_2 *deux variables aléatoires réelles, de variances bien définies, alors* $V(X_2) = V(\mathbb{E}[X_2|X_1])$ *si et seulement si* X_2 *est une fonction de* X_1.

Preuve. La preuve du Lemme 3 montre qu'il y a égalité si et seulement si :

$$\mathbb{E}[(X_2 - \mathbb{E}(X_2|X_1))^2] = 0,$$

c'est-à-dire si et seulement si $X_2 = \mathbb{E}(X_2|X_1)$ (voir Remarque à suivre). Ceci implique que X_2 est une fonction de X_1.

Réciproquement, si $X_2 = h(X_1)$, alors $\mathbb{E}(X_2|X_1) = h(X_1) = X_2$, d'où l'égalité des variances.

Théorème de Rao-Blackwell

Remarque. On a utilisé le résultat suivant dans la preuve précédente : *Soit X une v.a.r. telle que $\mathbb{E}(X^2) = 0$. Alors, $X = 0$.*

En toute rigueur, l'égalité est *presque sure*, mais on ne détaillera pas cet aspect dans ce cours.

Théorème de Rao-Blackwell

Théorème de Rao-Blackwell. *Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de densité $f_\theta(x)$. Soit $Y_1 = u_1(X_1, \ldots, X_n)$ une statistique exhaustive pour θ et soit $Y_2 = u_2(X_1, \ldots, X_n)$ un estimateur sans biais de $r(\theta)$, où r est une fonction quelconque de θ. Alors,*

1. *$V(\mathbb{E}[Y_2|Y_1]) \leq V(Y_2)$, l'inégalité étant stricte sauf si Y_2 est une fonction de Y_1.*
2. *$\mathbb{E}[\mathbb{E}(Y_2|Y_1)] = r(\theta)$.*

Théorème de Rao-Blackwell

Ainsi, *lorsqu'on cherche un estimateur sans biais de variance minimale de* $r(\theta)$*, on peut se limiter aux seules fonctions de la statistique exhaustive* Y_1.

Preuve. La propriété 1. est une conséquence des Lemmes 3 et 4. La propriété 2. découle du Lemme 1.

Exhaustivité et EMV

Soit $Y_1 = u_1(X_1, \ldots, X_n)$ une statisique exhaustive pour θ.

Théorème. *Si l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ est unique, alors $\widehat{\theta}_n$ est fonction de Y_1.*

Preuve. La densité conjointe des X_i peut être factorisée sous la forme :

$$f_\theta(x_1, \ldots, x_n) = g(\theta, z)h(x_1, \ldots, x_n)$$

où $z = u_1(x_1, \ldots, x_n)$. Soit θ_0 la valeur possible de θ qui maximise $g(\theta, z)$. Etant donné z, on peut déterminer θ_0 en examinant tous les $g(\theta, z)$; ainsi, θ_0 est fonction de $u_1(x_1, \ldots, x_n) = y_1$. Mais θ_0 maximise également $f_\theta(x_1, \ldots, x_n)$, d'où $\widehat{\theta} = \theta_0$ par unicité, où $\widehat{\theta}$ désigne *l'estimation* par maximum de vraisemblance de θ. On en déduit que l'EMV $\widehat{\theta}_n$ est fonction de Y_1.

Estimateurs USBVM

Les résultats précédents nous permettront de trouver des estimateurs USBVM (uniformément sans biais de variance minimale) de θ, ainsi que nous le verrons dans la suite.

Estimateurs USBVM

Soit Y une statistique exhaustive pour θ.

On dit que Y est *complète* s'il n'existe pas d'estimateur non trivial de 0 basé sur Y, c'est-à-dire si :

$$\mathbb{E}_\theta[g(Y)] = 0 \; \forall\theta \implies \mathbb{P}_\theta[g(Y) = 0] = 1 \; \forall\theta.$$

Estimateurs USBVM

Si Y est une statistique exhaustive complète pour θ et si on dispose de deux estimateurs sans biais de $r(\theta)$ basés sur Y, $\varphi(Y)$ et $\psi(Y)$, alors comme :

$$\mathbb{E}_\theta[\varphi(Y) - \psi(Y)] = 0 \ \forall\theta$$

on a nécessairement que

$$\mathbb{P}_\theta[\varphi(Y) = \psi(Y)] = 1 \ \forall\theta.$$

Par conséquent, si on trouve un estimateur sans biais de $r(\theta)$ basé sur Y, alors on les a essentiellement tous trouvés.

Estimateurs USBVM

Théorème de Lehmann-Scheffé. *Soit $Y_1 = u_1(X_1, \ldots, X_n)$ une statistique exhaustive complète pour θ. Si $\varphi(Y_1)$ est un estimateur sans biais de $r(\theta)$ basé sur Y_1, alors parmi tous les estimateurs sans biais de $r(\theta)$, $\varphi(Y_1)$ est de variance minimale, c'est-à-dire que pour tout estimateur sans biais Y_2 de $r(\theta)$:*

$$V_\theta[\phi(Y_1)] \leq V_\theta(Y_2) \quad \forall\theta.$$

On dit que $\varphi(Y_1)$ est un *estimateur uniformément sans biais de variance minimale* (ang. Uniformly Minimum Variance Unbiased Estimator) ou estimateur USBVM (ang. UMVUE) de $r(\theta)$.

Le terme "uniformément" est utilisé car le résultat est vrai pour *toutes les valeur possibles de θ*.

Estimateurs USBVM

Preuve. En vertu du théorème de Rao-Blackwell, si Y_2 est un estimateur sans biais quelconque de $r(\theta)$, alors $\mathbb{E}(Y_2|Y_1)$ est un estimateur sans biais de $r(\theta)$ avec

$$V[\mathbb{E}(Y_2|Y_1)] \leq V(Y_2).$$

Mais $\mathbb{E}(Y_2|Y_1)$ est fonction de Y_1, donc il coïncide essentiellement avec $\phi(Y_1)$ en vertu de la complétude de Y_1 et a donc même variance.

Ainsi, indépendamment de la valeur de θ, on a :

$$V_\theta[\phi(Y_1)] \leq V_\theta(Y_2).$$

Estimateurs USBVM

Il existe de nombreux modèles pour lesquels on peut trouver facilement des statistiques exhaustives complètes.

C'est le cas notamment de *la famille exponentielle*.

Celle-ci regroupe les densités de la forme :

$$f_\theta(x) = a(\theta)b(x)\,\text{exp}\left[\sum_{j=1}^{m} p_j(\theta)K_j(x)\right]$$

où $a(\theta) > 0$, $b(x) > 0$, $\alpha < x < \beta$, $\theta = (\theta_1, \ldots, \theta_k)$ avec $\gamma_j < \theta_j < \delta_j$, $1 \leq j \leq k$ ($\alpha, \beta, \gamma_j, \delta_j$ sont des constantes).

Estimateurs USBVM

Des *conditions de régularité* doivent en toute rigueur être imposées, mais elles seront toujours vérifiées dans les exemples que nous aurons à considérer, donc nous les omettons.

Dans tous les exemples considérés, on aura $k = m$. *C'est requis pour avoir la complétude*.

Estimateurs USBVM

Exemples. Les modèles suivant appartiennent à la famille exponentielle (le démontrer à titre d'exercice) :

(a) $\mathcal{B}(n, p)$, $\theta = p$

(b) $\mathcal{P}(\theta)$

(c) $\mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$

(d) $\gamma(\alpha, \beta)$, $\theta = (\alpha, \beta)$

(e) $\beta(a, b)$, $\theta = (a, b)$

Estimateurs USBVM

Nous allons à présent obtenir une statistique exhaustive complète pour les membres de la famille exponentielle.

Estimateurs USBVM

La densité des membres de la famille exponentielle s'écrit :

$$f_\theta(x) = a(\theta)b(x)\,\mathsf{exp}\left[\sum_{j=1}^{m} p_j(\theta)K_j(x)\right].$$

La densité conjointe de n observations i.i.d. est donc donnée par :

$$\begin{aligned} f_\theta(x_1,\ldots,x_n) &= a(\theta)^n \prod_{i=1}^{n} b(x_i)\,\mathsf{exp}\left[\sum_{j=1}^{m} p_j(\theta)K_j(x_1)\right] \\ &\qquad \cdots \mathsf{exp}\left[\sum_{j=1}^{m} p_j(\theta)K_j(x_n)\right]. \end{aligned}$$

Estimateurs USBVM

Mais, en vertu des propriétés de l'exponentielle, on a :

$$\exp\left[\sum_{j=1}^{m} p_j(\theta)K_j(x_1)\right]\cdots\exp\left[\sum_{j=1}^{m} p_j(\theta)K_j(x_n)\right]$$
$$= \exp\left[\sum_{j=1}^{m} p_j(\theta)\left(K_j(x_1)+\cdots+K_j(x_n)\right)\right]$$

Donc, en vertu du théorème de factorisation, avec $h(x) = \prod_{i=1}^{n} b(x_i)$, la statistique :

$$\left(\sum_{i=1}^{n} K_1(X_i), \ldots, \sum_{i=1}^{n} K_m(X_i)\right)$$

est exhaustive pour θ.

On admettra que cette statistique est également complète.

Estimateurs USBVM

Exemple. Soit $X_1, \ldots, X_n$ un échantillon associé à une v.a.r. X de loi $\mathcal{N}(\mu, \sigma^2)$, où σ^2 est connu.

Nous pouvons écrire :

$$
\begin{aligned}
f_\mu(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-\mu^2}{2\sigma^2}\right] \exp\left[-\frac{x^2}{2\sigma^2}\right] \exp\left[\frac{\mu x}{\sigma^2}\right] \\
&= a(\mu) b(x) \exp[p(\mu) K(x)]
\end{aligned}
$$

avec $p(x) = \mu/\sigma^2$ et $K(x) = x$. Ainsi, notre modèle appartient à la famille exponentielle et, par conséquent,

$$\sum_{i=1}^{n} X_i$$

ou, de façon équivalente, $\bar{X}$, est exhaustive et complète pour μ .

Estimateurs USBVM

Comme $\mathbb{E}(\bar{X}) = \mu$, il en découle que $\bar{X}$ est un estimateur $USBVM$ de μ, c'est-à-dire qu'il est uniformément sans biais de variance minimale.

Essayons à présent de trouver un estimateur sans biais de variance uniformément minimale de μ^2.

L'estimateur naturel $(\bar{X})^2$ n'est pas satisfaisant puisqu'on a :

$$V(\bar{X}) = \frac{\sigma^2}{n} = \mathbb{E}[(\bar{X})^2] - \mu^2$$

ce qui implique que :

$$\mathbb{E}[(\bar{X})^2] = \mu^2 + \frac{\sigma^2}{n},$$

c'est donc un estimateur biaisé de μ^2.

Estimateurs USBVM

Par contre,

$$(\bar{X})^2 - \frac{\sigma^2}{n}$$

est un estimateur sans biais de μ^2 d'après ce qui précède et, comme il est fonction de la statistique exhaustive complète $\bar{X}$, c'est un estimateur UMVUE de μ^2.

Estimateurs USBVM

Remarque. *Se limiter à considérer des estimateurs sans biais n'est pas toujours une bonne idée*. Pour l'illustrer, considérons un exemple : soit $X \sim \mathcal{P}(\theta)$ et soit X_1 un échantillon de taille 1 associé à X.

En mettant la densité de X sous forme exponentielle, on constate que X_1 est une statistique exhaustive complète pour θ.

Or :

$$\begin{aligned}
\mathbb{E}[(-1)^{X}_{1}] &= \sum_{k=0}^{\infty} (-1)^k \frac{e^{-\theta}\theta^k}{k!} \\
&= e^{-\theta} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \\
&= e^{-2\theta}.
\end{aligned}$$

Estimateurs USBVM

Ainsi, $(-1)^{X_1}$ est un estimateur uniformément sans biais de variance minimale de $e^{-2\theta}$.

Cependant $Y \equiv 1$ est clairement un *meilleur* estimateur, puisque 1 est plus proche de $e^{-2\theta}$ que -1.

En fait, estimer une quantité positive à l'aide d'une statistique qui peut prendre des valeurs négatives n'est pas raisonnable.

Notons également que l'EMV de θ n'est autre que X_1 (le démontrer à titre d'exercice).

Par conséquent, l'EMV de $e^{-2\theta}$ est e^{-2X_1}, qui semble plus raisonnable que Y.

Inégalité de Cramer-Rao

Soit donnée une densité $f_\theta(x)$, $x \in \mathbb{R}$, $a < \theta < b$.

On détermine les estimations par maximum de vraisemblance de θ en calculant :

$$\frac{\partial}{\partial\theta} \log f_\theta(x).$$

Si on remplace x par X dans cette expression, on obtient une *variable aléatoire* :

$$\frac{\partial}{\partial\theta} \log f_\theta(X).$$

Afin d'étudier celle-ci, commençons par un exemple.

Inégalité de Cramer-Rao

Supposons que X puisse prendre les valeurs x_1, x_2, x_3, x_4 avec la fonction de masse (densité) :

$p(x_1) = 0.5,\ p(x_2) = p(x_3) = 0.2,\ p(x_4) = 0.1.$

Alors, $p(X)$ est une variable aléatoire de loi :

$$\begin{aligned}
\mathbb{P}(p(X) = 0.5) &= 0.5 \\
\mathbb{P}(p(X) = 0.2) &= 0.4 \\
\mathbb{P}(p(X) = 0.1) &= 0.1.
\end{aligned}$$

Le cas continu est, en première approche plus simple : si X est de densité f et $X = x$, alors $f(X) = f(x)$. Mais quelle est la densité de $f(X)$? Nous n'aurons pas besoin de celle-ci mais elle sera déterminée en TD.

Inégalité de Cramer-Rao

Les deux lemmes suivants seront utilisés pour démontrer *l'inégalité de Cramer-Rao*.

Celle-ci peut être utilisée pour obtenir des estimateurs uniformément sans biais de varrance minimale (USBVM).

Dans la suite, on supposera systématiquement qu'on peut dériver sous le signe intégrale.

Inégalité de Cramer-Rao

Lemme 1. *Avec les notations précédentes, on a :*

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] = 0.$$

Preuve. L'espérance s'écrit :

$$\int_{-\infty}^{+\infty}\left[\frac{\partial}{\partial\theta}\log f_\theta(x)\right] f_\theta(x)dx = \int_{-\infty}^{+\infty}\frac{1}{f_\theta(x)}\frac{\partial f_\theta(x)}{\partial\theta}.f_\theta(x)dx$$

Elle est donc égale à :

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} f_\theta(x)dx = \frac{\partial}{\partial\theta}(1) = 0.$$

Inégalité de Cramer-Rao

Lemme 2. *Soit* $Y = g(X)$ *et supposons que* $\mathbb{E}_\theta(Y) = k(\theta)$*. Si on note* $k'(\theta) = dk(\theta)/d\theta$*, alors :*

$$k'(\theta) = \mathbb{E}_\theta \left[Y \frac{\partial}{\partial\theta} \log f_\theta(X) \right].$$

Preuve. On peut écrire :

$$\begin{aligned}
k'(\theta) &= \frac{\partial}{\partial\theta}\mathbb{E}_\theta[g(X)] \\
&= \frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} g(x) f_\theta(x) dx \\
&= \int_{-\infty}^{+\infty} g(x) \frac{\partial f_\theta(x)}{\partial\theta} dx \\
&= \int_{-\infty}^{+\infty} g(x) \frac{\partial f_\theta(x)}{\partial\theta} \frac{1}{f_\theta(x)} f_\theta(x) dx \\
&=. \int_{-\infty}^{+\infty} g(x) \left[\frac{\partial}{\partial\theta} \log f_\theta(x) \right] f_\theta(x) dx \\
&= \mathbb{E}_\theta[g(X) \frac{\partial}{\partial\theta} \log f_\theta(X)] \\
&= \mathbb{E}_\theta[Y \frac{\partial}{\partial\theta} \log f_\theta(X)].
\end{aligned}$$

Inégalité de Cramer-Rao

Théorème (Inégalité de Cramer-Rao). *Soit* $Y = g(X)$ *et supposons que* $\mathbb{E}_\theta(Y) = k(\theta)$*. Si on note* $k'(\theta) = dk(\theta)/d\theta$*, alors :*

$$V_\theta(Y) \geq \frac{[k'(\theta)]^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right]}.$$

Preuve. Soit U, V deux v.a.r. de variances bien définies. Alors, en vertu de l'inégalité de Cauchy-Schwarz, nous avons :

$$\begin{aligned}[\text{Cov}(U,W)]^2 &= (\mathbb{E}[(U-\mu_U)(W-\mu_W)])^2 \\ &\leq V(U)V(W).\end{aligned}$$

Par conséquent,

$$[\text{Cov}_\theta(Y, \frac{\partial}{\partial\theta}\log f_\theta(X))]^2 \leq V_\theta(Y)V_\theta\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)$$

Inégalité de Cramer-Rao

Mais, d'après le Lemme 1, on a :

$$\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right) = 0.$$

Par conséquent, l'inégalité précédente s'écrit :

$$\left(\mathbb{E}_\theta[Y\frac{\partial}{\partial\theta}\log f_\theta(X)]\right) \leq V_\theta(Y)\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2]$$

ce qui peut s'écrire, d'après le Lemme 2 :

$$[k'(\theta)]^2 \leq V_\theta(Y)\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2],$$

d'où le résultat.

Inégalité de Cramer-Rao pour un échantillon

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de densité commune $f_\theta(x)$. Notons :

$$X = (X_1, \ldots, X_n).$$

Alors, la densité conjointe de X est donnée par :

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

et on peut écrite, en vertu du Lemme 1 :

$$\begin{aligned}
\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2] &= V_\theta\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right) \\
&= V_\theta\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right) \\
&= n\,V_\theta\left(\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right) \\
&= n\,\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X_1))^2].
\end{aligned}$$

On en déduit le théorème suivant.

Information de Fisher

Théorème. *Soit* $X_1, \ldots, X_n$ *un échantillon i.i.d. de densité commune* $f_\theta(x)$ *et soit* $Y = g(X_1, \ldots, X_n)$ *un estimateur sans biais de* θ*. Alors :*

$$V_\theta(Y) \geq \frac{1}{n\,\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X_1))^2]},$$

ce qui peut s'écrire également :

$$V_\theta(Y) \geq \frac{1}{n\,I(\theta)}$$

où $I(\theta) = \mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X_1))^2]$ *est appelée l'information de Fisher.*

Preuve. Il suffit d'appliquer l'inégalité de Cramer-Rao, avec $k(\theta) = \theta$, donc $k'(\theta) = 1$.

Remarque 1. Il découle de l'inégalité précédente que si Y est un estimateur sans biais de θ pour lequel la borne de Cramer-Rao $1/nI(\theta)$ est *atteinte pour toute valeur possible de* θ, c'est-à-dire tel que :

$$V_\theta(Y) = \frac{1}{n\,I(\theta)} \quad \forall\theta$$

alors Y est nécessairement un estimateur uniformément sans biais de variance minimale de θ.

Information de Fisher

Remarque 2. Il est souvent plus simple de calculer l'information de Fisher de la façon suivante. On part de la conclusion du Lemme 1, qui s'écrit :

$$\int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial \theta} \log f_\theta(x) \right) f_\theta(x) dx = 0.$$

On différentie une fois pour obtenir :

$$\int_{-\infty}^{+\infty} \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} f_\theta(x) dx + \int_{-\infty}^{+\infty} \frac{\partial \log f_\theta(x)}{\partial \theta} \frac{\partial f_\theta(x)}{\partial \theta} dx = 0.$$

Ainsi :

$$\int_{-\infty}^{+\infty} \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} f_\theta(x) dx + \int_{-\infty}^{+\infty} \frac{\partial \log f_\theta(x)}{\partial \theta} \left[\frac{\partial f_\theta(x)}{\partial \theta} \frac{1}{f_\theta(x)} \right] f_\theta(x) dx = 0$$

Information de Fisher

Mais le terme entre crochet dans la deuxième intégrale n'est autre que : $\partial \log f_\theta(x)/\partial\theta$, d'où :

$$\int_{-\infty}^{+\infty} \frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} f_\theta(x)dx + \int_{-\infty}^{+\infty} \left(\frac{\partial \log f_\theta(x)}{\partial\theta}\right)^2 f_\theta(x)dx = 0$$

Par conséquent :

$$\mathbb{E}_\theta[(\frac{\partial}{\partial\theta} \log f_\theta(X_1))^2] = -\mathbb{E}_\theta \left[\frac{\partial^2 \log f_\theta(X_1)}{\partial\theta^2}\right]$$

Cette dernière expression est souvent plus facile à utiliser pour le calcul de $I(\theta)$.