Cours/TD/TP- Agrégation externe de mathématiques - Estimation - 2007/2008

Références :

Dacunha-Castelle D., Duflo M. Probabilités et statistiques. Tome 1 et 2 : problèmes à temps fixe/mobile. Masson, 1994.

Bickel-Docksum

Fourgeaud C., Fuchs A. Statistique. Dunod, 1967.

Saporta

Il est courant de vouloir estimer la valeur d'un paramètre θ figurant dans la loi de probabilité d'un phénomène aléatoire à l'aide d'une fonction T dépendant de tirages effectués sur le phénomène étudié. Les questions naturelles sont alors

– La statistique T est-elle non biaisée, *i.e.* $\mathbb{E}T = \theta$?

– L'estimateur T est-il efficace, c'est à dire de variance aussi petite possible ?

L'objectif de la première partie est de donner dans des cas simples des bornes sur la variance de l'estimation. Les autres parties seront plus illustratives dans des cas concrets.

1 Estimation et Borne de Cramer-Rao

1.1 Cas scalaire

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et X une variable aléatoire définie sur Ω à valeurs dans $\mathbb{R}$ de loi de densité p_θ où $\theta \in \Theta$ est un paramètre réel. On note E_θ le support de p_θ. On définit alors la quantité d'information de Fisher fournit par X sur θ notée $I_X(\theta)$:

$$I_X(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln p_\theta(X)\right)^2\right].$$

On a alors les propositions et théorème à démontrer dans le cadre préliminaire de ce cours/td.

Proposition 1.1 *Si E_θ ne dépend pas de θ, alors si p est suffisamment lisse on a*

$$I_X(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln p_\theta(X)\right] = V\left(\frac{\partial}{\partial\theta}\ln p_\theta(X)\right).$$

Proposition 1.2 (Propriété de tensorisation) *Soit $(X_1, \dots X_n)$ un n-échantillon i.i.d. de la variable X, si E_θ ne dépend pas de θ, alors*

$$I_{X_1,\dots X_n}(\theta) = nI_X(\theta).$$

Si on s'intéresse au problème d'estimer le paramètre θ à partir d'observations faites sur la variable X, le théorème suivant montre que l'information de Fisher permet de mesurer la qualité de tels estimateurs. Plus précisément, elle donne une borne inférieure pour la variance de tels estimateurs non biaisés.

Theorem 1.3 (Borne de Cramer-Rao) *Si E_θ ne dépend pas de θ, et que $T = \varphi(X_1, \dots X_n)$, alors on a sous des conditions génériques de régularité sur φ que*

$$V(T) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[T]\right)^2}{nI_X(\theta)}.$$

En particulier, si T est un estimateur sans biais de θ, alors on obtient $V(T) \geq \frac{1}{nI_X(\theta)}$.

Définition 1.4 (Estimateur efficace, asymptotiquement efficace) *T est un estimateur efficace si sa variance atteint la borne de Cramer-Rao. T est asymptotiquement efficace si sa variance est équivalente lorsque $n \to \infty$ à la borne de Cramer-Rao.*

Exercice 1.5 *Montrer que dans le cas d'un n-échantillon gaussien $\mathcal{N}(\mu, \sigma^2)$, l'estimateur de la moyenne empirique est un estimateur sans biais efficace de μ.*

1.2 Cas général

Bien entendu, lorsque le cadre n'est plus scalaire mais vectoriel, il faut donner un sens à la positivité des bornes obtenues. L'information de Fisher est tout d'abord définie comme une matrice :

$$I_X(\theta)_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta_i} \ln p_\theta(X)\right)\left(\frac{\partial}{\partial \theta_j} \ln p_\theta(X)\right)\right].$$

Et l'analogue du théorème précédent est

Theorem 1.6 *Soit $(X_1, \dots X_n)$ un n-échantillon de la variable X, soient $h_1, \dots h_s$ s fonctions réelles dépendant de la variable θ et $T_1, \dots T_s$ un estimateur sans biais de $h_1(\theta), \dots h_s(\theta)$. On note $V(T)$ la matrice de covariance du vecteur aléatoire $(T_1, \dots T_s)$ et $H'(\theta)$ la matrice des dérivées de la fonction $(h_1, \dots h_s)$:*

$$H'(\theta)_{i,j} = \frac{\partial}{\partial \theta_j} h_i(\theta).$$

Enfin, on suppose $I_X(\theta)$ inversible, alors sous les conditions d'existence et de régularités adéquates :

$$V(T) - \frac{1}{n} H'(\theta) I_X(\theta)^{-1} H'(\theta)^t \geq 0$$

Exercice 1.7 *Soit X une v.a. normale dans $\mathbb{R}^m$ de moyenne M et de covariance Γ. Montrer que la quantité d'information de Fisher de X sur M est*

$$I_X(M) = \Gamma^{-1}.$$

En déduire que l'estimateur de Monte-Carlo de la moyenne est efficace.

1.3 Maximum de vraisemblance

Définition 1.8 *X de densité $p_\theta(x)$ dépend de θ qu'on cherche à déterminer. La fonction de vraisemblance de la réalisation $X_1, \dots X_n$ est définie par*

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

On montre alors que l'estimateur du maximum de vraisemblance

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_\theta(X_1, \dots X_n)$$

vérifie le théorème suivant.

Theorem 1.9 *La suite $\hat{\theta}_n$ converge presque surement vers θ^*. La suite $\sqrt{n}(\hat{\theta}_n - \theta^*)$ converge en loi vers une loi normale centrée dont la matrice de covariance est l'inverse de la matrice d'information de Fisher :*

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \to_{\mathcal{L}} \mathcal{N}\left(0, I_X(\theta^*)^{-1}\right)$$

2.1 Statistiques exhaustives

On rappelle en vrac quelques définitions

Définition 2.1 *T est une statistique exhaustive si $L(x,\theta) = g(t,T)h(x)$ où L est la vraisemblance de l'échantillon $x = (x_1,\ldots,x_n)$. Une fois T connue, rien de plus n'apportera d'information sur θ.*

Le théorème suivant décrit précisément les lois possédant des statistiques exhaustives.

Theorem 2.2 *Si X est une v.a. tel que E_θ est indépendant de θ, une cns pour que X possède une statistique exhaustive est que sa densité soit de la forme*

$$f(x,\theta) = e^{a(x)\alpha(\theta)+b(x)+\beta(\theta)}$$

C'est à dire X appartient à la famille exponentielle.
Si en plus a bijective $\mathcal{C}^1$, alors une statistique exhaustive particulière est

$$T = \sum_{i=1}^{n} a(X_i).$$

Il y a de nombreux exemples : lois normales, exponentielles, Bernoulli, Poisson.

2.2 Estimation sans biais de variance minimale

Theorem 2.3 *S'il existe un estimateur sans biais de variance minimale, alors il est unique.*

Theorem 2.4 *Si T est un estimateur sans biais, et U une statistique exhaustive, alors $\mathbb{E}[T|U]$ est l'estimateur sans biais de variance minimale.*

2.3 Statistiques complètes

Définition 2.5 *On dit que U est une statistique complète si et seulement si*

$$\mathbb{E}h(U) = 0 \quad \forall\theta \qquad \Rightarrow \qquad h = 0$$

Theorem 2.6 (Lehmann-Scheffé) *Si T^* est un ESB de θ dépendant d'une statistique exhaustive complète U, alors T^* est l'ESBVM de θ. Des cas particuliers fréquents sont alors $T^* = \mathbb{E}[T|U]$.*

3 Maximum de vraisemblance et lois de Bernouilli

On lance une pièce mais des petites asymétries influencent la probabilité d'obtenir une face, qui n'est plus forcément 1/2.
On désigne $Y_1,\ldots Y_n$ les résultats d'essais indépendants et l'on note

$$P(Y_j = 1) = \theta \qquad P(Y_j = 0) = 1-\theta \qquad 0 \le \theta \le 1$$

1. Calculer la vraisemblance ainsi que la log-vraisemblance du modèle.
2. Calculer l'estimateur du maximum de vraisemblance.
3. Utiliser cet estimateur pour calculer dans ce cas $J(\hat{\theta}_n)$.
4. Programmer une fonction Matlab qui simule une suite de lancés de pièce biaisée de paramètre θ.
5. Tracer les vraisemblances obtenues en fonction de θ, pour différentes simulations de n pièces.
6. Illustrer le théorème de convergence de l'estimateur du maximum de vraisemblance.

On rappelle qu'une variable aléatoire X suit une loi de Pareto de paramètres α et θ si la densité s'écrit de la forme

$$f_{X,\alpha,\theta}(t) = kt^{-\alpha}\chi_{t\geq\theta}$$

4.1 Généralités

1. Montrer que la constante k de normalisation est donnée par

$$k = (\alpha - 1)\theta^{\alpha-1}$$

2. Montrer que si X suit une loi de Pareto (θ, α), alors

$$\mathbb{E}(X) = \frac{\alpha - 1}{\alpha - 2}\theta$$

et

$$V(X) = \frac{(\alpha - 1)\theta^2}{(\alpha - 3)(\alpha - 2)^2}$$

3. Montrer que la fonction de répartition $F_{X,\theta,\alpha}$ est donnée par

$$F_{X,\theta,\alpha}(t) = 1 - (\theta/t)^{\alpha-1}$$

4. Démontrer que si l'on effectue le changement de variable

$$U = (\alpha - 1)\ln\left(\frac{X}{\theta}\right)$$

alors U suit une loi exponentielle.

5. On note $Z = \min(X_1, \ldots, X_n)$ où $(X_1, \ldots, X_n)$ est un n-échantillon de loi $P(\theta, \alpha)$. Déterminer la loi de Z.
6. Écrire une fonction matlab pareto.m qui simule la loi de Pareto avec comme argument (θ, α). Écrire une fonction echantillon-pareto.m qui simule un n-échantillon de loi de Pareto. Tracer l'histogramme empirique obtenu.
7. Programmer la fonction min-pareto.m qui simule la variable Z en fonction de l'entier n taille de l'échantillon, de θ et α. Comparer l'histogramme empirique obtenu ici avec l'histogramme empirique construit à partir de p simulations de n-échantillons $(X_1, \ldots, X_n)$ de loi de Pareto θ, α.

4.2 Étude à θ connu

On suppose le paramètre θ connu, on cherche à estimer α.

1. Donner une statistique exhaustive pour α.
2. Estimer α par la méthode du maximum de vraisemblance. On notera $\hat{\alpha}$ l'EMV de α.
3. Écrire une fonction emv-alpha.m qui prend un argument un n-échantillon et qui donne l'EMV de α.
4. $\hat{\alpha}$ est-il sans biais ? Si non, construire un estimateur sans biais de α noté $\hat{\alpha}_1$. Modifier alors la fonction matlab précédente pour en déduire un ESB de α.
5. L'estimateur $\hat{\alpha}_1$ est-il efficace, asymptotiquement efficace ?
6. Construire l'histogramme empirique obtenu à partir de p simulations de n-échantillons. Comparer l'histogramme avec la borne de Cramer-Rao.

On suppose le paramètre α connu, on cherche à estimer θ.

1. Donner une statistique exhaustive de θ.
2. Estimer θ par la méthode du maximum de vraisemblance. On notera $\hat{\theta}$ l'EMV de θ.
3. Écrire une fonction emv-theta.m qui prend un argument un n-échantillon et qui donne l'EMV de θ.
4. Donner la loi exacte de $\hat{\theta}$.
5. Calculer l'information de Fisher $I_X(\theta)$ apportée par X sur θ. En déduire $I_n(\theta)$.
6. Calculer $\mathbb{E}(\hat{\theta})$. En déduire un estimateur sans biais $\tilde{\theta}$ de θ.
7. $\hat{\theta}$ est-il un estimateur efficace, asymptotiquement efficace de θ ?
8. Construire l'histogramme empirique obtenu à partir de p simulations de n-échantillons. Comparer l'histogramme avec la borne de Cramer-Rao.

5 Estimation autour de la loi de Poisson

L'objectif est de construire un estimateur sans biais, de variance minimale des probabilités $P(X = k)$ lorsque X suit une loi de Poisson de paramère λ. On note $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{P}(\lambda)$, où λ est un paramètre inconnu à estimer.

1. On cherche tout d'abord à estimer ce paramètre λ, calculer $\mathbb{E}(X_i)$ et proposer un estimateur $\hat{\lambda}_n$ de λ.
2. Montrer que la statistique $S = X_1 + \cdots + X_n$ est une statistique exhaustive et complète. Qu'en conclure pour l'estimateur de λ ?
3. Programmer en Matlab une fonction "*echantillon_poisson.m*" qui prend en arguments le paramètre λ et la taille de l'échantillon n à construire et qui construit un vecteur X, n-échantillon de loi $\mathcal{P}(\lambda)$.
4. De la même façon, programmer une fonction matlab "*estimateur_lambda.m* qui calcule $\hat{\lambda}_n$.
5. On cherche désormais à estimer la probabilité $p_k = P(\mathcal{P}(\lambda) = k)$.
   (a) Une première idée peut être d'utiliser l'estimation $\hat{\lambda}_n$. Que penser de l'estimateur
   $$\hat{p}^1_{n,k} = \frac{e^{-\hat{\lambda}_n} \hat{\lambda}_n^k}{k!}$$
   (b) Une seconde idée est d'utiliser l'estimateur naturel construit à partir des variables aléatoires $Y_i = \chi_{X_i = k}$. Donner un estimateur sans biais de p_k.
   (c) Programmer une fonction matlab "*esb_proba.m*" qui prend en argument l'entier k et un n-échantillon et qui construit l'estimateur sans biais $\hat{p}^2_{n,k}$ précédemment évoqué.
   (d) Calculer la variance de l'estimateur précédent.
   (e) On souhaite désormais utiliser l'amélioration de Rao-Blackwell pour augmenter la précision de l'esb précédent. Calculer l'esbvm $\hat{p}^3_{n,k}$ construit à partir de la statistique exhaustive et de l'esb précédent.
   (f) Programmer une fonction matlab "*esbvm_proba.m*" qui prend en argument l'entier k et un n-échantillon et qui construit l'estimateur sans biais de variance minimale de p_k.
   (g) Calculer la variance de l'estimateur.
   (h) Programmer une fonction matlab "*histo_poisson.m*" qui permet de travec les histogrammes obtenus en soustrayant aux estimateurs précédents la valeur théorique p_k avec comme arguments l'entier k, un nombre d'échantillons tirés N, la longueur n d'un échantillon et enfin le paramètre λ. Quel est le meilleur estimateur ?