FEUILLE DE TD N°8 Efficacité asymptotique

Exercice 40

Soit $(X_1,...,X_n)$ un n-échantillon de la loi $\mathcal{N}(m,\sigma^2)$, où $\theta=(m,\sigma^2)\in\mathbb{R}\times]0,\infty[$ est inconnu. On note $\widehat{\sigma}_n^2$ l'estimateur du maximum de vraisemblance de σ^2 et $R_\theta(\widehat{\sigma}_n^2)$ son risque quadratique en θ . Montrer que

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}_{\theta}} \mathcal{N}(0, 2\sigma^4)$$
 quand $n \to \infty$.

et que de plus, $\lim_{n\to\infty} nR_{\theta}(\widehat{\sigma}_n^2) = 2\sigma^4$. Solution. Pour montrer la normalité asymptotique, on pourrait utiliser le théorème de normalité asymptotique du max de vraisemblance, mais ici, il est plus simple d'utiliser un TLC usuel. En effet,

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2 .$$

Par le théorème de Cochran, $\hat{\sigma}_n^2$ est égal en loi à :

$$\frac{\sigma^2}{n} \sum_{i=1}^{n-1} N_i^2$$
,

où les N_i sont i.i.d de loi $\mathcal{N}(0,1)$. Le TLC nous dit que :

$$\frac{1}{\sqrt{n-1}} \sum_{i=1}^{n-1} N_i^2 - \sqrt{n-1} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0,2) ,$$

Donc (via Slutsky), en multipliant par $\sigma^2 \sqrt{n-1}/\sqrt{n}$:

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) = \sigma^2 \frac{\sqrt{n-1}}{\sqrt{n}} \left(\frac{1}{\sqrt{n-1}} \sum_{i=1}^{n-1} N_i^2 - \sqrt{n-1} \right) - \frac{\sigma^2}{\sqrt{n}} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, 2\sigma^4) .$$

Pour le risque, on remarque que :

$$R_{\theta}(\widehat{\sigma}_n^2) = \mathsf{Var}_{\theta}(\frac{\sigma^2}{n}\sum_{i=1}^{n-1}N_i^2) + \mathbb{E}_{\theta}[\frac{\sigma^2}{n}\sum_{i=1}^{n-1}N_i^2 - \sigma^2]^2 \ .$$

Or:

$$\operatorname{Var}_{\theta}(\frac{\sigma^2}{n}\sum_{i=1}^{n-1}N_i^2) = \frac{2\sigma^4(n-1)}{n^2} ,$$

et:

$$\mathbb{E}_{\theta} \left[\frac{\sigma^2}{n} \sum_{i=1}^{n-1} N_i^2 - \sigma^2 \right]^2 = \left(\frac{\sigma^2(n-1)}{n} - \sigma^2 \right)^2,$$

$$= \frac{\sigma^4}{n^2},$$

Donc $\lim_{n\to\infty} nR_{\theta}(\widehat{\sigma}_n^2) = 2\sigma^4$.

Exercice 42

Soit f une densité de probabilité de classe C^2 sur \mathbb{R} , telle que

- pour tout $x \in \mathbb{R}$, f(x) est strictement positif,
- les fonctions f'' et $(f')^2/f$ sont intégrables par rapport à la mesure de Lebesgue.
- la fonction $f''/f (f'/f)^2$ est uniformément continue sur \mathbb{R} .

Pour tout $\theta \in \mathbb{R}$ et $x \in \mathbb{R}$, on note

$$f_{\theta}(x) = f(x - \theta)$$
 et $l_{\theta}(x) = \log f_{\theta}(x)$.

Pour tout $x \in \mathbb{R}$, la fonction $\theta \mapsto l_{\theta}(x)$ est alors de classe \mathcal{C}^2 sur \mathbb{R} ; on note $l'_{\theta}(x)$ et $l''_{\theta}(x)$ ses dérivées première et seconde au point θ . On note de même $f'_{\theta}(x)$ la dérivée première de $\theta \mapsto f_{\theta}(x)$ au point θ . Enfin, on note X une variable aléatoire admettant la densité f_{θ_0} , $\theta_0 \in \mathbb{R}$.

- 1. Montrer les propriétés suivantes :
 - (a) $\int_{\mathbb{R}} f_{\theta_0}''(x) dx = 0,$

Solution. Tout d'abord, remarquons que $f'_{\theta}(x) = -f'(x-\theta)$ et $f''_{\theta}(x) = f''(x-\theta)$. Comme f'' est intégrable par rapport à la mesure de Lebesgue, f''_{θ} aussi, et avec un changement de variable trivial, on obtient :

$$\int_{\mathbb{R}} f_{\theta_0}''(x) \, dx = \int_{\mathbb{R}} f''(x) \; .$$

Comme f'' est intégrable, f' admet des limites en $+\infty$ et $-\infty$. Ces limites sont nécessairement nulles, car f est intégrable sur \mathbb{R} . Par conséquent,

$$\int_{\mathbb{R}} f_{\theta_0}''(x) \, dx = \int_{\mathbb{R}} f''(x) = \lim_{M \to +\infty} (f'(M) - f'(-M)) = 0 \, .$$

(b) Il existe un voisinage \mathcal{V} de θ_0 et une fonction mesurable réelle h dans $L^1(P_{\theta_0})$ telle que pour tout $\theta \in \mathcal{V}$ et tout $x \in \mathbb{R}$, $|l''_{\theta}(x)| \leq h(x)$,

Solution.

$$l_{\theta}''(x) = \frac{f_{\theta}''(x)}{f_{\theta}(x)} - \frac{f_{\theta}'^{2}(x)}{f_{\theta}(x)^{2}} = \frac{f''(x-\theta)}{f(x-\theta)} - \left(\frac{f'(x-\theta)}{f(x-\theta)}\right)^{2}.$$

Comme la fonction $f''/f - (f'/f)^2$ est uniformément continue sur \mathbb{R} , il existe un voisinage \mathcal{V} de θ_0 tel que pour tout $\theta \in \mathcal{V}$ et tout $x \in \mathbb{R}$,

$$|l''_{\theta}(x) - l''_{\theta_0}(x)| \leq 1$$
.

D'où, pour tout $\theta \in \mathcal{V}$ et tout $x \in \mathbb{R}$

$$|l_{\theta}''(x)| \leq |l_{\theta_0}''(x)| + 1$$
.

Posons $h: x \mapsto |l''_{\theta_0}(x)| + 1$. On peut voir que h est dans $L^1(P_{\theta_0})$, en effet :

$$\int_{\mathbb{R}} |l''_{\theta_0}(x)| f_{\theta_0}(x) dx \leqslant \int_{\mathbb{R}} \left[\left| \frac{f''(x - \theta_0)}{f(x - \theta_0)} \right| + \left(\frac{f'(x - \theta_0)}{f(x - \theta_0)} \right)^2 \right] f(x - \theta_0) dx ,$$

$$= \int_{\mathbb{R}} \left[\left| \frac{f''(x)}{f(x)} \right| + \left(\frac{f'(x)}{f(x)} \right)^2 \right] f(x) dx ,$$

$$= \int_{\mathbb{R}} \left[\left| f''(x) \right| + \frac{f'(x)^2}{f(x)} \right] dx ,$$

$$< \infty$$

par hypothèse.

(c) L'information de Fisher de l'observation X est finie et strictement positive. Elle ne dépend que de f et est égale à

$$I(f) = \int \frac{(f'(x))^2}{f(x)} dx.$$

Solution.

$$I(f) = \int_{\mathbb{R}} l'_{\theta_0}(x)^2 f_{\theta_0}(x) dx$$

$$= \int_{\mathbb{R}} \left(\frac{f'(x - \theta_0)}{f(x - \theta_0)}\right)^2 f(x - \theta_0) dx$$

$$= \int_{\mathbb{R}} \frac{f'(x)^2}{f(x)} dx$$

Exercice 43

Soit $(\xi_1, ..., \xi_n)$ un *n*-échantillon d'une loi admettant une densité strictement positive par rapport à la mesure de Lebesgue, de fonction de répartition $F(...-\theta)$, où $\theta \in \mathbb{R}$ est inconnu et F est connue et de classe C^2 . A chaque tirage, on observe seulement

$$X_i = (\mathbb{I}_{]-\infty,a]}(\xi_i), \mathbb{I}_{]a,b]}(\xi_i), \mathbb{I}_{]b,\infty[}(\xi_i)).$$

On souhaite construire un estimateur asymptotiquement efficace de θ sur la base des observations $(X_1, ..., X_n)$.

1. Calculer l'information de Fisher de l'échantillon $(X_1, ..., X_n)$.

Solution. Notons μ_n la mesure de comptage sur $\mathcal{X} = \{(1,0,0),(0,1,0),(0,0,1)\}^n$. Le modèle est donné par $(P_{n,\theta})_{\theta \in \mathbb{R}}$ avec :

$$\forall x \in \mathcal{X}, \ f_{n,\theta}(x) := \frac{dP_{n,\theta}}{d\mu_n}(x) = \prod_{i=1}^n F(a-\theta)^{X_i(1)} (F(b-\theta) - F(a-\theta))^{X_i(2)} (1 - F(b-\theta))^{X_i(3)}.$$

Soit, en notant $N_j = \sum_{i=1}^n X_i(j)$:

$$\forall x \in \mathcal{X}, \ f_{n,\theta}(x) = F(a-\theta)^{N_1} (F(b-\theta) - F(a-\theta))^{N_2} (1 - F(b-\theta))^{N_3}.$$

On sait que $I_n(\theta) = nI(\theta)$, où $I_n(\theta)$ est l'information de Fisher de l'échantillon $(X_1, ..., X_n)$ et $I(\theta)$ est l'information de Fisher de X_1 . On a :

$$l_{\theta}(X)^{2} = X_{1}(1)\frac{f(a-\theta)^{2}}{F(a-\theta)^{2}} + X_{1}(2)\frac{(f(b-\theta) - f(a-\theta))^{2}}{(F(b-\theta) - F(a-\theta))^{2}} + X_{1}(3)\frac{f(b-\theta)^{2}}{(1 - F(b-\theta))^{2}},$$

d'où:

$$I(\theta) = \frac{f(a-\theta)^2}{F(a-\theta)} + \frac{(f(b-\theta) - f(a-\theta))^2}{(F(b-\theta) - F(a-\theta))} + \frac{f(b-\theta)^2}{1 - F(b-\theta)}.$$

2. Soit $S_a = \sum_{i=1}^n \mathbb{I}_{X_i = (1,0,0)}$. On considère l'estimateur $\widetilde{\theta}_n = a - F^{-1}(S_a/n)$. Cet estimateur est-il asymptotiquement efficace?

Solution. D'après le TCL,

$$\sqrt{n}\left(\frac{S_a}{n} - F(a-\theta)\right) \xrightarrow[n\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$
,

où $\sigma^2 = F(a-\theta)(1-F(a-\theta))$. La méthode delta nous donne alors :

$$\sqrt{n}\left(F^{-1}(S_a/n)-(a-\theta)\right) \xrightarrow[n\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2((F^{-1})'(F(a-\theta)))^2)$$
,

c'est à dire:

$$\sqrt{n}\left(\theta - \widetilde{\theta}_n\right) \xrightarrow[n \infty]{\mathcal{L}} \mathcal{N}(0, F(a-\theta)(1-F(a-\theta))/f(a-\theta)^2)$$
.

Il faut alors comparer $f(a-\theta)^2/F(a-\theta)(1-F(a-\theta))$ et $I(\theta)$. En notant :

$$A = F(a - \theta) B = F(b - \theta) \alpha = f(a - \theta) \beta = f(b - \theta),$$

On a:

$$\begin{split} &I(\theta) - f(a-\theta)^2 / F(a-\theta)(1-F(a-\theta)) \\ &= \frac{\alpha^2}{A} - \frac{\alpha^2}{A(1-A)} + \frac{(\beta-\alpha)^2}{B-A} + \frac{\beta^2}{1-B} \,, \\ &= -\frac{\alpha^2}{1-A} + \frac{(\beta-\alpha)^2}{B-A} + \frac{\beta^2}{1-B} \,, \\ &= \frac{(\beta-\alpha)^2(1-A)(1-B) + \beta^2(B-A)(1-A) - \alpha^2(B-A)(1-B)}{(1-A)(B-A)(1-B)} \,, \\ &= \frac{(\beta-\alpha)^2(1-A)(1-B) + \beta^2((1-A)-(1-B))(1-A) - \alpha^2((1-A)-(1-B))(1-B)}{(1-A)(B-A)(1-B)} \,, \\ &= \frac{[(\beta-\alpha)^2 - \beta^2 - \alpha^2](1-A)(1-B) + \beta^2(1-A)^2 + \alpha^2(1-B)^2}{(1-A)(B-A)(1-B)} \,, \\ &= \frac{-2\alpha\beta(1-A)(1-B) + \beta^2(1-A)^2 + \alpha^2(1-B)^2}{(1-A)(B-A)(1-B)} \,, \\ &= \frac{[\beta(1-A) - \alpha(1-B)]^2}{(1-A)(B-A)(1-B)} \,, \end{split}$$

On en déduit donc que S_a/n est asymptotiquement efficace si et seulement si :

$$\forall \theta \in \mathbb{R}, \quad \frac{f(b-\theta)}{1-F(b-\theta)} = \frac{f(a-\theta)}{F(a-\theta)}.$$

Or ceci implique (par intégration) qu'il existe une constante C telle que :

$$\forall \theta \in \mathbb{R}, \quad \log(1 - F(b - \theta)) = \log(1 - F(a - \theta)) + C.$$

D'où:

$$\forall \theta \in \mathbb{R}, \quad 1 - F(b - \theta) = e^C (1 - F(a - \theta))$$
.

En faisant tendre θ vers $-\infty,$ on obtient $e^C=1.$ D'où :

$$\forall \theta \in \mathbb{R}, \quad 1 - F(b - \theta) = (1 - F(a - \theta)).$$

Mais ceci est bien sûr impossible si $a \neq b$, puisque F est une fonction de répartition (elle ne peut être périodique). En conclusion, si $a \neq b$, S_a/n n'est pas asymptotiquement efficace, et si a = b, S_a/n est efficace.

3. Construire un estimateur asymptotiquement efficace de θ .

Solution. On utilise la méthode de Newton. Il faut vérifier les hypothèses de régularité pour la normalité asymptotique du maximum de vraisemblance. Comme on a un espace d'observation discret, c'est assez facile, toutes les hypothèses nécessaires pour intervertir intégrale et dérivation sont satisfaites, il reste à voir que l'information de Fisher est inversible (conséquence de f > 0) et que $\theta \mapsto f_{\theta}$ est C^2 (conséquence de F C^2).