Examen de contrôle continu - mercredi 23 octobre 2024
Durée 2h - Seul le formulaire de probabilités est autorisé

Exercice 1 (4 pts) On considère un n -échantillon $x_1, \dots, x_n$ distribué suivant une loi $\mathcal{N}(\theta, 1)$ et la loi a priori impropre $\pi(\theta) \propto 1$.

On considère la fonction perte $L_a(\theta, d) = (\exp(a(\theta - d)) - a(\theta - d) - 1)$ avec $a > 0$, qui est une alternative aux fonctions classiques. Elle est dite LINEX (linear-exponential). Elle a été introduite par Varian en 1974.

1 (2 pts) Montrer que $L_a(\theta, d)$ ne prend que des valeurs positives et qu'elle est convexe. Décrire clairement comment elle pénalise les différentes erreurs d'estimation sur le paramètre θ .

2 (2 pts) Donner l'expression de l'estimateur bayésien de θ .

Exercice 2 (4 pts) On considère la variable aléatoire X admettant pour densité

$$f_X(x) \propto \exp(-x/2)\mathbb{I}_{[2,10]}(x).$$

Proposer deux méthodes de simulation exacte de X .

Exercice 3 (5 pts) On souhaite estimer

$$\theta = \int_0^\pi \left(\frac{1 + \exp(x)}{1 + \sin(x)} \right) dx.$$

1 (1 pt) Proposer un estimateur $\hat{\theta}$ de θ par une méthode de Monte Carlo.

2 (2 pts) Donner un intervalle de confiance de niveau asymptotiquement supérieur ou égal à 95% pour θ . Cet intervalle doit être calculable en pratique.

3 (2 pts) En utilisant une approximation asymptotique, déterminer le nombre de simulation nécessaire pour que l'erreur absolue relative de $\hat{\theta}$ soit inférieure à 1% avec une probabilité supérieure ou égale à 99%.

Exercice 4 (7 pts) On considère un n -échantillon $x_1, \dots, x_n$ distribué suivant une loi $\mathcal{N}(0, 1/\theta)$.

1 (2 pts) Donner la loi a priori de Jeffreys de θ .

2 (1 pts) En utilisant la loi de Jeffreys, donner un intervalle de crédibilité de niveau 95% pour θ .

On considère maintenant dans la suite la loi a priori $\pi(\theta) \propto \log(\theta)\theta^{-1}\mathbf{1}_{\theta \in [1,10]}$.

3 (2 pts) Proposer une méthode permettant d'obtenir des simulations suivant la loi a posteriori de θ .

4 (1 pt) Donner une stratégie permettant d'obtenir une approximation de l'estimateur bayésien de θ associé à la fonction de perte quadratique.

5 (1 pt) Donner une stratégie permettant d'obtenir une approximation d'un intervalle de crédibilité de θ de niveau 99%.

Correction examen contrôle
continu 23/10/2024
HAX 918 X

①

Exercice 1

1) Soit $f(x) = e^x - x - 1$

$$f'(x) = e^x - 1$$

Si $x > 0$, $f'(x) > 0$

$f(x)$ croissant et $f(0) = 0$

Donc $f(x) > 0$ si $x > 0$

Si $x < 0$, $f'(x) < 0$

$f(x)$ décroissant et $f(0) = 0$

Donc $f(x) > 0$ si $x < 0$

Ainsi $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

Et ainsi, pour $a > 0$,

(2)

$$L_a(\theta, d) \geq 0 \quad \forall \theta \in \mathbb{R} \\ \forall d \in \mathbb{R}$$

On verra, $f''(x) = e^x > 0$.

Ainsi, $L_a(\theta, d)$ est convexe
en (θ, d) .

Si on sous-estime fortement θ
alors la perte est exponentielle
et très forte.

Si on sur-estime fortement θ
alors la perte est linéaire et
moins forte.

Avec une perte LINEX, on cherche
à éviter de sous-estimer θ .

2] Sam trawon l'estimateur bayésien, on minimise la perte moyenne relativement à la loi a priori.

(3)

$$\begin{aligned}
 & \mathbb{E}^{\pi} [L_a(\theta, d) | \underline{\kappa}] \quad \underline{\kappa} = (\kappa_1, \dots, \kappa_n) \\
 &= \mathbb{E}^{\pi} [e^{a(\theta-d)} - a(\theta-d) - 1 | \underline{\kappa}] \\
 &= \mathbb{E}^{\pi} [e^{a(\theta-d)} | \underline{\kappa}] - a \mathbb{E}^{\pi} (\theta | \underline{\kappa}) + a d - 1 \\
 &= e^{-ad} \underbrace{\mathbb{E}^{\pi} [e^{a\theta} | \underline{\kappa}] - a \mathbb{E}^{\pi} (\theta | \underline{\kappa}) + ad - 1}
 \end{aligned}$$

$$h'(d) = -a e^{-ad} \mathbb{E}^{\pi} [e^{a\theta} | \underline{\kappa}] + a$$

$$h''(d) = a^2 e^{-ad} \mathbb{E}^{\pi} [e^{a\theta} | \underline{\kappa}] > 0$$

$h(\alpha)$ est une fonction strictement $\searrow$ (4)
convexe - On cherche donc
 α^* tel que

$$h'(\alpha^*) = 0$$

$$\Leftrightarrow e^{-\alpha \alpha^*} \mathbb{E}^\pi \left[e^{\alpha \theta} \mid \underline{\kappa} \right] = 1$$

$$\Leftrightarrow \alpha^* = \left(\frac{1}{\alpha} \right) \log \left[\mathbb{E}^\pi \left[e^{\alpha \theta} \mid \underline{\kappa} \right] \right]$$

L'estimateur bayésien est donc

$$\hat{\theta} = \left(\frac{1}{\alpha} \right) \log \left[\mathbb{E}^\pi \left[e^{\alpha \theta} \mid \underline{\kappa} \right] \right]$$

Mais nous avons donc calculé

$\mathbb{E}^\pi \left[e^{\alpha \theta} \mid \underline{\kappa} \right]$. Il nous a fallu
explicitement la loi a posteriori
de θ .

$$\pi(\theta | \underline{\kappa}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^M (\kappa_i - \theta)^2 \right\} \quad (5)$$

$$\pi(\theta | \underline{\kappa}) \propto \exp \left\{ -\frac{1}{2} \left(M\theta^2 - 2\theta \sum_{i=1}^M \kappa_i \right) \right\}$$

$$\pi(\theta | \underline{\kappa}) \propto \exp \left\{ -\frac{M}{2} \left(\theta - \bar{\kappa}_M \right)^2 \right\}$$

$$\Rightarrow \theta | \underline{\kappa} \sim \mathcal{N} \left(\bar{\kappa}_M, \frac{1}{M} \right)$$

$$\bar{\kappa}_M = \frac{1}{M} \sum_{i=1}^M \kappa_i$$

$$A_{i=1}^M \kappa_i, \quad \mathbb{E}^\pi \left[e^{u\theta} | \underline{\kappa} \right] = \int_{\mathbb{R}} e^{u\theta} \pi(\theta | \underline{\kappa}) d\theta$$

$$= \int_{\mathbb{R}} \frac{\sqrt{M}}{\sqrt{2\pi}} e^{-\frac{M}{2}(\theta - \bar{\kappa}_M)^2 + u\theta} d\theta$$

$$= \int_{\mathbb{R}} \frac{\sqrt{M}}{\sqrt{2\pi}} e^{-\frac{M}{2} \bar{\kappa}_M^2} \exp \left(-\frac{M}{2} \left[\theta^2 - 2\theta \bar{\kappa}_M - \frac{2u\theta}{M} \right] \right) d\theta$$

$$= e^{-\frac{M}{2} \bar{\kappa}_M^2} e^{\frac{M}{2} \left(\bar{\kappa}_M + \frac{u}{M} \right)^2} = e^{u \bar{\kappa}_M + \frac{u^2}{2M}}$$

Nous obtenons

$$\hat{\theta} = \bar{X}_M + \frac{a}{2M}$$

Estimateur qui exprime le fait que l'on ne veut surtout pas sous-estimer θ !

Exercice 2

Méthode 1 Utilisation de l'algorithme Acceptation-Rejet
On propose une simulation suivant une loi exponentielle de paramètre $\left(\frac{1}{2}\right)$ et on l'accepte si elle appartient à l'intervalle $[2, 10]$

Méthode 2 Utilisation l'inverse de la fonction de répartition

(6)

Calculons la constante de normalisation de $f_X(x)$: (7)

$$\int_{\mathbb{R}} e^{-x/2} \mathbb{1}_{[2,10]}(x) dx = \int_2^{10} e^{-x/2} dx$$

$$= \left[-2e^{-x/2} \right]_2^{10} = 2e^{-2} - 2e^{-5}$$

Ainsi, $f_X(x) = \frac{e^{-x/2} \mathbb{1}_{[2,10]}(x)}{2(e^{-2} - e^{-5})}$

Calculons maintenant la fonction de répartition de X

$$F_X(x) = \begin{cases} 0 & \text{si } x < 2 \\ 1 & \text{si } x > 10 \\ \int_2^x \frac{e^{-t/2}}{2(e^{-2} - e^{-5})} dt & \text{si } x \in [2, 10] \end{cases}$$

Si $x \in [2, 10]$, $F_X(x) = \left[\frac{e^{-2} - e^{-x/2}}{e^{-2} - e^{-5}} \right]$

Inversons la fonction de répartition

(8)

Soit $u \in]0, 1[$

$$u = \left[\frac{e^{-1} - e^{-\kappa/2}}{e^{-1} - e^{-5}} \right]$$

$$\Leftrightarrow u(e^{-1} - e^{-5}) = e^{-1} - e^{-\kappa/2}$$

$$\Leftrightarrow e^{-\kappa/2} = e^{-1} - u(e^{-1} - e^{-5})$$

$$\Leftrightarrow -\frac{\kappa}{2} = \log(e^{-1} - u(e^{-1} - e^{-5}))$$

$$\Leftrightarrow \kappa = -2 \log(e^{-1} - u(e^{-1} - e^{-5}))$$

Soit $U \sim \mathcal{U}[0, 1]$

$$X = -2 \log(e^{-1} - U(e^{-1} - e^{-5}))$$

est distribuée suivant $f_X(x)$.

Exercice 3

(9)

$$1) \hat{\theta}_N = \frac{1}{N} \sum_{i=1}^N \pi \left[\frac{1 + e^{X_i}}{1 + \sin(X_i)} \right]$$

où $X_1, \dots, X_N \stackrel{iid}{\sim} \mathcal{U}[0, \pi]$

2) D'après le TCL, comme $E(\hat{\theta}_N) = \theta$

$$(\hat{\theta}_N - \theta) \underset{N \text{ grand}}{\#} \mathcal{N}(0, V(\hat{\theta}_N))$$

$$V(\hat{\theta}_N) = \frac{1}{N} \int_{[0, \pi]} \left(\pi \left[\frac{1 + e^x}{1 + \sin(x)} \right] \right)^2 dx - \theta^2$$

$$V(\hat{\theta}_N) = \frac{1}{N} \left[\int_{[0, \pi]} \left(\frac{1 + e^x}{1 + \sin(x)} \right)^2 \pi^2 dx - \theta^2 \right]$$

Cherchons une borne supérieure de $V(\hat{\theta}_N)$

(10)

$$\forall \kappa \in [0, \pi]$$

$$1 + e^\kappa \leq 1 + e^\pi$$

$$1 + \cos(\kappa) \geq 1$$

$$\frac{1 + e^\kappa}{1 + \cos(\kappa)} \leq 1 + e^\pi$$

$$\forall \kappa \in [0, \pi]$$

$$\left(\frac{1 + e^\kappa}{1 + \cos(\kappa)} \right)^2 \leq (1 + e^\pi)^2 \quad \left| \quad 1 \leq \frac{1 + e^\kappa}{1 + \cos(\kappa)} \right.$$

$$\text{AIMM}, \quad \mathbb{E}_{\mathcal{U}_{[0, \pi]}} \left(\frac{\pi^2 (1 + e^x)^2}{(1 + \cos(x))^2} \right)$$

$$= \int_0^\pi \frac{\pi^2 (1 + e^\kappa)^2}{(1 + \cos(\kappa))^2} d\kappa \leq \pi^2 (1 + e^\pi)^2$$

$$\text{AIMM}, \quad \Theta = \int_0^\pi \left(\frac{1 + \cos(\kappa)}{1 + \cos(\kappa)} \right) d\kappa \geq \pi$$

$$\Leftrightarrow -\Theta^2 \leq -\pi^2$$

$$\text{AIMM}, \quad \forall (\hat{\Theta}_N) \leq \frac{1}{N} \left[\pi^2 \left((1 + e^\pi)^2 - 1 \right) \right]$$

$$\forall (\hat{\Theta}_N) \leq \frac{\pi^2 e^\pi}{N} (2 + e^\pi)$$

D'après la TCL, pour N grand (11)

$$\mathbb{P} \left[\sqrt{V(\hat{\theta}_N)} F_{N(0,1)}^{-1}(0,025) \leq \hat{\theta}_N - \theta \leq \sqrt{V(\hat{\theta}_N)} F_{N(0,1)}^{-1}(0,975) \right] = 0,95$$

Donc, comme $V(\hat{\theta}_N) \leq \frac{\pi^2 e^\pi}{N} (2+e^\pi)$

alors

$$\mathbb{P} \left[-\frac{\pi e^{\pi/2} \sqrt{2+e^\pi}}{\sqrt{N}} 1,96 \leq \hat{\theta}_N - \theta \leq \frac{\pi e^{\pi/2} \sqrt{2+e^\pi}}{\sqrt{N}} 1,96 \right] \geq 0,95$$

$$IC_{0,95}(\theta) = \left[\hat{\theta}_N \pm 1,96 \frac{\pi e^{\pi/2} \sqrt{2+e^\pi}}{\sqrt{N}} \right]$$

3) On cherche la plus petite valeur de N telle que

(12)

$$\mathbb{P} \left[\frac{|\hat{\theta}_N - \theta|}{|\theta|} \leq 0,01 \right] \geq 0,99$$

Comme $\theta > 0$,

$$\mathbb{P} \left[\frac{|\hat{\theta}_N - \theta|}{|\theta|} \leq 0,01 \right] =$$

$$\mathbb{P} \left[\left| \frac{\hat{\theta}_N - \theta}{\theta} \right| \leq 0,01 \right]$$

on a alors,

$$\frac{\hat{\theta}_N - \theta}{\theta} \underset{N \text{ grand}}{\#} \mathcal{N} \left(0, \frac{V(\hat{\theta}_N)}{\theta^2} \right)$$

Param N grand, mais on ne
 s'occupe

(13)

$$\mathbb{P} \left[\left| \frac{\hat{\theta}_N - \theta}{\theta} \right| \leq 0,01 \right] \stackrel{N \text{ grand}}{\approx} \mathbb{P} \left[|N(\theta)| \leq \sqrt{\frac{\theta^2}{V(\hat{\theta}_N)}} \cdot 0,01 \right]$$

$$\# 2 F_{N(\theta)} \left(0,01 \sqrt{\frac{\theta^2}{V(\hat{\theta}_N)}} \right) - 1$$

On cherche le plus petit valeur
 de N telle que

$$2 F_{N(\theta)} \left(0,01 \sqrt{\frac{\theta^2}{V(\hat{\theta}_N)}} \right) - 1 \geq 0,99$$

$$\Leftrightarrow 0,01 \sqrt{\frac{\theta^2}{V(\hat{\theta}_N)}} \geq 2,6 \quad \left(F_{N(\theta)}^{-1}(0,995) = 2,57 \dots \right)$$

$$\Leftrightarrow \sqrt{\frac{V(\hat{\theta}_N)}{\theta^2}} \leq (260)^{-2}$$

$$\text{Comme } \frac{V(\hat{\theta}_N)}{\theta^2} \leq \frac{e^\pi (2+e^\pi)}{N}$$

$$N \geq (260)^2 \times e^\pi \times (2+e^\pi)$$

$$N \geq 39\,327\,858$$

Esercizio 4

(14)

$$1) f(x|\theta) = \frac{\sqrt{\theta}}{\sqrt{2\pi}} e^{-\frac{\theta}{2}x^2}$$

$$\log(f(x|\theta)) = \frac{1}{2} \log(\theta) - \log(\sqrt{2\pi}) - \frac{\theta}{2}x^2$$

$$\frac{d \log f}{d\theta}(x|\theta) = \frac{1}{2\theta} - \frac{x^2}{2}$$

$$\frac{d^2 \log f}{(d\theta)^2}(x|\theta) = -\frac{1}{4\theta^2}$$

$$I_{\theta}(\theta) = \frac{1}{4\theta^2}$$

$$\pi^J(\theta) \propto \theta^{-2} \mathbb{1}_{\{\theta > 0\}}$$

ha una priori impropria

$$2) \pi^J(\theta|\underline{\kappa}) \propto \theta^{m/2} e^{-\frac{\theta}{2} \sum_{i=1}^m \kappa_i^2} \theta^{-2} \mathbb{1}_{\{\theta > 0\}}$$

$$\Rightarrow \theta|\underline{\kappa} \sim \text{Nta}\left(\frac{m}{2}, \frac{\sum_{i=1}^m \kappa_i^2}{2}\right)$$

En utilisant les quantiles de la loi Gamma, on obtient un intervalle de crédibilité pour θ au niveau 95%.

(15)

$$\left[F^{-1}\left(\frac{\alpha}{2}, \frac{\sum_{i=1}^m \frac{1}{z_i}\right), F^{-1}(0,975), \frac{\sum_{i=1}^m \frac{1}{z_i}}{2}\right)$$

3] $\pi(\theta | \mathcal{L}) \propto \theta^{m/2} e^{-\theta/2 \sum_{i=1}^m \frac{1}{z_i}} \propto \text{log}(\theta) \pi(\theta)$
[1,10]

let's check Acceptation-Rejet

for the proposition $\theta \in \left(\frac{m}{2}, \frac{\sum_{i=1}^m \frac{1}{z_i}}{2}\right)$

$$P(\theta) \propto \theta^{m/2-1} e^{-\theta/2 \sum_{i=1}^m \frac{1}{z_i}} \pi(\theta)$$

$$\frac{\pi(\theta | \mathcal{L})}{P(\theta)} = \frac{\text{log}(\theta) \pi(\theta)}{\pi(\theta)}$$

$$\leq \text{log}(20)$$

On simule une réalisation
 suivant une loi normale $(\frac{\mu}{2}, \frac{\sum \sigma_i^2}{2})$
 notée, on accepte cette simulation
 comme une simulation suivant
 $\pi(\theta | \underline{K})$ si $U \leq \frac{\log(\gamma) \pi(\gamma)}{\log(10) [1, 10]}$
 on a $U \sim \mathcal{U}[0, 1]$.

4) A l'issue de l'algorithme
 précédent, on obtient un
 N-échantillon $\theta^{(1)}, \dots, \theta^{(N)}$
 iid suivant $\pi(\theta | \underline{K})$
 On estime la moyenne a
 posteriori par $\frac{1}{N} \sum_{i=1}^N \theta^{(i)}$
 méthode de Monte Carlo

5) On utilise les
quantiles empiriques
de niveau $2,5\%$ et $97,5\%$
sur l'échantillon ind
général et le vide de
l'algorithme proposé à
la question 3].

(17)