Université de Montpellier
M2 Biostatistique - HAX918X

Année 2021-2022

Examen de rattrapage - mardi 21 juin 2021
Durée 1h30 - Documents non autorisés

Exercice 1 (4 pts)

Une variable aléatoire y distribuée suivant une loi Inverse Gamma de paramètres $\alpha > 0$ et $\beta > 0$ admet pour densité de probabilité

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\frac{\beta}{y}\right) \mathbb{I}_{y>0}\,.$$

Nous admettons que $\mathbb{E}(y) = \dfrac{\beta}{\alpha-1}$ et $\mathbb{V}(y) = \dfrac{\beta}{(\alpha-1)^2(\alpha-2)}$.

Nous modélisons la distance y à laquelle se trouve une particule de son point de départ (après un temps fixé) par une loi de Rayleigh de paramètre $\theta > 0$

$$f(y|\theta) = \frac{y}{\theta} \exp\left(-\frac{y^2}{2\theta}\right) \mathbb{I}_{y\geq 0}\,.$$

Nous admettons que $\mathbb{E}(y|\theta) = \sqrt{\dfrac{\theta\pi}{2}}$ et $\mathbb{E}(y^2|\theta) = 2\theta$.
Nous observons un n-échantillon $y_1, \ldots, y_n$ et souhaitons estimer θ. Nous nous plaçons dans le paradigme bayésien et supposons que θ est une variable aléatoire.

1) (1 pt) Montrer que la famille des lois Inverse Gamma est conjuguée.

2) (2 pts) Donner la loi a priori de Jeffreys pour θ. Est-elle propre ?

3) (1 pt) Pour la loi a priori de Jeffreys, donner l'estimateur associé à la fonction de perte $L(\theta, d) = (\theta - d)^2$.

Exercice 2 (4 pts)

Adam et Jean-Michel jouent aux fléchettes. Adam prétend pouvoir atteindre le centre de la cible 1 fois sur 2. Jean-Michel pense lui que Adam n'atteindra le centre que 1 fois sur 4. Adam effectue 10 essais et atteint le centre 3 fois.
Modéliser le problème et donner la réponse bayésienne (indication : $2^{-10} \times 3^7 \approx 2,14$).
Expliquer en quoi le résultat obtenu est intuitif.

Exercice 3 (6 pts)

Nous considérons la variable aléatoire X admettant pour densité

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

1 (4 pts) Proposer deux méthodes de simulation exacte de X.

2 (2 pts) Proposer une méthode de simulation approchée de X.

Exercice 4 (6 pts)

On souhaite estimer

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

1 (2 pts) Proposer un estimateur $\hat{\theta}$ de θ par une méthode de Monte Carlo.

2 (2 pts) Sans 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%.

3 (2 pts) Répondre à la même question que précédemment en utilisant une approximation asymptotique.

Correction examen de rattrapage - HAX918X
Statistique bayésienne
21 juin 2022

Exercice 1

1] $f(y|\theta) \propto \theta^{-1} e^{-\frac{y^2}{2\theta}}$

$\pi(\theta) \propto \theta^{-\alpha-1} e^{-\beta/\theta} \mathbb{1}\{\theta > 0\}$

$\pi(\theta|y) \propto \theta^{-\alpha-1-1} e^{-\frac{1}{\theta}(\beta + y^2/2)} \mathbb{1}\{\theta > 0\}$

$\Rightarrow \theta|y \sim IG(\alpha+1, \beta + y^2/2)$

loi conjuguée

2] Si $y > 0$

$\log(f(y|\theta)) = \log(y) - \log(\theta) - \frac{y^2}{2\theta}$

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

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

$$I_y(\theta) = \frac{\mathbb{E}(y^2|\theta)}{\theta^3} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

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

loi impropre

3] $\pi^J(\theta|y_1, ---, y_n)$

$$\propto \theta^{-n-1} e^{-\frac{1}{\theta}\sum_{i=1}^{n} y_i^2/2}$$

$$\Rightarrow \mathbb{E}^J(\theta|y_1, ---, y_n) = \frac{\sum_{i=1}^{n} y_i^2}{2(n-1)}$$

Exercice 2

$y|\theta \sim \mathcal{B}(10, \theta)$

$\theta \in \{\frac{1}{2}, \frac{1}{4}\}$ $\quad y = 3$

$\mathbb{P}(\theta = \frac{1}{2}) = \mathbb{P}(\theta = \frac{1}{4}) = \frac{1}{2}$

loi a priori uniforme

$\mathbb{P}(\theta = \frac{1}{2} | y = 3) \propto C_{10}^{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)$

$\mathbb{P}(\theta = \frac{1}{4} | y = 3) \propto C_{10}^{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^7 \left(\frac{1}{2}\right)$

$\Rightarrow \mathbb{P}(\theta = \frac{1}{2} | y = 3) = \dfrac{2^{-10}}{2^{-10} + 4^{-10} 3^7}$

$\mathbb{P}(\theta = \frac{1}{2} | y = 3) = 1 / (1 + 2^{-10} 3^7)$

$\# \dfrac{1}{3,14}$

$< \mathbb{P}(\theta = \frac{1}{4} | y = 3)$

On choisit $\hat{\theta} = \frac{1}{4}$

Intuitif car 0,3 est plus proche de 0,25 que de 0,5

4

Exercice 3

1) Méthode 1

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 \text{ si } x < 2 \\ 1 \text{ si } x \geq 10 \end{cases}$$

Soit $x \in [2, 10[$

$$F_X(x) = \int_2^x f_X(y)\, dy$$

$$= \left(\frac{2}{e^{-1} - e^{-5}}\right) \int_2^x e^{-y/2}\, dy$$

$$= \left[\frac{2(e^{-1} - e^{-x/2})}{e^{-1} - e^{-5}}\right]$$

Soit $y \in]0, 1[$

$$y = \left[\frac{2(e^{-1} - e^{-x/2})}{e^{-1} - e^{-5}}\right]$$

$$\Leftrightarrow \frac{(e^{-1} - e^{-5})\, y}{2} = e^{-1} - e^{-x/2}$$

$$\Leftrightarrow -\frac{x}{2} = \log\left(e^{-1} - \left(\frac{e^{-1} - e^{-5}}{2}\right) y\right)$$

$$\Leftrightarrow x = -2 \log\left[e^{-1} - \left[\frac{e^{-1} - e^{-5}}{2}\right] y\right]$$

Ainsi, si $U \sim \mathcal{U}_{[0,1]}$ alors

$$X = -2 \log\left[e^{-1} - \left[\frac{e^{-1} - e^{-5}}{2}\right] y\right]$$

$$\sim f_X(\cdot)$$

Méthode 2

Algorithme acceptation-rejet avec une loi exponentielle de paramètre $\frac{1}{2}$ comme loi instrumentale

On simule des réalisations suivant une loi $\mathcal{E}xp\left(\frac{1}{2}\right)$ et on accepte dès que la réalisation appartient à l'intervalle $[2, 10]$.

(6)

3] Algorithme de Hastings-Metropolis
loi de proposition marche aléatoire gaussienne de variance 0.1

Exercice 4

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

où $X_1, ----, X_N \overset{iid}{\sim} U_{[0,\pi]}$

2] $\mathbb{P}(|Y-\mathbb{E}(Y)| > \varepsilon) \leq \frac{\mathbb{V}(Y)}{\varepsilon^2}$

$\Rightarrow \mathbb{P}(|\widehat{\Theta}-\mathbb{E}(\widehat{\Theta})| > \varepsilon) \leq \frac{\mathbb{V}(\widehat{\Theta})}{\varepsilon^2}$

$\mathbb{E}(\widehat{\Theta}) = \theta$

$\mathbb{V}(\widehat{\Theta}) = \frac{\pi^2}{N} \mathbb{V}\left(\frac{1+e^X}{1+\sin(X)} \right)$

$\mathbb{V}(\widehat{\Theta}) = \frac{1}{N} \left(\mathbb{E}\left[\frac{\pi^2 (1+e^X)^2}{(1+\sin(X))^2} \right] - \theta^2 \right)$

$$\mathbb{P}\left(|\hat{\theta}-\mathbb{E}(\hat{\theta})| \leq \varepsilon\right) \geq 1 - \frac{\mathbb{V}(\hat{\theta})}{\varepsilon^2}$$

$$\Rightarrow \mathbb{P}\left(\frac{|\hat{\theta}-\mathbb{E}(\hat{\theta})|}{\theta} \leq 10^{-2}\right) \geq 1 - \frac{\mathbb{V}(\hat{\theta})}{10^{-4}\theta^2}$$

On cherche la plus petite valeur de N telle que

$$1 - \frac{\mathbb{V}(\hat{\theta})}{\theta^2 10^{-4}} \geq 0.99$$

telle que $\frac{\mathbb{V}(\hat{\theta})}{\theta^2} \leq 10^{-6}$

telle que $N \geq 10^6 \left[\frac{\pi^2 \mathbb{E}\left[\frac{(1+e^x)^2}{(1+\sin(x))^2}\right]}{\theta^2} - 1\right]$

$$\theta \geq \int_0^\pi \frac{2}{1+\sin(x)}\,dx$$

$$\theta \geq \int_0^\pi dx = \pi$$

$$\Rightarrow \frac{1}{\theta^2} \leq \frac{1}{\pi^2}$$

$$\frac{\pi^2 \, \mathbb{E}\left[\frac{(1+e^X)^2}{(1+\sin(X))^2}\right]}{\theta^2} - 1$$

$$\le \mathbb{E}\left[\frac{(1+e^X)^2}{(1+\sin(X))^2}\right] - 1$$

$$\le \frac{(1+e^\pi)^2}{1} - 1$$

$$\le 1 + e^{2\pi} + 2e^{\pi} - 1$$

$$\le e^{2\pi} + 2e^{\pi}$$

$$\Longrightarrow N \ge 10^6\left[e^{2\pi} + 2e^{\pi}\right]$$

3) $$\frac{\hat{\theta}-\theta}{\sqrt{\mathbb{V}(\hat{\theta})}} \; \# \; N(0,1)$$

$$\mathbb{P}\left(\frac{|\hat{\theta}-\theta|}{\theta} \le 10^{-2}\right) = \mathbb{P}\left(|N(0,1)| \le \frac{10^{-2}\theta}{\sqrt{\mathbb{V}(\hat{\theta})}}\right)$$

$$= F_{N(0,1)}\left(\frac{10^{-2}\theta}{\sqrt{\mathbb{V}(\hat{\theta})}}\right) - F_{N(0,1)}\left(\frac{-10^{-2}\theta}{\sqrt{\mathbb{V}(\hat{\theta})}}\right)$$

9

$$\mathbb{P}\left(\frac{|\hat{\theta}-\theta|}{\theta} \leq 10^{-2}\right) = -1 + 2F_{N(0,1)}\left(\frac{10^{-2}\theta}{\sqrt{\mathbb{V}(\hat{\theta})}}\right) \geq 0.99$$

$$\Rightarrow F_{N(0,1)}\left(\frac{10^{-2}\theta}{\sqrt{\mathbb{V}(\hat{\theta})}}\right) \geq 0.995$$

$$\Rightarrow N \geq 10^4 \left(F^{-1}_{N(0,1)}(0.995)\right)^2 \left[\frac{\mathbb{E}\left[\frac{\pi^2(1+e^X)^2}{(1+\sin(X))^2}\right] - \theta^2}{\theta^2}\right]$$

$$\Rightarrow N \geq \underbrace{10^4 \left(F^{-1}_{N(0,1)}(0.995)\right)^2}_{\ll 10^6} \left(e^{2\pi} + 2e^{\pi}\right)$$