Université de Montpellier Année 2017-2018
M1 BIOSTATISTIQUE & MIND - HMMA101

Examen partiel - 17 novembre 2017
Durée 1h30 - Documents interdits (sauf formulaire de probabilités)

Exercice 1 (8 pts) Le nombre y de requêtes arrivant à un serveur pendant une période de temps fixée suit une loi de Poisson de paramètre $\theta > 0$. Nous observons un n-échantillon $y_1, \ldots, y_n$ de nombres de requêtes. Nous nous plaçons dans le paradigme bayésien et supposons que le paramètre θ est stochastique.

1) (2 pts) Montrer que la famille des lois gamma est conjuguée pour θ.

2) (3 pts) Donner la loi a priori de Jeffreys $\pi^J(\theta)$ pour θ. Est-elle propre ? Comment peut-on la situer par rapport à la famille conjuguée ?

3) (3 pts) Pour la loi a priori de Jeffreys, donner l'estimateur bayésien associé à la fonction de perte $L(\theta, d) = \theta^2(\theta - d)^2$. Nous rappelons que $\Gamma(z+1) = z\Gamma(z)$.

Exercice 2 (6 pts) Nous considérons un n-échantillon $x_1, \ldots, x_n$ distribué suivant une loi gausienne d'espérance θ et de variance 1. Nous supposons que θ est distribué suivant une loi gaussienne centrée réduite.

1) (3 pts) Calculer l'estimateur bayésien associé à la fonction de perte $L(\theta, d) = (\theta - d)^2$.

2) (3 pts) Donner la région de crédibilité HPD associée à un niveau de risque de 10%.

Exercice 3 (6 pts) La même information binaire $\theta \in \{0, 2\}$ est transmise 2 fois consécutives vers un récepteur à travers un canal de transmission. Ces deux informations sont perturbées par un bruit supposé gaussien centré de variance 1. Le message reçu est stocké dans le vecteur $z = (z_1, z_2)$ avec z_1 et z_2 deux variables aléatoires indépendantes distribuées suivant des lois gaussiennes de moyenne θ et variance 1. Le problème consiste à retrouver le symbole émis θ à partir du message reçu $z = (z_1, z_2)$. Nous supposons que $\mathbb{P}(\theta = 0) = \mathbb{P}(\theta = 2) = 1/2$

1) (3 pts) Donner la loi a posteriori de θ, ie $\mathbb{P}(\theta = 0|z)$ et $\mathbb{P}(\theta = 2|z)$.

2) (3 pts) Calculer l'estimateur bayésien associé à la fonction de perte

$$L(\theta, d) = \begin{cases} 0 & \text{si} \quad \theta = d \\ 2 & \text{si} \quad \theta = 0 \quad \text{et} \quad d = 2 \\ 1 & \text{si} \quad \theta = 2 \quad \text{et} \quad d = 0 \end{cases}.$$

Correction examen
- contrôle continu -
statistique bayésienne
17/11/2017
HMMA 101

1

Exercice 1

1) $f(y_1, \dots, y_n | \theta)$

$$= \prod_{i=1}^{n} \left[e^{-\theta} \frac{\theta^{y_i}}{y_i!} \right] \mathbb{1}_{\mathbb{N}}(y_i)$$

$$f(y_1, \dots, y_n | \theta) = e^{-n\theta} \theta^{\sum y_i} \prod_{i=1}^{n} \left\{ \mathbb{1}_{\mathbb{N}}(y_i) / y_i! \right\}$$

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} e^{-b\theta} \theta^{a-1} \mathbb{1}\{\theta > 0\}$$

$a > 0$ et $b > 0$

2

$$\pi(\theta | y_1, ---, y_n)$$
$$\propto e^{-n\theta} \theta^{\sum_{i=1}^{n} y_i} e^{-b\theta} \theta^{a-1} \mathbb{1}_{\{\theta > 0\}}$$
$$\Rightarrow \theta | y_1, ---, y_n \sim \text{Gamma}\left(\sum_{i=1}^{n} y_i + a, n + b\right)$$

La famille des lois Gamma est donc conjuguée pour θ.

2] $\log(f(y|\theta))$
$$= -\theta + y \log(\theta) - \log(y!)$$
$\forall y \in \mathbb{N}$

$$\frac{d^2 \log f}{d\theta^2}(y|\theta) = -\frac{y}{\theta^2}$$
$\forall y \in \mathbb{N}$

Ainsi, $I^F_y(\theta) = \dfrac{\mathbb{E}(y)}{\theta^2} = \dfrac{1}{\theta}$

③

Et donc $I_n^F(\theta) = \frac{n}{\theta}$

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

$$\int_0^{+\infty} \theta^{-1/2} = \infty$$

C'est une loi a priori impropre.

Par ailleurs, la loi a priori de Jeffreys est un cas limite de la famille des lois Gamma, c'est une loi Gamma $\left(\frac{1}{2}, \varepsilon\right)$ avec $\varepsilon \to 0$

3) L'estimateur bayésien est celui qui minimise la fonction perte moyenne relativement

à la loi a posteriori. (4)

$$\mathbb{E}^{\Pi}\left(\theta^2(\theta-d)^2 \mid y_1,\dots,y_n\right)$$
$$= \mathbb{E}^{\Pi}(\theta^4 \mid \underline{y}) - 2d\,\mathbb{E}^{\Pi}(\theta^3 \mid \underline{y})$$
$$+ d^2\,\mathbb{E}^{\Pi}(\theta^2 \mid \underline{y}) \quad (*)$$

où $\underline{y} = (y_1,\dots,y_n)$

On souhaite minimiser $(*)$ en d.

La dérivée seconde de $(*)$ en d est strictement positive :

$$2\,\mathbb{E}^{\Pi}(\theta^2 \mid \underline{y})$$

Donc $(*)$ est strictement convexe. On doit donc résoudre

$$-2\,\mathbb{E}^{\Pi}(\theta^3 \mid \underline{y}) + 2d^*\,\mathbb{E}^{\Pi}(\theta^2 \mid \underline{y}) = 0$$

5

$$\Leftrightarrow d^* = \left\{ \frac{\mathbb{E}^{\pi}(\theta^3|\underline{y})}{\mathbb{E}^{\pi}(\theta^2|\underline{y})} \right\}$$

$$\Rightarrow \widehat{\theta_m} = \left[\frac{\mathbb{E}^{\pi}(\theta^3|\underline{y})}{\mathbb{E}^{\pi}(\theta^2|\underline{y})} \right]$$

$$\mathbb{E}^{\pi}(\theta^2|\underline{y}) = \mathbb{V}^{\pi}(\theta|\underline{y}) + \left(\mathbb{E}^{\pi}(\theta|\underline{y})\right)^2$$

Comme $\theta|\underline{y} \sim \text{Gamma}\left(\sum_{i=1}^{m} y_i + \frac{1}{2}, m\right)$

alors $\begin{cases} \mathbb{E}^{\pi}(\theta|\underline{y}) = \left[\dfrac{\sum_{i=1}^{m} y_i + 1/2}{m}\right] \\ \mathbb{V}^{\pi}(\theta|\underline{y}) = \left[\dfrac{\sum_{i=1}^{m} y_i + \frac{1}{2}}{m^2}\right] \end{cases}$

$$\Rightarrow \mathbb{E}^{\pi}(\theta^2|\underline{y}) = \frac{\Sigma y_i + 1/2 + (\Sigma y_i + \frac{1}{2})^2}{m^2} = (*)$$

Par ailleurs,

$$\mathbb{E}^{\pi}(\theta^3|\underline{y}) = \int_0^{+\infty} \frac{m^{\Sigma y_i + \frac{1}{2}}}{\Gamma(\Sigma y_i + \frac{1}{2})} \theta^3 \theta^{\Sigma y_i - 1/2} e^{-m\theta} d\theta$$

6

$$\mathbb{E}^{\pi}(\theta^3|\underline{y}) = \frac{m^{\sum y_i + \frac{1}{2}}}{\Gamma\left(\sum y_i + \frac{1}{2}\right)} \int_0^{\infty} \theta^{\sum y_i - \frac{1}{2} + 3} e^{-m\theta} \, d\theta$$

$$= \frac{m^{\sum y_i + \frac{1}{2}}}{\Gamma\left(\sum y_i + \frac{1}{2}\right)} \frac{\Gamma\left(\sum y_i + \frac{1}{2} + 3\right)}{m^{\sum y_i + \frac{1}{2} + 3}}$$

$$= \frac{\left(\sum y_i + \frac{1}{2} + 2\right) \Gamma\left(\sum y_i + \frac{1}{2} + 2\right)}{\Gamma\left(\sum y_i + \frac{1}{2}\right) m^3}$$

$$= \frac{\left(\sum y_i + \frac{1}{2} + 2\right)\left(\sum y_i + \frac{1}{2} + 1\right)\left(\sum y_i + \frac{1}{2}\right)}{m^3}$$

$$\hat{\theta}_m = \frac{\left(\sum y_i + \frac{1}{2} + 2\right)\left(\sum y_i + \frac{1}{2} + 1\right)\left(\sum y_i + \frac{1}{2}\right)}{\left(\sum y_i + \frac{1}{2} + 1\right)\left(\sum y_i + \frac{1}{2}\right) m}$$

$$\left[\mathbb{E}^{\pi}(\theta^2|\underline{y}) = \frac{m^{\sum y_i + \frac{1}{2}}}{\Gamma\left(\sum y_i + \frac{1}{2}\right)} \frac{\Gamma\left(\sum y_i + \frac{1}{2} + 2\right)}{m^{\sum y_i + \frac{1}{2} + 2}} \right.$$

$$\mathbb{E}^{\pi}(\theta^2|\underline{y}) = \frac{\left(\sum y_i + \frac{1}{2} + 1\right)\left(\sum y_i + \frac{1}{2}\right)}{m}$$

$$\left. = (*) \right]$$

On en déduit que

$$\widehat{\theta_n} = \frac{\left(\sum y_i + \frac{1}{2} + 2\right)}{n}$$

Exercice 2

$$x_1, ----, x_n \mid \theta \overset{iid}{\sim} N(\theta, 1)$$

$$\theta \sim N(0, 1)$$

1) $\widehat{\theta_n} = E^{\pi}(\theta \mid x_1, ----, x_n)$

car fonction de perte quadratique

$$L(\theta, d) = (\theta - d)^2$$

$$f(x_1, ----, x_n \mid \theta) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

$$\pi(\theta \mid x_1, ----, x_n) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2 - \frac{\theta^2}{2}\right\}$$

$$\pi(\theta \mid x_1, ----, x_n) \propto \exp\left\{-\frac{n\theta^2}{2} - \frac{\theta^2}{2} + \theta\sum_{i=1}^{n} x_i\right\}$$

$$\pi(\theta \mid x_1, ----, x_n) \propto \exp\left\{-\frac{n+1}{2}\left(\theta - \frac{1}{n+1}\sum x_i\right)^2\right\}$$

$$\Rightarrow \theta \mid x_1, ----, x_n \sim N\left(\frac{\sum x_i}{n+1}, \frac{1}{n+1}\right)$$

$$\Rightarrow \widehat{\theta}_n = \frac{1}{n+1} \sum_{i=1}^{n} x_i$$

$$\underline{x} = (x_1, ----, x_n)$$

2) $\left\{ \theta \,\middle|\, \pi(\theta | \underline{x}) \geq k_\alpha \right\}$

$$= \left\{ \theta \,\middle|\, -\frac{n+1}{2}\left(\theta - \frac{\sum x_i}{n+1}\right)^2 \geq k'_\alpha \right\}$$

$$= \left\{ \theta \,\middle|\, \left|\theta - \frac{\sum x_i}{n+1}\right| \leq k''_\alpha \right\}$$

On fixe k''_α tel que

$$\mathbb{P}^\pi\left[\left|\theta - \frac{\sum x_i}{n+1}\right| \leq k''_\alpha \,\middle|\, \underline{x}\right] = 1-\alpha$$

Pour $\alpha = 0{,}1$, on obtient

$$\mathbb{P}\left[-k''_\alpha(n+1) \leq N(0,1) \leq k''_\alpha(n+1)\right] = 0.9$$

$$\Leftrightarrow k''_\alpha = \frac{F^{-1}_{N(0,1)}(0.95)}{n+1}$$

On en déduit que l'intervalle de crédibilité HPD de niveau de risque de 10% pour θ est

$$\left[\frac{\sum x_i}{n+1} - \frac{F^{-1}_{\mathcal{N}(0,1)}(0.95)}{n+1} , \frac{\sum x_i}{n+1} + \frac{F^{-1}_{\mathcal{N}(0,1)}(0.95)}{n+1} \right]$$

Exercice 3

$$\theta \in \{0, 2\} = \Theta$$

$$Z_1, Z_2 \mid \theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$$

$$\Pi(\theta = 0) = \Pi(\theta = 2) = \frac{1}{2}$$

1] $z = (z_1, z_2)$

$$\Pi(\theta = 0 \mid z) \propto \exp\left\{ -\frac{1}{2} z_1^2 - \frac{1}{2} z_2^2 \right\}$$

$$\Pi(\theta = 2 \mid z) \propto \exp\left\{ -\frac{1}{2}(z_1 - \theta)^2 - \frac{1}{2}(z_2 - \theta)^2 \right\}$$

10

$$\mathbb{P}^{\Pi}(\theta=0|Z)=\frac{e^{-\frac{z_1^2}{2}-\frac{z_2^2}{2}}}{e^{-\frac{z_1^2}{2}-\frac{z_2^2}{2}}+e^{-\frac{1}{2}(z_1-2)^2-\frac{1}{2}(z_2-2)^2}}$$

$$\Leftrightarrow \mathbb{P}^{\Pi}(\theta=0|Z)=\left[\frac{1}{1+e^{2z_1+2z_2-4}}\right]$$

et ainsi

$$\mathbb{P}^{\Pi}(\theta=2|Z)=\left[\frac{e^{2z_1+2z_2-4}}{1+e^{2z_1+2z_2-4}}\right]$$

2) L'estimateur bayésien est celui qui minimise la perte intégrée relativement à la loi a posteriori

$$\mathbb{E}^{\Pi}\left[L(\theta,d)|Z\right]$$

$$= 2\,\mathbb{1}_{\{d=2\}}\mathbb{P}^{\Pi}(\theta=0|Z) + \mathbb{1}_{\{d=0\}}\mathbb{P}^{\Pi}(\theta=2|Z) \quad (**)$$

On cherche à minimiser (**) par rapport à d

Clairement on choisit

$d = 2 \iff 2\,\mathbb{P}^{\pi}(\theta = 0|Z) \leq \mathbb{P}^{\pi}(\theta = 2|Z)$

$d = 2 \iff 3\,\mathbb{P}^{\pi}(\theta = 0|Z) \leq 1$

$d = 2 \iff \mathbb{P}^{\pi}(\theta = 0|Z) \leq \frac{1}{3}$

$d = 2 \iff \mathbb{P}^{\pi}(\theta = 2|Z) \geq \frac{2}{3}$

On choisit $d = 0$

$\iff \mathbb{P}^{\pi}(\theta = 2|Z) < 2\,\mathbb{P}^{\pi}(\theta = 0|Z)$

$d = 0 \iff \mathbb{P}^{\pi}(\theta = 0|Z) > \frac{1}{3}$

$$\widehat{\theta} = \begin{cases} 2 \text{ si } e^{2Z_1 + 2Z_2 - 4} \geq 2 \\ 0 \text{ si } e^{2Z_1 + 2Z_2 - 4} < 2 \end{cases}$$