PC4 : Exercices Correction

Exercice 1 Après obtention d'une mesure z d'un paramètre θ, la densité de probabilité conditionnelle a posteriori de θ est de la forme $f(\theta|z) = \exp(z - \theta)$, pour $z \leq \theta$.

1. Calculer les estimateurs du risque moyen en valeur absolue, du risque moyen quadratique et du maximum de vraisemblance a posteriori.
   Le risque moyen en valeur absolue s'écrit : $r_{MVA} = \underset{\theta}{\mathrm{E}}|\theta - \hat{\theta}|$ ce qui s'écrit $r_{MVA} = \underset{\theta|Z}{\mathrm{E}}|\theta - \hat{\theta}|$ soit $\int_z^{+\infty} |\theta - \hat{\theta}| f(\theta|Z) d\theta$.
   Pour minimiser ce risque, on cherche $\hat{\theta}$ tel que $\frac{\partial}{\partial \hat{\theta}} \int_z^{+\infty} |\theta - \hat{\theta}| \exp(z - \theta) d\theta = 0$, soit $\frac{\partial}{\partial \hat{\theta}} \int_z^{\hat{\theta}} |\theta - \hat{\theta}| \exp(z - \theta) d\theta + \frac{\partial}{\partial \hat{\theta}} \int_{\hat{\theta}}^{+\infty} |\theta - \hat{\theta}| \exp(z - \theta) d\theta = 0$ qui se réduit à $-\frac{\partial}{\partial \hat{\theta}} \int_z^{\hat{\theta}} (\theta - \hat{\theta}) \exp(z - \theta) d\theta + \frac{\partial}{\partial \hat{\theta}} \int_{\hat{\theta}}^{+\infty} (\theta - \hat{\theta}) \exp(z - \theta) d\theta = 0$ soit $\hat{\theta}$ doit satisfaire $\int_z^{\hat{\theta}} \exp(z - \theta) d\theta = \int_{\hat{\theta}}^{+\infty} \exp(z - \theta) d\theta$, soit $\hat{\theta} = z + ln(2)$.
   On retrouve ici une propriété de l'estimateur pour le risque moyen en valeur absolue. En effet la médiane associée à la densité exponentielle $e^{-\theta}$ avec $\theta \geq 0$ est l'argument de $\frac{1}{2}$ soit $\theta = \ln(2)$. Après décalage de z, on retrouve bien l'estimateur du risque moyen en v. a. $\hat{\theta} = z + ln(2)$.
   Le risque quadratique s'écrit : $r_{MVA} = \underset{\theta}{\mathrm{E}}(\theta - \hat{\theta})^2$. Par le même raisonnement que précédemment, $\hat{\theta}$ minimisant ce risque doit satisfaire $\frac{\partial}{\partial \hat{\theta}} \int_z^{+\infty} (\theta - \hat{\theta})^2 \exp(z - \theta) d\theta = 0$ soit $2 \int_z^{+\infty} (\theta - \hat{\theta}) \exp(z - \theta) d\theta = 0$ soit $\hat{\theta} \int_z^{+\infty} \exp(z - \theta) d\theta = \int_z^{+\infty} \theta \exp(z - \theta) d\theta$ ou encore $\hat{\theta} = \int_z^{+\infty} \theta \exp(z - \theta) d\theta$. On retrouve le résultat classique sur l'estimateur bayésien du risque quadratique, à savoir que $\hat{\theta}$ est l'espérance a posteriori, qui vaut ici $z + 1$.
   On retrouve également le résultat classique de correspondance entre l'estimateur du risque moyen quadratique et l'estimateur de la moyenne a posteriori. La moyenne (espérance) associée à la densité $e^{-\lambda\theta}$ est égale à λ. Dans le cas présent, $\lambda = 1$, et après décalage de z, on retrouve bien l'estimateur de la moyenne a posteriori $\hat{\theta} = z + 1$, égal à l'estimateur du risque moyen quadratique.
   Pour le maximum de vraisemblance a posteriori, l'estimateur doit maximiser $f(\theta|z)$ donc maximiser $(z - \theta)$ avec $\theta \geq z$ donc $\hat{\theta} = z$

Exercice 2 Loi *a priori* de Jeffreys dans le cas gaussien. On considère la suite de variables aléatoires gaussiennes indépendantes de paramètres θ et σ^2, où θ est une variable aléatoire

1. $f(z|\theta)$ est la densité d'une loi gaussienne $\mathcal{N}(\theta, \sigma^2)$
   donc $f(z|\theta)$ est proportionnelle à $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\theta)^2}{2\sigma^2}\right)$

2. La vraisemblance s'écrit $L(\theta; z_1, ..., z_n) = f(z_1, ..., z_n|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i-\theta)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \theta)^2\right)$.
3. La log-vraisemblance s'écrit alors : $\ln L(\theta; z_1, ..., z_n) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \theta)^2$. Ainsi, la matrice d'information de Fisher s'écrit (σ est une constante déterministe) : $I(\theta) = -E(\frac{\partial^2 \ln L}{\partial^2\theta}) = E(\frac{n}{\sigma^2}) = \frac{n}{\sigma^2}$.
4. La loi a priori de Jeffreys sur θ s'exprime sous la forme $\pi(\theta) \propto \sqrt{\det(I(\theta))} = \frac{\sqrt{n}}{\sigma}$.

   Dans le cas considéré, la loi a priori de θ est la loi uniforme sur un intervalle que l'on choisira grand (σ constante)

Exercice 3 On considère une variable aléatoire représentant la réussite ou l'échec d'une réalisation. Cette variable, nommée X dans la suite suit une loi de Bernouilli de paramètre θ
Le paramètre θ est supposé représenter une variable aléatoire de loi *a priori* définie par $\mathbb{P}(\theta = \theta_1) = p$ et $\mathbb{P}(\theta = \theta_2) = 1 - p$

1. Loi $\pi(\theta|Z = z)$: $\mathbb{P}(Z = 0) = 1 - \theta$, $\mathbb{P}(Z = 1) = \theta$
   $\pi(\theta = \theta_1|Z = 0) = \frac{(1-\theta_1)p}{(1-\theta_1)p+(1-\theta_2)(1-p)}$
   $\pi(\theta = \theta_1|Z = 1) = \frac{\theta_1 p}{\theta_1 p+\theta_2(1-p)}$
   $\pi(\theta = \theta_2|Z = 0) = \frac{(1-\theta_2)(1-p)}{(1-\theta_1)p+(1-\theta_2)(1-p)}$
   $\pi(\theta = \theta_2|Z = 1) = \frac{\theta_2(1-p)}{\theta_1 p+\theta_2(1-p)}$
   ;
2. On considère l'estimateur défini par $T(0) = \mu_1$, $T(1) = \mu_2$. Pour $C(\theta, T) = (\theta - T)^2$ pour $z = 1$ et $z = 0$, $\mathbb{E}^\pi[C(\theta, T)|z = 0]$ est égal à
   $(\theta_1 - \mu_1)^2\pi(\theta = \theta_1|Z = 0) + (\theta_2 - \mu_1)^2\pi(\theta = \theta_2|Z = 0)$
   $\mathbb{E}^\pi[C(\theta, T)|z = 1]$ est égal à
   $(\theta_1 - \mu_2)^2\pi(\theta = \theta_1|Z = 1) + (\theta_2 - \mu_2)^2\pi(\theta = \theta_2|Z = 1)$
   $\mathbb{E}^\pi[C(\theta, T)|z = 0] = (\theta_1 - \mu_1)^2\lambda_1 + (\theta_2 - \mu_1)^2(1 - \lambda_1)$
   avec $\lambda_1 = \frac{(1-\theta_1)p}{(1-\theta_1)p+(1-\theta_2)(1-p)}$
   $\mathbb{E}^\pi[C(\theta, T)|z = 1] = (\theta_1 - \mu_2)^2\lambda_2 + (\theta_2 - \mu_2)^2(1 - \lambda_2)$
   avec $\lambda_2 = \frac{\theta_1 p}{\theta_1 p+\theta_2(1-p)}$
3. L'estimateur de Bayes associé à ce coût minimise $\mathbb{E}^\pi[C(\theta, T)|z = 0]$ et $\mathbb{E}^\pi[C(\theta, T)|z = 1]$.
   Pour $\mathbb{E}^\pi[C(\theta, T)|z = 0] = (\theta_1-\mu_1)^2\lambda_1+(\theta_2-\mu_1)^2(1-\lambda_1)$, l'expression s'écrit $\mathbb{E}^\pi[C(\theta, T)|z = 0] = \lambda(\theta_1^2 + \mu_1^2 - 2\theta_1\mu_1) + (1 - \lambda)(\theta_2^2 + \mu_1^2 - 2\theta_2\mu_1)(1 - \lambda)$

- $= \mu_1^2 - 2(\theta_1\lambda + \theta_2(1-\lambda)\mu_1 + \theta_1^2\lambda + \theta_2^2(1-\lambda)$
- $= (\mu_1 - (\theta_1\lambda + \theta_2(1-\lambda)))^2 + f(\theta_1, \theta_2, \lambda)$

On cherche $\widehat{\mu}_1$ qui minimise cette expression.
On obtient $\widehat{\mu}_1 = \theta_1\lambda_1 + \theta_2(1-\lambda_1)$
Les calculs sont identiques pour μ_2 et donne $\widehat{\mu}_2 = \theta_1\lambda_2 + \theta_2(1-\lambda_2)$

Exercice 4 Soit Z une v.a. telle que $Z \propto \mathcal{N}(\theta, 1)$ où θ est un paramètre qui suit une loi de probabilité $\mathcal{N}(0, \sigma^2)$ où σ^2 est inconnu et fixé. Pour estimer θ, on considère la fonction de perte quadratique : $\mathcal{R}(\theta, a) = (\theta - a)^2$.

1. La loi de $\theta|Z = z$ est définie par $\pi(\theta|Z = z) = \frac{f(Z|\theta)\pi(\theta)}{\int f(Z|t)\pi(t)dt}$ qui est proportionnelle au produit $\mathcal{N}(\theta, 1)\mathcal{N}(0, \sigma^2)$
   Le terme de l'exponentielle s'écrit $-(z-\theta)^2 - \frac{\theta^2}{\sigma^2}$ et par transformation, on peut montrer que la loi est proportionnelle à $\mathcal{N}(\frac{\sigma^2 z}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2})$ (les autres termes ne dépendent pas de θ.
   L'estimateur de Bayes est a qui minimise $\int_\Theta \mathcal{R}(\theta, a)d\mathbb{P}(\theta|Z = z)$
   a doit vérifier : $\frac{\partial}{\partial a}\int_\Theta \mathcal{R}(\theta, a)d\mathbb{P}(\theta|Z = z) = 0$ soit $\frac{\partial}{\partial a}\int_\Theta \mathcal{R}(\theta, a)f(\theta|Z = z)d\theta = 0$
   ce qui s'écrit $\int_\Theta(\theta - a)f(\theta|Z = z)d\theta = 0$
   on en déduit que $a = \frac{\int_\Theta \theta f(\theta|Z=z)d\theta}{\int_\Theta f(\theta|Z=z)d\theta}$, c'est donc la moyenne de $f(\theta|Z = z)$ soit $\frac{\sigma^2 z}{1+\sigma^2}$. Il faut vérifier que cela correspond bien à un minimum (dérivation).

Exercice 5 Soit l'estimateur $\widehat{\theta} = arg\max_\Theta \pi(\theta|x) = arg\min_\Theta \left(\frac{||x-M\theta||^2}{2\sigma^2} + \frac{||\theta||^2}{2\mu^2}\right)$

1. En calculant le gradient de $\frac{||x-M\theta||^2}{2\sigma^2} + \frac{||\theta||^2}{2\mu^2}$ et en cherchant la valeur de θ qui l'annule, on obtient $\widehat{\theta} = (M^tM + \lambda^2 I)^{-1}M^t x$ avec $\lambda = \frac{\sigma^2}{\mu^2}$
   Il s'agit ensuite de vérifier si cela correspond bien à un minimum.

Exercice 6 *Loi a posteriori*

1. — Fonction de vraisemblance :
   $$L(\theta|t_1, ..., t_n) = f_\theta(t_1, ..., t_n) = \prod_{i=1}^n f_\theta(t_i) = \theta^n \exp\left(-\theta \textstyle\sum_{i=1}^n t_i\right) \quad (1)$$
   — Fonction de log-vraisemblance :
   $$\ln L(\theta|t_1, ..., t_n) = n\ln(\theta) - \theta\sum_{i=1}^n t_i \quad (2)$$
   — En calculant les dérivées successives par rapport à θ, il vient :
   $$\begin{cases} \frac{\partial \ln L(\theta|t_1,...,t_n)}{\partial\theta} & = \frac{n}{\theta} - \sum_{i=1}^n t_i \\ \frac{\partial^2 \ln L(\theta|t_1,...,t_n)}{\partial\theta^2} & = -\frac{n}{\theta^2} \end{cases} \quad (3)$$

Ainsi, $\frac{\partial \ln L(\theta|t_1,...,t_n)}{\partial \theta} = 0$ équivaut à $\theta = \frac{n}{\sum_{i=1}^n t_i}$, et comme $\frac{\partial^2 \ln L(\theta|t_1,...,t_n)}{\partial \theta^2} < 0$ pour tout θ, il s'agit d'un maximum.
Ainsi, l'estimateur du maximum de vraisemblance est donné par : $\hat{\theta}^{(MV)} = \frac{n}{\sum_{i=1}^n t_i}$.

2. — Fonction de vraisemblance :

$$L(\theta|t_1,...,t_n) = \frac{f_\theta(t_1,...,t_n)g(\theta)}{f_{T_1,...,T_n}(t_1,...,t_n)} = \frac{g(\theta)}{f_{T_1,...,T_n}(t_1,...,t_n)} \prod_{i=1}^n f_\theta(t_i) = \frac{\lambda\theta^n}{f_{T_1,...,T_n}(t_1,...,t_n)} \exp\left(-\lambda\theta - \theta \sum_{i=1}^n t_i \tag{4}$$

— Fonction de log-vraisemblance :

$$\ln L(\theta|t_1,...,t_n) = -\ln(f_{T_1,...,T_n}(t_1,...,t_n)) + \ln(\lambda) + n\ln(\theta) - \lambda\theta - \theta\sum_{i=1}^n t_i \tag{5}$$

— En calculant les dérivées successives par rapport à θ, il vient :

$$\begin{cases} \frac{\partial \ln L(\theta|t_1,...,t_n)}{\partial \theta} & = \frac{n}{\theta} - \lambda - \sum_{i=1}^n t_i \\ \frac{\partial^2 \ln L(\theta|t_1,...,t_n)}{\partial \theta^2} & = -\frac{n}{\theta^2} \end{cases} \tag{6}$$

Ainsi, $\frac{\partial \ln L(\theta|t_1,...,t_n)}{\partial \theta} = 0$ équivaut à $\theta = \frac{n}{\lambda + \sum_{i=1}^n t_i}$, et comme $\frac{\partial^2 \ln L(\theta|t_1,...,t_n)}{\partial \theta^2} < 0$ pour tout θ, il s'agit d'un maximum.
Ainsi, l'estimateur du maximum a posteriori est donné par : $\hat{\theta}^{(MV)} = \frac{n}{\lambda + \sum_{i=1}^n t_i}$.

3.
$$\frac{\hat{\theta}^{(MV)}}{\hat{\theta}^{(MAP)}} = 1 + \frac{\lambda}{\sum_{i=1}^n} \to_{n\to\infty} 1. \tag{7}$$

Ainsi, les deux estimateurs sont équivalents pour n grand.