Corrigé : Examen novembre 2018

1. Maximum de vraisemblance

   (a) La fonction de répartition est continue et de classe C_1 par morceaux sur $\mathbb{R}$ donc la densité est obtenue en la dérivant, d'où la densité pour une observation :

   $$p_\theta(x) = \mathbb{1}_{\{x>u\}}\, \theta \frac{u^\theta}{x^{\theta+1}}$$

   Pour une observation, on a $\log p_\theta(x) = \log\theta + \theta\log u - (\theta+1)\log x$, d'où

   $$\log p_\theta^{\otimes n}(\mathbf{x}) = n(\log\theta + \theta\log u) - (\theta+1)\sum_{i=1}^n \log x_i$$

   Cette quantité est finie si et seulement si $x_i > 0$ pour $i \in \{1, \dots n\}$.

   (b) On obtient le MV est annulant la dérivée partielle en θ de la log vraisemblance (qui est bien différentiable et strictement concave sur $]0, +\infty[$, en tant que fonction de θ) :

   $$\frac{\partial}{\partial\theta}(\log p_\theta^{\otimes n}(\mathbf{x})) = n\left(\frac{1}{\theta} + \log u\right) - \sum \log X_i = \frac{n}{\theta} - \sum \log(X_i/u),$$

   d'où le résultat :

   $$\widehat{\theta}_{MV}(X) = \frac{n}{\sum_{i=1}^n \log\left(\frac{X_i}{u}\right)}.$$

2. Risque quadratique

   (a) Pour $y > 0$,

   $$\mathbb{P}_\theta(\log(X_i/u) > y) = \mathbb{P}_\theta(X_i > ue^y) = 1 - F_\theta(ue^y) = (u/(ue^y))^\theta = e^{-\theta y}$$

   donc $Y_i := \log(X_i/u) \sim \mathcal{E}(\theta)$.

   (b) D'après le point 2, $V_i := \frac{1}{n}\log(X_i/u) \sim \mathcal{E}(n\theta)$: on l'obtient en calculant $\mathbb{P}_\theta(V_i > y)$ pour $y > 0$. Ainsi, par le point 2. donné dans l'énoncé, $\frac{1}{n}\sum_{i=1}^n \log(X_i/u) \sim \mathcal{G}amma(n, n\theta)$, d'où le résultat.

   (c) Par définition le biais de $\widehat{\theta}_{MV}$ est

   $$\mathbb{E}_\theta(\theta_{MV}(X)) - \theta = \frac{n\theta}{n-1} - \theta = \frac{\theta}{n-1}$$

   où la deuxième égalité vient de l'expression de l'espérance d'un inverse Gamma donnée dans l'énoncé. Le risque quadratique de $\widehat{\theta}_{MV}$ est donc, d'après l'expression de la variance d'un inverse gamma,

   $$R(\theta, \widehat{\theta}_{MV}) = \text{biais}^2(\theta) + \text{var}_\theta(\widehat{\theta}_{MV}(X)) = \frac{\theta^2}{(n-1)^2} + \frac{n^2\theta^2}{(n-1)^2(n-2)} = \frac{\theta^2}{(n-1)^2}(1 + \frac{n^2}{n-2}).$$

(d) pour une observation,

$$I_1(\theta) = \mathbb{E}_\theta[(\frac{-\partial}{\partial\theta}\log p_\theta(X_1))^2] = \mathbb{E}_\theta[(\frac{1}{\theta} - \log(X_1/u))^2]$$
$$= \mathbb{V}\mathrm{ar}_\theta(Y)$$

où $Y \sim \exp(\theta)$ d'où

$$I_1(\theta) = \frac{1}{\theta^2}$$

puis

$$I_n(\theta) = \frac{n}{\theta^2}.$$

(e) Puisque $\widehat{\theta}_c = \frac{n-1}{n}\widehat{\theta}$, on a directement

$$\mathbb{V}\mathrm{ar}_\theta(\widehat{\theta}_c(X)) = \frac{(n-1)^2}{n^2}\mathbb{V}\mathrm{ar}_\theta(\widehat{\theta}_{MV}) = \frac{\theta^2}{n-2}$$

(f) L'estimateur $\widehat{\theta}_{MV}$ est biaisé : il ne peut pas être efficace. L'estimateur $\widehat{\theta}_c$ est non-biasé mais sa variance $\frac{\theta^2}{n-2}$ est strictement supérieure à la borne de Cramér-Rao $\frac{1}{I_n(\theta)} = \frac{\theta^2}{n}$. Il n'est donc pas non plus efficace.

(g) En comparant $R(\theta, \widehat{\theta}_{MV})$ et $R(\theta, \widehat{\theta}_c) = \mathbb{V}\mathrm{ar}_\theta(\widehat{\theta}_c(X))$, on constate que le premier est supérieur au second, pour tout θ, dès que $n \geq 4/3$. Or on a supposé $n > 2$ donc on choisit $\widehat{\theta}_c$.

3. Intervalle de confiance

(a) Puisque $\sum \log(X_i/u) \sim \mathcal{G}amma(n, \theta)$, on obtient par un calcul de densité par changement de variable,

$$Z = \varphi(X, \theta) = \theta \sum \log(X_i/u) \sim \mathcal{G}amma(n, 1).$$

(b) la densité de la loi Gamma est strictement positive et continue sur $\mathbb{R}^{+*}$ donc la fonction de répartition est continue et strictement croissante sur $\mathbb{R}^{+*}$. Elle est donc injective (d'où l'unicité du quantile) et admet un inverse $F^{-1} :]0, 1[\to \mathbb{R}^{+*}$. Le quantile existe donc et est donné par $q_n(\alpha = F^{-1}(\alpha)$.

(c) soit $\alpha_1 = \alpha/2$ et $\alpha_2 = (1 - \alpha/2)$, notons $q_1 = q_n(\alpha_1)$ et $q_2 = q_n(\alpha_2)$. On a,

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta(q_2 \leq Z \leq q_1) = 1 - \alpha,$$

c'est-à-dire en inversant $\varphi_X : \theta \mapsto \varphi(X, \theta)$ (et en utilisant le fait que $\sum_i \log(X_i/u) > 0$ avec probabilité 1,

$$\mathbb{P}_\theta\left(\frac{q_1}{\sum \log(X_i/u)} \leq \theta \leq \frac{q_2}{\sum \log(X_i/u)}\right) = 1 - \alpha$$

d'où l'intervalle cherché :

$$I(X) = \left[\frac{q_1}{\sum \log(X_i/u)}, \frac{q_2}{\sum \log(X_i/u)}\right]$$

4. Test d'hypothèses
   (a) Puisque l'on donne $s > u$, $g(\theta) = \mathbb{P}_\theta(X_1 > s) = 1 - F_\theta(s) = (u/s)^\theta$.
   (b) d'après le point précédent

$$g(\theta) \leq \rho_0 \iff \theta \log(u/s) \leq \log \rho_0 \iff \theta \geq \theta_0 = \frac{\log \rho_0}{\log(u/s)}$$

   (puisque $\log(u/s) < 0$)
   A.N : $\theta_0 = \log(1/1000)/\log(1/10) = 3/1 = 3$. $g(\theta) = 10^{-\theta}$.
   (c) d'après la question 1, en notant $p_1 = p_{\theta_1}^{\otimes n}$ et $p_0 = p_{\theta_0}^{\otimes n}$, le rapport de vraisemblance est

$$\begin{aligned}\Phi_{\theta_0,\theta_1}(X) &= \frac{p_1(X)}{p_0(X)} = \exp(\log p_1(X) - \log p_0(X)) \\ &= \exp\left(C_n(\theta_0,\theta_1) + (\theta_0 - \theta_1)\sum_{i=1}^n \log(X_i/u)\right)\end{aligned}$$

   où $C_n(\theta_0,\theta_1)$ est une constante ne dépendant pas de X. Le test de Neyman-Pearson, d'après le cours, est de type $\delta(X) = \mathbb{1}_{\Phi_{\theta_0,\theta_1}(X) \geq c'}$. Or, $\theta_0 > \theta_1$, donc le rapport de vraisemblance est une fonction strictement croissante de $W = \sum \log(X_i/u)$. Il existe donc $c \in \mathbb{R}$ tel que $\Phi_{\theta_0,\theta_1}(X) \geq c' \iff W \geq c$, d'où le résultat.
   (d) D'après la question 3., sous $\tilde{H}_0$, $\theta_0 W \sim \mathcal{G}amma(n,1)$. On a donc

$$\mathbb{P}_{\theta_0}(\theta_0 W > q_n(1-\alpha)) = \alpha.$$

   i.e.

$$\mathbb{P}_{\theta_0}(W > \frac{q_n(1-\alpha)}{\theta_0}) = \alpha.$$

   d'où $c = \frac{q_n(1-\alpha)}{\theta_0}$. Ce test est U.P.P. de niveau α d'après le cours.
   (e) Pour $t > \theta_0$, on a, sous $H_0(t)$, $tW \sim \mathcal{G}amma(n,1)$, d'où

$$\mathbb{P}_t(W > c) = \mathbb{P}_t(tW > \frac{t}{\theta_0} q_n(1-\alpha)) < \mathbb{P}_t(tW > q_n(1-\alpha))$$

   car $t/\theta_0 > 1$, et car la densité de la loi Gamma ne s'annulant pas sur $]0,\infty[$, $\mathbb{P}_t(tW \in [q_n(1-\alpha), \frac{t}{\theta_0} q_n(1-\alpha)]) > 0$.
   (f) D'après le point précédent, avec $\Theta_0 = [\theta_0, \infty[$, on a

$$\sup_{\theta \in \Theta_0} R(\theta, \delta) = \sup_{\theta \geq \theta} \mathbb{P}_\theta(W > c) = \mathbb{P}_{\theta_0}(W > c) = \alpha.$$

   Le niveau du test δ pour l'hypothèse composite H_0 est donc bien α. Soit δ' un autre test de niveau α pour H_0. en particulier, $R(\theta_0, \delta') \leq \alpha$ donc δ' est de niveau α pour $\tilde{H}_0$. Mais on sait que δ est U.P.P. pour $\tilde{H}_0$ donc $\beta(\theta_1, \delta') \leq \beta(\theta_1, \delta)$. Ceci étant vrai pour tout autre test δ' de niveau α pour H_0, le test δ et bien *U.P.P.* pour H_0 contre $\tilde{H}_1$.

(g) Le raisonnement ci-dessus étant valide pour tout $\theta_1 < \theta_0$, on a bien, pour tout autre test δ', pour tout $\theta_1 \in \Theta_1$, $\beta(\theta_1, \delta') \leq \beta(\theta_1, \delta)$. Le test δ est donc U.P.P de niveau α pour tester H_0 contre H_1.

5. (a) On résout

$$\begin{cases} \alpha/\lambda = 2 \\ \alpha/\lambda^2 = 100 \end{cases}$$

c'est à dire $(\alpha = 1/25, \lambda = 1/50)$.

(b) La densité a posteriori est proportionnelle à $p_\theta^{\otimes n}(x) \times \pi(\theta)$, c'est-à-dire, (à une constante multiplicative près ne dépendant pas de θ)

$$\begin{aligned} \pi(\theta|x) &\propto \theta^n \Big(\prod_{i=1}^n x_i \Big)^{-(\theta+1)} u^{n\theta} \times \theta^{a-1} e^{-\lambda\theta} \\ &\propto \theta^n e^{-(\theta+1)\Big(\sum_{i=1}^n \log(X_i) \Big) + n\theta \log u} \times \theta^{a-1} e^{-\lambda\theta} \\ &\propto \theta^{a+n-1} e^{-(\lambda + \sum_{i=1}^n \log(x_i/u))\theta} \\ &\propto f^{\mathcal{G}}_{\alpha',\lambda'}(\theta) \end{aligned}$$

Où $f^{\mathcal{G}}_{\alpha',\lambda'}$ désigne la densité de la loi $\mathcal{G}amma(\alpha', \lambda')$, et où

$$\alpha' = \alpha + n \quad ; \lambda' = \lambda + \sum_{i=1}^n \log(\frac{x_i}{u}).$$

(c) D'après la question précédente,

$$\mathbb{E}(\boldsymbol{\theta}|x) = \frac{\alpha'}{\lambda'} = \frac{\alpha + n}{\lambda + \sum_{i=1}^n \log(\frac{x_i}{u})}$$

(d) D'après l'expression de la loi a posteriori et la question 2.b, la loi conditionnelle de $\lambda'\boldsymbol{\theta}$ sachant $X = x$ est une $\mathcal{G}amma(\alpha', 1)$. On a donc, en notant $q_{\alpha'}(q)$ le quantile d'ordre p de la loi $\mathcal{G}amma(\alpha', 1)$,

$$\mathbb{P}_\pi(\lambda'\boldsymbol{\theta} < q_{\alpha'}(p) | X = x) = p,$$

avec α', λ' comme à la réponse de la question (b). D'où

$$\mathbb{P}_\pi(\boldsymbol{\theta} < q_{\alpha'}(p)/\lambda' | X = x) = p.$$

La borne inférieure cherchée est donc

$$m(x) = \frac{q_{a+n}(p)}{\lambda + \sum_{i=1}^n \log(x_i/u)}$$