Corrigé - statistique bayésienne
2022 (exercices 1 à 5)

Exercice 1

1)

On résout

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

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

2)

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)

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

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

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

3)

D'après la question précédente,

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

4)

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

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

Exercice 2

1. le modèle géométrique est dominé par la mesure discrète $\mu = \sum_{n\in\mathbb{N}^*} \delta_n$.

— La densité par rapport à cette mesure est la loi donnée dans l'énoncé,

$$p_\theta(x) = \mathbb{1}_{\mathbb{N}^*}(x)\mathbb{P}_\theta\{X = x\} = \theta(1-\theta)^{x-1}, x \in \mathbb{R}.$$

NB : Puisqu'on a choisi μ ne chargeant que $\mathbb{N}^*$, on peut omettre l'indicatrice dans l'expression de p_θ.

— Pour un échantillon i.i.d. de taille n,

$$\log p_\theta^{\otimes n}(x_1, \ldots, x_n) = n \log\theta + \sum_{i=1}^n (x_i - 1)\log(1-\theta).$$

2. **max de vraisemblance :** On le trouve en annulant la dérivée de la log-vraisemblance. Pour $x = (x_1, \ldots, x_n)$,

$$\frac{\partial}{\partial_\theta} \log p_\theta^{\otimes n}(x) = \frac{n}{\theta} - \frac{\sum x_i - n}{1-\theta} = \frac{n - \theta\sum x_i}{\theta(1-\theta)}$$

cette quantité s'annule en $\theta = n/\sum x_i$, elle est positive à gauche et négative à droite de cette valeur, qui est donc bien un maximiseur de la vraisemblance. Ainsi $\widehat{\theta}_{MV} = n/\sum x_i$.

3. **risque quadratique** pour $g(\theta) = 1/\theta$, on considère $g_n(X) = \frac{1}{n}\sum X_i$.
   - biais : on a $\mathbb{E}(g_n(X) = \mathbb{E}(X) = 1/\theta$. L'estimateur est donc non biaisé.
   - variance : $\mathbb{V}\mathrm{ar} g_n(X) = \frac{1}{n}\mathrm{var} X = \frac{1-\theta}{n\theta^2}$.

   Le risque vaut donc $R(g_n, \theta) = \frac{1-\theta}{n\theta^2}$.
   - Efficacité : g_n étant non biaisé, il est efficace si et seulement s'il atteint la borne de Cramér-Rao $B(\theta) = g'(\theta)^2/(nI_1(\theta))$ avec I_1 l'information de Fisher pour 1 observation. On a

     $$I_1(\theta) = \mathbb{E}_\theta[\partial_\theta \log p_\theta(X_1)^2]) = \mathbb{V}\mathrm{ar}_\theta[\partial_\theta \log p_\theta(X_1)]$$

     car d'après le cours, l'espérance du score est nulle. D'où

     $$I_1(\theta) = \mathbb{V}\mathrm{ar}\Big[\frac{1-\theta X_1}{\theta(1-\theta)}\Big] = \frac{\theta^2}{\theta^2(1-\theta)^2}\mathbb{V}\mathrm{ar}(X_1)\Big] = \frac{1}{\theta^2(1-\theta)}$$

     de plus $g'(\theta)^2 = 1/\theta^4$, d'où

     $$B(\theta) = \partial(1-\theta)n\theta^2 = R(g_n, \theta)$$

     g_n est donc efficace.

Approche bayésienne : prior $\pi(\,\cdot\,) = \mathcal{U}_{[0,1]}$.

4. La densité a posteriori peut se calculer à une constante de normalisation près (ne dépendant pas de θ)

   $$\pi(\theta|x) \propto \pi(\theta)p_\theta^{\otimes n}(x) = \mathbb{1}_{]0,1[}(\theta)\theta^n(1-\theta)^{\sum x_i - n} \propto \mathrm{beta}(\theta|a_n, b_n)$$

   où $\mathrm{beta}(\theta|a, b)$ est la densité de la loi Beta e paramètres a et b, et où

   $$a_n = n+1\ ;\ b_n = \sum x_i - n + 1$$

   Ainsi la loi a posteriori est une loi beta de paramètres a_n, b_n comme ci-dessus.
5. l'espérance a posteriori est

   $$\mathbb{E}(\theta|X = x) = \mathbb{E}(U)$$

   où $U \sim \pi(\,\cdot\,|x) = \mathcal{B}eta(a_n, b_n)$ D'après le résultat de l'encadré sur l'espérance des lois Beta, $\mathbb{E}(U) = a_n/(a_n + b_n)$, d'où

   $$d'où \widehat{\theta}_{EP}(x_1, \ldots, x_n) = \mathbb{E}(\theta|X = x) = \frac{n+1}{\sum x_i + 2}$$

6. D'après la loi des grands nombres, si $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\theta)$, on a, presque sûrement, $\frac{1}{n}\sum_{i=1}^n X_i \to \mathbb{E}_\theta(X_1) = 1/\theta$. Ainsi

   $$\widehat{\theta}_{EP}(X_1, \ldots, X_n) = \frac{1 + \frac{1}{n}}{\frac{1}{n}\sum X_i + \frac{2}{n}} \to \frac{1}{1/\theta} = \theta \quad \text{presque sûrement.}$$

Exercice 3 (modèle gaussien)

Dans cet exercice, $\mu_0, \sigma^2, \tau_0^2$ sont des « hyper-paramètres » et supposés connus à l'avance.

1. L'espace des paramètres est : $\Theta = \mathbb{R}$.

   Pour déterminer la loi a posteriori, on utilise la formule de la distribution a posteriori : $\pi(\theta|X = x) = \frac{\pi(\theta)p_\theta(x)}{m^X(x)}$, où $m^X(x)$ est la marginale de X en x et joue uniquement le rôle d'un facteur de normalisation pour $\pi(\theta|x)$ (fonction de θ).

   On « voit » alors que $\boldsymbol{\theta}|x$ suit une loi gaussienne. Nous calculons donc sa densité à une constante multiplicative (une quantité ne dépendant pas de θ) près : Dans toute la suite, la notation $\pi(\theta|x) \propto g(\theta)$ signifie « $\exists\lambda > 0 :\ \forall\theta, \pi(\theta|x) = \lambda g(\theta)$ », où λ peut dépendre de x mais pas de θ. La constante λ est une constante de normalisation qui pourrait être calculée explicitement grâce au fait que $\int_\theta \pi(\theta|x)\mathrm{d}\theta = 1$, mais dont on n'a pas besoin dans notre cas pour identifier la loi a posteriori.

$$\begin{aligned}
\pi(\theta|x) &\propto \pi(\theta)p_\theta(x) \\
&\propto e^{-\frac{(\theta-\mu_0)^2}{2\tau_0^2}} e^{-\frac{(\theta-x)^2}{2\sigma^2}} \\
&\propto e^{-\frac{1}{2}\left((\frac{1}{\tau_0^2}+\frac{1}{\sigma^2})\theta^2-2(\frac{\mu_0}{\tau_0^2}+\frac{x}{\sigma^2})\theta\right)} \\
&\propto e^{-\frac{1}{2}\frac{\tau_0^2+\sigma^2}{\tau_0^2\sigma^2}\left(\theta^2-2\frac{\mu_0\sigma^2+x\tau_0^2}{\tau_0^2+\sigma^2}\theta\right)}
\end{aligned}$$

   On identifie l'espérance et la variance de la loi grâce à la relation de proportionnalité $e^{-\frac{(\theta-m)^2}{2v^2}} \propto e^{-\frac{1}{2v^2}(\theta^2-2m\theta)}$ et on obtient :

$$\boldsymbol{\theta}|x \sim \mathcal{N}\left(\frac{\mu_0\sigma^2 + x\tau_0^2}{\tau_0^2 + \sigma^2}, \frac{\tau_0^2\sigma^2}{\tau_0^2 + \sigma^2}\right).$$

   Si on désire faire tous les calculs (y compris celui de la marginale), on peut utiliser l'identité suivante :

$$\int_{-\infty}^{+\infty} e^{-ax^2+bx+c}\,dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}+c}.$$

2. Puisque l'échantillon est *i.i.d.*, Y suit une loi produit. Le modèle bayésien pour Y est donc :

$$\begin{cases}\boldsymbol{\theta} \sim \mathcal{N}(\mu_0, \tau_0^2), \\ Y|\theta \sim \mathcal{N}(\theta, \sigma^2)^{\otimes n}.\end{cases}$$

   Pour déterminer la loi a posteriori, on utilise la formule de la distribution a posteriori : $\pi(\theta|y) = \frac{\pi(\theta)p_\theta(y)}{m^Y(y)}$, où m^Y est la marginale de Y. À une constante près cela donne :

$$\begin{aligned}
\pi(\theta|y) &\propto \pi(\theta)p_\theta(y) \\
&\propto e^{-\frac{(\theta-\mu_0)^2}{2\tau_0^2}} e^{-\frac{\sum_{i=1}^n(\theta-x_i)^2}{2\sigma^2}} \\
&\propto e^{-\frac{1}{2}\left((\frac{1}{\tau_0^2}+\frac{n}{\sigma^2})\theta^2-2(\frac{\mu_0}{\tau_0^2}+\frac{\sum_i x_i}{\sigma^2})\theta\right)} \\
&\propto e^{-\frac{1}{2}\frac{n\tau_0^2+\sigma^2}{\tau_0^2\sigma^2}\left(\theta^2-2\frac{\mu_0\sigma^2+\tau_0^2\sum_i x_i}{n\tau_0^2+\sigma^2}\theta\right)}
\end{aligned}$$

On reconnaît une loi gaussienne :

$$\boldsymbol{\theta}|y \sim \mathcal{N}\left(\frac{\mu_0\sigma^2+\tau_0^2\sum_i x_i}{n\tau_0^2+\sigma^2}, \frac{\tau_0^2\sigma^2}{n\tau_0^2+\sigma^2}\right).$$

Pour étudier le comportement lorsque $n \to +\infty$, on réécrit l'espérance et la variance de la loi en introduisant la moyenne des x_i, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$:

$$\boldsymbol{\theta}|y \sim \mathcal{N}\left(\frac{\mu_0\sigma^2+n\tau_0^2\bar{x}}{n\tau_0^2+\sigma^2}, \frac{\tau_0^2\sigma^2}{n\tau_0^2+\sigma^2}\right).$$

Lorsque $n \to +\infty$, $\mathbb{E}(\boldsymbol{\theta}|y) \sim \frac{n\tau_0^2\bar{x}}{n\tau_0^2} = \bar{x} \to \mathbb{E}(X_1|\theta) = \theta$ et $\mathbb{V}\mathrm{ar}(\boldsymbol{\theta}|y) \sim \frac{\tau_0^2\sigma^2}{n\tau_0^2} = \frac{\sigma^2}{n} \to 0$.

Exercice 4 (mélange d'opinions)

1. L'espace des paramètres est : $\Theta = \{\theta_1, \theta_2\}$.
   Le modèle statistique est : $\mathcal{P} = \{\mathcal{B}(\theta), \theta \in \Theta\}$, où $\mathcal{B}(\theta)$ est une loi de Bernoulli de paramètre θ.
2. Ici on définit un modèle bayésien pour X :

$$\begin{cases} \boldsymbol{\theta} \sim \pi, \\ X|\theta \sim \mathcal{B}(\theta). \end{cases}$$

   De manière générale, on a :

$$\pi(\theta|x) = \frac{\pi(\theta)p_\theta(x)}{\pi(\theta_1)p_{\theta_1}(x)+\pi(\theta_2)p_{\theta_2}(x)} = \frac{\pi(\theta)\theta^x(1-\theta)^{1-x}}{\pi(\theta_1)\theta_1^x(1-\theta_1)^{1-x}+\pi(\theta_2)\theta_2^x(1-\theta_2)^{1-x}}.$$

   Pour $\pi = \pi_1$ ($\pi_1(\theta_1) = \pi_2(\theta_2) = 0.5$), on obtient donc les valeurs $\pi_1(\theta|x)$:

θ/x	0	1
θ_1	$\frac{1-\theta_1}{1-\theta_1+1-\theta_2} = \frac{2}{3}$	$\frac{\theta_1}{\theta_1+\theta_2} = \frac{1}{4}$
θ_2	$\frac{1-\theta_2}{1-\theta_1+1-\theta_2} = \frac{1}{3}$	$\frac{\theta_2}{\theta_1+\theta_2} = \frac{3}{4}$

3. On a maintenant comme distribution a priori : $\pi_2(\theta_1) = 1/4,\ \pi_2(\theta_2) = 3/4$.
   On obtient donc les valeurs $\pi_2(\theta|x)$:

θ/x	0	1
θ_1	$\frac{1\times(1-\theta_1)}{1\times(1-\theta_1)+3\times(1-\theta_2)} = \frac{2}{5}$	$\frac{1\times\theta_1}{1\times\theta_1+3\times\theta_2} = \frac{1}{10}$
θ_2	$\frac{3}{5}$	$\frac{9}{10}$

4. La densité de la loi prédictive a posteriori est donnée par :
$$p(y) = \pi_2(\theta_1|x)p_{\theta_1}(y) + \pi_2(\theta_2|x)p_{\theta_2}(y).$$
   Pour $x = 1$ on obtient : $p(0) = \frac{1}{10} \times (1 - 0.2) + \frac{9}{10} \times (1 - 0.6) = \frac{11}{25}$ et $p(1) = \frac{14}{25}$.
5. Pour $X = (X_1, \ldots, X_n)$ échantillon *i.i.d.* :
$$\begin{aligned}
\pi(\theta|x) &= \frac{\pi(\theta)p_\theta(x)}{\pi(\theta_1)p_{\theta_1}(x) + \pi(\theta_2)p_{\theta_2}(x)} \\
&= \frac{\pi(\theta)\theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}}{\pi(\theta_1)\theta_1^{\sum_i x_i}(1-\theta_1)^{n-\sum_i x_i} + \pi(\theta_2)\theta_2^{\sum_i x_i}(1-\theta_2)^{n-\sum_i x_i}} \\
&= \frac{\pi(\theta)\theta^s(1-\theta)^{n-s}}{\pi(\theta_1)\theta_1^s(1-\theta_1)^{n-s} + \pi(\theta_2)\theta_2^s(1-\theta_2)^{n-s}}.
\end{aligned}$$
   Ainsi $\pi(\theta|x)$ ne dépend de x qu'à travers s.
6. On pose $\alpha = \frac{\theta_2(1-\theta_2)}{\theta_1(1-\theta_1)} = \frac{3}{2}$. Alors :
$$\pi(\theta_1|x) = \frac{\pi(\theta_1)}{\pi(\theta_1) + \alpha^{n/2}\pi(\theta_2)}$$
   et
$$\pi(\theta_2|x) = \frac{\pi(\theta_2)}{\alpha^{-n/2}\pi(\theta_1) + \pi(\theta_2)}.$$
   D'où, quand $n \to +\infty$, $\pi(\theta_1|x) \to 0$ et $\pi(\theta_2|x) \to 1$, indépendamment de la distribution a priori de $\boldsymbol{\theta}$. La loi conditionnelle a posteriori tend à privilégier le θ le plus proche de 0.5 (en effet $\theta_2 = 0.6 \approx 0.5$ et $\theta_1 = 0.2 \neq 0.5$).

Exercice 5 (calcul a posteriori)

1. On détermine la distribution a posteriori par la formule $\pi(\theta|x) = \frac{\pi(\theta)p_\theta(x)}{m^X(x)}$, où m^X est la marginale de X. Ici $\pi(\theta) = \mathbb{1}_{]0,1[}(\theta)$. On obtient donc :
$$\begin{aligned}
m^X(x) &= \int_{\mathbb{R}^+} \pi(\theta)p_\theta(x)\,d\theta \\
&= \int_0^1 p_\theta(x)\,d\theta \\
&= \int_x^1 \frac{2x}{\theta^2}\,d\theta \\
&= \left[\frac{-2x}{\theta}\right]_x^1 \\
&= 2(1-x) && \text{si } x > 0 \text{ et } 0 \text{ sinon.}
\end{aligned}$$

D'où : $\pi(\theta|x) = \frac{x}{1-x}\frac{1}{\theta^2}$ pour $\theta \in [x,1]$ et $x > 0$, et 0 sinon.

2. De la même manière pour $\pi(\theta) = 3\theta^2 \mathbb{1}_{]0,1[}(\theta)$: $m^X(x) = \int_x^1 3\theta^2 \frac{2x}{\theta^2}\, d\theta = 6x(1-x)$ (pour $x > 0$ et 0 sinon) et $\pi(\theta|x) = \frac{1}{1-x}$ pour $\theta \in [x,1]$ et $x > 0$, et 0 sinon.
3. L'espérance a posteriori est donnée par :
$$\mathbb{E}(\boldsymbol{\theta}|x) = \int_\Theta \theta \pi(d\theta|x) = \int_{\mathbb{R}^+} \theta \pi(\theta|x)\, d\theta.$$
Elle est systématiquement nulle pour $x \leq 0$. Dans la suite on suppose donc que $x > 0$.
Dans le premier cas ($\pi(\theta|x) = \frac{x}{1-x}\frac{1}{\theta^2}\mathbb{1}_{[x,1]}(\theta)$) : $\mathbb{E}(\boldsymbol{\theta}|x) = \int_x^1 \theta \frac{x}{1-x}\frac{1}{\theta^2}\, d\theta = -\frac{x\log(x)}{1-x}$.
Dans le second cas ($\pi(\theta|x) = \frac{1}{1-x}\mathbb{1}_{[x,1]}(\theta)$) : $\mathbb{E}(\boldsymbol{\theta}|x) = \int_x^1 \frac{\theta}{1-x}\, d\theta = \frac{1}{2}(1+x)$.
4. Dans cette question, on dispose d'un échantillon *i.i.d.* $X = (X_1, \ldots, X_n)$. X suit une loi produit. Ainsi, pour $\pi(\theta) = \mathbb{1}_{]0,1[}(\theta)$, la marginale de X est donnée par :
$$\begin{aligned} m^X(x) &= \int_{\mathbb{R}^+} \pi(\theta) p_\theta(x)\, d\theta \\ &= \int_0^1 p_\theta(x)\, d\theta \\ &= \int_{x_0}^1 \frac{2^n \prod_i x_i}{\theta^{2n}}\, d\theta \qquad\qquad (x_0 = \max_i x_i) \\ &= \left[-\frac{2^n \prod_i x_i}{(2n-1)\theta^{2n-1}} \right]_{x_0}^1 \\ &= \frac{2^n \prod_i x_i}{2n-1} \left(\frac{1}{x_0^{2n-1}} - 1 \right). \end{aligned}$$

En conséquence, la densité a posteriori est obtenue par la formule suivante :
$$\pi(\theta|x) = \frac{1}{m^X(x)} \frac{2^n \prod_i x_i}{\theta^{2n}} = \frac{2n-1}{\theta^{2n}} \frac{x_0^{2n-1}}{1 - x_0^{2n-1}}$$
pour $\theta \in [x_0, 1]$ et 0 sinon.