Corrigé : Feuille de travaux dirigés 4

Solution Exercice 1

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 $\mathbb{V}\mathrm{ar}(\boldsymbol{\theta}|y) \sim \frac{\tau_0^2\sigma^2}{n\tau_0^2} = \frac{\sigma^2}{n} \to 0$.

Solution Exercice 2

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 $\theta_1 = 0.2 \neq 0.5$).

Solution Exercice 3

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 :
$$\begin{aligned}m^X(x) &= \int_{\mathbb{R}^+} \pi(\theta)p_\theta(x)\,d\theta\\ &= \int_0^1 p_\theta(x)\,d\theta\\ &= \int_x^1 \frac{2x}{\theta^2}\,d\theta\\ &= \left[\frac{-2x}{\theta}\right]_x^1\\ &= 2(1-x) \qquad\qquad \text{si } x > 0 \text{ et } 0 \text{ sinon.}\end{aligned}$$

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.* $X = (X_1,\dots,X_n)$. X suit une loi produit. Ainsi, pour $\pi(\theta) = \mathbb{1}_{]0,1[}(\theta)$, la marginale de X est donnée par :
$$\begin{aligned}
m^X(x) &= \int_{\mathbb{R}^+} \pi(\theta)p_\theta(x)\,d\theta \\
&= \int_0^1 p_\theta(x)\,d\theta \\
&= \int_{x_0}^1 \frac{2^n\prod_i x_i}{\theta^{2n}}\,d\theta && (x_0 = \max_i x_i) \\
&= \left[-\frac{2^n\prod_i x_i}{(2n-1)\theta^{2n-1}}\right]_{x_0}^1 \\
&= \frac{2^n\prod_i x_i}{2n-1}\left(\frac{1}{x_0^{2n-1}} - 1\right).
\end{aligned}$$
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.

Solution Exercice 4

On remarque tout d'abord que $\sum_{j=1}^p X_j = n$.

1. Pour obtenir la loi multinomiale, le raisonnement est identique à celui d'une loi binomiale : dans l'arbre des possibles (tirages $(Y_1,\dots,Y_n)$) on choisit x_1 branches telles que $Y_j = 1$, x_2 branches telles que $Y_j = 2$, … puis on pondère par les probabilités $\theta_1^{x_1}$, $\theta_2^{x_2}$, … Les x_p dernières branches sont imposées. On obtient ainsi :
$$\begin{aligned}
\mathbb{P}(X_1 = x_1,\dots,X_p = x_p) &= \binom{n}{x_1}\dots\binom{n-\sum_{j=1}^{p-2} x_j}{x_{p-1}}\theta_1^{x_1}\dots\theta_{p-1}^{x_{p-1}}\theta_p^{x_p} \\
&= \frac{n!}{x_1!(n-x_1)!}\dots\frac{(n-\sum_{j=1}^{p-2} x_j)!}{x_{p-1}!(n-\sum_{j=1}^{p-1} x_j)!}\theta_1^{x_1}\dots\theta_p^{x_p} \\
&= \frac{n!}{x_1!\dots x_{p-1}!x_p!}\theta_1^{x_1}\dots\theta_{p-1}^{x_p} && (n-\sum_{j=1}^{p-1} x_j = x_p).
\end{aligned}$$

2. Le modèle bayésien est :

$$\begin{cases} \boldsymbol{\theta} \sim \mathrm{Diri}_{(a_1,\dots,a_p)}, \\ X|\theta \sim \mathcal{M}(\theta, n). \end{cases}$$

On a donc comme densité a posteriori :

$$\begin{aligned} \pi(\theta|x) &= \frac{\pi(\theta)p_\theta(x)}{m^X(x)} \\ &\propto \prod_{j=1}^{p} \theta_j^{a_j-1} \prod_{j=1}^{p} \theta_j^{x_j} \\ &\propto \prod_{j=1}^{p} \theta_j^{a_j+x_j-1}. \end{aligned}$$

On reconnaît que :

$$\boldsymbol{\theta}|x \sim \mathrm{Diri}_{(a_1+x_1,\dots,a_p+x_p)}.$$

3. En appliquant la formule donnée pour l'espérance, on obtient :

$$\mathbb{E}(\boldsymbol{\theta}|x) = \frac{1}{\sum_{j=1}^{p}(a_j+x_j)} \begin{pmatrix} a_1+x_1 \\ \vdots \\ a_p+x_p \end{pmatrix} = \frac{n}{n+\sum_{j=1}^{p} a_j} \begin{pmatrix} \frac{a_1}{n}+\frac{x_1}{n} \\ \vdots \\ \frac{a_p}{n}+\frac{x_p}{n} \end{pmatrix}.$$

Observons que : $\frac{x_j}{n} = \frac{1}{n}\sum_{i=1}^{n} z_i$, où $z_i = \mathbb{1}_{\{j\}}(y_i)$. On peut définir des variables aléatoires $Z_i = \mathbb{1}_{\{j\}}(Y_i)$. Les Z_i sont *i.i.d.* et suivent une loi de Bernoulli de paramètre $\mathbb{P}(Y_i = j) = \theta_j$. Ainsi $\frac{x_j}{n}$ est un estimateur sans biais de l'espérance de Z_1 et $\mathbb{E}(Z_1) = \theta_{0j}$. Finalement : $\mathbb{E}(\boldsymbol{\theta}|x) \to \theta_0$ lorsque $n \to +\infty$ ($\frac{a_j}{n} \to 0$).
De même, la variance conditionnelle de $\boldsymbol{\theta}$ est donnée par :

$$\mathbb{V}\mathrm{ar}(\boldsymbol{\theta}_j|x) = \frac{(a_j+x_j)(n+\sum_{\ell=1}^{p} x_\ell - a_j - x_j)}{(n+\sum_{\ell=1}^{p} x_\ell)^2(n+\sum_{\ell=1}^{p} x_\ell + 1)} \to 0.$$