TD 7 : Approche bayésienne

Exercice 1 (Lois conditionnelles et vecteur gaussien)
Soient X et Y deux variables aléatoires réelles de loi jointe de densité sur $\mathbb{R}^2$:

$$(x,y) \mapsto \frac{1}{2\pi} \exp\left(-x^2 + xy - \frac{y^2}{2}\right).$$

1. Déterminer :
   (a) les lois marginales de X et Y ;
   (b) la loi conditionnelle de Y sachant $X = x$ et celle de X sachant $Y = y$.
2. Vérifier que la loi de (X, Y) est celle d'un vecteur gaussien sur $\mathbb{R}^2$, dont on précisera la moyenne et la matrice de variance-covariance.

Exercice 2 (Lois conjuguées)
Montrer que les familles de lois a priori suivantes sont conjuguées, pour $n \geq 1$. De plus, on donnera dans chaque cas l'expression de la moyenne a posteriori, si celle-ci existe.

1. La famille des lois Gamma(a, b), $a > 0$, $b > 0$, pour $\mathcal{P} = \{\text{Gamma}(p, \lambda)^{\otimes n},\ \lambda > 0\}$, avec $p > 0$ fixé.
2. La famille des lois Beta(a, b), $a > 0$, $b > 0$, pour $\mathcal{P} = \{\text{Geom}(\theta)^{\otimes n},\ \theta \in]0, 1[\}$.
3. La famille des lois $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$, pour le modèle $\mathcal{P} = \{\mathcal{N}(\theta, v)^{\otimes n},\ \theta \in \mathbb{R}\}$, avec $v > 0$ fixé.

Exercice 3 (A priori impropre) Un a priori Π est dit impropre si Π est une mesure positive sur Θ, de masse infinie, soit

$$\Pi(\Theta) = +\infty.$$

Un a priori impropre Π n'est donc pas une mesure de probabilité sur Θ. On suppose néanmoins que Π est une mesure σ-finie, absolument continue par rapport à ν, et l'on note π sa dérivée de Radon-Nikodym par rapport ν. Dans le cadre d'une expérience $(\mathbf{X}, \mathcal{P})$ avec $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$, et $\mathrm{d}P_\theta = p_\theta \,\mathrm{d}\mu$, si l'on met un a priori impropre Π sur Θ avec $\mathrm{d}\Pi = \pi \,\mathrm{d}\nu$, et si $\int_\Theta p_\theta(\mathbf{X}) \,\mathrm{d}\Pi(\theta)$ est finie p.s., alors on peut former la loi a posteriori correspondante $\Pi[\cdot \,|\, \mathbf{X}]$. Celle-ci est de densité par rapport à ν égale à

$$\theta \mapsto \pi(\theta \,|\, \mathbf{X}) = \frac{p_\theta(\mathbf{X})\pi(\theta)}{\displaystyle\int_\Theta p_\theta(\mathbf{X})\pi(\theta)\,\mathrm{d}\nu(\theta)}.$$

1. Dans le modèle gaussien $(\mathcal{N}(\theta, 1)^{\otimes n})_{\theta \in \mathbb{R}}$, on considère l'a priori Π donné par la mesure de Lebesgue sur $\mathbb{R}$. Montrer que la loi a posteriori est bien définie et la déterminer.
2. Un voyageur arrive dans une ville où les tramways sont numérotés de 1 jusqu'au nombre total de tramways. Il voit passer un tramway numéroté 100. On cherche à savoir dans quelle mesure cela peut l'aider à avoir une idée du nombre total de tramways dans cette ville.
   (a) Modéliser ce problème dans un cadre bayésien.
   (b) Si l'on choisit l'a priori impropre donné par la mesure de comptage sur $\mathbb{N}^*$, la loi a posteriori est-elle bien définie ?
   (c) Qu'en est-il si le voyageur voit passer $k \geq 2$ tramways indépendants ?

Exercice 4 (Bayésien empirique) Pour déterminer une loi a priori pour un problème donné, une approche très utilisée en pratique est la méthode dite « bayésienne empirique ». On dispose d'observations $\mathbf{X} = (X_1, \ldots, X_n)$ issues d'un modèle $(P_\theta^{\otimes n})_{\theta\in\Theta}$, avec $\mathrm{d}P_\theta = p_\theta \,\mathrm{d}\mu$. Pour choisir une loi a priori sur Θ, on commence par restreindre le choix à une famille paramétrée $(\Pi_\alpha)_{\alpha\in\mathcal{A}}$, avec $\mathrm{d}\Pi_\alpha = \pi_\alpha \,\mathrm{d}\nu$, pour une certaine mesure ν sur Θ. On cherche alors à estimer α à partir de $\mathbf{X}$. Pour cela, on forme, pour tout $\alpha \in \mathcal{A}$, la vraisemblance marginale de $\mathbf{X}$ sous l'a priori Π_α :

$$f_\alpha(\mathbf{X}) = \int_\Theta p_\theta(\mathbf{X}) \,\mathrm{d}\Pi_\alpha(\theta),$$

et l'on estime α par :

$$\widehat{\alpha}(\mathbf{X}) \in \arg\max_{\alpha\in\mathcal{A}} f_\alpha(\mathbf{X}).$$

C'est le principe de la méthode du maximum de vraisemblance marginale. Si la loi a posteriori pour la loi a priori Π_α est donnée par $\Pi_\alpha[\cdot \,|\, \mathbf{X}]$, alors la « loi a posteriori » donnée par la méthode bayésienne empirique est obtenue par *plug-in* : on remplace α par $\widehat{\alpha}(\mathbf{X})$ et l'on considère la loi $\Pi_{\widehat{\alpha}(\mathbf{X})}[\cdot \,|\, \mathbf{X}]$. Cette loi est généralement bien définie mais il y a abus de langage à l'appeler loi a posteriori : on a fait comme si cela ne changeait pas le modèle de prendre une loi a priori qui dépend des données. On l'appellera plutôt pseudo-loi a posteriori.

1. Modèle exponentiel : on se place dans le modèle $(\mathcal{E}(\theta)^{\otimes n})_{\theta>0}$ et l'on cherche à choisir une loi a priori parmi l'ensemble des lois $\Pi_\lambda = \mathcal{E}(\lambda)$, $\lambda > 0$.
   (a) Soit $\lambda > 0$. Montrer que, lorsque $\boldsymbol{\theta} \sim \Pi_\lambda$, la loi marginale de X_1 a pour densité $f_\lambda(x) = \frac{\lambda}{(\lambda+x)^2}\mathbf{1}_{x\geq 0}$.
   (b) Soit $\lambda > 0$ et $\boldsymbol{\theta} \sim \Pi_\lambda$. Calculer la densité marginale de $\mathbf{X} = (X_1, \ldots, X_n)$. Rappel : si $T \sim \mathcal{E}(\lambda)$, alors pour tout $k \geq 1$, $\mathbb{E}[T^k] = k!/\lambda^k$.
   (c) En déduire que l'estimateur du maximum de vraisemblance marginal est $\widehat{\lambda}(\mathbf{X}) = \overline{X}_n$. Quelle est la pseudo-loi a posteriori suggérée par la méthode bayésienne empirique ?
2. Modèle Poisson : on se place dans le modèle $(\mathcal{P}(\theta)^{\otimes n})_{\theta>0}$ et l'on cherche à choisir une loi a priori parmi l'ensemble des lois $\Pi_\lambda = \mathcal{E}(\lambda)$ pour $\lambda > 0$.
   (a) Montrer que la loi marginale de X_1 est une loi géométrique à valeurs dans $\mathbb{N}$ de paramètre $\lambda/(\lambda+1)$, autrement dit, que $\mathbb{P}(X_1 = k) = \left(\frac{1}{\lambda+1}\right)^k \frac{\lambda}{\lambda+1}$ pour tout entier positif k.
   (b) Soit $\lambda > 0$ et $\boldsymbol{\theta} \sim \Pi_\lambda$. Calculer la densité marginale de $\mathbf{X} = (X_1, \ldots, X_n)$.
   (c) Quelle est la pseudo-loi a posteriori suggérée par la méthode bayésienne empirique ?
3. Modèle gaussien : on se place dans le modèle $(\mathcal{N}(\theta,1)^{\otimes n})_{\theta\in\mathbb{R}}$ et l'on cherche à choisir une loi a priori parmi l'ensemble des lois $\Pi_\mu = \mathcal{N}(\mu, 1)$, $\mu \in \mathbb{R}$.
   (a) Déterminer $\widehat{\mu}$, l'estimateur du maximum de vraisemblance marginal.
   (b) Quelle est la pseudo-loi a posteriori suggérée par la méthode bayésienne empirique ?

Exercice 5 (Famille gaussienne conjuguée dans $\mathbb{R}^d$)
Soient μ un vecteur de $\mathbb{R}^d$, Λ et Σ deux matrices de covariance $d \times d$ inversibles fixées.

1. Montrer que si $\boldsymbol{\theta} \sim \Pi = \mathcal{N}(\mu, \Lambda)$ et

$$\mathbf{X} = (X_1, \ldots, X_n) \,\big|\, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)^{\otimes n},$$

   alors la loi a posteriori $\Pi[\cdot \,|\, \mathbf{X}]$ est $\mathcal{N}(m_\mathbf{X}, \Sigma_\mathbf{X})$, avec

$$\begin{aligned} m_\mathbf{X} &= \Sigma_\mathbf{X}(\Lambda^{-1}\mu + n\Sigma^{-1}\overline{X}_n) \\ \Sigma_\mathbf{X} &= (n\Sigma^{-1} + \Lambda^{-1})^{-1}. \end{aligned}$$

2. Qu'en concluez-vous ?

Exercice 6 (A priori de Jeffreys) Soit Θ est un intervalle ouvert de $\mathbb{R}$ et $\mathcal{P} = \{P_\theta,\, \theta \in \Theta\}$ un modèle régulier, dominé par la mesure de Lebesgue sur Θ, d'information de Fisher $\theta \mapsto \mathbf{I}(\theta)$. On appelle *a priori de Jeffreys* la mesure sur Θ de densité π par rapport à $\nu = \mathrm{Leb}_{|\Theta}$ proportionnelle à $\sqrt{\mathbf{I}(\cdot)}$. Plus précisément, pour tout $\theta \in \Theta$, $\pi(\theta) = \frac{1}{\Lambda}\sqrt{\mathbf{I}(\theta)}$ avec

$$\Lambda = \begin{cases} \int_\Theta \sqrt{\mathbf{I}(\theta)}\,\mathrm{d}\theta & \text{si } \int_\Theta \sqrt{\mathbf{I}(\theta)}\,\mathrm{d}\theta < +\infty, \\ 1 & \text{si } \int_\Theta \sqrt{\mathbf{I}(\theta)}\,\mathrm{d}\theta = +\infty. \end{cases}$$

Cet a priori possède la proprité d'être invariant par re-paramétrisation lisse du modèle, comme le montre la question suivante.

1. Soit Π l'a priori de Jeffreys dans le modèle $\mathcal{P} = \{P_\theta,\, \theta \in \Theta\}$, et soit φ un $\mathcal{C}^1$-difféomorphisme de Θ dans $\varphi(\Theta)$. Montrer que $\Pi \circ \varphi^{-1}$, la mesure image de Π par φ, est l'a priori de Jeffreys dans le modèle

$$\mathcal{Q} = \{P_{\varphi^{-1}(\eta)},\, \eta \in \varphi(\Theta)\}\,.$$

2. Déteminer l'a priori de Jeffreys dans les modèles suivants :
   (a) $\mathcal{P} = \{\mathcal{B}(\theta),\ \theta \in (0,1)\}$ (modèle de Bernoulli) ;
   (b) $\mathcal{P} = \{\mathcal{N}(\theta, 1),\ \theta \in \mathbb{R}\}$ (modèle gaussien) ;
   (c) $\mathcal{P} = \{\mathcal{P}(\theta),\ \theta > 0\}$ (modèle Poisson).

Exercice 7 ($\star$) (Famille de Dirichlet et modèle multinomial)
On rappelle que pour $a = (a_1, \dots, a_K) \in {\mathbb{R}_+^*}^K$, la loi de Dirichlet $\mathrm{Dir}(a)$ est la loi sur le simplexe $\mathcal{S}_K = \{x = (x_1, \dots, x_K) \in [0,1]^K,\ \sum_{k=1}^K x_k = 1\}$ de densité (par rapport à la mesure de Lebesgue sur $\mathcal{S}_K$)

$$x \mapsto \frac{1}{B(a)} \prod_{k=1}^K x_k^{a_k - 1}, \qquad B(a) = \frac{\prod_{k=1}^K \Gamma(a_k)}{\Gamma(\sum_{k=1}^K a_k)}.$$

1. (a) Montrer que si $Z = (Z_1, \dots, Z_K) \sim \mathrm{Dir}(a)$, alors Z_1 suit une loi Beta que l'on déterminera (on pourra considérer le changement de variables $x_i = (1 - x_1) y_i$, pour $2 \le i \le K-1$).
   (b) Montrer que la famille des lois de Dirichlet $\mathrm{Dir}(a_1, \dots, a_K)$ est conjuguée pour le modèle

$$\mathcal{P} = \left\{ \left(\sum_{k=1}^K p_k \delta_k \right)^{\otimes n},\ p = (p_1, \dots, p_K) \in \mathcal{S}_K \right\}.$$

2. Soient $X_1, \dots, X_n$ des variables i.i.d. à valeurs dans $\{1, 2, \dots, K\}$, de loi $p = (p_1, \dots, p_K)$ (i.e. $\mathbb{P}(X_1 = j) = p_j$). Notons, pour $j = 1, \dots, K$,

$$N_j = \sum_{i=1}^n \mathbb{1}_{X_i = j}.$$

   (a) Quelle est la loi de N_j pour $j \in \{1, \dots, K\}$?
   (b) Montrer que la loi jointe des N_j est donnée par

$$\mathbb{P}\left(N_j = n_j,\ 1 \le j \le K\right) = \binom{n}{n_1, \dots, n_K} \prod_{j=1}^K p_j^{n_j} \mathbb{1}_{\{\sum_{j=1}^K n_j = n\}},$$

   où $\binom{n}{n_1, \dots, n_K} = n!/(n_1! \dots n_K!)$ désigne le nombre de façons de regrouper n éléments en K groupes de tailles respectives $n_1, \dots, n_K$.
   On dit que le vecteur $(N_1, \dots, N_K)$ suit la loi multinomiale de paramètres $(n, p_1, \dots, p_K)$.
   (c) Montrer que la famille des lois de Dirichlet $\mathrm{Dir}(a)$, $a \in (\mathbb{R}_+^*)^K$, est conjuguée pour le modèle des lois multinomiales $\mathcal{P} = \{\mathrm{Mult}(n, p_1, \dots, p_K),\ p = (p_1, \dots, p_K) \in \mathcal{S}_K\}$.

Exercice 8 (Estimateur de Laplace-Bayes)

Une urne contient N boules, dont r boules rouges et $N-r$ boules noires, avec r inconnu. On tire n boules sans remise dans l'urne, $n \leq N$, et l'on note X le nombre de boules rouges tirées. On considère le cadre bayésien suivant :

$$\begin{aligned} R &\sim \mathrm{Unif}(\{0,1,\ldots,N\}) \\ X \,\big|\, R &\sim P_R. \end{aligned}$$

1. Sachant R, déterminer la loi P_R qui correspond au modèle décrit.
2. Déterminer la loi a posteriori de R sachant X. <u>Indication</u> : avec les conventions usuelles sur les coefficients binomiaux, on a la relation
$$\sum_{k=0}^{N} \binom{k}{m}\binom{N-k}{n-m} = \binom{N+1}{n+1}.$$
3. Quelle est la probabilité, sachant X, que le $(n+1)^{\text{ième}}$ tirage donne une boule rouge ?

Exercice 9 (Identifiabilité)
On s'intéresse à une colonie d'insectes pondeurs d'œufs. On suppose que pour le site observé, le nombre X d'œufs suit une loi de Poisson de paramètre $\lambda > 0$ et que chaque œuf pondu donne, indépendamment des autres et de X, un insecte avec une probabilité $p \in]0,1[$. On note Y le nombre de naissances, de sorte qu'on peut écrire

$$Y = \sum_{i=1}^{X} \varepsilon_i$$

où les ε_i sont i.i.d. de loi de Bernoulli de paramètre p.

1. On suppose qu'on observe à la fois X et Y.
   (a) Déterminer la loi conditionnelle de Y sachant $X = x$.
   (b) Écrire le modèle statistique associé. Est-il identifiable ?
   (c) Supposons que l'on observe des couples indépendants $(X_1, Y_1), \ldots, (X_n, Y_n)$, de même loi que (X, Y). Proposer un estimateur de λ et de p.
2. On n'observe plus que Y.
   (a) Quelle est la loi de Y ?
   (b) Le modèle associé est-il identifiable si le paramètre d'intérêt est toujours le couple (λ, p) ?

Exercice 10 (Lois conditionnelles)
Soient X et Y deux variables aléatoires indépendantes et $S = X+Y$. Dans chacun des trois cas suivants, déterminer la loi de S, la loi de X sachant S, ainsi que l'espérance conditionnelle $\mathbb{E}[X \,|\, S]$.

1. $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$, avec $\lambda, \mu > 0$.
2. $X \sim \Gamma(r, \lambda)$ et $Y \sim \Gamma(t, \lambda)$, avec $r, t, \lambda > 0$.
3. $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(m, p)$, avec $n, m \in \mathbb{N}^*$ et $p \in]0,1[$.

<u>Indications</u> :
- Si X suit une loi de Poisson de paramètre λ (respectivement une loi Gamma de paramètres (r, λ)), alors sa fonction caractéristique vaut $\exp(\lambda(e^{it}-1))$ (respectivement $(1-it/\lambda)^{-r}$).
- Si $p \in [0,1]$ et $0 \leq n \leq N$ sont deux entiers naturels avec pN entier, on dit que X suit une loi hypergéométrique de paramètres (N, p, n) si $\mathbb{P}(X = k) = \binom{pN}{k}\binom{(1-p)N}{n-k}/\binom{N}{n}$ si $\max(0, n-(1-p)N) \leq k \leq \min(pN, n)$, auquel cas son espérance vaut np.