Examen 2018 : Modélisation et statistique bayésienne computationnelle

13 avril 2018

L'examen dure 3h et est noté sur 30. Tous les supports de cours sont autorisés. Il est attendu un code R commenté a minima pour les parties computationnelles, et un support papier peut être utilisé pour la partie formelle. Lisez bien tout le document, certaines questions peuvent être traitées indépendamment du reste de l'exercice.

1 Fonction de coût (9 pts)

Soit $\theta \in \Theta = I\!R$ le paramètre d'un modèle, sur lequel on dispose d'une loi *a priori* $\pi(\theta)$ et de données $x_1, \ldots, x_n$. On suppose que la loi a posteriori de densité $\pi(\theta|x_1, \ldots, x_n)$ est propre et telle que $\mathrm{E}_\pi[\exp(k\theta)|x_1, \ldots, x_n] < \infty$ pour tout $k \in I\!R$. On considère la fonction de coût pour l'estimation δ de θ définie sur $I\!R$ par

$$L_a(\theta, \delta) \quad = \quad \exp(a(\theta - \delta) - a(\theta - \delta) - 1$$

où a est un réel.

1. Montrer que $L_a(\theta, \delta) \geq 0$ pour tout $\theta \in \Theta$ et pour tout a et qu'elle est convexe en θ ; représenter cette fonction de coût comme une fonction de $(\theta - \delta)$ lorsque $a = \{0.1, 0.5, 1, 2\}$.
2. On suppose que $a > 0$. À quelles conditions cette fonction pénalise-t-elle les coûts de sous-estimation et de surestimation de θ de façon similaire ? Au contraire, à quelles conditions cette fonction pénalise-t-elle les coûts de sous-estimation et de surestimation de θ de façon très dissymétrique ?
3. On suppose que $a \neq 0$. Donner l'expression de l'estimateur de Bayes $\hat{\delta}_a$ sous cette fonction de coût.
4. Supposons que les données sont issues de $\mathcal{N}(\theta, 1)$ et que $\pi(\theta) \propto 1$; donnez l'estimateur de Bayes associé.

2 Élicitation d'*a priori* non informatif (7 pts)

On considère le problème suivant

$$x_i \quad \sim \quad \mathcal{N}(\mu_i, \sigma^2) \quad \text{pour } i = 1, \ldots, n$$

où les x_i sont indépendants.

1. Quelle est la densité jointe des données $x_1, \ldots, x_n$?
2. Calculer la matrice d'information I de Fisher pour ce jeu de données
3. En déduire la mesure *a priori* de Jeffreys $\pi^J(\theta)$ pour $\theta = (\mu_1, \ldots, \mu_n, \sigma)$
4. Que peut-on dire de $\pi^J(\sigma^2 | x_1, \ldots, x_n, \mu_1, \ldots, \mu_n)$? Est-ce une loi vue en cours ?

3 Élicitation et calcul bayésien pour un problème de Gumbel (14 pts)

La loi de Gumbel, de fonction de répartition

$$P(X < x|\theta) \quad = \quad \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\} \quad \text{avec } \sigma > 0,\ \mu \in I\!R \text{ et } x \in I\!R$$

et $\theta = (\mu, \sigma)$, est souvent utilisée en météorologie pour modéliser le comportement d'un échantillon de *maxima* d'une variable environnementale. Son espérance vaut $\mathrm{E}[X|\theta] = \mu + \sigma\gamma$ où γ est la constante d'Euler. On suppose connaître un échantillon de données de pluies (en mm) $\mathbf{x_n} = (x_1, \ldots, x_n)$ suivant cette loi. Elles sont fournies dans la table 1 et correspondent aux années 1987 à 2013. Par ailleurs on dispose d'une expertise *a priori* qui s'exerce sur la loi *a priori* prédictive de X, et est spécifiée statistiquement sous la forme $P(X < 75) = 25\%$, $P(X < 100) = 50\%$, $P(X < 150) = 75\%$.

107.6	72.4	204.5	83.8	142	95.5	316.1	177.9	87.3
81.9	109.1	89.5	150.7	122.1	98.2	113.2	104.4	66.9
136.4	275.4	125	199.8	51.2	75	168.2	106	72.8

TABLE 1 – Données de pluviométrie extrême.

On considère la mesure *a priori*

$$\pi(\mu, \sigma) \quad \propto \quad \sigma^{-m} \exp\left(m\frac{(\mu - \bar{\bar{x}}_m)}{\sigma} - \sum_{i=1}^{m} \exp\left\{-\frac{\tilde{x}_i - \mu}{\sigma}\right\}\right)$$

où les hyperparamètres $(m, \bar{\bar{x}}_m, \tilde{x}_1, \ldots, \tilde{x}_m)$ correspondent respectivement à la taille d'un échantillon de données *a priori* (virtuelles), sa moyenne et les données elles-mêmes (supposées calibrables).

1. Ecrivez la densité de la loi *a posteriori* conditionnelle aux données réelles $\mathbf{x_n}$. La loi *a priori* est-elle conjuguée ?
2. Produisez un algorithme qui simule la loi *a priori* prédictive de X en fonction des hyperparamètres et estime les quantiles prédictifs *a priori*. En fixant $m = 3$ et $(\tilde{x}_1, \tilde{x}_2) = (81, 93)$, testez les valeurs de $\tilde{x}_3$ suivantes : 97,101,110,120. Quelle calibration vous semble la plus adéquate vis-à-vis de l'expertise *a priori* ?

3. Pour les calibrations des hyperparamètres précédentes, écrivez un algorithme qui produit un tirage de la loi *a posteriori* de θ ainsi qu'une représentation (densité empirique) de la loi *a posteriori prédictive* sur X. Comparez avec un histogramme des données $\mathbf{x_n}$.

4. **Cette question peut être traitée indépendamment du reste.** On pose à présent $\mu > 0$ et on cherche à définir une nouvelle loi *a priori* $\pi_2(\theta)$ par maximum d'entropie qui est telle que les contraintes linéaires suivantes soient respectées :

$$\begin{aligned} \mathrm{E}[X] &= 100, \\ \mathrm{E}_\pi[\log \sigma] &= 1. \end{aligned}$$

   Formalisez et résolvez numériquement (possiblement graphiquement) le problème de maximum d'entropie en supposant que la mesure de référence est la mesure de Jeffreys $\pi_0(\theta) \propto \sigma^{-2}$ (valable pour le modèle de Gumbel). Sous quelles contraintes sur les multiplicateurs de Lagrange pouvez-vous trouver une loi jointe propre ? Celle-ci appartient-elle à une classe de lois connues ?

   Rappel : *Si Y suit une loi gamma $\mathcal{G}(a,b)$, alors E$[Y] = \Psi(a) - \log(b)$ où Ψ est la fonction digamma (*`digamma`* en R).*

5. Adaptez le code produit à la question 3 pour produire un nouveau calcul *a posteriori*, en utilisant $\pi_2(\theta)$.