Master Data science: examen.
Statistiques bayésiennes.

Anne Sabourin

Documents autorisés: Une feuille A4 manuscrite. Tout autre document et appareil électronique est interdit.

Notations et annexe: Les résultats fournis en annexe peuvent être utilisés sans justification supplémentaire. La notation $\log$ fait référence au logarithme népérien, *i.e.* $\log(\exp(x)) = 1$.

Exercice 1 (Données censurées)**.** On considère un durée de vie $Y > 0$ distribuée selon une loi exponentielle de paramètre θ inconnu. On choisit pour simplifier un prior Gamma de paramètres (a, λ). On n'observe pas la valeur exacte de Y mais seulement que $Y \geq 100$.

1. Calculer la loi a posteriori de θ
2. Calculer la moyenne et la variance a posteriori de θ.
3. On observe dans une autre expérience $Y = 100$. Calculez la loi a posteriori, l'espérance et la variance a posteriori. Comparez avec les résultats des questions précédentes.

Exercice 2 (Prior de Jeffreys)**.** Calculez le prior de Jeffreys et donnez la loi a posteriori pour n observations indépendantes dans

1. Le modèle de Bernoulli $Ber(p)$, de paramètre p.
2. Le modèle gaussien univarié avec variance connue $\sigma^2 = 1$ et moyenne inconnue.

Exercice 3 (Meilleure approximation au sens de Kullback-Leiber)**.** Soit $p(x)$ une densité de probabilité donnée sur $\mathbb{R}^d$ possédant des moments d'ordre 2, c'est-à-dire $\int_{\mathbb{R}^d} x_i^2 \ p(x) \, \mathrm{d}x < \infty$ pour $i \in \{1, \ldots, d\}$. Soit $X \sim p$. On souhaite approcher p par la densité de probabilité $q_{(\mu,\Lambda)}$ d'une loi Gaussienne de moyenne $\mu \in \mathbb{R}^d$ et matrice de covariance $\Sigma = \Lambda^{-1}$.

1. Calculer $KL(p||q_{(\mu,\Lambda)})$ en fonction de μ Λ et d'espérances sous la loi p de fonctions de X.
2. En calculant les gradients de $KL(p||q_{(\mu,\Lambda)})$ par rapport à μ et Λ, calculer les les paramètres μ^*, Λ^* minimisant la distance $KL(p||q)$, en fonction de l'espérance et de la matrice de covariance de X.

   Indication: si $u \in \mathbb{R}^d$ et $M \in \mathbb{R}^{d\times d}$, on a

   $$u^\top M u = Trace(u^\top M u) = Trace(u u^\top M)$$

   formulaire On pourra utiliser les résultats suivants sur les calculs de gradient par rapport à une matrice carrée $A \in \mathbb{R}^{d\times d}$:

(a) $\nabla_A \log \det(A) = (A^{-1})^\top$ (si A est inversible)

(b) $\nabla_A Trace(WA) = W$, avec $W \in \mathbb{R}^{d\times d}$.

Exercice 4 (Loi stationnaire)**.** On considère une chaîne de Markov sur $\mathbb{R}$ définie par

$$X_{t+1} = \rho X_t + \epsilon_t\,, \quad \text{where } (\epsilon_t)_{t\geq 1} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1) \text{ and } -1 < \rho < 1\,.$$

1. Calculez la densité du noyau de transition par rapport à la mesure de Lebesgue.
2. Montrez que la loi stationnaire pour la chaîne de Markov est une loi normale dont vous préciserez la moyenne et la variance.

ANNEXE

1. **Loi Exponentielle:** La densité de la loi exponentielle de paramètre θ est dobnnée par $p_\theta(y) = \theta e^{-y\theta}$, $y \geq 0$.
2. **Loi gamma** La densité de la loi gamma de paramètres (a, λ) est
$$p_{(a,\lambda)}(y) = \mathbb{1}_{y>0}\frac{\lambda^a}{\Gamma(a)} y^{a-1} e^{-\lambda y}.$$
L'espérance d'une variable aléatoire Z distribuée selon cette loi sont
$$\mathbb{E}_{a,\lambda}(Z) = \frac{a}{\lambda} \quad ; \quad \mathbb{V}\mathrm{ar}_{a,\lambda}(Z) = \frac{a}{\lambda^2}.$$
3. **Information de Fisher et prior de Jeffreys.** Dans le cas d'un paramètre $\theta \in \mathbb{R}$, l'information de Fisher est
$$I(\theta) = \mathbb{E}_\theta\Big(\big(\frac{\partial}{\partial\theta}\log p_\theta(X)\big)^2\Big)$$
où $X \sim P_\theta$ et $p_\theta(\,\cdot\,)$ est la densité de P_θ. Le prior de Jeffreys est la mesure ayant pour densité $\pi(\theta) \propto \sqrt{I(\theta)}$.
4. La **divergence de Kullback-Leibler** (KL) entre deux densités de probabilité π, q sur $\mathbb{R}^d$ est donnée par
$$KL(q||\pi) = \int_{\mathbb{R}^d} q(x) \log\left(\frac{q(x)}{\pi(x)}\right) \mathrm{d}x.$$