M1 - Statistiques bayésiennes

Examen

à rendre avant samedi 23 mai 12h à anna.ben-hamou@upmc.fr

Exercice 1

Soient $\delta, r > 0$. On considère le cadre bayésien suivant :

$$\boldsymbol{\theta} \sim \Pi = \mathcal{P}(\delta, r)$$
$$\mathbf{X} = (X_1, \dots, X_n) \,|\, \boldsymbol{\theta} \sim P_{\boldsymbol{\theta}}^{\otimes n} = \mathrm{Unif}[0, \boldsymbol{\theta}]^{\otimes \mathrm{n}}\,,$$

où $\mathcal{P}(\delta, r)$ correspond à la loi de Pareto de densité

$$z \mapsto \delta r^{\delta} z^{-\delta-1} \mathbb{1}_{[r,+\infty[}(z)\,.$$

1. Montrer que la loi a posteriori de $\boldsymbol{\theta}$ sachant $\mathbf{X}$ est une loi de Pareto $\mathcal{P}(\delta_{\mathbf{X}}, r_{\mathbf{X}})$ pour deux paramètres $\delta_{\mathbf{X}}$, $r_{\mathbf{X}}$ que l'on déterminera.
2. Quel est le mode a posteriori ?
3. Déterminer, pour tout $z \in \mathbb{R}$, la probabilité $\mathbb{P}(\boldsymbol{\theta} > z \,|\, \mathbf{X})$.
4. En déduire une région de crédibilité de plus haute densité a posteriori de niveau $1 - \alpha$, pour $\alpha \in]0, 1[$.
5. Dans cette question, on s'intéresse à la consistance de la loi a posteriori.
   (a) Montrer que si $0 < \theta_0 < r$, alors la loi a posteriori n'est pas consistante en θ_0.

   On suppose maintenant que $\theta_0 \geq r$ et l'on se place sous l'hypothèse $\mathbf{X} \sim P_{\theta_0}^{\otimes n}$.

   (b) Montrer que pour tout $\varepsilon > 0$, P_{θ_0}-presque sûrement,
   $$\mathbb{P}\left(|\boldsymbol{\theta} - \theta_0| > \varepsilon \,|\, \mathbf{X}\right) = \left(1 - \left(\frac{r_{\mathbf{X}}}{\theta_0 - \varepsilon}\right)^{\delta+n}\right) \mathbb{1}_{\{r_{\mathbf{X}} \leq \theta_0 - \varepsilon\}} + \left(\frac{r_{\mathbf{X}}}{\theta_0 + \varepsilon}\right)^{\delta+n}.$$
   (c) En déduire que pour tout $\varepsilon > 0$, P_{θ_0}-presque sûrement,
   $$\mathbb{P}\left(|\boldsymbol{\theta} - \theta_0| > \varepsilon \,|\, \mathbf{X}\right) \leq \mathbb{1}_{\{X_{(n)} \leq \theta_0 - \varepsilon\}} + \left(\frac{\theta_0}{\theta_0 + \varepsilon}\right)^{\delta+n},$$
   où $X_{(n)} = \max\{X_1, \dots, X_n\}$.

   (d) Montrer que la loi a posteriori converge à vitesse $\frac{1}{n}$ en tout point $\theta_0 \geq r$, i.e. que pour toute suite M_n qui tend vers $+\infty$ quand $n \to +\infty$, on a , sous $P_{\theta_0}^{\otimes n}$,
   $$\mathbb{P}\left(|\boldsymbol{\theta} - \theta_0| > \frac{M_n}{n} \,|\, \mathbf{X}\right) \xrightarrow{\mathbb{P}} 0\,.$$

Exercice 2

Soit $n \geq 1$ un entier connu. Pour $\theta \in \mathbb{R}_+^*$, on note P_θ la loi sur $\mathbb{N}$ donnée par

$$\forall k \geq 0\,,\ P_\theta(\{k\}) = \binom{n+k-1}{k} \theta^k (\theta+1)^{-(n+k)}\,.$$

On admet que si $X \sim P_\theta$, alors $\mathbb{E}_\theta[X] = n\theta$ et $\mathrm{Var}_\theta(X) = n\theta(\theta+1)$. On considère la fonction de perte donnée par

$$\forall \theta, t \in \mathbb{R}_+^*\,,\ \ell(\theta, t) = \frac{(t-\theta)^2}{\theta(\theta+1)}\,,$$

et, pour un estimateur T, on note $\theta \mapsto \mathbf{R}(\theta, T)$ la fonction de risque associée à la perte ℓ.

1. Soient T_0 et T_1 deux estimateurs de θ définis par

$$T_0(X) = \frac{X}{n} \qquad \text{et} \qquad T_1(X) = \frac{X}{n+1} \, .$$

Calculer les fonctions de risque de T_0 et de T_1. L'estimateur T_0 est-il admissible ?

Soient $a, b > 0$. On prend comme loi a priori $\Pi_{a,b}$ de densité donnée par

$$\forall \theta \in \mathbb{R} \, , \; \pi_{a,b}(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (\theta+1)^{-(a+b)} \mathbb{1}_{\mathbb{R}_+^*}(\theta) \, .$$

On se place alors dans le cadre bayésien : $\boldsymbol{\theta} \sim \Pi_{a,b}$ et $X \,|\, \boldsymbol{\theta} \sim P_{\boldsymbol{\theta}}$.

2. Déterminer la loi a posteriori $\Pi_{a,b}[\cdot \,|\, X]$.
3. Montrer que pour tous réels p, q tels que $a + p > 0$ et $b - p - q > 0$, on a

$$\mathbb{E}\left[\boldsymbol{\theta}^p (\boldsymbol{\theta}+1)^q\right] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+p)\Gamma(b-p-q)}{\Gamma(a+b-q)} \, .$$

4. Donner un estimateur de Bayes pour $\Pi_{a,b}$ et la perte ℓ. On le notera $T^\star$.
5. Montrer que $T^\star$ est admissible.
6. Que vaut le risque de Bayes $\mathbf{R}_B(\Pi_{a,b})$?
7. Calculer le risque maximal de T_1 et montrer que T_1 est minimax.

Exercice 3

Un dé a une probabilité θ_j de tomber sur la face j, pour $j \in \{1, \ldots, 6\}$, avec $\theta_j \in [0,1]$ et $\sum_{j=1}^6 \theta_j = 1$. Pour $\theta = (\theta_1, \ldots, \theta_6)$, on note P_θ la loi du résultat d'un lancer de ce dé et p_θ sa densité contre la mesure de comptage sur $\{1, \ldots, 6\}$, i.e.

$$\forall k \in \{1, \ldots, 6\} \, , \; p_\theta(k) = \theta_k \, .$$

On se place dans un cadre bayésien en choisissant une loi a priori Π sur l'ensemble des vecteurs de probabilité de longueur 6 et en posant

$$\begin{gathered} \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_6) \sim \Pi \\ \mathbf{X} = (X_1, \ldots, X_n) \,|\, \boldsymbol{\theta} \sim P_{\boldsymbol{\theta}}^{\otimes n} \, . \end{gathered}$$

1. Montrer que la famille des lois de Dirichlet est conjuguée pour ce modèle.
2. On souhaite savoir si le dé est équilibré ou non. On note $\theta_0 = \left(\frac{1}{6}, \ldots, \frac{1}{6}\right)$ et l'on se propose de tester

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta \neq \theta_0 \, .$$

Pour cela, on choisit comme loi a priori

$$\Pi = \frac{1}{2}\delta_{\theta_0} + \frac{1}{2}\mathrm{Dir}(1, \ldots, 1) \, ,$$

et l'on considère la fonction de perte équilibrée. En notant $N_j = \sum_{i=1}^n \mathbb{1}_{X_i = j}$, montrer que le test de Bayes s'écrit

$$\varphi(\mathbf{X}) = \mathbb{1}_{\left\{\prod_{j=1}^6 N_j! \geq r_n\right\}} \, ,$$

pour une certaine valeur r_n ne dépendant que de n que l'on explicitera.

Exercice 4

Pour $a, b, p > 0$ trois réels strictement positifs fixés, on considère le cadre bayésien suivant :

$$\boldsymbol{\theta} \sim \Pi = \Gamma(a, b)$$
$$\mathbf{X} = (X_1, \dots, X_n) \,|\, \boldsymbol{\theta} \sim P_{\boldsymbol{\theta}}^{\otimes n} = \Gamma(p, \boldsymbol{\theta})^{\otimes n}\,.$$

1. Quelle est la loi a posteriori $\Pi[\cdot \,|\, \mathbf{X}]$? Donner la moyenne a posteriori, notée $m_{\mathbf{X}}$, et la variance a posteriori, notée $v_{\mathbf{X}}$.
2. Déterminer la densité marginale de $\mathbf{X}$, notée f.
3. Montrer que la loi a posteriori est consistante en tout point $\theta_0 > 0$.
4. Déterminer l'estimateur du maximum de vraisemblance pour ce modèle.
5. Montrer que le modèle $(P_\theta)_{\theta>0}$ est régulier et donner l'information de Fisher de ce modèle.
6. Enoncer le théorème de Bernstein von Mises pour ce modèle, en vérifiant que les hypothèses sont vérifiées.