Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

1er décembre 2025
Aucun document autorisé
Durée : 1 heure

Contrôle Continu

Rappels

— Il sera tenu compte du soin apporté à la présentation et à la rédaction.
— Si vous n'arrivez pas à démontrer un résultat, vous pouvez l'admettre pour la suite de l'exercice.

Exercice 1 (5 points)

Soient $\alpha > 0$ et $r > 0$. On considère le cadre bayésien suivant :

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

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

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

1. (a) Représenter la densité de la loi $\mathcal{P}(\alpha, r)$.
Correction. Sur $]-\infty, r[$ la densité est nulle tandis que sur $[r, +\infty[$ elle décroît de α/r à 0.
(b) Déterminer, pour tout $t \in \mathbb{R}$, la probabilité $\mathbb{P}(\boldsymbol{\theta} > t)$.
Correction. Pour tout $t < r$, on a $\mathbb{P}(\boldsymbol{\theta} > t) = 1$, et pour $t \geq r$, on a
$$\mathbb{P}(\boldsymbol{\theta} > t) = \int_t^{+\infty} \alpha r^\alpha \theta^{-\alpha-1} d\theta = \left(\frac{r}{t}\right)^\alpha .$$
(c) Soit $\delta \in]0, 1[$. Trouver $c_\delta \in \mathbb{R}$ tel que $\mathbb{P}(\boldsymbol{\theta} > c_\delta) = \delta$.
Correction. Pour tout $t \geq r$, on a
$$\mathbb{P}(\boldsymbol{\theta} > t) = \delta \iff \left(\frac{r}{t}\right)^\alpha = \delta \iff t = \frac{r}{\delta^{1/\alpha}} .$$
Ainsi $c_\delta = \frac{r}{\delta^{1/\alpha}}$.
(d) Pour $\boldsymbol{\theta} \sim \mathcal{P}(\alpha, r)$ avec $\alpha > 1$, calculer $\mathbb{E}[\boldsymbol{\theta}]$.
Correction.
$$\mathbb{E}[\boldsymbol{\theta}] = \int_r^{+\infty} \alpha r^\alpha \theta^{-\alpha} d\theta = \alpha r^\alpha \left[\frac{\theta^{-\alpha+1}}{-\alpha+1}\right]_r^{+\infty} = \frac{\alpha}{\alpha-1} r.$$

2. Montrer que la loi a posteriori de $\boldsymbol{\theta}$ sachant $\mathbf{X}$ est une loi de Pareto $\mathcal{P}(\alpha_{\mathbf{X}}, r_{\mathbf{X}})$ pour deux paramètres $\alpha_{\mathbf{X}}$, $r_{\mathbf{X}}$ que l'on déterminera.
Correction. On a, pour tout $\theta \in \mathbb{R}$,
$$\begin{aligned}\pi(\theta \,|\, \mathbf{X}) &\propto \theta^{-\alpha-1} \mathbb{1}_{[r,+\infty[}(\theta) \frac{1}{\theta^n} \mathbb{1}_{[X_{(n)},+\infty[}(\theta) \\ &\propto \theta^{-\alpha-n-1} \mathbb{1}_{\theta \geq \max(r, X_{(n)})} ,\end{aligned}$$
où $X_{(n)} = \max(X_1, \ldots, X_n)$. Ainsi la loi a posteriori est une loi de Pareto $\mathcal{P}(\alpha_{\mathbf{X}}, r_{\mathbf{X}})$ avec $\alpha_{\mathbf{X}} = \alpha + n$ et $r_{\mathbf{X}} = \max(r, X_{(n)})$.

3. Donner la région de crédibilité de plus haute densité a posteriori de niveau $1-\delta$, pour $\delta \in]0,1[$.
Correction. D'après la question 1(a), la densité a posteriori est nulle sur $]-\infty, r_{\mathbf{X}}[$ et positive décroissante sur $[r_{\mathbf{X}}, +\infty[$, donc trouver la région de plus haute densité a posteriori de niveau $(1-\delta)$ revient à trouver $c_\delta \geq r_{\mathbf{X}}$ tel que
$$\mathbb{P}\left(\boldsymbol{\theta} > c_\delta \,\middle|\, \mathbf{X}\right) = \delta\,.$$
Par la question 1(c), on a $c_\delta = \frac{r_{\mathbf{X}}}{\delta^{\frac{1}{\alpha_{\mathbf{X}}}}}$. Ainsi la région de crédibilité de plus haute densité a posteriori au niveau $(1-\delta)$ est donnée par l'intervalle $\left[r_{\mathbf{X}}\,,\, r_{\mathbf{X}}\delta^{-\frac{1}{\alpha_{\mathbf{X}}}}\right]$.

4. On considère la fonction de perte $\ell(\theta, t) = \left(\frac{t}{\theta} - 1\right)^2$. Déterminer un estimateur de Bayes.
Correction. Par minimisation du risque a posteriori $\rho(\Pi, T|\mathbf{X})$ et la question 1(d), on trouve
$$T^\star(\mathbf{X}) = \frac{\alpha + n + 2}{\alpha + n + 1} r_{\mathbf{X}}\,.$$

Exercice 2 (5 points)

Soit $\varepsilon > 0$. On considère le cadre bayésien suivant :
$$\begin{aligned} \boldsymbol{\theta} &\sim \Pi_\varepsilon = \mathrm{Beta}(1+\varepsilon, 1+\varepsilon) \\ \mathbf{X} = (X_1, \ldots, X_n) \,\big|\, \boldsymbol{\theta} = \theta &\sim \mathcal{B}(\theta)^{\otimes n}\,, \end{aligned}$$
où $\mathcal{B}(\theta)$ est la loi de Bernoulli de paramètre θ.

1. Quelle est la loi a posteriori ?
Correction. C'est la loi $\mathrm{Beta}(n\overline{X}_n + 1 + \varepsilon, n - n\overline{X}_n + 1 + \varepsilon)$.
2. On considère la perte $\ell(\theta, t) = \frac{(t-\theta)^2}{\theta(1-\theta)}$.
   (a) Déterminer un estimateur de Bayes pour Π_ε et la perte ℓ. On le notera $T^\star$.
   Correction. On a
$$\mathbb{E}[\ell(\boldsymbol{\theta}, T) \,|\, \mathbf{X}] \propto \int_0^1 (\theta - T(\mathbf{X}))^2 \theta^{n\overline{X}_n + \varepsilon - 1}(1-\theta)^{n - n\overline{X}_n + \varepsilon - 1} d\theta\,.$$
   On est ramené à la perte quadratique pour la loi $\mathrm{Beta}(n\overline{X}_n + \varepsilon, n - n\overline{X}_n + \varepsilon)$. Le risque a posteriori est donc minimisé en $T^\star(\mathbf{X}) = \frac{n\overline{X}_n + \varepsilon}{n + 2\varepsilon}$.
   (b) Calculer sa fonction de risque.
   Correction. Pour tout $\theta \in]0,1[$, on a
$$\begin{aligned} \mathbf{R}(\theta, T^\star) &= \frac{1}{\theta(1-\theta)} \mathbb{E}_\theta\left[\left(\frac{n\overline{X}_n + \varepsilon}{n+2\varepsilon} - \theta\right)^2\right] \\ &= \frac{1}{\theta(1-\theta)} \mathbb{E}_\theta\left[\left(\frac{n}{n+2\varepsilon}(\overline{X}_n - \theta) + \frac{\varepsilon(1-2\theta)}{n+2\varepsilon}\right)^2\right] \\ &= \frac{1}{\theta(1-\theta)}\left(\left(\frac{n}{n+2\varepsilon}\right)^2 \mathrm{Var}_\theta(\overline{X}_n) + \left(\frac{\varepsilon(1-2\theta)}{n+2\varepsilon}\right)^2\right) \\ &= \frac{n}{(n+2\varepsilon)^2} + \frac{\varepsilon^2}{(n+2\varepsilon)^2}\frac{(1-2\theta)^2}{\theta(1-\theta)}\,, \end{aligned}$$
   où l'on a utilisé le fait que $\mathrm{Var}_\theta(\overline{X}_n) = \frac{\theta(1-\theta)}{n}$.

(c) En utilisant le fait que si $\boldsymbol{\theta} \sim \text{Beta}(1+\varepsilon, 1+\varepsilon)$, alors $\mathbb{E}\left[\frac{(1-2\boldsymbol{\theta})^2}{\boldsymbol{\theta}(1-\boldsymbol{\theta})}\right] = \frac{2}{\varepsilon}$, montrer que le risque de Bayes vaut

$$\mathbf{R}_B(\Pi_\varepsilon) = \frac{1}{n+2\varepsilon} \cdot$$

Correction. En utilisant l'indication on obtient

$$\mathbf{R}_B(\Pi_\varepsilon) = \frac{1}{n+2\varepsilon} \cdot$$

(d) L'estimateur $\overline{X}_n$ est-il minimax pour la perte ℓ ?
Correction. Pour tout $\theta \in]0,1[$, on a $\mathbf{R}(\theta, \overline{X}_n) = \frac{1}{n}$. Ainsi

$$\mathbf{R}_{\max}(\overline{X}_n) = \frac{1}{n} = \lim_{\varepsilon \to 0} \mathbf{R}_B(\Pi_\varepsilon) \,.$$

Par un résultat du cours, on en déduit que l'estimateur $\overline{X}_n$ est minimax.

3. On souhaite tester

$$H_0 : \boldsymbol{\theta} \leq \frac{1}{2} \quad \text{contre} \quad H_1 : \boldsymbol{\theta} > \frac{1}{2} \cdot$$

Déterminer le test de Bayes $\varphi^\star$ pour l'a priori $\Pi = \text{Unif}([0,1])$ et la perte équilibrée, et vérifier que

$$\varphi^\star(\mathbf{X}) = \mathbb{1}_{\{\overline{X}_n \geq \frac{1}{2}\}} \,.$$

Correction. Le test de Bayes s'écrit

$$\varphi^\star(\mathbf{X}) = \mathbb{1}_{\{\Pi(\Theta_1 \,|\, \mathbf{X}) \geq \Pi(\Theta_0 \,|\, \mathbf{X})\}} \,,$$

avec $\Theta_0 = [0, 1/2]$ et $\Theta_1 =]1/2, 1]$. On a

$$\begin{aligned}
\Pi(\Theta_1 \,|\, \mathbf{X}) \geq \Pi(\Theta_0 \,|\, \mathbf{X}) &\Leftrightarrow \int_{1/2}^{1} \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n} d\theta \geq \int_0^{1/2} \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n} d\theta \\
&\Leftrightarrow \int_0^{1/2} (1-\theta)^{n\overline{X}_n}\theta^{n-n\overline{X}_n} d\theta \geq \int_0^{1/2} \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n} d\theta
\end{aligned}$$

Or, pour tout $\theta \in]0, 1/2[$, on a $1-\theta > \theta$ et

$$(1-\theta)^{n\overline{X}_n}\theta^{n-n\overline{X}_n} \geq \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n} \Leftrightarrow \left(\frac{1-\theta}{\theta}\right)^{2n\overline{X}_n - n} \geq 1 \Leftrightarrow \overline{X}_n \geq \frac{1}{2}.$$

4. (Bonus) Démontrer la formule admise ci-dessus : $\mathbb{E}\left[\frac{(1-2\boldsymbol{\theta})^2}{\boldsymbol{\theta}(1-\boldsymbol{\theta})}\right] = \frac{2}{\varepsilon}$ si $\boldsymbol{\theta} \sim \text{Beta}(1+\varepsilon, 1+\varepsilon)$. On rappelle que $\text{Var}(\text{Beta}(a,b)) = ab/((a+b)^2(a+b+1))$.
Correction. Si on note $Y \sim \text{Beta}(\varepsilon, \varepsilon)$, alors le Théorème de transfert donne

$$\mathbb{E}\left[\frac{(1-2\boldsymbol{\theta})^2}{\boldsymbol{\theta}(1-\boldsymbol{\theta})}\right] = \frac{B(\varepsilon,\varepsilon)}{B(1+\varepsilon, 1+\varepsilon)} \mathbb{E}\left[(1-2Y)^2\right] = 4\frac{B(\varepsilon,\varepsilon)}{B(1+\varepsilon,1+\varepsilon)} \mathbb{E}\left[\left(Y - \frac{1}{2}\right)^2\right],$$

or $\mathbb{E}[Y] = 1/2$ donc

$$\mathbb{E}\left[\frac{(1-2\boldsymbol{\theta})^2}{\boldsymbol{\theta}(1-\boldsymbol{\theta})}\right] = 4\frac{B(\varepsilon,\varepsilon)}{B(1+\varepsilon,1+\varepsilon)} \text{Var}(Y).$$

Il reste à tenir compte du fait que $\text{Var}(Y) = \frac{1}{4(1+2\varepsilon)}$, $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ et $\Gamma(a+1) = a\Gamma(a)$ pour en déduire le résultat.