Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

Lundi 3 janvier 2022
Aucun document autorisé
Durée : 3 heures

Examen

Rappel : 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 (8 points)
On considère une variable aléatoire $Y \sim \mathcal{N}(0,1)$ de fonction de répartition notée Φ, un paramètre inconnu $\theta \in \mathbb{R}$, et la variable aléatoire $X = \theta + |Y|$. De plus, on considère $X_1, \ldots, X_n$ i.i.d. et de même loi que X.

1. Montrer que X admet pour densité
$$f_\theta(x) = \sqrt{\frac{2}{\pi}}\, e^{-\frac{(x-\theta)^2}{2}}\, \mathbf{1}_{x \geq \theta}.$$
2. Montrer que $\mathbb{E}[|Y|] = \sqrt{\frac{2}{\pi}}$ et $\mathrm{Var}(|Y|) = \frac{\pi-2}{\pi}$.
3. Par la méthode des moments, en déduire un estimateur consistant $\hat{\theta}_n$ de θ et préciser sa loi asymptotique.
4. Pour $0 < \alpha < 1$, construire un intervalle de confiance de niveau asymptotique $(1-\alpha)$.
5. Montrer que la médiane de X est $x_{1/2} = \theta + q_{3/4}$, où $q_{3/4} = \Phi^{-1}(3/4)$ est le quantile d'ordre $3/4$ de la loi normale centrée réduite.
6. En déduire un nouvel estimateur consistant $\tilde{\theta}_n$ de θ et donner sa loi asymptotique.
7. Montrer que l'estimateur du maximum de vraisemblance de θ est $X_{(1)} = \min_{1 \leq i \leq n} X_i$.
8. Donner le développement limité de Φ à l'ordre 1 en 0, c'est-à-dire déterminer les réels a et b dans l'expression $\Phi(x) = a + bx + o(x)$, et en déduire le développement limité de $2(1-\Phi(x))$ à l'ordre 1 en 0.
9. En déduire que $n\left(X_{(1)} - \theta\right)$ tend en loi vers une loi exponentielle de paramètre λ que vous préciserez.
10. Proposer un nouvel intervalle de confiance pour θ de niveau asymptotique $(1-\alpha)$ de la forme
$$\left[X_{(1)} - \frac{c_\alpha}{n} ; X_{(1)}\right],$$
où c_α est une constante positive que l'on précisera.
11. Des trois estimateurs précédents, lequel choisissez-vous pour n grand ?
12. Le modèle est-il régulier ? Si oui, préciser l'information de Fisher pour une observation.

Exercice 2 (2 points)
Soient $\mu \in \mathbb{R}$ et le modèle bayésien à une observation :

$$\boldsymbol{\theta} \sim \Pi = \mathcal{N}(\mu, \sigma^2)$$
$$\mathbf{X} = X_1 \mid \boldsymbol{\theta} = \theta \sim \mathcal{N}(\theta, 1).$$

1. Dans le cas où $\sigma^2 = 1$, montrer que la loi a posteriori s'écrit $\Pi[\cdot \mid \mathbf{X}] = \mathcal{N}\left(\frac{\mu + X_1}{2}, \frac{1}{2}\right)$.

2. Plus généralement, calculer $\Pi[\cdot \mid \mathbf{X}]$ pour $\sigma^2 > 0$ quelconque.

3. Montrer qu'un estimateur de Bayes pour l'a priori Π et la perte quadratique s'écrit

$$\hat{\theta} = \mu + \alpha(X_1 - \mu), \qquad \text{où } \alpha = \frac{\sigma^2}{\sigma^2 + 1}.$$

Exercice 3 (10 points)
Soient $a > 0, b > 0$ et une variable aléatoire $U \sim \Gamma(a, b)$. On rappelle que la loi Gamma de paramètres a et b, notée $\Gamma(a, b)$, a pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$x \mapsto \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{x>0},$$

et vérifie

$$\mathbb{E}[U] = \frac{a}{b}, \quad \mathrm{Var}(U) = \frac{a}{b^2} \quad \text{et} \quad \mathbb{E}[e^{tU}] = \left(1 - \frac{t}{b}\right)^{-a},$$

pour tout $t < b$. De plus, pour $\lambda > 0$, la loi de Poisson $\mathcal{P}(\lambda)$ est la loi sur les entiers de densité par rapport à la mesure de comptage sur $\mathbb{N}$: $x \mapsto \frac{\lambda^x}{x!} e^{-\lambda}$.

On considère le cadre bayésien à une observation suivant, où l'on supposera $a > 1$ (sauf pour la question 4.) et $b > 0$:

$$\boldsymbol{\theta} \sim \Pi = \Gamma(a, b)$$
$$\mathbf{X} = X_1 \mid \boldsymbol{\theta} = \theta \sim \mathcal{P}(\theta),$$

ainsi que la fonction de perte $L(\theta, T) = \frac{(T-\theta)^2}{\theta}$.

1. Déterminer le risque ponctuel $\mathbf{R}(\theta, T_M)$ pour la perte L de l'estimateur $T_M = X_1$.

2. Déterminer la loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$.

3. (a) Déterminer $\mathbb{E}[\boldsymbol{\theta} \mid \mathbf{X}]$ et $\mathbb{E}[\boldsymbol{\theta}^{-1} \mid \mathbf{X}]$ en fonction de a, b et X_1.
   (b) Montrer qu'un estimateur de Bayes pour la fonction de perte L et la loi a priori Π est

$$T^*_{a,b} = \frac{a + X_1 - 1}{b + 1}.$$

4. On suppose, dans cette question uniquement, que $a = 1$.

   (a) Montrer que le risque ponctuel $\mathbf{R}(\theta, T^*_{1,b})$ pour la perte L de l'estimateur $T^*_{1,b}$ s'écrit

$$\mathbf{R}(\theta, T^*_{1,b}) = \frac{b^2\theta + 1}{(b+1)^2}.$$

(b) Comparer ce risque ponctuel à celui de l'estimateur T_M. Un estimateur est-il préférable à l'autre ?

5. (a) Montrer que, pour tout $a > 1$ et $b > 0$,

$$\mathbf{R}_B(\Pi, T^*_{a,b}) = \frac{1}{b+1}.$$

Utiliser par exemple que pour tout estimateur T

$$\mathbf{R}_B(\Pi, T) = \mathbb{E}[L(\boldsymbol{\theta}, T(\mathbf{X}))] = \mathbb{E}[\mathbb{E}[L(\boldsymbol{\theta}, T(\mathbf{X})) \mid \mathbf{X}]]$$

ainsi que le calcul effectué à la question 3.(b).

(b) Comparer les risques bayésiens $\mathbf{R}_B(\Pi, T^*_{a,b})$ et $\mathbf{R}_B(\Pi, T_M)$ des estimateurs $T^*_{a,b}$ et T_M. Commenter.

6. On rappelle que si un estimateur T est tel qu'on puisse trouver $(\Pi_k)_{k\geq 1}$ suite de lois a priori avec

$$\mathbf{R}_{\max}(T) = \lim_{k\to\infty} \mathbf{R}_B(\Pi_k),$$

alors T est minimax. En déduire que $T_M = X_1$ est minimax pour le modèle des lois de Poisson.

On dispose maintenant de n observations $\mathbf{X} = (X_1, \ldots, X_n)$ et l'on suppose $\mathbf{X} \mid \boldsymbol{\theta} = \theta \sim \mathcal{P}(\theta)^{\otimes n}$. La loi a priori Π est toujours une loi $\Gamma(a, b)$. On note à présent $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

7. On souhaite réaliser l'analyse fréquentiste de la loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$. Pour ce faire, on se donne $\theta_0 > 0$ fixé.

(a) Déterminer la loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$.

(b) Montrer que pour tout $t > 0$ la suite de variables aléatoires

$$\mathbb{E}\left[e^{t\sqrt{n}(\boldsymbol{\theta}-\bar{X}_n)} \mid \mathbf{X}\right]$$

converge en probabilité, sous $\mathbb{P}_{\theta_0}$, vers $e^{t^2\theta_0/2}$. On pourra utiliser le développement limité $\log(1-u) = -u - u^2/2 + o(u^2)$.

(c) Pour un $v > 0$, on dit que la loi a posteriori (recentrée par $\bar{X}_n$ et renormalisée par $\sqrt{n}$) *converge en loi (en probabilité sous* $\mathbb{P}_{\theta_0}$*)* vers une loi $\mathcal{N}(0, v)$, si pour tout $t > 0$:

$$\mathbb{E}\left[e^{t\sqrt{n}(\boldsymbol{\theta}-\bar{X}_n)} \mid \mathbf{X}\right] \xrightarrow[n\to\infty]{\mathbb{P}_{\theta_0}} \mathbb{E}\left[e^{tZ}\right],$$

pour Z de loi $\mathcal{N}(0, v)$. Montrer que c'est bien le cas ici pour un réel v à déterminer.

(d) En rappelant l'estimateur du maximum de vraisemblance et l'information de Fisher dans le modèle des lois de Poisson, rapprocher cette conclusion d'un résultat du cours. Quelle différence éventuelle voyez-vous avec ce résultat ?

8. Dans un modèle où la loi a posteriori ne serait pas explicite,

(a) Donner un nom d'algorithme possible de simulation approchée de la loi a posteriori en expliquant son principe (en 1 ou 2 lignes maximum).

(b) Proposer une méthode de calcul approché de l'espérance de la question 7. (b) (pour $n \geq 1$ fixé) à partir de variables $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ simulées à partir d'un tel algorithme.