Sorbonne Université — 11 janvier 2024
Master Mathématiques et Applications — Aucun document autorisé
MU4MA015 : Statistique — Durée : 3 heures

Examen

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.
- Pour $a > 0$ et $b > 0$, la loi Gamma(a, b) a pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ la fonction
$$x \mapsto \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbb{1}_{x>0}.$$
La loi inverse-Gamma IG(a, b) est la loi de l'inverse d'une variable aléatoire de loi Gamma(a, b), et a pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ la fonction
$$x \mapsto \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x} \mathbb{1}_{x>0}.$$

Exercice 1 (9 points)

Soient $a > 0$ et $n \geq 2$. On considère le cadre bayésien suivant :

$$\boldsymbol{\theta} \sim \Pi = \mathrm{IG}(a, 1)$$
$$\mathbf{X} = (X_1, \dots, X_n) \,\big|\, \boldsymbol{\theta} \sim \mathcal{E}\left(\frac{1}{\boldsymbol{\theta}}\right)^{\otimes n},$$

et la fonction de perte $\ell : (\mathbb{R}_+^*)^2 \to \mathbb{R}_+$ définie par

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

1. Soit $Z \sim \mathrm{IG}(c, d)$, avec $c > 0$ et $d > 0$.
   (a) Montrer que si $c > 2$, alors $\mathbb{E}[Z] = \frac{d}{c-1}$ et $\mathrm{Var}(Z) = \frac{d^2}{(c-1)^2(c-2)}$.
   (b) Quelle est la loi de $\frac{1}{Z}$? En déduire $\mathbb{E}\left[\frac{1}{Z}\right]$ et $\mathrm{Var}\left(\frac{1}{Z}\right)$.
2. Donner la loi a posteriori $\Pi[\cdot \,|\, \mathbf{X}]$. Vérifier que l'espérance a posteriori $m_{\mathbf{X}}$ vaut $\frac{1+n\overline{X}_n}{a+n-1}$ et préciser la variance a posteriori $v_{\mathbf{X}}$.
3. Déterminer un estimateur de Bayes pour la fonction de perte ℓ et l'a priori Π. On le notera T^*.
4. Montrer que pour tout $\theta > 0$, on a
$$\mathbf{R}(\theta, T^*) = \frac{1}{(a+n+1)^2}\left(n + \left(\frac{1}{\theta} - (a+1)\right)^2\right),$$
où $\mathbf{R}$ est la fonction de risque associée à la fonction de perte ℓ.

5. En déduire le risque de Bayes $\mathbf{R}_B(\Pi)$ pour l'a priori Π et la perte ℓ.
6. Pour $\gamma > 0$, on pose $T_\gamma(\mathbf{X}) = \gamma \overline{X}_n$.
   (a) Calculer la fonction de risque de T_γ (pour la perte ℓ). Que vaut $\mathbf{R}_{\max}(T_\gamma)$?
   (b) Déterminer $\gamma^* = \arg\min_{\gamma>0} \mathbf{R}_{\max}(T_\gamma)$.
   (c) Montrer que T_{γ^*} est minimax.
7. Soit $\theta_0 > 0$.
   (a) Montrer que, sous $P_{\theta_0}^{\otimes n}$, on a
   $$n v_{\mathbf{X}} \xrightarrow{\mathbb{P}} \theta_0^2 .$$
   (b) Montrer que, sous $P_{\theta_0}^{\otimes n}$, on a
   $$\sqrt{n}\,(m_{\mathbf{X}} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta_0^2) .$$
   (c) Montrer que pour toute suite $(M_n)_{n\geq 1}$ avec $M_n \to +\infty$, on a, sous $P_{\theta_0}^{\otimes n}$:
   $$\mathbb{P}\left(|\boldsymbol{\theta} - \theta_0| > \frac{M_n}{\sqrt{n}} \,|\, \mathbf{X} \right) \xrightarrow{\mathbb{P}} 0 .$$

Exercice 2 (7,5 points)

On considère une variable aléatoire X de densité f_θ définie par
$$f_\theta(x) = \frac{3x^2}{\theta^3} \mathbf{1}_{[0,\theta]}(x),$$
avec $\theta > 0$ un paramètre inconnu que l'on souhaite estimer. Pour ce faire, on dispose de variables aléatoires i.i.d. $X_1, \ldots, X_n$ de même loi que X.

1. Le modèle est-il régulier ?
2. Calculer et représenter la fonction de répartition F de la variable aléatoire X.
3. Déterminer la médiane de la loi de X et en déduire un estimateur $\widehat{\theta}_n$ de θ. Préciser sa normalité asymptotique. Pour $0 < \alpha < 1$, donner un intervalle de confiance de niveau asymptotique $(1 - \alpha)$.
4. Déterminer l'estimateur du maximum de vraisemblance.
5. Soit $X_{(n)} = \max_{1\leq i\leq n} X_i$. Construire un intervalle de confiance de niveau $(1 - \alpha)$ de la forme $[X_{(n)}; c_{\alpha,n} X_{(n)}]$, où $c_{\alpha,n} \geq 1$ est une constante que l'on précisera.
6. Déduire de la question précédente un test au niveau α pour les hypothèses :
   $$H_0 : \theta = 1 \quad \text{contre} \quad H_1 : \theta \neq 1.$$
   Soit $x_{(n)}$ une réalisation de $X_{(n)}$. Calculer la p-valeur associée.
7. Montrer que $n(\theta - X_{(n)})$ tend en loi vers une loi exponentielle dont on précisera le paramètre.
8. Soient $\lambda > 0$ et $T \sim \mathcal{E}\left(\frac{\lambda}{\theta}\right)$. Déterminer $q_{\alpha,\theta}$ tel que $\mathbb{P}(0 \leq T \leq q_{\alpha,\theta}) = 1 - \alpha$. Déduire de ceci et de la question précédente un intervalle de confiance de niveau asymptotique $(1 - \alpha)$ pour θ.
   On prendra pour convention $[A, B] = \emptyset$ lorsque $B < A$ et on vérifiera que l'intervalle de confiance obtenu est de la forme $[X_{(n)}; d_{\alpha,n} X_{(n)}]$, où $d_{\alpha,n}$ est une constante (supérieure à 1 à partir d'un certain rang) que l'on précisera.
9. Pour $\alpha \in]0, 1[$ fixé et lorsque n tend vers l'infini, donner des développements asymptotiques de la forme $c_{\alpha,n} \sim 1 + c_\alpha/n$ et $d_{\alpha,n} \sim 1 + d_\alpha/n$. Commenter.

10. Entre $X_{(n)}$ et $\widehat{\theta}_n$, quel estimateur choisissez-vous ?

Exercice 3 (3,5 points)

Les questions 1. et 2. de cet exercice sont indépendantes.

1. On dispose du tableau de données suivant, représentant des réalisations $x_1, \dots, x_n$ de variables aléatoires réelles $X_1, \dots, X_n$ i.i.d., et on souhaite tester (au niveau $\alpha \in]0,1[$)

$$H_0 : \theta \leq 1 \qquad \text{contre} \qquad H_1 : \theta > 1$$

dans le modèle $\mathcal{P} = \{\mathcal{N}(\theta, 1), \theta \in \mathbb{R}\}$.

```
print(x.shape)
print(x[:10])
```

```
(200,)
[ 3.25617103 -1.24290386  0.21805856 -0.72901134  2.81435003  2.29702943
  1.68190239  1.17001241 -0.76642552  1.7031285 ]
```

(a) On note Φ la fonction de répartition de $\mathcal{N}(0,1)$. Par quoi faut-il remplacer `???` et `$$$` pour calculer la réalisation du test $T(x) = \mathbb{1}_{\overline{x}_n > 1 + \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha)}$ (avec $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$) et la p-valeur associée $\alpha_0(x) = 1 - \Phi(\sqrt{n}(\overline{x}_n - 1))$?

```
alpha = 0.05
threshold = 1 + stats.norm.ppf(1-alpha) / np.sqrt(n)
test = ??? > threshold
alpha0 = 1 - $$$
print(test, alpha0)
```

```
False 0.4135433062246249
```

(b) Au vu du résultat, rejette-t-on H_0 ?

2. (a) Expliquer en quoi le script suivant échantillonne approximativement suivant la loi de densité proportionnelle à la fonction $f : (x, y) \mapsto y^2 \, \mathrm{e}^{-(1+x)y} \mathbb{1}_{x \geq 0,\, y \geq 0}$. On détaillera la réponse.

```
n = 5000

x, y = 0, 1
sample = [(x, y)]
for i in range(n):
    x = stats.expon(scale=1/y).rvs()
    y = stats.gamma(a=3, scale=1/(1+x)).rvs()
    sample.append((x, y))
sample = np.asarray(sample)
```

(b) Justifier l'illustration ci-dessous.

```
u = np.linspace(0, 10, num=200)
```

```
plt.hist(sample[:, 1], density=True, bins='auto')
plt.plot(u, stats.gamma(a=2, scale=1).pdf(u))
```