Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

9 Janvier 2025
Aucun document autorisé
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.

Exercice 1 (8 points)
Soit $\theta \in \mathbb{R}$ un paramètre inconnu et X de densité

$$f_\theta(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{((\log x)-\theta)^2}{2}} \mathbf{1}_{x>0}.$$

On rappelle que si $Y \sim \mathcal{N}(0,1)$, alors pour tout réel s, on a $\mathbb{E}[e^{sY}] = e^{\frac{s^2}{2}}$.

1. Montrer que si $Y \sim \mathcal{N}(0,1)$, alors X a même loi que la variable aléatoire $e^{\theta+Y}$. Grâce au rappel, en déduire $\mathbb{E}_\theta[X]$ et $\mathrm{Var}_\theta(X)$.
2. On suppose désormais disposer d'un échantillon i.i.d. $(X_1, \ldots, X_n)$ de même loi que X. Déduire de la méthode des moments un estimateur $\widehat{\theta}_n$ de θ et établir sa normalité asymptotique.
3. On note comme d'habitude Φ la fonction de répartition de la loi $\mathcal{N}(0,1)$. Grâce à la première question, donner l'expression de la fonction de répartition $F_\theta(x)$ de la variable X en fonction de Φ, θ et x. En déduire la médiane de la loi de X.
4. Déduire de la question précédente un nouvel estimateur $\widetilde{\theta}_n$ de θ et établir sa normalité asymptotique.
5. Déterminer l'estimateur du maximum de vraisemblance $\check{\theta}_n$. Grâce à la première question, préciser sa loi.
6. Le modèle $(f_\theta)_{\theta\in\mathbb{R}}$ est-il régulier ? Si oui, calculer l'information de Fisher $I(\theta)$ pour une donnée. Préciser si l'un des estimateurs précédents est asymptotiquement efficace.
7. Soit $\alpha \in]0,1[$. Donner un intervalle de confiance de niveau $(1-\alpha)$ pour θ.
8. Proposer un test de niveau α pour décider entre $H_0 : \theta = 0$ et $H_1 : \theta \neq 0$. Déterminer la p-value $\alpha_0(\mathbf{x})$ associée à une réalisation $\mathbf{x} = (x_1, \ldots, x_n)$.

Exercice 2 (9 points)
Pour $a > 0$ et $b > 0$, la loi inverse-Beta(a,b), notée $\mathrm{IB}(a,b)$, est la loi dont la densité est donnée par

$$\theta \mapsto \frac{1}{B(a,b)} \theta^{-a-b} (\theta-1)^{b-1} \mathbb{1}_{]1,+\infty[}(\theta)\,,$$

où $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ et Γ est la fonction gamma, vérifiant, pour tout $x > 0$, $\Gamma(x+1) = x\Gamma(x)$. La distribution $\mathrm{IB}(a,b)$ est la loi de l'inverse d'une variable aléatoire de loi $\mathrm{Beta}(a,b)$. Par ailleurs, si $Y \sim$

$\mathrm{IB}(a,b)$ avec $a > 2$, on a

$$\mathbb{E}[Y] = \frac{a+b-1}{a-1} \qquad \text{et} \qquad \mathrm{Var}(Y) = \frac{(a+b-1)b}{(a-1)^2(a-2)}\,.$$

On rappelle que si X suit la loi géométrique de paramètre $p \in]0,1[$, notée $\mathcal{G}(p)$, alors pour tout $k \in \mathbb{N}^*$, $\mathbb{P}(X = k) = (1-p)^{k-1}p$. De plus, $\mathbb{E}[X] = \frac{1}{p}$ et $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

On fixe maintenant $a > 0$ et un entier $n \geq 2$, et l'on considère le modèle :

$$\boldsymbol{\theta} \sim \Pi := \mathrm{IB}(a,2)$$
$$\mathbf{X} = (X_1, \ldots, X_n) \,\big|\, \boldsymbol{\theta} = \theta \sim \mathcal{G}\left(\frac{1}{\theta}\right)^{\otimes n}.$$

1. Montrer que la loi a posteriori est donnée par $\Pi[\,\cdot \,|\, \mathbf{X}] = \mathrm{IB}(a+n, n\overline{X}_n - n + 2)$, où $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
2. Donner l'espérance a posteriori, notée $m_{\mathbf{X}}$, et la variance a posteriori, notée $v_{\mathbf{X}}$.

On considère la fonction de perte ℓ définie par

$$\forall \theta > 1\,,\ t > 1\,,\ \ell(\theta, t) = \frac{(t-\theta)^2}{\theta(\theta-1)}\,.$$

3. Donner un estimateur de Bayes, noté T^*, pour la loi a priori Π et la perte ℓ. On pourra introduire la loi $\widetilde{\Pi}_{\mathbf{X}} = \mathrm{IB}(a+n+2, n\overline{X}_n - n + 1)$.
4. Calculer le risque a posteriori de T^*, et en déduire que le risque de Bayes $\mathbf{R}_B(\Pi)$ vaut $\frac{1}{n+a+1}$.
5. On pose $T'(\mathbf{X}) = \frac{n\overline{X}_n + 1}{n+1}$.
   (a) Pour $\theta > 1$, calculer le risque ponctuel $\mathbf{R}(\theta, T')$ de T' (toujours pour la perte ℓ).
   (b) En déduire le risque maximal de T', noté $\mathbf{R}_{\max}(T')$.
   (c) Montrer que T' est minimax.
6. Soit $\theta_0 > 1$.
   (a) Montrer que, sous $\mathcal{G}\left(\frac{1}{\theta_0}\right)^{\otimes n}$, on a

$$n v_{\mathbf{X}} \xrightarrow[n\to+\infty]{\mathbb{P}} \theta_0(\theta_0 - 1) \qquad \text{et} \qquad \sqrt{n}(m_{\mathbf{X}} - \theta_0) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \theta_0(\theta_0-1))\,.$$

   (b) En déduire que pour toute suite positive $(M_n)_{n\geq 1}$ avec $M_n \xrightarrow[n\to\infty]{} +\infty$, on a, sous $\mathcal{G}\left(\frac{1}{\theta_0}\right)^{\otimes n}$,

$$\mathbb{P}\left(|\boldsymbol{\theta} - \theta_0| > \frac{M_n}{\sqrt{n}} \,\Big|\, \mathbf{X}\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0\,.$$

Exercice 3 (3 points)
Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}_+$, définie par :

$$\forall (u,v) \in \mathbb{R}^2, \quad f(u,v) = \frac{(\cos|u+v|)^2}{100 + u^4 + v^4}.$$

On admet que $\iint_{\mathbb{R}^2} f(u,v)\,\mathrm{d}u\mathrm{d}v < \infty$.

1. Par quoi faut-il remplacer `???` et `$$$` pour que l'algorithme de Metropolis-Hastings ci-dessous, où `f` représente la fonction f, simule une suite de variables aléatoires (X_i) dont la loi limite est de densité proportionnelle à f ?

[3]:
```python
unif = stats.uniform()
kernel = stats.multivariate_normal(cov=1e-2 * np.eye(2))

n = 10000

threshold = unif.rvs(size=n)  # Array of the n thresholds U_i
move = kernel.rvs(size=n)  # Array of the n moves from the current position

sample = np.zeros((n+1, 2))
x = sample[0]
for i in range(n):
    y = x + move[i]
    crit = f(y) / f(x)
    dec = ???
    if dec:
        sample[i+1] = y
        x = y
    else:
        $$$
```

2. Expliquer cet algorithme grâce à des éléments théoriques.
3. Voici une représentation des courbes de niveau de f et d'une réalisation de la suite $(X_i)_{0\leq i\leq n}$ construite par l'algorithme précédent (`Sample` sur la figure). Quelle est votre analyse de ce résultat ? Quelle correction peut-on apporter ?