Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

Janvier 2023
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.
- Si X_1 et X_2 sont deux vecteurs gaussiens indépendants et de même dimension, avec $X_1 \sim \mathcal{N}(m_1, \Sigma_1)$ et $X_2 \sim \mathcal{N}(m_2, \Sigma_2)$, alors $Y = X_1 + X_2$ est aussi un vecteur gaussien et $Y \sim \mathcal{N}(m_1 + m_2, \Sigma_1 + \Sigma_2)$.
- On rappelle que la loi $\mathcal{E}(\lambda)$ correspond à la loi $\Gamma(1, \lambda)$.

Exercice 1 (6 points)
Soit $\theta > 0$ un paramètre inconnu et X de densité

$$f_\theta(x) = \frac{1}{\theta}(1-x)^{\frac{1}{\theta}-1}\mathbf{1}_{]0,1[}(x).$$

Dans la suite, on suppose disposer d'un échantillon $(X_1, \dots, X_n)$ i.i.d. de même loi que X.

1. Calculer la fonction de répartition F_θ de X. En déduire la médiane de la loi de X.
2. En déduire un estimateur $\hat{\theta}_n$ de θ. Etablir sa normalité asymptotique.
3. Donner un intervalle de confiance de niveau asymptotique $(1-\alpha)$ pour θ.
4. Déterminer l'estimateur du maximum de vraisemblance $\tilde{\theta}_n$ de θ.
5. Montrer que $T = -\log(1-X)$ suit une loi exponentielle dont on précisera le paramètre.
6. D'un point de vue asymptotique, entre $\hat{\theta}_n$ et $\tilde{\theta}_n$, quel estimateur choisissez-vous ?
7. Montrer que $\tilde{\theta}_n/\theta$ suit une loi Gamma $\Gamma(r, \lambda)$ dont on précisera les paramètres r et λ (qui ne dépendent pas de θ). En déduire un test de niveau (non asymptotique) α pour décider entre $H_0 : \theta \leq 1$ et $H_1 : \theta > 1$.

Exercice 2 (4 points)
On suppose disposer de deux ensembles $T_1, \dots, T_\ell$ et $U_1, \dots, U_m$ de variables aléatoires réelles. On note $T = [T_1, \dots, T_\ell]' \in \mathbb{R}^\ell$ et $U = [U_1, \dots, U_m]' \in \mathbb{R}^m$ les vecteurs correspondants. On suppose que les vecteurs T et U sont observés et donnés par

$$T = V\beta + \sigma\eta \qquad \text{et} \qquad U = W\gamma + \tau\varepsilon$$

où β, $\gamma \in \mathbb{R}^p$ sont inconnus, V et W sont des matrices connues de tailles $\ell \times p$ et $m \times p$ et de rang p (avec $\ell > p$ et $m > p$), σ et τ sont des réels positifs inconnus, et η, ε sont deux vecteurs gaussiens standards indépendants non observés et de dimensions respectives ℓ et m. On désigne par $\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}^2$ et $\hat{\tau}^2$ les estimateurs usuels des moindres carrés de β, γ, σ^2 et τ^2 respectivement.

1. Rappeler la loi de $(\ell - p)\hat{\sigma}^2/\sigma^2$ ainsi que son espérance et sa variance. Quelle est la loi de $(\tau^2\hat{\sigma}^2)/(\sigma^2\hat{\tau}^2)$?
2. On veut tester $\tau = \sigma$ contre l'alternative $\tau \neq \sigma$. Proposer un test de niveau α.
3. Quelle est la loi de la variable $(\ell - p)\frac{\hat{\sigma}^2}{\sigma^2} + (m - p)\frac{\hat{\tau}^2}{\tau^2}$?
4. Jusqu'à la fin de l'exercice, on suppose que $\tau = \sigma$. Déduire de la question précédente un estimateur sans biais $\tilde{\sigma}^2$ de σ^2. Préciser sa variance. Comment se compare-t-il à $\hat{\sigma}^2$ en terme de risque quadratique ?
5. Quelle est la loi du vecteur $\hat{\beta} - \hat{\gamma}$?
6. Soit $X \sim \mathcal{N}(m, \Gamma)$ un vecteur gaussien en dimension d, avec Γ inversible. Rappeler (sans le justifier) la loi de la variable $(X - m)'\Gamma^{-1}(X - m)$.
7. Construire un test de niveau α de l'hypothèse $\beta = \gamma$ contre l'alternative $\beta \neq \gamma$.

Exercice 3 (7 points)
Soient $\delta, r > 0$. La loi de Pareto $\mathcal{P}(\delta, r)$ est la loi de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$
$$x \mapsto \delta r^\delta x^{-\delta-1}\mathbf{1}_{[r,+\infty[}(x)\,.$$
On considère le cadre bayésien suivant:
$$\boldsymbol{\theta} \sim \Pi = \mathcal{P}(\delta, r)$$
$$\mathbf{X} = (X_1, \ldots, X_n) \mid \boldsymbol{\theta} \sim P_{\boldsymbol{\theta}}^{\otimes n} = \mathrm{Unif}[0, \boldsymbol{\theta}]^{\otimes \mathrm{n}}\,.$$

1. Représenter la densité de la loi $\mathcal{P}(\delta, r)$ et donner sa fonction de répartition.
2. 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.
3. Donner une région de crédibilité de plus haute densité a posteriori de niveau $1 - \alpha$, pour $\alpha \in]0, 1[$.
4. 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 on rappelle que $\mathbf{X}|\boldsymbol{\theta} = \theta_0 \sim P_{\theta_0}^{\otimes n}$.

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

Soit $(M_n)_{n\geq 1}$ une suite de réels positifs qui tend vers $+\infty$ quand $n \to +\infty$.

(d) Montrer que

$$\mathbb{P}_{\theta_0}\left(X_{(n)} \leq \theta_0 - \frac{M_n}{n}\right) \underset{n\to\infty}{\longrightarrow} 0\,.$$

On pourra utiliser l'inégalité: $\forall u \in [0,1]$, $(1-u)^n \leq e^{-un}$.

(e) Montrer que

$$\left(\frac{\theta_0}{\theta_0 + \frac{M_n}{n}}\right)^{\delta+n} \underset{n\to\infty}{\longrightarrow} 0\,.$$

(f) En déduire que, sous $P_{\theta_0}^{\otimes n}$, on a

$$\mathbb{P}\left(|\boldsymbol{\theta} - \theta_0| > \frac{M_n}{n} \mid \mathbf{X}\right) \overset{\mathbb{P}}{\underset{n\to\infty}{\longrightarrow}} 0\,.$$

Exercice 4 (3 points)
Expliquer en termes mathématiques précis ce que fait chacune des parties de l'illustration numérique suivante.

Partie 1

[2]:
```
a, b = 2, 3
n = 30

prior = stats.gamma(a=a, scale=1/b)
theta0 = prior.rvs()
ptheta0 = stats.norm(loc=theta0)
X = ptheta0.rvs(size=n)
```

Partie 2

[3]:
```
def likelihood(theta):
    return stats.norm(loc=theta).pdf(X).prod()

est_emv = X.mean()

N = 1000
thetas = prior.rvs(size=N)
likelihoods = np.array([likelihood(theta) for theta in thetas])
est_mc = (thetas*likelihoods).sum() / likelihoods.sum()

print(est_emv)
print(est_mc)
```

```
0.1743723482282432
0.27901366347949874
```

Partie 3

```
[4]: M = 2000
thetas = prior.rvs(size=M)
likelihoods = np.array([likelihood(theta) for theta in thetas])
print(((thetas-est_emv)**2 * likelihoods).sum() / likelihoods.sum())
print(((thetas-est_mc)**2 * likelihoods).sum() / likelihoods.sum())
```

```
0.025305233550915397
0.017664889523622763
```

Partie 4

```
[5]: unif = stats.uniform()

m = 5000
l = 2*m

sample = []
while len(sample) < m:
    candidates = prior.rvs(size=l)
    thresholds = unif.rvs(size=l)
    crit = np.exp(-n/2 * (est_emv - candidates)**2)
    sample += list(candidates[crit > thresholds])
sample = sample[:m]

est_abs = np.quantile(sample, q=0.5, interpolation='lower')

print((np.abs(thetas-est_emv) * likelihoods).sum() / likelihoods.
 ↪sum())
print((np.abs(thetas-est_mc) * likelihoods).sum() / likelihoods.sum())
print((np.abs(thetas-est_abs) * likelihoods).sum() / likelihoods.
 ↪sum())
```

```
0.12283677437841431
0.10911696680300526
0.1066337249029515
```