École Normale Supérieure
FIMFA — Statistique 2016–2017

Examen

1. INFÉRENCE GAUSSIENNE

Dans cet exercice $X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n$ sont des variables aléatoires gaussienne indépendantes de variance σ^2. Pour chaque i, X_i et Y_i ont même espérance μ_i.

(a) (1 point) On suppose d'abord les μ_i connus, explicitez l'estimateur au maximum de vraisemblance $\widehat{\sigma}_n^2$ de σ^2. Est-il sans biais ? Calculez son risque quadratique. Calculez l'information de Fisher du modèle. La suite des estimateur

(b) (2 points) On suppose maintenant les μ_i inconnus. Explicitez l'estimateur au maximum de vraisemblance $\widetilde{\sigma}_n^2$ de σ^2. Explicitez sa distribution. Cet estimateur est-il sans biais ? Calculez son risque quadratique.

(c) (1 point) Toujours en supposant les μ_i inconnus, l'estimateur $\widetilde{\sigma}^2$ est-il admissible ? Si on considère le sous-modèle formé par $\{\sigma^2 : \sigma^2 \leq 1\}$ atteint-il le risque quadratique minimax ?

2. RÉGRESSION (NON GAUSSIENNE)

Dans cet exerice, $\mathbf{X}$ est une matrice de dimensions $n \times p$ avec $n > p$ et de rang p. La matrice $\mathbf{X}$ est supposée connue. On observe le vecteur aléatoire $Y := \mathbf{X}\theta + \sigma\epsilon$, où $\theta \in \mathbb{R}^p$, σ est une quantité positive, et ϵ est un vecteur dont les composantes sont i.i.d. et distribuées selon une loi symétrique : soit une loi de Rademacher $\epsilon = \pm 1$ avec probabilité $1/2$, soit une loi de Laplace (densité $\frac{1}{2}\exp(-|x|)$).

(a) (1 point) Proposer un estimateur sans biais de θ.

(b) (2 points) Proposer un estimateur sans biais de $\mathrm{var}(\sigma\epsilon)$.

(c) (2 points) Si on suppose que les ϵ_i sont des Rademacher indépendantes ($|\epsilon_i| = 1$ avec probabilité 1), calculer la variance de votre estimateur de σ^2.

(d) (2 points) Si on suppose que les ϵ_i sont distribuées selon une loi de Laplace ($\mathbb{P}\{|\epsilon_i| > x\} = \exp(-x)$ pour tout $x > 0$), calculer la variance de votre estimateur.

(e) (1 point) Toujours en supposant les ϵ_i sont distribuées selon des lois de Laplace, quel serait l'estimateur de θ obtenu en appliquant la méthode du maximum de vraisemblance ? Quelle méthode algorithmique vous semble appropriée pour calculer l'estimateur ?

3. LOI GÉOMÉTRIQUE

Soit $p \in]0,1[$. On dit que la variable aléatoire X suit la loi géométrique de paramètre p (et on note $X \sim Geom(p)$) si X est à valeurs dans $\mathbb{N}^*$ et $P[X = k] = (1-p)^{k-1}p$.

(a) (1 point) Soit $X \sim Geom(p)$. Calculer l'espérance et la variance de X.

(b) (1 point) On se donne un échantillon $X_1, \ldots, X_n$ de variables iid géométriques de paramètre p inconnu et à estimer. Calculer l'estimateur du maximum de vraisemblance, noté $\hat{p}_n$. Cet estimateur est-il sans biais ? Convergent ?

(c) (1 point) Donner la loi asymptotique de $\hat{p}_n$ lorsque $n \to \infty$.

(d) (1 point) Construire un intervalle de confiance de niveau $\alpha \in]0,1[$ pour p.

(e) (1 point) Inférence Bayésienne. On rappelle que la loi $Beta(a,b)$ ($a > 0$, $b > 0$) est la loi de probabilité à valeurs réelles de densité par rapport à la mesure de Lebesgue

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{I}_{[0,1]}(x).$$

Vérifier que la famille des lois $Beta$ est conjuguée pour le modèle géométrique. Si on prend comme loi a priori sur p la loi $Beta(a,b)$, quelle est la loi a posteriori ?

(f) (1 point) Dans le cadre de la question précédente, on prend comme estimateur ponctuel $\hat{p}'_n$ l'espérance a posteriori de p. Étudier le comportement asymptotique de $\hat{p}'_n$ lorsque $n \to \infty$.

(g) (1 point) Soit $X_n \sim Geom(\frac{1}{n})$ et $Y_n = \frac{X_n}{n}$. Donner la limite en loi de Y_n lorsque $n \to \infty$.

4. Modèle autorégressif (gaussien)

Un processus autorégressif gaussien $(X_t)_{t\in\mathbb{N}}$ est défini par une récurrence stochastique

$$X_{t+1} = \theta X_t + Z_{t+1} \qquad \text{pour } t \geq 0$$

avec $\theta \in \mathbb{R}$ et $Z_1, Z_2, \ldots, Z_t, \ldots \sim_{i.i.d.} \mathcal{N}(\mu, \sigma^2)$ et $X_0 \sim \mathcal{N}(\nu, \eta^2)$.
On observe $X_0, X_1, \ldots, X_n$. Les innovations $Z_1, \ldots, Z_n$ ne sont pas observées.

(a) (1 point)
  i. $(X_0, X_1, \ldots, X_n)^T$ est-il un vecteur gaussien ?
  ii. Calculer la matrice de covariance de $(X_0, X_1, \ldots, X_n)^T$.
  iii. Calculer l'espérance de $(X_0, X_1, \ldots, X_n)^T$.
  iv. A quelle condition sur θ, ν, η^2, les X_t peuvent-ils être identiquement distribués ?
  Si les X_t sont identiquement distribués, on dit que le processus est stationnaire.

(b) (1 point) Quelle est la loi conditionnelle de X_n sachant $X_0, X_1, \ldots, X_{n-1}$?

(c) (1 point) On considère le modèle paramétré par (θ, μ, σ^2) avec $|\theta| < 1, \nu = \mu/(1-\theta)$ et $\eta^2 = \sigma^2/(1-\theta^2)$. En supposant θ, σ^2 connus, proposez l'estimateur au maximum de vraisemblance $\widehat{\mu}_n$ de μ. Caractérisez sa loi.

(d) (1 point) Toujours sous les hypothèses de la question précédente, proposez un intervalle de niveau de confiance $1-\alpha$ pour μ.

(e) (2 points) En supposant θ connu et $\mu = 0$, calculer l'estimateur au maximum de vraisemblance $\widehat{\sigma}^2_n$ de σ^2. Montrer que la suite $\widehat{\theta}_n$ est consistante. Caractériser la loi limite de $(\sqrt{n}(\widehat{\sigma}^2_n - \sigma^2))_n$.

5. Estimation dans un modèle de irrégulier simplifié

On note

$$f_0(x) = 1 \qquad \text{pour} \quad x \in [0,1]$$

la densité de la loi uniforme sur $[0,1]$. La densité $f_{0,\mu}$ est définie par

$$f_{0,\mu}(x) = f_0(x-\mu).$$

On s'intéresse au modèle échantillonné Θ_0 défini par les densités $\{f_{0,\mu} : \mu \in \mathbb{R}\}$ (il s'agit d'un modèle de translation).

On se souviendra ou on admettra la propriété suivante : si $Y_1, \ldots, Y_{n+1}$ sont des variables aléatoires exponentielles indépendamment et identiquement distribuées, alors les variables aléatoires $X_i := ((\sum_{j=1}^{i} Y_j)/(\sum_{j=1}^{n+1} Y_j))$ pour $i \in 1, \ldots, n$ sont distribuées comme les statistiques d'ordre d'un n-échantillon de la loi uniforme sur $[0,1]$.

On cherche à estimer le paramètre de translation μ à partir d'un échantillon i.i.d. $X_1, \ldots, X_n$ collecté selon $f_{0,\mu}$.

(a) (1 point) Calculer les estimateurs au maximum de vraisemblance $\widehat{\mu}_n$.

(b) (2 points) Choisissez parmi les estimateurs au maximum de vraisemblance un estimateur sans biais, étudier sa variance.

(c) (1 point) Déterminer la loi limite de cet estimateur après recentrage et renormalisation.

(d) (1 point) Proposer un intervalle de niveau de confiance asymptotique $1-\alpha$ pour le paramètre de localisation μ.

(e) (1 point) Donner un équivalent de $H^2(f_{0,0}, f_{0,\mu})$ lorsque μ tend vers 0.

(f) (1 point) Déterminer la loi limite du rapport de vraisemblance entre $f_{0,0}$ et $f_{0,\mu/n}$ (sous $f_{0,0}$).