EXAMEN STATISTIQUE - 1MF2E

Lundi 22 mars 2021, 8h15-10h00

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1 : Estimation (12 points)

On considère n variables aléatoires indépendantes $X_1, ..., X_n$ suivant la même loi normale $\mathcal{N}(0, \sigma^2)$ où $\sigma^2 > 0$ est un paramètre inconnu. L'objectif de cet exercice est d'étudier différentes méthodes d'estimation du paramètre $\theta = \sigma^2$.

1. (2pts) Déterminer la vraisemblance de $X_1, ..., X_n$ et montrer qu'elle admet un unique maximum global pour une valeur de σ^2 que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance du paramètre σ^2 noté $\widehat{\sigma}^2_{\text{MV}}$ en fonction des variables X_i.

2. (2pts) Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre σ^2 à partir de l'observation des variables $X_1, ..., X_n$. L'estimateur $\widehat{\sigma}^2_{\text{MV}}$ est-il l'estimateur efficace de σ^2 ?
Indication : on rappelle que si $X \sim \mathcal{N}(0, \sigma^2)$, on a $E[X^4] = 3\sigma^4$.

3. Les données $x_1, ..., x_n$ sont soumises à une non-linéarité quadratique et on ne peut qu'observer les données transformées $y_i = x_i^2$. On désire estimer le paramètre θ à partir des données $y_1, ..., y_n$.
   - (2pts) Montrer que la densité de la variable aléatoire $Y_i = X_i^2$ est une loi gamma $\Gamma\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$.
   - (1pt) Déterminer l'estimateur du maximum de vraisemblance du paramètre σ^2 noté $\widetilde{\sigma}^2_{\text{MV}}$ construit à partir des variables Y_i.
   - (1pt) Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre σ^2 à partir de l'observation des variables $Y_1, ..., Y_n$.
   - (1pt) Peut-on espérer obtenir de meilleurs estimateurs de σ^2 en utilisant les données $y_1, ..., y_n$ plutôt que les données $x_1, ..., x_n$?

4. (3pts) On suppose maintenant qu'on dispose d'une information a priori sur le paramètre σ^2 résumée dans la loi a priori $p(\sigma^2)$ définie comme une loi inverse gamma de paramètres $\theta = \sigma_0^2$ et $\nu = 2$ (la moyenne de cette loi a priori est égale à σ_0^2), c'est-à-dire

$$p(\sigma^2) = \frac{1}{\sigma^6} \exp\left(-\frac{\sigma_0^2}{\sigma^2}\right) \mathcal{I}_{\mathbb{R}^+}(\sigma^2)$$

   où $\mathcal{I}_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($\mathcal{I}_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et $\mathcal{I}_{\mathbb{R}^+}(x) = 0$ si $x \leq 0$). Montrer que la loi a posteriori de $\sigma^2 | y_1, ... y_n$ est une loi inverse gamma dont on déterminera les paramètres. En déduire que l'estimateur de la moyenne a posteriori du paramètre σ^2 construit à partir des données $y_1, ..., y_n$ s'écrit

$$\widehat{\sigma}^2_{\text{MMSE}} = \beta(n) \left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right) + [1 - \beta(n)]\, \sigma_0^2$$

   où $\beta(n)$ est une fonction de n qu'on déterminera. Expliquer le comportement de $\beta(n)$ lorsque $n \to \infty$ et $n \to 0$.

Exercice 2 : Test Statistique (10 **points)**

On considère n variables aléatoires $Y_1, ..., Y_n$ indépendantes de densités

$$f(y_i; \theta) = \frac{1}{\sqrt{2\pi\theta y_i}} \exp\left(-\frac{y_i}{2\theta}\right) \mathbb{I}_{\mathbb{R}^+}(y_i)$$

avec $\theta > 0$ et où $\mathbb{I}_{\mathbb{R}^+}$ indique la fonction indicatrice sur $\mathbb{R}^+$. On veut à l'aide d'observations $y_1, ..., y_n$ de ces variables aléatoires tester les deux hypothèses suivantes

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \qquad \text{avec } \theta_1 > \theta_0.$$

1. (2pts) À l'aide du théorème de Neyman-Pearson, calculer la statistique T_n du test le plus puissant et indiquer la région critique de ce test (comme d'habitude, on retiendra pour T_n la fonction seule des observations).

2. (2pts) Montrer que $\frac{Y_i}{\theta}$ possède une loi du χ^2_1. En déduire que $\frac{T_n}{\theta}$ suit une loi du χ^2_n.

3. (2pts) En déduire une expression des risques de première et seconde espèce α et β du test en fonction des paramètres θ_0 et θ_1 et de la fonction de répartition d'une loi du χ^2_n notée F_n.

4. (2pts) Déterminer les courbes COR associées à ce test et expliquer comment les performances de ce test dépendent des valeurs de θ_0 et θ_1.

5. (2pts) On désire à partir d'un ensemble d'observations $(y_1, ..., y_n)$ déterminer s'il est raisonnable de supposer que ces observation sont issues de la densité $f(y_i; \theta)$. Expliquer le principe du test de Kolmogorov pour résoudre ce problème: on précisera avec soin 1) comment calculer la statistique de ce test notée D_n, 2) la région critique du test, 3) l'expression du seuil du test en fonction du risque de première espèce α et de la loi asymptotique de $\sqrt{n}D_n$ et 4) la manière de prendre en compte le fait que θ soit inconnu.

LOIS DE PROBABILITÉ CONTINUES

$\mathbf{m}$: moyenne σ^2 : variance **F. C. :** fonction caractéristique

LOI	Densité de probabilité	m	σ^2	F. C.
Uniforme	$f(x) = \frac{1}{b-a}$ $x \in]a, b[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\dfrac{e^{itb} - e^{ita}}{it(b-a)}$
Gamma $\mathcal{G}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0, \ \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n! \ \forall n \in \mathbb{N}$	$\dfrac{\nu}{\theta}$	$\dfrac{\nu}{\theta^2}$	$\dfrac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
Inverse gamma $\mathcal{IG}(\nu, \theta)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0, \ \nu > 0$ $x \geq 0$ avec $\Gamma(n+1) = n! \ \forall n \in \mathbb{N}$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{-\lvert x \rvert}, \quad x \in \mathbb{R}$	0	2	$\dfrac{1}{1+t^2}$
Normale univariée $\mathcal{N}(m, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\boldsymbol{m}, \boldsymbol{\Sigma})$	$f(x) = K e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{m})}$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T \boldsymbol{m} - \frac{1}{2} \boldsymbol{u}^T \boldsymbol{\Sigma} \boldsymbol{u}}$
Khi$_2$ χ^2_ν $\Gamma\left(\frac{1}{2}, \frac{\nu}{2}\right)$	$f(x) = k e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}$ $k = \dfrac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}$ $\nu \in \mathbb{N}^*, \ x \geq 0$	ν	2ν	$\dfrac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda,\alpha}$	$f(x) = \dfrac{1}{\pi\lambda\left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0, \ \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda\lvert t \rvert}$
Beta $B(a, b)$	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0, \ b > 0$ $x \in]0, 1[$ avec $\Gamma(n+1) = n! \ \forall n \in \mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

LOIS DE PROBABILITÉ DISCRÈTES

$\mathbf{m}$: moyenne σ^2 : variance **F. C. :** fonction caractéristique

$p_k = P[X = k]$ $p_{1,\dots,m} = P[X_1 = k_1, \dots, X_m = k_m]$

LOI	Probabilités	m	σ^2	F. C.
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, \dots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}\left(1 - e^{itn}\right)}{n\left(1 - e^{it}\right)}$
Bernoulli	$p_1 = P[X = 1] = p$ $p_0 = P[X = 0] = q$ $p \in [0, 1] \quad q = 1 - p$	p	pq	$pe^{it} + q$
Binomiale $B(n, p)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \{0, 1, \dots, n\}$	np	npq	$\left(pe^{it} + q\right)^n$
Binomiale négative	$p_k = C_{n+k-1}^{n-1} p^n q^k$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}$	$n\dfrac{q}{p}$	$n\dfrac{q}{p^2}$	$\left(\dfrac{p}{1 - qe^{it}}\right)^n$
Multinomiale	$p_{1,\dots,m} = \frac{n!}{k_1!\dots k_m!} p_1^{k_1} \dots p_m^{k_m}$ $p_j \in [0, 1] \quad q_j = 1 - p_j$ $k_j \in \{0, 1, \dots, n\}$ $\sum_{j=1}^m k_j = n \quad \sum_{j=1}^m p_j = 1$	np_j	Variance : $np_j q_j$ Covariance : $-np_j p_k$	$\left(\sum_{j=1}^m p_j e^{it}\right)^n$
Poisson $P(\lambda)$	$p_k = e^{-\lambda} \dfrac{\lambda^k}{k!}$ $\lambda > 0 \quad k \in \mathbb{N}$	λ	λ	$\exp\left[\lambda\left(e^{it} - 1\right)\right]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1 - qe^{it}}$