EXAMEN STATISTIQUE - 1MF2E

Vendredi 8 mars 2019

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

Exercice 1 : Estimation (8 points)

On considère n variables aléatoires indépendantes $X_1, ..., X_n$ suivant la même loi normale $\mathcal{N}(\mu, \sigma_1^2)$ et m variables aléatoires $Y_1, ..., Y_m$ suivant la même loi normale $\mathcal{N}(\mu, \sigma_2^2)$, où $\mu \in \mathbb{R}$ est un paramètre inconnu et σ_1^2, σ_2^2 sont deux paramètres connus. On supposera de plus que les vecteurs $(X_1, ..., X_n)^T$ et $(Y_1, ..., Y_m)^T$ sont indépendants. Une situation pratique dans laquelle on peut avoir ces deux ensembles d'observations $(X_1, ..., X_n)^T$ et $(Y_1, ..., Y_m)^T$ correspond par exemple au cas où $(X_1, ..., X_n)^T$ sont obtenues par un premier capteur dont l'incertitude dépend de la variance σ_1^2 et où $(Y_1, ..., Y_m)^T$ sont obtenues par un second capteur dont l'incertitude dépend de la variance σ_2^2. L'objectif de cet exercice est d'étudier des estimateurs de μ basés sur les vecteurs $(X_1, ..., X_n)^T$ et $(Y_1, ..., Y_m)^T$.

1. On appelle vraisemblance conjointe de ces deux ensembles d'observations le produit des vraisemblances des deux échantillons $(X_1, ..., X_n)^T$ et $(Y_1, ..., Y_m)^T$ notée $L(\boldsymbol{x}, \boldsymbol{y}; \mu)$ avec $\boldsymbol{x} = (x_1, ..., x_n)$ et $\boldsymbol{y} = (y_1, ..., y_m)$. Montrer que cette vraisemblance admet un unique maximum global pour une valeur de μ que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance (conjointe) du paramètre μ noté $\widehat{\mu}_{\text{MV}}$ en fonction des variables X_i et Y_j et des paramètres σ_1^2 et σ_2^2.

2. Montrer que
$$\widehat{\mu}_{\text{MV}} = \alpha \bar{X} + (1 - \alpha)\bar{Y}$$
avec $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$, et où $\alpha \in]0, 1[$ est un paramètre qu'on exprimera en fonction de n, m, σ_1^2 et σ_2^2. L'estimateur $\widehat{\mu}_{\text{MV}}$ est-il sans biais ? Déterminer la variance de $\widehat{\mu}_{\text{MV}}$ et montrer qu'elle tend vers 0 lorsque $n = m$ et que n tend vers l'infini.

3. Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre μ à partir de l'observation des variables $(X_1, ..., X_n)^T$ et $(Y_1, ..., Y_m)^T$. L'estimateur $\widehat{\mu}_{\text{MV}}$ est-il l'estimateur efficace du paramètre μ ?

4. En comparant les expressions des bornes de Cramér-Rao pour $\sigma_1^2 = \sigma_2^2 = \sigma^2$ et pour $\sigma_1^2 = 2\sigma_2^2 = \sigma^2$, déterminer quelle situation est la plus favorable pour l'estimation du paramètre μ.

Exercice 2 : Test Statistique (7 points)

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes de densités

$$f(x_i; \theta) = \frac{3}{\theta} x_i^2 \exp\left(-\frac{x_i^3}{\theta}\right) \mathbb{I}_{\mathbb{R}^+}(x)$$

avec $\theta > 0$ et où $\mathbb{I}_{\mathbb{R}^+}(x)$ indique la fonction indicatrice sur $\mathbb{R}^+$. On veut à l'aide d'observations $x_1, ..., x_n$, 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. 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. On retiendra pour T_n la fonction seule des observations.

2. Vérifier que la loi de $Y_i = \frac{2}{\theta}X_i^3$ est une loi du chi-deux à 2 degrés de liberté. Montrer que si Z et T sont deux variables aléatoires indépendantes suivant des lois du chi-deux à n_Z et n_T degrés de liberté, c'est-à-dire, $Z \sim \chi^2_{n_Z}$ et $T \sim \chi^2_{n_T}$, alors on a $Z + T \sim \chi^2_{n_Z+n_T}$.

3. En utilisant les résultats de la question précédente, exprimer le seuil du test de Neyman-Pearson en fonction du risque de première espèce α et de l'inverse de la fonction de répartition d'une loi du chi-deux dont on prendra soin de déterminer le nombre de degrés de liberté. On notera $F_\nu(x)$ la fonction de répartition d'une loi du chi-deux à ν degrés de liberté.

4. Déterminer la puissance du test en fonction du seuil du test de Neyman-Pearson, de θ_1 et de F_ν, où ν est le nombre de degrés de liberté trouvé à la question précédente.

5. Déterminer les courbes COR associées à ce test et analyser leur comportement en fonction de θ_0 et θ_1.

Exercice 3 : Test d'adéquation (5 points)

Pour tester si une variable X possède la densité

$$f(x;\theta) = \frac{3}{\theta}x^2 \exp\left(-\frac{x^3}{\theta}\right) \mathbb{I}_{\mathbb{R}^+}(x)$$

où $\theta > 0$ est un paramètre connu, on considère un échantillon $(X_1, ..., X_n)$ de la var X et teste si la loi de $Y = \frac{2}{\theta}X^3$est une loi du chi-deux à deux degrés de liberté. En d'autres termes, on considère le test d'hypothèses suivant :

$$\begin{aligned} H_0 &: Y \sim \chi^2_2 \\ H_1 &: \text{non } H_0 \end{aligned}$$

Malheureusement, on n'a pu obtenir que $n = 5$ réalisations de la variable aléatoire X qui donnent les valeurs suivantes de Y :

$y_1 = 1.1$	$y_2 = 2.3$	$y_3 = 5.4$	$y_4 = 7.8$	$y_5 = 10.2$

On décide de faire un test de Kolmogorov. Un logiciel de statistique (comme Matlab) nous a permis de calculer les valeurs de la fonction de répartition aux points y_i :

$F(y_1) = 0.42$	$F(y_2) = 0.68$	$F(y_3) = 0.93$	$F(y_4) = 0.98$	$F(y_5) = 0.99$

Effectuer un test de Kolmogorov avec les risques de première espèce $\alpha = 0.01$ et $\alpha = 0.05$ (les seuils associés à ces valeurs de α sont $S_{0.01} = 0.352$ et $S_{0.05} = 0.294$) et répondre aux questions suivantes

1. Quelle est la statistique du test de Kolmogorov ?

2. Comment conclure au vu des résultats des tests précédents ?

3. Expliquer pourquoi on a $S_{0.05} < S_{0.01}$.

4. Peut-on calculer la puissance du test ? Pourquoi ?

LOIS DE PROBABILITÉ CONTINUES

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 $\Gamma(\theta, \nu)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0,\ \nu > 0$ $x \geq 0$	$\dfrac{\nu}{\theta}$	$\dfrac{\nu}{\theta^2}$	$\dfrac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$
Inverse gamma IG(θ, ν)	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\frac{\theta}{x}} \frac{1}{x^{\nu+1}}$ $\theta > 0,\ \nu > 0$ $x \geq 0$	$\frac{\theta}{\nu-1}$ si $\nu > 1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu > 2$	(*)
Première loi de Laplace	$f(x) = \frac{1}{2} e^{-\lvert x \rvert}$	0	2	$\dfrac{1}{1+t^2}$
Normale $\mathcal{N}\left(m, \sigma^2\right)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
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	$f(x) = k x^{a-1} (1-x)^{b-1}$ $k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a > 0,\ b > 0,\ x \in]0, 1[$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

LOIS DE PROBABILITÉ DISCRÈTES

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

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

LOI	Probabilités	m	σ^2	F. C.
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, ..., n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}\left(1 - e^{itn}\right)}{n\left(1 - e^{it}\right)}$
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 $B(n, p)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \{0, 1, ..., n\}$	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	$p_{1,...,m} = \frac{n!}{k_1!...k_m!} p_1^{k_1} ... p_m^{k_m}$ $p_j \in [0, 1] \quad q_j = 1 - p_j$ $k_j \in \{0, 1, \ldots, n\}$ $\sum_{j=1}^m k_j = n \quad \sum_{j=1}^m p_j = 1$	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	$p_k = pq^{k-1}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1 - qe^{it}}$