EXAMEN STATISTIQUE - 1SN

Mardi 19 janvier 2021 (13h45-15h15)

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

Exercice 1 : Estimation (11 points)

On considère n observations $x_1, ..., x_n$ issues d'un vecteur $(X_1, ..., X_n)$ de n variables aléatoires indépendantes de lois normales telles que $X_k \sim \mathcal{N}(km, \sigma^2)$, où σ^2 est un paramètre connu et m est un paramètre inconnu que l'on cherche à estimer. On remarquera que le vecteur $(X_1, ..., X_n)$ n'est pas un échantillon car il est constitué de variables aléatoires n'ayant pas la même loi (la moyenne de X_k dépend de k). On rappelle que la densité d'une loi normale $\mathcal{N}(km, \sigma^2)$ est

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-km)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}$$

et que cette loi est de moyenne km et de variance σ^2.

1. On construit l'estimateur
$$\widetilde{m}_n = \frac{1}{n}\sum_{k=1}^{n} \frac{X_k}{k}.$$
Montrer que $\widetilde{m}_n$ est un estimateur non biaisé et convergent du paramètre m.
Indication : pour la convergence, on pourra montrer que la variance de $\widetilde{m}_n$ notée var$(\widetilde{m}_n)$ est inférieure à $\frac{\sigma^2}{n}$.

2. Montrer que la vraisemblance de $(x_1, ..., x_n)$ admet un unique maximum global pour une valeur de m que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance du paramètre m noté $\widehat{m}_{\text{MV}}$.
Remarque : on rappelle que $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \triangleq s_n$.

3. Montrer que $\widehat{m}_{\text{MV}}$ est un estimateur sans biais et convergent du paramètre m. Donner un équivalent de la variance de $\widehat{m}_{\text{MV}}$ lorsque $n \to \infty$ et commenter ce résultat.

4. Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre m. L'estimateur $\widehat{m}_{\text{MV}}$ est-il l'estimateur efficace du paramètre m ? Quel estimateur doit-on préférer entre $\widetilde{m}_n$ et $\widehat{m}_{\text{MV}}$?

5. On suppose désormais que le paramètre m est muni d'une loi a priori normale de moyenne m_0 et de variance σ_0^2 (i.e., $m \sim \mathcal{N}(m_0, \sigma_0^2)$), où m_0 et σ_0^2 sont connus. Déterminer l'estimateur du maximum a posteriori du paramètre m noté $\widehat{m}_{\text{MAP}}$ et montrer que
$$\widehat{m}_{\text{MAP}} = \beta(n)\widehat{m}_{\text{MV}} + [1-\beta(n)]m_0$$
où $\beta(n)$ est une quantité dépendant de n à déterminer. Analyser le comportement de $\widehat{m}_{\text{MAP}}$ lorsque n est petit ($n \to 0$) ou grand ($n \to \infty$).

Exercice 2 : Tests Statistiques (9 points)

On considère n observations $x_1, ..., x_n$ issues issues d'un vecteur $(X_1, ..., X_n)$ de n variables aléatoires indépendantes de lois normales telles que $X_k \sim \mathcal{N}(km, \sigma^2)$, où σ^2 est un paramètre connu. On désire utiliser les observations $x_1, ..., x_n$ pour déterminer si $m = 0$ ou si $m = m_1 < 0$. On considère donc le test d'hypothèses

$$H_0 : m = 0, \quad H_1 : m = m_1 \qquad \text{avec } m_1 < 0.$$

1. Montrer que la statistique de test du théorème de Neyman-Pearson est

   $$T_n = \sum_{k=1}^{n} kX_k$$

   et indiquer la région critique de ce test.

2. Montrer que la loi de T_n est une loi normale sous les deux hypothèses H_0 et H_1 avec des paramètres à préciser sous chaque hypothèse.

   Indication : on rappelle que $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \triangleq s_n$.

3. On note F la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$. Exprimer les risques de première et seconde espèce α et β en fonction du seuil du test de Neyman-Pearson noté S, de F et de σ et m_1.

4. Déterminer les caractéristiques opérationnelles du récepteur (courbes COR) pour ce test et analyser les performances en fonction des valeurs de σ et m_1. Donner l'allure de ces courbes pour m_1 fixé et pour différentes valeurs de σ (par exemple, $\sigma \in \{0.1, 1, 10\}$). Faire de même pour différentes valeurs de m_1 (à σ fixé), par exemple $m_1 \in \{-10, -1, -0.1\}$.

5. On désire vérifier que l'hypothèse $X_k \sim \mathcal{N}(km, \sigma^2)$ est correcte (avec m et σ^2 connus). Pour cela, on construit les variables $Y_k = \frac{X_k - km}{\sigma}$ et on cherche à vérifier si ces variables suivent une loi normale centrée réduite $\mathcal{N}(0, 1)$ à l'aide d'un test du χ^2. Expliquer le principe de ce test (avec $K = 4$ classes).

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	$f(x) = \frac{1}{2} e^{-\|x\|}, \quad x \in \mathbb{R}$	0	2	$\dfrac{1}{1+t^2}$
Normale univariée $\mathcal{N}\left(m, \sigma^2\right)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\boldsymbol{m}, \boldsymbol{\Sigma})$	$f(x) = K \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{m})\right]$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T\boldsymbol{m} - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u}}$
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 $c_{\lambda,\alpha}$	$f(x) = \dfrac{1}{\pi\lambda\left(1 + \left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda > 0,\ \alpha \in \mathbb{R}$	(-)	(-)	$e^{i\alpha t - \lambda\|t\|}$
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

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

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	$p_k = pq^{k-1}$ $p \in [0, 1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1 - qe^{it}}$