EXAMEN STATISTIQUE - 1SN

Lundi 17 janvier 2022 (8h00-9h30)

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

Exercice 1 : Estimation (9 points)

On considère n observations $x_1, ..., x_n$ issues d'un vecteur $(X_1, ..., X_n)$ de n variables aléatoires indépendantes de lois de Laplace de densité

$$f(x_i; m, b) = \frac{1}{2b} \exp\left(-\frac{|x_i - m|}{b}\right), \quad x_i \in \mathbb{R}$$

avec $m \in \mathbb{R}$ et $b > 0$. On admettra qu'une telle loi est de moyenne $E[X_i] = m$ et de variance $\text{var}[X_i] = 2b^2$. On suppose dans cet exercice que m est connu et on cherche à estimer le paramètre b.

1. (2pts) Déterminer les variations de la vraisemblance de $X_1, ..., X_n$ en fonction de b. En déduire l'estimateur du maximum de vraisemblance du paramètre b noté $\widehat{b}_{\text{MV}}$.

2. (2pts) Montrer que $Y_i = |X_i - m|$ suit une loi gamma dont on déterminera les paramètres. En déduire que $\widehat{b}_{\text{MV}}$ est un estimateur sans biais et convergent du paramètre b.

3. (1pt) Déterminer la borne de Cramer-Rao pour un estimateur non biaisé du paramètre b. L'estimateur $\widehat{b}_{\text{MV}}$ est-il l'estimateur efficace du paramètre b ?

4. (4pts) On suppose désormais que $m = 0$ et que le paramètre b est muni d'une loi a priori inverse-gamma de paramètres α et β (i.e., $b \sim \mathcal{IG}(\alpha, \beta)$), où α et β sont connus.

   - Montrer que la loi a posteriori de $b|x_1, ..., x_,$ est une loi inverse gamma dont on déterminera les paramètres.
   - Déterminer l'estimateur du maximum a posteriori du paramètre b noté $\widehat{b}_{\text{MAP}}$ et montrer que

   $$\widehat{b}_{\text{MAP}} = a_\alpha(n)\widehat{b}_{\text{MV}} + [1 - a_\alpha(n)]\frac{\beta}{\alpha + 1}$$

   où $a_\alpha(n)$ est une quantité dépendant de n et de α à déterminer. Analyser le comportement de $\widehat{b}_{\text{MAP}}$ lorsque n est petit ($n \to 0$) ou grand ($n \to \infty$).

Exercice 2 : Tests Statistiques (7 points)

On considère n observations $x_1, ..., x_n$ issues d'un vecteur $(X_1, ..., X_n)$ de n variables aléatoires indépendantes de lois de Laplace de densité

$$f(x_i; m, b) = \frac{1}{2b} \exp\left(-\frac{|x_i - m|}{b}\right), \quad x_i \in \mathbb{R}$$

On admettra qu'une telle loi est de moyenne $E[X_i] = m$ et de variance $\text{var}[X_i] = 2b^2$. On désire utiliser les observations $x_1, ..., x_n$ pour déterminer si $b = b_0$ ou si $b = b_1 < b_0$ dans le cas où le paramètre m est connu. On considère donc le test d'hypothèses

$$H_0 : b = b_0, \quad H_1 : b = b_1 \qquad \text{avec } b_1 < b_0.$$

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

$$T_n = \sum_{k=1}^{n} |X_k - m|$$

et indiquer la région critique de ce test.

2. (2pts) On admet que $Y_k = |X_k - m|$ suit une loi exponentielle de paramètre $1/b$ (i.e., une loi gamma $\Gamma\left(1, \frac{1}{b}\right)$) de densité

$$p(y_k) = \begin{cases} \frac{1}{b} \exp\left(-\frac{y_k}{b}\right) \text{ si } y_k > 0 \\ 0 \text{ sinon} \end{cases}$$

Déterminer la loi de T_n sous les deux hypothèses H_0 et H_1 et son approximation résultant de l'utilisation du théorème de la limite centrale. Dans la suite de cet exercice, on utilisera la loi approchée de T_n car elle permet d'obtenir des résultats plus simples.

3. (2pts) On note F la fonction de répartition d'une loi du normale $\mathcal{N}(0, 1)$. Exprimer les risques de première et seconde espèce α et β du test Neyman-Pearson en fonction du seuil de ce test noté S, de F, n, b_0 et b_1.

4. (2pts) Déterminer les caractéristiques opérationnelles du récepteur (courbes COR) pour ce test en utilisant la loi de T_n issue de l'application du théorème de la limite centrale. Expliquer comment les performances du test varient, d'une part quand n augmente, d'autre part quand $\frac{b_0}{b_1}$ augmente.

Exercice 3 : Test de Kolmogorov (5 points)

On désire vérifier si l'hypothèse d'une loi de Laplace pour un ensemble de n variables aléatoires $X_1, ..., X_n$ est correcte ou non. Pour cela, on centre et on réduit les variables X_i et on cherche à vérifier si les variables centrées réduites notées $Y_1, ..., Y_n$ suivent une loi de Laplace de moyenne nulle et de variance unité à l'aide d'un test de Kolmogorov. La loi de Laplace de moyenne nulle et de variance unité a pour densité de probabilité

$$\pi(y_i) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|y_i|\right), \quad y_i \in \mathbb{R} \tag{1}$$

1. (1pt) Expliquer les opérations à mener pour centrer et réduire les données, i.e., comment déterminer les variables Y_i à partir des variables X_i.

2. (1pt) Déterminer la fonction de répartition de la loi de densité π définie dans (1).

3. (2pt) Rappeler la statistique du test de Kolmogorov que l'on exprimera en fonction des données ordonnées $y_1^* \leq y_1^* \leq ... \leq y_n^*$. Quelle est la région critique de ce test ?

4. (1pts) Déterminer le seuil du test de Kolmogorov en fonction du risque de première espèce α et de l'inverse de la fonction de répartition de la loi de Kolmogorov.

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 $\mathcal{N}\left(m, \sigma^2\right)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	m	σ^2	$e^{imt - \frac{\sigma^2 t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\boldsymbol{m}, \boldsymbol{\Sigma})$	$f(x) = K e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{m})}$ $K = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}}$ $x \in \mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T\boldsymbol{m} - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u}}$
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 $\quad \sigma^2$: variance $\quad$ **F. C. :** fonction caractéristique

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

LOI	Probabilités	$\mathbf{m}$	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	$p_1 = P\left[X=1\right] = p$ $p_0 = P\left[X=0\right] = q$ $p \in [0,1] \quad q = 1-p$	p	pq	$pe^{it} + q$
Binomiale $B\left(n,p\right)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0,1] \quad q = 1-p$ $k \in \{0, 1, \dots, n\}$	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 $P\left(\lambda\right)$	$p_k = e^{-\lambda}\dfrac{\lambda^k}{k!}$ $\lambda > 0 \quad k \in \mathbb{N}$	λ	λ	$\exp\left[\lambda\left(e^{it}-1\right)\right]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0,1] \quad q = 1-p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1-qe^{it}}$