LSTAT2040: Examen

Durée 3 heures

Les feuilles blanches sont vos feuilles de réponses. Inscrivez vos nom et prénom sur chacune d'entre elles et numérotez-les dans l'ordre de lecture. Seules ces feuilles seront corrigées. Les feuilles de couleur vous serviront de brouillon.
À la fin de l'examen, vous devez rendre toutes les feuilles que vous avez reçues, y compris les feuilles de brouillon et les feuilles vierges que vous n'avez pas utilisées.
Vous pouvez consulter le syllabus du cours.
Tout échange d'informations, sous quelque forme que ce soit, est interdit et sera considéré comme une tricherie.

Excercice 1 (16 pts)

$X_1, \ldots, X_n$, et $Y_1, \ldots, Y_n$ sont des variables indépendantes de lois exponentielles (Exp). La densité de X_i est $f_i(x) = \lambda_i \theta \exp(-\lambda_i \theta x) I(x > 0)$, et la densité de Y_i est $g_i(y) = \lambda_i \exp(-\lambda_i y) I(y > 0)$. $\lambda_1, \ldots, \lambda_n$ et θ sont des paramètres positifs.

Dans tout ce qui suit, *sauf indication contraire*, nous supposons que $\lambda_1, \ldots, \lambda_n$ et θ sont tous inconnus et que l'échantillon que nous observons est (X_i, Y_i), $i = 1, \ldots, n$.

Pour information, la fonction de répartition (cdf) d'une $Exp(\lambda)$ est $(1 - \exp(-\lambda x)) I(x > 0)$. Sa moyenne est de $1/\lambda$ et sa variance de $1/\lambda^2$.

1 (1 pt)

Le modèle serait-il identifiable si nous observons seulement les X_i mais pas les Y_i ? Justifiez.

2 (1 pt)

Supposons que les λ_i soient tous connus. Trouver l'estimateur du maximum de vraisemblance (EMV) de θ basé sur les X_i.

3 (2 pts)

Montrez que $\lambda_i X_i \sim Exp(\theta)$. Utilisez ce résultat pour trouver la distribution asymptotique de l'estimateur décrit en (2).

4 (2 pts)

Montrez que $\hat{\theta}$, l'EMV de θ, basée sur (X_i, Y_i), découle de l'équation suivante

$$\frac{n}{\hat{\theta}} = 2\sum_{i=1}^{n} \frac{R_i}{1+\hat{\theta}R_i},$$

où $R_i = X_i/Y_i$. Quel est l'EMV de λ_i ?

5 (2 pts)

Montrez que la cdf de R_i est donnée par $F_R(r) = 1 - (1+\theta r)^{-1}$, pour $r > 0$.
Montrez que l'EMV de θ basé sur $R_1, \ldots, R_n$ est la même que celui défini ci-dessus.

6 (2 pts)

Montrez que l'information de Fisher contenue dans $R_1, \ldots, R_n$ à propos de θ est donnée par

$$\frac{n}{3\theta^2}.$$

Indice : Pensez à utiliser une intégration par changement de variable de type $u = \theta r + 1$. Aussi, sachez que $\int_1^\infty \frac{(1-u)^2}{u^4} du = 1/3$.

7 (2 pts)

Quelle est la distribution asymptotique de $\hat{\theta}$?
Proposez, si possible, une fonction k pour laquelle $\sqrt{n}(k(\hat{\theta}) - k(\theta)) \xrightarrow{d} N(0,1)$. Si vous estimez que c'est impossible, veuillez expliquer pourquoi.

8 (1 pt)

Nous voulons utiliser l'algorithme de Newton-Raphson (NR) pour calculer $\hat{\theta}$. Proposer une valeur (estimateur) de départ appropriée. Justifiez votre réponse.
Décrivez le déroulement de l'algorithme NR et préciser les calculs successifs à réaliser pour parvenir à la solution.

9 (1 pt)

Nous avons relevé les 20 observations suivantes.

x	0.7	11.3	2.1	30.7	4.6	20.2	0.3	0.9	0.7	2.3	1.1	1.9	0.5	0.8	1.2	15.2	0.2	0.7	0.4	2.3
y	3.8	4.6	2.1	5.6	10.3	2.8	1.9	1.4	0.4	0.9	2.8	3.2	8.5	14.5	14.4	8.8	7.6	1.3	2.2	4.0

Ces observations ont été utilisées pour calculer $\hat{\theta}$ via l'algorithme NR. Le tableau suivant présente les itérations successives, ainsi que quelques détails relatifs aux calculs effectués (dans ces formules $r_i = x_i/y_i$).

k	θ_k	$\sum_i r_i/(\theta_k r_i + 1)$	$\sum_i r_i^2/(\theta_k r_i + 1)^2$
0	2.333333	4.502554	1.307859
1		5.121220	1.743069
2		4.983727	1.639954
3	2.010147	4.974742	1.633345
4	2.010169	4.974705	1.633318
5	2.010169	4.974705	1.633318

Utilisez ces information pour compléter ce tableau en calculant θ_1 et θ_2, i.e. les valeurs de θ à l'itération 1 et 2, respectivement, de l'algorithme NR.

8 (2 pts)

Proposer deux estimations de l'écart-type asymptotique de $\hat{\theta}_{20}$, l'EMV de θ basée sur l'échantillon observé.

Excercice 2 (4 pts)

Commençons par définir la loi Gamma et par donner quelques informations à son sujet.

X suit une loi Gamma de paramètres k et $\theta > 0$, i.e. $X \sim \text{Gamma}(k, \theta)$, si sa densité est donnée par

$$f_X(x) = \frac{x^{k-1}\exp(-x/\theta)}{\Gamma(k)\theta^k} \mathrm{I}(0 \leq x < \infty),\ x \in \mathbb{R},$$

où Γ est la fonction gamma $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-x)\,dx$, $\alpha \in \mathbb{R}$. Cette fonction possède un certain nombre de propriétés, notamment le fait que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, $\forall \alpha > 0$ et $\Gamma(n+1) = n!$, $\forall n \in \mathbb{N}$. Aussi, sachez que $E(X^r) = \frac{\Gamma(k+r)}{\Gamma(k)}\theta^r$, et que $\text{Gamma}(k = n/2, \theta = 2) \equiv \chi^2_n$.

Soit $Y_1, \ldots, Y_n$ un échantillon iid de taille n $(n \geq 2)$ d'une $N(0, \sigma^2)$.

1 (1 pt)

Soit $U = \sum_{i=1}^n Y_i^2$. Montrez que $U \sim \text{Gamma}(n/2, 2\sigma^2)$? Expliquez vos calculs de manière détaillée.

2 (1 pt)

Développez un *estimateur sans biais* pour $\theta = \sigma^r$ (r est un entier positif connu), sous la forme $\hat{\theta} = aU^b$, où a et b doivent être déterminés explicitement.

3 (2 pts)

Dérivez la CRLB (Cramer-Rao Lower bound) pour $\theta = \sigma^r$. Pour $r = 2$, montrez que $\hat{\theta}$ est efficace pour θ.