MST2 : Devoir Maison

26 mars 2020

Nicolas Brunel, Nawal Belaid, Salim Amoukou

Consigne : Le DM se fait seul, et donnera lieu à un document au format pdf que vous déposerez sur le serveur exam.ensiie.fr, MST2-DM. Il faut le rendre pour **le 12 avril au plus tard**.
Vous devez rédiger avec soin vos réponses, y compris les calculs nécessaires (l'usage de Latex pour la rédaction de votre devoir est en cela presque nécessaire).
La partie couverte du programme couverte par le DM concerne l'intégralité des parties I (modélisation statistique) et II (estimation ponctuelle et région de confiance).

Exercice 1

Soit X une variable aléatoire de densité sur $\mathbb{R}$:

$$f(x) = \begin{cases} \frac{1}{2\theta\sqrt{x}} \exp(\frac{-\sqrt{x}}{\theta}) & x > 0 \\ 0 & x \leq 0 \end{cases}$$

où θ est un paramètre inconnu, strictement positif.
Nous pouvons obtenir de l'information sur θ, grâce à un échantillon i.i.d $(X_1, \ldots, X_n)$.

1. Trouver la loi de la variable aléatoire $Y = \frac{2}{\theta}\sqrt{X}$.
2. Donner une statistique exhaustive minimale pour l'échantillon $(X_1, \ldots, X_n)$.
3. Trouver l'Estimateur du Maximum de Vraisemblance de θ (calculé avec l'échantillon $(X_1, \ldots, X_n)$). Cet estimateur est-il efficace ?
4. Déduire un intervalle de confiance asymptotique pour θ avec un niveau de confiance $1 - \alpha$.

Exercice 2

Partie 1

La durée de vie d'une pièce industrielle est couramment modélisée par une variable aléatoire X ayant une loi Exponentielle $\mathcal{E}(a)$, de paramètre inconnu $a > 0$. Pour mémoire, la densité est

$$f(x) = \begin{cases} a \exp(-ax) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

1. Trouver l'Estimateur de Max de Vraisemblance (EMV) de a, pour un échantillon i.i.d de taille n.
2. Calculer l'information de Fisher apportée sur a par un échantillon de taille n.

Partie 2

Un central téléphonique comporte n lignes numériques numérotées de 1 à n. On suppose que ce central, initialement en panne, est remis en marche à un instant τ inconnu. On note T_i le premier instant postérieur à τ ou la ligne i commence à être occupée. On suppose que les variables aléatoires $X_i = T_i - \tau$ $(i = 1, ..., n)$ sont exponentielles de même paramètre connu a et indépendantes.

1. A partir de l'échantillon $(T_1, T_2, ..., T_n)$, donnez l'EMV de τ, que l'on notera W.
2. Calculer l'espérance de W et en déduire un estimateur sans biais de τ que l'on notera $\hat{\tau}_1$.

3. Calculer la variance de $\hat{\tau}_1$.
4. Calculer l'espérance mathématique de $\overline{T} = \frac{1}{n}\sum_{i=1}^n T_i$, et en déduire un estimateur $\hat{\tau}_2$ sans biais de τ.
5. Calculer la variance de $\hat{\tau}_2$.
6. De $\hat{\tau}_1$ et de $\hat{\tau}_2$, quel est l'estimateur de τ qui vous semble préférable ?

Problème

Le problème suivant va s'intéresser à la construction d'intervalle de confiance et d'estimation de quantile. Pour cela, nous allons avoir besoin d'abord montrer qu'un quantile empirique s'estime "bien" à partir d'un échantillon, avec une précision que l'on peut calculer [1]. La partie 1 consiste à utiliser une technique de type "méthode-delta" pour déterminer la loi d'un quantile empirique. Dans la partie 2, nous nous intéressons à l'estimation d'un paramètre, de plusieurs manières différentes, et à la construction d'un intervalle de confiance. Les 2 parties peuvent être indépendamment.

Partie 1 : Normalité asymptotique du quantile empirique

Soit $(X_1, ..., X_n)$ un échantillon i.i.d dont la fonction de répartition est notée $F(x) = P(X \leq x)$. Pour $\alpha \in]0,1[$, nous notons q_α le quantile d'ordre α de F (i.e. par définition $F(q_\alpha) = \alpha$).
On note

$$x \mapsto F_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{]-\infty,x]}(X_i)$$

la fonction de répartion empirique et $q_\alpha(n)$ le α-quantile empirique. Enfin, on note $F_n^{[-1]}$ l'inverse généralisée associée, si bien que par définition nous avons $F_n^{[-1]}(\alpha) = q_\alpha(n)$.
Pour rappel, l'inverse généralisée d'une fonction croissante F, est la fonction définie par

$$F^{[-1]}(\alpha) = \inf\{x \in \mathbb{R} | F(x) > \alpha\}.$$

L'essentiel des difficultés arrive quand la fonction n'est pas strictement croissante, mais possède des plateaux, ce qui est exactement le cas de la fonction de répartition empirique F_n...

Dans cette partie, nous allons montrer que si q_α est un quantile d'ordre $\alpha \in]0,1[$ fixé (de la fonction de répartition F), alors si F est dérivable en q_α, de dérivée $f(q_\alpha) > 0$ alors :

$$\sqrt{n}(q_\alpha(n) - q_\alpha) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{\alpha(1-\alpha)}{f(q_\alpha)^2})$$

1. En utilisant le fait que les sauts de la fonction de répartition empirique sont d'amplitude au moins $\frac{1}{n}$, montrer

$$\mathbb{P}(\sqrt{n}(q_\alpha(n) - q_\alpha) \leq x) = \mathbb{P}(F_n^{[-1]}(\alpha) \leq q_\alpha + \frac{x}{\sqrt{n}}) = \mathbb{P}(F_n(q_\alpha + \frac{x}{\sqrt{n}}) > \frac{\lceil n\alpha \rceil - 1}{n})$$

où $\lceil n\alpha \rceil$ désigne la partie entière supérieure de $n\alpha$.

2. Nous introduisons maintenant la variable aléatoire $Y_n = \sqrt{n}(F_n(q_\alpha + \frac{x}{\sqrt{n}}) - F(q_\alpha + \frac{x}{\sqrt{n}}))$, de fonction de répartition G_n.

(a) Pour un y_n que l'on précisera, montrer que

$$\mathbb{P}(\sqrt{n}(q_\alpha(n) - q_\alpha) \leq x) = 1 - G_n(y_n)$$

(b) En utilisant un développement limité, montrer que

$$y_n \xrightarrow[n\to+\infty]{} -f(q_\alpha)x$$

1. pour estimer une espérance avec une moyenne empirique, c'est assez facile, notamment grâce au théorème central limite. Par contre, les quantiles sont définis par un problème d'ordonnancement de l'échantillon...or les problèmes de classement / ordre sont des notions très "discrètes", il vaut mieux donc y regarder à deux fois

3. En posant $Y_n = Z_n + (Y_n - Z_n)$ avec $Z_n = \sqrt{n}(F_n(q_\alpha) - F(q_\alpha))$
   (a) Montrer que $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \alpha(1-\alpha))$ et $Y_n - Z_n \xrightarrow{P} 0$
   (b) En déduire à l'aide du lemme de Slutsky que $Y_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \alpha(1-\alpha))$.
   Tips pour (a) : TCL et trouver la loi de $|Y_n - Z_n|$
4. Conclure en montrant que $\mathbb{P}(\sqrt{n}(q_\alpha(n) - q_\alpha) \leq x) \xrightarrow[n \to +\infty]{} \boldsymbol{\phi}(\frac{f(q_\alpha)}{\sqrt{\alpha(1-\alpha)}} x)$ avec $\boldsymbol{\phi}$ la fonction de répartition d'une loi normale centrée réduite.
5. Question bonus (**résultat admis pour la suite**) : Montrer que si F est strictement croissant en q_α, alors $q_\alpha(n) \xrightarrow{ps} q_\alpha$

Partie 2 : Comparaison d'estimateurs

Soit $\theta > 0$, un paramètre inconnu. On considère la densité f_θ pa rapport à la mesure de Lesbesgue sur $\mathbb{R}$ définie par

$$f_\theta(x) = \frac{1}{2\theta}(\mathbf{1}_{[0,\theta]}(x) + \mathbf{1}_{[2\theta,3\theta]}(x))$$

Dans toute la suite, on suppose que nous avons un échantillon $X_1, \ldots, X_n$ i.i.d, de densité f_θ.

1. Calculer $\mathbb{E}(X_1)$ et $Var(X_1)$. Déduire de la méthode des moments un estimateur $\hat{\theta}_{M,n}$ de θ. Préciser la loi limite (après une renormalisation convenable) de $\hat{\theta}_{M,n} - \theta$
2. Représenter f_θ. Calculer la fonction de répartition F_θ de X_1 et la représenter.
3. Que vaut le premier quartile $q_{1/4} = F_\theta^{-1}(1/4)$? En déduire un estimateur $\hat{\theta}_{q,n}$ consistant pour θ et préciser la loi limite (après une renormalisation convenable) de $\hat{\theta}_{q,n} - \theta$
4. Que vaut le troisième quartile $q_{3/4} = F_\theta^{-1}(3/4)$? En déduire un estimateur $\hat{\theta}_{Q,n}$ consistant pour θ et préciser la loi limite (après une renormalisation convenable) de $\hat{\theta}_{Q,n} - \theta$
5. Parmi $\hat{\theta}_{M,n}, \hat{\theta}_{q,n}, \hat{\theta}_{Q,n}$, quel estimateur choisiriez-vous asymptotiquement ?
6. Soit $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Pour $x \in]0, \theta[$, que vaut la probabilité $\mathbb{P}(X_{(n)} \leq 3\theta - x)$? En déduire un estimateur $\hat{\theta}_{S,n}$ consistant pour θ.
7. Montrer que $\mathbb{P}(\hat{\theta}_{S,n} \leq \theta - \frac{t}{n})$ converge, pour tout $t \geq 0$, vers une limite non nulle que l'on calculera. Préciser la loi limite de $n(\theta - \hat{\theta}_{S,n})$
8. Si vous avez un grand nombre d'observations, lequel des estimateurs précédents choisir ?
9. Que vaut la médiane $q_{1/2} = F_\theta^{-1}(1/2)$? Que dire de la médiane empirique comme estimateur de θ.
10. On veut trouver un intervalle de confiance bilatère non asymptotique pour θ au niveau $(1-\alpha)$. Proposer une méthode.