Université A/Mira de Béjaia
Faculté des Sciences Exactes
Département de Mathématiques
$2^{\text{ème}}$ année Licence STID.

2019/2020

Série N°2 - Statistique Inférentielle

Exercice 1

On considère les modèles suivants :

- modèle de Poisson $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$;
- modèle de la loi de exponentielle $(\mathbb{R}^+, \mathcal{B}_{R^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$;
- modèle gaussien avec σ^2 positif connu : $(\mathbb{R}, \mathcal{B}_R, \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}\})$;
- modèle gaussien avec μ dans $\mathbb{R}$ connu : $(\mathbb{R}, \mathcal{B}_R, \{\mathcal{N}(\mu, \sigma^2) : \sigma^2 > 0\})$;
- modèle gaussien : $(\mathbb{R}, \mathcal{B}_\mathbb{R}, \{\mathcal{N}(\mu, \sigma^2)\} : \mu \in \mathbb{R}, \sigma^2 > 0)$.

1. Pour chacun de ces modèles donner l'expression d'une statistique exhaustive (éventuellement vectorielle).
2. Retrouver le résultat pour le modèle de Poisson en utilisant une autre méthode.

Exercice 2

On considère une famille exponentielle générale de statistique canonique $T(X)$ où X est la variable générique dans ce modèle.

1. Montrer que $\sum_{i=1}^n T(X_i)$ est une statistique exhaustive pour le modèle d'échantillonnage associé.
2. Montrer que la moyenne empirique $\overline{X_n}$ est une statistique exhaustive dans un modèle d'échantillonnage de la loi Binomiale.

Exercice 3

On considère le modèle de la loi uniforme $(\mathbb{R}^+, \mathcal{B}_{R^+}, \{\mathcal{U}[0, \theta] : \theta > 0\})$ et un échantillon $X_1, ..., X_n$ dans ce modèle. On se propose d'améliorer, si possible, l'estimateur $\theta = X_{(1)} + X_{(n)}$ avec $X_{(1)} = \inf\limits_{1 \leq i \leq n} X_i$ et $X_{(n)} = \sup\limits_{1 \leq i \leq n} X_i$.

1. Donner une statistique exhaustive dans ce modèle pour le paramètre θ.
2. Calculer la densité de la loi de $X_{(1)}$ conditionnelle à $\{X_{(n)} = x_n\}$. En déduire l'expression de $\mathbb{E}\left(X_{(1)} | X_{(n)} = x_n\right)$.
3. Déterminer alors $\check{\theta}$, estimateur amélioré de θ par le théorème de Rao-Blackwell. La statistique $X_{(n)}$ est-elle complète ? Conclure.

Exercice 4

1. On suppose qu'une v.a.r. X suit une loi exponentielle $\mathcal{E}(\lambda)$. Calculer $\mathbb{E}(X^2)$.
2. Soit $t_0 > 0$. Écrire $F(t_0) = \mathbb{P}(X > t_0)$ sous forme d'une espérance.

3. On considère le modèle de la loi exponentielle $(\mathbb{R}^+; \mathcal{B}_R^+; \{\mathcal{E}(\lambda) : \lambda > 0\})$. En vous inspirant des résultats des deux questions précédentes et en utilisant à chaque fois la méthode des moments, proposer deux autres estimateurs du paramètre λ.

Exercice 5
On considère le modèle de la loi uniforme $(\mathbb{R}^+, \mathcal{B}_{R^+}, \{\mathcal{U}[0,\theta] : \theta > 0\})$ et un échantillon $X_1, ..., X_n$ dans ce modèle. On pose $X_{(1)} = \inf_{1\leq i\leq n} X_i$ et $X_{(n)} = \sup_{1\leq i\leq n} X_i$ (la première et la dernière statistique d'ordre). On considère les estimateurs suivants :

$$\begin{aligned}\hat{\theta}_1 &= X_{(n)}\\ \hat{\theta}_2 &= \frac{n+1}{n}X_{(n)}\\ \hat{\theta}_3 &= X_{(1)} + X_{(n)}\\ \hat{\theta}_4 &= 2\overline{X}_{(n)}.\end{aligned}$$

1. Pour chacun d'entre eux, étudier la consistance, le biais et donner l'expression de son risque quadratique.
2. Comparer les fonctions de risque quadratique. Qu'en conclure ?

Exercice 6
On considère le Modèle Statistique de la loi Gamma $(\mathbb{R}^+; \mathcal{B}_R^+; \{G(\alpha;\beta) : \alpha > 0; \beta > 0\})$. On rappelle que la densité d'une v.a. X de loi $G(\alpha;\beta)$ est :

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}1_{\mathbb{R}^+}(x).$$

1. Calculer $\mathbb{E}(X)$ et $Var(X)$.
2. Par la méthode des moments, donner un estimateur du paramètre bidimensionnel (α, β) du modèle, basé sur l'observation d'un échnatillon $X_1, ..., X_n$.
3. Déterminer des estimateurs de α et β en utilisant conjointement des estimateurs empiriques des moments.

Exercice 7 Un agriculteur possède un champ carré dont il veut estimer la superficie. Quand il mesure un côté de son champ, il suppose, que l'erreur expérimentale de la mesure est une variable aléatoire de loi normale centrée et de variance σ^2. Il réalise une première mesure de ce côté et trouve une valeur $x_1 = 510$ mètres. Il en déduit une superficie de $s_1 = 26.01$ hectares. Il réalise une deuxième mesure et trouve alors $x_2 = 490$, d'où une valeur de la superficie $s_2 = 24.01$. Il abandonne ses mesures et réfléchit pour savoir quelle est la bonne façon de procéder. Doit-il adopter comme estimation de la surface s_1, s_2, ou une estimation combinant les deux mesures, telle que

$$\begin{aligned}s_3 &= x_1x_2 = 24.99;\\ s_4 &= \frac{s_1+s_2}{2} = 25.01;\\ s_5 &= \left(\frac{x_1+x_2}{2}\right) = 25.\end{aligned}$$

Faut-il recommencer ses mesures jusqu'à ce qu'il trouve deux résultats identiques, ou bien combiner intelligemment n mesures pour construire des estimations du type s_4 ou s_5 (généralisées à ces n mesures) ?

1. On se propose d'aider l'agriculteur à résoudre son problème. Préciser le modèle considéré ainsi que la fonction $q(\theta)$ que l'on cherche à estimer. Étudier les estimateurs proposés. On calculera notamment leurs biais, variances et risques quadratiques moyens. (Ind. si $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ alors $Var(X^2) = 2(\sigma^4 + 2m^2\sigma^2)$. A l'aide de ces calculs, aider l'agriculteur à choisir l'estimateur qui vous semble préférable aux autres.
2. Donner les estimateurs qui généralisent s_4 et s_5 au cas où l'agriculteur a pu faire n mesures du coté de son champ. Effectuer la même étude qu'à la question 1 pour ces estimateurs.
3. Donner l'estimateur du maximum de vraisemblance et l'étudier s'il est différent de ceux considérés précédemment.

Exercice 8

Un matériel a une durée de vie modélisée par une v.a. X de fonction de répartition F. Un étudiant en Licence de Mathématiques sait qu'il devra l'utiliser pendant un temps x_0. Il souhaite naturellement qu'il n'y ait pas de panne durant cette période. Cet étudiant, ayant suivi le module de Statistique Inférentielle, cherche en premier lieu à estimer la loi (en fait la f.d.r.) de cette durée de vie, c'est à dire estimer F(x) pour tout x de $\mathbb{R}^+$. Il a alors l'idée de faire fonctionner, sur banc d'essai, n machines identiques à celle qu'il utilisera dans l'avenir. Il note $x_1, ..., x_n$ les n temps de panne observés, qui sont donc les réalisations des v.a. $X_1, ..., X_n$ i.i.d. de même loi que X.

1. Par la méthode des moments il propose un estimateur de $F(x)$, pour tout x dans $\mathbb{R}^+$. Pouvez-vous en faire autant ?
2. Son estimateur est-il consistant ? Que dire de son biais et de son risque quadratique ?
3. Se souvenant de ses cours, il sait que, pour être précis, il aurait dû, au préalable, introduire un modèle paramétrique. Quel(s) modèle(s) pourrait-il proposer ? Que sont les observations sous ce(s) modèle(s) ? Une estimation par maximum de vraisemblance nous donnerait-elle quelque chose de différent dans ce modèle ?
4. Que dire alors de l'efficacité de l'estimateur proposé dans la première question ?

Exercice 9

On s'intéresse à la durée de vie X d'un matériel électronique. Il est raisonnable de considérer que cette durée de vie est aléatoire et que sa loi soit exponentielle. En revanche, on ignore la valeur du paramètre λ de cette loi.

1. Écrire le modèle statistique engendré par X. Donner également le modèle d'échantillonnage associé.
2. Donner l'estimateur du maximum de vraisemblance pour une observation $x_1, ..., x_n$ d'un échantillon $X_1, ..., X_n$ de durées de vie de ces matériels.
3. Donner une estimation par maximum de vraisemblance de la quantité $\alpha = \mathbb{P}(X > t_0)$, où t_0 est un temps fixé.
4. Quels estimateurs de λ et de α obtient-on si on utilise la méthode des moments ?

On suppose que l'on réalise n fois cette expérience et on note $K_1, ..., K_n$ les nombres de pannes observées dans chaque intervalle $[0, t_0]$.

1. Donner le modèle statistique associé à ces observations.
2. Donner par la méthode du maximum de vraisemblance un autre estimateur du paramètre λ, basé cette fois sur les observations $K_1, ..., K_n$.
3. Qu'obtient-on comme estimateur de λ si, dans ce modèle, on utilise la méthode des moments ?

Exercice 10

On considère deux modèles :

- celui de la loi de Poisson $(\mathbb{N}; \mathcal{P}(\mathbb{N}); \{\mathcal{P}(\lambda) : \lambda > 0\})$, où $\mathcal{P}(\lambda)$ désigne la loi de Poisson de paramètre λ ;
- celui de la loi de exponentielle $(\mathbb{R}^+; \mathcal{B}_R^+; \{\mathcal{E}(\lambda) : \lambda > 0\})$, où $\mathcal{E}(\lambda)$ désigne la loi exponentielle de paramètre λ.

Pour chacun de ces modèles, répondre à l'ensemble des questions suivantes. On considérera à chaque fois l'observation d'un échantillon $X_1, ..., X_n$.

1. Rappeler l'expression de l'estimateur du maximum de vraisemblance dans ce modèle.
2. Étudier la consistance, le biais et le risque quadratique de cet estimateur.
3. Si cet estimateur est biaisé, est-il asymptotiquement sans biais ? Donner, si nécessaire, un estimateur sans biais. L'estimateur sans biais (l'initial ou le second) est-il efficace ? Est-il consistant ?