Statistique Inférentielle
Devoir Maison

Les questions ayant des numéros en vert sont faciles. De plus, une telle question peut être traité même si vous n'avez pas réussi toutes les questions précédentes (bien que vous pouvez avoir besoin des résultats qui y sont énoncés). Notons que ces questions rapportent 12 points à elles seules.

Loi de Pareto

Soit $X_1, \dots, X_n$ un échantillon de variables aléatoires d'une loi de Pareto $\mathscr{P}(\lambda, \theta)$, où $\lambda, \theta \in \mathbb{R}_+^*$, ayant pour densité de probabilité la fonction

$$f(x) = \frac{\lambda\,\theta^\lambda}{x^{\lambda+1}}\,\mathbb{1}_{\{x>\theta\}}, \qquad x \in \mathbb{R}.$$

Quelques propriétés de la loi de Pareto

1. Calculer la fonction de répartition de la loi $\mathscr{P}(\lambda, \theta)$ et en déduire que, pour tout $\varepsilon \in \,]0,1[$, le quantile d'ordre ε de cette loi est $q_\varepsilon = \frac{\theta}{(1-\varepsilon)^{1/\lambda}}$.
2. Soit $p > 0$. Calculer le moment d'ordre p de la loi $\mathscr{P}(\lambda, \theta)$ en précisant sous quelle condition sur λ ce moment existe. En déduire que $\mathbf{E}\,X_1 = \frac{\lambda}{\lambda-1}\,\theta$ et $\mathbf{Var}\,X_1 = \frac{\lambda}{(\lambda-1)^2(\lambda-2)}\,\theta^2$ (lorsqu'elles existent).
3. Montrer que $Y = \ln\big(\frac{X_1}{\theta}\big) \hookrightarrow \mathscr{E}(\lambda)$ et, par conséquent $S_n = \sum_{i=1}^n \ln\big(\frac{X_i}{\theta}\big) \hookrightarrow \Gamma(n, \lambda)$.
   Indication : *il suffit de calculer la fonction de répartition de la variable aléatoire Y.*
4. Montrer que $X_{(1)} = \min\{X_1, \dots, X_n\} \hookrightarrow \mathscr{P}(\lambda n, \theta)$.
   Indication : *il suffit de calculer la fonction de répartition de la variable aléatoire $X_{(1)}$.*

Inférence sur le paramètre λ

Dans cette partie on suppose que $\theta > 0$ est connu et le paramètre inconnu est $\lambda \in \mathbb{R}_+^*$.

5. Calculer la log-vraisemblance. En déduire que l'EMV de λ est $\widehat{\lambda}_n = \frac{n}{S_n}$.
6. Montrer que $\widehat{\lambda}_n$ est fortement consistant et asymptotiquement normal. Préciser sa variance limite.
7. Construire (en supposant que $\lambda > 1$) un estimateur λ_n^* de λ à l'aide de la méthode des moments en utilisant le moment d'ordre 1.
8. Montrer (en supposant que $\lambda > 2$) que λ_n^* est fortement consistant et asymptotiquement normal. Préciser sa variance limite.
9. Lequel des estimateurs $\widehat{\lambda}_n$ et λ_n^* converge plus vite ?

T.S.V.P.

10. Calculer l'information de Fisher apportée par l'échantillon $X_1, \dots, X_n$ sur le paramètre λ et énoncer proprement la borne de Cramér-Rao.

11. Calculer (en supposant que $n > 1$) l'espérance de l'estimateur $\widehat{\lambda}_n$. En déduire le biais de $\widehat{\lambda}_n$, ainsi qu'un estimateur sans biais $\widetilde{\lambda}_n$ de λ.
Indication : *la formule donnant le moment m_p d'ordre p de la loi Gamma (page 2 de la brochure des lois et des tables) est valable également pour les moments d'ordres négatifs (lorsqu'ils existent) ; on peut donc l'utiliser pour le moment d'ordre $p = -1$.*

12. Sans calculs supplémentaires, conclure que l'estimateur $\widehat{\lambda}_n$ n'est pas efficace (au sens de Cramèr-Rao). D'ailleurs, existe-t-il un estimateur efficace dans ce problème ?

13. Calculer la variance de l'estimateur $\widetilde{\lambda}_n$. Quelle propriété de $\widetilde{\lambda}_n$ peut-on en déduire ? Comparer la variance avec la borne de Cramér-Rao et discuter.
Indication : *on peut utiliser la même formule que dans la question 11, mais pour $p = -2$.*

On souhaite maintenant tester l'hypothèse "$\lambda = \lambda_0$" contre l'hypothèse "$\lambda > \lambda_0$", où $\lambda_0 > 0$. Comme statistique de test on choisit d'utiliser l'estimateur $\widehat{\lambda}_n$ et, pour n grand, on approche sa loi par la loi $\mathcal{N}(\lambda, {}^{\lambda^2}\!/_{n})$ (grâce à la question 3).

14. Choisir (en justifiant votre choix) la forme de la zone de rejet parmi $]-\infty, k]$ et $[k, +\infty[$ et déterminer $\big($en fonction de $\alpha \in]0,1[\big)$ la valeur $k = k(\alpha)$ qui garantie que le test soit de niveau α.

15. Exprimer $\big($en fonction de $\lambda \in]\lambda_0, +\infty[\big)$ la puissance $\gamma = \gamma(\lambda)$ de ce test à l'aide de la fonction de répartition Φ de la loi normale centrée réduite.

16. Calculer la limite de la puissance $\gamma(\lambda)$ lorsque $n \longrightarrow +\infty$ (pour $\lambda \in]\lambda_0, +\infty[$ fixe). Que peut-on en déduire ?

Inférence sur le paramètre θ

Dans cette partie on suppose que $\lambda > 0$ est connu et le paramètre inconnu est $\theta \in \mathbb{R}_+^*$.

17. Calculer la vraisemblance. En déduire que l'EMV de θ est $\widehat{\theta}_n = X_{(1)}$.

18. En étudiant la limite de la fonction de répartition de $\widehat{\theta}_n$, ainsi que celle de $n(\widehat{\theta}_n - \theta)$, montrer que l'estimateur $\widehat{\theta}_n$ est consistant et que sa vitesse de convergence est ${}^1\!/_n$. Préciser sa loi limite.

19. À votre avis, est-il intéressant de construire des estimateurs par la méthode des moments dans ce problème ? Pourquoi ?

20. Calculer (en précisant sous quelle condition sur n elle existe) l'espérance de l'estimateur $\widehat{\theta}_n$. En déduire le biais de $\widehat{\theta}_n$, ainsi qu'un estimateur sans biais $\widetilde{\theta}_n$ de θ.

21. Calculer (en précisant sous quelle condition sur n elles existent) les variances et les risques quadratiques des estimateurs $\widehat{\theta}_n$ et $\widetilde{\theta}_n$ et comparer ces estimateurs.

22. Est-il possible d'établir une borne de Cramér-Rao pour ce problème ? Pourquoi ?

23. Soit $p \geqslant 0$. Calculer (en précisant sous quelle condition sur n il existe) l'estimateur pseudo-bayésien de θ avec la pseudo-densité $q(t) = \frac{1}{t^p} \mathbb{1}_{\{t>0\}}$. Pour quelle valeur de p retrouve-t-on l'estimateur $\widetilde{\theta}_n$?

24. En se basant sur $\widehat{\theta}_n$, construire un intervalle de confiance de risque $\alpha \in]0,1[$ pour θ (ne pas oublier de tracer l'allure de la courbe de densité pour choisir correctement la forme de l'intervalle).