Statistique Inférentielle Avancée

Durée : 3 heures.

Documents autorisés : notes manuscrites.

Les résultats vus en cours ou en TD peuvent être utilisés sans être redémontrés.

Les deux parties sont indépendantes.

Barème indicatif - Partie 1 : 12 pts, Partie 2 : 8 pts.

Première partie

On considère un échantillon de taille n de la loi uniforme sur $[\theta, 2\theta]$, avec $\theta > 0$. On a vu en TD que (X_1^*, X_n^*) est une statistique exhaustive et que l'estimateur de maximum de vraisemblance de θ est $X_n^*/2$.

1. Montrer que les densités de X_1^* et X_n^* sont respectivement :
$$f_{X_1^*}(x) = \frac{n}{\theta^n}(2\theta - x)^{n-1}\mathbb{1}_{[\theta,2\theta]}(x) \quad \text{et} \quad f_{X_n^*}(x) = \frac{n}{\theta^n}(x - \theta)^{n-1}\mathbb{1}_{[\theta,2\theta]}(x).$$

2. Calculer la densité du couple (X_1^*, X_n^*).

3. Calculer l'espérance de la loi uniforme sur $[\theta, 2\theta]$. En déduire l'estimateur de θ par la méthode des moments. Montrer qu'il est sans biais.

4. Calculer $E[X_1^*]$. En déduire un estimateur sans biais de θ ne dépendant que de X_1^*.

5. Calculer $E[X_n^*]$. En déduire un estimateur sans biais de θ ne dépendant que de X_n^*. Vérifier que l'estimateur de maximum de vraisemblance est biaisé mais asymptotiquement sans biais.

6. Montrer que la statistique exhaustive n'est pas complète.

7. On cherche un estimateur de θ de la forme $a_nX_1^* + b_nX_n^*$. Calculer son espérance et en déduire une condition reliant a_n et b_n pour que cet estimateur soit sans biais.

8. Calculer cet estimateur dans le cas où $a_n = b_n$.

9. On admet que $Var(X_1^*) = Var(X_n^*) = \dfrac{n\theta^2}{(n+1)^2(n+2)}$ et que
$Cov(X_1^*, X_n^*) = \dfrac{\theta^2}{(n+1)^2(n+2)}$.
Calculer le meilleur estimateur sans biais de θ de la forme $a_nX_1^* + b_nX_n^*$.

Deuxième partie

Soient $X_1, \dots, X_n$ n variables aléatoires réelles indépendantes et de même loi, de variance $Var[X] = \mu_2$, de fonction de répartition F et telle que $E[X^4] < \infty$. On a vu en cours que :

$$\sqrt{n}\,\frac{S'^2_n - \mu_2}{\sqrt{\mu_4 - \mu_2^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

où $\mu_4 = E[(X - E[X])^4]$, ce qui permet de montrer qu'un intervalle de confiance asymptotique de seuil α pour μ_2 est :

$$\Big[S'^2_n - \frac{u_\alpha}{\sqrt{n}}\sqrt{\mu_4^e - S'^4_n}\ ,\ S'^2_n + \frac{u_\alpha}{\sqrt{n}}\sqrt{\mu_4^e - S'^4_n}\Big].$$

où $\mu_4^e = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x}_n)^4$.

Quand n est trop petit, il est possible que la borne inférieure de cet intervalle soit négative, ce qui pose problème puisque la quantité à estimer, la variance, est positive. Pour résoudre ce problème, on utilise la variante de la méthode delta suivante.

Si $\{Y_n\}_{n\geq 1}$ est une suite de variables aléatoires réelles telle que

$$\sqrt{n}\,(Y_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

alors pour toute fonction dérivable φ, on a

$$\sqrt{n}\ [\varphi(Y_n) - \varphi(\theta)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2\,\varphi'(\theta)^2).$$

1. Déterminez la loi asymptotique de $\sqrt{n}\ \big(\ln S'^2_n - \ln \mu_2\big)$.
2. Montrer que $\dfrac{S'^2_n}{\sqrt{\mu_4^e - S'^4_n}}$ converge en probabilité vers $\dfrac{\mu_2}{\sqrt{\mu_4 - \mu_2^2}}$.
3. En utilisant les 2 résultats précédents et le théorème de Slutsky, donnez un nouvel intervalle de confiance asymptotique de seuil α pour μ_2, dont la borne inférieure est toujours positive.
4. De même, un intervalle de confiance asymptotique de seuil α pour $F(x)$ est

   $$\Big[\mathbb{F}_n(x) - \frac{u_\alpha}{\sqrt{n}}\sqrt{\mathbb{F}_n(x)(1-\mathbb{F}_n(x))}, \mathbb{F}_n(x) + \frac{u_\alpha}{\sqrt{n}}\sqrt{\mathbb{F}_n(x)(1-\mathbb{F}_n(x))}\Big]$$

   Il peut arriver que la borne inférieure de cet intervalle soit négative et que la borne supérieure soit supérieure à 1.
   En utilisant la méthode delta avec la fonction logit $\varphi(p) = \ln \dfrac{p}{1-p}$, construire un nouvel intervalle de confiance asymptotique de seuil α pour $F(x)$. Montrer que les bornes de cet intervalle sont forcément comprises entre 0 et 1.