ENSIMAG $2^{\text{ème}}$ année Mai 2014

Statistique Inférentielle Avancée

Durée : 3 heures. Tous documents autorisés.

Les deux parties sont indépendantes.

Les résultats vus en cours ou en TD n'ont pas à être redémontrés.

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

Première partie

Soit $\theta \in \mathbb{N}^*$. Une variable aléatoire X est de loi uniforme (discrète) sur $\{1, \ldots, \theta\}$ si elle est à valeurs dans $\{1, \ldots, \theta\}$ et que

$$P(X = k) = \frac{1}{\theta} \mathbb{1}_{\{1,\ldots,\theta\}}(k).$$

1. Donner la fonction de répartition et l'espérance mathématique de X.

2. Soient $X_1, \ldots, X_n$ n variables aléatoires indépendantes et de même loi uniforme sur $\{1, \ldots, \theta\}$. Soit $T = X_n^*$. Donner la fonction de répartition de T et les probabilités élémentaires $P(T = t) = P(T \leq t) - P(T \leq t-1)$, $\forall t \in \{1, \ldots, \theta\}$.

3. Montrer que $\forall i \in \{1, \ldots, n\}$, $\forall t \in \{1, \ldots, \theta\}$, la loi de probabilité conditionnelle de X_i sachant $[T = t]$ est donnée par :

$$P(X_i = k | T = t) = \begin{cases} \dfrac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} & \text{si } k \in \{1, \ldots, t-1\} \\ \dfrac{t^{n-1}}{t^n - (t-1)^n} & \text{si } k = t \\ 0 & \text{sinon} \end{cases}$$

4. Un récipient contient un nombre inconnu θ d'objets identiques numérotés de 1 à θ, que l'on désire estimer. Pour cela, on effectue n tirages au hasard dans le récipient. Après chaque tirage, on note le numéro de l'objet obtenu et on remet celui-ci dans le récipient. Quel est le modèle statistique associé au résultat $x_1, \ldots, x_n$ de cette expérience ?

5. Ecrire la fonction de vraisemblance de ce modèle paramétrique et calculer l'estimateur de maximum de vraisemblance de θ. Montrer sans calculs qu'il est biaisé.

6. Montrer que la statistique de maximum de vraisemblance est exhaustive et complète.

7. Calculer l'estimateur de θ par la méthode des moments. Montrer qu'il est sans biais.

8. En déduire explicitement l'estimateur sans biais et de variance minimale de θ. Quel est son inconvénient et comment y remédier ?

9. Pour $n = 20$, estimer θ quand $t = 10$, $t = 40$ et $t = 100$.

Deuxième partie

Soient $X_1, \ldots, X_n$ 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 :

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

ce qui permet de montrer qu'un intervalle de confiance asymptotique de seuil α pour μ_2 est :

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

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

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

1. Déterminez la loi asymptotique de $\sqrt{n}\,\left(\ln S'^2_n - \ln \mu_2\right)$.

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

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

   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.