Statistiques

Décembre 2017, Sans document, 1h30

Préambule : Le sujet est composé de quatre exercices indépendants. La qualité de la rédaction sera prise en compte. Toutes les réponses seront données sur la copie (ne pas rendre le sujet).

Exercice 1

1. Enoncer la loi forte des grands nombres (hypothèses et résultats).
2. Enoncer le théorème central limite (hypothèses et résultats).
3. Soit $X_1, \ldots, X_n$ n variables aléatoires réelles de loi de Poisson de paramètre $\lambda > 0$. Appliquer le théorème central limite à ce modèle.
4. Soit $X_1, \ldots, X_n$ n variables aléatoires réelles de loi uniforme sur $[\theta, 1+\theta]$ avec $\theta > 0$.
   (a) Donner la densité de X_1.
   (b) Calculer l'estimateur des moments de θ.
   (c) Calculer le biais de cet estimateur.
5. Soit $X_1, \ldots, X_n$ n variables aléatoires réelles indépendantes de loi $\mathcal{N}(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ inconnus.
   (a) Enoncer le théorème de Cochran.
   (b) En déduire un intervalle de confiance de niveau $1-\alpha$ avec $\alpha \in]0,1[$ pour μ. On prendra soin de justifier **toutes les étapes** de construction de cet intervalle de confiance à partir du théorème de Cochran.

Exercice 2

Soit $X_1, \ldots, X_n$ n variables aléatoires de loi de Bernoulli de paramètre $p \in]0,1[$.

1. Rappeler la définition de la fonction caractéristique d'une variable aléatoire réelle X.
2. Calculer la fonction caractéristique de X_1.
3. En déduire la fonction caractéristique d'une variable aléatoire Y de loi binomiale $\mathcal{B}(n,p)$.
4. Soit $(Y_n)_n$ une suite de variable aléatoire réelle de loi binomiale $\mathcal{B}(n, p_n)$. On suppose que $np_n \to \lambda > 0$ lorsque $n \to \infty$. On rappelle que la fonction caractéristique de la loi de Poisson de paramètre $\lambda > 0$ est donnée par
$$\varphi(t) = \exp(\lambda(\exp(it) - 1)).$$
   Montrer que Y_n converge en loi vers une variable aléatoire de loi de Poisson $\mathcal{P}(\lambda)$.

Exercice 3

Soit $X_1, \ldots, X_n$ un n échantillon composé de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi admettant pour densité

$$f(x) = \frac{1}{2\sqrt{x\theta}} \mathbf{1}_{]0,\theta]}(x)$$

où θ est un paramètre strictement positif que l'on cherche à estimer dans cet exercice. On considère l'estimateur $\widehat{\theta} = \max(X_1, \ldots, X_n)$.

1. $\widehat{\theta}$ est-il l'estimateur du maximum de vraisemblance ? Justifier.
2. Calculer la fonction de répartition de $\widehat{\theta}$ et en déduire la densité de $\widehat{\theta}$.
3. Calculer le biais de $\widehat{\theta}$ et la variance de $\widehat{\theta}$. En déduire son risque quadratique.
4. Montrer que $n(\theta - \widehat{\theta})$ converge en loi vers une loi à préciser (**indication :** on pourra calculer la fonction de répartition de $n(\theta - \widehat{\theta})$). On en déduira un intervalle de confiance asymptotique de niveau $1-\alpha$ pour θ.

Exercice 4

Soit $X_1, \dots, X_n$ n variables aléatoires réelles i.i.d de densité f inconnue. On fixe un point x dans $\mathbb{R}$ et on cherche à estimer f au point x. On suppose que f est bornée par B et on se donne le noyau $K : \mathbb{R} \to \mathbb{R}^+$ défini par

$$K(u) = \frac{1}{2}\mathbf{1}_{[-1,1]}(u).$$

On considère l'estimateur de noyau K et de fenêtre $h > 0$ défini par

$$\widehat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right).$$

Dans cet exercice, on s'interessera au biais de cet estimateur. On suppose que f est dérivable et que sa dérivée vérifie :

$$|f'(x) - f'(y)| \leq L|x - y|, \quad \forall x, y \in \mathbb{R}.$$

1. Calculer les 3 intégrales suivantes

$$\int K(u)\,\mathrm{d}u, \quad \int uK(u)\,\mathrm{d}u \quad \text{et} \quad \int u^2 K(u)\,\mathrm{d}u.$$

2. On admettra qu'il existe $\tau \in]0, 1[$ tel que $\forall u$

$$f(x + uh) = f(x) + (uh)f'(x + \tau uh).$$

   Montrer que le biais de $\widehat{f}(x)$ vérifie

$$b(\hat{f}(x)) = \int K(u)(uh)[f'(x + \tau uh) - f'(x)]\,\mathrm{d}u.$$

3. En déduire

$$|b(\hat{f}(x))| \leq Ch^\alpha$$

   où α est un entier positif à préciser et C est une constante réelle positive à préciser.