Informatique
TP 9

Estimation

Partie 1 : comparaison de plusieurs estimateurs

Exercice 1

Étant donné un n-échantillon $(X_1, \dots, X_n)$ i.i.d. de loi uniforme sur $[0, \theta]$, où $\theta > 0$ est un paramètre inconnu, on a montré en TD que

$$S_n = \frac{2}{n} \sum_{k=1}^{n} X_k \quad \text{et} \quad T_n = \frac{n+1}{n} \max(X_1, \dots, X_n)$$

sont deux estimateurs sans biais de θ.
Représenter la loi de ces estimateurs pour $\theta = 1$ et différentes valeurs de n afin d'en comparer les performances.

Partie 2 : niveau réel d'un intervalle de confiance

Le niveau réel d'un intervalle de confiance $[U_n, V_n]$ pour $g(\theta)$ est le réel

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta\big(U_n \leqslant g(\theta) \leqslant V_n\big).$$

On peut estimer ce niveau par la méthode de Monte-Carlo.

Exercice 2

Étant donnée une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires i.i.d. de loi $\mathcal{B}(p)$ où $p \in \,]0, 1[$ est un paramètre inconnu, on considère l'intervalle de confiance asymptotique

$$J_n = \left[\overline{X}_n - \frac{1{,}96}{2\sqrt{n}}, \overline{X}_n + \frac{1{,}96}{2\sqrt{n}}\right]$$

de p au niveau de confiance 95%.

1. Simuler 100 réalisations de l'intervalle J_{500} et les représenter verticalement au dessus des abscisses entières de 1 à 100 en faisant apparaître en bleu ceux qui contiennent p et en rouge ceux qui ne contiennent pas p.
   Pour représenter le segment d'extremités $A = (x_A, y_A)$ et $B = (x_B, y_B)$, on pourra utiliser l'instruction `plot([xA;xB],[yA;yB],opt)` où `opt='b-+'` (resp. `opt='r-+'`) pour un segment bleu (resp. rouge).
2. Déterminer le niveau réel de l'intervalle J_{500} par la méthode de Monte-Carlo.

Partie 3 : comparaison de différents intervalles de confiance

Étant donnée une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires i.i.d. de loi $\mathcal{B}(p)$ où $p \in \,]0, 1[$ est un paramètre inconnu, on a construit en classe les trois intervalles de confiance (asymptotiques pour les deux derniers) pour p au niveau de confiance $1 - \alpha$:

$$I_n = \left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}, \overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right], \quad J_n = \left[\overline{X}_n - \frac{z_{1-\alpha/2}}{2\sqrt{n}}, \overline{X}_n + \frac{z_{1-\alpha/2}}{2\sqrt{n}}\right]$$

$$\text{et} \quad \mathrm{K}_n = \left[\overline{\mathrm{X}}_n - z_{1-\alpha/2}\sqrt{\frac{\overline{\mathrm{X}}_n(1-\overline{\mathrm{X}}_n)}{n}}, \overline{\mathrm{X}}_n + z_{1-\alpha/2}\sqrt{\frac{\overline{\mathrm{X}}_n(1-\overline{\mathrm{X}}_n)}{n}}\right]$$

où $z_{1-\alpha/2}$ désigne le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$ et

$$\overline{\mathrm{X}}_n = \frac{1}{n}\sum_{k=1}^{n}\mathrm{X}_k,$$

dont on souhaite comparer les performances.

Exercice 3

Pour différentes valeurs de $p \in \left]0, \frac{1}{2}\right]$, de $\alpha \in \left[\frac{1}{2}, 1\right[$ et $n = 500$, simuler $\mathrm{N} = 10^3$ intervalles de confiance des trois types ci-dessus et comparer :

- le niveau de confiance réel de chaque intervalle ;
- la demi-largeur moyenne de chaque intervalle.

Commenter.