TP 3 : Estimateurs de la moyenne et de la variance

Ce TP n°3 reprend les illustrations de la Loi des Grands Nombres (LGN) et du Théorème Central Limite (TCL), et aborde tranquillement l'estimation.

Apéritif Illustrer la LGN et le TCL pour des variables de Poisson, du paramètre de votre choix.

Exercice 1 : Etude de données réelles, estimateurs de la moyenne et de la variance. Nous cherchons à modéliser la longueur d'un cours d'eau (fleuve ou rivière) comme étant une variable aléatoire de loi "exponentielle décalée", c'est-à-dire de densité h définie par

$$h(x, \alpha, \tau) := \begin{cases} \alpha \exp(-\alpha(x - \tau)) & \text{si } x \geq \tau, \\ 0 & \text{si } x < \tau. \end{cases}$$

1. Ecrire une fonction `h` prenant en paramètres `x, alpha` et `tau`, et qui renvoie $h(x, \alpha, \tau)$. Représenter graphiquement la fonction $x \mapsto h(x, \alpha, \tau)$ pour des paramètres α, τ de votre choix.

2. Ce modèle est-il paramétrique ? Si oui, quels sont ces paramètres ?

On suppose que l'on dispose de $X_1, \dots, X_n$ réalisations de la loi étudiée.

3. Rappeler la définition de l'estimateur de la moyenne, noté $m_n(X_1, \dots, X_n)$, ainsi que l'estimateur de la variance sans biais, noté $v_n(X_1, \dots, X_n)$.

4. En appliquant la loi des grands nombres, vers quoi convergent (presque sûrement) m_n et v_n ? Donner leur limite en fonction de α et τ. En déduire des estimateurs $\hat{\alpha}$ et $\hat{\tau}$ des paramètres α et τ. Ces estimateurs sont-ils consistants ?

L'étude *US Geological Survey* a recensé la longueur (en miles) de 141 rivières d'Amérique du Nord. Le vecteur `rivers` (présent par défaut dans R) contient ces 141 mesures.

5. Calculer m_n, v_n et $\hat{\alpha}$ et $\hat{\tau}$ à l'aide de ces données. Afficher les valeurs de ces estimateurs.

6. Représenter la fonction de répatition empirique de `rivers` (*Indication : on utilisera* `plot` *et* `ecdf`*, comme dans le TP1*) et superposer la fonction de répartition associée à h pour les paramètres $\alpha = \hat{\alpha}$ et $\tau = \hat{\tau}$. Commentez.

Pour celleux qui veulent aller plus loin : une autre loi universelle On considère une matrice aléatoire M de taille $n \times n$ telles que

- M est symétrique : pour tous $1 \leq i, j \leq n, M_{ij} = M_{ji}$,
- Les entrées M_{ij} pour $i < j$ sont indépendantes et de même loi d'espérance nulle et de variance 1. Nous prendrons des lois normales centrées réduites.
- Les éléments diagonaux ont une variance et une espérance finie. Nous prendrons des lois normales $\mathcal{N}(0, 2)$.

1. Vérifier qu'il suffit de prendre

$$M = \frac{A + A^T}{\sqrt{2}},$$

ou A est de taille $n \times n$ avec des entrées i.i.d. Gaussiennes centrées réduites.

2. Simuler une telle matrice M pour $n = 2000$. *Indication : on pourra mettre un vecteur sous la forme d'une matrice grâce à la fonction `matrix(vecteur, ncol=...)`, et la fonction `t` qui transpose une matrice.*

3. Calculer le spectre de $M/\sqrt{n}$ (fonction `eigen` avec les options `symmetric=TRUE, only.values=TRUE`), et afficher l'histogramme des valeurs propres (avec la colonne `values` du résultat). Que se passe-t-il ? A votre avis, comment appelle-t-on cette loi ?