CY-Tech	Statistiques	Mathématiques
2021 - 2022	TD 1	ING2-GI

Exercice 1. -
Soit X le temps d'attente avant d'être servi au RU. Une série d'observations journalières donne l'échantillon suivant (en min)

10	11	15	8	4	22	13	2	10	10	7	1	19	2	3
16	1	8	15	3	1	19	2	3	19	9	8	15	5	9

On calcule la moyenne $\overline{x} = 9$ min et l'écart-type empirique $s^* = 6.3$ min.

1. Quelle loi suit X ?
2. Comment déterminer le paramètre de cette loi à partir de l'échantillon ?
3. En déduire la probabilité qu'un étudiant attende plus de 15 min avant d'être servi.

Exercice 2. -
Soit X une variable aléatoire dont la densité f est définie par $f(x) = \begin{cases} \dfrac{4x^3}{\theta^4} & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$
où θ est un paramètre strictement positif.
On se donne un n-échantillon de X, $(X_1, X_2, ..., X_n)$ et on définit l'estimateur T_n de θ par :
$T_n = a_n(X_1 + X_2 + ... + X_n)$.

1. Vérifier que $E(X) = \dfrac{4\theta}{5}$ et que $V(X) = \dfrac{2\theta}{75}$.
2. Calculer a_n pour que T_n soit sans biais, puis calculer son risque quadratique.

Exercice 3. -
Pour estimer la proportion p d'individus d'une région atteints par une affection, deux médecins ont examinés respectivement $n_1 = 40$ et $n_2 = 60$ personnes choisies au hasard, de manière indépendante.
Les proportions de personnes atteintes dans les deux échantillons sont notées $\overline{x_1}$ et $\overline{x_2}$.

1. Les proportions $\overline{x_1}$ et $\overline{x_2}$ sont des réalisations des variables aléatoires $\overline{X_1}$ et $\overline{X_2}$.
   $\overline{X_1}$ et $\overline{X_2}$ sont-ils des estimateurs sans biais de p ? Sont-ils indépendants ?
2. Les estimateurs suivants sont-ils sans biais ?
   - $T_3 = \dfrac{\overline{X_1} + \overline{X_2}}{2}$.
   - $T_4 = \dfrac{2\overline{X_1} + 3\overline{X_2}}{5}$.
3. Parmi les estimateurs suivants lequel a le plus faible risque quadratique ?
   - $T_1 = \overline{X_1}$.
   - $T_2 = \overline{X_2}$.
   - $T_3 = \dfrac{\overline{X_1} + \overline{X_2}}{2}$.
   - $T_4 = \dfrac{2\overline{X_1} + 3\overline{X_2}}{5}$.
4. Si le premier médecin trouve 25 personnes atteintes (sur 40) et le second en trouve 38 (sur 60), quelle est la meilleure estimation de la proportion d'individus atteints dans la population ?

Exercice 4. -
Soit X une variable aléatoire de densité :
$f(x;\theta) = \begin{cases} \frac{2}{\theta}\left(1 - \frac{x}{\theta}\right) & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}$

On peut vérifier que : $E(X) = \dfrac{\theta}{3}$ et $V(X) = \dfrac{\theta^2}{18}$.
On cherche à estimer θ à partir d'un échantillon X_1 , ..., X_n de même loi que X. On utilise pour cela l'estimateur :

$$T_n = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

1. Calculer le biais de cet estimateur. Pouvez-vous éliminer ce biais ?
2. Etablir la convergence en probabilité de T_n(modifié).
3. Déterminer le risque quadratique puis la convergence en moyenne quadratique.

Exercice 5. -
Soit X_1 , ..., X_n un échantillon de loi normale d'espérance θ et de variance $\theta(1-\theta)$ où $\theta \in]0,1[$ est un paramètre inconnu. On considère les estimateurs

$$T_1 = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \text{et} \qquad T_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

1. Etude théorique des propriétés des estimateurs
   (a) Les estimateurs sont-ils sans biais ?
   (b) Sont-ils convergents ?
   (c) Quel est le meilleur des deux ?
   N.B. On donne $V\left(X_i^2\right) = 2\theta^2\left(1-\theta^2\right)$
2. Illustration des propriétés à l'aide du logiciel R
   (a) Simuler un échantillon de taille $n = 50$ avec $\theta = 0,9$ (vérifier la simulation en calculant la moyenne, la variance et en traçant la distribution de l'échantillon). Calculer les valeurs de
   T_1 et T_2 . Relancer la simulation et observer les valeurs obtenues pour les estimateurs
   (dans la fenêtre environment).
   (b) Simuler 30 échantillons de taille $n = 50$ (matrice à 30 lignes et 50 colonnes de réalisation de $\mathcal{N}(\theta, \theta(1-\theta))$). Calculer les 30 valeurs de T_1 et T_2 . Comparer leur dispersion à l'aide d'un boxplot.
   (c) Que se passe-t-il si $\theta = 0.05$?
   (d) Que se passe-t-il si $n = 5000$?