Feuille n° 2
Statistiques Inférentielles

On admettra le résultat suivant (s'il n'a pas été précédemment démontré).

Théorème 1 (Inégalité de Jensen) *Soit ϕ une fonction convexe sur un intervalle réel I et X une variable aléatoire à valeurs dans I, dont l'espérance existe. Alors,*

$$\phi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left[\phi(X)\right] . \tag{1}$$

Si ϕ est strictement convexe (dérivée seconde strictement positive), alors l'inégalité dans (1) *est stricte.*

Exercice 1

Soit $X_1, \ldots, X_n$ un échantillon de loi uniforme sur $[-\theta, \theta]$, où $\theta > 0$ est inconnu.

1. Montrer que
$$T = \frac{3}{n}\left(X_1^2 + X_2^2 + \ldots + X_n^2\right)$$
est un estimateur sans biais de θ^2.
2. $\sqrt{T}$ est-il un estimateur sans biais de θ ? Si non, quel est le signe de son biais ? *Indication* On pourra utiliser l'inégalité de Jensen.

Exercice 2

Soit $X_1, X_2, \ldots, X_n$ des variables aléatoires de même espérance μ.

1. La variable aléatoire
$$S = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_6$$
est-elle un estimateur sans biais de μ ?
2. Sous quelles conditions sur les réels a et b, la variable aléatoire
$$T = a(X_1 + X_2 + \ldots + X_n) + b$$
est-elle un estimateur sans biais de μ ?
3. Sous quelles conditions sur les réels $a_1, a_2, \ldots, a_n$, la variable aléatoire
$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$
est-elle un estimateur sans biais de μ ?

Exercice 3

En 1986, deux chercheurs ont étudié le nombre de cycles menstruels nécessaires aux femmes pour tomber enceinte, à partir du moment où elles ont décidé d'avoir un enfant. On modélise le nombre des cycles par une variable aléatoire X et on note p la probabilité qu'elle tombe enceinte pendant un cycle donné, indépendemment des cycles précédents ?

1. Déterminer la loi de la variable aléatoire X et calculer $\mathbb{E}X$.
2. Montrer, grâce la loi des grands nombres, que
$$T = \frac{1}{\overline{X}_n}$$
est un estimateur naturel de p.

3. Vérifier que T est un estimateur biaisé et déterminer le signe de son biais.

Exercice 4

Soit $X_1, \ldots, X_n$ un échantillon de loi $\mathcal{E}xp(\lambda)$, où $\lambda > 0$ est un paramètre inconnu. On note M_n le minimum de l'échantillon.

Déterminer la constante c_n telle que $T = c_n M_n$ soit un estimateur sans biais de $1/\lambda$.

Exercice 5

On considère le jeu de données de la durée de vie (en heures) de roulements à billes.

6278	3113	5236	11584	12628	7725	8604	14266	6125	9250
3212	9003	3523	12888	9460	13431	17809	2812	11825	2398

On s'intéresse à l'estimation de la durée de vie minimale et on modélise le jeu de données comme la réalisation d'un échantillon $X_1, \ldots, X_n$. Chaque variable X_i est représentée par

$$X_i = \delta + Y_i$$

où Y_i est distribuée selon une loi $\mathcal{E}xp(\lambda)$ et $\delta > 0$ est un paramètre inconnu modélisant la durée de vie minimale. L'objectif est de construire un estimateur sans biais de δ.

On note

$$M_n = \min(X_1, \ldots, X_n) \quad \text{et} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,.$$

1. Calculer $\mathbb{E}M_n$ et $\mathbb{E}\overline{X}_n$.
2. Montrer que
$$T = \frac{n}{n-1}\left(\overline{X}_n - M_n\right)$$
est un estimateur sans biais de $1/\lambda$.
3. Construire un estimateur sans biais de δ.
4. Proposer une estimation de δ grâce au jeu de données. On pourra utiliser que la durée de vie moyenne des données vaut 8563.5

Exercice 6

En génétique, les feuilles sont divisées en 4 catégories différentes : amidonnée et verte, sucrée et blanche, amidonnée et blanche, et sucrée et verte. Chaque catégorie a une probabilité d'occurence donnée par $p_1 = (\theta+2)/4$, $p_2 = \theta/4$, $p_3 = (1-\theta)/4$ et $p_4 = (1-\theta)/4$ respectivement, où $0 < \theta < 1$ est inconnu.

Supposons que l'on dispose de n feuilles.

Catégorie	Nombre
Amidonnée Verte	1997
Sucrée Blanche	32
Amidonnée Blanche	906
Sucrée Verte	904

Le nombre de feuilles amidonnées et vertes est modélisée par une variable N_1 de loi binomiale $B(n, p_1)$, le nombre de feuilles sucrées et blanches par N_2 de loi binomiale $B(n, p_2)$. Considérons les deux estimateurs suivants pour θ

$$T_1 = \frac{4}{n}N_1 - 2 \quad \text{et} \quad T_2 = \frac{4}{n}N_2\,.$$

1. Montrer que T_1 et T_2 sont des estimateurs sans biais de θ.
2. Calculer l'estimation de θ donnée par chaque estimateur.