DEZs **Ex 1** - *Estimateur et intervalle de confiance*

Soient $X_1, ..., X_n$ des variables aléatoires indépendantes et de même loi ayant pour densité :

$$f_\theta : x \mapsto \begin{cases} \frac{1}{2}(1+\theta x) & \text{si } -1 \leqslant x \leqslant 1 \\ 0 & \text{sinon.} \end{cases}$$

où $\theta \in [-1;1]$ est un paramètre.

1. Montrer que pour tout $\theta \in [-1;1]$, f_θ est une densité de probabilité.

Le fait que $\theta \in [-1;1]$ garantit que $f_\theta(x) \geqslant 0$ pour tout $x \in [-1;1]$. De plus,

$$\begin{aligned} \int f_\theta(x)dx &= \int_{-1}^{1} \frac{1}{2}(1+\theta x)dx \\ &= \int_{-1}^{1} \frac{1}{2}(1+0)dx \quad \text{par imparité de x} \\ &= 1 \end{aligned}$$

Donc f_θ définit une bien une densité de probabilité.

2. Calculer $\mathbb{E}(X_1)$ et $\mathbb{V}(X_1)$. En déduire l'espérance et la variance de la variable aléatoire $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

On calcule les moments d'ordre 1 et 2 de la variable à densité X_1 :

$$\begin{aligned} \mathbb{E}(X_1) &= \int x f_\theta(x)dx \\ &= \int_{-1}^{1} \frac{1}{2}(x+\theta x^2)dx \\ &= \int_{-1}^{1} \frac{1}{2}(\theta x^2)dx \quad \text{par imparité de } x \\ &= 2\int_{0}^{1} \frac{1}{2}(\theta x^2)dx \quad \text{par parité de } x^2 \\ &= \frac{\theta}{3} \end{aligned}$$

De même,

$$\begin{aligned} \mathbb{E}(X_1^2) &= \int x^2 f_\theta(x)dx \\ &= \int_{-1}^{1} \frac{1}{2}(x^2+\theta x^3)dx \\ &= \int_{-1}^{1} \frac{1}{2}(x^2)dx \quad \text{par imparité de } x^3 \\ &= \int_{0}^{1} x^2 dx \quad \text{par parité de } x^2 \\ &= \frac{1}{3} \end{aligned}$$

Avec la formule de Koenig Huygens, on en déduit que

$$\mathbb{V}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \frac{1}{3} - \frac{\theta^2}{9} = \frac{3-\theta^2}{9}$$

3. On pose $T_n = \frac{3}{n}\sum_{i=1}^{n} X_i$. Montrer que T est un estimateur sans biais et convergent de θ.

On calcule par linéarité de l'espérance : $\mathbb{E}(T_n) = \frac{3}{n} \times n\mathbb{E}(X_1) = \theta$. Donc le biais de T_n est $B(T_n) = \mathbb{E}(T_n - \theta) = \theta - \theta = 0$.
De plus, par propriétés de la variance et indépendance,

$$\mathbb{V}(T_n) = \frac{9}{n^2} \times n \times \mathbb{V}(X_1) = \frac{3-\theta^2}{n}$$

Or $EQM(T_n) = \mathbb{V}(T_n) + B(T_n)^2$ donc $EQM(T_n) = \frac{3-\theta^2}{n} \xrightarrow[n\to+\infty]{} 0$: cela prouve que l'estimateur T_n est convergent.

4. À l'aide du Théorème Central Limite, donner une approximation de la loi de T_n.

On pose $\mu = \mathbb{E}(X_1)$ et $\sigma = \sqrt{\mathbb{V}(X_1)}$. Les variables (X_i) sont iid, admettent une espérance et une variance donc d'après le théorème central limite, la variable

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

converge en loi vers une loi normale $\mathcal{N}(0,1)$. En réécrivant, cela revient à dire que

$$\frac{\frac{3}{n}\sum_{i=1}^{n} X_i - 3\mu}{3\frac{\sigma}{\sqrt{n}}} = \frac{T_n - \theta}{\sqrt{\frac{3-\theta^2}{n}}}$$

converge en loi vers une loi normale $\mathcal{N}(0,1)$.
Si n est grand ($n \geqslant 30$), cela revient à dire que T_n suit approximativement une loi normale $\mathcal{N}(\theta, \sigma^2 = \frac{3-\theta^2}{n})$.

5. Démontrer qu'il existe une constante M_n ne dépendant pas de θ telle que si $\lambda > 0$,

$$\mathrm{P}(|T_n - \theta| < \lambda) \geqslant 1 - \frac{M_n}{\lambda^2}$$

D'après l'inégalité de Bienaymé Tchebichev,

$$\mathrm{P}(|T_n - \mathbb{E}(T_n)| \geqslant \lambda) \leqslant \frac{\mathbb{V}(T_n)}{\lambda^2}$$

d'où

$$\mathrm{P}(|T_n - \theta| \geqslant \lambda) \leqslant \frac{3-\theta^2}{n\lambda^2} \leqslant \frac{3}{n\lambda^2} = \frac{M_n}{\lambda^2}$$

en posant $M_n = \frac{3}{n}$. Par passage au complémentaire, on obtient finalement :

$$\mathrm{P}(|T_n - \theta| < \lambda) \geqslant 1 - \frac{3}{n\lambda^2}$$

6. Déterminer un intervalle de confiance permettant d'estimer θ avec une confiance d'au moins 95%.

On cherche un intervalle I tel que $\mathrm{P}(\theta \in I) \geqslant 0{,}95$. Or d'après ce qui précède,

$$\begin{aligned} \mathrm{P}(|T_n - \theta| < \lambda) \geqslant 1 - \frac{3}{n\lambda^2} &\iff \mathrm{P}(-\lambda < T_n - \theta < \lambda) \geqslant 1 - \frac{3}{n\lambda^2} \\ &\iff \mathrm{P}(\theta \in]T_n - \lambda; T_n + \lambda[) \geqslant 1 - \frac{3}{n\lambda^2} \end{aligned}$$

Or $1 - \frac{3}{n\lambda^2} = 0{,}95 \iff \frac{3}{n\lambda^2} = 0{,}05 \iff \lambda^2 = \frac{3}{0{,}05n}$ donc

$$\mathrm{P}\left(\theta \in \left]T_n - \sqrt{\frac{3}{0{,}05n}}; T_n + \sqrt{\frac{3}{0{,}05n}}\right[\right) \geqslant 0{,}95$$

d'où l'intervalle de confiance $I = \left]T_n - \sqrt{\frac{3}{0{,}05n}}; T_n + \sqrt{\frac{3}{0{,}05n}}\right[$.