Devoir en temps libre n$^{\text{o}}$ 2

Exercice 1. Quantiles et comparaison d'estimateurs

Définition 0.1. *Soit X une variable aléatoire réelle dont nous supposons la fonction de répartition F continue et strictement croissante. Pour tout $p \in]0,1[$ nous appelons quantile d'ordre p et nous notons q_p la racine (unique d'après les hypothèses) de l'équation en x : $F(x) = p$, soit $q_p = F^{-1}(p)$. En particulier si $p = 1/2$, $q_{1/2}$ est appelée médiane et sera notée $me(X)$.*

Définition 0.2. *Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire de loi parente X et $(Y_1, \ldots, Y_n)$ l'échantillon ordonné associé. Si $p \in]0,1[$ est tel que $n \times p$ n'est pas entier nous appelons quantile empirique d'ordre p et nous notons $Q_p(n)$ la variable aléatoire $Y_{[n\times p]+1}$, où $[n \times p]$ désigne la partie entière de $[n \times p]$.*

Remarque 0.1. Si $[n \times p]$ est entier nous appellerons quantile d'ordre p toute variable aléatoire comprise entre $Y_{n\times p}$ et $Y_{n\times p+1}$.

1. Soit X une variable aléatoire réelle de fonction de répartition F et $(X_1, \ldots, X_n)$ un échantillon aléatoire de loi parente X. Soit E_x l'événement $\{X < x\}$, $x \in \mathbb{R}$. Que vaut $\mathbb{P}[E_x]$? Nous posons $R_n(x)$ le nombre de répétitions de E_x en n expériences indépendantes. Pour $x \in \mathbb{R}$ fixé, quelle est la loi de $R_n(x)$?

Définition 0.3. *En fonction de $x \in \mathbb{R}$, $F_n(x) = R_n(x)/n$, où $R_n(x)$ est défini à la question **1.**, définit une fonction aléatoire appelée fonction de répartition empirique associée à X basée sur l'échantillon $(X_1, \ldots, X_n)$.*

2. En utilisant la loi forte des grands nombres, montrez que, pour $x \in \mathbb{R}$ fixé, $F_n(x) \stackrel{p.s.}{\to} F(x)$.

3. Montrer que si X est une variable aléatoire réelle dont la fonction de répartition F est continue strictement croissante alors :

$$Q_p(n) \stackrel{p.s.}{\to} q_p, \quad n \to +\infty.$$

Vous pourrez utiliser le théorème de Glivenko-Cantelli dont l'énoncé est le suivant.

Théorème 0.1. *Pour $\mathbb{P}-$presque tout ω, la suite des fonctions de répartition $F_n(\cdot, \omega)$ converge uniformément vers F, autrement dit, nous avons*

$$\mathbb{P} - p.s. \quad \lim_n \sup_{x\in\mathbb{R}} |F_n(x, \cdot) - F(x)| = 0.$$

4. Soit X une variable aléatoire admettant une densité continue $f > 0$, alors la variable aléatoire $T_p = \sqrt{n}\,(Q_p - q_p)$ possède la propriété suivante :

$$T_p \stackrel{\mathcal{L}}{\to} \mathcal{N}\left(0, \frac{\sqrt{p(1-p)}}{f}(q_p)\right).$$

5. Déduire de la question **4.** que si l'échantillon aléatoire $(X_1, \ldots, X_n)$ a pour loi parente X une loi normale centrée réduite alors $\sqrt{n}Q_{1/2}$ suit asymptotiquement une normale $\mathcal{N}(0, \sqrt{\pi/2})$. Montrez que $\overline{X}_n = \widehat{\mu}_n$ et $Q_{1/2}$ sont deux estimateurs convergents de la moyenne d'une loi normale $\mathcal{N}(\mu, \sigma)$. Lequel est-il le plus efficace ?

6. Qu'en est-il pour la loi logistique, dont la fonction de densité est définie pour $x \in \mathbb{R}$ par $f(x) = \exp(-x)/(1 + \exp(-x))^2$, et la loi de Laplace , dont la fonction de densité est définie pour $x \in \mathbb{R}$ par $f(x) = \exp(-|x|)/2$.