TD n° 1 : Représentation graphique et analyse de lois de variables aléatoires et d'échantillons

Exercice 1. (Estimation de densité par un histogramme) On pose

$$f(x) = 2x \cdot \mathbf{1}_{x\in[0,1]}.$$

1. Tracer f et montrer que f est la densité d'une loi de probabilité.

On note $(X_1,\dots,X_n)$ un échantillon de taille n de densité f. On fixe $K \in \mathbb{N}^*$ et on définit

$$Z_k = \sum_{i=1}^n \mathbf{1}_{X_i\in]\frac{k-1}{K},\frac{k}{K}]}, \quad k = 1,\dots,K$$

$$\hat{f}_n(x) = \begin{cases} \frac{K}{n} Z_k, & \text{si } x \in]\frac{k-1}{K},\frac{k}{K}] \\ 0, & \text{sinon} \end{cases}$$

$$g_K(x) = \begin{cases} K \int_{(k-1)/K}^{k/K} f(y)\,dy, & \text{si } x \in]\frac{k-1}{K},\frac{k}{K}] \\ 0, & \text{sinon.} \end{cases}$$

Alors $\hat{f}_n$ est l'estimateur de la densité donnée par l'histogramme avec K classes de l'échantillon $(X_1,\dots,X_n)$.

2. Pour $k \in \{1,\dots,K\}$, donner la loi de Z_k et en déduire que $\mathbb{E}[Z_k] = n\int_{(k-1)/K}^{k/K} f(y)\,dy$ et $\mathrm{Var}(Z_k) \le n\int_{(k-1)/K}^{k/K} f(y)\,dy$.
3. Montrer que $\mathbb{E}[\|\hat{f}_n - g_K\|_2^2] \le \frac{K}{n}$, où $\|g-h\|_2 = \sqrt{\int_0^1 (g(x)-h(x))^2\,dx}$.
4. En déduire que $\mathbb{E}[\|\hat{f}_n - g_K\|_2] \le \sqrt{\frac{K}{n}}$.
5. Montrer que $\|g_K - f\|_2 = \frac{1}{\sqrt{3K^2}}$.
6. En déduire que $\mathbb{E}[\|\hat{f}_n - f\|_2] \le \frac{1}{\sqrt{3K^2}} + \sqrt{\frac{K}{n}}$.
7. Montrer que la quantité à droite de la dernière inégalité est minimisée pour $K = \left\lfloor \sqrt[3]{\frac{4}{3}n} \right\rfloor$ ou $K = \left\lfloor \sqrt[3]{\frac{4}{3}n} \right\rfloor + 1$. *Conseil : minimiser d'abord sur $K > 0$ réel, puis utiliser un argument de convexité.*

Exercice 2. (Fonctions inverses généralisées)

1. Montrer que pour toute fonction $F : \mathbb{R} \to \mathbb{R}$ croissante et continue à droite, on a
$$\forall u, x \in \mathbb{R} : F(F^-(u)) \ge u \quad \text{et} \quad F^-(F(x)) \le x.$$
En déduire que
$$\forall u, x \in \mathbb{R} : F^{-1}(u) \le x \iff u \le F(x).$$
2. Soit F la fonction de répartition d'une loi de probabilité μ sur $\mathbb{R}$. Soit $U \sim \mathrm{Unif}(0,1)$. Montrer que la v.a. $X = F^{-1}(U)$ suit la loi μ.
3. *Bonus :* Démontrer le théorème de Glivenko-Cantelli (montrer au préalable qu'il suffit de le montrer pour $X_1,\dots,X_n$ suivant la loi uniforme sur $[0,1]$).