Estimation non paramétrique
Master 1 MIMSE – Université de Bordeaux

– TP 1 : Fonction de répartition empirique –

Exercice 1. Le but de cet exercice est d'illustrer les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle.

1) Générer un échantillon de taille $n = 100$ de variables aléatoires i.i.d. selon une loi exponentielle $\mathcal{L}$ de paramètre λ à choisir.
2) Soit F la fonction de répartition de la loi $\mathcal{L}$. Estimer F à partir de la fonction de répartition empirique $\hat{F}_n$ en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3) Représenter graphiquement F et $\hat{F}_n$.
4) Illustrer la convergence presque sûre de $\hat{F}_n(x)$ vers $F(x)$.
5) Illustrer également la normalité asymptotique associée à $\hat{F}_n(x)$.
6) Illustrer graphiquement la convergence de $\|\hat{F}_n - F\|_\infty$ quand $n \to +\infty$.

Exercice 2. Mélanges de gaussiennes.

1) Simuler un échantillon de taille $n = 100$ de variables aléatoires i.i.d. selon le mélange gaussien suivant à K classes défini par la densité :

$$f(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2\sigma_k^2}(x-m_k)^2}, x \in \mathbb{R}$$

où les π_k sont des poids positifs tels que $\sum_{k=1}^{K} \pi_k = 1$. Le choix des paramètres dans ce modèle est libre.

2) Représenter graphiquement la densité f pour différentes valeurs des paramètres du modèle.
3) Représenter graphiquement la fonction de répartition F associée à ce modèle, ainsi que la fonction de répartition empirique $\hat{F}_n$ associée à un échantillon.
4) Est-il possible de retrouver facilement les modes de la densité f à partir de $\hat{F}_n$? Le résultat devient-il plus facile à interpréter lorsque n augmente ?