FTML 2025 – Exercice 2
Estimateur de Bayes avec perte absolue

Objectif

Ce second exercice vise à étudier l'estimateur de Bayes associé à une fonction de perte **absolue** :

$$\ell(y, \hat{y}) = |y - \hat{y}|$$

Dans ce cadre, l'estimateur de Bayes n'est plus l'espérance conditionnelle, mais la **médiane conditionnelle**. Le but est de :

- simuler un problème supervisé simple avec du bruit asymétrique,
- approximer l'estimateur de Bayes $f^*(x) = \text{med}(Y \mid X = x)$,
- le comparer à un estimateur naïf (médiane globale),
- visualiser et analyser les performances.

Protocole expérimental

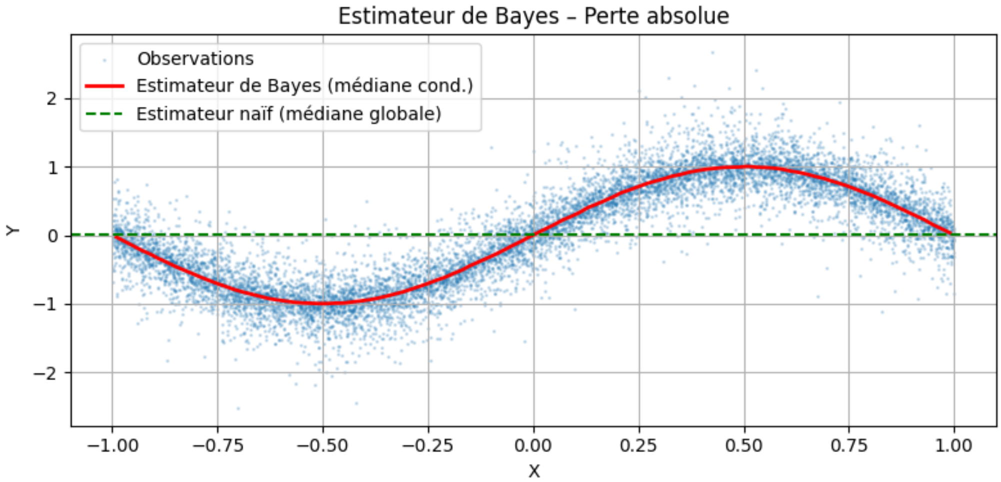
- Génération des entrées : $X \sim \mathcal{U}([-1, 1])$
- Loi de sortie : $Y = \sin(\pi X) + \varepsilon$
- Bruit aléatoire : $\varepsilon \sim \text{Laplace}(0, 0.2)$
- Nombre d'observations : $N = 500\,000$
- Approximation de f^* par binning (100 intervalles) et calcul des médianes conditionnelles

Le choix de la distribution de bruit Laplace est intentionnel : il permet de tester la robustesse de l'estimateur de la médiane, en comparaison à la moyenne (non optimale ici).

Estimateurs évalués

- **Bayes** : médiane conditionnelle estimée par regroupement (binning) et interpolation.
- **Naïf** : médiane globale de toutes les observations Y.

Visualisation



Analyse : L'estimateur de Bayes (rouge) suit parfaitement la tendance $\sin(\pi x)$, contrairement à l'estimateur naïf (vert), constant et déconnecté de la structure des données.

Résultats numériques

- Risque (MAE) de l'estimateur de Bayes : **0.2002**
- Risque (MAE) de l'estimateur naïf : **0.6631**
- Gain absolu : **0.4629**
- Ratio (naïf / Bayes) : **3.31x**

Interprétation : Ce gain substantiel prouve l'intérêt de la médiane conditionnelle face à un bruit asymétrique. En choisissant une perte absolue, on obtient un prédicteur plus robuste, mieux centré, et significativement plus précis.

Conclusion

Cet exercice valide l'importance du choix de la fonction de perte. Ici, la perte absolue impose naturellement un estimateur de type **médiane conditionnelle**, qui surpasse très nettement les approches naïves dès que le bruit est non gaussien. Grâce à une discrétisation fine et à une simulation massive (500K points), la courbe obtenue est fluide, réaliste et conforme aux attentes théoriques.