Devoir 1

Exercice 1 : estimation avec peu de valeurs

On a réalisé 4 mesures d'un même élément ; on a obtenu les résultats suivants :

$$10,23 \quad 10,86 \quad 10,91 \quad 10,50.$$

Après avoir précisé les hypothèses nécessaires, proposer un intervalle de confiance de la valeur mesurée à l'aide de la loi de Student.

Afin de décrire la précision de l'appareil de mesure, donner également un intervalle de confiance pour l'écart-type.

Exercice 2 : estimateur de la médiane

On considère une variable aléatoire X de densité continue f inconnue. On souhaite en estimer la médiane notée M. On rappelle qu'elle est définie par la propriété

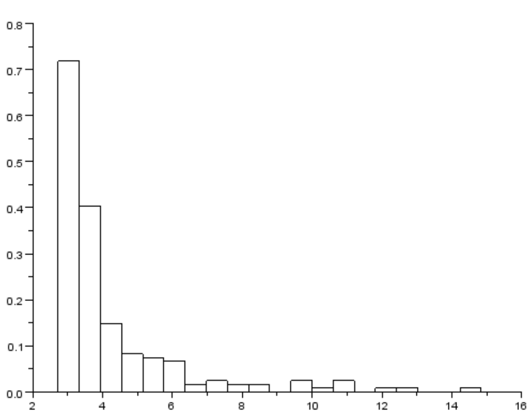
$$\mathbf{P}(X \geqslant M) = \frac{1}{2}.$$

On considère un échantillon $X_1, \dots, X_n$ de taille n. On propose d'estimer la médiane M de X par la médiane de l'échantillon, c'est-à-dire par la valeur centrale de l'échantillon, une fois ordonné. Si l'échantillon est de taille n paire, on considère la moyenne des deux valeurs centrales de l'échantillon.

Quitte à considérer les X_i ordonnés, on définit donc l'estimateur ainsi :

$$\overline{M} = \begin{cases} X_{\frac{n+1}{2}} & \text{si } n \text{ impair} \\ \frac{1}{2}(X_{\frac{n}{2}} + X_{\frac{n+1}{2}}) & \text{si } n \text{ pair} \end{cases}$$

Pour étudier notre estimateur, nous disposons d'un certain échantillon de taille $n = 200$. Ses valeurs sont représentées ci-dessous dans un histogramme normalisé (l'aire totale vaut 1).



Estimation de la moyenne avec cet échantillon :

$$\overline{X} = \frac{\sum_{i=1}^{200} x_i}{200} = 4,197.$$

Estimation de la médiane avec cet échantillon : comme 200 est pair, on estime la médiane par la moyenne de la 100$^{\text{ème}}$ et de la 101$^{\text{ème}}$ valeur de l'échantillon. On obtient

$$\overline{M} = \frac{3,478 + 3,485}{2} = 3,481.$$

1. **Intervalle de confiance de la moyenne**
   Déterminer un intervalle de confiance à 95% de la moyenne de X.
   On fournit la valeur suivante : $\sum_{i=1}^{200}(x_i - 4,197)^2 \approx 794$.

2. **Robustesse de l'estimateur**
   La robustesse d'un estimateur est sa capacité à ne pas être modifié par une petite modification dans les données. En particulier, un estimateur robuste doit voir sa valeur faiblement perturbée par la présence de valeurs aberrantes dans l'échantillon.

   Pour l'illustrer, imaginons qu'on ajoute à nos 200 valeurs de l'échantillon une $201^{\text{ème}}$ valeur excessivement grande de 30.

   (a) Quelles sont alors les nouvelles estimations de la moyenne et de la médiane ?
   (b) Comparer la robustesse de nos deux estimateurs.

3. **Maximum de vraisemblance**
   Nous allons justifier que notre estimateur de la médiane est en un sens un estimateur du maximum de vraisemblance.
   Notons Y le nombre de valeurs de l'échantillon supérieures à la médiane M.

   (a) La variable Y suit une loi binomiale. Quels en sont les paramètres ?
   (b) Quelle est la valeur de Y ayant la probabilité maximale ?
   (c) En déduire que notre estimateur $\overline{M}$ est un estimateur du maximum de vraisemblance.

4. **Intervalle de confiance de la médiane**
   On admet le résultat suivant : si la densité f de X est continue et strictement positive, alors la variable aléatoire $2f(M)\sqrt{n}(\overline{M} - M)$ converge en loi quand n tend vers $+\infty$ vers la loi normale centrée réduite.

   En supposant ces hypothèses satisfaites pour notre exemple, on peut raisonnablement considérer que pour n assez grand,

   $$2f(M)\sqrt{n}\left(\overline{M} - M\right) \sim \mathcal{N}(0,1).$$

   (a) En déduire que notre estimateur de la médiane est convergent.
   (b) Estimer graphiquement la valeur de $f(M)$.
   Ce n'est pas la façon la plus rigoureuse de procéder mais nous nous en contenterons.
   (c) Déterminer un intervalle de confiance à 95% de M.
   (d) Interpréter et expliquer brièvement le rôle du terme $f(M)$ pour la taille de l'intervalle de confiance.