Filière 3EA
1^{re} **année**

Statistique

BE n°1 - Estimation paramétrique

1 Estimation de la moyenne

L'objectif de ce premier exercice est d'étudier le comportement de l'estimateur de la moyenne d'une variable aléatoire réelle X calculé à partir de n observations de cette variable :

$$\widehat{m}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Pour des données distribuées selon une loi normale centrée réduite $\mathcal{N}(0,1)$, on représentera, en fonction de n :

- sur une premier graphique : deux réalisations de l'estimateur calculée sur n observations pour n variant de 10 à 200 par pas de 10,
- sur un deuxième graphique : la moyenne estimée de l'estimateur à partir de K de ses réalisations (on prendra par exemple $K = 500$) ainsi que la moyenne estimée $\pm$ l'écart type estimé sur les K réalisations, pour n variant également de 10 à 200 par pas de 10.

On comparera les résultats obtenus par simulation aux valeurs théoriques de la moyenne et la variance de l'estimateur. Réaliser les mêmes tracés en considérant maintenant une loi de Cauchy (on pourra utiliser pour la génération l'inverse de la fonction de répartition d'une loi de Cauchy). Commenter.

2 Estimation de l'amplitude d'un signal

Lors de sa transmission, un signal (représenté ici sous sa forme échantillonnée) $\mathbf{s} = (s_1, s_2, \ldots, s_n)^{\mathsf{T}}$ subit une atténuation d'un facteur θ (constant sur toute la durée du signal) ainsi que des perturbations modélisées par un bruit additif $\mathbf{b} = (b_1, b_2, \ldots, b_n)^{\mathsf{T}}$, de telle sorte que le signal reçu $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathsf{T}}$ s'écrit sous la forme :

$$y_i = \theta s_i + b_i \ , \ \ 1 \leq i \leq n.$$

On s'intéresse ici à l'estimation de l'atténuation θ.

On suppose que les échantillons de bruit $b_1, b_2, \ldots, b_n$ sont indépendants et identiquement distribués selon une loi normale, de moyenne nulle et de variance σ^2. Le signal $\mathbf{s} = (s_1, s_2, \ldots, s_n)^{\mathsf{T}}$ est supposé connu.

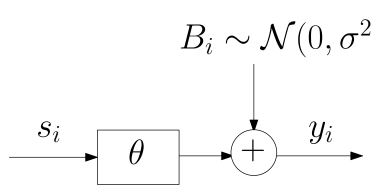

Figure 1 – Transmission d'un signal par un canal bruité.

1. Montrer alors que l'estimateur du maximum de vraisemblance de θ est donné par :

$$\widehat{\theta}_{\text{MV}} = \frac{\mathbf{s}^{\mathsf{T}}\mathbf{y}}{\mathbf{s}^{\mathsf{T}}\mathbf{s}}$$

Il est possible de montrer que :

(a) $\widehat{\theta}_{\text{MV}}$ est un estimateur non biaisé :

$$\mathbb{E}\left[\widehat{\theta}_{\text{MV}}\right] = \theta$$

(b) la variance de $\widehat{\theta}_{\text{MV}}$ est donnée par :

$$\text{Var}(\widehat{\theta}_{\text{MV}}) = \frac{\sigma^2}{\mathbf{s}^{\mathsf{T}}\mathbf{s}}$$

(c) $\widehat{\theta}_{\text{MV}}$ est l'estimateur efficace, c'est-à-dire que sa variance est égale à la borne de Cramer-Rao des estimateurs non biaisés de θ.

On souhaite vérifier ces différentes propriétés à l'aide de simulations numériques effectuées sous Matlab. Ainsi, pour chaque longueur de signal n, on doit générer K signaux $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathsf{T}}$, et calculer l'estimateur du maximum de vraisemblance sur chacun de ces K signaux. On peut alors estimer, sur K de ses réalisations, la moyenne et la variance de l'estimateur à partir de n observations (par réalisation) et comparer avec les grandeurs théoriques.

2. Ecrire une fonction `Y = generer(teta,signal,sigma2,K)` qui renvoie une matrice `Y` de taille $n \times K$, dont chaque colonne contient une réalisation du signal $\mathbf{y}_j = (y_{1,j}, y_{2,j}, \ldots, y_{n,j})^{\mathsf{T}}$. Les variables d'entrée sont :

   — `teta` : paramètre θ modélisant l'atténuation ;
   — `signal` : vecteur des n échantillons du signal supposé connu $\mathbf{s} = (s_1, s_2, \ldots, s_n)^{\mathsf{T}}$;
   — `sigma2` : variance σ^2 du bruit additif supposée connue ;
   — `K` : nombre de réalisations du vecteur des observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathsf{T}}$.

   On génèrera une matrice `B` de taille $n \times K$, dont chaque colonne contient une réalisation du bruit $\mathbf{b}_j = (b_{1,j}, b_{2,j}, \ldots, b_{n,j})^{\mathsf{T}}$. Pour ajouter à chaque colonne de cette matrice le signal $\theta\mathbf{s} = (\theta s_1, \theta s_2, \ldots, \theta s_n)^{\mathsf{T}}$, on pourra utiliser la fonction `repmat` de Matlab (utiliser `help`).
   Tester cette fonction avec $\theta = 4$, $n = 100$, $s_i = \log i$, $1 \leq i \leq n$, $\sigma^2 = 1$, $K = 500$.
   Tracer une réalisation du vecteur signal y ainsi que la moyenne des K réalisations.

3. Ecrire une fonction `[teta_est,BRC] = estimateur_mv(teta,Y,signal,sigma2)` qui renvoie K valeurs de l'estimateur $\widehat{\theta}_{\text{MV}}$ pour chacune des K réalisations du vecteur des observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathsf{T}}$, à partir de la matrice `Y` construite à la question 1, ainsi que la borne de Cramer-Rao correspondant à `signal` et `sigma2`.

4. Calculer la moyenne empirique et la variance empirique de $\left(\widehat{\theta}_{\text{MV}}(k)\right)_{k=1,\ldots,K}$ et comparer avec les valeurs théoriques.

5. Recommencer l'expérience pour $n = 100, 150, 200, \ldots, 500$. Afficher les courbes représentant la moyenne et la variance estimées de l'estimateur en fonction de n et superposer (`hold on`) les courbes des bornes de Cramer-Rao théoriques correspondantes.