Prof. Kamal GHOUMID Année universitaire $2020-2021$

$4^{\text{ème}}$ année "**Ingénierie Data Sciences & Cloud Computing**"

Cours "**Détection, Estimation & Information pour les Data Sciences**"

Série de TD $N^o.4$

Ex-1- Estimation, Maximum de vraisemblance.

On considère une variable aléatoire X (data échantillons) $X_1, ..., X_n$, indépendantes et identiquement distribuées (iid).
Estimer en utilisant la méthode du maximum de vraisemblance :

- $\star$ Le paramètre θ d'une loi de Poisson : $X \sim \mathbb{P}(x, \theta) = \dfrac{e^{-\theta}\ \theta^x}{x!} \quad , \quad x \in \mathbb{N}.$
- $\star$ Le paramètre $\dfrac{1}{\mu}$ d'une loi exponentielle : $X \sim \epsilon(\frac{1}{\mu}) = \begin{cases} \frac{1}{\mu}\ e^{\frac{-x}{\mu}} & \text{x} > 0 \\ 0 & \text{sinon} \end{cases}$

Ex-2- Estimation, Maximum de vraisemblance.

Soient $x_1, ..., x_n$ les data de n observations d'un caractère qui peut être modélisé par une variable aléatoire X de densité de probabilité :

$$f_\theta(x) = \begin{cases} \frac{\theta}{1-\theta}\ x^{\frac{2\theta-1}{1-\theta}} & \text{si } x \in [0\,,\,1] \\ 0 & \text{sinon} \end{cases}$$

Avec $\theta \in]\frac{1}{2}, 1[$ est un paramètre inconnu.
Déterminer l'estimateur $\widehat{\theta}_{MV}$ le plus vraisemblable du paramètre θ pour $x_1, ..., x_n$.

Ex-3- Biais, Convergence, Information de Fisher, Borne de Cramér-Rao, MSE.

On considère la distribution de Bernoulli qui prend la valeur 1 "succès" avec la probabilité p et 0 "échec" avec la probabilité $q = 1 - p$, donnée par :

$$f_\theta(x) = \theta^x\ (1-\theta)^{1-x}, \qquad x \in \{0\,,1\},\ \ 0 < \theta < 1$$

On cherche à partir des data d'un échantillon aléatoire de n tests, obtenir une estimation ponctuelle du paramètre θ.

1. Calculer $E[X], E[X^2]$ et $Var(X)$.

2. Donner l'estimeteur $\widehat{\theta}_{MV}$ du paramètre θ, obtenu selon la méthode du maximum de vraisemblance.
3. Estimer la probabilité du succès dans le cas d'un échantillon aléatoire contenant les 15 tests suivants $\{1\,;0\,;0\,;1\,;1\,;0\,;1\,;1\,;0\,;1\,;1\,;0\,;1\,;0\,;1\}$.
4. Calculer $E[\widehat{\theta}_{MV}]$, puis en déduire que $\widehat{\theta}_{MV}$ est non biaisé.
5. Calculer l'information de Fisher$I_n(\theta)$ relative au paramètre θ.
6. Exprimer la variance $Var(\widehat{\theta}_{MV})$ en fonction l'information de Fisher$I_n(\theta)$. Conclure concernant la borne de Cramér-Rao. L'estimateur $\widehat{\theta}_{MV}$ est-il parfait ?
7. Montrer que l'erreur quadratique moyenne MSE (Mean Square Error) est nulle pour un nombre de test n très grand.

Ex-4- Max. de vraisemblance, Efficacité d'un estimateur, Borne de Cramér-Rao.

Dans un traitement de données massives, on considère n variables aléatoires $\{X_1, ..., X_n\}$ iid qui suivent la même loi normale $\mathcal{N}(0, \sigma^2)$ centrée et de variance σ^2 inconnu.

1. Déterminer la fonction de vraisemblance de $\{X_1, ..., X_n\}$ et montrer qu'elle admet un maximum unique pour une valeur de σ^2 à déterminer. En déduire ensuite l'estimateur $\widehat{\sigma^2}_{MV}$ le plus vraisemblable du paramètre σ.
2. Déterminer la borne de Cramér-Rao pour un estimateur non biaisé du paramètre σ^2 à partir de l'observation des $\{X_1, ..., X_n\}$.
3. Montrer que $\widehat{\sigma^2}_{MV}$ est un estimateur efficace de σ^2.

Ex-5- Eestimation d'une constante dans le bruit, Borne de Cramér-Rao.

Dans une transmission bruitée on cherche à estimer une constante. Pour cela, on considère le n-échantillon suivant :

$$r_i = \theta + n_i$$

Où les n_i reprśentent des variables aléatoires Gaussiennes centrées indépendantes de même variances σ^2. On propose alors d'estimer la constante inconnue θ au moyen de l'estimateur suivant :

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} r_i$$

1. Calculer l'espérance $E[\widehat{\theta}]$ puis en déduire que cet estimateur est sans biais.
2. Calculer la variance $\sigma^2_{\widehat{\theta}} = Var(\widehat{\theta})$.
3. Déterminer la borne de Cramér-Rao pour cet estimateur non biaisé puis en déduire qu'il est consistant.