Fiche de TD n°3 :

$\boxed{1}$ Soit $X_1, ...X_n$ un n-échantillon de la population de moyenne ν et de variance σ^2.

[a] Montrer que l'estimateur $\sum_{i=1}^n a_i X_i$ est un estimateur non biaisé de ν si $\sum_{i=1}^n a_i = 1$.

[b] Parmi tous les estimateurs de cette forme (appelé estimateur linéaire non biaisé) trouvez celui avec la variance minimale et la déterminer.

1 Intervalle de confiance

$\boxed{2}$ Dans une ville on donne la répartition du nombre de jours sans accident, avec un accident, etc. parmi 50 jours d'observation au cours d'une même année :

Nbre accidents	0	1	2	3	4
Nbre jours	21	18	7	3	1

On suppose que le nombre d'accidents par jour suit une loi de Poisson. Donner un intervalle de confiance de niveau 0,95 pour le nombre moyen d'accidents par jour (on utilisera une approximation asymptotique).

$\boxed{3}$ Un sondage auprès de 1 500 ménages tirés au hasard dans la population française a indiqué que 20 % de ceux-ci prévoient d'acheter une nouvelle voiture dans les douze prochains mois. Estimer par un intervalle de confiance asymptotique à 95 % le pourcentage de ménages de la population française prévoyant d'acheter une nouvelle voiture dans les douze mois.

$\boxed{4}$ Un stock comporte 10 000 pièces. Pour évaluer le nombre de pièces défectueuses dans le stock on tire au hasard 400 pièces dont on constate que 45 sont défectueuses. Donner un intervalle de confiance asymptotique à 99 % pour le nombre total de pièces défectueuses.

$\boxed{5}$ Julien, qui travaille dans un magasin de bricolage, vient de recevoir une importante commande de vis. Pour estimer la proportion de vis défectueuses, il effectue n expériences indépendantes. Au cours de chaque expérience, il tire des vis avec remise, jusqu'à ce qu'il en tire une défectueuse. Elle est alors remise dans le lot, et Julien passe à l'expérience suivante.

[a] Déterminer le modèle statistique associé à cette expérience.

[b] Construire un estimateur de la proportion de vis défectueuses.

[c] Soit $\alpha \in]0, 1[$. En déduire un intervalle de confiance asymptotique (sur le nombre d'expériences) au niveau $(1 - \alpha)$ pour la proportion de vis défectueuses.

[d] Donner un estimateur sans biais du nombre moyen de vis que Julien a tirées au cours de ces n expériences.

[e] Soit $\alpha \in]0, 1[$. Construire un intervalle de confiance asymptotique (sur le nombre d'expériences) au niveau $(1 - \alpha)$ pour le nombre moyen de vis tirées.

$\boxed{6}$ Lois exponentielles On considère le modèle statistique $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathcal{E}(\lambda)^{\otimes n}\}_{\lambda>0})$. Dans la suite, $(X_1, X_2, \ldots, X_n) \sim \mathcal{E}(\lambda)^{\otimes n}$ et $\alpha \in]0, 1[$.

[a] Montrer que la v.a. $\lambda \max_{1\leq i\leq n} X_i$ est pivotale, i.e. sa loi ne dépend pas de λ. En déduire un intervalle de confiance pour le paramètre λ au niveau de confiance $1-\alpha$.

[b] Construire un intervalle de confiance pour λ au niveau $1-\alpha$ en utilisant un pivot relié à $\bar{X}_n$.

[c] Déterminer la vitesse de l'estimateur $\hat{\lambda} = 1/\bar{X}_n$. En déduire un intervalle de confiance asymptotique pour le paramètre λ au niveau $1-\alpha$.

$\boxed{7}$ Soit $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{Q_\theta^{\otimes n}\}_{\theta\in]-1,1[})$ un modèle statistique tel que, pour tout $\theta \in]-1,1[$, Q_θ désigne la loi sur $\mathbb{R}$ de densité

$$f_\theta(x) = \frac{1}{2}(1+\theta x)\mathbb{1}_{]-1,1[}(x).$$

[a] Construire un estimateur $\hat{\theta}$ de θ en utilisant la méthode des moments.

[b] Calculer son biais et son risque quadratique moyen.

[c] Soit $\alpha \in]0,1[$. Donner un intervalle de confiance asymptotique pour θ au niveau de confiance $1-\alpha$.

2 Estimateur du maximum de vraisemblance.

$\boxed{8}$ On considère l'expérience statistique $\Upsilon = (\{0,1\}, \mathcal{P}(\{0,1\}), (\mathbb{P}_\theta)_{\theta\in\{\frac{1}{10},\frac{8}{10}\}})$, où

$$\mathbb{P}_\theta(dx) = \theta\delta_{\{0\}}(dx) + (1-\theta)\delta_{\{1\}}(dx).$$

Montrer que le modèle est dominé , identifiable, calculer une version de sa fonction de vraisemblance. En déduire l'estimateur du maximum de vraisemblance. Généraliser pour un modèle discret.

$\boxed{9}$ Calculer l'estimateur du maximum de vraisemblance (s'il existe et s'il est bien défini) et étudier ses propriétés (vitesse de convergence , intervalle de confiance) dans les cas suivants :

[a] On observe $(X_1, \ldots, X_n)$ i.i.d. $\mathcal{P}(\lambda)$, où $\lambda > 0$.

[b] On observe $X \sim \mathcal{B}(n,p)$, où $p \in]0,1[$ et n est connu.

[c] On observe $(X_1, \ldots, X_n)$ i.i.d. $\mathcal{N}(\nu, \sigma^2)$, où $\nu \in \mathbb{R}$ et $\sigma > 0$.

[d] On observe $(X_1, \ldots, X_n)$ i.i.d. de loi uniforme sur $[0,\theta]$, où $\theta > 0$.

$\boxed{10}$ Un étang contient N poissons où N est inconnu. M (connu) poissons ont été marqués. On pêche (sans remise) jusqu'à ce qu'on obtienne le premier poisson marqué. Soit X le nombre de poissons qu'on doit ainsi pêcher.

Donner la loi de X en supposant un tirage aléatoire sans remise.
En déduire l'équation de vraisemblance de N associée à une (seule) observation x de X.
Application : résoudre numériquement avec $M = 100$ et $x = 3$ pour donner une estimation de N.

$\boxed{11}$ Soit $g : [0,1] \to]0,\infty[$ une densité continue connue. Pour $n \geq 1$, on observe $(Y_1, ..., Y_n)$ où :

$$Y_i = \theta + \frac{1}{\sqrt{g(i/n)}}\epsilon_i, \quad i = 1, ..., n,$$

et $\epsilon_1, ..., \epsilon_n$ sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$ avec σ^2 paramètre connu.

[a] Ecrire le modèle statistique correspondant. Montrer qu'il est identifiable et dominé.

[b] Calculer le maximum de vraisemblance, noté $\hat{\theta}^{MV}$ de θ.

[c] Calculer $\mathbb{E}((\hat{\theta}^{MV} - \theta)^2)$.

$\boxed{12}$ Pour chaque $\theta \in \mathbb{R}$, on note Q_θ la loi sur $\mathbb{R}$ de densité

$$\theta^2 x e^{-\theta x} \mathbb{1}_{\mathbb{R}_+}(x).$$

On observe un n-échantillon de loi Q_θ.

[a] Calculer le maximum de vraisemblance, noté $\hat{\theta}^{MV}$ de θ.

[b] Etudier le biais et le risque quadratique de $\hat{\theta}^{MV}$.

[c] Déterminer la loi limite de $\sqrt{n}(\hat{\theta}^{MV} - \theta)$.

$\boxed{13}$ On suppose que les v.a. $Y_1, ..., Y_n$ satisfassent :

$$Y_i = \beta x_i + \epsilon_i, \ \ i = 1, ..., n,$$

où les $x_1, ..., x_n$ sont des constantes fixés et connues, et $\epsilon_1, ..., \epsilon_n$ sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$ avec σ^2 paramètre inconnu.

[a] Calculer le maximum de vraisemblance, noté $\hat{\beta}^{MV}$ de β et montrez que c'est un estimateur non biaisé de β.

[b] Quelle est la loi de $\hat{\beta}^{MV}$?

[c] Montrer que

$$\hat{\beta}^2 = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}$$

est un estimateur non biaisé de β.

[d] Calculer la variance de $\hat{\beta}^2$. Quel estimateur préférez vous entre $\hat{\beta}^2$ et $\hat{\beta}^{MV}$?

[e] Montrer que

$$\hat{\beta}^3 = \frac{\sum_{i=1}^n \frac{Y_i}{x_i}}{n}$$

est un estimateur non biaisé de β.

[f] Calculer la variance de $\hat{\beta}^3$.

[g] Quel estimateur préférez vous entre les 3 précédents ?

3 Estimateur de la médiane

$\boxed{14}$ Soit une famille de lois $\{f(x, \theta)\}$ telle que $f(x, \theta)$ s'écrive $f(x - \theta)$ où f ne dépend pas de θ. (On dit que θ est un paramètre de positionnement et on note qu'on a également $F(x, \theta) = F(x - \theta)$ pour la fonction de répartition). Supposons, de plus, que f soit une fonction paire (lois symétriques par rapport à θ). Les résultats suivants seront établis pour n impair (mais sont valables pour $n = 2k$ en définissant la médiane de façon unique par $\frac{1}{2}(X_{(k)} + X_{(k+1)})$).

[a] Etablir la fonction de répartition, puis la densité de la médiane empirique d'un n-échantillon.

[b] Montrer que sa loi est également symétrique par rapport à θ.

[c] En déduire que la médiane empirique est un estimateur sans biais pour θ.

$\boxed{15}$ Lois de Cauchy Pour chaque $\theta \in \mathbb{R}$, on note Q_θ la loi sur $\mathbb{R}$ de densité

$$f_\theta(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}.$$

L'objectif est de définir et d'étudier un estimateur du paramètre du modèle $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{Q_\theta^{\otimes n}\}_{\theta\in\mathbb{R}})$.

[a] Pour quelle raison ne peut-on pas utiliser la méthode des moments pour construire un estimateur de θ ?

[b] Soit F_θ la fonction de répartition de la loi Q_θ. Calculer

$$\inf\left\{x \in \mathbb{R} \ : \ F_\theta(x) \geq \frac{1}{2}\right\}$$

puis en déduire un estimateur $\hat{\theta}$ de θ.

[c] Dans la suite, la notation $(X_{(1)}, \ldots, X_{(n)})$ désigne la statistique d'ordre associée à $(X_1, \ldots, X_n)$ i.e. le réarrangement croissant de l'échantillon. Montrer qu'il est unique.

[d] On suppose que n est pair. Montrer que $\hat{\theta} = X_{(n/2)}$.

[e] Calculer la fonction de répartition de $\hat{\theta}$.

[f] Prouver que $\hat{\theta}$ est consistant.

[g] On note F_n la fonction de répartition empirique de l'échantillon :

$$x \mapsto F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}.$$

(a) Pour tout $x \in \mathbb{R}$, montrer que

$$\sqrt{n}\left(F_n(\theta + \frac{x}{\sqrt{n}}) - F_n(\theta)\right) \xrightarrow[Q_\theta^{\otimes n}]{} \frac{x}{\pi}.$$

(b) En déduire la loi limite de $\sqrt{n}(F_n(\theta + x/\sqrt{n}) - 1/2)$ pour tout $x \in \mathbb{R}$.

(c) Conclure.