Examen de Statistique Mathématique

Les téléphones portables doivent être éteints et rangés.
L'usage des calculatrices n'est pas autorisé.
Les notes de cours et autres documents sont interdits.
La qualité de la rédaction sera très largement prise en compte.

Exercice 1.
Pour chacun des modèles suivants, dire si le modèle est dominé (auquel cas exhiber une mesure dominante), donner le support de la loi de X noté $X(\Omega)$, donner l'espace naturel des paramètres noté Θ et dire si le modèle est régulier :

1. $\mathbb{P}_{a,\lambda}[X = x] = e^{-\lambda}\dfrac{\lambda^{x-a}}{(x-a)!}$ pour $x \in \{a, a+1, ...\}$ avec $\lambda > 0$ et $a \in \mathbb{N}$.
2. $F_{\alpha,\lambda}(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\alpha-\lambda x} & \text{si } x \geq 0 \end{cases}$ avec $\alpha, \lambda > 0$.
3. $F_a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < a \\ \dfrac{(a-1)x - a}{a-2} & \text{si } a \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$ avec $1.5 \leq a < 2$.

Exercice 2.
Afin de mieux dimensionner la capacité d'un serveur informatique, un ingénieur réalise une étude sur la durée de connexion des clients au serveur. On admet que la durée (exprimée en minutes) de connexion d'un client au serveur peut être représentée par une variable aléatoire X positive de loi exponentielle de paramètre $\lambda > 0$ à estimer. Pour cela, l'ingénieur constitue un échantillon $(X_1, ..., X_n)$ de variables aléatoires i.i.d. distribuées comme la variable parente X. On rappelle que la fonction de répartition de la loi exponentielle est donnée pour $x > 0$ par

$$F_\lambda(x) = 1 - e^{-\lambda x}.$$

1. Déterminer $\widehat{\lambda}_n$ un estimateur de λ.
2. Etudier le comportement asymptotique de $\widehat{\lambda}_n$.
3. Déterminer un intervalle de confiance asymptotique pour λ de niveau de confiance 95%.
4. Proposer un estimateur consistant de la probabilité qu'un client se connecte au serveur pour une durée comprise entre 15 et 20 minutes. Etablir son comportement asymptotique. En déduire un intervalle de confiance asymptotique de niveau de confiance 95% pour la probabilité qu'un client se connecte au serveur pour une durée comprise entre 15 et 20 minutes.

5. Estimer, parmi 1000 clients pris au hasard, le nombre moyen de clients se connectant au serveur pour une durée inférieure à 15 minutes. Etablir la consistance et le comportement asymptotique de l'estimateur choisi. En déduire un intervalle de confiance asymptotique de niveau de confiance 95% pour le nombre de clients se connectant au serveur pour une durée inférieure à 15 minutes parmi 1000 clients pris au hasard.

Exercice 3.

Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de variable parente X dont la distribution admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$f_\theta(x) = \frac{2x}{\theta^2} I(0 \leq x \leq \theta) \quad \text{avec } \theta > 0.$$

1. Vérifier qu'il s'agit bien d'une densité de probabilité.
2. Déterminer la fonction de répartition associée ainsi que l'espérance et la variance de X.
3. Déterminer l'information de Fisher du modèle probabiliste (à une observation).
4. Déterminer $\widetilde{\theta}_n$ l'estimateur de θ obtenu par la méthode des moments.
5. Déterminer $\widehat{\theta}_n$ l'estimateur de θ obtenu par la méthode du maximum de vraisemblance.
6. Déterminer le biais, la variance et le risque quadratique de $\widetilde{\theta}_n$ et $\widehat{\theta}_n$. Comparer $\widetilde{\theta}_n$ et $\widehat{\theta}_n$.
7. La borne de Fréchet-Darmois-Cramer-Rao s'applique-t-elle ici ? Pourquoi ?
8. Etudier le comportement asymptotique de $\widetilde{\theta}_n$ et de $\widehat{\theta}_n$.

Exercice 4.

Soit $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ un vecteur aléatoire de loi $\mathcal{N}\left(\begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -1 \\ 2 & -1 & 3 \end{pmatrix}\right)$.

1. Calculer $\mathbb{E}[X - Z | Y - X]$. En déduire la loi de $\mathbb{E}[X - Z | Y - X]$.
2. Déterminer l'ensemble des variables $\eta_{a,b,c} = aX + bY + cZ$ avec $a, b, c \in \mathbb{R}$ indépendante de $U = Y + Z - X$.
3. Déterminer la loi de $\begin{pmatrix} Y - X \\ Z - X \end{pmatrix}$.
4. Déterminer la loi de $\begin{pmatrix} Y \\ X \\ Z \end{pmatrix}$.

Exercice 5.

Soient U, V et Y des variables aléatoires représentant des quantités que l'on suspecte être liées. Plus particulièrement, on se demande s'il existe une influence de U et V sur Y. Soient (U_i, V_i, Y_i) pour $i = 1, ..., n$ des triplets i.i.d. distribués comme (U, V, Y).

1. On souhaite ajuster un modèle de régression linéaire de la forme $Y = aU + bV + c + \varepsilon$ où ε est une variable aléatoire centrée indépendante de (U, V) et de variance σ^2. Comment interpréter ce modèle ? Que valent $\mathbb{E}[Y|U, V]$ et $\text{Var}(Y|U, V)$?
2. Déterminer $\widehat{a}$, $\widehat{b}$ et $\widehat{c}$ les estimateurs de a, b et c respectivement obtenus par la méthode des moindres carrés i.e. satisfaisant le problème d'optimisation suivant :

$$(\widehat{a}, \widehat{b}, \widehat{c}) = \text{argmin}_{(a,b,c) \in \mathbb{R}^3} \sum_{i=1}^{n} (Y_i - aU_i - bV_i - c)^2.$$