Corrigé Examen final Stat.nonparametrique MASTER II : Mathématiques .Appliquée& Statistique.)

Exercice 01 : (06 pts)

1. Donner la difference entre l'estimation paramétrique et l'estimation non paramétrique.
La statistique paramétrique est le cadre "classique" de la statistique. Le modèle statistique est dècrit par un nombre fini de paramètres avec une famille de loi connu modèle Gaussien, modèle loi Poisson....Par opposition, en statistique non paramétrique, le modèle n'est pas décrit par un nombre fini de paramètres.et de loi ou famille de loi inconnu.

2. Enoncer le principe du test de wilcoxon ? Pour la statistique de wilcoxon W^+, calculez espérance et la variance de W^+.
le test de signe et de rang Pour déterminer la présence d'un effet positif d'un échantillon on utilise le test de Wilcoxon , Pour cela calculez la statistique $W^+ = \sum R_{(d_i>0)}$, son espérance et sa variance. On rappelle que
$$E[W^+] = \frac{n(n+1)}{4} \quad , Var(W^+) = \frac{n(n+1)(2n+1)}{24}$$

3. Donner la difference entre la régression paramétrique et la régression non-paramétrique ?
Un modèle de régression consiste à déterminer la façon dont l'espérance d'une variable dépendante Y dépend d'un ensemble de variables explicatives X. Le problème consiste donc à déterminer pour chaque réalisation de la variable x; la valeur de la fonction r(x), dite fonction de lien. *un modèle de régression paramétrique.* On suppose que cette fonction peut s'écrire comme une fonction explicite des valeurs de X: Cette fonction peut être linéaire, logarithmique, non-linéaire etc. Dans un modèle de *régression non-paramétrique*, la fonction de lien (r(x)) n'a pas de forme explicite et ne pas peut s'écrire en fonction d'un nombre réduit de paramètres.

4. Est ce que on peut estimer la densité de probabilité par la dérivé de la fonction de répartition empérique ?justifier?
Non – puisque la la fonction de répartition empérique est discontinue donc n'est pas dérivable

5. On disposer de n observations i.i.d. $X_1, ..., X_n$ de densité f_θ par rapport à la mesure de Lebesgue sur $\mathbb{R}$ définie par $f_\theta(x) = 2\theta x e^{-\theta x^2} \mathbb{1}_{x>0}$. et $\theta > 0$ un paramètre inconnu.
   - Calculer la fonction de répartition F_θ de X_1.? $F_\theta(x) = \int_0^x 2\theta t e^{-\theta t^2} dt = \left(1 - e^{-\theta x^2}\right) \mathbb{1}_{x>0}$
   - Quelle est la médiane $q_{0.5}$ de la loi de X_1? $F_\theta(q_{0.5}) = 0.5 \Longrightarrow \left(1 - e^{-\theta q_{0.5}^2}\right) = 0.5 \Longrightarrow q_{0.5} = \sqrt{(\log(2))/\theta}$

Exercice 02 : (05 pts)

Soit (X, Y) un couple de V.A.R de fonction densité conjointe $f_{X,Y}$ définie par
$$f(x,y) = \begin{cases} \frac{4}{5}(x+3y)e^{-x-2y}, & \text{if } 0 \leq x \text{ and } 0 \leq y; \\ 0, & \text{otherwise.} \end{cases}$$

1. Déterminer les lois marginales de X et de Y .
   - $f_X(x) = \int_0^{+\infty} \frac{4}{5}(x+3y)e^{-x-2y} dy = \frac{4}{5}\int_0^{+\infty} xe^{-x-2y} + 3ye^{-x-2y} dy =$
   $\frac{2}{5}\left[xe^{-x} \underbrace{\int_0^{+\infty} 2e^{-2y}dy}_{\mathcal{E}(2)} + 3e^{-x} \underbrace{\int_0^{+\infty} 2e^{-2y}dy}_{\mathcal{E}(2)} \right] \Longrightarrow f_X(x) = \frac{2}{5}\left[e^{-x}(x+3)\right] \mathbb{1}_{x\geq 0}$

- $f_Y(y) = \int_0^{+\infty} \frac{4}{5}(x+3y)e^{-x-2y}dx = \frac{4}{5}\int_0^{+\infty} xe^{-x-2y} + 3ye^{-x-2y}dy =$
$\frac{4}{5}\left[e^{-2y}\underbrace{\int_0^{+\infty} xe^{-x}dx}_{\mathcal{E}(1)\Longrightarrow\mathbb{E}(x)=1} + 3ye^{-2y}\underbrace{\int_0^{+\infty} e^{-x}dx}_{\mathcal{E}(1)}\right] \Longrightarrow f_Y(y) = \frac{4}{5}\left[e^{-2y}(1+3y)\right]\mathbb{1}_{y\geq 0}$

2. Calculer $cov(x,y)$ et étudier l'independance de X et Y.

   - $cov(x,y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
     (a) $\mathbb{E}(XY) = \frac{4}{5}\int_0^{+\infty}\int_0^{+\infty} xy(x+3y)e^{-x-2y}dxdy$
     $= \frac{4}{5}\int_0^{+\infty}\int_0^{+\infty} x^2ye^{-x-2y} + 3xy^2e^{-x-2y}dxdy =$
     $\frac{4}{5}\left[\underbrace{\int_0^{+\infty} ye^{-2y}dy}_{Y\sim\mathcal{E}(2)\Longrightarrow\mathbb{E}(y)=1}\underbrace{\int_0^{+\infty} x^2e^{-x}dx}_{X\sim\mathcal{E}(1)\Longrightarrow var(x)=1} + \underbrace{\int_0^{+\infty} 3y^2e^{-2y}dy}_{Y\sim\mathcal{E}(2)\Longrightarrow var(y)=3/8}\underbrace{\int_0^{+\infty} xe^{-x}dx}_{X\sim\mathcal{E}(1)\Longrightarrow\mathbb{E}(x)=1}\right] \Longrightarrow \mathbb{E}(XY) = \frac{4}{5}(1\times$
     $1 + \frac{3}{8}\times 1) = \frac{44}{40} = \frac{11}{10}$
     (b) $\mathbb{E}(X) = \int_0^{+\infty} xf_X(x)dx = \frac{2}{5}\int_0^{+\infty} x\left[e^{-x}(x+3)\right]dx = \frac{2}{5}\int_0^{+\infty} x\left[x^2e^{-x} + 3xe^{-x}\right]dx = \frac{8}{5}$
     (c) $\mathbb{E}(Y) = \int_0^{+\infty} yf_Y(y)dy = \frac{4}{5}\int_0^{+\infty} y\left[e^{-2y}(1+3y)\right]dy = \frac{2}{5}\int_0^{+\infty}\left[y2e^{-2y} + 3y^22e^{-2y})\right]dy =$
     $\frac{2}{5}(1/2+3/4) = 1/2 \Longrightarrow cov(x,y) = \frac{11}{10} - \frac{8}{10} = \frac{3}{10}$
   - L'independance de X et Y. $\Longrightarrow f_{X,Y}(x,y) \neq f_X(x)\times f_Y(y) \Longrightarrow X\perp Y$

3. Calculer $m(x) = \mathbb{E}(Y/X=x)$? $\Longrightarrow \mathbb{E}(Y/X=x)\int_0^{+\infty} yf_{Y/X}(y)dy = \int_0^{+\infty}\frac{1}{x+3}y(x+3y)2e^{-2y}dy =$
$\frac{1}{x+3}\int_0^{+\infty}\left[xy2e^{-2y} + 3y^22e^{-2y}\right]dy = \frac{1}{x+3}\left[\frac{x}{2}+\frac{3}{4}\right]$

Exercice 03 : ____ (09 pts)

I— Soit $(X_1,...,X_n)$ n variables aléatoires indépendantes et de même loi uniforme sur l'intervalle $[0,\theta]$.

(a) - Donnez la forme explicite de la fonction de répartition ?

Soit la V.A.R. continue X qui suit $U([0,\theta])$,sa fonction de répartition notée F_X est donnée par :

$$F_X(x) = \begin{cases} 0, & \text{if } x<0;\\ \frac{x}{\theta}, & \text{if } 0\leq x\leq\theta;\\ 1, & \text{if } x\geq\theta;\end{cases}$$

(b) - Soit $Y = \max(X_1,...,X_n)$ diterminer la densité de Y ? $\Longrightarrow F_Y(y) = P(Y\leq y) = P(\max(X_1,...,X_n)\leq y) =^{ind} P(X_1)\leq y)P(X_2\leq y)...\times P(X_n\leq y) = (P(X_1\leq y))^n = \left(\frac{y}{\theta}\right)^n \Longrightarrow f_Y(y) = \frac{n}{\theta^n}y^{n-1}\mathbb{I}_{[0,\theta]}(y)$.

(c) - Soit $M_n = \frac{1}{y}\mathbb{I}_{[x,\theta]}(y)$ Montrer que pour $y\in[x,\theta]$ $\quad\mathbb{E}(M_n) = \frac{1}{\theta}\frac{n}{n-1}\left(1-\left(\frac{x}{\theta}\right)^{n-1}\right)$

$$\mathbb{E}(M_n) = \int_0^{+\infty}\frac{1}{y}f_Y(y)dy = \int_0^{+\infty}\left(\frac{n}{\theta^n}\right)y^{n-2}dy = \frac{n}{\theta^n}\left[\frac{y^{n-1}}{n-1}\right]_x^\theta \Longrightarrow \mathbb{E}(M_n) = \frac{1}{\theta}\frac{n}{n-1}\left[1-\left(\frac{x}{\theta}\right)^{n-1}\right]$$

II— • Soit $(X_1, ..., X_n)$ des variables aléatoires indépendantes et de même loi qui admet pour densité f. On suppose que les variables sont dans L^1.

(a) Rappeler la définition d'un noyau sommatif ?
un noyau statistique K est une densité de probabilité. vérifie

$$\int_{\mathbb{R}} K(x)dx = 1 \quad \int_{\mathbb{R}} xK(x)dx = 0 \quad , \int_{\mathbb{R}} K^2(x)dx = \tau^2 < +\infty.$$

(b) Rappeler la formule de $\widehat{f}_n$ l'estimateur à noyau pour le noyau uniforme (cet estimateur est appelé l'histogramme mobile)
$\widehat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X-x_i}{h}\right)$ ..

$$\text{K un noyau uniforme} \Longleftrightarrow K(u) = \begin{cases} 1, & \text{if } u \in [-1/2, 1/2]; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 1/2, & \text{if } u \in [-1, 1]; \\ 0, & \text{otherwise.} \end{cases}$$

(c) montrer que $\widehat{f}_n$ est une function densité de probabilité .?

$$\int_{-\infty}^{+\infty} \widehat{f}_n(t)dt = \int_{-\infty}^{+\infty} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-x_i}{h}\right) dt = \int_{-\infty}^{+\infty} K(u)\, du = 1$$

(d) On définit l'intégrale

$$I(X_1, ..., X_n) = \int x\widehat{f}_n(x)dx$$

a- Que représente cette intégrale ?
Cette intégrale représente $\mathbb{E}(x)$ telque $x \sim \widehat{f}_n(x)$

b- Calculer $I(X_1, ..., X_n)$. Commenter le résultat.

$$I(X_1, ..., X_n) = \int x\widehat{f}_n(x)dx = \int x\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X-x_i}{h}\right) dx =_{\text{K noyau uniforme}} \Longrightarrow \int_{-1}^{1} x\frac{1}{2}dx = 0$$

$\Longrightarrow \mathbb{E}(x) = 0$ centré

(e) On suppose que les variables sont dans L^4 . On définit l'intégrale

$$J(X_1, ..., X_n) = \int x^2\widehat{f}_n(x)dx$$

Que représente cette intégrale ? Calculer $J(X_1, ..., X_n)$.

Cette intégrale représente le moment d'ordre 2 $\mathbb{E}(x^2)$ telque $x \sim \widehat{f}_n(x)$

$$J(X_1, ..., X_n) = \int x^2\widehat{f}_n(x)dx = \int x^2\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X-x_i}{h}\right) dx =_{\text{K noyau uniforme}} \Longrightarrow \int_{-1}^{1} x^2\frac{1}{2}dx = \frac{1}{3}$$

Rappel : $x \sim Unif_{[a,b]}(x) \Longrightarrow \mathbb{E}(x) = \frac{a+b}{2}$ $Var(x) = \frac{(b-a)^2}{12}$

• Nous envisageons maintenant le lissage de $F_n(x)$ pour le cas où l'on sait pouvoir se restreindre à une fonction de répartition dérivable jusqu'à un certain ordre.et l'estimer avec la noyau intégré.

(a) Donner l'expression générale d'un estimateur à noyau de la densité de probabilité $\widehat{f}_n(x)$?
l'expression générale d'un estimateur à noyau de la densité de probabilité

$$\widehat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X-x_i}{h}\right).$$

telque K un noyau statistique (K est une densité de probabilité.h paramètre de lissage) vérifie

$$\int_{\mathbb{R}} K(x)dx = 1 \quad \int_{\mathbb{R}} xK(x)dx = 0 \quad , \int_{\mathbb{R}} K^2(x)dx = \tau^2 < +\infty.$$

(b) Déterminer $\mathbb{E}(\widehat{f}_n(x)) - f(x)$. ? Commenter le résultat obtenu.
sous les conditions général du noyau K on a

$$\mathbb{E}(\widehat{f}_n(x)) - f(x) = \frac{h^2}{2} f''(x) \int_{\mathbb{R}} u^2 K(u) du + O(h^2)$$

. $\mathbb{E}(\widehat{f}_n(x)) - f(x)) \longrightarrow_{h \longrightarrow 0} 0.$ donc $\widehat{f}_n(x)$ est un estimateur sans biais.

(c) Soit la noyau intégré définie par:

$$H(u) = \int_{-\infty}^{u} K(t) dt.$$

a- Déterminer $H(u)$ pour K soit bien un noyau uniforme ?

$$\text{K un noyau uniforme} \Longleftrightarrow K(u) = \begin{cases} 1, & \text{if } u \in [-1/2, 1/2]; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 1/2, & \text{if } u \in [-1, 1]; \\ 0, & \text{otherwise.} \end{cases}$$

par intégration

$$H(x) = \begin{cases} 0, & \text{if } u < -1; \\ \frac{1}{2}(u+1), & \text{if } -1 \leq u \leq 1; \\ 1, & \text{if } u > 1; \end{cases}$$

(d) Determiner l'estimateur de la fonction F en intégrant l'estimation par noyau de la densite f?
Pour trouver l'estimateur de F en intégrant l'estimation par noyau de la densite f *ie* :

$$\widehat{F}_n(x) = \int_{-\infty}^{x} \widehat{f}_n(t) dt.$$

$$\widehat{F}_n(x) = \int_{-\infty}^{X} \widehat{f}_n(t) dt = \frac{1}{nh} \int_{-\infty}^{X} \sum_{i=1}^{n} K\left(\frac{t - x_i}{h}\right) dt = \frac{1}{nh} \sum_{i=1}^{n} \int_{-\infty}^{X} K\left(\frac{t - x_i}{h}\right) dt.$$

$$\widehat{F}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \int_{-\infty}^{\frac{X - x_i}{h}} K(v) dv = \frac{1}{n} \sum_{i=1}^{n} H\left(\frac{X - x_i}{h}\right).$$

(e) Montrer que

$$\mathbb{E}(\widehat{F}_n(x)) - F(x) = \frac{h^2}{2} f'(x) \int_{\mathbb{R}} u^2 K(u) du + o(h^2).$$

Commenter les résultats obtenus.,

$$\mathbb{E}(\widehat{F}_n(x)) - F(x) = \int (\mathbb{E}(\widehat{f}_n(t)) - f(t)) dt = \frac{h^2}{2} f'(x) \int_{\mathbb{R}} u^2 K(u) du + o(h^2).$$

$\mathbb{E}(\widehat{F}_n(x)) - F(x) \longrightarrow_{h \rightarrow 0} 0$
On déduire que la fonction de répartition empirique $\widehat{F}_n$ est sans biais,

Fin