Université Hassiba Benbouali Chlef
Faculté des Sciences Exactes & Informatique
Département de Mathématiques

Année Universitaire 2021/2022
Master II : Maths.Appliquée& Statistique.
Module : Statistique.nonparametrique

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

Exercice 01 : (09 pts)

I- Soit $X_1, X_2, ..., X_n$ une suite de n v.a. i.i.d.. La fonction de répartition de X_i est $F(x)$. Soit $\widehat{F}_n(x)$ la fonction de répartition empirique .

01 1. Quelle est l'intérêt de l'estimation non parametrique ?
On utilise l'estimation non paramétrique,lorsque le modèle n'est pas décrit par un nombre fini de paramètres c'est a dire la loi ou famille de loi est inconnu.ou lorsque l'estimation paramétrique n'est pas satisfait pour le modéle estimé

0.5 2. Donner la définition explicite de la fonction de répartition empirique $\widehat{F}_n$ associée aux variables $X_1, ..., X_n$.

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} = \begin{cases} 0, & \text{if } x < X_{(1)}; \\ \frac{i}{n}, & \text{if } X_{(i)} \leq x < X_{(i+1)}; \\ 0, & \text{if } x \geq X_n; \end{cases}$$

1.5 3. Quelle est la loi de $nF_n(x)$ et la loi limite de $\sqrt{n}(F_n(x) - F(x))$ pour un élément x fixé dans R.
$nF_n(x)$ suit une loi binomiale de paramètre $(n, F(x))$.
$\mathbb{E}(F_n(x)) = F(x)$, pour tout x, $F_n(x)$ est un estimateur sans biais de $F(x)$.
$Var(nF_n(x)) = nF(x)(1 - F(x))$
On déduit du TCL que pour tout x tel que $F(x)(1 - F(x)) \neq 0$
$\sqrt{n}(F_n(x) - F(x)) \longrightarrow^{Loi} N(0, F(x)(1 - F(x)))$.

0.5 4. Démontrer que pour tout $x \in \mathbb{R}$, $\widehat{F}_n(x) \longrightarrow^{p}_{n\longrightarrow\infty} F(x)$.
$E[(F_n(x) - F(x))^2] = Biais^2 + Variance$ pour un élément x fixé dans R.
$\mathbb{E}(F_n(x)) - F(x) = 0$ et $Var(F_n(x)) = \frac{F(x)(1-F(x))}{n}$
$\mathbb{E}[(F_n(x) - F(x))^2] = \frac{F(x)(1-F(x))}{n}$
$MSE \longrightarrow 0 \Longrightarrow \widehat{F}_n(x) \longrightarrow^{p}_{n\longrightarrow\infty} F(x)$

II- On veux Tester l'homogénéité des deux forêts pour cela on a relevé les hautes en mètres de 5 et 6 arbres respectivement. On obtient le tableau suivant :

X	22.5	22.9	23.4	24	
Y	23	24.5	24.6	25.7	26

0.5 **(a)** Montrer qu'il faut utiliser un test non paramétrique.
Test d'homogénéité entre deux echantillons (X,Y) de loi Inconnue donc on utilise un test non paramétrique.

02 **(b)** Soit $\hat{F}_n(x)$ et $\hat{G}_n(y)$ la fonction de répartition empérique de X et Y respectivement. Calculer $D^+ = \sup |\hat{F}_n(x) - \hat{G}_n(y)|$.

Les Obs	22.5	22.9	23	23.4	24	24.5	24.6	25.7	26
$\hat{F}_n(x) = \frac{i}{4}$	1/4	2/4	2/4	3/4	4/4=1	1	1	1	1
$\hat{G}_n(y) = \frac{i}{5}$	0	0	1/5	1/5	1/5	2/5	3/5	4/5	5/5=1
$\lvert\hat{F}_n(x) - \hat{G}_n(y)\rvert$	1/4	2/4	3/10	4/5	4/5	3/5	2/5	1/5	0

$$\Longrightarrow D_n^+ = 4/5 = 0.8$$

III- Le graphisme suivante illustre la densité f et la Fonction de répartition F d'une variable aléatoire Z.

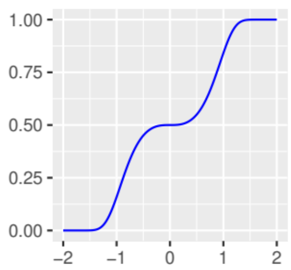

Figure 01

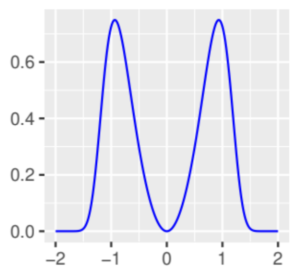

Figure 02

01 1. Préciser le graphe de la densité et la Fonction de répartition de Z ?
- Figure 01 $\Longrightarrow$ la Fonction de répartition.
- Figure 02 $\Longrightarrow$ la Fonction de densité de probabilité.

01 2. Est ce que Z gaussienne? Est ce que f est bimodale?
Z n'est pas gaussienne (n'est pas symétrique en 0) et la densité est bimodale selon le graphe de la densité.

01 3. On suppose que la densité f s'ecrit sous la forme $f(x) = \dfrac{1}{x^2}$ pour $x \geq 1$. Déterminer $\mathbb{P}(X = 2)$ et la médiane m ?
$\mathbb{P}(X = 2) = \displaystyle\int_2^2 \frac{1}{x^2} dx = 0$
La médiane $m \Longrightarrow F(m) = \frac{1}{2} \Longrightarrow \displaystyle\int_1^m \frac{1}{x^2} dx = 1 - \frac{1}{m} = \frac{1}{2} \Longrightarrow m = 2$

Exercice 02 : ______________________________ (11 pts)

A-) Soit K un noyau statistique. et Soit un échantillon $(X_1, ..., X_n)$ iid dont la loi parente est de densité f . On rappelle que l'estimateur non paramétrique de Parzen-Rosenblatt de f est donné par $\widehat{f}_h(x)$, pour tout $x \in \mathbb{R}$,

01 1. Donner la définition explicite d'un noyau sommatif? donner deux exemples?
Soit un noyau statistique K est une densité de probabilité. vérifie

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

exemple : $K(u) = 15/16(1 - u^2)^2_{I_{\{-1 \leq u \leq 1\}}}$, $K(u) = 1/2, I_{u \in [-1,1]}$;

01 2. Montrer que, pour tout $h > 0$ et tout $x \in \mathbb{R}$, la fonction définie par

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

est également un noyau statistique (au sens de la définition rappelée ci-dessus).

$\int_{\mathbb{R}} K_h(x)dx = \int_{\mathbb{R}} \frac{1}{h} K\left(\frac{x}{h}\right) dx = 1$

01 3. Donner la définition d'un estimateur non paramétrique de Parzen-Rosenblatt de f et Montrer que $\widehat{f}_h(x)$ est une densité de probabilité ?

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

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

donc $\widehat{f}_n$ est bien une function densité de probabilité.

01 4. Montrer que le $biais(\widehat{f}_n(x)) \longrightarrow 0$ as $h \longrightarrow 0$.?

Déterminer $\mathbb{E}(\widehat{f}_n(x)) - f(x)$.

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.

01 **B-) I-)** Soit les variables aléatoires X et Y $\in \mathbb{R}^2$ et $K(u)$ est un noyau sommatif . montrer que :

$$\int_{\mathbb{R}} \frac{y}{h} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right) dy = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i.$$

Voir le cours .

II-) La densité conjointe des variables aléatoires X et Y est sous la forme suivante:

$$f(x,y) = \begin{cases} 2xy + \frac{3}{2}y^2, & 0<\text{x}< 1 \text{ et } 0<\text{y} <1; \\ 0, & \text{sinon.} \end{cases}$$

1.5 1. Vérifier que $f(x,y)$ est une densité?.Trouver les densités marginales $f_X(x)$ et $f_Y(y)$.

- $\int\int f(x,y)dxdy = \int_0^1 \int_0^1 (2xy + \frac{3}{2}y^2)dxdy = 1,$
- $f_X(x) = \int f(x,y)dy = \int_0^1 (2xy + \frac{3}{2}y^2)dy = \left[2x\frac{y^2}{2} + \frac{3}{6}y^3\right]_0^1 = x + \frac{1}{2} I_{x\in[0,1]}$
- $f_Y(y) = \int f(x,y)dx = \int_0^1 (2xy + \frac{3}{2}y^2)dx = \left[2y\frac{x^2}{2} + \frac{3}{2}y^2 x\right]_0^1 = y + \frac{3}{2}y^2 I_{y\in[0,1]}$

01 2. Trouver la densité conditionnelle $f_{Y|X=x}(y)$.

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{2xy + \frac{3}{2}y^2}{x + \frac{1}{2}} \Longrightarrow f_{Y|X=x}(y) = \frac{4xy + 3y^2}{2x + 1} . I_{y\in[0,1]}$$

1.5 3. Trouver $\mathbb{P}((X,Y) \in [0,1/2]\times[0,1/2])$. , $\mathbb{P}(X<Y)$. et $\mathbb{E}(Y|X=x)$.

- $\mathbb{P}((X,Y) \in [0,1/2]\times[0,1/2]) = \int_0^{1/2}\int_0^{1/2}(2xy+\frac{3}{2}y^2)dxdy = \frac{1}{16}$,
- $\mathbb{P}(X<Y) = \int_0^1\int_0^y(2xy+\frac{3}{2}y^2)dxdy = \frac{5}{8}$,
- $\mathbb{E}(Y|X=x) = \int_0^1 y f_{Y|X=x}(y)dy = \int_0^1 y\frac{4xy+3y^2}{x+2}dy = \frac{1}{12}.\frac{16x+9}{2x+1}$,

4. Soit la variable aléatoire $Z = \mathbb{E}(Y|X)$.

01 (a) Quelle est la distribution de Z ?

$$F_Z(z) = \mathbb{P}(Z<z) = \mathbb{P}(\frac{1}{12}.\frac{16x+9}{2x+1} < z) \Longrightarrow \mathbb{P}\left(X < \frac{12z-9}{16-24z}\right) = F_X\left(\frac{12z-9}{16-24z}\right)$$

$$\Longrightarrow f_Z(z) = \frac{3}{64(2-3z)^3}.I_{\frac{25}{36}<z<\frac{3}{4}},$$

01 (b) Trouver $\mathbb{E}(Z)$.

$$\mathbb{E}(Z) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y) = \int_0^1(y^2+\frac{3}{2}y^3)dy = \frac{17}{24},$$

Fin