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

Année Universitaire 2021/2022
Master II : Math.Appliq& Statistique.
Module : Stat.non paramétrique

Corrigée Examen de Rattrapage (Master2)

Exercice 01 : ______ (06 pts)

1. Donner la difference entre l'estimation paramétrique et l'estimation non paramétrique.

01

le test paramétrique suppose que les données sont issues d'une loi de probabilité de forme connue dont seuls les paramètres sont inconnus et on cherche a estimer. Dans ce cas, l'estimation de densité se résume à l'estimation des paramètres de la distribution. Mais si la loi de probabilité est inconnue,on utilise l'approche du test non-paramétrique ou s'il s'agit justement de trouver la forme de cette loi sans à priori pour ensuite en réaliser une test paramétrique, on doit se tourner vers une méthode de test non-paramétrique dans laquelle les données parlent d'elles mêmes

2. Soit F une fonction de répartition. Donner les propriétés de F.?

01

On appelle fonction de répartition associée au n échantillon $X_1, ..., X_n$ la fonction

$F(x) = P(X \leq x)$ est croissante, continue à droite , limite a gauche (cadlag) $\lim_{x \longrightarrow -\infty} F(x) = 0 \quad \lim_{x \longrightarrow +\infty} F(x) = 1$

3. On considère n variables $X_1, ..., X_n$ i.i.d et de même loi de fonction de répartition F.

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

   0.5

   On appelle fonction de répartition empirique associée au n échantillon $X_1, ..., X_n$ la fonction

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

   (b) Enoncer le théorème de Glivenko-Cantelli.?

   0.5

   Soit $(X_1, ..., X_n)$ un échantillon de loi L sur $\mathbb{R}$, de fonction de répartition F inconnu . Pour tout $x \in \mathbb{R}$, on a $F_n(x) \longrightarrow^{p.s.}_{n \longrightarrow \infty} F(x)$, et cette convergence est uniforme sur $\mathbb{R}$

$$\|F_n(x) - F(x)\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \longrightarrow^{p.s.}_{n \longrightarrow \infty} 0$$

4. Quelle la défférence entre l'estimation de la densité de probabilité par l'histogramme mobile et l'estimation noyau.?

01

l'histogramme mobile utilise noyau uniforme et l'estimation de la densité de probabilité par l'histogramme mobile

5. Donner la définition de l'Estimateur de Nadaraya Watson.?

01

$$r_n(\chi) := \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

6. On note X une variable aléatoire à valeurs réelles de fonction de répartition $F(.)$. Calculer la médiane m de F?

01

$$F(x) = \begin{cases} 1 - \exp(-4x^2), & \text{if } 0 \leq x; \\ 0, & \text{otherwise.} \end{cases}$$

$F(m) = 1/2 \Longrightarrow m = \frac{\sqrt{\ln 2}}{2}$

Exercice 02 : ____(06 pts)

1. Consider la loi conjoint avec pdf

$$f(x,y) = \begin{cases} e^x + y, & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

(a) Déterminer la densité marginale $f_X(x)$

01

$$f(x) = \int_0^1 f(x,y)dy = \int_0^1 (e^x + y)dy = [e^x + \frac{1}{2}]I(x)_{[0,1]}$$

(b) Déterminer la densité marginale $f_Y(y)$

01

$$f(y) = \int_0^1 f(x,y)dx = \int_0^1 (e^x + y)dy = [e + y - 1]I(y)_{[0,1]}$$

01

(c) Les v.a.r. X et Y sont-elles indépendantes ?
Le domaine Δ de $f_{X,Y}$ n'est pas un pavé donc on ne peut pas avoir $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

$\boxed{\text{Les v.a.r. } X \text{ et } Y \text{ ne sont pas indépendantes}}$

(d) Calculer l'estimateur de Nadaraya-Watson de la function $r(x) = \mathbb{E}(Y/X = x)$

01

$$r(x) = \mathbb{E}(Y/X = x) = \int_0^1 yf(y/x)dy = \int_0^1 y\frac{f(y,x)}{f_X(x)}dy = \int_0^1 y\frac{e^x + y}{[e^x + \frac{1}{2}]}dy = \frac{[\frac{e^x}{2} + \frac{1}{3}]}{[e^x + \frac{1}{2}]}$$

2. Montrer que l'estimateur de Nadaraya-Watson est la solution d'optimization de $L(c)$ pour un x fixé. telque

$$L(c) = \arg\min_{c \in \mathbb{R}} \sum_{i=1}^{n} K\left(\frac{X - x_i}{h}\right)(y_i - c)^2$$

01

La dérrivé partial

$$\frac{\partial L(c)}{\partial c} = 0 \Longrightarrow c = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)} = \widehat{r}_n(X)$$

Exercice 03 : ____ (08 pts)

Soit $X_1, ..., X_n$ un échantillon issu de la loi de fonction de répartition inconnue F .

I— On note F_n la fonction de répartition empirique associée au $X_1, ..., X_n$.

01 (a) 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 $F(x)(1 - F(x)) \neq 0$

$\sqrt{n}(F_n(x) - F(x)) \longrightarrow^{Loi} N(0, F(x)(1 - F(x)))$.

01 (b) Calculer $E[(F_n(x) - F(x))^2]$ pour un élément x fixé dans R.

$\mathbb{E}(F_n(x)) - F(x) = 0$ et $Var(F_n(x)) = \frac{F(x)(1-F(x))}{n}$

$\mathbb{E}[(F_n(x) - F(x))^2] = \frac{F(x)(1-F(x))}{n}$

01 (c) Déduire que $F_n(x)$ converge en moyenne quadratique vers $F(x)$ lorsque n tend vers l'infini. Donner la vitesse de convergence.

$\mathbb{E}(F_n(x)) - F(x) = 0$ et $Var(F_n(x)) = \frac{F(x)(1-F(x))}{n} \longrightarrow_{n\longrightarrow\infty} 0$

$\mathbb{E}[(F_n(x) - F(x))^2] \longrightarrow_{n\longrightarrow\infty} 0$ la convergence en moyen quadratique. la vitesse de convergence: on a $\mathbb{E}[(F_n(x) - F(x))^2] = O(\frac{1}{n})$ donc la vitesse de convergence est $\frac{1}{n} = n^{-1}$

II— Le but de cet partie est d'estimer la fonction de survie $S(x) = 1 - F(x)$ telque $x > 0$ est fixé.

01 (a) Par la méthode empirique, proposez un estimateur $\widehat{S}_n(.)$ de $S(.)$?

Comme $S(x) = 1 - F(x) = P(X > x)$ d'après le cours $\widehat{S}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>x\}}$

01 (b) Quelle est la loi limite de $\sqrt{n}(S_n(x) - S(x))$ pour un élément x fixé dans R.

On applique le TLC sur $\mathbb{1}_{\{X_i>x\}} \sim$ *Bernoulli(S(x)* et on trouve $\sqrt{n}(S_n(x)-S(x)) \longrightarrow^{Loi} N(0, S(x)(1 - S(x)))$.

01 (c) Calculer $E[(S_n(x) - S(x))^2]$ pour un élément x fixé dans R.

$E[(S_n(x) - S(x))^2] = \mathbb{E}[(F_n(x) - F(x))^2]$ puisque $S(x) = 1 - F(x)$

III— Soit la noyau intégré définie par:

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

01 **a-** Déterminer $H(u)$ pour K soit bien un noyau biweight ($K(u) = \frac{15}{16}(1 - u^2)^2$. $I_{\{-1\leq u\leq 1\}}$

par intégration

$$H(x) = \begin{cases} 0, & \text{if } u \leq -1; \\ \frac{1}{16}(8 + 15u - 10u^3 + 3u^5), & \text{if } -1 \leq u \leq 1; \\ 1, & \text{if } u \geq 1; \end{cases}$$

b- Donner l'expression générale d'un estimateur à noyau de la densité de probabilité et Montrer que cette exprèssion est une densité de probabilité ?

0.5 $\widehat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X-x_i}{h}\right)$. Sous les condition générale du noyau K on a

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

c- Determiner l'estimateur de F en intégrant l'estimation par noyau de la densite f *ie* :

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

0.5

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

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

e- Déterminer $\mathbb{E}(\widehat{f}_n(x)) - f(x)$. ? et déduire que

01

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

sous les conditions général du noyau K on a

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

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

$\mathbb{E}(\widehat{F}_n(x)) - F(x) \longrightarrow_{h\to 0} 0$ donc $\widehat{F}_n$ est sans biais,

Fin