corrigée Examen terminal Stat.nonparametrique MASTER II : Mathématiques .Appliquée& Statistique.)

Exercice 01 : (5.5 pts)

1. Quelles est la différence entre le test paramétrique et le test non paramétrique.? donner des exemples des deux tests?
   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. **des exemples : parametrique** (1- test de student 2-test de ficher) **nonparametrique** (par exemple 1-test de signe 2-test de kolmogorov)

2. Enoncer la méthode d'estimation de densité par histogramme.? Quelles sont les conditions retenues pour l'estimation d'une densité par l'histogramme.?
   L'histogramme est 1' estimateur non paramétrique de la densité le plus ancien et le plus simple, il est considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini. Cette méthode consiste à estimer la densité f en un point x par la proportion des variables aléatoires $(X_1, X_2, ..., X_n)$, qui se trouvent dans un intervalle de longueur d'un paramètre de lissage h et qui contient x. Elle est donc basé sur le choix d'un point d'origine a_0 et d'une partition $A_k = ([a_k, a_{k+1}[)$, en P intervalles du support de X. Si nous notons n_k le nombre de variables dans la classe $[a_k, a_{k+1}[$ et $h = a_{k+1} - a_k$, l'estimateur de f sur $[a_k, a_{k+1}[$ du type histogramme est $f_n(x) = \frac{n_k}{nh} = \frac{\sum_{i=1}^{n} \mathbb{1}_{[a_k,a_{k+1}[}(X_i)}{nh}$ pour $x \in [a_k, a_{k+1}[$
   parmis les inconvinients de cet estimateur **est la condition et le choix d'un point d'origine et le pas ou bien le paramètre de lissage pour chaque classe doit ètre finie et petit.**

3. Enoncer le principe du test de Kruskal Wallis ?

4. Enoncer le but et le principe d'une procédure de type validation croisée?
   La méthode de validation croisée est une méthode de type moindre carrée permet d'obtenire le paramètre de lissage basé sur le fait qu'une partie de l'échantillon est utilisé pour obtenir l'information sur une autre partie Elle consiste à choisir la valeur de h qui minimise l'expression de EQM avec l'utilisation $\sum_{i=1}^{n} f^n_{-i}(x)$ la densité de l'échantillon $X_1, X_1, ..., X_{i-1}, X_{i+1}..., X_n$ avec l'élimination de l'observation X_i.

5. Est ce que on peut estimer la densité de probabilité par la dérivé de la fonction de répartition empérique ?justifier?

   Non , fonction discontinue.

Exercice 02 : (6.5 pts)

• 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.Notons que la noyau intégré définie par: $H(u) = \int_{-\infty}^{u} K(t)dt$. ou K(t) est un noyau sommatif .

1. Rappeler la définition d'un noyau sommatif ?

Soit un noyau statistique K est une densité de probabilité. vérifie

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

2. Est ce que en peut supposer que $H(u)$ est une fonction de répartition.? justifier ?
   Oui en peut supposer que $H(u)$ est une fonction de répartition. puisque K(u) est noyau(densité) est l'intégrale d'une densité de probabilité est une fonction de répartition.

3. 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

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

4. Déterminer $H(u)$ pour K soit bien un noyau biweight $(K(u) = 15/16(1-u^2)^2_{I_{\{-1 \leq u \leq 1\}}}$

   $(K(u) = 15/16(1-u^2)^2_{I_{\{-1 \leq u \leq 1\}}}$ par intégration

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

5. Calculer $\int K^2(u)du$, $\int u^2 K(u)du$ pour un noyau biweight .

$$\int_{-1}^{1} K^2(u) = \int_{-1}^{1} (15/16(1-u^2)^2)^2 du = \frac{225}{256} \times \frac{256}{315} = \frac{5}{7}$$

$$\int_{-1}^{1} u^2 K(u)dudu = \int_{-1}^{1} u^2(15/16(1-u^2)^2)du = \frac{1}{7}$$

6. Rappeler la formule de $\widehat{f}_n$ l'estimateur à noyau et montrer que elle est une function 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.

7. 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.

Exercice 03 : ____(07 pts)

On Considère la fonction de densité de probabilité conjointe suivante .

$$f(x,y) = \begin{cases} 3\exp(-x-3y), & \text{if } \mathbb{I}_{]0,+\infty[}(x) \text{ and } \mathbb{I}_{]0,+\infty[}(y); \\ 0, & \text{otherwise.} \end{cases}$$

1. Montrer que f est la densité de probabilité du couple(X,Y)?

$$\int_0^\infty \int_0^\infty f(s,t)dsdt = \int_0^\infty \exp(-x)dx \int_0^\infty 3\exp(-3y)dy = 1$$

$x \sim \mathcal{E}(1)$, $y \sim \mathcal{E}(3)$

2. Déterminer la distribution conjointe $F(x,y)$?

$$\bullet F(x,y) = \int_0^x \int_0^y f(s,t)dsdt \implies \boxed{F(x,y) = [1-e^{-x}] \times [1-e^{-3x}]_{[0,+\infty[}(x,y)}$$

3. Déterminer la distribution conjointe $F(x,y)$?

$$\bullet F(x,y) = \int_0^x \int_0^y f(s,t)dsdt \implies \boxed{F(x,y) = [1-e^{-x}] \times [1-e^{-3x}]_{[0,+\infty[}(x,y)}$$

4. Déterminer les densités de probabilités marginales $f_X(x)$ et $f_Y(y)$?

$f_X(x) = \int f(x,y)dy = e^{-x}_{[0,+\infty[}(x)$, $f_Y(y) = \int f(x,y)dx = e^{-3x}_{[0,+\infty[}(y)$

5. Les v.a.r. X et Y sont-elles indépendantes ? $f_{X,Y}(x,y) = f_X(x) \times f_Y(y)$. donc Les v.a.r. X et Y sont indépendantes

6. Calculer P[X>Y] et cov (X,Y)? $\int_0^\infty \left(\int_y^\infty f(x,y)dx\right) dy = \frac{3}{4}$ et cov(X,Y)=0

7. Quelle est la fonction de densité conditionellle de Y sachant $X=x$ et celle de X sachant $Y=y$?

$f_{X/y}(x) = \frac{f(x,y)}{f_Y(y)} = f_X(x)$ et $f_{Y/X}(y) = \frac{f(x,y)}{f_X(x)} = f_Y(y)$ puisque Les v.a.r. X et Y sont indépendantes

8. Déterminer $m(x) = E[Y|X=x]$.?

$$m(x) = \mathbb{E}(Y/X=x) = \int yf(y/x)dy = \int_{1/2}^1 y\frac{f(y,x)}{f_X(x)}dy = \mathbb{E}(Y) = 1/3$$

Exercice 04 : ____(01 pts)

Soit $K(u)$ 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 .