Université Abou Bekr Belkaid – Tlemcen.
Faculté des Sciences.
Département de Mathématiques.

Année Universitaire 2021/2022
2ère année Master.
Semestre 3.

Examen Finale "Statistiques des Processus".

Aucun document n'est autorisé.
Durée : 1h30mn.
16 Janvier 2022.

Exercice 1.

Soit $X_1, ..., X_n$ une suite de variables aléatoires i.i.d de loi de Bernoulli de paramètre p, $p \in]0,1[$.

1. Proposer un estimateur $\hat{p}$ de p.
2. En utilisant l'inégalité de Tchébychef, construire un intervalle de confiance pour p de niveau $1-\alpha$ pour $0 < \alpha < 1$.
3. Déduire de l'inégalité de Hoeffding, un intervalle de confiance pour p de niveau $1-\alpha$.
4. Comparer la précision de ces intervalles de confiances pour $n = 100$ et $\alpha = 5\%$.

- **Rappel** : Soit $\alpha \in]0,1[$, un intervalle de confiance pour un paramètre θ de niveau $1-\alpha$ est un couple d'estimateurs $\left(\underline{\theta}_n, \bar{\theta}_n\right)$ tel que $\mathbb{P}\left(\theta \in \left[\underline{\theta}_n, \bar{\theta}_n\right]\right) \geq 1-\alpha$.
- **Indication** : Pour tout $x \in]0,1[$, $x(1-x) \leq \frac{1}{4}$.

Exercice 2.

On observe un n-échantillon $(X_1, Y_1), ..., (X_n, Y_n)$ de même loi qu'un couple de variables aléatoires réelles (X, Y). On suppose que le vecteur (X, Y) admet pour densité la fonction f telle que pour tout $(x, y) \in \mathbb{R}^2$, $f(x, y) > 0$. On considère $\hat{f}_h$ l'estimateur de f défini pour tout $(x, y) \in \mathbb{R}^2$ par :

$$\hat{f}_h(x,y) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_1}, \frac{y - Y_i}{h_2}\right)$$

où $h = (h_1, h_2)$ avec $h_1 > 0$ et $h_2 > 0$. $K : \mathbb{R}^2 \to \mathbb{R}$ est un noyau tel que :

$$\iint |v|\,|K(u,v)|\,dudv < \infty, \quad \iint |u|^{1/2}\,|K(u,v)|\,dudv < \infty \quad et \quad \|K\|_2^2 < \infty$$

On pose pour tout $(x, y) \in \mathbb{R}^2$: $K_h(x,y) = \frac{1}{h_1h_2} K\left(\frac{x}{h_1}, \frac{y}{h_2}\right)$

1. Montrer que pour $x \in \mathbb{R}$ et $y \in \mathbb{R}$,

$$\mathbb{E}[\hat{f}_h(x,y)] = (K_h * f)(x,y).$$

2. Calculer pour $x \in \mathbb{R}$ et $y \in \mathbb{R}$, $Var\left(\hat{f}_h(x,y)\right)$ en fonction de n, $K_h * f$ et $K_h^2 * f$.
3. Montrer que si f est bornée par M, alors

$$Var\left(\hat{f}_h(x,y)\right) \leqslant \frac{M\|K\|_2^2}{nh_1h_2}.$$

4. Montrer que si f vérifie de plus la propriété suivante : pour tous $(u, v) \in \mathbb{R}^2$ et $(u', v') \in \mathbb{R}^2$
$$|f(u, v) - f(u', v')| \leqslant |u - u'|^{1/2} + |v - v'|,$$
alors le biais de $\hat{f}_h(x, y)$ vérifie pour tout $(x, y) \in \mathbb{R}^2$
$$\left|Biais\left(\hat{f}_h(x, y)\right)\right| \leqslant C_K(h_1^{1/2} + h_2)$$
où C_K est une constante ne dépendant que de K.
5. Déduire une borne pour le risque quadratique ponctuel.
6. On note à présent f_X la densité de X. En utilisant $\hat{f}_h$, proposer un estimateur de f_X. Donner son expression si on suppose qu'il existe deux noyaux F et G telles que pour tous $x \in \mathbb{R}$ et $y \in \mathbb{R}$
$$K(x, y) = F(x)G(y).$$

► **Rappel** : Le produit de convolution des fonctions f_1 et f_2 de deux variables est défini par :
$$\forall (x, y) \in \mathbb{R}^2, (f_1 * f_2)(x, y) = \iint f_1(x - u, y - v) f_2(u, v) du dv$$

Bonne courage.

"Corrigé Epreuve finale"
"Statistique des Processus"

Exercice n°1 :

$X_i \hookrightarrow \mathcal{B}(p)$, $i=\overline{1,n}$ avec X_i i.i.d. $p \in]0,1[$.

1°/ $X_i \hookrightarrow \mathcal{B}(p)$ donc $\mathbb{E}(X_i) = p \quad \forall i = \overline{1,n}$

0,1 Alors, $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$

2°/ Inégalité de Tchébychef :

0,5 $\mathbb{P}(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| > x) \leq \frac{Var(\bar{X}_n)}{x^2}$ pour tout $x \in \mathbb{R}^{*+}$.

- $\mathbb{E}(\bar{X}_n) = p$.
- $Var(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{p(1-p)}{n}$

Donc :

0,5 $$\mathbb{P}(|\bar{X}_n - p| > x) \leq \frac{p(1-p)}{n x^2} \leq \frac{1}{4nx^2}$$

i.e.

$$\mathbb{P}(|\bar{X}_n - p| \leq x) \geqslant 1 - \frac{1}{4nx^2}$$

$$\mathbb{P}(\hat{p} - x \leq p \leq \hat{p} + x) \geqslant 1 - \frac{1}{4nx^2} = 1 - \alpha.$$

0,5 Donc pour $\alpha = \frac{1}{4nx^2} \Leftrightarrow x = \frac{1}{2\sqrt{n\alpha}}$, on aura :

$$\mathbb{P}\left(\hat{p} - \frac{1}{2\sqrt{n\alpha}} \leq p \leq \hat{p} + \frac{1}{2\sqrt{n\alpha}}\right) \geqslant 1 - \alpha$$

3°/ Inégalité de Hoeffding : $S_n = \sum_{i=1}^{n} X_i$

0,5 $\forall x > 0$, $\mathbb{P}(|S_n - \mathbb{E}(S_n)| > x) \leq 2 \exp\left(\frac{-2x^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$.

Comme $X_i \hookrightarrow \mathcal{B}(p)$ alors $0 \leq X_i \leq 1$.

Donc : $\mathbb{P}(|\hat{p} - p| > \frac{x}{n}) \leq 2 e^{\frac{-2x^2}{n}}$. i.e :

0,75 $$\mathbb{P}(\hat{p} - \frac{x}{n} \leq p \leq \hat{p} + \frac{x}{n}) \geqslant 1 - 2e^{\frac{-2x^2}{n}} = 1 - \alpha.$$

Alors pour $\alpha = 2\exp(\frac{-2x^2}{n}) \Leftrightarrow x = \sqrt{\frac{n \ln(2/\alpha)}{2}}$.

on a :

$$\mathbb{P}\left(\hat{p} - \sqrt{\frac{\ln(2/\alpha)}{2n}} \leq p \leq \hat{p} + \sqrt{\frac{\ln(2/\alpha)}{2n}}\right) \geqslant 1 - \alpha$$

4°/ $n = 100$ et $\alpha = 5\%$.

. I.C. Tchebychef :

$$P\left(\hat{p} - \frac{1}{2\sqrt{100 \cdot 0.05}} \leq p \leq \hat{p} + \frac{1}{2\sqrt{100 \cdot 0.05}} \right) \geq 1 - 0.05 = 95\%$$

Donc la précision de l'intervalle est :

$$2 \cdot \frac{1}{2\sqrt{5}} \approx 0.447$$

(02) . I.C. Hoeffding :

$$2 \times \sqrt{\frac{\ln(2/0.05)}{2 \cdot 100}} \approx 0,271$$

On a bien l'inégalité de Hoeffding donne un intervalle de confiance plus précis pour la valeur p.

Exercice n° 2 : $\hat{f}_h(x,y) = \frac{1}{n h_1 h_2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_1}, \frac{y - Y_i}{h_2} \right)$.

1°/ $\forall (x,y) \in \mathbb{R}^2$, $\mathbb{E}\left[\hat{f}_h(x,y) \right] = (K_h * f)(x,y)$.

$$\mathbb{E}\left(\hat{f}_h(x,y) \right) = \frac{1}{n h_1 h_2} \sum_{i=1}^{n} \iint_{\mathbb{R}^2} K\left(\frac{x-u}{h_1}, \frac{y-v}{h_2} \right) f(u,v)\, du\, dv$$

(01)

$$= \iint_{\mathbb{R}^2} K_h(x-u, y-v)\, f(u,v)\, du\, dv$$

$$= (K_h * f)(x,y)$$

2°/ $\mathrm{Var}\left(\hat{f}_h(x,y) \right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left(K_h(x - X_i, y - Y_i) \right)$

$$= \frac{1}{n} \mathbb{E}\left(K_h^2(x-X, y-Y) \right) - \frac{1}{n} \mathbb{E}^2\left(K_h(x-X, y-Y) \right)$$

(01)

$$= \frac{1}{n} (K_h^2 * f)(x,y) - \frac{1}{n} (K_h * f)^2(x,y).$$

3°/ $\mathrm{Var}\left(\hat{f}_h(x,y) \right) \leq \frac{M \|K\|_2^2}{n h_1 h_2}$.

01 $\mathrm{Var}\left(\hat{f}_h(x,y) \right) \leq \frac{1}{n} (K_h^2 * f)(x,y)$

05 $\leq \frac{M}{n h_1^2 h_2^2} \iint_{\mathbb{R}^2} K^2\left(\frac{x-u}{h_1}, \frac{y-v}{h_2} \right) du\, dv$

$(u', v') = \left(\frac{x-u}{h_1}, \frac{y-v}{h_2} \right)$ $\quad \langle du', dv' \rangle = \left(-\frac{du}{h_1}, -\frac{dv}{h_2} \right)$

01 Alors, $\mathrm{Var}\left(\hat{f}_h(x,y) \right) \leq \frac{M}{n h_1 h_2} \iint_{\mathbb{R}^2} K^2(u', v')\, du'\, dv'$

$$= \frac{M \|K\|_2^2}{n h_1 h_2} \quad \text{c.q.f.d}$$

4/ $|\text{biais}(\hat{f}_h(x,y))| = |\mathbb{E}(f_h(x,y)) - f(x,y)|$ (0,1)

$= |(K_h * f)(x,y) - f(x,y)|$.

comme k est le noyau (0,2) $\longrightarrow = \left| \frac{1}{h_1 h_2} \iint_{\mathbb{R}^2} K\left(\frac{x-u}{h_1}, \frac{y-v}{h_2}\right) f(u,v)\,du\,dv - \frac{1}{h_1 h_2} \iint_{\mathbb{R}^2} K\left(\frac{x-u}{h_1}, \frac{y-v}{h_2}\right) f(x,y)\,du\,dv \right|$

(0,5) $\leq \frac{1}{h_1 h_2} \iint_{\mathbb{R}^2} \left|K\left(\frac{x-u}{h_1}, \frac{y-v}{h_2}\right)\right| |f(u,v) - f(x,y)|\,du\,dv$

$\leq \frac{1}{h_1 h_2} \iint_{\mathbb{R}^2} \left|K\left(\frac{x-u}{h_1}, \frac{y-v}{h_2}\right)\right| \left(|u-x|^{1/2} + |v-y|\right) du\,dv$

Après un changement de variable : $(u', v') = \left(\frac{x-u}{h_1}, \frac{y-v}{h_2}\right)$.

(0,1) $|\text{biais}(\hat{f}_h(x,y))| \leq h_1^{1/2} \iint_{\mathbb{R}^2} \underbrace{|K(u',v')|}_{<\infty} |u'|^{1/2}\,du'dv' + h_2 \iint_{\mathbb{R}^2} \underbrace{|K(u',v')|}_{<\infty} |v'|\,du'dv'$

(0,1) $:= C_{1K} h_1^{1/2} + C_{2K} h_2$.

$\leq \max(C_{1,K}, C_{2K})(h_1^{1/2} + h_2)$

$\| \quad C_K \cdot (h_1^{1/2} + h_2)$ c.q.f.d

5/ En utilisant successivement la décomposition "biais au carré - variance", les bornes obtenues dans les questions 4 et 5 et $(a+b)^2 \leq 2(a^2+b^2)$ (0,2) par tout $a, b \geq 0$, nous obtenons :

(0,1) $MSE_{\hat{f}_h}(h,x) := \text{biais}^2 + \text{variance}$.

$\leq \max\left(2C_K^2, M\|K\|_2^2\right)\left(h_1 + h_2^2 + \frac{1}{nh_1h_2}\right)$ (0,1)

6/ f_X densité de X.

$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy$. donc,

$\hat{f}_X(x) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} \int_{\mathbb{R}} K\left(\frac{x-X_i}{h_1}, \frac{y-Y_i}{h_2}\right) dy$.

(0,2) Si $\forall (x,y) \in \mathbb{R}^2$, $K(x,y) = F(x) \cdot G(y)$ alors :

$\hat{f}_X(x) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} \int_{\mathbb{R}} F\left(\frac{x-X_i}{h_1}\right) G\left(\frac{y-Y_i}{h_2}\right) dy$

$= \frac{1}{nh_1h_2} \sum_{i=1}^{n} F\left(\frac{x-X_i}{h_1}\right) \int_{\mathbb{R}} G(u) \cdot h_2\,du$

$\hat{f}_X(x) = \frac{1}{nh_1} \sum_{i=1}^{n} F\left(\frac{x-X_i}{h_1}\right)$