Université de Sidi Bel Abbès
Faculté des Sciences
Département de Mathématique

Année 2011/2012.
Module: Statistique et application.
3-ème cycle.

Concours de Statistique Non Paramétrique

Solution du Problème

1. La démonstration repose sur les résultats suivants

$$\mathbb{E}f_n(x) - f(x) = O(h_n^k). \tag{1}$$

$$f_n(x) - \mathbb{E}f_n(x) = O\left(\sqrt{\frac{\log n}{nh_n}}\right), p.co. \tag{2}$$

Concernant l'équation (1) on a

$$\mathbb{E}f_n(x) = \frac{1}{h_n}\mathbb{E}K^{(1)}\left(\frac{x-X}{h_n}\right) = \frac{1}{h_n}\int K^{(1)}\left(\frac{x-u}{h_n}\right)f(u)du.$$

Pour calculer cette intégrale on pose $z = (x-u)/h_n$ pour arriver à:

$$\mathbb{E}f_n(x) = \int K^{(1)}(z)f(x - zh_n)dz.$$

En utilisant l'hypothèse de f (qu'elle est strictement positive, et de classe C^k au voisinage de x), il suffit de développer la fonction f au voisinage de x.

Ceci s'écrit, pour θ_z entre x et $x + zh_n$ on a:

$$f(x - zh_n) = f(x) + \sum_{j=1}^{k-1} \frac{(-1)^j (zh_n)^j}{j!} f^{(j)}(x) + \frac{(-1)^k (zh_n)^k}{k!} f^{(k)}(\theta_z).$$

D'après l'hypothèse faite sur le noyau K on aboutit à

$$\mathbb{E}f_n(x) = f(x) + (-1)^k h_n^k \frac{\int z^k K'(z) f^{(k)}(\theta_z)dz}{k!}.$$

La continuité de f^k et la compacité du support compact de K' assurent la convergence uniforme en z de $f^{(k)}(\theta_z)$ vers $f^k(x)$, ainsi

$$\mathbb{E}f_n(x) = f(x) + (-1)^k h_n^k \int z^k K'(z)dz \frac{f^k(x)}{k!} + O(h_n^k).$$

Concernant l'équation (2), l'idée essentielle est d'utiliser une inégalité exponentielle de type Bernstein dont on applique cette inégalité aux variables:

$$\Delta_i = \frac{1}{h_n}\left[K^{(1)}\left(\frac{x-X_i}{h_n}\right) - \mathbb{E}K^{(1)}\left(\frac{x-X_i}{h_n}\right)\right]$$

pour cela il faut majorer $|\Delta_i|$ ainsi que $\mathbb{E}\Delta_i^2$.

L'hypothèse faite sur le noyau K permet de construire la première borne, ainsi $\exists C$ une constante finie telle que:

$$|\Delta_i| \leq \frac{C}{h_n}.$$

Posons

$$\Gamma_i = \frac{1}{h_n}K^{(1)}\left(\frac{x-X_i}{h_n}\right)$$

calculons le moment d'ordre 2

$$\mathbb{E}\Gamma_i^2 = \frac{1}{h_n}\mathbb{E}\left(\frac{1}{h_n}K'^2\left(\frac{x-X_i}{h_n}\right)\right) = \frac{1}{h_n^2}\int K'^2\left(\frac{x-u}{h_n}\right)f(u)du$$

on pose $z = (x-u)/h_n$ pour aboutir à

$$\mathbb{E}\Gamma_i^2 = \frac{1}{h_n}\int K'^2(z)f(x-zh_n)dz.$$

Puisque f est bornée car continue sur le support compact de K, on a l'existence d'une constante finie C telle que:

$$\mathbb{E}\Gamma_i^2 \leq \frac{C}{h_n}$$

d'une manière évidente on a l'existence d'une constante finie C telle que :

$$\mathbb{E}\Delta_i^2 \leq \frac{C}{h_n}.$$

Ainsi, pour ε suffisament petit on a:

$$\mathbb{P}\left(|\mathbb{E}f_n(x) - f_n(x)| > \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2 h_n}{4C}\right) \tag{3}$$

On applique (3) à $\varepsilon = \varepsilon_0\sqrt{\frac{\log n}{nh_n}}$, on aura pour tout ε_0 il existe une constante $C > 0$ telle que

$$\mathbb{P}\left(|\mathbb{E}f_n(x) - f_n(x)| > \varepsilon_0\sqrt{\frac{\log n}{nh_n}}\right)] \leq 2\exp\left(-C\varepsilon_0^2\log n\right).$$

Pour ε_0 bien choisi, le terme à droite est celui d'une série convergente.

Concernant la preuve de $\mathbb{P}\left(f_n(x) < \delta\right) < \infty$, les résultats (1) et (2) entraînent en particulier la convergence presque complète de $f_n(x)$ vers $f(x)$. Ainsi pour tout $\varepsilon > 0$ on a

$$\sum_{i\geq 1} \mathbb{P}\left(|f_n(x) - f(x)| > \varepsilon\right) < \infty.$$

Il suffit maintenant de poser $\delta = f(x)/2$ (ce qui est toujours possible puisque $f(x) > 0$) pour obtenir $\mathbb{P}\left(f_n(x) < \delta\right) < \infty$.

Ceci achève la preuve du premier résultat.

2. La preuve repose sur les résultats suivants

$$\mathbb{E}F_n(x) - F(x) = O(h_n^k). \tag{4}$$

$$F_n(x) - \mathbb{E}F_n(x) = O\left(\sqrt{\frac{\log n}{nh_n}}\right), p.co. \tag{5}$$

Concernant l'équation (4) La preuve est similaire à celle de la preuve de l'équation (1) en remplaçant f par F et $\mathbb{E}f_n$ par $\mathbb{E}F_n$

Concernant l'équation (5) la preuve est basée sur les mêmes arguments de la preuve de l'équation (2),tout en posant

$$\Delta_i = K\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}K\left(\frac{x - X_i}{h_n}\right)$$

comme précédement sous l'hypothèse faite sur le noyau K, on arrive à l'existence d'une constante finie $C_1 > 0$ telle que

$$|\Delta_i| \leq C_1$$

posons

$$Y_i = K\left(\frac{x - X_i}{h_n}\right)$$

et calculons le moment d'ordre 2 de Y_i

$$\mathbb{E}Y_i^2 = \mathbb{E}\left(K^2\left(\frac{x - X_i}{h_n}\right)\right) = \int K^2\left(\frac{x - u}{h_n}\right) f(u)du$$

on effectue le changement de variable suivant $z = (x - u)/h_n$ pour aboutir à:

$$\mathbb{E}Y_i^2 = \int K^2(z) f(x - zh_n)dz.$$

f est bornée (continue sur le support compact de K), donc il existe une constante finie $C_1 > 0$ telle que:

$$\mathbb{E}Y_i^2 \leq C_1.$$

par suite

$$\mathbb{E}\Delta_i^2 \leq C_1$$

l'application de l'inégalité de type Bernestein, pour ε suffisament petit nous donne

$$\mathbb{P}\left(|\mathbb{E}F_n(x) - F_n(x)| > \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{4C_1}\right)$$

pour $\varepsilon = \varepsilon_0\sqrt{\frac{\log n}{n}}$, on aura pour tout ε_0 il existe une constante $C_1 > 0$ telle que:

$$\mathbb{P}\left(|\mathbb{E}F_n(x) - F_n(x)| > \varepsilon_0\sqrt{\frac{\log n}{n}}\right) \leq 2\exp\left(-C\varepsilon_0^2\log n\right).$$

ainsi pour ε_0 bien choisi, le terme à droite est celui d'une série convergente, et cela achève la preuve du résultat 2.

3. Nous avons déjà démontré la convergence presque complète de $F_n(x)$ vers $F(x)$. Autrement dit, nous avons

$$\forall\varepsilon > 0 \qquad \sum_{n=1}^{\infty}\mathbb{P}\{|F_n(x) - F(x)| > \varepsilon\} < \infty.$$

D'autre part, on a par hypothèse $F(x) < 1$, c'est à dire

$$1 - F_n(x) \geq F(x) - F_n(x),$$

ainsi,

$$\inf_{x\in\mathbb{R}}|1 - F_n(x)| \leq (1 - \sup_{x\in\mathbb{R}}F(x))/2 \Rightarrow \sup_{x\in\mathbb{R}}|F_n(x) - F(x)| \geq (1 - \sup_{x\in\mathbb{R}}F(x))/2.$$

En terme de probabilité, on obtient

$$\mathbb{P}\{\inf_{x\in\mathbb{R}}|1 - F_n(x)| < \delta\} \leq \mathbb{P}\{\sup_{x\in\mathbb{R}}|F_n(x) - F(x)| \geq (1 - \sup_{x\in\mathbb{R}}F(x))/2\} < \infty.$$

Finalement, il suffit de prendre $\delta = (1 - \sup_{x\in\mathbb{R}}F(x))/2$ pour achever la démonstration du résultat 3.