Solution du Problème.

(1) (i) *Etant donné $X_1, X_2, \ldots, X_n$ une suite de variable aléatoire réelle de même loi que X, l'estimateur de la densité par la méthode du noyau noté $f_n(x)$, défini par:*

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \forall x \in \mathbb{R}$$

1 pt

où K est le noyau de cet estimateur et h est la fenêtre.

(ii) *Commençons par calculer le biais : en faisant le changement de variable suivant*

$$y = x + uh, \quad dy = hdu$$

$$\begin{aligned}
\mathbb{E}\left(f_n(x)\right) &= \frac{1}{nh} \sum_{i=1}^{n} \mathbb{E}\left[K\left(\frac{X_i - x}{h}\right)\right] \\
&= \frac{1}{nh} \sum_{i=1}^{n} \int_{\mathbb{R}} K\left(\frac{y - x}{h}\right) f(y) dy \\
&= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{y - x}{h}\right) f(y) dy \\
&= \int_{\mathbb{R}} K(u) f(x + uh) du.
\end{aligned}$$

En effectuant un dévelopement limité à l'ordre 2, avec $\zeta_u \in [x, x + uh]$, il vient

$$\begin{aligned}
\mathbb{E}\left(f_n(x)\right) &= \int_{\mathbb{R}} K(u) f(x + uh) du \\
&= \int_{\mathbb{R}} K(u)[f(x) + (uh) f'(x) + \frac{(uh)^2}{2} f''(\zeta_u)] du \\
&= f(x) \underbrace{\int_{\mathbb{R}} K(u) du}_{=1} + hf'(x) \underbrace{\int_{\mathbb{R}} uK(u) du}_{=0} + \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) f''(\zeta_u) du. \\
&= f(x) \underbrace{\int_{\mathbb{R}} K(u) du}_{=1} + hf'(x) \underbrace{\int_{\mathbb{R}} uK(u) du}_{=0} + \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) f''(x) + O(h^2) du. \\
&= \frac{h^2}{2} f''(x) \int_{\mathbb{R}} u^2 K(u) du + O(h^2).
\end{aligned}$$

2 pts

(iii) *Il en résulte que*

$$\begin{aligned}
|Biais\,(f_n(x))| &= |\mathbb{E}[f_n(x)] - f(x)| \\
&\leq \frac{h^2}{2} \left| \int_{\mathbb{R}} u^2 K(u) f''(\zeta_u) du \right| \\
&\leq \frac{h^2}{2} \int_{\mathbb{R}} u^2 |K(u)| |f''(\zeta_u)| du \\
&\leq h^2 \underbrace{\frac{\max_x |f''(x)|}{2} \int_{\mathbb{R}} u^2 |K(u)| du}_{C_1}
\end{aligned}$$

2 pts

d'où la première partie.

(iv) *Pour prouver la seconde partie, on utilise le faite que les variables aléatoires $Y_i = K\left((X_i - x)/h\right)$, $i = 1 \ldots, n$ sont i.i.d. et que la variance de la somme de variables* *indépendantes coïncide avec la somme des variances :*

$$\begin{aligned}
Var\left[f_n(x)\right] &= \frac{1}{(nh)^2} Var\left[\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\right] \\
&= \frac{1}{(nh)^2} \sum_{i=1}^{n} Var\left[K\left(\frac{X_i - x}{h}\right)\right] \\
&= \frac{1}{(nh)^2} \times n \times Var\left[K\left(\frac{X_1 - x}{h}\right)\right] \\
&\leq \frac{1}{(nh)^2} \mathbb{E}\left[K\left(\frac{X_1 - x}{h}\right)^2\right] \\
&= \frac{1}{(nh)^2} \int_{\mathbb{R}} K^2\left(\frac{y - x}{h}\right) f(y)dy \\
&= \frac{1}{nh} \int_{\mathbb{R}} K^2(u) f(x - uh)du \\
&\leq \frac{1}{nh} \underbrace{\sup_{z\in\mathbb{R}} f(z) \int_{\mathbb{R}} K^2(u)du}_{C_2}.
\end{aligned}$$

2 pts

C'est exactement ce qu'il fallait démontrer. 2 pts

(2) *Soit X une variable aléatoire définie sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, à valeur dans $\mathbb{R}$. Pour tout $x \in \mathbb{R}$, on désigne par F la fonction de répartition de X, et par f la fonction de densité.*

Etant donné $X_1, X_2, \ldots, X_n$ une suite de variable aléatoire réelle de même loi que X, l'estimateur de la fonction de répartition par la méthode du noyau noté $F_n(x)$, défini par:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \qquad \forall x \in \mathbb{R}$$

où K est noyau et λ_n est une suite de réels positifs. On déduit de F_n un estimateur de la fonction de répartition et de la densité, noté f_n, défini par

$$\begin{aligned}
f_n(x) &= F_n^{(1)}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K^{(1)}\left(\frac{x - X_i}{h_n}\right) \\
&= \frac{1}{nh_n} \sum_{i=1}^{n} K^{(1)}\left(\frac{x - X_i}{h_n}\right)
\end{aligned}$$

(i) *En vertu de sa définition, l'estimateur natural de la fonction de hasard, noté $\lambda_n(x)$, est définie par:*

$$\lambda_n(x) = \frac{f_n(x)}{1 - F_n(x)} = \frac{\sum_{i=1}^{n} K^{(1)}\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^{n} \int_{x}^{+\infty} K^{(1)}\left(\frac{t - X_i}{h_n}\right) dt}, \qquad \forall x \in \mathbb{R}$$

1 pts

(ii) *On considère la décomposition suivante*

$$\begin{aligned} \lambda_n(x) - \lambda(x) &= \frac{f_n(x)}{1-F_n(x)} - \frac{f(x)}{1-F(x)} \\ &= \frac{f_n(x) - f_n(x)F(x) - f(x) + f(x)F_n(x)}{(1-F_n(x))(1-F(x))} \\ &= \frac{1}{1-F_n(x)}\left[(f_n(x)-f(x)) + \frac{f(x)}{1-F(x)}(F_n(x)-F(x))\right]. \end{aligned} \tag{1}$$

2 pts

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

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

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,

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

En terme de probabilité, on obtient

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

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

2 pts

(iv) *On démontre la convergence presque complète uniforme sur un compact réel de $\lambda_n(x)$ vers $\lambda(x)$, autrement dit*

$$\sup_{x\in S} |\lambda_n(x) - \lambda(x)| = O(h_n^k) + O\left(\sqrt{\frac{\log n}{nh_n}}\right) \qquad p.co. \tag{2}$$

En utilisant la décomposition (1) on peut déduire que,

$$\sup_{x\in S} |\lambda_n(x) - \lambda(x)| \leq$$

$$\frac{1}{\inf_{x\in S} |1 - F_n(x)|}\left[\sup_{x\in S} |f_n(x) - f(x)| + \frac{\sup_{x\in S} |f(x)|}{\inf_{x\in S} |1 - F(x)|} \sup_{x\in S} |F_n(x) - F(x)|\right]$$

Ainsi, la démonstration de (2) repose sur les résultats suivants

$$\sup_{x\in S} |F_n(x) - F(x)| = O(h_n^k) + O\left(\sqrt{\frac{\log n}{n}}\right), \qquad p.co \tag{3}$$

3 pts

$$\sup_{x\in S} |f_n(x) - f(x)| = O(h_n^k) + O\left(\sqrt{\frac{\log n}{nh_n}}\right), \qquad p.co \tag{4}$$

3 pts

$$\exists \delta > 0 \quad telque \quad \sum_{n=1}^{\infty} \mathbb{P}\left\{\inf_{x\in S} |1 - F_n(x)| < \delta\right\} < \infty.$$

de même la preuve de (3) (resp. (4)) est basée respectivement sur les décompositions suivantes

$$F_n(x) - F(x) = F_n(x) - \mathbb{E}[F_n(x)] + \mathbb{E}[F_n(x)] - F(x)$$

$$f_n(x) - f(x) = f_n(x) - \mathbb{E}[f_n(x)] + \mathbb{E}[f_n(x)] - f(x)$$

La démonstration de (4) repose sur les résultats suivants

(5) $$\mathbb{E}[f_n(x)] - f(x) = O(h_n^k).$$

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

Concernant l'équation (5) on a

$$\mathbb{E}[f_n(x)] = \mathbb{E}\left[\frac{1}{nh_n}\sum_{i=1}^{n} K^{(1)}\left(\frac{x-X_i}{h_n}\right)\right] = \frac{1}{nh_n}\sum_{i=1}^{n}\mathbb{E}\left[K^{(1)}\left(\frac{x-X_i}{h_n}\right)\right]$$

comme les variables X_i sont i.i.d alors

$$\mathbb{E}\left[K^{(1)}\left(\frac{x-X_1}{h_n}\right)\right] = \mathbb{E}\left[K^{(1)}\left(\frac{x-X_2}{h_n}\right)\right] = \ldots = \mathbb{E}\left[K^{(1)}\left(\frac{x-X_i}{h_n}\right)\right]$$

donc

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

Pour calculer cette intégrale on considére le changement des variables on pose $z = (x - u)/h_n \Rightarrow dz = -du/h_n \Rightarrow du = -h_n dz$ et $u = x - zh_n$ pour arriver à:

$$\begin{aligned}\mathbb{E}[f_n(x)] &= -\int_{+\infty}^{-\infty} K^{(1)}(z) f(x - zh_n)dz \\ &= \int_{-\infty}^{+\infty} K^{(1)}(z) f(x - zh_n)dz.\end{aligned}$$

En utilisant l'hypothèse que la densité est de classe C^k, il suffit de développer f au voisinage de x (D.L.T).
Ceci s'écrit, pour θ_z entre x et $x + zh_n$ on a:

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

Puisque K est un noyau de densité borné intégrable et d'ordre k on aboutit à

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

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 (6), on applique l'inégalité de type Bernestein aux variables:

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

pour cela il faut majorer $|\Delta_i|$ ainsi que $\mathbb{E}[\Delta_i^2]$.
Le fait que K est un noyau de densité borné intégrable et d'ordre k, ça nous 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:*

$$\mathbb{E}[\Gamma_i^2] \le \frac{C}{h_n}$$

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

$$\mathbb{E}[\Delta_i^2] \le \frac{C}{h_n}$$

Ainsi, pour ϵ *suffisamment petit on a:*

$$\mathbb{P}\left(|f_n(x) - \mathbb{E}[f_n(x)]| > \epsilon\right) \le 2\exp\left(-\frac{n\epsilon^2}{4C}\right) \tag{7}$$

On applique (7) à $\epsilon = \epsilon_0\sqrt{\frac{\log n}{nh_n}}$*, on aura pour tout* ϵ_0, $\exists\ C > 0$*:*

$$\mathbb{P}\left(|f_n(x) - \mathbb{E}[f_n(x)]| > \epsilon_0\sqrt{\frac{\log n}{nh_n}}\right) \le 2\exp(-C\epsilon_0^2\log n).$$

Pour ϵ_0 *bien choisi, le terme à droite est celui d'une série convergente.*
La preuve (3) repose sur les résultats suivants

$$\mathbb{E}[F_n(x)] - F(x) = O(h_n^k). \tag{8}$$

$$F_n(x) - \mathbb{E}[F_n(x)] = O\left(\sqrt{\frac{\log n}{nh_n}}\right), \qquad p.co. \tag{9}$$

Concernant l'équation (8) La preuve est similaire à celle de la preuve de l'équation (5) en remplaçant f *par* F *et* $\mathbb{E}[f_n]$ *par* $\mathbb{E}[F_n]$.
Concernant l'équation (9) la preuve est basée sur les mêmes arguments de la preuve de l'équation (6), tout en posant

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

comme précédement sous les hypothèses du noyau K *on arrive à l'existence d'une constante finie* $C_1 > 0$ *telle que*

$$|\Delta_i| \le C_1$$

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 *suffisamment petit nous donne*

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

pour $\epsilon = \epsilon_0\sqrt{\frac{\log n}{n}}$, *on aura pour tout* ϵ_0 *il existe une constante* $C_1 > 0$ *telle que:*

$$\mathbb{P}\left(|F_n(x) - \mathbb{E}[F_n(x)]| > \epsilon_0\sqrt{\frac{\log n}{n}}\right) \leq 2\exp(-C\epsilon_0^2 \log n).$$

Ainsi pour ϵ_0 *bien choisi, le terme à droite est celui d'une série convergente, et celà achève la preuve.*