Estimation semi-paramétrique de Quasi-Score

Michel Bonneu Michel Gba

Abstract

The method of maximum quasi-likelihood estimation gives sa-tisfactory results in a parametric regression model, where the link function r and the variance function V are well specified. In semiparametric models, when the functions r and V are unknown, this method fails. Nevertheless, it is possible to define the quasi-score function and its estimation, computed from kernel regression estimators of the functions r and V. We propose an estimator for the regression coefficients based on a one step Newton-Raphson iteration in a maximum quasi-likelihood optimization starting from an initial $\sqrt{n}$-consistent estimate and using the estimated quasi score. We derive the asymptotic properties of this estimator and its semi parametric efficiency.

Résumé

Dans le cadre paramétrique, la méthode d'estimation du ma-ximum de quasi-vraisemblance donne des résultats satisfaisants pour des modèles de régression où la fonction de lien r et la fonction de variance V sont correctement spécifiées. Dans un cadre semi-paramétrique, il n'en est plus de même, quand les fonctions r et V sont inconnues. Néanmoins il est possible de définir la fonction de quasi-score, ainsi que son estimation calculée à partir d'estimateurs non paramétriques à noyau des deux fonctions r et V inconnues. Nous proposons un estimateur des coefficients de régression basé sur la méthode de Newton-Raphson à un pas, dans l'optimisation du maximum de la quasi-vraisemblance, à partir d'un estimateur semi-paramétrique initial $\sqrt{n}$-consistant en utilisant le quasi-score estimé. Nous étudions les propriétés aymptotiques de cet estimateur, et prouvons notamment qu'il est asymptotiquement efficace au sens semi-paramétrique.

Received by the editors February 1997.

Communicated by M. Hallin.

Key words and phrases : modèles à direction révélatrice unique, quasi-vraisemblance, estimateur à noyau, quasi-score, estimation semi-paramétrique, efficacité semi-paramétrique.

1 Introduction

On considère n vecteurs aléatoires $Z_1, Z_2, ..., Z_n$, indépendants et identiquement distribués, à valeurs dans un espace $\mathcal{Z}$, de même loi de probabilité P_o inconnue qui admet une densité par rapport à une mesure σ-finie. Soit $\mathcal{P}$ un ensemble suffisamment grand de mesures de probabilités contenant P_o. Dans le cadre de la statistique semi-paramétrique, on considère la paramétrisation $(\theta, g) \longrightarrow P_{(\theta,g)} \in \mathcal{P}$ avec :

- un paramètre d'intérêt θ appartenant à un sous-espace paramétrique Θ euclidien de dimension finie,
- un paramètre fonctionnel $g \in \mathcal{G}$, dit paramètre de nuisance, où $\mathcal{G}$ est un ensemble fonctionnel spécifié pour le modèle statistique considéré.

Nous considérons un modèle de régression pour les variables $Z_i = (Y_i, X_i)$ à valeurs dans $R \times R^p$, dont la loi de probabilité P_o n'est connue que partiellement, à travers les deux hypothèses suivantes H_1 et H_2.

H_1 **:** Il existe une fonction $r_o \in L_2(I)$, où I est une partie de R , et un vecteur $\theta_o \in \Theta \subset R^p$ tels que :

$$E\left(Y_i|X_i\right) = r_o(< X_i, \theta_o >).$$

H_2 **:** Il existe une fonction V_o strictement positive telle que :

$$Var\left(Y_i|X_i\right) = V_o\left(E(Y_i|X_i)\right).$$

Dans la formulation de H_1 et H_2 et dans tout ce qui suivra les espérances sont relatives à la loi P_o et on notera $x\theta$ le produit scalaire $< x, \theta >$.

Le modèle de régression défini par H_1 est appelé Single Index Model (S.I.M) ou modèle à direction révélatrice unique. Dans cette famille de modèles définis par H_1, les modèles linéaires généralisés sont un cas particulier, car ils sont définis en plus de l'hypothèse H_1, par une deuxième hypothèse supplémentaire qui stipule que la loi de $Y_i|X_i$ appartient à la famille exponentielle linéaire. Ce qui implique que la variance conditionnelle de $Y_i|X_i$ est une fonction de l'espérance conditionnelle $\mu_i = E\left(Y_i|X_i\right)$. Cette propriété constitue la deuxième hypothèse H_2. Sans perdre de généralité, l'hypothèse H_2 peut aussi s'écrire :

$$Var\left(Y_i|X_i\right) = \sigma^2 V_o(\mu_i)$$

où σ^2 est un autre paramètre réel de nuisance. Dans la suite, on considère que $\sigma^2 = 1$. L'hypothèse H_2 pourra être relaxée par l'hypothèse $H_2^{'}$ suivante qui est moins restrictive.

$H_2^{'}$: Il existe une fonction $w_o \in L_2(I)$, strictement positive telle que

$$Var(Y_i|X_i) = Var(Y_i|X_i\theta_o) = w_o(X_i\theta_o).$$

Les hypothèses H_2 et $H_2^{'}$ sont équivalentes quand la fonction r_o est bijective. L'objectif, dans ces modèles de régression semi-paramétriques est d'estimer le paramètre d'intérêt θ_o en présence du paramètre fonctionnel de nuisance $g = (r_o, V_o)$ ou $g = (r_o, w_o)$. Dans un premier paragraphe, nous rappelons la méthodologie de l'estimation de quasi-vraisemblance dans le cas paramétrique. Dans ce cas "purement" paramétrique, où g est connue, on précise les propriétés de tels estimateurs. Trois exemples sont détaillés dans la suite afin d'illustrer de tels modèles de régression relatifs à l'hypothèse $\mathbf{H_2}$ ou $\mathbf{H_2'}$. Le point de vue semi-paramétrique est considéré dans le cas où les paramètres fonctionnels du modèle sont inconnus, notamment quand la fonction qui lie la moyenne au prédicteur linéaire est inconnue. Dans le paragraphe 3, on présente les estimateurs des coefficients de régression quand la fonction variance, qui lie la variance à la moyenne, est connue ou inconnue. Quand le paramètre (r_o, w_o) est inconnu, on propose un estimateur en deux étapes défini à partir du quasi-score. Le paragraphe 4 est consacré aux propriétés asymptotiques de cet estimateur.

2 La quasi-vraisemblance

2.1 Point de vue paramétrique

Dans le cadre paramétrique, le modèle est paramétré par $\theta \in \Theta$, défini par les hypothèses H_1 et H_2, pour les fonctions r_o et V_o connues. Par analogie à la fonction score dans les modèles linéaires généralisés, on définit la fonction quasi-score $U_n(\theta)$ pour les observations $(y_i, x_i)_{i=1\cdots n}$, en $\theta \in \Theta$ par

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i - \mu_i(\theta)}{V_o(\mu_i(\theta))}\frac{\partial \mu_i(\theta)}{\partial \theta} \tag{1}$$

où $\mu_i(\theta) = r_o(x_i\theta)$ et $x_i\theta =< x_i, \theta >$. La fonction quasi-score vérifie les propriétés usuelles de la dérivée de la log-vraisemblance

$$E(U_n(\theta_o)) = 0,$$

$$E(U_n(\theta_o)^t(U_n(\theta_o))) + E\left(\frac{\partial U_n(\theta_o)}{{}^t\partial\theta}\right) = 0.$$

On peut définir une fonction analogue à la vraisemblance, en posant

$$K(u,t) = \int_u^t \frac{u-s}{V_o(s)}ds$$

et en définissant le logarithme de la quasi-vraisemblance (Wedderburn 1974) du modèle conditionnellement aux variables aléatoires X_i par

$$K_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} K(y_i, \mu_i(\theta)). \tag{2}$$

La fonction quasi-score s'explicite alors

$$U_n(\theta) = \frac{\partial K_n(\theta)}{\partial \theta}.$$

Par définition, on appelle estimateur du maximum de quasi-vraisemblance, l'estimateur $\hat{\theta}_n$ solution de $U_n(\theta) = 0$ et défini par :

$$\hat{\theta}_n = arg\max_{\theta\in\Theta} K_n(\theta). \tag{3}$$

L'existence de cet estimateur est présenté par Mac Cullagh (1983), dans lequel les propriétés asymptotiques de l'estimateur $\hat{\theta}_n$ sont établies. Sous des hypothèses adéquates, on obtient le résultat de convergence, qui fournit la $\sqrt{n}$-normalité de $\hat{\theta}_n$

$$\sqrt{n}(\hat{\theta}_n - \theta_o) \xrightarrow{d} N\left(0, \Gamma(\theta_o)^{-1}\right)$$

où $\Gamma(\theta_o) = \lim_{n\to\infty} E(-U_n'(\theta_o))$. Néanmoins on n'obtient pas un estimateur asymptotiquement efficace. Ce résutat est illustré dans l'exemple 1 suivant.

Exemple 1 : Soit le modèle à erreurs multiplicatives, $y_i = r_o(x_i\theta_o)u_i$ pour $i = 1, n$, où u_i est une variable aléatoire gaussienne indépendante de x_i telle que $E(u_i) = 1$ et $Var(u_i) = 1$. Les n v.a. (Y_i, X_i) sont supposées i.i.d.. Ce modèle est également défini par les deux hypothèses H_1 et H_2, qui s'écrivent $E\left(Y_i|X_i = x_i\right) = r_o(x_i\theta_o) = \mu_i(\theta_o)$ et $Var\left(Y_i|X_i = x_i\right) = \left(r_o(x_i\theta_o)\right)^2 = V_o(\mu_i(\theta_o))$.
Pour chaque $\theta \in \Theta$ la fonction quasi-score au point θ est donnée par

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{(y_i - r_o(x_i\theta))}{(r_o(x_i\theta))^2} r_o'(x_i\theta)^t x_i.$$

Dans ce cas, comme $V_o(x) = x^2$, la fonction $U_n(\theta)$ n'est autre que le score d'un modèle linéaire généralisé pour la loi gamma. La matrice $E(-U_n'(\theta_o))$ admet pour limite $\Gamma(\theta_o) = E\left(\frac{r_o'(X\theta_o)^2}{r_o(X\theta_o)^2} X^t X\right)$ quand $n \longrightarrow \infty$, où X est le vecteur aléatoire de R^p de même loi que chaque X_i. Le modèle spécifié pour la loi de probabilité P_o, revient à considérer que la v.a. $Y_i|X_i = x_i$ suit une loi normale de moyenne $r_o(x_i\theta_o)$ et de variance $(r_o(x_i\theta_o))^2$. La log-vraisemblance conditionnellement aux v.a. X_i, s'écrit

$$L_n(\theta) = -\frac{1}{2n}\sum_{i=1}^{n} \frac{1}{(r_o(x_i\theta))^2}(y_i - r_o(x_i\theta))^2 - \frac{1}{n}\sum_{i=1}^{n} log(r_o(x_i\theta)) + constante.$$

Après calcul, on trouve que $E\left(-\frac{\partial^2 L_n(\theta_o)}{\partial\theta^2}\right) = 3E(-U_n'(\theta_o))$. On en déduit donc que la variance asymptotique de l'estimateur du maximum de quasi-vraisemblance est égale à 3 fois celle de l'estimateur du maximum de vraisemblance $\hat{\theta}_{n,mv}$, qui vérifie

$$\sqrt{n}(\hat{\theta}_{n,mv} - \theta_o) \xrightarrow{d} N\left(0, \frac{1}{3}(\Gamma(\theta_o))^{-1}\right).$$

Dans cet exemple simple, il est possible de mesurer la perte d'efficacité, quand on utilise la quasi-vraisemblance, au lieu de la vraisemblance du modèle correctement spécifié. Par contre, l'estimateur de quasi-vraisemblance sera préférable à l'estimateur du maximum de vraisemblance dans un modèle mal spécifié défini par une loi autre que la loi gaussienne.

Exemple 2 : Soit les v.a $(Y_i, X_i)_{i=1,n}$ i.i.d. de même loi que (Y, X) telles que la loi conditionnelle de $Y|X$ est une loi continue de type Euler, de densité

$$f^{Y|X=x}(y) = \frac{1}{\Gamma(r_o(x\theta_o)}y^{r_o(x\theta_o)-1}e^{-y}1_{\{y\geq 0\}}.$$

Ce modèle est également défini par les hypothèses H_1 et H_2 qui s'écrivent $E(Y|X = x) = r_o(x\theta_o)$ et $Var(Y|X = x) = r_o(x\theta_o)$.
La quasi-vraisemblance de la v.a. continue $Y|X$ n'est autre que la log-vraisemblance d'un modèle discret poissonnien, de la famille exponentielle linéaire.

Exemple 3 : Soit le modèle tobit, défini par

$$Y_i = \begin{cases} Y_i^* & si\ Y_i^* > 0 \\ 0 & sinon \end{cases}$$

avec $Y_i^* = \theta_{o1}X_{1i} + \theta_{o2}X_{2i} + u_i$. Les v.a. $u = (u_1 \ldots u_n)'$, $X_1 = (X_{11} \ldots X_{1n})'$ et $X_2 = (X_{21} \ldots X_{2n})'$ sont supposées indépendantes et u suit la loi normale centrée réduite avec $E(u_iu_j) = \delta_{ij}$. Soit $X = (X_1, X_2)$ et $X_i = (X_{1i}, X_{2i})$. On désigne par Φ (resp φ) la fonction de répartition (resp la densité) de la loi normale centrée réduite. Ce modèle est défini par les hypothèses H_1 et $H_2^{'}$ qui s'écrivent dans ce cas, $E(Y_i|X_i) = E(Y_i^*1_{\{Y_i^*>0\}}|X_i) = r_o(X_i\theta_o)$ et
$Var(Y_i|X_i) = X_i\theta_oE(Y_i|X_i) - E^2(Y_i|X_i) + \Phi(X_i\theta_o) = w_o(X_i\theta_o)$, pour les fonctions $r_o(x) = x\Phi(x)+\varphi(x)$ et $w_o(x) = -r_o^2(x)+xr_o(x)+\Phi(x)$. Dans ce troisième exemple, il n'est plus possible d'expliciter la quasi-vraisemblance comme vraisemblance d'un modèle de la famille exponentielle linéaire.

2.2 Point de vue semi paramétrique

Dans la pratique, les fonctions r_o, V_o et w_o sont souvent inconnues, de même que la densité f du vecteur aléatoire X_i. Un choix erroné de r_o, V_o ou de w_o entraîne, de surcroît, une estimation moins performante du paramètre θ_o. Le cadre des modèles S.I.M. définis par l'hypothèse H_1, a été largement étudié quand la fonction r_o est inconnue (Ichimura (1993), Newey et Stocker (1993), Bonneu, Delecroix et Hristache (1995), Scherman (1994). Les différents auteurs précités proposent des estimateurs convergents de θ_o, de loi limite $\sqrt{n}$-normalité, dans le cas où on utilise un estimateur non paramétrique de r_o. On se placera, dans la suite, dans la même situation que les auteurs précités, à savoir r_o est supposée inconnue et on considérera la fonction $r_\theta(t)$ définie pour chaque $\theta \in \Theta$ fixé et pour tout $t \in \mathcal{W}_\theta$, par

$$r_\theta(t) = E(Y_i|X_i\theta = t), \tag{4}$$

où $\mathcal{W}_\theta = \{z = x\theta\ ; x \in S\}$, S désignant le support de X_i. De même quand w_o est inconnue, on sera ramené à utiliser la fonction w_θ définie sur $\mathcal{W}_\theta$ par $w_\theta(t) = Var(Y_i|X_i\theta = t)$. Les fonctions $r_\theta(.)$ et $w_\theta(.)$ vérifient respectivement $r_{\theta_o}(x\theta_o) = r_o(x\theta_o)$ et $w_{\theta_o}(x\theta_o) = w_o(x\theta_o)$. Quand r_o et w_o sont inconnues, on construit des estimateurs de θ_o en utilisant des estimateurs non paramétriques des fonctions $r_\theta(.)$ et $w_\theta(.)$. Cette construction est présentée dans le paragraphe suivant.

3 Estimateurs semi-paramétriques

Le but de ce paragraphe est de présenter dans ces modèles uniquement définis par les hypothèses sur les moments (H_1 et H_2) ou (H_1 et $H_2^{'}$), une méthode d'estimation du paramètre θ_o. Dans cette méthode, la composante paramétrique du modèle étant préalablement fixée, on estime la composante non paramétrique qui apparaît explicitement dans l'expression des moments, par un estimateur de type à noyau. Il est ensuite utilisé pour construire un estimateur de la composante paramétrique de façon analogue à la méthode de quasi-score, selon que V_o est connue ou inconnue. Dans cette approche semi-paramétrique, quand r_o est inconnue, nous considérons les conditions suivantes qui assurent l'identifiabilité du modèle défini par H_1 :
1) l'espace paramétrique Θ est de la forme $\Theta = \{(\theta_1 \cdots \theta_p) \in R^p \,/\theta_1 = 1\}$
2) $r_o \in C^1(R)$.
Ces conditions seront intégrées aux hypothèses nécessaires aux propriétés démontrées dans le paragraphe 3.

3.1 Estimateur $\hat{\theta}_{n,1}$ quand V_o est supposée connue

Dans un premier temps nous considérons le modèle relatif aux variables $Z_1, .., Z_n$, où $Z_i = (Y_i, X_i)$ est défini par les hypothèses H_1 et H_2, quand la fonction r_o est inconnue, pour la paramétrisation $(\theta, r_\theta, f) \longrightarrow P_{(\theta, r_\theta, f)} \in \mathcal{P}$. On rappelle que f est la densité de X_i. Pour définir l'estimateur de θ_o, nous allons, pour chaque θ fixé, estimer la fonction $r_\theta(.)$ définie par (4), par l'estimateur non paramétrique à noyau de Nadaraya et Watson

$$\hat{r}_\theta(x_i\theta) = \frac{\sum_{j\neq i}^n Y_j K\left(\frac{X_j\theta - x_i\theta}{b_n}\right)}{\sum_{j\neq i}^n K\left(\frac{X_j\theta - x_i\theta}{b_n}\right)} \tag{5}$$

où K est un noyau de densité et b_n est une suite de réels strictement positifs. On modifie alors la quasi-vraisemblance $K_n(\theta)$ définie en (2), en remplaçant dans son expression $\mu_i(\theta) = r_\theta(x_i\theta)$ par $\hat{r}_\theta(x_i\theta)$; on obtient la fonction

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^n K(y_i, \hat{r}_\theta(x_i\theta)). \tag{6}$$

On définit alors un premier estimateur semi-paramétrique qui maximise cette pseudo quasi-vraisemblance :

$$\hat{\theta}_{n,1} = arg \max_{\theta\in\Theta} Q_n(\theta) \tag{7}$$

et qui est solution du quasi-score nul, c'est à dire $Q_n^{'}(\hat{\theta}_{n,1}) = 0$, où

$$Q_n^{'}(\theta) = \frac{\partial Q_n(\theta)}{\partial\theta} = \sum_{i=1}^n \frac{(y_i - \hat{r}_\theta(x_i\theta))}{V_o(\hat{r}_\theta(x_i\theta))}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta}.$$

L'estimateur $\hat{\theta}_{n,1}$ rentre dans le cadre des M-estimateurs semi-paramétriques pour les S.I.M., obtenu en maximisant la fonction $\sum_{i=1}^n \Psi(y_i, \hat{r}_\theta(x_i\theta))$ (cf. M. Delecroix et M. Hristache (1995)). En effet la fonction objective, Ψ, dans ce cas n'est autre que $\Psi(y_i, r) = \int_{y_i}^r \frac{y_i - s}{V_o(s)} ds$. Cette fonction est une log-densité de la famille exponentielle linéaire, ce qui est une hypothèse suffisante dans le cadre des M-estimateurs des auteurs précités. Ainsi les propriétés de convergence presque sûre, de $\sqrt{n}$-normalité et d'efficacité, établies par ces auteurs s'adaptent à l'estimateur $\hat{\theta}_{n,1}$. Le concept d'efficacité dans ces modèles, est précisé dans le paragraphe 4.3. Dans le cas où V_o est connue, une alternative aux M-estimateurs, est un estimateur en deux étapes, basé sur le quasi-score (cf. S. Weisberg et A. H. Welsh (1994). Cette approche est considérée dans le paragraphe suivant pour construire un estimateur de θ_o, quand V_o est inconnue.

3.2 Estimateur $\hat{\theta}_{n,2}$ quand V_o est inconnue

Quand V_o est inconnue, on peut élargir l'hypothèse H_2 à l'hypothèse H_2'. Nous considérons donc le modèle relatif aux variables $Z_1, .., Z_n$ défini par les hypothèses H_1 et H_2' quand les fonctions r_o et w_o sont inconnues. Dans ce cas il n'est pas possible de définir une quasi-vraisemblance qui appartienne à la famille exponentielle linéaire. La méthode générale des M-estimateurs ne s'applique plus. Pour définir un estimateur du paramètre d'intérêt θ_o, on définit, pour chaque $\theta \in \Theta$ fixé et $\forall t \in \mathcal{W}_\theta$,

$$w_\theta(t) = Var\left(Y_i|X_i\theta = t\right), \tag{8}$$

$$\psi_\theta(t) = E\left(Y_i^2|X_i\theta = t\right). \tag{9}$$

On considère ainsi la paramétrisation $(\theta, r_\theta, w_\theta, f) \longrightarrow P_{(\theta, r_\theta, w_\theta)} \in \mathcal{P}$, avec $P_o = P_{\theta_o, r_o, w_o, f}$. D'après l'égalité $w_\theta(t) = \psi_\theta(t) - r_\theta^2(t)$, on définit l'estimateur de w_θ par

$$\hat{w}_\theta(x_i\theta) = \hat{\psi}_\theta(x_i\theta) - \hat{r}_\theta^2(x_i\theta) \tag{10}$$

et

$$\hat{\psi}_\theta(x_i\theta) = \frac{\sum_{j\neq i}^n Y_j^2 K(\frac{X_j\theta - x_i\theta}{b_n})}{\sum_{j\neq i}^n K(\frac{X_j\theta - x_i\theta}{b_n})}. \tag{11}$$

La suite b_n est une suite de réels strictement positifs et K est un noyau de densité. On choisit pour l'estimation de $r_\theta(.)$ et $\psi_\theta(.)$ la même largeur de fenêtre, ce qui permet de simplifier les calculs et qui se justifie quand on prend b_n de la forme $cn^{-\delta}$ avec $\delta > 0$. On définit la fonction quasi-score par

$$K_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)} \frac{\partial r_\theta(x_i\theta)}{\partial\theta} \tag{12}$$

qui n'est pas le gradient d'une quasi-vraisemblance. Pour définir l'estimateur de θ_o, on modifie le quasi-score en remplaçant les quantités inconnues par leurs estimateurs et on obtient

$$Q_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{y_i - \hat{r}_\theta(x_i\theta)}{\hat{w}_\theta(x_i\theta)} \frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta}. \tag{13}$$

Soit la matrice définie par

$$\hat{M}(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\hat{w}_\theta(x_i\theta)} \frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} \frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta'}. \tag{14}$$

L'estimateur semi-paramétrique $\hat{\theta}_{n,2}$, est construit à partir du quasi-score modifié $Q_n^*(\theta)$. Pour ce faire, on se donne un estimateur initial θ_n, qui soit $n^{1/2}$-convergent de θ_o et on définit alors

$$\hat{\theta}_{n,2} = \theta_n + \hat{M}(\theta_n)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{r}_{\theta_n}(x_i\theta_n)}{\hat{w}_{\theta_n}(x_i\theta_n)}\left(\frac{\partial\hat{r}_\theta(x_i\theta)}{\partial\theta}\right)|_{\theta=\theta_n}. \tag{15}$$

Un choix possible d'estimateur initial de θ_0 (celui que nous utiliserons par la suite) est l'estimateur semi-paramétrique des moindres carrés défini par Ichimura (1993),

$$\theta_n = arg\min_{\theta\in\Theta}(\sum_{i=1}^{n}(y_i - \hat{r}_\theta(x_i\theta))^2). \tag{16}$$

4 Propriétés asymptotiques de l'estimateur $\hat{\theta}_{n,2}$

Pour les modèles définis par H_1, les auteurs précités dans le paragraphe 2.2, ont proposé des estimateurs de θ_o basés sur l'estimation préliminaire non paramétrique de la fonction de régression $E(Y_1|_{X_1\theta})$ pour tout θ. Il ressort de leurs travaux la nécessité, pour obtenir la convergence de ces estimateurs, d'imposer un élagage(“trimming”) des valeurs X_i, soit fixes comme dans Ichimura (1993), soit dépendantes des observations comme dans Sherman (1994). Vu la complexité des conditions imposées dans le cas d'un élagage aléatoire, la solution qui consiste à ne retenir que les observations appartenant à un compact S, apparait satisfaisante, tant du point de vue théorique que pratique. En effet, les valeurs observées X_i sont, par nature, pour un échantillon de taille n fixé, contenues dans un compact. On supposera par la suite, que le support des X_i est compact. Dans le développement des propriétés asymptotiques de l'estimateur $\hat{\theta}_{n,2}$, nous présentons d'abord quelques résultats préliminaires et les hypothèses requises.

4.1 Hypothèses et résultats préliminaires

Pour obtenir la convergence presque sûre de l'estimateur $\hat{\theta}_{n,2}$ vers θ_o, nous présentons dans cette partie quelques résultats préliminaires obtenus sous les hypothèses suivantes :

A_1. L'espace des paramètres, Θ, est un compact de R^p défini par $\Theta = \{(\theta_1 \ldots \theta_p) \in R^p\ /\theta_1 = 1\}$. Le support de X_i, noté S, est un compact de R^p et θ_o appartient à l'interieur de Θ.

A_2. $E|Y_i|^8 < \infty$.

A_3. $\forall\theta \in \Theta$, la variable $X_i\theta$ est absolument continue de densité $h_\theta(.)$ et $\inf_{(x,\theta)\in S\times\Theta} h_\theta(x\theta) > 0$. Les fonctions définies par $R(x,\theta) = r_\theta(x\theta)$ et $D(x,\theta) = h_\theta(x\theta)$ sont deux fois continûment dérivables.

A_4. La largeur de la fenêtre $b_n = cn^{-\delta}$ avec $c > 0$ et $\delta \in (\frac{1}{8}\ \frac{1}{7})$.

A_5. Le noyau K est unitaire, deux fois dérivable. Les fonctions K, K', K” sont lipchitziens, bornés, vérifiant des conditions supplémentaires données en appendice.

Les hypothèses A_5 faites sur, le noyau, sont satisfaites pour le noyau gaussien. Dans l'hypothèse A_1 la première composante du paramètre θ_o a été contrainte à être égale à un ce qui permet l'identifiabilité du modèle, explicitée dans la propriété suivante.

Propriété 1. *Sous les hypothèses,* $\Theta = \{(\theta_1 \dots \theta_p) \in R^p \ / \ \theta_1 = 1\}$ *et la fonction* $r_\theta(z)$ *est dérivable en tout point de* $z \in \{x\theta \ / \ x \in S, \ \theta \in \Theta\}$ *pour tout* $\theta \in \Theta$*, on a*

$$\forall(\theta, \theta^*) \in \Theta^2, r_\theta(x\theta) = r_{\theta^*}(x\theta^*) \Longrightarrow \theta = \theta^*.$$

Dans la suite, on désignera par $|a|$ la norme supérieure de tout vecteur ou toute matrice de dimension finie, c'est-à-dire $|a|$ est le maximum des valeurs absolues des éléments de a. Pour établir la convergence de l'estimateur $\hat{\theta}_{n,2}$, il est nécessaire d'abord de démontrer la convergence des estimateurs $\hat{r}_\theta(x_i\theta)$ vers $r_\theta(x_i\theta)$ uniformement en θ et en x_i, ainsi que celle de leurs dérivées. C'est l'objet du lemme qui suit dont la démonstration figure dans Bonneu, Delecroix et Hristache (1995).

Lemme 1. *Sous les hypothèses* $A_1 - A_5$ *, on établit les résultats suivants*

$$\lim_{n\to\infty} \sup_{(x,\theta)\in S\times\Theta} |\hat{r}_\theta(x\theta) - r_\theta(x\theta)| = 0 \qquad P_o \ ps,$$

$$n^{1/4} \sup_{(x,\theta)\in S\times\Theta} |\hat{r}_\theta(x\theta) - r_\theta(x\theta)| = o_{P_o}(1), \tag{17}$$

$$\lim_{n\to\infty} \sup_{(x,\theta)\in S\times\Theta} \left|\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{\partial r_\theta(x\theta)}{\partial\theta}\right| = 0 \qquad P_o \ ps,$$

$$n^{1/4} \sup_{(x,\theta)\in S\times\Theta} \left|\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{\partial r_\theta(x\theta)}{\partial\theta}\right| = o_{P_o}(1). \tag{18}$$

4.2 Convergence de l'estimateur $\hat{\theta}_{n,2}$

4.2.1 Convergence presque sûre

La convergence presque sûre de $\hat{\theta}_{n,2}$ nécessite celles des estimateurs $\hat{w}_\theta(x_i\theta)$ donnés par (10), uniformement en θ et en x_i vers $w_\theta(x_i\theta)$. Cette convergence est énoncée dans le lemme suivant sous l'hypothèse A_6.

A_6**.** La fonction $\Psi(x, \theta) = \psi_\theta(x\theta)$ donnée par (9) est une fois continûment différentiable sur $S \times \Theta$ et $\inf_{(x,\theta)\in S\times\Theta} w_\theta(x\theta) > 0$.

Lemme 2. *Sous les hypothèses* $A_1 - A_6$ *, on obtient*

$$\lim_{n\to\infty} \sup_{(x,\theta)\in S\times\Theta} |\hat{\psi}_\theta(x\theta) - \psi_\theta(x\theta)| = 0 \qquad P_0 \ ps,$$

$$n^{1/4} \sup_{(x,\theta)\in S\times\Theta} |\hat{\psi}_\theta(x\theta) - \psi_\theta(x\theta)| = o_{P_o}(1).$$

La preuve est une adaptation directe de la démonstration du Lemme 1. Les Lemmes 1 et 2 permettent d'établir le Lemme suivant dont la preuve est immédiate.

Lemme 3. *Sous les hypothèses* $A_1 - A_6$ *, on a*

$$\lim_{n\to\infty} \sup_{(x,\theta)\in S\times\Theta} |\hat{w}_\theta(x\theta) - w_\theta(x\theta)| = 0 \qquad P_o \ ps,$$

$$n^{1/4} \sup_{(x,\theta)\in(S\times\Theta)} |\hat{w}_\theta(x\theta) - w_\theta(x\theta)| = o_{P_o}(1). \tag{19}$$

Soit

$$Q^*(\theta) = E\left(\frac{Y_i - r_\theta(X_i\theta)}{w_\theta(X_i\theta)}\frac{\partial r_\theta(X_i\theta)}{\partial\theta}\right). \tag{20}$$

Pour montrer la convergence presque sûre de $\hat{\theta}_{n,2}$, nous montrerons d'abord que la fonction $Q_n^*(\theta)$ converge uniformément vers $Q^*(\theta)$. Cette propriété est énoncée dans la proposition suivante.

Proposition 1. *Sous les hypothèses* $A_1 - A_6$ *, on a*

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|Q_n^*(\theta) - K_n^*(\theta)| = 0 \qquad P_o\ ps,$$

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|K_n^*(\theta) - Q^*(\theta)| = 0 \qquad P_o\ ps.$$

Preuve : Voir appendice A_1.

Sous les hypothèses $A_1 - A_6$, l'estimateur θ_n est $n^{1/2}$-convergent et converge presque sûrement vers θ_o (cf. Ichimura 1990).

Soit

$$M(\theta) = E_X\left(\frac{1}{Var(Y_i|X_i)}\frac{\partial r_\theta(X_i\theta)}{\partial\theta}\frac{\partial r_\theta(X_i\theta)}{\partial\theta'}\right) \tag{21}$$

où E_X désigne l'espérance par rapport à la loi de la variable aléatoire X, qui a la même loi que X_i.

Lemme 4. *Sous les hypothèses* $A_1 - A_6$ *, la matrice* $\hat{M}(\theta_n)$ *donnée par (14) converge presque sûrement vers la matrice* $M(\theta_o)$*, donnée par (21).*

La preuve se déduit de celle d'un résultat plus général énoncé au paragraphe suivant, dans le théorème 2. L'estimateur $\hat{\theta}_n$ est défini en (16). Nous énonçons enfin le théorème qui établit la convergence presque sûre de l'estimateur $\hat{\theta}_{n,2}$ sous l'hypothèse A_7 suivante.

A_7 **.** La matrice $M(\theta_o)$ donnée par (21) est inversible.

Théorème 1. *Sous les hypothèses* $A_1 - A_7$**,** *l'estimateur* $\hat{\theta}_{n,2}$ *converge presque sûrement vers* θ_o *.*

Preuve du théorème 1 : On peut majorer $|\hat{\theta}_{n,2} - \theta_o|$ par $|\theta_n - \theta_o| +$ $|\hat{M}(\theta_n)^{-1}|\sup_{\theta\in\Theta}|Q_n^*(\theta) - Q^*(\theta)| + |\hat{M}(\theta_n)^{-1}||Q^*(\theta_n)|$. En utilisant la Proposition 1, le Lemme 4, la continuité de $Q^*(\theta)$ et la convergence de θ_n vers θ_o qui est un zéro de $Q^*(\theta)$, on a la convergence de $\hat{\theta}_{n,2}$ vers θ_o.

4.2.2 Normalité asymptotique

La normalité asymptotique de $\hat{\theta}_{n,2}$ nécessite les résultats préliminaires que nous énonçons dans un cadre un peu plus général, dans le théorème suivant.

Théorème 2. *Sous les hypothèses* $A_1 - A_7$ *,* $\forall\ \alpha \in (0,\infty)$*, on a*

$$\sup_{||\theta-\theta_o||\le n^{-1/2}\alpha}|\sqrt{n}(Q_n^*(\theta) - K_n^*(\theta_o)) + M(\theta_o)\sqrt{n}(\theta - \theta_o)| = o_{P_o}(1),$$

$$\sup_{||\theta-\theta_o||\le n^{-1/2}\alpha}|\frac{1}{n}\sum_{i=1}^n \frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial\hat{r}_\theta(x_i\theta)}{\partial\theta}\frac{\partial\hat{r}_\theta(x_i\theta)}{{}^t\partial\theta} - M(\theta_o)| = o(1) \qquad P_o\ ps.$$

Preuve : La démonstration de ce théorème étant technique, nous la ferons en appendice A_2. Les fonctions $K_n^*(\theta)$ et $Q_n^*(\theta)$ sont définies respectivement en (12) et en (13).
Finalement on obtient la normalité asymptotique de $\hat{\theta}_{n,2}$ comme conséquence du Théorème 2.

Théorème 3. *Sous les hypothèses* $A_1 - A_7$ *, l'estimateur* $\hat{\theta}_{n,2}$ *défini par (15) vérifie*

$$\sqrt{n}(\hat{\theta}_{n,2} - \theta_o) \xrightarrow{d} N(0, M(\theta_o)^{-1}).$$

Preuve : En utilisant le théorème 2, on obtient

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{n,2} - \theta_o) &= \hat{M}(\theta_n)^{-1}\left(\hat{M}(\theta_n)\sqrt{n}(\theta_n - \theta_o) + \sqrt{n}Q_n^*(\theta_n)\right) \\ &= \hat{M}(\theta_n)^{-1}((\hat{M}(\theta_n) - M(\theta_o))\sqrt{n}(\theta_n - \theta_o)\end{aligned}$$

$$+M(\theta_o)\sqrt{n}(\theta_n - \theta_o) + \sqrt{n}(Q_n^*(\theta_n) - K_n^*(\theta_o)) + \hat{M}(\theta_n)^{-1}\sqrt{n}K_n^*(\theta_o)$$

$$= M(\theta_o)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\left(\frac{\partial r_\theta(x_i\theta)}{\partial\theta}\right)|_{\theta=\theta_o} + o_{P_o}(1).$$

On retrouve le résutat par application du théorème central limite.
La 2^{eme} partie du Théorème 2 fournit un estimateur convergent $\hat{M}(\hat{\theta}_{n,2})$ de la matrice de variance covariance de de l'estimateur $\hat{\theta}_{n,2}$. Ce qui peut être utile pour construire des tests d'hypothèses ou des intervalles de confiance. Ces résultats asymptotiques énoncés plus haut permettent de conclure que l'estimateur $\hat{\theta}_{n,2}$ est adaptatif, en ce sens que sa distribution asymptotique ne dépend pas de la manière dont la fonction de lien et la fonction de variance ont été estimées, et ne dépend pas, non plus, de l'estimateur initial.

4.3 Bornes d'efficacité

Dans les modèles semi-paramétriques à indice simple définis par l'hypothèse H_1 où la variance conditionnelle de $Y_i|X_i$ bornée et la fonction de lien est supposée continûment dérivable, la borne d'efficacité de tout estimateur régulier est donnée par

$$B_{\theta_o} = E\left(\frac{1}{Var(Y_1|X_1)} r_o^{'}(X_1\theta_o)^2 (X_1 - E_v(X_1\theta_o))(X_1 - E_v(X_1\theta_o))^{'}\right)$$

où $E_v(X_1\theta_o) = \frac{E(X_1 Var(Y_1|X_1)^{-1}|X_1\theta_o)}{E(Var(Y_1|X_1)^{-1}|X_1\theta_o)}$ (cf. Newey (1990), Hristache (1995)). La matrice B_{θ_o} est la borne d'efficacité d'un estimateur obtenu à partir d'une fonction estimante linéaire en $Y_i - E(Y_i|X_i)$ et dans ce cas, le calcul de cette borne n'utilise pas l'hypothèse H_2 ou $H_2^{'}$ spécifiant la variance. Cet argument a été largement étudié par W. Wefelmeyer (1996), qui, pour étendre des estimateurs basés sur le quasi-score, considère une fonction estimante linéaire en $Y_i - E(Y_i|_{X_i})$ et en $(Y_i - E(Y_i|_{X_i}))^2$. Il obtient alors, pour cette classe d'estimateurs une borne d'efficacité qui tient compte de l'hypothèse sur la variance. Ce n'est pas le cadre de notre travail, car l'estimateur que l'on propose, est basé sur le quasi-score. Dans le cas où la variance conditionnelle de $Y_1|X_1$ ne dépend que de $X_1\theta_o$, ce qui est le cas des hypothèses H_2 ou $H_2^{'}$, l'expression ci-dessus se simplifie et on a $E_v(X_1\theta_o) = E(X_1|X_1\theta_o)$. Dans les modèles S.I.M., on obtient

$$\frac{\partial r_\theta(X_1\theta)}{\partial\theta}|\theta = \theta_o = r_o^{'}(X_1\theta_o)(X_1 - E(X_1|X_1\theta_o)). \tag{22}$$

La démonstration de cette égalité est fondée sur le raisonnement suivant. Soit la variable (Y, X), ayant la même loi que les variables (Y_i, X_i), on a les égalités $r_\theta(X\theta) = E(Y|X\theta) = E(E(Y|X)|X\theta) = E(r_o(X\theta_o)|X\theta)$. Rappelons que $r_o(X\theta_o) = E(Y|X) = E(Y|X\theta_o)$.
Soit $(\theta_1, \theta_2) \in \Theta^2$ et la fonction g définie par $g(\theta_1, \theta_2) = E(r_o(X\theta_o - X\theta_1 + X\theta_2)|X\theta_2)$. Il en résulte que $r_\theta(X\theta) = g(\theta, \theta)$ et

$$\frac{\partial r_\theta(X\theta)}{\partial\theta}|_{\theta=\theta_o} = \frac{\partial g(\theta_1, \theta_o)}{\partial\theta_1}|_{\theta_1=\theta_o} + \frac{\partial g(\theta_o, \theta_2)}{\partial\theta_2}|_{\theta_2=\theta_o}$$

$$= \frac{\partial}{\partial\theta_1} E\left(r_o(2X\theta_o - X\theta_1)|X\theta_o\right)|_{\theta_1=\theta_o} + \frac{\partial}{\partial\theta_2} E\left(r_o(X\theta_2)|X\theta_2\right)|_{\theta_2=\theta_o}$$

$$= -r_o^{'}(X\theta_o)E(X|X\theta_o) + Xr_o^{'}(X\theta_o) = r_o^{'}(X\theta_o)(X - E(X|X\theta_o)).$$

Ainsi la borne d'efficacité devient

$$B_{\theta_o} = E\left(\frac{1}{Var(Y_1|X_1)}\left(\frac{\partial r_\theta(X_1\theta)}{\partial\theta}|_{\theta=\theta_o}\right)\left(\frac{\partial r_\theta(X_1\theta}{\partial\theta'}|_{\theta=\theta_o}\right)\right).$$

On conclut donc que l'estimateur $\hat{\theta}_{n,2}$ atteint la borne d'efficacité semi-paramétrique dans les modèles S.I.M., puisque $B_{\theta_o} = M(\theta_o)$.

5 Appendices

Dans les démonstrations on notera de façon équivalente y_i ou Y_i. Les conditions sur le noyau K qui permettent d'établir les Lemmes 1 et 2 sont : K est deux fois dérivable et K, K', K'' sont bornés et lipschitziens,$\int K(u)du = 1$, $\int u^p K'(u)du = -\delta_1^p$ pour $p \in \{0,1,2\}$ et $\int u^p K''(u)du = 2\delta_2^p$ pour $p \in \{0,1,2,3\}$. Aussi, avec les hypothèses sur la largeur de la fenêtre on a b_n, nb_n^8 et $\sqrt{n}b_n^4$ qui convergent vers zéro quand n tend vers l'infini et nb_n, nb_n^2, nb_n^4 et $|\frac{nb_n}{\log(b_n)}|$ qui convergent vers l'infini quand n tend vers l'infini. Ainsi pour la démonstration des lemmes 1 et 2 on peut indifféremment se reporter à Delecroix, Bonneu, Hristache (1995) ou à Ichimura (1993).

Appendice A_1

Preuve de la proposition 1. Pour la premère équation, on a

$$|Q_n^*(\theta) - K_n^*(\theta)| = |\frac{1}{n}\sum_{i=1}^n \left(\frac{y_i - \hat{r}_\theta(x_i\theta)}{\hat{w}_\theta(x_i\theta)} \frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - \frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)} \frac{\partial r_\theta(x_i\theta)}{\partial\theta} \right)|$$

$$\begin{aligned}
&\leq |\frac{1}{n}\sum_{i=1}^n (r_\theta(x_i\theta) - \hat{r}_\theta(x_i\theta))(\frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta})| \\
&+ |\frac{1}{n}\sum_{i=1}^n (y_i - r_\theta(x_i\theta)) \left(\frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta} \right)| \\
&+ |\frac{1}{n}\sum_{i=1}^n \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta} \left((r_\theta(x_i\theta)) - \hat{r}_\theta(x_i\theta)) \right) |.
\end{aligned}$$

Notons A, B, C le premier, le deuxième et le trosième terme du membre de droite de l'inégalité, on a

$$|A| \leq \sup_{(x,\theta)\in S\times\Theta} |\hat{r}_\theta(x\theta) - r_\theta(x\theta)| \sup_{(x,\theta)\in S\times\Theta} |\frac{1}{\hat{w}_\theta(x\theta)}\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta}|$$

$$\begin{aligned}
|B| &\leq \sup_{(x,\theta)\in S\times\Theta} |\frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta}| \sup_{(x,\theta)\in S\times\Theta} |\hat{r}_\theta(x\theta) - r_\theta(x\theta)| \\
|C| &\leq \sup_{(x,\theta)\in S\times\Theta} |\frac{1}{\hat{w}_\theta(x\theta)}\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta}| \left(\frac{1}{n}\sum_{i=1}^n (|y_i| + c) \right).
\end{aligned}$$

D'après A_3, A_6 et les Lemmes 1 et 3, on a le resultat cherché.
Pour la seconde équation de la proposition 1, on rapelle que

$$K_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}.$$

Posons

$$g_i(\theta) = \frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta} - E(\frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}).$$

Il nous faut montrer que $\lim_{n\to\infty}\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{i=1}^{n} g_i(\theta)| = 0$. On divise Θ en J_n cubes disjoints C_r^n de volume A_o/k_n^d avec $J_n = k_n^d$ où $\lim_{n\to\infty} k_n = \infty$. Choisissons un point θ_r dans chaque cube C_r^n, on peut alors écrire :

$$\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{i=1}^{n} g_i(\theta)| \leq \sup_{\theta\in C_r^n\ 1\leq r\leq J_n}|\frac{1}{n}\sum_{i=1}^{n}(g_i(\theta)-g_i(\theta_r))| + \sup_{1\leq r\leq J_n}|\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_r)|.$$

Remarquons alors que

$$|\frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta} - \frac{y_i - r_{\theta_r}(x_i\theta_r)}{w_{\theta_r}(x_i\theta_r)}\frac{\partial r_\theta(x_i\theta_r)}{\partial\theta}|$$

$$\begin{aligned}
&\leq |(r_{\theta_r}(x_i\theta_r) - r_\theta(x_i\theta))(\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_{\theta_r}(x_i\theta_r)}\frac{\partial r_\theta(x_i\theta_r)}{\partial\theta})| \\
&+ |(y_i - r_{\theta_r}(x_i\theta_r))(\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_{\theta_r}(x_i\theta_r)}\frac{\partial r_\theta(x_i\theta_r)}{\partial\theta})| \\
&+ |\frac{1}{w_{\theta_r}(x_i\theta_r)}\frac{\partial r_\theta(x_i\theta_r)}{\partial\theta}(r_\theta(x_i\theta) - r_{\theta_r}(x_i\theta_r))| \\
&\leq O(||\theta-\theta_r||^2) + O(||\theta-\theta_r||)(|y_i| + c),
\end{aligned}$$

où c est une constante. Ainsi

$$|E(\frac{y_i - r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta} - \frac{y_i - r_{\theta_r}(x_i\theta_r)}{w_{\theta_r}(x_i\theta_r)}\frac{\partial r_\theta(x_i\theta_r)}{\partial\theta})| = O(||\theta-\theta_r||^2) + O(||\theta-\theta_r||).$$

Comme $\theta \in C_r^n$ on a $||\theta-\theta_r|| = O(k_n^{-1})$ et $\sup_{1\leq r\leq J_n\ \theta\in C_r^n}|\frac{1}{n}\sum_{i=1}^{n}(g_i(\theta)-g_i(\theta_r))| = O(k_n^{-2}) + O(k_n^{-1})(\frac{1}{n}\sum_{i=1}^{n}(|y_i|+c))$.
Cette dernière égalité assure la convergence vers zéro du terme de gauche. D'autre part les égalités $|g_i(\theta)| = O(|y_i|+c)$ et $|g_i^2(\theta)| = O(|y_i|^2+c)$ permettent de déduire que $E(g_i(\theta))^2 \leq m_o < \infty$.
D'après l'inégalité exponentielle on a

$$P(\sup_{1\leq r\leq J_n}|\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_r)| > \epsilon) \leq J_n P(|\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_r)| > \epsilon) \leq$$
$$2J_n exp(\frac{-n\epsilon^2}{4m_o}) = 2ck_n^{-d} exp(\frac{-n\epsilon^2}{4m_o}).$$

Pour $k_n = n^\alpha$, avec $\alpha > 0$ la série $\sum_{n=1}^{\infty} 2cn^{-d\alpha} exp(\frac{-n\epsilon^2}{4m_o})$ converge.
D'où on obtient $\lim_{n\to\infty}\sup_{1\leq r\leq J_n}\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_r) = 0$. Ce qui achève la preuve.

Appendice 2

Preuve du théorème 2. Nous commencons par montrer la première partie du théorème. Notons $\bar\Theta = \{\theta/\ \ |\theta-\theta_o| \leq n^{-1/2}\alpha\}$, et soit la variable

$$\begin{aligned}
T &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{y_i - \hat r_\theta(x_i\theta)}{\hat w_\theta(x_i\theta)}\frac{\partial\hat r_\theta(x_i\theta)}{\partial\theta} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{y_i - r_{\theta_o}(x_i\theta_o)}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}|_{\theta=\theta_o} \\
&+ M(\theta_o)\sqrt{n}(\theta-\theta_o).
\end{aligned}$$

Alors T se décompose sous la forme $T = T_1 + T_2 + T_3$, avec

$$\begin{aligned}
T_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - r_{\theta_o}(x_i\theta_o))(\frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta})|_{\theta=\theta_o}),\\
T_2 &= M(\theta_o)\sqrt{n}(\theta-\theta_o) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(r_\theta(x_i\theta) - r_{\theta_o}(x_i\theta_o))\frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta},\\
T_3 &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta))\frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta}.
\end{aligned}$$

Il nous suffit de montrer que $\sup_{\theta\in\bar{\Theta}}|T_m| = o_{P_o}(1)$ pour $m = 1, 2, 3$.

Posons $W(x_i,\theta) = \frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - (\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta})|_{\theta=\theta_o}$.

On a la majoration suivante :

$$\begin{aligned}
\sup_{(x,\theta)\in S\times\bar{\Theta}}|W(x,\theta)| \leq &\sup_{(x,\theta)\in S\times\bar{\Theta}}|\frac{1}{\hat{w}_\theta(x\theta)}\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta}|\\
&+ \sup_{(x,\theta)\in S\times\bar{\Theta}}|\frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta} - (\frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta})|_{\theta=\theta_o}|.
\end{aligned}$$

Par le théorème des accroissements finis le deuxième terme est un $O(n^{-1/2})$. D'après les lemmes 1 et 3 on a

$$n^{1/4}\sup_{(x,\theta)\in S\times\bar{\Theta}}|W(x,\theta)| = o_{P_o}(1), \tag{23}$$

$$\lim_{n\to\infty}\sup_{(x,\theta)\in S\times\bar{\Theta}}|W(x,\theta)| = 0 \qquad P_o\ ps. \tag{24}$$

Montrons dans un premier temps que $\sup_{\theta\in\bar{\Theta}}|T_3| = o_{P_o}(1)$.
On a la majoration suivante

$$\begin{aligned}
|T_3| \leq |\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta)W(x_i,\theta)| + &\\
&|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}|_{\theta=\theta_o}(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta)).
\end{aligned}$$

Soit la variable

$$T_3(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta))W(x_i,\theta).$$

On a, $|T_3(1)| \leq n^{1/4}\sup_{(x,\theta)\in S\times\bar{\Theta}}|\hat{r}_\theta(x\theta) - r_\theta(x\theta)|n^{1/4}\sup_{(x,\theta)\in S\times\bar{\Theta}}|W(x,\theta)|$ et comme $\sup_{(x,\theta)\in S\times\bar{\Theta}}|W(x,\theta)| = O_p(b_n^2) + O_p(n^{-1/2})$, le Lemme 1 et (23) permettent de conclure que $T_3(1) = o_{P_o}(1)$. D'autre part en notant $\epsilon_i^1 = \frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}|_{\theta=\theta_o}$, $E(\epsilon_i^1|_{x_i\theta_o}) = 0$ grâce à (22). Dès lors,

$$T_3(2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}|_{\theta=\theta_o}(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta))$$

se traite comme T_1.
Montrons en suite que $\sup_{\theta\in\bar\Theta}|T_2| = o_{P_o}(1)$. On a,

$$\begin{aligned}
T_2 &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial \hat r_\theta(x_i\theta)}{\partial\theta}(r_\theta(x_i\theta)-r_{\theta_o}(x_i\theta_o)) + M(\theta_o)\sqrt{n}(\theta-\theta_o)\\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial \hat r_\theta(x_i\theta)}{\partial\theta}(r_\theta(x_i\theta)-r_{\theta_o}(x_i\theta_o))\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'}\sqrt{n}(\theta-\theta_o)+o_{P_o}(1)
\end{aligned}$$

Pour $\theta\in\bar\Theta$, on obtient, $r_\theta(x_i\theta)-r_{\theta_o}(x_i\theta_o) = \frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'}(\theta-\theta_o)+O(n^{-1})$. Par suite

$$\begin{aligned}
T_2 &= -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial \hat r_\theta(x_i\theta)}{\partial\theta}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'}\sqrt{n}(\theta-\theta_o)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'}\sqrt{n}(\theta-\theta_o)\\
&- O(n^{-1})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial \hat r_\theta(x_i\theta)}{\partial\theta}+o_{P_o}(1).
\end{aligned}$$

On a,

$$\begin{aligned}
T_2 = -\frac{1}{n}\sum_{i=1}^{n}W(x_i,\theta)\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'}\sqrt{n}(\theta-\theta_o)+\\
O(n^{-1/2})\frac{1}{n}\sum_{i=1}^{n}W(x_i,\theta)-O(n^{-1/2})\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta}+o_{P_o}(1).
\end{aligned}$$

D'où on obtient,
$\sup_{\theta\in\bar\Theta}|T_2| \le M \sup_{(x,\theta)\in\bar\Theta\times S}|W(x,\theta)| \sup_{(x,\theta)\in S\times\bar\Theta}|\frac{\partial r_\theta(x\theta)}{\partial\theta}|+$
$O(n^{-1/2}) \sup_{(x,\theta)\in S\times\bar\Theta}|W(x,\theta)| + O(n^{-1/2})+o_{P_o}(1)$.
Comme $\lim_{n\to\infty}\sup_{(x,\theta)\in S\times\bar\Theta}|W(x,\theta)|=0$, on a $\sup_{\theta\in\bar\Theta}|T_2| = o_{P_o}(1)$.
Montrons enfin que $\sup_{\theta\in\bar\Theta}|T_1| = o_{P_o}(1)$. Revenant à la définition de T_1, on a

$$T_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-r_{\theta_o}(x_i\theta_o))W(x_i,\theta),$$

avec $W(x_i,\theta) = \frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial\hat r_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}|_{\theta=\theta_o}$. On a par le théorème des accroissements finis, $W(x_i,\theta) = \frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial\hat r_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}+$
$D_{\tilde\theta}(x_i\tilde\theta)(\theta-\theta_o)$, avec $|\theta-\tilde\theta|\le|\theta-\theta_o|$ et $D_{\tilde\theta}(x_i\tilde\theta) = \frac{\partial}{\partial\theta}\left(\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}\right)|\theta=\tilde\theta$. On obtient ainsi

$$\begin{aligned}
T_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-r_{\theta_o}(x_i\theta_o))\left(\frac{1}{\hat w_\theta(x_i\theta)}\frac{\partial\hat r_\theta(x_i\theta)}{\partial\theta}-\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}\right)\\
&+ \left(\frac{1}{n}\sum_{i=1}^{n}(y_i-r_{\theta_o}(x_i\theta_o))D_{\tilde\theta}(x_i\tilde\theta)\right)\sqrt{n}(\theta-\theta_o).
\end{aligned}$$

Pour $\sqrt{n}(\theta-\theta_o) = O(1)$, le dernier terme est un $o_{P_o}(1)$ par la loi des grands nombres car $E(y_i - r_{\theta_o}(x_i\theta_o))D_{\tilde{\theta}}(x_i\tilde{\theta})) = E(E(y_i - r_{\theta_o}(x_i\theta_o)|x_i)D_{\tilde{\theta}}(x_i\tilde{\theta}) = 0$.
Il nous reste donc à montrer que

$$T_1(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - r_{\theta_o}(x_i\theta_o))(\frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}) = o_{P_o}(1).$$

Le développement à l'ordre 2 s'écrit

$$\frac{1}{\hat{w}_\theta(x\theta)}\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} = \frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta} + \frac{1}{w_\theta(x\theta)}(\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{\partial r_\theta(x\theta)}{\partial\theta}) - \frac{1}{w^2_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta}(\hat{w}_\theta(x\theta) - w_\theta(x\theta)) + R_n,$$

où R_n regroupe les termes du second ordre en $(\hat{w}_\theta(x\theta) - w_\theta(x\theta))$ et $(\frac{\partial \hat{r}_\theta(x\theta)}{\partial\theta} - \frac{\partial r_\theta(x\theta)}{\partial\theta})$.

Les termes d'ordre deux se traitant par simple majoration grâce aux Lemmes 1 et 2, il suffit donc de montrer que

$$T_1(2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - r_{\theta_o}(x_i\theta_o))\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}(\hat{w}_\theta(x_i\theta) - w_\theta(x_i\theta)) = o_{P_o}(1),$$

$$T_1(3) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - r_{\theta_o}(x_i\theta_o))\frac{1}{w_\theta(x_i\theta)}\left(\frac{\partial \hat{r}_\theta(x_i\theta)}{\partial\theta} - \frac{\partial r_\theta(x_i\theta)}{\partial\theta}\right) = o_{P_o}(1).$$

Mais

$$\hat{w}_\theta(x_i\theta) - w_\theta(x_i\theta) = \hat{\varphi}_\theta(x_i\theta) - \varphi_\theta(x_i\theta) - (\hat{r}^2_\theta(x_i\theta) - r^2_\theta(x_i\theta)) = (\hat{\varphi}_\theta(x_i\theta) - \varphi_\theta(x_i\theta)) - (\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta))^2 - 2r_\theta(x_i\theta)(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta)).$$

Pour montrer donc que $T_1(2) = o_{P_o}(1)$, il suffit de prouver que

$$T_1(4) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - r_{\theta_o}(x_i\theta_o))\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}(\hat{\varphi}_\theta(x_i\theta) - \varphi_\theta(x_i\theta)) = o_{P_o}(1),$$

$$T_1(5) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - r_{\theta_o}(x_i\theta_o))\frac{2r_\theta(x_i\theta)}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}(\hat{r}_{\theta_o}(x_i\theta) - r_\theta(x_i\theta)) = o_{P_o}(1).$$

En effet le terme en $(\hat{r}_\theta(x_i\theta) - r_\theta(x_i\theta))^2$ se traite par simple majoration en vertu de la deuxième partie du lemme 1. Les variables $T_1(3)$, $T_3(2)$, $T_1(4)$ et $T_1(5)$ se traitent de la même façon. Nous nous contenterons de montrer $T_1(4) = o_{P_o}(1)$.

Rappelons que $\hat{\varphi}_\theta(x_i\theta) = \dfrac{\sum_{j\neq i}^{n} y_j^2 K(\frac{x_j\theta - x_i\theta}{b_n})}{\sum_{j\neq i}^{n} K(\frac{x_j\theta - x_i\theta}{b_n})} = \dfrac{\hat{N}_\theta(x_i\theta)}{\hat{D}_\theta(x_i\theta)}$.

Soient $N_\theta(x_i\theta)$ et $D_\theta(x_i\theta)$ les limites respectives de $\hat{N}_\theta(x_i\theta)$ et $\hat{D}_\theta(x_i\theta)$. On a $\hat{\varphi}_\theta(x_i\theta) - \varphi_\theta(x_i\theta) = \frac{1}{D_\theta(x_i\theta)}(\hat{N}_\theta(x_i\theta) - N_\theta(x_i\theta)) - \frac{\varphi_\theta(x_i\theta)}{D_\theta(x_i\theta)}(\hat{D}_\theta(x_i\theta) - D_\theta(x_i\theta)) + R_n$, où R_n regroupe les termes du second ordre en $(\hat{N}_\theta(x_i\theta) - N_\theta(x_i\theta))$ et $(\hat{D}_\theta(x_i\theta) - D_\theta(x_i\theta))$.

Pour montrer que $T_1(4) = o_{P_o}(1)$, il suffit donc de montrer que

$$T_1(6) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i\frac{1}{w_\theta(x_i\theta)D_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}(\hat{N}_\theta(x_i\theta) - N_\theta(x_i\theta)) = o_{P_o}(1),$$

$$T_1(7) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}\frac{\varphi_\theta(x_i\theta)}{D_\theta(x_i\theta)}(\hat{D}_\theta(x_i\theta) - D_\theta(x_i\theta)) = o_{P_o}(1)$$

avec $\epsilon_i = y_i - r_{\theta_o}(x_i\theta_o)$.

Les termes du second ordre en $(\hat{N}_\theta(x_i\theta) - N_\theta(x_i\theta))$ et $(\hat{D}_\theta(x_i\theta) - D_\theta(x_i\theta))$ non considérés se traitent par simple majoration. Les variables $T_1(6)$ et $T_1(7)$ se traitent de façon analogue, aussi, nous nous contenterons de montrer que $T_1(6) = o_{p_o}(1)$. Posons $\omega(x_i\theta) = \frac{1}{w_\theta(x_i\theta)D_\theta(x_i\theta)}$ $\times \frac{\partial r_\theta(x_i\theta)}{\partial\theta}$. L'écriture explicite de $\hat{N}_\theta(x_i\theta)$, permet d'écrire
$T_1(6) = \dfrac{1}{(n-1)\sqrt{n}b_n}\displaystyle\sum_{i=1}^n\sum_{j\neq i}^n \epsilon_i\omega(x_i\theta)(y_j^2K(\frac{x_j\theta - x_i\theta}{b_n}) - b_nN_\theta(x_i\theta)).$

On a,

$$E(\epsilon_i\omega(x_i\theta)(y_j^2K(\frac{x_j\theta - x_i\theta}{b_n}) - b_nN_\theta(x_i\theta))) =$$
$$E(E(\epsilon_i\omega(x_i\theta)(y_j^2K(\frac{x_j\theta - x_i\theta}{b_n}) - b_nN_\theta(x_i\theta))|x_i)) =$$
$$E\{E(\epsilon_i|_{x_i})E(\omega(x_i\theta)(y_j^2K(\frac{x_j\theta - x_i\theta}{b_n}) - b_nN_\theta(x_i\theta))|x_i)\}$$

car $(y_i, x_i)_{i\geq 1}$ sont indépendants. Mais comme $E(\epsilon_i|x_i) = 0$, on a $E(T_1(6)) = 0$ et il nous suffit donc de montrer que la variance de $T_1(6)$ tend vers zéro quand n tend vers l'infini. Posons

$$T_1(6) = \frac{1}{(n-1)\sqrt{n}b_n}\sum_{i=1}^n\sum_{j\neq i}^n \epsilon_i\psi_{ij}$$

avec $\psi_{ij} = \omega(x_i\theta)(y_j^2K(\frac{x_j\theta - x_i\theta}{b_n}) - b_nN_\theta(x_i\theta))$. Rappelons que dans la démonstration du lemme 2, on montre d'abord que $E(\frac{1}{b_n}y_j^2K(\frac{x_j\theta - x_i\theta}{b_n})|x_i) - N_\theta(x_i\theta) = O(b_n^2)$ et $E(\frac{1}{b_n}K(\frac{x_j\theta_o - x_i\theta_o}{b_n})|x_i) - D_{\theta_o}(x_i\theta_o) = O(b_n^2)$. La variance de $T_1(6)$ est égale à

$$\frac{1}{(n-1)^2nb_n^2}E\left(\sum_{i=1}^n\sum_{j\neq i}^n\sum_{k=1}^n\sum_{l\neq k}^n \epsilon_i\epsilon_k\psi_{ij}\psi_{kl}\right)$$

$$= \frac{(n-2)}{(n-1)b_n^2}E(\epsilon_i^2\psi_{ij}\psi_{ik}) + \frac{1}{(n-1)b_n^2}E(\epsilon_i^2\psi_{ij}^2) + \frac{1}{(n-1)b_n^2}E(\epsilon_i\epsilon_j\psi_{ij}\epsilon_{ji})$$

avec i, j, et k tous différents. On a $\frac{1}{b_n}E(\epsilon_i^2\psi_{ij}\psi_{ik}) = E\{\epsilon_i^2E(\frac{1}{b_n}\psi_{ij}|(y_i, x_i))$ $\times E(\frac{1}{b_n}\psi_{ik}|(y_i, x_i))\} = O(b_n^4)$. Aussi par l'inégalité de Cauchy Swartz on a $E(\epsilon_i\epsilon_j\psi_{ij}\psi_{ji})$ $\leq (E(\epsilon_i^2\psi_{ij}^2))^{1/2}(E(\epsilon_j^2\psi_{ji}^2))^{1/2}$ et $\frac{1}{b_n}E(\epsilon_i^2\psi_{ij}^2) = E\{\epsilon_i^2\omega^2(x_i\theta)$ $\times E(\frac{1}{b_n}y_j^2K(\frac{x_j\theta - x_i\theta}{b_n} - N(x_i\theta))^2|(y_i, x_i)\}$. Par la décomposition en variance et en biais et avec l'hypothèse $E(y_j^4|x_j\theta = z)$ est une fonction deux fois continûment dérivable on obtient $E((\frac{1}{b_n}y_j^2K(\frac{x_j\theta - x_i\theta}{b_n})^2|(y_i, x_i)) = C_1 + \frac{C_2}{b_n} + O(b_n) + O(b_n^2)$. En fin $\omega(x_i\theta)$ et la variance de y_i sachant $x_i\theta_o$ étant bornés et comme b_n converge vers zéro, nb_n et nb_n^2 convergent vers l'infini, la variance de $T_1(6)$ tend vers zéro. L'inégalité de Chebyshev complète la preuve.
Pour la seconde partie du théorème, on a
$\dfrac{1}{n}\displaystyle\sum_{i=1}^n \frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial\hat{r}_\theta(x_i\theta)}{\partial\theta}\frac{\partial\hat{r}_\theta(x_i\theta)}{\partial\theta'} - M(\theta_o)$
$= \dfrac{1}{n}\displaystyle\sum_{i=1}^n \frac{1}{\hat{w}_\theta(x_i\theta)}\frac{\partial\hat{r}_\theta(x_i\theta)}{\partial\theta} frac\partial\hat{r}_\theta(x_i\theta)\partial\theta' - \frac{1}{n}\sum_{i=1}^n \frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\theta}\frac{\partial r_\theta(x_i\theta)}{\partial\theta'}$

$$+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}\frac{\partial r_\theta(x_i\theta)}{\partial\theta'}-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta_o}-M(\theta_o).$$

Par la loi forte des grands nombres le dernier terme est un $o(1)$. Le premier terme est majoré par $\sup_{(x,\theta)\in S\times\Theta}|\frac{1}{\hat{w}_\theta(x\theta)}\frac{\partial\hat{r}_\theta(x\theta)}{\partial\theta}\frac{\partial\hat{r}_\theta(x\theta)}{\partial\theta'}-\frac{1}{w_\theta(x\theta)}\frac{\partial r_\theta(x\theta)}{\partial\theta}\frac{\partial r_\theta(x\theta)}{\partial\theta'}|$, qui est un $o_p(1)$ par le Lemme 1 et Lemme 3. D'autre part S et $\bar{\Theta}$ étant compact, par l'inégalité des accroissements finis il existe C_1 telle que
$\sup_{\theta\in\bar{\Theta}}|\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{w_\theta(x_i\theta)}\frac{\partial r_\theta(x_i\theta)}{\partial\theta}\frac{\partial r_\theta(x_i\theta)}{\partial\theta'}-\frac{1}{w_{\theta_o}(x_i\theta_o)}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta}\frac{\partial r_\theta(x_i\theta_o)}{\partial\theta'})|$
$\leq C_1 n^{-1/2}\alpha$, ce qui achève la preuve.

Références

[1] Bonneu, M., Delecroix, M. and Hristache, M. (1995) Semiparametric estimation of generalized linear models and related models, *soumis à statistics.*

[2] Delecroix, M. et Hristache, M. (1995) M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique, *soumis à Bulletin de la Société Mathématique de Belgique.*

[3] Hristache, M. (1995) Calcul des bornes d'efficacité dans les modèles de régression semi-paramétriques, *Preprint, CREST, Paris.*

[4] Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single index models, *Journal of Econometrics,* **58***, 71-120.*

[5] Mac, Cullagh, P.(1983) Quasi-likelihood functions,*Annals of Statistics,* **11***, 59-67.*

[6] Newey, W., K. and Stoker, T., M. (1993) efficiency of weighted average derivative estimators in index models, *Econometrica ,* **61***, 1199-1223.*

[7] Newey, W., K. (1990) Semiparametric efficiency bounds, *Journal of Applied Econometrics,* **5***, 99-135.*

[8] Sherman, R., P. (1994) U-process in the analysis of a generalized semiparametric regression estimator, *Econometric theory,* **10***, 372-395.*

[9] Wedderburn, R., W., M. (1974) Quasi-likelihood functions, generalizedlinear models and the Newton model, *Biometrika,* **61***, 439-447.*

[10] Wefelmeyer, W. (1996) Quasi-likelihood models and optimal inference, *the annals of statistics,* **24***, 1, 405-422.*

[11] Weisberg, S. and Welsh, A., H. (1994) Adapting for the missing link, *Annals of Statistics,* **22***, 4, 1674-1700.*

IRMAR, Université de Rennes 1