LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

CONSISTENCY OF A MINIMUM-ENTROPY ESTIMATOR OF LOCATION

Eric Wolsztynski, Eric Thierry, Luc Pronzato

Projet TOPMODEL

Rapport de recherche
ISRN I3S/RR–2004-38–FR

Novembre 2004

LABORATOIRE I3S: Les Algorithmes / Euclide B – 2000 route des Lucioles – B.P. 121 –
06903 Sophia-Antipolis Cedex, France – Tél. (33) 492 942 701 – Télécopie : (33) 492 942 898
http://www.i3s.unice.fr/I3S/FR/

RÉSUMÉ :

MOTS CLÉS :
aberrations, efficacité, entropie, estimation adaptative, estimation paramétrique, modèle semi-paramétrique, robustesse

ABSTRACT:

KEY WORDS :
adaptive estimation, efficiency, entropy, parameter estimation, semi-parametric models, robustness, outliers

Consistency of a Minimum-Entropy Estimator of Location *

Eric Wolsztynski† Eric Thierry† Luc Pronzato†

November 10, 2004

Abstract

In regression problems where the density f of the errors is not known, maximum likelihood is unapplicable, and the use of alternative techniques (least squares, robust M-estimation,...) generally results in inefficient estimation of the parameters. We consider here an alternative parametric estimator that was first presented in [13, 12], using an empirical estimate of the entropy of the residuals as the criterion to minimize. Adaptivity of this estimator, that is, the property for the estimator of remaining asymptotically efficient independently of the knowledge of f (see in particular [16, 17, 4] and the review paper [10]), is discussed in [14], where we consider a direct approach, as opposed to the Stone-Bickel 2-step estimator, which involves a preliminary $\sqrt{n}$-consistent estimator. In the following we give a detailed proof of consistency of the estimator and discuss the steps mentioned in [14] to prove its adaptivity.

Keywords: adaptive estimation, efficiency, entropy, parameter estimation, semi-parametric models, robustness, outliers

1 Semi-parametric minimum-entropy estimation in nonlinear regression : problem statement

We consider the problem of parameter estimation in general nonlinear regression models. The observations Y_i for such models correspond to the measures made from a parametric model $\eta(\theta, x)$, which is a known function of θ and the design variable $x \in \mathcal{X} \subset \mathbb{R}^d$, corrupted by some additive noise ε of density f :

$$Y_i = \eta(\bar{\theta}, X_i) + \varepsilon_i\,, \quad i = 1, \ldots, n\,, \tag{1}$$

[2]{wolsztyn,et,pronzato}@i3s.unice.fr, Laboratoire I3S, 2000 route des Lucioles, 06903 Sophia Antipolis

*This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

where $\bar{\theta}$ denotes the unknown value of the model parameters $\theta \in \Theta \subset I\!R^p$.
We first introduce some notations. For F a function $\Theta \to I\!R$, $\nabla F(\theta)$ and $\nabla^2 F(\theta)$ will denote its first and second order derivatives with respect to θ, respectively a p-dimensional vector and a $p \times p$ symmetric matrix. For g a function $I\!R \to I\!R$, the first, second and third order derivatives are simply denoted g', g'' and g'''. Also, the following assumptions will hold throughout the paper. We suppose that $\bar{\theta} \in \text{int}(\Theta)$, $\Theta = \overline{\text{int}(\Theta)}$, and that $\eta(\theta, x)$ is bounded on $\Theta \times \mathcal{X}$ and two times continuously differentiable w.r.t. $\theta \in \text{int}(\Theta)$ for any $x \in \mathcal{X}$, $\nabla\eta(\theta, x)$ and $\nabla^2\eta(\theta, x)$ being bounded on $\text{int}(\Theta) \times \mathcal{X}$. The additive noise (ε_i) forms a sequence of independently and identically distributed (i.i.d.) random variables with probability density function (p.d.f.) f (with respect to the Lebesgue measure) that we suppose to be symmetric about zero and of unbounded support. f is also assumed two times continuously differentiable, having bounded derivatives $f'(.)$, $f''(.)$ and $f'''(.)$, and we suppose that the Fisher information for location

$$\mathcal{I}(f) = \int_{-\infty}^{\infty} [f'(u)]^2 / f(u)\, du$$

exists. For a given measure μ on the design variable x, the Fisher information matrix $\mathbf{M}_F(\theta)$ associated with f and θ is given by

$$\mathbf{M}_F(\theta) = \mathcal{I}(f) \int_{\mathcal{X}} \nabla\eta(\theta, x)[\nabla\eta(\theta, x)]^\top \mu(dx)\,. \tag{2}$$

We suppose that $\mathbf{M}_F(\bar{\theta})$ has full rank and that the identifiability condition

$$\int_{\mathcal{X}} [\eta(\theta, x) - \eta(\bar{\theta}, x)]^2 \mu(dx) = 0 \;\Rightarrow\; \theta = \bar{\theta} \tag{3}$$

is satisfied.

We know that, under standard assumptions, the Maximum Likelihood (ML) estimator $\hat{\theta}^n_{ML}$ is asymptotically efficient, that is,

$$\sqrt{n}(\hat{\theta}^n_{ML} - \bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{M}_F^{-1}(\bar{\theta}))\,.$$

In the present context, f is not known further to the assumptions made above, which implies that we are not in the classical situation where ML estimation is possible, and we need an estimator that does not require the knowledge of f. The model (1) is then called *semi-parametric* (as Stein first termed it in [16]) with θ and f respectively its parametric and non-parametric parts, and f can be considered as an infinite-dimensional nuisance parameter for the estimation of θ. The ultimate goal is then to define an estimator of θ that is *asymptotically efficient*, i.e. asymptotically normal with minimum variance without requiring the knowledge of f. In general this proves to be difficult : the presence of a nuisance parameter of infinite dimension induces the loss of efficiency.

One can refer to [14] for a review of the developments on adaptive estimation. Our present goal is to provide a proof of consistency (along with other technical results) of an estimator we presented earlier, which consists in minimizing the entropy of a kernel estimate constructed from the symmetrized residuals in the regression model (1). In the next section, we recall some definitions for p.d.f. estimation and entropy estimation and we introduce some notations used later. Section 3 contains the main result on the consistency of the estimator for the simple location model. We also detail further technical points like convergence of the second order derivative of the estimation criterion to the Fisher information matrix (subsection 3.2 provides a brief reminder of the rather standard approach we first considered for proving adaptivity, which motivated these developments). Finally, some perspectives are given in section 4, where extension to the general non-linear regression model is considered, although not in full details.

2 Estimation by minimizing the entropy of the residuals

2.1 Entropy of the true density

Consider the residuals $e_i(\theta)$ obtained from the observations in the regression model (1),

$$e_i(\theta) = Y_i - \eta(\theta, X_i) = \varepsilon_i + \eta(\bar{\theta}, X_i) - \eta(\theta, X_i)\,.$$

When f is known, the Maximum Likelihood estimator $\hat{\theta}^n_{ML}$ minimizes

$$\bar{H}_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\log f[e_i(\theta)] \tag{4}$$

with respect to $\theta \in \Theta$. Since $\bar{H}_n(\bar{\theta}) = -(1/n)\sum_{i=1}^{n}\log f(\varepsilon_i)$ is an empirical version of the (Shannon) entropy

$$H(f) = -\int_{-\infty}^{\infty}\log[f(u)]f(u)du\,,$$

an intuitive idea is to base the estimation criterion on entropy (minimizing the entropy of the residuals forces them to gather). The density of $e_i(\theta)$, given X_i, is

$$f_{e,\,X_i}(u) = f(u - \eta(\bar{\theta}, X_i) + \eta(\theta, X_i)).$$

Since entropy is invariant by translation, we shall consider the $2n$ symmetrized residuals $e_i(\theta), -e_i(\theta)$, with corresponding density given X_i

$$f^s_{e,\,X_i}(u) = \frac{1}{2}\left[f(u - \eta(\bar{\theta}, X_i) + \eta(\theta, X_i)) + f(u + \eta(\bar{\theta}, X_i) - \eta(\theta, X_i))\right]. \tag{5}$$

Other techniques could be considered; we could for instance use un-symmetrized residuals with the constraint that their median, or their mean, is zero; however, symmetrization is conceptually simpler and requires less numerical calculation.

We can show that the entropy of the marginal distribution of the symmetrized residuals,

$$f_e^s(u) = \int_{\mathcal{X}} f_{e,\,x}^s(u)\mu(dx) \tag{6}$$

is minimum for $\theta = \bar{\theta}$. For this purpose we can use the following lemma

Lemma 1 ([1], Lemma 8.3.1 p.238) *Let $p(x)$ and $q(x)$ be arbitrary probability density functions.*

a. *If $-\int_{-\infty}^{\infty} p(x)\log q(x)dx$ is finite, then $-\int_{-\infty}^{\infty} p(x)\log p(x)dx$ exists, and furthermore*

$$-\int_{-\infty}^{\infty} p(x)\log p(x)dx \leq -\int_{-\infty}^{\infty} p(x)\log q(x)dx\,,$$

with equality if and only if $p(x) = q(x)$ for almost all x (with respect to Lebesgue measure).

b. *If $-\int_{-\infty}^{\infty} p(x)\log p(x)dx$ is finite, then $-\int_{-\infty}^{\infty} p(x)\log q(x)dx$ exists, and the above inequality holds.*

It gives

$$\begin{aligned} H(f_e^s) &= -\frac{1}{2}\int_{\mathcal{X}}\left[\int_{-\infty}^{\infty} f[u-\eta(\bar{\theta},x)+\eta(\theta,x)]\log[f_e^s(u)]\,du\right]\mu(dx) \\ &\quad -\frac{1}{2}\int_{\mathcal{X}}\left[\int_{-\infty}^{\infty} f[u+\eta(\bar{\theta},x)-\eta(\theta,x)]\log[f_e^s(u)]\,du\right]\mu(dx) \\ &\geq -\int_{\mathcal{X}}\left[\int_{-\infty}^{\infty} f(u)\log[f(u)]\,du\right]\mu(dx) = H(f)\,. \end{aligned}$$

Equality is obtained only if for μ-almost all x and almost all u (Lebesgue)

$$f[u-\eta(\bar{\theta},x)+\eta(\theta,x)] = f[u+\eta(\bar{\theta},x)-\eta(\theta,x)] = f_e^s(u)\,.$$

From the identifiability condition (3), this implies $\theta = \bar{\theta}$. The same is true for the conditional entropy of the symmetrized residuals

$$\begin{aligned} \mathtt{E}_\mu\{H(f_{e,X}^s)\} &= -\int_{\mathcal{X}}\left[\int_{-\infty}^{\infty} f_{e,x}^s(u)\log[f_{e,x}^s(u)]\,du\right]\mu(dx) \\ &\geq -\int_{\mathcal{X}}\left[\int_{-\infty}^{\infty} f(u)\log[f(u)]\,du\right]\mu(dx) = H(f)\,, \end{aligned}$$

with equality attained only if $H(f_{e,x}^s) = H(f)$ for μ-almost all x, which again implies $\theta = \bar{\theta}$. Recall now the following classical result in information theory, see, e.g., [1] p. 239. Consider a random variable B with conditional p.d.f. given X $f_{e,x}^s$ and unconditional p.d.f. f_e^s. Then, $H(B|X) = \mathtt{E}_\mu[H(f_{e,X}^s)] \leq H(B) = H(f_e^s)$ (conditioning reduces entropy). We thus have

$$\mathtt{E}_\mu\{H(f_{e,X}^s)\} \leq H(f_e^s)\,.$$

Finally, we can perform a local study of $H(f_e^s)$ around $\theta = \bar{\theta}$. Direct calculations give

$$\nabla f_e^s(u)_{|\theta=\bar{\theta}} = \mathbf{0}\,,\ \text{and}\ \nabla^2 f_e^s(u)_{|\theta=\bar{\theta}} = f''(u) \int_{\mathcal{X}} \nabla\eta(\bar{\theta},x)\,[\nabla\eta(\bar{\theta},x)]^\top\,\mu(dx).$$

The entropy of f_e^s satisfies

$$\begin{aligned}
\nabla H(f_e^s) &= -\int_{-\infty}^{\infty} [1+\log f_e^s(u)]\,\nabla f_e^s(u)\,du,\\
\nabla^2 H(f_e^s) &= -\int_{-\infty}^{\infty} \frac{1}{f_e^s(u)} \nabla f_e^s(u) [\nabla f_e^s(u)]^\top\,du\\
&\quad -\int_{-\infty}^{\infty} [1+\log f_e^s(u)]\,\nabla^2 f_e^s(u)\,du\,.
\end{aligned}$$

Since $(f\log f)'' = (1+\log f)f'' + (f')^2/f$, we obtain

$$\nabla H(f_e^s)_{|\theta=\bar{\theta}} = \mathbf{0}\,,\ \nabla^2 H(f_e^s)_{|\theta=\bar{\theta}} = \mathbf{M}_F(\bar{\theta})\,, \tag{7}$$

that is, the entropy $H(f_e^s)$ is locally concave with zero derivative at $\theta = \bar{\theta}$. Similar results are obtained for the conditional entropy $\mathtt{E}_\mu\{H(f_{e,X}^s)\}$,

$$\forall x \in \mathcal{X}\,,\ \nabla f_{e,x}^s(u)_{|\theta=\bar{\theta}} = \mathbf{0}\,,\quad \nabla^2 f_{e,x}^s(u)_{|\theta=\bar{\theta}} = f''(u)\,\nabla\eta(\bar{\theta},x)\,[\nabla\eta(\bar{\theta},x)]^\top\,,$$

and

$$\nabla \mathtt{E}_\mu\{H(f_{e,X}^s)\}_{|\theta=\bar{\theta}} = \mathbf{0}\,,\ \nabla^2 \mathtt{E}_\mu\{H(f_{e,X}^s)\}_{|\theta=\bar{\theta}} = \mathbf{M}_F(\bar{\theta})\,.$$

2.2 Estimating the entropy of the residuals

Neither $\bar{H}_n(\theta)$ given by (4), $H(f_e^s)$ nor $\mathtt{E}_\mu\{H(f_{e,X}^s)\}$ can be used as criteria for parameter estimation, since f and $\bar{\theta}$ are unknown. We thus need to define a criterion approaching $H(f_e^s)$. We can construct an estimate of this last quantity by simply plugging a symmetric kernel estimate $\hat{f}_n^\theta$ of f_e^s into the expression of the exact entropy $H(f_e^s)$. For technical reasons (related to the proof of consistency of the associated estimators) we introduce a truncation to $[-A_n, A_n]$ and the criterion to be minimized for the estimation of θ is

$$\hat{H}_n(\theta) = -\int_{-A_n}^{A_n} \log[\hat{f}_n^\theta(u)]\hat{f}_n^\theta(u)du\,, \tag{8}$$

where (A_n) is a suitably (slowly) increasing sequence of positive numbers (to be chosen in accordance with the decrease of the bandwidth h_n of the kernel estimate $\hat{f}_n^\theta$, see [13, 12]). Similarly, when a kernel estimate $f_{n_i}^{i,\theta}$ of the conditional distribution f_{e,X_i}^s can be constructed (that is, for designs with replications, see [14]), we can plug it in H to approach $\mathtt{E}_\mu\{H(f_{e,X}^s)\}$.

A second estimator can be defined based on the empirical estimate $(1/n)\sum_i \log f(X_i)$ which converges to $H(f)$ from the law of large numbers. We also plug in the kernel estimate of the density f, yielding the following criterion to be minimized:

$$\hat{H}_n(\theta) = -\frac{1}{n}\sum_i \log \hat{f}_n^\theta(X_i). \tag{9}$$

Noting that using cross-validated kernels would reduce the bias for finite samples, we can also define a variant to that last estimator by

$$\hat{H}_n(\theta) = -\frac{1}{n}\sum_{i=1}^n \log \hat{f}_{n,\,i}^\theta(X_i)\,, \tag{10}$$

where $\hat{f}_{n,\,i}$ uses all but the i^{th} kernel.

Justifications of the estimators defined above can be found in [14]. In what follows we only consider the estimator (10). Extension to (8) and (9) can be obtained following similar steps. Only the case of the location model is considered, extension to a general nonlinear regression model will be discussed elsewhere.

3 Estimation of location

3.1 Model and further assumptions

The observations satisfy $Y_i = \bar{\theta} + \varepsilon_i$, $i = 1, \ldots, n$, which are i.i.d. with density $f(x + \bar{\theta})$. We consider the parameters $\theta \in \Theta$ with Θ satisfying the following assumption :

A1 Θ is a compact set with $\Theta \subset \overline{\text{int}(\Theta)}$, $\bar{\theta} \in \text{int}(\Theta)$.

We denote $L = \max_{(\theta,\,\theta')\in\Theta} |\theta - \theta'|$ the diameter of Θ.

For any $\theta \in \Theta$ we form the n residuals $e_i(\theta) = Y_i - \theta$, $i = 1, \ldots, n$, and use the symmetrized sample $-e_i(\theta)$, $e_i(\theta)$ which yields the density

$$f_e^s(x) = f_e^{s,\,\theta}(x) = \frac{1}{2}[f(x + \bar{\theta} - \theta) + f(x - \bar{\theta} + \theta)]$$

(f_e^s depends on θ since the residuals are functions of the parameter, but we omit the index in further notations whenever it does not affect clarity). We construct the following estimate of $f_e^s(u)$

$$\hat{f}_n^\theta(u) = \frac{1}{2}\left[k_n^\theta(u) + k_n^\theta(-u)\right] \tag{11}$$

to be used to compute $\hat{H}_n(\theta)$ given by (10), with

$$k_n^\theta(u) = \frac{1}{n\,h_n}\sum_{i=1}^n K\left(\frac{u - e_i(\theta)}{h_n}\right)$$

where $K(.)$ satisfies the following standard conditions (see for instance [15]) :

Conditions K

K1 $K(.)$ is symmetric about zero ;

K2 $\int_{-\infty}^{\infty} |u|K(u)du < \infty$;

K3 K is two times continuously differentiable with derivatives of bounded variation ;

K4 the bandwidth h_n of the kernels satisfies $h_n \to 0$, $nh_n \to \infty$ as $n \to \infty$.

$\hat{f}_n^\theta$ is then a kernel density estimate based on the $2n$ symmetrized residuals $\pm e_i(\theta)$. A classical choice is the normal density $K(u) = 1/\sqrt{2\pi}\, \exp(-u^2/2)$.

We also define a smooth truncation for large residuals. Let U_n be such that $U_n(z) = U(|z|/A_n - 1)$ with

$$\begin{cases} U(z) = 1 & \text{for } z \leq 0, \\ U(z) = 0 & \text{for } z \geq 1 \\ U(z) & \text{varying smoothly between 0 and 1}, \end{cases} \tag{12}$$

with $U'(0) = U'(1) = 0$, $\max_z |U'(z)| = d_1 < \infty$, $\max_z |U''(z)| = d_2 < \infty$, and (A_n) a sequence of (slowly) increasing positive numbers. We consider a truncated version of the estimator (10) using (12),

$$\hat{H}_n(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_{n,\,i}^\theta[e_i(\theta)]\, U_n[e_i(\theta)] \tag{13}$$

where $\hat{f}_{n,\,i}^\theta$ is similar to (11), but does not use $e_i(\theta)$, that is,

$$\hat{f}_{n,\,i}^\theta(u) = \frac{1}{2} \left[k_{n,\,i}^\theta(u) + k_{n,\,i}^\theta(-u)\right] \tag{14}$$

with

$$k_{n,\,i}^\theta(u) = \frac{1}{(n-1)\,h_n} \sum_{j=1,\, j\neq i}^{n} K\left(\frac{u - e_j(\theta)}{h_n}\right),\ i = 1, \ldots, n\,.$$

We shall make the following assumptions on f.

Conditions F

F0 f is symmetric about 0, $f(x) > 0\ \forall x \in \mathbb{R}$ and $f(x)$ and $|f'(x)|$ are bounded.

F1 $\int |\log f(x)| f(x-a)dx < \infty\ \forall a \in [-2L, 2L]$.

F2 there exists a positive constant C such that for any $x > C$, f satisfies $f(x) < 1$, $f(x)$ and $|f'(x)|$ are decreasing.

F2' $|f''(x)|$ and $|f'''(x)|$ are bounded for $x \in I\!R$ and decreasing for $x > C$.

F3 there exists a strictly increasing function B such that for all $u \in I\!R$, $B(u) \geq \sup_{|y|<u} 1/f(y)$.

F4 $\int (f'/f)^2 f < \infty$.

F5 $\int (f'(x-a)/f(x))^2 f(x) < \infty$, $\int (f'(x-a)/f(x))^4 f(x) < \infty$, $\int (f''(x-a)/f(x))^2 f(x) < \infty \; \forall a \in [0, 2L]$.

F6 $\int |f'''(x)| dx < \infty$.

Notice that $\hat{H}_n(\theta)$ is two times continuously differentiable w.r.t. $\theta \in \text{int}(\Theta)$. Convergence in probability when $n \to \infty$ will be denoted $\xrightarrow{\text{p}}$ ($\overset{\theta,\text{p}}{\rightsquigarrow}$ will be used when the convergence is uniform with respect to θ), and convergence in distribution will be denoted $\xrightarrow{\text{d}}$.

3.2 Three steps towards adaptivity and technical issues

Under common measurability conditions (see, *e.g.*, Lemmas 2 and 3 of [8]) the standard, and rather general, approach for proving the consistency and asymptotic normality of $\hat{\theta}^n$ minimizing some criterion $\hat{H}_n(\theta)$ can be decomposed into three steps:

A) show that $\hat{H}_n(\theta) \overset{\theta,\text{p}}{\rightsquigarrow} H(\theta)$, $n \to \infty$, with $\hat{H}_n(\theta)$ continuous in θ for any n, and that $H(\bar{\theta}) < H(\theta)$ for any $\theta \neq \bar{\theta}$;

B) show that $\nabla^2 \hat{H}_n(\theta) \overset{\theta,\text{p}}{\rightsquigarrow} \nabla^2 H(\theta)$, $n \to \infty$, with $\nabla^2 H(\bar{\theta})$ positive definite ($\succ 0$);

C) decompose $\nabla \hat{H}_n(\bar{\theta})$ into $\nabla \bar{H}_n(\bar{\theta}) + \Delta_n(\bar{\theta})$, with $\sqrt{n} \nabla \bar{H}_n(\bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{M}_1)$ and $\sqrt{n} \Delta_n(\bar{\theta}) \xrightarrow{\text{p}} \mathbf{0}$ as $n \to \infty$.

The first two points are proved by Theorems 1 and 2, given in section 3.3 below. As mentioned in [14], the uniform convergence in (A) proves the weak consistency of $\hat{\theta}^n$ ($\hat{\theta}^n \xrightarrow{\text{p}} \bar{\theta}$). (A) and (B) imply that the second order derivative of the criterion taken at $\hat{\theta}^n$ converges to $\nabla^2 \hat{H}_n(\hat{\theta}^n) \xrightarrow{\text{p}} \mathbf{M}_2 = \nabla^2 H(\bar{\theta})$ as $n \to \infty$, and $\nabla^2 H(\bar{\theta}) = \mathbf{M}_F(\bar{\theta})$ from (7).

Finally, consider the following Taylor expansion of $\nabla \hat{H}_n(\theta)$ at $\theta = \hat{\theta}^n$,

$$\nabla \hat{H}_n(\hat{\theta}^n) = \mathbf{0} = \nabla \hat{H}_n(\bar{\theta}) + (\hat{\theta}^n - \bar{\theta})^\top \nabla^2 H[\alpha_n \hat{\theta}^n + (1 - \alpha_n)\bar{\theta}] \,,$$

with $\alpha_n \in [0, 1]$ (see [8] who uses a similar approach for LS estimation). (C) then implies asymptotic normality, that is, $\sqrt{n}(\hat{\theta}^n - \bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{M}_2^{-1} \; \mathbf{M}_1 \; \mathbf{M}_2^{-1})$. The adaptivity of $\hat{\theta}^n$ would then directly follow from $\mathbf{M}_2^{-1} \mathbf{M}_1 \mathbf{M}_2^{-1} = \mathbf{M}_F^{-1}(\bar{\theta})$, the inverse of the Fisher information matrix (2).

One may notice that step (C) allows some freedom in the choice of the function $\bar{H}_n(\theta)$, even though it would be natural to pick (4), for which the asymptotic normality $\sqrt{n} \nabla \bar{H}_n(\bar{\theta}) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{M}_1)$ holds under standard assumptions, with

$\mathbf{M}_1 = \mathbf{M}_F(\bar{\theta})$ (asymptotic properties of the ML estimator). Also, according to the review [2], $\sqrt{n}$-consistency of $\hat{H}_n(\theta)$ is difficult to obtain, but notice that it is not a prerequisite for $\sqrt{n}$-consistency of $\hat{\theta}^n$ (we only need $\sqrt{n}\Delta_n(\bar{\theta}) \xrightarrow{\mathrm{p}} 0$, $n \to \infty$).

A key step to prove adaptivity of $\hat{\theta}^n$ at step (C) would be to show that

$$-\frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{(k_{n,i}^{\bar{\theta}})'(-\varepsilon_i)}{k_{n,i}^{\bar{\theta}}(\varepsilon_i) + k_{n,i}^{\bar{\theta}}(-\varepsilon_i)} U_n(\varepsilon_i) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathcal{I}(f)) ,$$

the term on the left-hand side being the major contribution to $\sqrt{n}\nabla\hat{H}_n(\bar{\theta})$ when $\hat{H}_n(\theta)$ is given by (13). The conditions required on the functions f, K and U for this to hold are currently under investigation. The rest of this paper is dedicated to proving points A and B.

3.3 Main results

The first result concerns the consistency of the estimator defined by the criterion (13) and is given by the following theorem. Its proof is given in Appendix C.

Theorem 1 *Consider observations having the density $f(x - \bar{\theta})$, $\bar{\theta} \in \Theta$ with Θ satisfying* **A1**, *and with f satisfying conditions* **F0-F2,F3,F4**. *Let (δ_n) be a positive sequence such that $\delta_n \to 0$, $n^\gamma \delta_n \to \infty$ $\forall\, \gamma > 0$. Let $\hat{H}_n$ be as defined in (13) where the kernels $K(.)$ satisfy* **K0-K4** *with $h_n = n^{-\alpha}\delta_n^2$ in* **K4** *and where U_n satisfies (12) with $B(3A_n) = n^\alpha$ with $B(.)$ defined in* **F3**. *Then, for $\alpha < 1/3$, $\hat{H}_n(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} H(\theta)$, $n \to \infty$.*

Since $\hat{H}_n(\theta)$ is continuous in θ for any n and $H(\bar{\theta}) < H(\theta)$ for any $\theta \neq \bar{\theta}$, Theorem 1 implies that the estimator minimizing (13) is consistent : $\hat{\theta}^n \xrightarrow{p} \bar{\theta}$ when $n \to \infty$.

Similarly, with slightly stronger conditions on f, we can prove the following concerning the second order derivative of the criterion. The proof is given in Appendix C.

Theorem 2 *Consider observations having the density $f(x - \bar{\theta})$, $\bar{\theta} \in \Theta$ with Θ satisfying* **A1**, *and with f satisfying all conditions* **F**. *Let (δ_n) be a positive sequence such that $\delta_n \to 0$, $n^\gamma \delta_n \to \infty$ $\forall\, \gamma > 0$. Let $\hat{H}_n$ be as defined by (13), where the kernels $K(.)$ satisfy* **K** *with $h_n = n^{-\alpha}\delta_n$ in* **K4** *and where U_n satisfies (12) with $B(3A_n) = n^{\alpha/3}$ with $B(.)$ defined in* **F3**. *Then, for $\alpha < 1/7$ the criterion $\hat{H}_n$ satisfies $\nabla^2\hat{H}_n(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} \nabla^2 H(\theta)$ as $n \to \infty$.*

Note that under the conditions of Theorems 1 and 2,

$$\nabla^2\hat{H}_n(\hat{\theta}^n) \xrightarrow{p} \nabla^2 H(\bar{\theta}) = \mathbf{M}_F^{-1}(\bar{\theta}),$$

see (7).

4 Conclusion and extensions

Extension of the present approach to general non-linear models should be similar to the results given here, and will be developed elsewhere. As explained in [14] two situations can be considered in this case, involving either the conditional entropy of residuals resulting from repetitions at fixed points, or the entropy of all residuals mixed together. The main difference with the location problem is in the introduction of dependency of the densities on the model function $\eta(\theta, X)$ and thus on the regressors X. The general approach should however remain the same.

The dependence of the kernel estimates (11) or (14) in θ makes the derivation of the asymptotic properties of $\hat{\theta}^n$ minimizing (8) or (13) much more difficult than for the minimizer of an approximation of the score-function used in ML estimation, or equivalently the ML criterion (4). In particular, the adaptivity of $\hat{\theta}^n$ is still an open question (see point C in section 3.2).

Several methods exist for entropy estimation, and each of them could be used to define a minimum-entropy estimator. Some do not require kernel smoothing of the empirical density of residuals, which could be considered as an advantage over the plug-in estimates used in this paper, see, *e.g.*, [18] for a sample-spacing method and [9] for an approach relying on nearest neighbors (in particular, the latter applies for samples in any dimension k, and could be used for minimum-entropy estimation in multiple regression where Y_i is then a k-dimensional vector). However, investigating the asymptotic properties of their associated minimum-entropy estimators seems a very difficult task. Another direction would be to consider recent developments in parametric estimation via divergence minimization. In a parametric context (which, for regression models, means that f is known), the asymptotically efficient estimator of [3], based on minimizing Hellinger distance, requires smoothing of the empirical distribution in order to compute its distance to a distribution with density. On the other hand, the approach used by [5, 6] is based on a duality property that permits to estimate the divergences of interest without requiring smoothing. The application to the semi-parametric problem considered in the paper is an open and motivating issue.

Appendix

Appendix A : useful expressions and inequalities

We indicate below the expressions of the first two derivatives of the true density of the residuals and of the kernel density estimators of f_e^s, $\hat{f}_n^\theta(e_i(\theta))$ and $\hat{f}_{n,i}^\theta(e_i(\theta))$ w.r.t. θ, along with an upper bound on their distances. We also give the first two derivatives of the kernel estimator

$$\hat{g}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - Y_i}{h_n}\right) \quad (15)$$

of the true density f w.r.t. x. For $e_i(\theta) = Y_i - \theta$, we have :

$$\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\} = -f'(Y_i + \bar{\theta} - 2\theta) \tag{16}$$

$$\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\} = 2f''(Y_i + \bar{\theta} - 2\theta) \tag{17}$$

$$\frac{\partial}{\partial\theta}\left\{\hat{f}_n^\theta(e_i(\theta))\right\} = -\frac{1}{nh_n^2}\sum_{j=1}^n K'\left(\frac{Y_i + Y_j - 2\theta}{h_n}\right) \tag{18}$$

$$\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\} = -\frac{1}{(n-1)h_n^2}\sum_{j=1,\,j\neq i}^n K'\left(\frac{Y_i + Y_j - 2\theta}{h_n}\right) \tag{19}$$

$$\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_n^\theta(e_i(\theta))\right\} = \frac{2}{nh_n^3}\sum_{j=1}^n K''\left(\frac{Y_i + Y_j - 2\theta}{h_n}\right) \tag{20}$$

$$\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\} = \frac{2}{(n-1)h_n^3}\sum_{j=1,\,j\neq i}^n K''\left(\frac{Y_i + Y_j - 2\theta}{h_n}\right) \tag{21}$$

$$\hat{g}_n'(x) = \frac{1}{nh_n^2}\sum_{j=1}^n K'\left(\frac{x - Y_j}{h_n}\right) \tag{22}$$

$$\hat{g}_n''(x) = \frac{1}{nh_n^3}\sum_{j=1}^n K''\left(\frac{x - Y_j}{h_n}\right) \tag{23}$$

Notice that

$$\hat{f}_n^\theta(x) - \hat{f}_{n,\,i}^\theta(x) = \frac{1}{2nh_n}\left[K\left(\frac{x - e_i(\theta)}{h_n}\right) + K\left(\frac{x + e_i(\theta)}{h_n}\right)\right] - \frac{\hat{f}_{n,\,i}^\theta(x)}{n},$$

yielding

$$\sup_{i,\,\theta}|\hat{f}_n^\theta(e_i(\theta)) - \hat{f}_{n,\,i}^\theta(e_i(\theta))| \leq \frac{K(0)}{nh_n} + \frac{K(0)}{nh_n} = \frac{2K(0)}{nh_n}\,. \tag{24}$$

In the same way we obtain :

$$\sup_{i,\,\theta}|\frac{\partial}{\partial\theta}\left\{\hat{f}_n^\theta(e_i(\theta))\right\} - \frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\}| \leq \frac{K'(0)}{nh_n^2} + \frac{K'(0)}{nh_n^2} = \frac{2K'(0)}{nh_n^2} \tag{25}$$

and

$$\sup_{i,\,\theta}|\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_n^\theta(e_i(\theta))\right\} - \frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\}| \leq 2\frac{K''(0)}{nh_n^3} + 2\frac{K''(0)}{nh_n^3} = \frac{4K''(0)}{nh_n^3}\,. \tag{26}$$

Appendix B : two useful lemmas

In this appendix we give two lemmas that we refer to in the proofs of Theorems 1 and 2.

Lemma 2 (Dmitriev and Tarasenko, 1973, p.632) *If f and its first $r+1$ derivatives are bounded, and $\{\varepsilon_n\}$ is a sequence of positive numbers such that $h_n = o(\varepsilon_n)$, there exists a positive constant $c < \infty$ such that for sufficiently large n,*

$$P\left\{\sup_y \left|\hat{g}_n^{(r)}(y) - f^{(r)}(y-\bar{\theta})\right| > \varepsilon_n\right\} \leq \frac{c}{nh_n^{2r+1}\varepsilon_n^2}.$$

Lemma 3 (Newey, 1991, Corollary 3.1) *Consider a sequence of functions $q_t(Y_t,\theta)$ of the observations Y_t and parameter vector $\theta \in \Theta$. Define $\hat{Q}_n(\theta) = \sum_{t=1}^n q_t(Y_t,\theta)/n$ and $\bar{Q}_n(\theta) = \sum_{t=1}^n \mathtt{E}[q_t(Y_t,\theta)]/n$. Suppose that Θ is a compact metric space and that for each $\theta \in \Theta$, $\hat{Q}_n(\theta) - \bar{Q}_n(\theta) = o_p(1)$. Let $h(.)$ denote a function $h : [0,\infty) \to [0,\infty)$ with $h(0) = 0$ and h continuous at 0, and suppose there exists a function $b_t(Y_t)$ such that $\sum_{t=1}^n \mathtt{E}[b_t(Y_t)]/n = \mathcal{O}(1)$, and $|q_t(Y_t,\tilde{\theta}) - q_t(Y_t,\theta)| \leq b_t(Y_t)h(d(\tilde{\theta},\theta))$ for $\tilde{\theta}$, $\theta \in \theta$. Then $\sup_{\theta\in\Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$.*

We shall use this lemma with $h(d(\tilde{\theta},\theta)) = d(\tilde{\theta},\theta) = |\tilde{\theta} - \theta|$.

Appendix C : proofs of Theorems 1 and 2

Proof of Theorem 1. The proof is based on Lemmas 2 and 3. We can break down the difference $\Delta_n(\theta) = \hat{H}_n(\theta) - H(\theta)$ into three separate terms $\Delta_n(\theta) = \Delta_n^1(\theta) + \Delta_n^2(\theta) + \Delta_n^3(\theta)$, with

$$\begin{aligned}
\Delta_n^1(\theta) &= -\frac{1}{n}\sum_{i=1}^n \left[\log \hat{f}_{n,\,i}^{\theta}(e_i(\theta)) - \log f_e^s(e_i(\theta))\right] U_n(e_i(\theta)) \\
\Delta_n^2(\theta) &= -\frac{1}{n}\sum_{i=1}^n \log f_e^s(e_i(\theta))\,[U_n(e_i(\theta)) - 1] \\
\Delta_n^3(\theta) &= -\frac{1}{n}\sum_{i=1}^n \log f_e^s(e_i(\theta)) + \int f_e^s(x)\log f_e^s(x)dx
\end{aligned}$$

a) Uniform convergence of $\Delta_n^1(\theta)$ to 0.
Applying Lemma 2, for $h_n = o(\varepsilon_n)$, there exists a positive constant c such that for $\hat{g}_n$ as defined by (15),

$$P\left\{\sup_z |\hat{g}_n(z) - f(z-\bar{\theta})| > \varepsilon_n\right\} \leq \frac{c}{nh_n\varepsilon_n^2}.$$

Since the convergence is uniform over Θ, shifting the densities $\hat{g}_n$ and f by θ does not change the last probability, and we can also write

$$P\left\{\sup_{x,\,\theta} |\hat{f}_n^{\theta}(x) - f_e^s(x)| > \varepsilon_n\right\} \leq \frac{c}{nh_n\varepsilon_n^2},$$

which yields for $\varepsilon'_n = \varepsilon_n + 2K(0)/(nh_n)$, see (24),

$$P\left\{\sup_{x,\,\theta,\,i} |\hat{f}_{n,\,i}^{\theta}(x) - f_e^s(x)| > \varepsilon'_n\right\} \leq \frac{c}{nh_n\varepsilon_n^2}. \tag{27}$$

Suppose now that $\sup_{x,\,\theta,\,i} |\hat{f}^{\theta}_{n,\,i}(x) - f^s_e(x)| < \varepsilon'_n$. We have

$$\begin{aligned}\sup_{\theta} |\Delta^1_n(\theta)| &\leq \sup_{\theta,\,i} \left\{ \left| \log \hat{f}^{\theta}_{n,\,i}(e_i(\theta)) - \log f^s_e(e_i(\theta)) \right| U_n(e_i(\theta)) \right\} \\ &\leq \sup_{\theta,\,i,\,|x|<2A_n} \left| \log \hat{f}^{\theta}_{n,\,i}(x) - \log f^s_e(x) \right| .\end{aligned}$$

For $|x| < 2A_n$, $|x \pm (\theta - \bar{\theta})| < 2A_n + L$, and $f^s_e(x) \geq 1/B(2A_n + L) > 1/B(3A_n) = 1/B_n$ for n large enough (since $A_n \to \infty$ and B is increasing). Recalling that $|\log X| < |1/X - 1| + |X - 1|$, we thus have

$$\begin{aligned}\sup_{\theta} |\Delta^1_n(\theta)| &\leq \sup_{\theta,\,i,\,|x|<2A_n} \left| \hat{f}^{\theta}_{n,\,i}(x) - f^s_e(x) \right| \left[\frac{1}{\hat{f}^{\theta}_{n,\,i}(x)} + \frac{1}{f^s_e(x)} \right] \\ &\leq \varepsilon'_n \left(\frac{1}{1/B_n - \varepsilon'_n} + B_n \right) = B_n \varepsilon'_n \left(1 + \frac{1}{1 - B_n \varepsilon'_n} \right) \\ &= B_n \varepsilon'_n \left(2 + \frac{B_n \varepsilon'_n}{1 - B_n \varepsilon'_n} \right)\end{aligned}$$

and finally, for large n, assuming that $B_n \varepsilon'_n \to 0$, $\sup_\theta |\Delta^1_n(\theta)| \leq 3 B_n \varepsilon'_n$. It now only remains to set ε'_n so that $B_n \varepsilon'_n \to 0$ as $n \to \infty$. Let $B_n = n^{\alpha}$, for some fixed positive α. Then for $\delta_n \to 0$ as $n \to \infty$, $B_n \varepsilon_n \to 0$ for any $\varepsilon_n = n^{-\alpha} \delta_n$. Fix now $h_n = n^{-\alpha} \delta^2_n$, so that $h_n = o(\varepsilon_n)$ as required by Lemma 2. Then $B_n (n h_n)^{-1} = (n^{1-2\alpha} \delta^2_n)^{-1} \to 0$ for $\alpha < 1/2$. It is thus sufficient to take $\alpha < 1/2$ in order to obtain $B_n \varepsilon'_n = B_n \varepsilon_n + B_n K(0)/(n h_n) \to 0$. Finally, α must also satisfy $c/(n h_n \varepsilon^2_n) \to 0$ in (27), i.e. $n^{1-3\alpha} \delta^4_n \to \infty$, which is the case when $\alpha < 1/3$. Therefore, for $\alpha < 1/3$, $\sup_\theta |\Delta^1_n(\theta)| \xrightarrow{p} 0$, $n \to \infty$.

b) Uniform convergence of $\Delta^2_n(\theta)$ to 0.
Let us consider the expectation of the supremum of $|\Delta^2_n(\theta)|$.

$$\begin{aligned}\mathtt{E}\left[\sup_{\theta} |\Delta^2_n(\theta)| \right] &\leq \mathtt{E}\left[\sup_{\theta} \frac{1}{n} \sum_{i=1}^{n} |\log f^s_e(e_i(\theta))| \, |U_n(e_i(\theta)) - 1| \right] \\ &\leq \mathtt{E}\left[\frac{1}{n} \sum_{i=1}^{n} \sup_{\theta} |\log f^s_e(e_i(\theta))| \, |U_n(e_i(\theta)) - 1| \right] \\ &= R_n = \int \sup_{\theta} |\log f^s_e(x)| \, |U_n(x) - 1| \, f(x - \bar{\theta} + \theta) dx \,.\end{aligned}$$

Since $U_n(x) \to 1$, $n \to \infty$, in order to prove that $R_n \to 0$, we simply show that $\int \sup_\theta |\log f^s_e(x)| f(x - \bar{\theta} + \theta) dx$ is finite. Writing $z = \bar{\theta} - \theta$, we have $|z| < L$, and from **F2**, for $x > C + L$,

$$h(x) = \sup_{|z|<L} \left| \log \frac{f(x-z) + f(x+z)}{2} \right| f(x - z) < |\log f(x + L)| f(x - L) \,.$$

F1 then implies

$$\int_{|x|>C+L} h(x)dx < \infty\,.$$

For $|x| < C + L$, we have

$$h(x) < \sup_{|x|<C+2L} \{f(x)\} \left| \log \left(\min_{|x|<C+2L} f(x) \right) \right| .$$

$h(x)$ thus has finite expectation since f is bounded and strictly positive from **F0**. Therefore, $\mathtt{E}\left[\sup_\theta |\Delta_n^2(\theta)|\right]$ tends to 0 when $n \to \infty$, and Chebyshev's inequality implies $\sup_\theta \Delta_n^2(\theta) \xrightarrow{p} 0$, that is, $\Delta_n^2(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$.

c) Uniform convergence of $\Delta_n^3(\theta)$ to 0.
For a fixed θ, for $n \to \infty$, from **F1** the law of large numbers implies the following pointwise consistency result

$$\frac{1}{n}\sum_{i=1}^{n} \log f_e^s(-e_i(\theta)) \to \int \log f_e^s(x) f(x - \bar{\theta} + \theta)dx = \int \log\left(f_e^s(x)\right) f_e^s(x)dx\,.$$

Therefore, for any θ, $\Delta_n^3(\theta) \xrightarrow{p} 0$. We now prove the uniformity in θ of the convergence using Lemma 3. For any couple $(\theta, \theta') \in \Theta^2$,

$$\log f_e^s(e_i(\theta)) - \log f_e^s(e_i(\theta')) = \frac{\partial}{\partial\theta}\left\{\log f_e^s(e_i(\theta))\right\}\Big|_{\tilde{\theta}} (\theta - \theta')\,,$$

for some $\tilde{\theta} = (1-\alpha)\,\theta + \alpha\,\theta'$, $\alpha \in (0,1)$, which yields

$$\begin{aligned} |\log f_e^s(e_i(\theta)) - \log f_e^s(e_i(\theta'))| &< \max_{\tilde{\theta}\in\Theta} \frac{|f'(Y_i - 2\tilde{\theta} + \bar{\theta})|}{f_e^{s,\tilde{\theta}}(e_i(\tilde{\theta}))} |\theta - \theta'| \\ &= \max_{\tilde{\theta}\in\Theta} \frac{2|f'(Y_i - 2\tilde{\theta} + \bar{\theta})|}{f(Y_i - \bar{\theta}) + f(Y_i + \bar{\theta} - 2\tilde{\theta})} |\theta - \theta'| \end{aligned}$$

and, writing $z = 2(\bar{\theta} - \tilde{\theta})$,

$$|\log f_e^s(e_i(\theta)) - \log f_e^s(e_i(\theta'))| < \max_{|z|<2L} \frac{2|f'(\varepsilon_i + z)|}{f(\varepsilon_i) + f(\varepsilon_i + z)} |\theta - \theta'|\,.$$

In order to apply Lemma 3, we need to show that the function

$$b(\varepsilon_i) = \max_{|z|<2L} \frac{2|f'(\varepsilon_i + z)|}{f(\varepsilon_i) + f(\varepsilon_i + z)}$$

has finite expectation. From **F2**, for $x > C + 2L$,

$$b(x) < \frac{2|f'(x - 2L)|}{f(x)}\,.$$

Assumption **F4** implies $\int |f'(x)|dx < \infty$, so that $\int_{|x|>C+2L} b(x)f(x)dx < \infty$. On the other hand, $\int_{|x|<C+2L} b(x)f(x)dx < \infty$ is finite since $|f'(x)|$ is bounded from **F0**. We thus obtain $\Delta_n^3(\theta) \overset{\theta,\text{p}}{\rightsquigarrow} 0$.

■

Proof of Theorem 2. The proof uses similar techniques to the ones used for the proof of Theorem 1.

The first order derivative of $\hat{H}_n(\theta)$ is given by

$$\nabla \hat{H}_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\left\{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right\}\frac{Un(e_i(\theta))}{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))} - \log \hat{f}^{\theta}_{n,\,i}(e_i(\theta))\frac{\partial}{\partial\theta}\left\{U_n(e_i(\theta))\right\}\right].$$

We can break down the expression of the second order derivative of the criterion $\hat{H}_n(\theta)$ into four separate terms $\nabla^2 \hat{H}_n(\theta) = \nabla_n^{2\,(1)}(\theta) + \nabla_n^{2\,(2)}(\theta) + \nabla_n^{2\,(3)}(\theta) + \nabla_n^{2\,(4)}(\theta)$, with :

$$\begin{aligned}
\nabla_n^{2\,(1)}(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right\}\frac{U_n(e_i(\theta))}{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))}\,,\\
\nabla_n^{2\,(2)}(\theta) &= +\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\left\{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right\}\right]^2\frac{U_n(e_i(\theta))}{\left(\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right)^2}\,,\\
\nabla_n^{2\,(3)}(\theta) &= -\frac{2}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left\{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right\}\frac{1}{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))}\frac{\partial}{\partial\theta}\left\{U_n(e_i(\theta))\right\}\,,\\
\nabla_n^{2\,(4)}(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\log \hat{f}^{\theta}_{n,\,i}(e_i(\theta))\frac{\partial^2}{\partial\theta^2}\left\{U_n(e_i(\theta))\right\}\,.
\end{aligned}$$

a) Uniform convergence of $\nabla_n^{2\,(1)}(\theta)$ to $-\int \frac{2f''(x+\bar{\theta}-\theta)}{f_e^s(x)}f(x+\bar{\theta}-\theta)dx$.
We can in turn decompose the first term of the sum into four sub-terms :

$$\nabla_n^{2\,(1)}(\theta) = E_1^n(\theta) + E_2^n(\theta) + E_3^n(\theta) + E_4(\theta),$$

with

$$\begin{aligned}
E_1^n(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right\}}{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))} - \frac{\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}}{f_e^s(e_i(\theta))}\right]U_n(e_i(\theta))\,,\\
E_2^n(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}}{f_e^s(e_i(\theta))}\left(U_n(e_i(\theta))-1\right)\,,\\
E_3^n(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}}{f_e^s(e_i(\theta))} + \int\frac{2f''(x+\bar{\theta}-\theta)}{f_e^s(x)}f(x-\bar{\theta}+\theta)dx\,,\\
E_4(\theta) &= -\int\frac{2f''(x+\bar{\theta}-\theta)}{f_e^s(x)}f(x-\bar{\theta}+\theta)dx\,.
\end{aligned}$$

We show that $E_n^j(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$ for $j = 1, 2, 3$.
From Lemma 2, for $h_n = o(\varepsilon_{n_2})$, there exists a positive constant c_2 such that

$$P\left\{\sup_z |\hat{g}_n''(z) - f''(z - \bar{\theta})| > \varepsilon_{n_2}\right\} \le \frac{c_2}{nh_n^5 \varepsilon_{n_2}^2}.$$

Here again (as in the proof of Theorem 1, and using (17), (20) and (23)), we can also write

$$P\left\{\sup_{i,\,\theta} |\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_n^\theta(e_i(\theta))\right\} - \frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}| > 2\varepsilon_{n_2}\right\} \le \frac{c_2}{nh_n^5 \varepsilon_{n_2}^2},$$

which yields for $\varepsilon'_{n_2} = 2\varepsilon_{n_2} + 4K''(0)/(nh_n^3)$ (see (26))

$$P\left\{\sup_{i,\,\theta} |\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\} - \frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}| > \varepsilon'_{n_2}\right\} \le \frac{c_2}{nh_n^5 \varepsilon_{n_2}^2}.$$

Suppose now that $\sup_{i,\,\theta} |\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\} - \frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}| < \varepsilon'_{n_2}$. Suppose also, as in the proof of Theorem 1, that $\sup_{i,\,\theta} |\hat{f}_{n,\,i}^\theta(e_i(\theta)) - f_e^s(e_i(\theta))| < \varepsilon'_n = \varepsilon_n + 2K(0)/(nh_n)$. For $|x| < 2A_n$, $|x \pm (\theta - \bar{\theta})| < 2A_n + L$, and $f_e^s(x) \ge 1/B(2A_n + L) > 1/B(3A_n) = 1/B_n$ for n large enough (since $A_n \to \infty$ and B is increasing). Let $S = \sup_z f''(z)$. Then, recalling (17),

$$\begin{aligned}
\sup_\theta |E_n^1(\theta)| &\le \sup_{i,\,\theta}\left|\frac{\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\}}{\hat{f}_{n,\,i}^\theta(e_i(\theta))} - \frac{\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}}{f_e^s(e_i(\theta))}\right| U_n(e_i(\theta)) \\
&\le \sup_{i,\,\theta}\left|\frac{\partial^2}{\partial\theta^2}\left\{\hat{f}_{n,\,i}^\theta(e_i(\theta))\right\} - \frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}\right| \frac{1}{\hat{f}_{n,\,i}^\theta(e_i(\theta))} U_n(e_i(\theta)) \\
&\qquad + \sup_{i,\,\theta}\left|\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(e_i(\theta))\right\}\right| \left|\frac{1}{\hat{f}_{n,\,i}^\theta(e_i(\theta))} - \frac{1}{f_e^s(e_i(\theta))}\right| U_n(e_i(\theta)) \\
&\le \sup_{i,\,\theta} \frac{\varepsilon'_{n_2}}{\hat{f}_{n,\,i}^\theta(e_i(\theta))} U_n(e_i(\theta)) \\
&\qquad + \sup_{i,\,\theta} 2\left|f''(e_i(\theta) + \bar{\theta} - \theta)\right| \left|\frac{1}{\hat{f}_{n,\,i}^\theta(e_i(\theta))} - \frac{1}{f_e^s(e_i(\theta))}\right| U_n(e_i(\theta)) \\
&\le \sup_{\theta,\,i,\,|z|<2A_n} \frac{\varepsilon'_{n_2}}{\hat{f}_{n,\,i}^\theta(z)} + \sup_{\theta,\,i,\,|z|<2A_n} 2\left|f''(z + \bar{\theta} - \theta)\right| \left|\frac{1}{\hat{f}_{n,\,i}^\theta(z)} - \frac{1}{f_e^s(z)}\right| \\
&\le \varepsilon'_{n_2} \frac{1}{(\frac{1}{B_n} - \varepsilon'_n)} + \frac{2S\varepsilon'_n B_n}{(\frac{1}{B_n} - \varepsilon'_n)} \\
&= B_n \varepsilon'_{n_2} \frac{1}{1 - B_n\varepsilon'_n} + 2S\varepsilon'_n \frac{B_n^2}{1 - B_n\varepsilon'_n} \\
&= (B_n\varepsilon'_{n_2} + 2S\varepsilon'_n B_n^2)(1 - B_n\varepsilon'_n)^{-1}.
\end{aligned}$$

Considering this last expression, we need to set B_n, ε_{n_2} and ε'_{n_2} so that $B_n\varepsilon'_{n_2} \to 0$, $B_n\varepsilon'_n \to 0$ and $B_n^2\varepsilon'_n \to 0$ as $n \to \infty$. For this, let δ_n satisfy $\delta_n \to 0$ and $n^\gamma\delta_n \to \infty$ for any $\gamma > 0$. For some $\alpha > 0$, we then write $\varepsilon_n = \varepsilon_{n_2} = n^{-\alpha}\delta_n$, so that $\varepsilon_n \to 0$ as $n \to \infty$. Since $h_n = o(\varepsilon_n)$ is necessary for Lemma 2, we fix $h_n = n^{-\alpha}\delta_n^2$.
Then, by fixing $B_n = n^{\alpha/3}$, $B_n\varepsilon'_n$ is of the same order as $n^{-2\alpha/3}\delta_n + (n^{1-4\alpha/3}\delta_n^2)^{-1}$. Taking $\alpha < 3/4$ is sufficient for this term to tend to 0 as $n \to \infty$.
Using the same approach, for large n, $B_n^2\varepsilon'_n$ behaves like $n^{-\alpha/3}\delta_n + (n^{1-5\alpha/3}\delta_n^2)^{-1}$, in which case choosing $\alpha < 3/5$ leads to $(n^{1-5\alpha/3}\delta_n^2)^{-1} \to 0$, i.e. to $B_n^2\varepsilon'_n \to 0$. Also, $B_n\varepsilon'_{n_2}$ behaves like $n^{-2\alpha/3}\delta_n + (n^{1-10\alpha/3}\delta_n^6)^{-1}$. In this case, taking $\alpha < 3/10$ allows convergence to 0 of this term.
On the other hand, α must also satisfy $c/(nh_n\varepsilon_n^2) \to 0$, i.e. $n^{1-3\alpha}\delta_n^4$, and $c_2/(nh_n^5\varepsilon_{n_2}^2) \to 0$, i.e. $n^{1-7\alpha}\delta_n^7 \to \infty$. Taking $\alpha < 1/7$ is sufficient for these conditions to be satisfied. Taking $\alpha < 1/7$ is therefore suitable for all terms to tend to 0, that is, for $E_n^1(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$.

We obtain uniform convergence of $E_2^n(\theta)$ to 0 using the same approach as in point *(b)* of the proof of Theorem 1, showing that $\mathtt{E}[\sup_\theta |E_2^n(\theta)|]$ converges to 0. We have

$$
\begin{aligned}
&\mathtt{E}\left[\sup_\theta |E_2^n(\theta)|\right] \\
&\leq \mathtt{E}\left[\sup_\theta \frac{1}{n}\sum_{i=1}^n \left|\frac{\partial^2}{\partial\theta^2}\{f_e^s(e_i(\theta))\}\right| \frac{1}{f_e^s(e_i(\theta))}\left|U_n(e_i(\theta)) - 1\right|\right] \\
&\leq \mathtt{E}\left[\frac{1}{n}\sum_{i=1}^n \sup_\theta \left|\frac{\partial^2}{\partial\theta^2}\{f_e^s(e_i(\theta))\}\right| \frac{1}{f_e^s(e_i(\theta))}\left|U_n(e_i(\theta)) - 1\right|\right] \\
&= \int 4\sup_\theta \frac{\left|f''(Y_i + \bar\theta - 2\theta)\right|}{f(Y_i - \bar\theta) + f(Y_i + \bar\theta - 2\theta)}\left|U_n(Y_i - \theta) - 1\right| f(Y_i - \bar\theta)dY_i
\end{aligned}
$$

where we used (17). Since $U_n(x) \to 1$, $x \to \infty$, we only need to show that

$$
\mathcal{I} = \int \sup_\theta \frac{\left|f''(Y_i + \bar\theta - 2\theta)\right|}{f(Y_i - \bar\theta) + f(Y_i + \bar\theta - 2\theta)} f(Y_i - \bar\theta)dY_i < \infty.
$$

Let $z = 2(\bar\theta - \theta)$; we have

$$
\mathcal{I} = \int \sup_{z<2L} \frac{|f''(x+z)|}{f(x) + f(x+z)} f(x)dx.
$$

Using the assumptions **F2** and **F2'** that $f(x)$ and $|f''(x)|$ are decreasing for $x > C$, we obtain for $x > C + 2L$

$$
\sup_{|z|<2L} \frac{|f''(x+z)|}{f(x) + f(x+z)} < \frac{|f''(x-2L)|}{f(x)}.
$$

This term has finite expectation since f'' is integrable. The integral for $|x| < C + 2L$ is also finite since $f(x) > 0$ and $|f''|$ is bounded. We can then conclude using Chebyshev's inequality that $E_2^n(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$.

Pointwise convergence of $E_3^n(\theta)$ to 0 directly follows from the weak law of large numbers : $E_3^n(\theta) \xrightarrow{p} 0$ when $n \to \infty$. To prove uniform convergence using Lemma 3, we consider for any $\theta, \theta' \in \Theta$ the difference

$$\begin{aligned}
\delta_i &= \frac{\frac{\partial^2}{\partial\theta^2}\{f_e^s(e_i(\theta))\}}{f_e^s(e_i(\theta))} - \frac{\frac{\partial^2}{\partial\theta^2}\{f_e^s(e_i(\theta'))\}}{f_e^s(e_i(\theta'))} \\
&= \frac{4f''(Y_i - 2\theta + \bar{\theta})}{f(Y_i - \bar{\theta}) + f(Y_i - 2\theta + \bar{\theta})} - \frac{4f''(Y_i - 2\theta' + \bar{\theta})}{f(Y_i - \bar{\theta}) + f(Y_i - 2\theta' + \bar{\theta})} \\
&= g_i(\theta) - g_i(\theta') \\
&= g_i'(\tilde{\theta})(\theta - \theta'),
\end{aligned}$$

for some $\tilde{\theta} = (1-\alpha)\,\theta + \alpha\,\theta'$, $\alpha \in (0,1)$. We have

$$\begin{aligned}
|\delta_i| &\leq \max_{\tilde{\theta}} |g_i{}'(\tilde{\theta})||\theta - \theta'| \\
&= \max_{\tilde{\theta}} \left| \frac{-8f'''(Y_i - 2\tilde{\theta} + \bar{\theta})}{f(Y_i - \bar{\theta}) + f(Y_i - 2\tilde{\theta} + \bar{\theta})} + \frac{8f''(Y_i - 2\tilde{\theta} + \bar{\theta})f'(Y_i - 2\tilde{\theta} + \bar{\theta})}{[f(Y_i - \bar{\theta}) + f(Y_i - 2\tilde{\theta} + \bar{\theta})]^2} \right| |\theta - \theta'| \\
&= \Delta_i |\theta - \theta'|.
\end{aligned}$$

Now, rewriting $z = 2(\bar{\theta} - \tilde{\theta})$,

$$\Delta_i \;<\; 8 \max_{|z|<2L} \frac{|f'''(\varepsilon_i + z)|}{f(\varepsilon_i) + f(\varepsilon_i + z)} + 8 \max_{|z|<2L} \frac{|f''(\varepsilon_i + z)||f'(\varepsilon_i + z)|}{[f(\varepsilon_i) + f(\varepsilon_i + z)]^2}.$$

Let

$$b_1(x) = \max_{|z|<2L} \frac{|f'''(x+z)|}{f(x) + f(x+z)}, \qquad b_2(x) = \max_{|z|<2L} \frac{|f''(x+z)f'(x+z)|}{[f(x) + f(x+z)]^2}.$$

To apply Lemma 3, we thus need to show that these two functions have finite expectation over x, i.e. that

$$\int b_1(x)f(x)dx < \infty, \qquad \int b_2(x)f(x)dx < \infty.$$

Consider first the function b_1. Since $|f'''(x)|$ is decreasing for $x > C$ from **F2'**, we obtain for $x > C + 2L$

$$b_1(x) < \frac{|f'''(x - 2L)|}{f(x)}.$$

and $\int_{|x|>C+2L} b_1(x)f(x)dx < \infty$ from **F6** and **F2'**. The integral of $b_1(x)f(x)$ for $|x| < C + 2L$ is also finite since $f(x) > 0$ and $|f'''|$ is bounded. Consider now the second function b_2. Since $|f'(x)|$ and $|f''(x)|$ are decreasing functions for $x > C$ from **F2** and **F2'**, we have for $x > C + 2L$

$$b_2(x) < \frac{|f''(x-2L)f'(x-2L)|}{(f(x))^2}\,.$$

The Cauchy-Schwartz inequality gives

$$\int \frac{|f''(x-2L)f'(x-2L)|}{(f(x))^2} f(x)dx$$

$$< \left[\int \left(\frac{f''(x-2L)}{f(x)} \right)^2 f(x)dx \int \left(\frac{f'(x-2L)}{f(x)} \right)^2 f(x)dx \right]^{1/2}$$

which is finite from **F5**. The integral of $b_2(x)f(x)$ for $|x| < C + 2L$ is also finite since $f(x) > 0$ and $|f'|$ and $|f''|$ are bounded. Lemma 3 then implies

$$\nabla_n^{2(1)} \overset{\theta,\mathrm{p}}{\rightsquigarrow}_{n\to\infty} -2 \int \frac{f''(x+\bar\theta-\theta)}{f_e^s(x)} f(x-\bar\theta+\theta)dx\,.$$

b) Uniform convergence of $\nabla_n^{2\,(2)}(\theta)$ *to* $\int \left(\frac{f'(x+\bar\theta-\theta)}{f_e^s(x)} \right)^2 f(x-\bar\theta+\theta)dx$.

The second term $\nabla_n^{2(2)}(\theta)$ of the decomposition of $\nabla^2 \hat H_n(\theta)$ can be treated using the same approach as for the first term. We can decompose $\nabla_n^{2(2)}(\theta)$ into $F_1^n(\theta) + F_2^n(\theta) + F_3^n(\theta) + F_4(\theta)$, with

$$\begin{aligned}
F_1^n(\theta) &= \frac{1}{n}\sum_{i=1}^n \left[\left(\frac{\frac{\partial}{\partial\theta}\left\{ \hat f_{n,i}^\theta(e_i(\theta)) \right\}}{\hat f_{n,i}^\theta(e_i(\theta))} \right)^2 - \left(\frac{\frac{\partial}{\partial\theta}\left\{ f_e^s(e_i(\theta)) \right\}}{f_e^s(e_i(\theta))} \right)^2 \right] U_n(e_i(\theta))\,,\\
F_2^n(\theta) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{\frac{\partial}{\partial\theta}\left\{ f_e^s(e_i(\theta)) \right\}}{f_e^s(e_i(\theta))} \right)^2 (U_n(e_i(\theta)) - 1)\,,\\
F_3^n(\theta) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{\frac{\partial}{\partial\theta}\left\{ f_e^s(e_i(\theta)) \right\}}{f_e^s(e_i(\theta))} \right)^2 - \int \left(\frac{f'(x+\bar\theta-\theta)}{f_e^s(x)} \right)^2 f(x-\bar\theta+\theta)dx\,,\\
F_4(\theta) &= \int \left(\frac{f'(x+\bar\theta-\theta)}{f_e^s(x)} \right)^2 f(x-\bar\theta+\theta)dx\,.
\end{aligned}$$

First, it can be shown for $n \to \infty$ that $F_1^n(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$, similarly to what was done for the term $E_1^n(\theta)$. From Lemma 2, for $h_n = o(\varepsilon_{n_1})$, there exists a positive constant c_1 such that

$$P\left\{ \sup_z |\hat g_n'(z) - f'(z-\bar\theta)| > \varepsilon_{n_1} \right\} \le \frac{c_1}{n h_n^3 \varepsilon_{n_1}^2}.$$

Using (16), (18) and (22), we can also write

$$P\left\{\sup_{i,\,\theta}|\frac{\partial}{\partial\theta}\left\{\hat{f}_n^{\theta}(e_i(\theta))\right\}-\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}|>\varepsilon_{n_1}\right\}\leq\frac{c_1}{nh_n^3\varepsilon_{n_1}^2},$$

which yields for $\varepsilon'_{n_1}=\varepsilon_{n_1}+2K'(0)/(nh_n^2)$ (see (25))

$$P\left\{\sup_{i,\,\theta}|\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}-\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}|>\varepsilon'_{n_1}\right\}\leq\frac{c_1}{nh_n^3\varepsilon_{n_1}^2}.$$

Suppose now that

$$\sup_{i,\,\theta}|\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}-\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}|<\varepsilon'_{n_1}. \tag{28}$$

From here, supposing, as in the proof of Theorem 1, that $\sup_{i,\,\theta}|\hat{f}_{n,\,i}^{\theta}(e_i(\theta))-f_e^s(e_i(\theta))|<\varepsilon'_n=\varepsilon_n+2K(0)/(nh_n)$, the calculations are similar to those of point (a) for the term $E_n^1(\theta)$. We have

$$\begin{aligned}
&\sup_{\theta}\left|F_n^1(\theta)\right|\\
\leq\;&\sup_{i,\,\theta}\left|\left(\frac{\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}}{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))}\right)^2-\left(\frac{\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}}{f_e^s(e_i(\theta))}\right)^2\right|U_n(e_i(\theta))\\
\leq\;&\sup_{i,\,\theta}\left|\left(\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}\right)^2-\left(\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}\right)^2\right|\frac{U_n(e_i(\theta))}{\left|\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right|^2}\\
&+\sup_{i,\,\theta}\left|\frac{1}{\left(\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right)^2}-\frac{1}{\left(f_e^s(e_i(\theta))\right)^2}\right|\left|\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}\right|^2U_n(e_i(\theta))\\
=\;&\sup_{i,\,\theta}\left|\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}-\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}\right|\\
&\times\left|\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}+\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}\right|\frac{U_n(e_i(\theta))}{\left|\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right|^2}\\
&+\sup_{i,\,\theta}\left|\frac{1}{\left(\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right)^2}-\frac{1}{\left(f_e^s(e_i(\theta))\right)^2}\right|\left|f'(e_i(\theta)+\bar{\theta}-\theta)\right|^2U_n(e_i(\theta))\\
\leq\;&\sup_{i,\,\theta}\varepsilon'_{n_1}\left(\left|\frac{\partial}{\partial\theta}\left\{\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right\}\right|+\left|\frac{\partial}{\partial\theta}\left\{f_e^s(e_i(\theta))\right\}\right|\right)\frac{U_n(e_i(\theta))}{\left|\hat{f}_{n,\,i}^{\theta}(e_i(\theta))\right|^2}
\end{aligned}$$

$$
\begin{aligned}
&+\sup_{i,\,\theta}\left|\frac{1}{\left(\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right)^2}-\frac{1}{\left(f^s_e(e_i(\theta))\right)^2}\right|\left|f'(e_i(\theta)+\bar{\theta}-\theta)\right|^2 U_n(e_i(\theta))\\
\leq\ &\sup_{i,\,\theta}\varepsilon'_{n_1}\left(\varepsilon'_{n_1}+\left|f'(e_i(\theta)+\bar{\theta}-\theta)\right|+\left|f'(e_i(\theta)+\bar{\theta}-\theta)\right|\right)\frac{U_n(e_i(\theta))}{\left|\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right|^2}\\
&+\sup_{i,\,\theta}\left|\frac{1}{\left(\hat{f}^{\theta}_{n,\,i}(e_i(\theta))\right)^2}-\frac{1}{\left(f^s_e(e_i(\theta))\right)^2}\right|\left|f'(e_i(\theta)+\bar{\theta}-\theta)\right|^2 U_n(e_i(\theta))\\
\leq\ &\sup_{i,\,\theta,\,|z|<2A_n}\varepsilon'_{n_1}\left(\varepsilon'_{n_1}+2\left|f'(z+\bar{\theta}-\theta)\right|\right)\frac{1}{\left|\hat{f}^{\theta}_{n,\,i}(z)\right|^2}\\
&+\sup_{i,\,\theta,\,|z|<2A_n}\left|\frac{1}{\left(\hat{f}^{\theta}_{n,\,i}(z)\right)^2}-\frac{1}{\left(f^s_e(z)\right)^2}\right|\left|f'(z+\bar{\theta}-\theta)\right|^2 .
\end{aligned}
$$

Notice that for $|x|<2A_n$

$$
\begin{aligned}
\left|\frac{1}{(\hat{f}^{\theta}_{n,\,i}(x))^2}-\frac{1}{(f^s_e(x))^2}\right| &= \left|\frac{1}{\hat{f}^{\theta}_{n,\,i}(x)}-\frac{1}{f^s_e(x)}\right|\left|\frac{1}{\hat{f}^{\theta}_{n,\,i}(x)}+\frac{1}{f^s_e(x)}\right|\\
&= \frac{\left|f^s_e(x)-\hat{f}^{\theta}_{n,\,i}(x)\right|}{\hat{f}^{\theta}_{n,\,i}(x)f^s_e(x)}\left|\frac{1}{\hat{f}^{\theta}_{n,\,i}(x)}+\frac{1}{f^s_e(x)}\right|\\
&< \frac{\varepsilon'_n B_n^3}{1-\varepsilon'_n B_n}\left(1+\frac{1}{1-\varepsilon'_n B_n}\right).
\end{aligned}
$$

With $F=\sup_z|f'(z)|$, $|\frac{\partial}{\partial\theta}\{f^s_e(e_i(\theta))\}|\leq F$ and

$$
\sup_{\theta}|F^n_1(\theta)|\leq\frac{3F\varepsilon'_{n_1}B_n^2}{(1-\varepsilon'_n B_n)^2}+\frac{F^2\varepsilon'_n B_n^3}{1-\varepsilon'_n B_n}\left(1+\frac{1}{1-\varepsilon'_n B_n}\right).
$$

We thus need to show that $\varepsilon'_n B_n\to 0$, $\varepsilon'_n B_n^3\to 0$ and $\varepsilon'_{n_1}B_n^2\to 0$ when $n\to\infty$. As previously, we define δ_n such that $\delta_n\to 0$, $n^{\gamma}\delta_n\to\infty$ as $n\to\infty$. Fixing $\alpha>0$, we write $\varepsilon_{n_1}=\varepsilon_n=n^{-\alpha}\delta_n$ and $h_n=n^{-\alpha}\delta_n^2$ so that $h_n/\varepsilon_n\to 0$ as necessary for Lemma 2. Now, let $B_n=n^{\alpha/3}$.
$B_n^3\varepsilon'_n$ behaves for large n like $\delta_n+(n^{1-2\alpha}\delta_n^2)^{-1}$. Choosing $\alpha<1/2$ then ensures that $B_n^3\varepsilon'_n\to 0$.
The second term $B_n^2\varepsilon'_{n_1}$ behaves like $n^{-\alpha/3}\delta_n+(n^{1-8\alpha/3}\delta_n^4)^{-1}$. We then need to take $\alpha<3/8$ to obtain $B_n^2\varepsilon'_{n_1}\to 0$.
Also, $B_n\varepsilon'_n$ tends to 0 for $\alpha<3/4$ as before.
On the other hand, α must also satisfy $c/(nh_n\varepsilon_n^2)\to 0$, i.e. $n^{1-3\alpha}\delta_n^4\to\infty$, and $c_1/(nh_n^3\varepsilon_{n_1}^2)\to 0$, i.e. $n^{1-5\alpha}\delta_n^8\to\infty$. Taking $\alpha<1/5$ is sufficient for these conditions to be satisfied.
Taking $\alpha<1/5$ therefore gives $F^1_n(\theta)\overset{\theta,\mathrm{p}}{\rightsquigarrow}0$. Notice that $\alpha<1/7$ from point (a) is therefore admissible here.

The proof for uniform convergence of $F_n^2(\theta)$ to 0 is similar to that of $E_n^2(\theta)$. Here,

$$
\begin{aligned}
&\mathtt{E}\left[\sup_\theta |F_2^n(\theta)|\right] \\
&\leq \mathtt{E}\left[\sup_\theta \frac{1}{n}\sum_{i=1}^n \left|\left(\frac{\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}}{f_e^s(e_i(\theta))}\right)^2 (U_n(e_i(\theta))-1)\right|\right] \\
&\leq \mathtt{E}\left[\frac{1}{n}\sum_{i=1}^n \sup_\theta \left|\left(\frac{\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}}{f_e^s(e_i(\theta))}\right)^2 (U_n(e_i(\theta))-1)\right|\right] \\
&= \int \sup_\theta 4\left(\frac{f'(Y_i+\bar\theta-2\theta)}{f(Y_i-\bar\theta)+f(Y_i+\bar\theta-2\theta)}\right)^2 |U_n(Y_i-\theta)-1|\, f(Y_i-\bar\theta)dY_i
\end{aligned}
$$

where we used (16). Since $U_n(x) \to 1$, $x \to \infty$, we only need to show that

$$\mathcal{I} = \int \sup_\theta \left(\frac{f'(Y_i+\bar\theta-2\theta)}{f(Y_i-\bar\theta)+f(Y_i+\bar\theta-2\theta)}\right)^2 f(Y_i-\bar\theta)dY_i < \infty.$$

Writing $z = 2(\bar\theta-\theta)$, we have

$$\mathcal{I} = \int \sup_{|z|<2L} \left(\frac{f'(x+z)}{f(x)+f(x+z)}\right)^2 f(x)dx$$

and for $x > C+2L$ we have from **F2** that

$$\sup_{|z|<2L} \left(\frac{f'(x+z)}{f(x)+f(x+z)}\right)^2 < \left(\frac{f'(x-2L)}{f(x)}\right)^2$$

which has finite expectation from **F5**. The integral for $|x| < C+2L$ is also finite since $f(x) > 0$ and $|f'|$ is bounded. We can then conclude using Chebyshev's inequality that $F_n^2(\theta) \overset{\theta;\mathrm{p}}{\rightsquigarrow} 0$.

The proof for uniform convergence of $F_n^3(\theta)$ to 0 is similar to that of $E_n^3(\theta)$. Here again, the pointwise convergence of $F_n^3(\theta)$ to 0 directly follows from the weak law of large numbers. In order to prove uniform convergence using Lemma 3, we consider the following difference δ_i

$$
\begin{aligned}
\delta_i &= \left[\frac{\partial^2}{\partial\theta^2}\{f_e^s(e_i(\theta))\}\right]^2 \frac{1}{(f_e^s(e_i(\theta)))^2} - \left[\frac{\partial^2}{\partial\theta^2}\{f_e^s(e_i(\theta'))\}\right]^2 \frac{1}{(f_e^s(e_i(\theta')))^2} \\
&= \frac{4[f'(Y_i-2\theta+\bar\theta)]^2}{[f(Y_i-\bar\theta)+f(Y_i+\bar\theta-2\theta)]^2} - \frac{4[f'(Y_i-2\theta'+\bar\theta)]^2}{[f(Y_i-\bar\theta)+f(Y_i+\bar\theta-2\theta')]^2} \\
&= g_i(\theta) - g_i(\theta') \\
&= g_i'(\tilde\theta)(\theta-\theta'),
\end{aligned}
$$

for some $\tilde{\theta} = (1-\alpha)\,\theta + \alpha\,\theta'$, $\alpha \in (0,1)$. We have :

$$\begin{aligned}
|\delta_i| &\leq \max_{\tilde{\theta}} |g_i'(\tilde{\theta})|\,|\theta-\theta'| \\
&= \max_{\tilde{\theta}} \left| \frac{-16 f'(Y_i - 2\tilde{\theta} + \bar{\theta}) f''(Y_i - 2\tilde{\theta} + \bar{\theta})}{[f(Y_i-\bar{\theta}) + f(Y_i - 2\tilde{\theta} + \bar{\theta})]^2} \right. \\
&\quad \left. + \frac{16\left[f'(Y_i - 2\tilde{\theta} + \bar{\theta})\right]^3}{[f(Y_i-\bar{\theta}) + f(Y_i - 2\tilde{\theta} + \bar{\theta})]^3} \right| |\theta-\theta'| \\
&= \Delta_i\,|\theta-\theta'|.
\end{aligned}$$

Rewriting $z = 2(\bar{\theta} - \tilde{\theta})$, we have

$$\Delta_i \quad < \quad 16 \max_{|z|<2L} \frac{|f'(\varepsilon_i+z)||f''(\varepsilon_i+z)|}{[f(\varepsilon_i)+f(\varepsilon_i+z)]^2} + 16 \max_{|z|<2L} \frac{|f'(\varepsilon_i+z)|^3}{[f(\varepsilon_i)+f(\varepsilon_i+z)]^3}.$$

Let

$$b_3(x) = \max_{|z|<2L} \frac{|f'(x+z)||f''(x+z)|}{[f(x)+f(x+z)]^2}, \qquad b_4(x) = \max_{|z|<2L} \frac{|f'(x+z)|^3}{[f(x)+f(x+z)]^3}.$$

The function b_3 is identical to the function b_2 in point (a), which has finite expectation under **F2**, **F2'** and **F5**. From **F2**, for $x > C + 2L$, we have

$$b_4(x) < \frac{|f'(x-2L)|^3}{(f(x))^3}.$$

The Cauchy-Schwartz inequality then gives

$$\int b_4(x) f(x) dx \quad < \quad \left[\int \left(\frac{f'(x-2L)}{f(x)} \right)^2 f(x) dx \int \left(\frac{(f'(x-2L))^2}{(f(x))^2} \right)^2 f(x) dx \right]^2$$

which is also finite from **F5**. The integral of $b_4(x)f(x)$ for $|x| < C + 2L$ is also finite since $f(x) > 0$ and $|f'|$ is bounded. The difference δ_i is thus bounded by a function that has finite expectation. Lemma 3 then implies $F_3^n(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$.
We can therefore conclude that

$$\nabla_n^{2(2)}(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow}_{n\to\infty} \int \left(\frac{f'(x+\bar{\theta}-\theta)}{f_e^s(x)} \right)^2 f(x-\bar{\theta}+\theta) dx.$$

c) Uniform convergence of $\nabla_n^{2\,(3)}(\theta)$ to 0.
We can break down this term into

$$\begin{aligned}
\nabla_n^{2\,(3)}(\theta) = \quad & - \quad \frac{2}{n} \sum_{i=1}^n \left[\frac{\frac{\partial}{\partial\theta}\left\{ \hat{f}^{\theta}_{n,\,i}(e_i(\theta)) \right\}}{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))} - \frac{\frac{\partial}{\partial\theta}\left\{ f_e^s(e_i(\theta)) \right\}}{f_e^s(e_i(\theta))} \right] \frac{\partial}{\partial\theta} \left\{ U_n(e_i(\theta)) \right\} \\
& - \quad \frac{2}{n} \sum_{i=1}^n \frac{\frac{\partial}{\partial\theta}\left\{ f_e^s(e_i(\theta)) \right\}}{\hat{f}^{\theta}_{n,\,i}(e_i(\theta))} \frac{\partial}{\partial\theta} \left\{ U_n(e_i(\theta)) \right\}.
\end{aligned}$$

By construction, for $A'(x) < d_1$ as assumed with (12), $\left|\frac{\partial}{\partial\theta}\{U_n(e_i(\theta))\}\right| < d_1/A_n$. From this,

$$\sup_\theta \left|\nabla_n^{2\,(3)}(\theta)\right| \leq \frac{2d_1}{A_n}\sup_{i,\,\theta}\left|\frac{\frac{\partial}{\partial\theta}\left\{\hat{f}^\theta_{n,\,i}(e_i(\theta))\right\}}{\hat{f}^\theta_{n,\,i}(e_i(\theta))} - \frac{\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}}{f_e^s(e_i(\theta))}\right| \times I_{|e_i(\theta)|<2A_n}$$
$$+\frac{2d_1}{A_n}\frac{1}{n}\sum_{i=1}^n \sup_\theta\left|\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}\right|\frac{1}{f_e^s(e_i(\theta))} \times I_{|e_i(\theta)|<2A_n} \quad (29)$$

From the weak law of large numbers, for $n\to\infty$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \sup_\theta\left|\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}\right|\frac{1}{f_e^s(e_i(\theta))} &\overset{p}{\to} \int \sup_\theta \frac{|f'(x+\bar\theta-\theta)|}{f_e^s(x)} f(x-\bar\theta+\theta)dx\\ &< \int \sup_\theta \frac{|f'(x+\bar\theta-\theta)|}{f(x-\bar\theta+\theta)} f(x-\bar\theta+\theta)dx\\ &= \int \sup_\theta |f'(x+\bar\theta-\theta)|dx\end{aligned}$$

which is finite from **F2**. The second term of the right-hand-side of (29) therefore converges to 0 with A_n^{-1} when $n\to\infty$. Assume as in the proof of Theorem 1 that $\sup_{i,\,\theta}|\hat{f}^\theta_{n,\,i}(e_i(\theta)) - f_e^s(e_i(\theta))| < \varepsilon'_n = \varepsilon_n + 2K(0)/(nh_n)$, and suppose that (28) holds. Using again $F = \sup_z |f'(z)|$, $|\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}| \leq F$ from (16), and

$$\left|\frac{1}{\hat{f}^\theta_{n,\,i}(x)} - \frac{1}{f_e^s(x)}\right| = \left|\frac{\hat{f}^\theta_{n,\,i}(x) - f_e^s(x)}{\hat{f}^\theta_{n,\,i}(x) f_e^s(x)}\right|,$$

and the first term of (29) satisfies

$$\begin{aligned}\frac{2d_1}{A_n}\ \sup\nolimits_{i,\,\theta}\ &\left|\frac{\frac{\partial}{\partial\theta}\left\{\hat{f}^\theta_{n,\,i}(e_i(\theta))\right\}}{\hat{f}^\theta_{n,\,i}(e_i(\theta))} - \frac{\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}}{f_e^s(e_i(\theta))}\right| \times I_{|e_i(\theta)|<2A_n}\\
\leq\ &\frac{2d_1}{A_n}\left[\sup_{i,\,\theta}\left|\frac{\partial}{\partial\theta}\left\{\hat{f}^\theta_{n,\,i}(e_i(\theta))\right\} - \frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}\right|\frac{1}{\hat{f}^\theta_{n,\,i}(e_i(\theta))}\right.\\
&\left.+\sup_{i,\,\theta}\left|\frac{1}{\hat{f}^\theta_{n,\,i}(e_i(\theta))} - \frac{1}{f_e^s(e_i(\theta))}\right|\left|\frac{\partial}{\partial\theta}\{f_e^s(e_i(\theta))\}\right|\right] \times I_{|e_i(\theta)|<2A_n}\\
\leq\ &\frac{2d_1}{A_n}\left[\sup_{i,\,\theta}\varepsilon'_{n_1}\frac{1}{\hat{f}^\theta_{n,\,i}(e_i(\theta))} \times I_{|e_i(\theta)|<2A_n}\right.\\
&\left.+\sup_{i,\,\theta}\left|\frac{1}{\hat{f}^\theta_{n,\,i}(e_i(\theta))} - \frac{1}{f_e^s(e_i(\theta))}\right|\left|f'(e_i(\theta)+\bar\theta-\theta)\right| \times I_{|e_i(\theta)|<2A_n}\right]\\
\leq\ &\frac{2d_1}{A_n}\left[\sup_{i,\,\theta,\,|x|<2A_n}\varepsilon'_{n_1}\frac{1}{\hat{f}^\theta_{n,\,i}(x)}\right.\end{aligned}$$

$$
\begin{aligned}
& + \sup_{i,\,\theta,\,|x|<2A_n} \left| \frac{1}{\hat{f}^{\theta}_{n,\,i}(x)} - \frac{1}{f^s_e(x)} \right| \left| f'(x+\bar{\theta}-\theta) \right| \Bigg] \\
\leq \quad & \frac{2d_1}{A_n} \left[\frac{\varepsilon'_{n_1} B_n}{1 - B_n \varepsilon'_n} + \frac{\varepsilon'_n B_n^2}{1 - B_n \varepsilon'_n} F \right]
\end{aligned}
$$

which converges to 0 in probability with A_n^{-1} when $n \to \infty$, provided that $\varepsilon'_n B_n \to 0$ and $\varepsilon'_n B_n^2 \to 0$ as in points (a) and (b). We can therefore conclude that $\nabla_n^{2\,(3)}(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$ for the choices of B_n, ε_n, h_n made before.

d) Uniform convergence of $\nabla_n^{2\,(4)}(\theta)$ to 0.
We decompose this term into

$$
\begin{aligned}
\nabla_n^{2\,(4)}(\theta) \quad = \quad & -\frac{1}{n} \sum_{i=1}^{n} \left[\log \hat{f}^{\theta}_{n,\,i}(e_i(\theta)) - \log f^s_e(e_i(\theta)) \right] \frac{\partial^2}{\partial \theta^2} \left\{ U_n(e_i(\theta)) \right\} \\
& -\frac{1}{n} \sum_{i=1}^{n} \log f^s_e(e_i(\theta)) \frac{\partial^2}{\partial \theta^2} \left\{ U_n(e_i(\theta)) \right\} . \qquad (30)
\end{aligned}
$$

For $n \to \infty$, it follows directly from the weak law of large numbers that

$$
\frac{1}{n} \sum_{i=1}^{n} \sup_{\theta} |\log f^s_e(e_i(\theta))| \xrightarrow{p} \int \sup_{\theta} |\log f^s_e(x)| f(x+\bar{\theta}-\theta) dx \,,
$$

which is finite, as shown in point (b) of the proof of Theorem 1. By construction, with $A''(x) < d_2$ as assumed with (12), $|\frac{\partial^2}{\partial \theta^2} \left\{ U_n(e_i(\theta)) \right\}| < d_2/A_n^2$. Therefore, considering the second term of the right-hand-side of (30),

$$
\sup_{\theta} \left| \frac{1}{n} \sum_{i=1}^{n} \log f^s_e(e_i(\theta)) \frac{\partial^2}{\partial \theta^2} \left\{ U_n(e_i(\theta)) \right\} \right| < \frac{d_2}{A_n^2} \left| \frac{1}{n} \sum_{i=1}^{n} \sup_{\theta} |\log f^s_e(e_i(\theta))| \right| \xrightarrow{p} 0
$$

with A_n^{-2} when $n \to \infty$. Consider now the first term of the right-hand-side of (30). We have

$$
\sup_{\theta} \left| \frac{1}{n} \sum_{i=1}^{n} \left[\log \hat{f}^{\theta}_{n,\,i}(e_i(\theta)) - \log f^s_e(e_i(\theta)) \right] \frac{\partial^2}{\partial \theta^2} \left\{ U_n(e_i(\theta)) \right\} \right|
$$

$$
\begin{aligned}
\leq \quad & \frac{d_2}{A_n^2} \sup_{i,\,\theta,\,|x|<2A_n} \left| \log \hat{f}^{\theta}_{n,\,i}(x) - \log f^s_e(x) \right| \\
\leq \quad & \frac{d_2}{A_n^2} \sup_{i,\,\theta,\,|x|<2A_n} \left| \hat{f}^{\theta}_{n,\,i}(x) - f^s_e(x) \right| \left| \frac{1}{\hat{f}^{\theta}_{n,\,i}(x)} + \frac{1}{f^s_e(x)} \right| \\
\leq \quad & \frac{d_2}{A_n^2} {\varepsilon_n}' \left(\frac{1}{\frac{1}{B_n} - {\varepsilon_n}'} + B_n \right)
\end{aligned}
$$

which converges to 0, as shown in the proof of Theorem 1. In conclusion, both terms of the breakdown of $\sup_\theta \left|\nabla_n^{2\,(4)}(\theta)\right|$ tend to 0 as $n \to \infty$, which implies $\nabla_n^{2\,(4)}(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} 0$ for the choices of B_n, ε_n, h_n made before.

e) Uniform convergence of $\nabla_n^{2\,(1)}(\theta) + \nabla_n^{2\,(2)}(\theta)$ to $\nabla^2 H(\theta)$.

So far we have proved that $\nabla_n^{2\,(1)}(\theta) + \nabla_n^{2\,(2)}(\theta) \overset{\theta,\mathrm{p}}{\rightsquigarrow} E_4(\theta) + F_4(\theta)$. It remains to show that $E_4(\theta) + F_4(\theta) = \nabla^2 H(\theta)$. We have

$$
\begin{aligned}
& E_4(\theta) + F_4(\theta) \\
&= -2 \int \frac{f''(x+\bar\theta-\theta)}{f_e^s(x)} f(x-\bar\theta+\theta)dx + \int \left[\frac{f'(x+\bar\theta-\theta)}{f_e^s(x)}\right]^2 f(x-\bar\theta+\theta)dx\,.
\end{aligned}
$$

Writing $z = x + \bar\theta - \theta$, $z - v = x - \bar\theta + \theta$, we obtain

$$
E_4(\theta) + F_4(\theta) = -\int \frac{4f''(z)f(z-v)}{f(z)+f(z-v)}dz + \int \frac{4(f'(z))^2 f(z-v)}{(f(z)+f(z-v))^2}dz. \tag{31}
$$

Define now

$$
\begin{aligned}
A(z) &= \frac{d}{dz}\left\{2\frac{f'(z)f(z-v)}{f(z)+f(z-v)}\right\} \\
&= \frac{2f''(z)f(z-v)}{f(z)+f(z-v)} + \frac{2f'(z)f'(z-v)}{f(z)+f(z-v)} - \frac{2(f'(z))^2 f(z-v)}{[f(z)+f(z-v)]^2} \\
&\quad - \frac{2f'(z)f'(z-v)f(z-v)}{[f(z)+f(z-v)]^2}\,,
\end{aligned}
$$

yielding

$$
\begin{aligned}
-\int \frac{4f''(z)f(z-v)}{f(z)+f(z-v)}dz &= -2\int A(z)dz + 4\int \frac{f'(z)f'(z-v)}{f(z)+f(z-v)}dz \\
&\quad -4\int \frac{(f'(z))^2 f(z-v)}{[f(z)+f(z-v)]^2}dz \\
&\quad -4\int \frac{f'(z)f'(z-v)f(z-v)}{[(f(z)+f(z-v))/2]^2}dz.
\end{aligned}
$$

Substituting this last expression in (31), we obtain

$$
\begin{aligned}
E_4(\theta) + F_4(\theta) &= -2\int A(z)dz + 4\int \frac{f'(z)f'(z-v)}{f(z)+f(z-v)}dz \\
&\quad - \int \frac{4f'(z)f'(z-v)}{[f(z)+f(z-v)]^2} f(z-v)dz\,.
\end{aligned}
$$

Now, since, from **F0**, **F2**,

$$
\frac{f'(x)f(x-v)}{f(x)+f(x-v)} = \frac{f'(x)}{1+\frac{f'(x)}{f'(x-v)}} \to_{|x|\to\infty} 0\,,
$$

we have

$$\int_{-\infty}^{x} A(z)dz = \left[2\frac{f'(z)f(z-v)}{f(z)+f(z-v)}\right]_{-\infty}^{x} \to_{x\to\infty} 0\,.$$

Furthermore, using the mapping $y = -z + v$, and since by hypothesis f is symmetric about 0 (and thus f' is antisymmetric, i.e. $f'(-x) = -f'(x)$), we have

$$\mathcal{I} \;=\; \int \frac{4f'(z)f'(z-v)}{[f(z)+f(z-v)]^2} f(z-v)dz = \int \frac{4f'(y)f'(y-v)}{[f(y)+f(y-v)]^2} f(y)dy,$$

which, rewriting, gives

$$\mathcal{I} = \int \frac{4f'(z)f'(z-v)}{[f(z)+f(z-v)]^2} f(z)dz\,.$$

From this we obtain

$$\mathcal{I} = \frac{\mathcal{I}+\mathcal{I}}{2} = \int \frac{2f'(z)f'(z-v)}{f(z)+f(z-v)}dz.$$

From the last expression of $E_4(\theta) + F_4(\theta)$, we thus have

$$\begin{aligned} E_4(\theta) + F_4(\theta) &= \int \frac{4f'(z)f'(z-v)}{f(z)+f(z-v)}dz - \int \frac{4f'(z)f'(z-v)}{[f(z)+f(z-v)]^2} f(z-v)dz \\ &= 4\int \frac{f'(z)f'(z-v)}{f(z)+f(z-v)}dz - 2\int \frac{f'(z)f'(z-v)}{f(z)+f(z-v)}dz \\ &= 2\int \frac{f'(z)f'(z-v)}{f(z)+f(z-v)}dz\,, \end{aligned}$$

i.e.

$$E_4(\theta) + F_4(\theta) = \int \frac{f'(x-\bar{\theta}+\theta)f'(x+\bar{\theta}-\theta)}{f_e^s(x)}dx\,. \tag{32}$$

Consider now the second order derivative of $H(\theta)$ w.r.t. θ

$$\nabla_\theta^2 H(\theta) = -\int \frac{\partial^2}{\partial\theta^2}\left\{f_e^s(x)\right\}\left(1+\log f_e^s(x)\right)dx - \int \frac{\left(\frac{\partial}{\partial\theta}\left\{f_e^s(x)\right\}\right)^2}{f_e^s(x)}dx.$$

Notice that

$$\frac{\partial^2}{\partial\theta^2}\left\{f_e^s(x)\right\} = \frac{1}{2}\left[f''(x-\bar{\theta}+\theta) + f''(x+\bar{\theta}-\theta)\right] = \frac{d^2}{dx^2}\left\{f_e^s(x)\right\}.$$

Using this last result we can write

$$\int \frac{\partial^2}{\partial\theta^2}\left\{f_e^s(x)\right\}\left(1+\log f_e^s(x)\right)dx$$

$$
\begin{aligned}
&= \int \frac{d^2}{dx^2}\left\{f_e^s(x)\right\}\left(1+\log f_e^s(x)\right)dx \\
&= \left[\frac{d}{dx}\left\{f_e^s(x)\right\}\left(1+\log f_e^s(x)\right)\right]_{-\infty}^{+\infty} - \int \left(\frac{d}{dx}\left\{f_e^s(x)\right\}\right)^2 \frac{1}{f_e^s(x)}dx. \quad (33)
\end{aligned}
$$

We can rewrite the first term of this last expression as

$$
\begin{aligned}
&\left[\frac{d}{dx}\left\{f_e^s(x)\right\}\left(1+\log f_e^s(x)\right)\right]_{-\infty}^{+\infty} \\
&\quad = \left[\frac{d}{dx}\left\{f_e^s(x)\right\}\right]_{-\infty}^{+\infty} + \left[\left(\frac{\frac{d}{dx}\left\{f_e^s(x)\right\}}{\sqrt{f_e^s(x)}}\right)\left(2\sqrt{f_e^s(x)}\log\sqrt{f_e^s(x)}\right)\right]_{-\infty}^{+\infty}.
\end{aligned}
$$

Here, $\left[\frac{d}{dx}\left\{f_e^s(x)\right\}\right]_{-\infty}^{+\infty} = 0$ from **F0** and **F2**, and within the second term of the sum, the square of the first factor integrates to the Fisher information, which is finite, while the second factor tends to zero since $x\log x \to 0$ as $x \to 0$ and $f_e^s(x) \to 0$ as $|x| \to \infty$. Therefore $\frac{d}{dx}\left\{f_e^s(x)\right\}\log f_e^s(x) \to 0$ as $|x| \to \infty$. Finally, it remains from (33) that

$$
\begin{aligned}
\nabla_\theta^2 H(\theta) &= \int \frac{\left(\frac{d}{dx}\left\{f_e^s(x)\right\}\right)^2}{f_e^s(x)}dx - \int \frac{\left(\frac{\partial}{\partial\theta}\left\{f_e^s(x)\right\}\right)^2}{f_e^s(x)}dx \\
&= \int \frac{(f'(x-\bar\theta+\theta)+f'(x+\bar\theta-\theta))^2}{4f_e^s(x)}dx \\
&\quad - \int \frac{(f'(x-\bar\theta+\theta)-f'(x+\bar\theta-\theta))^2}{4f_e^s(x)}dx \\
&= \int \frac{f'(x-\bar\theta+\theta)f'(x+\bar\theta-\theta)}{f_e^s(x)}dx.
\end{aligned}
$$

Comparing this last result to (32), we obtain

$$E_4(\theta) + F_4(\theta) = \nabla_\theta^2 H(\theta),$$

i.e.

$$\nabla_\theta^2 \hat{H}_n(\theta) \overset{\theta;\mathrm{P}}{\rightsquigarrow} \nabla_\theta^2 H(\theta)$$

with the choices of B_n, h_n indicated in the theorem.

■

References

[1] Ash, R., 1965. Information Theory. Wiley, New York, (Republished by Dover, New York, 1990).

[2] Beirlant, J., Dudewicz, E., Györfi, L., van der Meulen, E., 1997. Nonparametric entropy estimation; an overview. Intern. J. Math. Stat. Sci. 6 (1), 17–39.

[3] Beran, R., 1977. Minimum Hellinger distance estimates for parametric models. Annals of Statistics 5 (3), 445–463.

[4] Bickel, P., 1982. On adaptive estimation. Annals of Statistics 10, 647–671.

[5] Broniatowski, M., 2003. Estimation through Kullback-Leibler divergence. Mathematical Methods of Statistics (to appear).

[6] Broniatowski, M., Keziou, A., 2004. Parametric estimation and testing through divergences. Prepublication 2004-1, L.S.T.A., Universit Paris 6.

[7] Dmitriev, Y., Tarasenko, F., 1973. On the estimation of functionals of the probability density and its derivatives. Theory of Probability and its Applications 18 (3), 628–633.

[8] Jennrich, R., 1969. Asymptotic properties of nonlinear least squares estimation. Annals of Math. Stat. 40, 633–643.

[9] Kozachenko, L., Leonenko, N., 1987. On statistical estimation of entropy of random vector. Problems Infor. Transmiss. 23 (2), 95–101, (translated from Problemy Peredachi Informatsii, in Russian, vol. 23, No. 2, pp. 9-16, 1987).

[10] Manski, C., 1984. Adaptive estimation of nonlinear regression models. Econometric Reviews 3 (2), 145–194.

[11] Newey, W., 1991. Uniform convergence in probability and stochastic equicontinuity. Econometrica 9 (4), 1161–1167.

[12] Pronzato, L., Thierry, E., 2001(a). Entropy minimization for parameter estimation problems with unknown distribution of the output noise. In: Proc. ICASSP'2001. Salt Lake City.

[13] Pronzato, L., Thierry, E., 2001(b). A minimum-entropy estimator for regression problems with unknown distribution of observation errors. In: Mohammad-Djafari, A. (Ed.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Proc. 20th Int. Workshop, Gif-sur-Yvette, France, July 2000. Am. Inst. of Physics, New York, pp. 169–180.

[14] Pronzato, L., Thierry, E., Wolsztynski, E., 2004. Minimum entropy estimation in semi parametric models: a candidate for adaptive estimation? In: Di Bucchianico, A., Läuter, H., Wynn, H. (Eds.), mODa'7 – Advances in Model–Oriented Design and Analysis, Proceedings of the 7th Int. Workshop, Heeze (Netehrlands). Physica Verlag, Heidelberg, to appear.

[15] Schuster, E., 1969. Estimation of a probability density function and its derivatives. Annals of Math. Stat. 40, 1187–1195.

[16] Stein, C., 1956. Efficient nonparametric testing and estimation. In: Proc. 3rd Berkeley Symp. Math. Stat. Prob. Vol. 1. University of California Press, Berkeley, pp. 187–196.

[17] Stone, C., 1975. Adaptive maximum likelihood estimators of a location parameter. Annals of Statistics 3 (2), 267–284.

[18] Vasicek, O., 1976. A test for normality based on sample entropy. Journal of Royal Statistical Society B38 (1), 54–59.