Efficient Estimation of Sensitivity Indices

Sébastien da Veiga, Fabrice Gamboa

Journal of Nonparametric Statistics
Vol. 00, No. 00, Month 200x, 1–41

RESEARCH ARTICLE

Efficient Estimation of Sensitivity Indices

Sébastien Da Veiga and Fabrice Gamboa
(*Received 00 Month 200x; in final form 00 Month 200x*)

In this paper we address the problem of efficient estimation of Sobol sensitivy indices. First, we focus on general functional integrals of conditional moments of the form $\mathbb{E}(\psi(\mathbb{E}(\varphi(Y)|X)))$ where (X, Y) is a random vector with joint density f and ψ and φ are functions that are differentiable enough. In particular, we show that asymptotical efficient estimation of this functional boils down to the estimation of crossed quadratic functionals. An efficient estimate of first-order sensitivity indices is then derived as a special case. We investigate its properties on several analytical functions and illustrate its interest on a reservoir engineering case.

Keywords: density estimation, semiparametric Cramér-Rao bound, global sensitivity analysis.

AMS Subject Classification: 2G20, 62G06, 62G07, 62P30

1. Introduction

In the past decade, the increasing interest in the design and analysis of computer experiments motivated the development of dedicated and sharp statistical tools (Santner et al.. 2003). Design of experiments, sensitivity analysis and proxy models are examples of research fields where numerous contributions have been proposed. More specifically, global Sensitivity Analysis (SA) is a key method for investigating complex computer codes which model physical phenomena. It involves a set of techniques used to quantify the influence of uncertain input parameters on the variability in numerical model responses. Recently, sensitivity studies have been applied in a large variety of fields, ranging from chemistry (Cukier et al. 1973; Turanyi 1990) or oil recovery (Iooss et al. 2011) to space science (Carrasco et al. 2007) and nuclear safety (Iooss et al. 2006).
In general, global SA refers to the probabilistic framework, meaning that the uncertain

IFP Energies nouvelles 1 & 4, avenue de Bois-Préau F-92852 Rueil-Malmaison Cedex
sebastien.da-veiga@ifpen.fr
Institut de Mathématiques Université Paul Sabatier F-31062 Toulouse Cedex 9
http://www.lsp.ups-tlse.fr/Fp/Gamboa. gamboa@math.univ-toulouse.fr.

ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

input parameters are modelled as a random vector. By propagation, every computer code output is itself a random variable. Global SA techniques then consists in comparing the probability distribution of the output with the conditional probability distribution of the output when some of the inputs are fixed. This yields in particular useful information on the impact of some parameters. Such comparisons can be performed by considering various criteria, each one of them providing a different insight on the input-output relationship. For example, some criteria are based on distances between the probability density functions (e.g. L^1 and L^2 norms (Borgonovo (2007)) or Kullback-Leibler distance (Liu et al. (2006)), while others rely on functionals of conditional moments. Among those, variance-based methods are the most widely used (Saltelli et al. 2000). They evaluate how the inputs contribute to the output variance through the so-called Sobol sensitivity indices (Sobol' 1993), which naturally emerge from a functional ANOVA decomposition of the output (Antoniadis 1984; Hoeffding 1948; Owen 1994). Interpretation of the indices in this setting makes it possible to exhibit which input or interaction of inputs most influences the variability of the computer code output. This can be typically relevant for model calibration (Kennedy and O'Hagan 2001) or model validation (Bayarri et al. 2007).

Consequently, in order to conduct a sensitivity study, estimation of such sensitivity indices is of great interest. Initially, Monte-Carlo estimates have been proposed (McKay 1995; Sobol' 1993). Recent work also focused on their asymptotic properties (Janon et al. 2012). However, in many applications, calls to the computer code are very expensive, from several minutes to hours. In addition, the number of inputs can be large, making Monte-Carlo approaches untractable in practice. To overcome this problem, recent work focused on the use of metamodeling techniques. The complex computer code is approximated by a mathematical model, referred to as a "metamodel", which should be as representative as possible of the computer code, with good prediction capability. Once the metamodel is built and validated, it is used in the extensive Monte-Carlo sampling instead of the complex numerical model. Several metamodels can be used: polynomials, Gaussian process metamodels (Oakley and O'Hagan (2004), Iooss et al. (2011)) or local polynomials (Da Veiga et al. (2009)). However, in these papers, the approach is generally empirical in the sense that no convergence study is performed and do not provide any insight about the asymptotic behavior of the sensitivity indices estimates. The only exception is the work of Da Veiga et al. (2009), where the authors investigate the convergence of a local-polynomial based estimate using the work of Fan and Gijbels (1996) and Wand and Jones (1994). In particular, this plug-in estimate achieves a nonparametric convergence rate.

In this paper, we go one step further and propose the first asymptotically efficient estimate for sensitivity indices. More precisely, we investigate the problem of efficient estimation of some general nonlinear functional based on the density of a pair of random variables. Our approach follows the work of Laurent (1996, 2005), and we also refer to Levit (1978) and Kerkyacharian and Picard (1996) for general results on nonlinear functionals estimation. Such functionals of a density appear in many statistical applications and their efficient estimation remains an active research field (Chacón and Tenreiro 2011; Giné and Nickl 2008; Giné and Mason 2008). However we consider functionals involving conditional densities, which necessitate a specific treatment. The estimate obtained here can be used for global SA involving general conditional moments, but it includes as a special case Sobol sensitivity indices. Note also that an extension of the approach developed in our work is simultaneously proposed in the context of sliced inverse regression (Loubes et al. 2011).

The paper is organized as follows. Section 2 first recaps variance-based methods for global SA. In particular, we point out which type of nonlinear functional appears in sensitivity indices. Section 3 then describes the theoretical framework and the proposed methodology for building an asymptotically efficient estimator. In Section 4, we focus on Sobol sensitivity indices and study numerical examples showing the good behavior of the proposed estimate. We also illustrate its interest on a reservoir engineering example, where uncertainties on the geology propagate to the potential oil recovery of a reservoir. Finally, all proofs are postponed to the appendix.

2. Global sensitivity analysis

In many applied fields, physicists and engineers are faced with the problem of estimating some sensitivity indices. These indices quantify the impact of some input variables on an output. The general situation may be formalized as follows.
The output $Y \in \mathbb{R}$ is a nonlinear regression of input variables $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_l)$ ($l \geq 1$ is generally large). This means that Y and $\boldsymbol{\tau}$ satisfy the input-output relationship

$$Y = \Phi(\boldsymbol{\tau}) \tag{1}$$

where Φ is a known nonlinear function. Usually, Φ is complicated and has not a closed form, but it may be computed through a computer code (Oakley and O'Hagan 2004). In general, the input $\boldsymbol{\tau}$ is modelled by a random vector, so that Y is also a random variable. A common way to quantify the impact of input variables is to use the so-called Sobol sensitivity indices (Sobol' 1993). Assuming that all the random variables are square integrable, the Sobol index for the input τ_j ($j = 1, \ldots, l$) is

$$\Sigma_j = \frac{\mathrm{Var}(\mathbb{E}(Y|\tau_j))}{\mathrm{Var}(Y)}. \tag{2}$$

Observing an i.i.d. sample $(Y_1, \boldsymbol{\tau}^{(1)}), \ldots, (Y_n, \boldsymbol{\tau}^{(n)})$ (with $Y_i = \Phi(\boldsymbol{\tau}^{(i)})$, $i = 1, \ldots, n$), the goal is is then to estimate Σ_j ($j = 1, \ldots, l$). Obviously, (2) may be rewritten as

$$\Sigma_j = \frac{\mathbb{E}(\mathbb{E}(Y|\tau_j)^2) - \mathbb{E}(Y)^2}{\mathrm{Var}(Y)}.$$

Thus, in order to estimate Σ_j, the hard part is $\mathbb{E}(\mathbb{E}(Y|\tau_j)^2)$. In this paper we will provide an asymptotically efficient estimate for this kind of quantity. More precisely we will tackle the problem of asymptotically efficient estimation of some general nonlinear functional. Let us specify the functionals we are interested in. Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be a sample of i.i.d. random vectors of $\mathbb{R}^2$ having a *regular* density f (see Section 3 for the precise frame). We will study the estimation of the nonlinear functional

$$\begin{aligned} T(f) &= \mathbb{E}\Big(\psi\big(\mathbb{E}(\varphi(Y)|X)\big)\Big) \\ &= \iint \psi\left(\frac{\int \varphi(y) f(x,y) dy}{\int f(x,y) dy}\right) f(x,y) dx dy \end{aligned}$$

where ψ and φ are regular functions. Hence, the Sobol indices are the particular case obtained with $\psi(\xi) = \xi^2$ and $\varphi(\xi) = \xi$.
The method developed in order to obtain an asymptotically efficient estimate for $T(f)$ follows the one developed by Laurent (1996). Roughly speaking, it involves a preliminary estimate $\hat{f}$ of f built on a small part of the sample. This preliminary estimate is used in a Taylor expansion of $T(f)$ up to the second order in a neighbourhood of $\hat{f}$. This expansion allows to remove the bias that occurs when using a direct plug-in method. Hence, the bias correction involves a quadratic functional of f. Due to the form of T, this quadratic functional of f may be written as

$$\theta(f) = \iiint \eta(x, y_1, y_2) f(x, y_1) f(x, y_2) dx dy_1 dy_2.$$

This kind of functional does not fall in the frame treated in Laurent (1996) or Giné and Nickl (2008) and have not been studied to the best of our knowledge. We study this problem in Section 3.1 where we build an asymptotically efficient estimate for θ. Efficient estimation of $T(f)$ is then investigated in Section 3.2.

3. Model frame and method

Let $a < b$ and $c < d$, $L^2(dxdy)$ will denote the set of square integrable functions on $[a, b] \times [c, d]$. Further, $L^2(dx)$ (resp. $L^2(dy)$) will denote the set of square integrable functions on $[a, b]$ (resp. $[c, d]$). For sake of simplicity, we work in the whole paper with the Lebesgue measure as reference measure. Nevertheless, most of the results presented can be obtained for a general reference measure on $[a, b] \times [c, d]$. Let $(\alpha_{i_\alpha}(x))_{i_\alpha \in D_1}$ (resp. $(\beta_{i_\beta}(y))_{i_\beta \in D_2}$) be a countable orthonormal basis of $L^2(dx)$ (resp. of $L^2(dy)$). We set $p_i(x, y) = \alpha_{i_\alpha}(x)\beta_{i_\beta}(y)$ with $i = (i_\alpha, i_\beta) \in D := D_1 \times D_2$. Obviously $(p_i(x, y))_{i \in D}$ is a countable orthonormal (tensor) basis of $L^2(dxdy)$. We will also use the following subset of $L^2(dxdy)$:

$$\mathcal{E} = \left\{ \sum_{i \in D} e_i p_i : (e_i)_{i \in D} \text{ is a sequence with } \sum_{i \in D} \left| \frac{e_i}{c_i} \right|^2 \leq 1 \right\},$$

here $(c_i)_{i \in D}$ is a given fixed positive sequence.
Let (X, Y) having a bounded joint density f on $[a, b] \times [c, d]$ from which we have a sample $(X_i, Y_i)_{i=1,\ldots,n}$. We will also assume that f lies in the ellipsoid $\mathcal{E}$. Recall that we wish to estimate a conditional functional

$$\mathbb{E}\Big(\psi\big(\mathbb{E}(\varphi(Y)|X)\big)\Big)$$

where φ is a measurable bounded function with $\chi_1 \leq \varphi \leq \chi_2$ and $\psi \in C^3([\chi_1, \chi_2])$ the set of thrice continuously differentiable functions on $[\chi_1, \chi_2]$. This last quantity can be

expressed in terms of an integral depending on the joint density f:

$$T(f) = \iint \psi\left(\frac{\int \varphi(y) f(x,y)dy}{\int f(x,y)dy}\right) f(x,y)dxdy.$$
$$= \iint \psi(m(x)) f(x,y)dxdy$$

where $m(x) = \int \varphi(y)f(x,y)dy / \int f(x,y)dy$ is the conditional expectation of $\varphi(Y)$ given $(X = x)$. We suggest as a first step to consider a preliminary estimator $\hat{f}$ of f, and to expand $T(f)$ in a neighborhood of $\hat{f}$. To achieve this goal we first define $F : [0,1] \to \mathbb{R}$:

$$F(u) = T(uf + (1-u)\hat{f}) \quad (u \in [0,1]).$$

The Taylor expansion of F between 0 and 1 up to the third order is

$$F(1) = F(0) + F'(0) + \frac{1}{2}F''(0) + \frac{1}{6}F'''(\xi)(1-\xi)^3 \tag{3}$$

for some $\xi \in]0,1[$. Here, we have

$$F(1) = T(f)$$

and

$$F(0) = T(\hat{f}) = \iint \psi\left(\frac{\int \varphi(y)\hat{f}(x,y)dy}{\int \hat{f}(x,y)dy}\right) \hat{f}(x,y)dxdy$$
$$= \iint \psi(\hat{m}(x))\hat{f}(x,y)dxdy$$

where $\hat{m}(x) = \int \varphi(y)\hat{f}(x,y)dy / \int \hat{f}(x,y)dy$. Straightforward calculations also give higher-order derivatives of F :

$$F'(0) = \iint \Big(\big[\varphi(y) - \hat{m}(x)\big]\dot{\psi}(\hat{m}(x)) + \psi(\hat{m}(x))\Big)\Big(f(x,y) - \hat{f}(x,y)\Big)dxdy$$

$$F''(0) = \iiint \frac{\ddot{\psi}(\hat{m}(x))}{\Big(\int \hat{f}(x,y)dy\Big)}\big(\hat{m}(x) - \varphi(y)\big)\big(\hat{m}(x) - \varphi(z)\big)$$
$$\Big(f(x,y) - \hat{f}(x,y)\Big)\Big(f(x,z) - \hat{f}(x,z)\Big)dxdydz$$

$$
\begin{aligned}
F'''(\xi) = \iiint & \frac{\Big(\int \hat{f}(x,y)dy\Big)^2}{\Big(\int \xi f(x,y) + (1-\xi)\hat{f}(x,y)dy\Big)^5} \\
& \big[\big(\hat{m}(x) - \varphi(y)\big)\big(\hat{m}(x) - \varphi(z)\big)\big(\hat{m}(x) - \varphi(t)\big) \\
& \left(\int \hat{f}(x,y)dy\right) \dddot{\psi}\left(\hat{r}(\xi,x)\right) - 3\big(\hat{m}(x) - \varphi(y)\big)\big(\hat{m}(x) - \varphi(z)\big) \\
& \left(\int [\xi f(x,y) + (1-\xi)\hat{f}(x,y)]dy\right) \ddot{\psi}\left(\hat{r}(\xi,x)\right)\Big] \\
& \Big(f(x,y) - \hat{f}(x,y)\Big)\Big(f(x,z) - \hat{f}(x,z)\Big) \\
& \Big(f(x,t) - \hat{f}(x,t)\Big)dxdydzdt
\end{aligned}
$$

where $\hat{r}(\xi,x) = \dfrac{\int \varphi(y)[\xi f(x,y) + (1-\xi)\hat{f}(x,y)]dy}{\int [\xi f(x,y) + (1-\xi)\hat{f}(x,y)]dy}$ and $\dot{\psi}$, $\ddot{\psi}$ and $\dddot{\psi}$ denote the three first derivatives of ψ.

Plugging these expressions into (3) yields the following expansion for $T(f)$:

$$
\begin{aligned}
T(f) = & \iint H(\hat{f},x,y)f(x,y)dxdy \\
& + \iiint K(\hat{f},x,y,z)f(x,y)f(x,z)dxdydz + \Gamma_n
\end{aligned} \tag{4}
$$

where

$$
\begin{aligned}
H(\hat{f},x,y) &= \big[\varphi(y) - \hat{m}(x)\big]\dot{\psi}(\hat{m}(x)) + \psi(\hat{m}(x)), \\
K(\hat{f},x,y,z) &= \frac{1}{2}\frac{\ddot{\psi}(\hat{m}(x))}{\Big(\int \hat{f}(x,y)dy\Big)}\big(\hat{m}(x) - \varphi(y)\big)\big(\hat{m}(x) - \varphi(z)\big), \\
\Gamma_n &= \frac{1}{6}F'''(\xi)(1-\xi)^3
\end{aligned}
$$

for some $\xi \in]0,1[$. Notice that the first term is a linear functional of the density f, it will be estimated with

$$\frac{1}{n_2}\sum_{j=1}^{n_2} H(\hat{f}, X_j, Y_j).$$

The second one involves a crossed term integral which can be written as

$$\iiint \eta(x,y_1,y_2)f(x,y_1)f(x,y_2)dxdy_1dy_2 \tag{5}$$

where $\eta : \mathbb{R}^3 \to \mathbb{R}$ is a bounded function verifying $\eta(x,y_1,y_2) = \eta(x,y_2,y_1)$ for all $(x,y_1,y_2) \in \mathbb{R}^3$. In summary, the first term can be easily estimated, unlike the second one

which deserves a specific study. In the next section we then focus on the asymptotically efficient estimation of such crossed quadratic functionals. In Section 3.2, these results are finally used to propose an asymptotically efficient estimator for $T(f)$.

3.1. *Efficient estimation of quadratic functionals*

In this section, our aim is to build an asymptotically efficient estimate for

$$\theta = \iiint \eta(x, y_1, y_2) f(x, y_1) f(x, y_2) dx dy_1 dy_2.$$

We denote $a_i = \int f p_i$ the scalar product of f with p_i as defined at the beginning of Section 3. We will first build a projection estimator achieving a bias equal to

$$-\iiint \left[S_M f(x, y_1) - f(x, y_1)\right] \left[S_M f(x, y_2) - f(x, y_2)\right] \eta(x, y_1, y_2) dx dy_1 dy_2$$

where $S_M f = \sum_{i \in M} a_i p_i$ and M is a subset of D. Thus, the bias would only be due to projection. Developing the previous expression leads to a goal bias equal to

$$\begin{aligned} & 2 \iiint S_M f(x, y_1) f(x, y_2) \eta(x, y_1, y_2) dx dy_1 dy_2 \\ & - \iiint S_M f(x, y_1) S_M f(x, y_2) \eta(x, y_1, y_2) dx dy_1 dy_2 \\ & - \iiint f(x, y_1) f(x, y_2) \eta(x, y_1, y_2) dx dy_1 dy_2. \end{aligned} \tag{6}$$

Consider now the estimator $\hat{\theta}_n$ defined by

$$\begin{aligned} \hat{\theta}_n = & \frac{2}{n(n-1)} \sum_{i \in M} \sum_{j \neq k=1}^{n} p_i(X_j, Y_j) \int p_i(X_k, u) \eta(X_k, u, Y_k) du \\ & - \frac{1}{n(n-1)} \sum_{i, i' \in M} \sum_{j \neq k=1}^{n} p_i(X_j, Y_j) p_{i'}(X_k, Y_k) \\ & \int p_i(x, y_1) p_{i'}(x, y_2) \eta(x, y_1, y_2) dx dy_1 dy_2. \end{aligned} \tag{7}$$

This estimator achieves the desired bias :

Lemma 3.1: *The estimator $\hat{\theta}_n$ defined in (7) estimates θ with bias equal to*

$$-\iiint [S_M f(x, y_1) - f(x, y_1)][S_M f(x, y_2) - f(x, y_2)] \eta(x, y_1, y_2) dx dy_1 dy_2.$$

Since we will carry out an asymptotic analysis, we will work with a sequence $(M_n)_{n \geq 1}$ of subsets of D. We will need an extra assumption concerning this sequence:

A1. For all $n \geq 1$, we can find a subset $M_n \subset D$ such that $\left(\sup_{i \notin M_n} |c_i|^2\right)^2 \approx \frac{|M_n|}{n^2}$ ($A_n \approx B_n$ means $\lambda_1 \leq A_n/B_n \leq \lambda_2$ for some positive constants λ_1 and λ_2). Furthermore, $\forall t \in L^2(dxdy)$, $\int (S_{M_n} t - t)^2 dxdy \to 0$ when $n \to \infty$.

The following theorem gives the most important properties of our estimate $\hat{\theta}_n$:

Theorem 3.2 : *Assume A1 hold. Then $\hat{\theta}_n$ has the following properties:*

(i) If $|M_n|/n \to 0$ when $n \to \infty$, then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \to \mathcal{N}\left(0, \Lambda(f,\eta)\right), \tag{8}$$

$$\left|\mathbb{E}\left(\hat{\theta}_n - \theta\right)^2 - \Lambda(f,\eta)\right| \leq \gamma_1 \left[\frac{|M_n|}{n} + \|S_{M_n} f - f\|_2 + \|S_{M_n} g - g\|_2\right], \tag{9}$$

where $g(x,y) := \int f(x,u)\eta(x,y,u)du$ and

$$\Lambda(f,\eta) = 4\left[\iint g(x,y)^2 f(x,y)dxdy - \left(\iint g(x,y)f(x,y)dxdy\right)^2\right].$$

(ii) Otherwise

$$\mathbb{E}\left(\hat{\theta}_n - \theta\right)^2 \leq \gamma_2 \frac{|M_n|}{n},$$

where γ_1 and γ_2 are constants depending only on $\|f\|_\infty$, $\|\eta\|_\infty$ and Δ_Y (with $\Delta_Y = d - c$). Moreover, these constants are increasing functions of these quantities.

Remark 1 : Since in our main result (to be given in the next section) η will depend on n through the preliminary estimator $\hat{f}$, we need in (9) a bound that depends explicitly on n. Note however that (9) implies

$$\lim_{n\to\infty} n\mathbb{E}\left(\hat{\theta}_n - \theta\right)^2 = \Lambda(f,\eta).$$

The asymptotic properties of $\hat{\theta}_n$ are of particular importance, in the sense that they are optimal as stated in the following theorem.

Theorem 3.3 : *Consider the estimation of*

$$\theta = \theta(f) = \iiint \eta(x,y_1,y_2) f(x,y_1) f(x,y_2) dxdy_1dy_2.$$

Let $f_0 \in \mathcal{E}$. Then, for all estimator $\hat{\theta}_n$ of $\theta(f)$ and every family $\mathcal{V}(f_0)$ of vicinities of f_0, we have

$$\inf_{\{\mathcal{V}(f_0)\}} \liminf_{n\to\infty} \sup_{f \in \mathcal{V}(f_0)} n\mathbb{E}(\hat{\theta}_n - \theta(f_0))^2 \geq \Lambda(f_0,\eta).$$

In other words, the optimal asymptotic variance for the estimation of θ is $\Lambda(f_0, \eta)$. As our estimator defined in (7) achieves this variance, it is therefore asymptotically efficient. We are now ready to use this result to propose an efficient estimator of $T(f)$.

3.2. *Main Theorem*

In this section we come back to our main problem of the asymptotically efficient estimation of

$$T(f) = \iint \psi\left(\frac{\int \varphi(y) f(x,y) dy}{\int f(x,y) dy}\right) f(x,y) dx dy.$$

Recall that we have derived in (4) an expansion for $T(f)$. The key idea is to use here the previous results on the estimation of crossed quadratic functionals. Indeed we have provided an asymptotically efficient estimator for the second term of this expansion, conditionally on $\hat{f}$. A natural and straightforward estimator for $T(f)$ is then

$$\begin{aligned}
\widehat{T}_n = & \frac{1}{n_2} \sum_{j=1}^{n_2} H(\hat{f}, X_j, Y_j) \\
& + \frac{2}{n_2(n_2-1)} \sum_{i \in M} \sum_{j \neq k=1}^{n_2} p_i(X_j, Y_j) \int p_i(X_k, u) K(\hat{f}, X_k, u, Y_k) du \\
& - \frac{1}{n_2(n_2-1)} \sum_{i,i' \in M} \sum_{j \neq k=1}^{n_2} p_i(X_j, Y_j) p_{i'}(X_k, Y_k) \\
& \int p_i(x, y_1) p_{i'}(x, y_2) K(\hat{f}, x, y_1, y_2) dx dy_1 dy_2.
\end{aligned}$$

In the above expression, one can note that the remainder Γ_n does not appear : we will see in the proof of the following theorem that it is negligible comparing to the two first terms.

In order to study the asymptotic properties of $\widehat{T}_n$, some assumptions are required concerning the behavior of the joint density f and its preliminary estimator $\hat{f}$:

A2. $\text{supp} f \subset [a,b] \times [c,d]$ and $\forall (x,y) \in \text{supp} f$, $0 < \alpha \leq f(x,y) \leq \beta$ with $\alpha, \beta \in \mathbb{R}$

A3. One can find an estimator $\hat{f}$ of f built with $n_1 \approx n/\log(n)$ observations, such that

$$\forall (x,y) \in \text{supp} f,\ 0 < \alpha - \epsilon \leq \hat{f}(x,y) \leq \beta + \epsilon.$$

Moreover,

$$\forall 2 \leq q < +\infty,\ \forall l \in \mathbb{N}^*,\ \mathbb{E}_f \|\hat{f} - f\|_q^l \leq C(q,l) n_1^{-l\lambda}$$

for some $\lambda > 1/6$ and some constant $C(q,l)$ not depending on f belonging to the ellipsoid $\mathcal{E}$.

Here supp$f$ denotes the set where f is different from 0. Assumption A2 is restrictive in the sense that only densities with compact support can be considered, excluding for example a Gaussian joint distribution.
Assumption A3 imposes to the estimator $\hat{f}$ a convergence fast enough towards f. We will use this result to control the remainder term Γ_n.

We can now state the main theorem of the paper. It investigates the asymptotic properties of $\widehat{T}_n$ under assumptions A1, A2 and A3.

Theorem 3.4 : *Assume that A1, A2 and A3 hold. Then $\widehat{T}_n$ has the following properties if $\frac{|M_n|}{n} \to 0$:*

$$\sqrt{n}\left(\widehat{T}_n - T(f)\right) \to \mathcal{N}\left(0, C(f)\right), \tag{10}$$

$$\lim_{n\to\infty} n\mathbb{E}\left(\widehat{T}_n - T(f)\right)^2 = C(f), \tag{11}$$

where $C(f) = \mathbb{E}\Big(Var(\varphi(Y)|X)\Big[\dot{\psi}\big(\mathbb{E}(Y|X)\big)\Big]^2\Big) + Var\Big(\psi\big(\mathbb{E}(\varphi(Y)|X)\big)\Big)$.

We can also compute as in the previous section the semiparametric Cramér-Rao bound for this problem.

Theorem 3.5 : *Consider the estimation of*

$$T(f) = \iint \psi\left(\frac{\int \varphi(y) f(x,y)dy}{\int f(x,y)dy}\right) f(x,y)dxdy = \mathbb{E}\Big(\psi\big(\mathbb{E}(\varphi(Y)|X)\big)\Big)$$

for a random vector (X,Y) with joint density $f \in \mathcal{E}$. Let $f_0 \in \mathcal{E}$ be a density verifying the assumptions of Theorem 3.4. Then, for all estimator $\widehat{T}_n$ of $T(f)$ and every family $\mathcal{V}(f_0)$ of vicinities of f_0, we have

$$\inf_{\{\mathcal{V}(f_0)\}} \liminf_{n\to\infty} \sup_{f\in\mathcal{V}(f_0)} n\mathbb{E}(\widehat{T}_n - T(f_0))^2 \geq C(f_0).$$

Combination of theorems 3.4 and 3.5 finally proves that $\widehat{T}_n$ is asymptotically efficient.

4. Application to the estimation of sensitivity indices

Now that we have built an asymptotically efficient estimate for $T(f)$, we can apply it to the particular case we were initially interested it: the estimation of Sobol sensitivity indices. Let us then come back to model (1) :

$$Y = \Phi(\boldsymbol{\tau})$$

where we wish to estimate (2):

$$\Sigma_j = \frac{\mathrm{Var}(\mathbb{E}(Y|\tau_j))}{\mathrm{Var}(Y)} = \frac{\mathbb{E}(\mathbb{E}(Y|\tau_j)^2) - \mathbb{E}(Y)^2}{\mathrm{Var}(Y)} \quad j = 1, \dots, l.$$

To do so, we have an i.i.d. sample $(Y_1, \boldsymbol{\tau}^{(1)}), \dots, (Y_n, \boldsymbol{\tau}^{(n)})$. We will only give here the procedure for the estimation of Σ_1 since it will be the same for the other sensitivity indices. Denoting $X := \tau_1$, this problem is equivalent to estimating $\mathbb{E}(\mathbb{E}(Y|X)^2)$ with an i.i.d. sample $(Y_1, X_1), \dots, (Y_n, X_n)$ with joint density f. We can hence apply the estimate we developed previously by letting $\psi(\xi) = \xi^2$ and $\varphi(\xi) = \xi$:

$$\begin{aligned} T(f) &= \mathbb{E}(\mathbb{E}(Y|X)^2) \\ &= \iint \left(\frac{\int y f(x,y) dy}{\int f(x,y) dy} \right)^2 f(x,y) dx dy. \end{aligned}$$

The Taylor expansion in this case becomes

$$\begin{aligned} T(f) = &\iint H(\hat{f}, x, y) f(x,y) dx dy \\ &+ \iiint K(\hat{f}, x, y, z) f(x,y) f(x,z) dx dy dz + \Gamma_n \end{aligned}$$

where

$$\begin{aligned} H(\hat{f}, x, y) &= 2y\hat{m}(x) - \hat{m}(x)^2, \\ K(\hat{f}, x, y, z) &= \frac{1}{\left(\int \hat{f}(x,y) dy \right)} \big(\hat{m}(x) - y \big) \big(\hat{m}(x) - z \big) \end{aligned}$$

and the corresponding estimator is

$$\begin{aligned} \widehat{T}_n = &\frac{1}{n_2} \sum_{j=1}^{n_2} H(\hat{f}, X_j, Y_j) \\ &+ \frac{2}{n_2(n_2-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n_2} p_i(X_j, Y_j) \int p_i(X_k, u) K(\hat{f}, X_k, u, Y_k) du \\ &- \frac{1}{n_2(n_2-1)} \sum_{i, i' \in M} \sum_{j \neq k = 1}^{n_2} p_i(X_j, Y_j) p_{i'}(X_k, Y_k) \\ &\int p_i(x, y_1) p_{i'}(x, y_2) K(\hat{f}, x, y_1, y_2) dx dy_1 dy_2. \end{aligned}$$

for some preliminary estimator $\hat{f}$ of f, an orthonormal basis $(p_i)_{i \in D}$ of $L^2(dxdy)$ and a subset $M \subset D$ verifying the hypotheses of Theorem 3.4.

We propose now to investigate the practical behavior of this estimator on two analytical models and on a reservoir engineering test case. In all subsequent simulation studies,

Table 1. Conditional moments for analytical model (12). Mean and standrad deviation of $\widehat{T}_n$ for different values of n.

Inputs	$\mathbb{E}(\mathbb{E}(Y\|\tau_j)^2)$	$\widehat{T}_n$ $n = 100$	$\widehat{T}_n$ $n = 10000$
Configuration (a)			
τ_1	0.5733	0.5894 +/- 0.052	0.5729 +/- 0.005
τ_2	0.5611	0.5468 +/- 0.054	0.5611 +/- 0.005
Configuration (b)			
τ_1	314.04	305.98 +/- 52.1	318.27 +/- 7.52
τ_2	779.85	814.04 +/- 10.3	787.82 +/- 0.53
Configuration (c)			
τ_1	16258	18414 +/- 3759	16897 +/- 427
τ_2	44034	44667 +/- 82.6	44073 +/- 8.17

the preliminary estimator $\hat{f}$ will be a kernel density estimator with bounded support built on $n_1 = [\log(n)/n]$ observations. Moreover, we choose the Legendre polynomials on $[a, b]$ and $[c, d]$ to build the orthonormal basis $(p_i)_{i\in D}$ and we will take $|M| = \sqrt{n}$. Finally, the integrals in $\widehat{T}_n$ are computed with an adaptive Simpson quadrature.

4.1. *Simulation study on analytical functions*

The first model we investigate is

$$Y = \tau_1 + \tau_2^4 \tag{12}$$

where three configurations are considered (τ_1 and τ_2 being independent):

(a) $\tau_j \sim \mathcal{U}(0, 1)$, $j = 1, 2$;
(b) $\tau_j \sim \mathcal{U}(0, 3)$, $j = 1, 2$;
(c) $\tau_j \sim \mathcal{U}(0, 5)$, $j = 1, 2$.

For each configuration, we report the results obtained with $n = 100$ and $n = 10000$ in Table 1. Note that we repeat the estimation 100 times with different different random samples of (τ_1, τ_2).

The asymptotically efficient estimator $\widehat{T}_n$ gives a very accurate approximation of sensitivity indices when $n = 10000$. But surprisingly, it also gives a reasonably accurate estimate when n only equals 100, whereas it has been built to achieve the best symptotic rate of convergence.

It is then interesting to compare it with other estimators, more precisely two nonparametric estimators that have been specifically built to give an accurate approximation of sensitivity indices when n is not large. The first one is based on a Gaussian process metamodel (Oakley and O'Hagan 2004), while the other one involves local polynomial

Table 2. Comparison between efficient estimation and nonparametric estimates on analytical model (13).

	True value	Oakley-O'Hagan	Local polynomials	$\widehat{T}_n$
$\mathrm{Var}(\mathbb{E}(Y\|X^1))$	1.0932	1.0539	1.0643	1.1701
$\mathrm{Var}(\mathbb{E}(Y\|X^2))$	0.0729	0.1121	0.0527	0.0939

estimators (Da Veiga et al. 2009). The comparison is performed on the following model :

$$Y = 0.2\exp(\tau_1 - 3) + 2.2|\tau_2| + 1.3\tau_2^6 - 2\tau_2^2 - 0.5\tau_2^4 - 0.5\tau_1^4 + 2.5\tau_1^2 + 0.7\tau_1^3 + \frac{3}{(8\tau_1 - 2)^2 + (5\tau_2 - 3)^2 + 1} + \sin(5\tau_1)\cos(3\tau_1^2) \qquad (13)$$

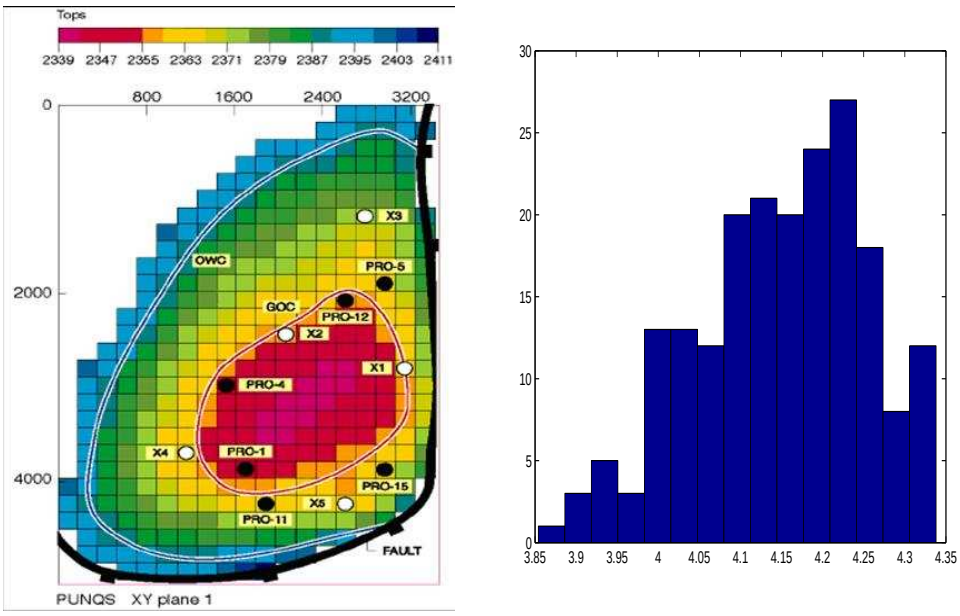
where τ_1 and τ_2 are independent and uniformly distributed on $[-1, 1]$. This nonlinear function is interesting since it presents a peak and valleys. We estimate the sensitivity indices with a sample of size $n = 100$, the results are given in Table 2.

Globally, the best estimates are given by the local polynomials technique. However, the accuracy of the asymptotically efficient estimator $\widehat{T}_n$ is comparable to that of the nonparametric ones. These results confirm that $\widehat{T}_n$ is a valuable estimator even with a rather complex model and a small sample size (recall that here $n = 100$).

4.2. *Reservoir engineering example*

The PUNQ test case (Production forecasting with UNcertainty Quantification) is an oil reservoir model derived from real field data (Manceau et al. 2001). The considered reservoir is surrounded by an aquifer in the north and the west, and delimited by a fault in the south and the east. The geological model is composed of five independent layers, three of good quality and two of poorer quality. Six producer wells (PRO-1, PRO-4, PRO-5, PRO-11, PRO-12 and PRO-15) have been drilled, and production is supported by four additional wells injecting water (X1, X2, X3 and X4). The geometry of the reservoir and the well locations are given in Figure 1, left.

In this setting, 7 variables which are characteristic of media, rocks, fluids or aquifer activity, are considered as uncertain: the coefficient of aquifer strength (AQUI), horizontal and vertical permeability multipliers in good layers (MPV1 and MPH1, respectively), horizontal and vertical permeability multipliers in poor layers (MPV2 and MPH2, respectively), residual oil saturation after waterflood and after gas flood (SORW and SORG, respectively). We focus here on the cumulative production of oil of this field during 12 years. In practice, a fluid flow simulator is used to forecast this oil production for every value of the uncertain parameters we might want to investigate. The uncertain parameters are assumed to be uniformly distributed, with ranges given in Table 3. We draw a random sample of size $n = 200$ of these 7 parameters, and perform the corresponding fluid-flow simulations to compute the cumulative oil production after 12 years. The histogram of the production obtained with this sampling is depicted in Figure 1, right. Clearly, the impact of the uncertain parameters on oil production is



(a) (b)

Figure 1. Left: top view of the PUNQ reservoir. Producer and injector wells are indicated by black and white circles, respectively. OWC and GOW stand for Oil Water Contact and Gas Oil Contact. Right: histogram of the cumulative production after 12 years (10^6 m^3).

Table 3. Range of variation and estimated first-order sensitivity index of the uncertain parameters of the PUNQ model.

Parameter	Range of variation	Estimated sensitivity index with $\widehat{T}_n$ (%)
AQUI	0.2 - 0.3	7.206
MPH1	0.8 - 1.2	40.929
MPH2	0.8 - 1.2	0.419
MPV1	0.8 - 1.2	0.041
MPV2	0.8 - 1.2	0.693
SORG	0.15 - 0.2	0.338
SORW	0.15 - 0.25	49.4

large, since different values yield forecats varying by tens of thousands of oil barrels. In this context, reservoir engineers aim at identifying which parameters affect the most the production. This help them design strategies in order to reduce the most influential uncertainties, which will reduce, by propagation, the uncertainty on production forecasts.

In this context, computation of sensivity indices is of great interest. Starting from the random sample of size $n = 200$, we then estimate the first-order sensitivity index of each parameter with the estimator $\widehat{T}_n$. Results are given in Table 3. As expected, the most influential parameters are the horizontal permeability multiplier in the good

reservoir units MPH1 and the residual oil saturation after waterflood SORW. Indeed, fluid dispacement towards the producer wells is mainly driven by the permeability in units with good petrophysical properties and by water injection. More interestingly, vertical permeability multipliers do not seem to impact oil production in this case. This means that fluid displacements are mainly horizontal in this reservoir.

5. Discussion and conclusions

In this paper, we developed a framework to build an asymptotically efficient estimate for nonlinear conditional functionals. This estimator is both practically computable and has optimal asymptotic properties. In particular, we show how Sobol sensitivty indices appear as a special case of our estimator. We investigate its practical behavior on two analytical functions, and illustrate that it can compete with metamodel-based estimators. A reservoir engineering application case is also studied, where geological and petrophysical uncertain parameters affect the forecasts on oil production. The methodology developed here will be extended to other problems in forthcoming work. A very attractive extension is the construction of an adaptive procedure to calibrate the size of M_n as done in Laurent (2005) for the L^2 norm. However, this problem is non obvious since it would involve treating refined inequalities on U-statistics such as presented in Houdret and Reynaud (2002). From a sensitivity analysis perspective, we will also investigate efficient estimation of other indices based on entropy or other norms. Ideally, this would give a general framework for building estimates in global sensitivity analysis.

Acknowledgements

Many thanks are due to A. Antoniadis, B. Laurent and F. Wahl for helpful discussion. This work has been partially supported by the French National Research Agency (ANR) through COSINUS program (project COSTA-BRAVA ANR-09-COSI-015).

References

Antoniadis, A. (1984). Analysis of variance on function spaces. *Math. Oper. Forsch. und Statist.*, series Statistics, 15(1):59–71.

Bayarii, M.J., Berger, J., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C., and Tu, J. (2007). A framework for validation of computer models. *Technometrics*, 49:138–154.

Borgonovo E. (2007). A New Uncertainty Importance Measure. *Reliability Engineering and System Saftey*, 92:771–784.

Carrasco, N., Banaszkiewicz, M., Thissen, R., Dutuit, O., and Pernot, P. (2007). Uncertainty analysis of bimolecular reactions in Titan ionosphere chemistry model. *Planetary and Space Science*, 55:141–157.

Chacón, J.E. and Tenreiro C. (2011) Exact and Asymptotically Optimal Bandwidths for Kernel Estimation of Density Functionals. *Methodol Comput Appl Probab*, DOI 10.1007/s11009-011-9243-x.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., and Schaibly, J.H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59:3873–3878.

Da Veiga, S., Wahl, F., and Gamboa, F. (2006). Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 59(4):452–463.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.

Ferrigno, S. and Ducharme, G.R. (2005). Un test d'adéquation global pour la fonction de répartition conditionnelle. *Comptes rendus. Mathématique*, 341:313–316.

Giné, E. and Nickl, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli*, 14(1):47–61

Giné, E. and Mason, D.M (2008). Uniform in Bandwidth Estimation of Integral Functionals of the Density Function *Scandinavian Journal of Statistics,*, 35:739–761

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The annals of Mathematical Statistics*, 19:293–32 5.

Houdré, C. and Reynaud, P. (2002). Stochastic inequalities and applications. In *Euroconference on Stochastic inequalities and applications*. Birkhauser.

Ibragimov, I.A. and Khas'minskii, R.Z. (1991). Asymptotically normal families of distributions and efficient estimation. *The Annals of Statistics*,19:1681–1724.

Iooss, B., Marrel, A., Da Veiga, S. and Ribatet, M. (2011). Global sensitivity analysis of stochastic computer models with joint metamodels *Stat Comput*,DOI 10.1007/s11222-011-9274-8.

Iooss, B., Van Dorpe, F. and Devictor, N. (2006). Response surfaces and sensitivity analyses for an environmental model of dose calculations. *Reliability Engineering and System Safety*, 91:1241-1251.

Janon, A., Klein, T., Lagnoux-Renaudie, A., Nodet, M. and Prieur, C. (2012). Asymptotic normality and efficiency of two Sobol index estimators. HAL e-prints, http://hal.inria.fr/hal-00665048.

Kennedy, M. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society*, 63(3):425–464.

Kerkyacharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using haar wavelets. *The Annals of Statistics*, 24:485–507.

Laurent, B. (1996). Efficient estimation of integral functionals of a density. *The Annals of Statistics*, 24:659–681.

Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM: Probability and Statistics*, 9:1–19.

Leonenko N. and Seleznjev O. (2010). Statistical inference for the ϵ-entropy and the quadratic Rnyi entropy. *Journal of Multivariate Analysis*, 101:1981–1994.

Levit, B.Y. (1978). Asymptotically efficient estimation of nonlinear functionals. *Problems Inform. Transmission*, 14:204–209.

Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86:316–327.

Liu, H., Chen, W. and Sudjianto, A. (2006). Relative entropy based method for probabilistic sensitivity analysis in engineering design. *Journal of Mechanical Design*, 128(2):326–336.

Loubes, J.-M. and Marteau, C. and Solis, M. and Da Veiga, S. (2011). Efficient estimation of conditional covariance matrices for dimension reduction. ArXiv e-prints, http://adsabs.harvard.edu/abs/2011arXiv1110.3238L.

Manceau, E., Mezghani, M., Zabalza-Mezghani, I., and Roggero, F. (2001). Combination of experimental design and joint modeling methods for quantifying the risk associated with deterministic and stochastic uncertainties - An integrated test study. *2001 SPE Annual Technical Conference and Exhibition, New Orleans, 30 September-3 October*,

paper SPE 71620.

McKay, M.D. (1995). Evaluating prediction uncertainty. Tech. Rep. NUREG/CR-6311, U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory.

Oakley, J.E. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models : a bayesian approach. *Journal of the Royal Statistical Society Series B*, 66:751–769.

Owen, A.B. (1994). Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays. *The Annals of Statistics*, 22:930–945.

Saltelli, A., Chan, K., and Scott, E., editors (2000). *Sensitivity analysis*. Wiley Series in Probability and Statistics. Wiley.

Santner T., Williams B. and Notz W. (2003). The design and analysis of computer experiments. New York: Springer Verlag.

Sobol', I M. (1993). Sensitivity estimates for nonlinear mathematical models. *MMCE*, 1:407–414.

Turanyi, T. (1990). Sensitivity analysis of complex kinetic systems. *Journal of Mathematical Chemistry*, 5:203–248.

Van Der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Wand, M. and Jones, M. (1994). *Kernel Smoothing* London: Chapman and Hall.

Appendix A. Proofs of Theorems

A.1. *Proof of Lemma 3.1*

Let $\hat{\theta}_n = \hat{\theta}_n^1 - \hat{\theta}_n^2$ where

$$\hat{\theta}_n^1 = \frac{2}{n(n-1)} \sum_{i \in M} \sum_{j \neq k=1}^{n} p_i(X_j, Y_j) \int p_i(X_k, u) \eta(X_k, u, Y_k) du$$

and

$$\hat{\theta}_n^2 = \frac{1}{n(n-1)} \sum_{i,i' \in M} \sum_{j \neq k=1}^{n} p_i(X_j, Y_j) p_{i'}(X_k, Y_k) \int p_i(x, y_1) p_{i'}(x, y_2) \eta(x, y_1, y_2) dx dy_1 dy_2.$$

Let us first compute $\mathbb{E}(\hat{\theta}_n^1)$:

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_n^1) &= 2\sum_{i\in M}\iint p_i(x,y)f(x,y)dxdy \iiint p_i(x,y)\eta(x,u,y)f(x,y)dxdydu \\
&= 2\sum_{i\in M} a_i \iiint p_i(x,y)\eta(x,u,y)f(x,y)dxdydu \\
&= 2\iiint \left(\sum_{i\in M} a_i p_i(x,y)\right)\eta(x,u,y)f(x,y)dxdydu \\
&= 2\iiint S_M f(x,y)\eta(x,u,y)f(x,y)dxdydu.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_n^2) &= \sum_{i,i'\in M}\iint p_i(x,y)f(x,y)dxdy \iint p_{i'}(x,y)f(x,y)dxdy \\
&\quad \int p_i(x,y_1)p_{i'}(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2 \\
&= \sum_{i,i'\in M} a_i a_{i'} \int p_i(x,y_1)p_{i'}(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2 \\
&= \int \left(\sum_{i\in M} a_i p_i(x,y_1)\right)\left(\sum_{i'\in M} a_{i'} p_{i'}(x,y_2)\right)\eta(x,y_1,y_2)dxdy_1dy_2 \\
&= \int S_M f(x,y_1) S_M f(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2.
\end{aligned}$$

Finally, $\mathbb{E}(\hat{\theta}_n) - \theta = \mathbb{E}(\hat{\theta}_n^1) - \mathbb{E}(\hat{\theta}_n^2) - \theta$ and we get the desired bias with (6).

A.2. *Proof of Theorem 3.2*

We will write M instead of M_n for readability and denote $m = |M|$. We want to bound the precision of $\hat{\theta}_n$. We first write

$$\mathbb{E}\left(\hat{\theta}_n - \iiint \eta(x,y_1,y_2)f(x,y_1)f(x,y_2)dxdy_1dy_2\right)^2 = \mathrm{Bias}^2(\hat{\theta}_n) + \mathrm{Var}(\hat{\theta}_n).$$

The first term of this decomposition can be easily bounded, since $\hat{\theta}_n$ has been built to achieve a bias equal to

$$\begin{aligned}
\mathrm{Bias}(\hat{\theta}_n) = -\iiint &[S_M f(x,y_1) - f(x,y_1)][S_M f(x,y_2) - f(x,y_2)] \\
&\eta(x,y_1,y_2)dxdy_1dy_2.
\end{aligned}$$

We then get the following lemma :

Lemma A.1: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$|Bias(\hat{\theta}_n)| \leq \Delta_Y \|\eta\|_\infty \sup_{i \notin M} |c_i|^2.$$

Proof:

$$\begin{aligned}
|\text{Bias}(\hat{\theta}_n)| &\leq \|\eta\|_\infty \int \left(\int |S_M f(x,y_1) - f(x,y_1)| dy_1 \right) \\
&\qquad \left(\int |S_M f(x,y_2) - f(x,y_2)| dy_2 \right) dx \\
&\leq \|\eta\|_\infty \int \left(\int |S_M f(x,y) - f(x,y)| dy \right)^2 dx \\
&\leq \Delta_Y \|\eta\|_\infty \iint (S_M f(x,y) - f(x,y))^2 dxdy \\
&\leq \Delta_Y \|\eta\|_\infty \sum_{i \notin M} |a_i|^2 \leq \Delta_Y \|\eta\|_\infty \sup_{i \notin M} |c_i|^2.
\end{aligned}$$

Indeed, $f \in \mathcal{E}$ and the last inequality follows from Hölder inequality. □

Bounding the variance of $\hat{\theta}_n$ is however less straightforward. Let A and B be the $m \times 1$ vectors with components

$$\begin{aligned}
a_i &:= \iint f(x,y) p_i(x,y) dxdy \quad i = 1, \ldots, m \\
b_i &:= \iiint p_i(x,y_1) f(x,y_2) \eta(x,y_1,y_2) dxdy_1 dy_2 \\
&= \iint g(x,y) p_i(x,y) dxdy \quad i = 1, \ldots, m
\end{aligned}$$

where $g(x,y) = \int f(x,u)\eta(x,y,u)du$ for each $i \in M$. a_i et b_i are the components of f and g onto the ith component of the basis. Let Q and R be the $m \times 1$ vectors of the centered functions $q_i(x,y) = p_i(x,y) - a_i$ and $r_i(x,y) = \int p_i(x,u)\eta(x,u,y)du - b_i$ for $i = 1, \ldots, m$. Let C be the $m \times m$ matrix of constants $c_{ii'} = \iiint p_i(x,y_1) p_{i'}(x,y_2) \eta(x,y_1,y_2) dxdy_1 dy_2$ for $i, i' = 1, \ldots, m$. Take care that here $c_{ii'}$ is double subscript unlike in the (c_i) sequence appearing in the definition of the ellipsoid $\mathcal{E}$. We denote by U_n the process $U_n h = \dfrac{1}{n(n-1)} \sum_{j \neq k=1}^{n} h(X_j, Y_j, X_k, Y_k)$ and by P_n the empirical measure $P_n f = \dfrac{1}{n} \sum_{j=1}^{n} f(X_j, Y_j)$. With the previous notation, $\hat{\theta}_n$ has the following Hoeffding's decomposition (see chapter 11 of Van Der Vaart (1998)):

$$\hat{\theta}_n = U_n K + P_n L + 2^t AB - {}^t ACA \tag{A1}$$

where

$$K(x_1,y_1,x_2,y_2) = 2^t Q(x_1,y_1)R(x_2,y_2) - {}^tQ(x_1,y_1)CQ(x_2,y_2),$$
$$L(x_1,y_1) = 2^t AR(x_1,y_1) + 2^t BQ(x_1,y_1) - 2^t ACQ(x_1,y_1).$$

Then $\mathrm{Var}(\hat{\theta}_n) = \mathrm{Var}(U_nK) + \mathrm{Var}(P_nL) + 2\ \mathrm{Cov}(U_nK, P_nL)$. We have to get bounds for each of these terms : they are given in the three following lemmas.

Lemma A.2: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$Var(U_nK) \leq \frac{20}{n(n-1)}\|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_Y^2 (m+1).$$

Proof: Since U_nK is centered, $\mathrm{Var}(U_nK)$ equals

$$\mathbb{E}\left(\frac{1}{(n(n-1))^2}\sum_{j\neq k=1}^{n}\sum_{j'\neq k'=1}^{n} K(X_j,Y_j,X_k,Y_k)K(X_{j'},Y_{j'},X_{k'},Y_{k'})\right)$$
$$= \frac{1}{n(n-1)}\mathbb{E}(K^2(X_1,Y_1,X_2,Y_2) + K(X_1,Y_1,X_2,Y_2)K(X_2,Y_2,X_1,Y_1)).$$

By the Cauchy-Schwarz inequality,

$$\mathrm{Var}(U_nK) \leq \frac{2}{n(n-1)}\mathbb{E}(K^2(X_1,Y_1,X_2,Y_2)).$$

Moreover, the inequality $2|\mathbb{E}(XY)| \leq \mathbb{E}(X^2) + \mathbb{E}(Y^2)$ leads to

$$\mathbb{E}(K^2(X_1,Y_1,X_2,Y_2)) \leq 2\left[\mathbb{E}\left((2Q'(X_1,Y_1)R(X_2,Y_2))^2\right)\right.$$
$$\left.+\mathbb{E}\left((Q'(X_1,Y_1)CQ(X_2,Y_2))^2\right)\right].$$

We have to bound these two terms. The first one is

$$\mathbb{E}\left((2Q'(X_1,Y_1)R(X_2,Y_2))^2\right) = 4(W_1 - W_2 - W_3 + W_4)$$

where

$$W_1 = \iiiiint \sum_{i,i'} p_i(x,y)p_{i'}(x,y)p_i(x',u)p_{i'}(x',v\eta(x',u,y')\eta(x',v,y') \\ f(x,y)f(x',y')dudvdxdydx'dy'$$

$$W_2 = \iint \sum_{i,i'} b_i b_{i'} p_i(x,y)p_{i'}(x,y)f(x,y)dxdy$$

$$W_3 = \iiint \sum_{i,i'} a_i a_{i'} p_i(x,u)p_{i'}(x,v)\eta(x,u,y)\eta(x,v,y)f(x,y)dxdy$$

$$W_4 = \sum_{i,i'} a_i a_{i'} b_i b_{i'}.$$

Straightforward manipulations show that $W_2 \geq 0$ and $W_3 \geq 0$. This implies that

$$\mathbb{E}\left((2Q'(X_1,Y_1)R(X_2,Y_2))^2\right) \leq 4(W_1 + W_4).$$

On the one hand,

$$\begin{aligned}
W_1 &= \iiiint \sum_{i,i'} p_i(x,y)p_{i'}(x,y) \int p_i(x',u)\eta(x',u,y')du \int p_{i'}(x',v)\eta(x',v,y')dv f(x,y)f(x',y')dxdydx'dy' \\
&\leq \iiiint \left(\sum_i p_i(x,y) \int p_i(x',u)\eta(x',u,y')du\right)^2 f(x,y)f(x',y')dxdydx'dy' \\
&\leq \|f\|_\infty^2 \iiiint \left(\sum_i p_i(x,y) \int p_i(x',u)\eta(x',u,y')du\right)^2 dxdydx'dy' \\
&\leq \|f\|_\infty^2 \iiiint \sum_{i,i'} p_i(x,y)p_{i'}(x,y) \int p_i(x',u)\eta(x',u,y')du \int p_{i'}(x',v)\eta(x',v,y')dvdxdydx'dy' \\
&\leq \|f\|_\infty^2 \sum_{i,i'} \iint p_i(x,y)p_{i'}(x,y)dxdy \iint \left(\int p_i(x',u)\eta(x',u,y')du\right)\left(\int p_{i'}(x',v)\eta(x',v,y')dv\right)dx'dy' \\
&\leq \|f\|_\infty^2 \sum_i \iint \left(\int p_i(x',u)\eta(x',u,y')du\right)^2 dx'dy'
\end{aligned}$$

since the p_i are orthonormal. Moreover,

$$\begin{aligned}
\left(\int p_i(x',u)\eta(x',u,y')du\right)^2 &\leq \left(\int p_i(x',u)^2du\right)\left(\int \eta(x',u,y')^2du\right) \\
&\leq \|\eta\|_\infty^2 \Delta_Y \int p_i(x',u)^2du,
\end{aligned}$$

and then

$$\iint \left(\int p_i(x',u)\eta(x',u,y')du \right)^2 dx'dy' \leq \|\eta\|_\infty^2 \Delta_Y^2 \iint p_i(x',u)^2 dudx'$$
$$\|\eta\|_\infty^2 \Delta_Y^2.$$

Finally,

$$W_1 \leq \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_Y^2 m.$$

On the other hand,

$$W_4 = \left(\sum_i a_i b_i \right)^2 \leq \sum_i a_i^2 \sum_i b_i^2 \leq \|f\|_2^2 \|g\|_2^2 \leq \|f\|_\infty \|g\|_2^2.$$

By the Cauchy-Scharwz inequality we have $\|g\|_2^2 \leq \|\eta\|_\infty^2 \|f\|_\infty \Delta_Y^2$ and then

$$W_4 \leq \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_Y^2$$

which leads to

$$\mathbb{E}\left((2Q'(X_1,Y_1)R(X_2,Y_2))^2 \right) \leq 4\|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_Y^2 (m+1).$$

Let us bound now the second term $\mathbb{E}\left((Q'(X_1,Y_1)CQ(X_2,Y_2))^2 \right) = W_5 - 2W_6 + W_7$ where

$$W_5 = \iiiint \sum_{i,i'} \sum_{i_1,i_1'} c_{ii'} c_{i_1 i_1'} p_i(x,y) p_{i_1}(x,y) p_{i'}(x',y') p_{i_1'}(x',y') f(x,y) f(x',y') dxdydx'dy'$$
$$W_6 = \sum_{i,i'} \sum_{i_1,i_1'} \iint c_{ii'} c_{i_1 i_1'} a_i a_{i_1} p_{i'}(x,y) p_{i_1'}(x,y) f(x,y) dxdy$$
$$W_7 = \sum_{i,i'} \sum_{i_1,i_1'} c_{ii'} c_{i_1 i_1'} a_i a_{i_1} a_{i'} a_{i_1'}.$$

Following the previous manipulations, we show that $W_6 \geq 0$. Thus,

$$\mathbb{E}\left((Q'(X_1,Y_1)CQ(X_2,Y_2))^2 \right) \leq W_5 + W_7.$$

First, observe that

$$\begin{aligned}
W_5 &= \iiiint \left(\sum_{i,i'} c_{ii'} p_i(x,y) p_{i'}(x',y') \right)^2 f(x,y) f(x',y') dx dy dx' dy' \\
&\leq \|f\|_\infty^2 \iiiint \left(\sum_{i,i'} c_{ii'} p_i(x,y) p_{i'}(x',y') \right)^2 dx dy dx' dy' \\
&\leq \|f\|_\infty^2 \sum_{i,i'} \sum_{i_1,i'_1} c_{ii'} c_{i_1 i'_1} \iiiint p_i(x,y) p_{i_1}(x,y) \\
&\quad p_{i'}(x',y') p_{i'_1}(x',y') dx dy dx' dy' \\
&\leq \|f\|_\infty^2 \sum_{i,i'} c_{ii'}^2
\end{aligned}$$

since the p_i are orthonormal. Besides,

$$\begin{aligned}
\sum_{i,i'} c_{ii'}^2 &= \iint \sum_{i_\alpha, i'_\alpha} \alpha_{i_\alpha}(x) \alpha_{i'_\alpha}(x) \alpha_{i_\alpha}(x') \alpha_{i'_\alpha}(x') \sum_{i_\beta, i'_\beta} \left(\iint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x, y_1, y_2) dy_1 dy_2 \right) \\
&\quad \left(\iint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x', y_1, y_2) dy_1 dy_2 \right) dx dx' \\
&= \iint \left(\sum_{i_\alpha} \alpha_{i_\alpha}(x) \alpha_{i_\alpha}(x') \right)^2 \sum_{i_\beta, i'_\beta} \left(\iint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x, y_1, y_2) dy_1 dy_2 \right) \\
&\quad \left(\iint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x', y_1, y_2) dy_1 dy_2 \right) dx dx'.
\end{aligned}$$

But

$$\begin{aligned}
&\sum_{i_\beta, i'_\beta} \left(\iint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x, y_1, y_2) dy_1 dy_2 \right) \\
&\left(\iint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x', y_1, y_2) dy_1 dy_2 \right) \\
&= \sum_{i_\beta, i'_\beta} \iiiint \beta_{i_\beta}(y_1) \beta_{i'_\beta}(y_2) \eta(x, y_1, y_2) \beta_{i_\beta}(y'_1) \beta_{i'_\beta}(y'_2) \eta(x', y'_1, y'_2) dy_1 dy_2 dy'_1 dy'_2 \\
&= \iint \sum_{i_\beta} \left(\int \beta_{i_\beta}(y_1) \eta(x, y_1, y_2) dy_1 \right) \beta_{i_\beta}(y'_1) \sum_{i'_\beta} \left(\int \beta_{i'_\beta}(y'_2) \eta(x', y'_1, y'_2) dy'_2 \right) \beta_{i'_\beta}(y_2) dy'_1 dy_2 \\
&= \iint \eta(x, y'_1, y_2) \eta(x', y'_1, y_2) dy'_1 dy_2 \\
&\leq \Delta_Y^2 \|\eta\|_\infty^2
\end{aligned}$$

using the fact that (β_i) is an orthonormal basis. We then get

$$\begin{aligned}
\sum_{i,i'} c_{ii'}^2 &\leq \Delta_Y^2 \|\eta\|_\infty^2 \iint \left(\sum_{i_\alpha} \alpha_{i_\alpha}(x)\alpha_{i_\alpha}(x') \right)^2 dxdx' \\
&\leq \Delta_Y^2 \|\eta\|_\infty^2 \iint \sum_{i_\alpha, i'_\alpha} \alpha_{i_\alpha}(x)\alpha_{i'_\alpha}(x)\alpha_{i_\alpha}(x')\alpha_{i'_\alpha}(x')dxdx' \\
&\leq \Delta_Y^2 \|\eta\|_\infty^2 \sum_{i_\alpha, i'_\alpha} \left(\int \alpha_{i_\alpha}(x)\alpha_{i'_\alpha}(x)dx \right)^2 \\
&\leq \Delta_Y^2 \|\eta\|_\infty^2 \sum_{i_\alpha} \left(\int \alpha_{i_\alpha}(x)^2 dx \right)^2 \\
&\leq \Delta_Y^2 \|\eta\|_\infty^2 m
\end{aligned}$$

since the α_i are orthonormal. Finally,

$$W_5 \leq \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_Y^2 m.$$

Besides,

$$W_7 = \left(\sum_{i,i'} c_{ii'} a_i a_{i'} \right)^2$$

with

$$\begin{aligned}
\left| \sum_{i,i'} c_{ii'} a_i a_{i'} \right| &\leq \|\eta\|_\infty \iiint |S_M f(x,y_1) S_M f(x,y_2)| dx dy_1 dy_2 \\
&\leq \|\eta\|_\infty \iint \left(\int |S_M f(x,y_1) S_M f(x,y_2)| dx \right) dy_1 dy_2.
\end{aligned}$$

By using the Cauchy-Schwarz inequality twice, we get

$$\begin{aligned}\left(\sum_{i,i'} c_{ii'}a_i a_{i'}\right)^2 &\leq \Delta_Y^2\|\eta\|_\infty^2 \iint \left(\int |S_M f(x,y_1)S_M f(x,y_2)|dx\right)^2 dy_1 dy_2 \\ &\leq \Delta_Y^2\|\eta\|_\infty^2 \iint \left(\int S_M f(u,y_1)^2 du\right)\left(\int S_M f(v,y_2)^2 dv\right) dy_1 dy_2 \\ &\leq \Delta_Y^2\|\eta\|_\infty^2 \iiiint S_M f(u,y_1)^2 S_M f(v,y_2)^2 du dv dy_1 dy_2 \\ &\leq \Delta_Y^2\|\eta\|_\infty^2 \left(\iint S_M f(x,y)^2 dx dy\right)^2 \\ &\leq \Delta_Y^2\|\eta\|_\infty^2\|f\|_\infty^2.\end{aligned}$$

Finally,

$$\mathbb{E}\left((Q'(X_1,Y_1)CQ(X_2,Y_2))^2\right) \leq \|\eta\|_\infty^2\|f\|_\infty^2\Delta_Y^2(m+1).$$

Collecting this inequalities, we obtain

$$\mathrm{Var}(U_n K) \leq \frac{20}{n(n-1)}\|\eta\|_\infty^2\|f\|_\infty^2\Delta_Y^2(m+1)$$

which concludes the proof of Lemma A.2. □

Let us now deal with the second term of the Hoeffding's decomposition of $\hat{\theta}_n$:

Lemma A.3: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$Var(P_n L) \leq \frac{36}{n}\Delta_Y^2\|f\|_\infty^2\|\eta\|_\infty^2.$$

Proof: First note that

$$\mathrm{Var}(P_n L) = \frac{1}{n}\mathrm{Var}(L(X_1,Y_1)).$$

We can write $L(X_1, Y_1)$ as

$$\begin{aligned}
L(X_1,Y_1) &= 2A'R(X_1,Y_1) + 2B'Q(X_1,Y_1) - 2A'CQ(X_1,Y_1) \\
&= 2\sum_i a_i \left(\int p_i(X_1,u)\eta(X_1,u,Y_1)du - b_i \right) \\
&\quad +2\sum_i b_i(p_i(X_1,Y_1) - a_i) - 2\sum_{i,i'} c_{ii'}a_{i'}(p_i(X_1,Y_1) - a_i) \\
&= 2\int \sum_i a_i p_i(X_1,u)\eta(X_1,u,Y_1)du + 2\sum_i b_i p_i(X_1,Y_1) \\
&\quad -2\sum_{i,i'} c_{ii'}a_{i'}p_i(X_1,Y_1) - 4A'B + 2A'CA \\
&= 2\int S_M f(X_1,u)\eta(X_1,u,Y_1)du + 2S_M g(X_1,Y_1) \\
&\quad -2\sum_{i,i'} c_{ii'}a_{i'}p_i(X_1,Y_1) - 4A'B + 2A'CA.
\end{aligned}$$

Let $h(x,y) = \int S_M f(x,u)\eta(x,u,y)du$, we have

$$\begin{aligned}
S_M h(z,t) &= \sum_i \left(\iint h(x,y)p_i(x,y)dxdy \right) p_i(z,t) \\
&= \sum_i \left(\iiint S_M f(x,u)\eta(x,u,y)p_i(x,y)dudxdy \right) p_i(z,t) \\
&= \sum_{i,i'} \left(\iiint a_{i'}p_{i'}(x,u)\eta(x,u,y)p_i(x,y)dudxdy \right) p_i(z,t) \\
&= \sum_{i,i'} c_{ii'}a_{i'}p_i(z,t)
\end{aligned}$$

and we can write

$$L(X_1,Y_1) = 2h(X_1,Y_1) + 2S_M g(X_1,Y_1) - 2S_M h(X_1,Y_1) - 4A'B + 2A'CA.$$

Thus,

$$\begin{aligned}
\mathrm{Var}(L(X_1,Y_1)) &= 4\mathrm{Var}[h(X_1,Y_1) + S_M g(X_1,Y_1) - S_M h(X_1,Y_1)] \\
&\le 4\mathbb{E}[(h(X_1,Y_1) + S_M g(X_1,Y_1) - S_M h(X_1,Y_1))^2] \\
&\le 12\mathbb{E}[(h(X_1,Y_1))^2 + (S_M g(X_1,Y_1))^2 + (S_M h(X_1,Y_1))^2].
\end{aligned}$$

Each of these three terms has to be bounded :

$$\begin{aligned}
\mathbb{E}((h(X_1,Y_1))^2) &= \iint \left(\int S_M f(x,u)\eta(x,u,y)du \right)^2 f(x,y)dxdy \\
&\leq \Delta_Y \iiint S_M f(x,u)^2 \eta(x,u,y)^2 f(x,y)dxdydu \\
&\leq \Delta_Y^2 \|f\|_\infty \|\eta\|_\infty^2 \iint S_M f(x,u)^2 dxdu \\
&\leq \Delta_Y^2 \|f\|_\infty \|\eta\|_\infty^2 \|S_M f\|_2^2 \\
&\leq \Delta_Y^2 \|f\|_\infty \|\eta\|_\infty^2 \|f\|_2^2 \\
&\leq \Delta_Y^2 \|f\|_\infty^2 \|\eta\|_\infty^2
\end{aligned}$$

$$\mathbb{E}((S_M g(X_1,Y_1))^2) \leq \|f\|_\infty \|S_M g\|_2^2 \leq \|f\|_\infty \|g\|_2^2 \leq \Delta_Y^2 \|f\|_\infty^2 \|\eta\|_\infty^2$$

$$\mathbb{E}((S_M h(X_1,Y_1))^2) \leq \|f\|_\infty \|S_M h\|_2^2 \leq \|f\|_\infty \|h\|_2^2 \leq \Delta_Y^2 \|f\|_\infty^2 \|\eta\|_\infty^2$$

from previous calculations. Finally,

$$\mathrm{Var}(L(X_1,Y_1)) \leq 36\Delta_Y^2 \|f\|_\infty^2 \|\eta\|_\infty^2.$$

□

The last term of the Hoeffding's decomposition can also be controled :

Lemma A.4: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$Cov(U_n K, P_n L) = 0.$$

Proof : Since $U_n K$ et $P_n L$ are centered, we have

$$\begin{aligned}
\mathrm{Cov}(U_n K, P_n L) &= \mathbb{E}(U_n K P_n L) \\
&= \mathbb{E}\left[\frac{1}{n^2(n-1)} \sum_{j \neq k=1}^{n} K(X_j,Y_j,X_k,Y_k) \sum_{i=1}^{n} L(X_i,Y_i) \right] \\
&= \frac{1}{n} \mathbb{E}(K(X_1,Y_1,X_2,Y_2)(L(X_1,Y_1)+L(X_2,Y_2))) \\
&= 0
\end{aligned}$$

since K, L, Q and R are centered. □

The four previous lemmas give the expected result on the precision of $\hat{\theta}_n$:

Lemma A.5: *Assuming the hypotheses of Theorem 3.2 hold, we have :*

- *If* $m/n \to 0$,

$$\mathbb{E}(\hat{\theta}_n - \theta)^2 = O\left(\frac{1}{n}\right),$$

- *Otherwise,*

$$\mathbb{E}(\hat{\theta}_n - \theta)^2 \leq \gamma_2(m/n^2)$$

where γ_2 *only depends on* $\|f\|_\infty$, $\|\eta\|_\infty$ *and* Δ_Y.

Proof : Lemmas A.2, A.3 and A.4 imply

$$\mathrm{Var}(\hat{\theta}_n) \leq \frac{20}{n(n-1)}\Delta_Y^2\|f\|_\infty^2\|\eta\|_\infty^2(m+1) + \frac{36}{n}\Delta_Y^2\|f\|_\infty^2\|\eta\|_\infty^2.$$

Finally, for n large enough and a constant $\gamma \in \mathbb{R}$,

$$\mathrm{Var}(\hat{\theta}_n) \leq \gamma\Delta_Y^2\|f\|_\infty^2\|\eta\|_\infty^2\left(\frac{m}{n^2} + \frac{1}{n}\right).$$

Lemma A.1 gives

$$\mathrm{Bias}^2(\hat{\theta}_n) \leq \Delta_Y^2\|\eta\|_\infty^2\left(\sup_{i\notin M}|c_i|^2\right)^2$$

and by assumption $\left(\sup_{i\notin M}|c_i|^2\right)^2 \approx m/n^2$. If $m/n \to 0$, then $\mathbb{E}(\hat{\theta}_n - \theta)^2 = O(\frac{1}{n})$. Otherwise $\mathbb{E}(\hat{\theta}_n - \theta)^2 \leq \gamma_2(m/n^2)$ where γ_2 only depends on $\|f\|_\infty$, $\|\eta\|_\infty$ and Δ_Y. □

The lemma we just proved gives the result of Theorem 3.2 when m/n does not converge to 0. Let us now study more precisely the semiparametric case, that is when $\mathbb{E}(\hat{\theta}_n - \theta)^2 = O(\frac{1}{n})$, to prove the asymptotic normality (8) and the bound in (9). We have

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \sqrt{n}(U_nK) + \sqrt{n}(P_nL) + \sqrt{n}(2A'B - A'CA).$$

We will study the asymptotic behavior of each of these three terms. The first one is easily treated :

Lemma A.6: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$\sqrt{n}U_nK \to 0$$

in probability when $n \to \infty$ *if* $m/n \to 0$.

Proof : Since $\mathrm{Var}(\sqrt{n}U_nK) \leq \frac{20}{(n-1)}\|\eta\|_\infty^2\|f\|_\infty^2\Delta_Y^2(m+1)$, $\sqrt{n}U_nK$ converges to 0 in probability when $n \to \infty$ if $m/n \to 0$. □

The random variable P_nL will be the most important term for the central limit theorem. Before studying its asymptotic normality, we need the following lemma concerning the asymptotic variance of $\sqrt{n}(P_nL)$:

Lemma A.7: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$n\,Var(P_nL) \to \Lambda(f,\eta)$$

where

$$\Lambda(f,\eta) = 4\left[\iint g(x,y)^2 f(x,y)dxdy - \left(\iint g(x,y)f(x,y)dxdy\right)^2\right].$$

Proof: We proved in Lemma A.3 that

$$\begin{aligned}\mathrm{Var}(L(X_1,Y_1)) &= 4\mathrm{Var}[h(X_1,Y_1) + S_Mg(X_1,Y_1) - S_Mh(X_1,Y_1)]\\ &= 4\mathrm{Var}[A_1+A_2+A_3]\\ &= 4\sum_{i,j=1}^{3}\mathrm{Cov}(A_i,A_j).\end{aligned}$$

We will show that $\forall i,j \in \{1,2,3\}^2$, we have

$$\left|\mathrm{Cov}(A_i,A_j) - \epsilon_{ij}\left[\iint g(x,y)^2 f(x,y)dxdy - \left(\iint g(x,y)f(x,y)dxdy\right)^2\right]\right| \leq \gamma\left[\|S_Mf - f\|_2 + \|S_Mg - g\|_2\right] \quad \text{(A2)}$$

where $\epsilon_{ij} = -1$ if $i = 3$ or $j = 3$ and $i \neq j$ and $\epsilon_{ij} = 1$ otherwise, and where γ depends only on $\|f\|_\infty$, $\|\eta\|_\infty$ and Δ_Y.
We shall give the details only for the case $i = j = 3$ since the calculations are similar for the other configurations. We have

$$\mathrm{Var}(A_3) = \iint S_M^2[h(x,y)]f(x,y)dxdy - \left(\iint S_M[h(x,y)]f(x,y)dxdy\right)^2$$

We first study the quantity

$$\left|\iint S_M^2[h(x,y)]f(x,y)dxdy - \iint g(x,y)^2 f(x,y)dxdy\right|.$$

It is bounded by prout prout prout prout prout prout prout prout prout prout

$$\begin{aligned}&\iint \left|S_M^2[h(x,y)]f(x,y) - S_M^2[g(x,y)]f(x,y)\right|dxdy\\ &+ \iint \left|S_M^2[g(x,y)]f(x,y) - g(x,y)^2 f(x,y)\right|dxdy\\ &\leq \|f\|_\infty\|S_Mh + S_Mg\|_2\|S_Mh - S_Mg\|_2 + \|f\|_\infty\|S_Mg + g\|_2\|S_Mg - g\|_2.\end{aligned}$$

Using the fact that S_M is a projection, this sum is bounded by

$$\begin{aligned}&\|f\|_\infty\|h+g\|_2\|h-g\|_2+2\|f\|_\infty\|g\|_2\|S_Mg-g\|_2\\&\leq\|f\|_\infty(\|h\|_2+\|g\|_2)\|h-g\|_2+2\|f\|_\infty\|g\|_2\|S_Mg-g\|_2.\end{aligned}$$

We saw previously that $\|g\|_2\leq\Delta_Y\|f\|_\infty^{1/2}\|\eta\|_\infty$ and $\|h\|_2\leq\Delta_Y\|f\|_\infty^{1/2}\|\eta\|_\infty$. The sum is then bounded by

$$2\Delta_Y\|f\|_\infty^{3/2}\|\eta\|_\infty\|h-g\|_2+2\Delta_Y\|f\|_\infty^{3/2}\|\eta\|_\infty\|S_Mg-g\|_2$$

We now have to deal with $\|h-g\|_2$:

$$\begin{aligned}\|h-g\|_2^2&=\iint\left(\int\left(S_Mf(x,u)-f(x,u)\right)\eta(x,u,y)du\right)^2dxdy\\&\leq\iint\left(\int(S_Mf(x,u)-f(x,u))^2du\right)\left(\int\eta(x,u,y)^2du\right)dxdy\\&\leq\Delta_Y^2\|\eta\|_\infty^2\|S_Mf-f\|_2^2.\end{aligned}$$

Finally, the sum is bounded by

$$2\Delta_Y\|f\|_\infty^{3/2}\|\eta\|_\infty\left(\Delta_Y\|\eta\|_\infty\|S_Mf-f\|_2+\|S_Mg-g\|_2\right).$$

Let us now study the second quantity

$$\left|\left(\iint S_M[h(x,y)]f(x,y)dxdy\right)^2-\left(\iint g(x,y)f(x,y)dxdy\right)^2\right|.$$

It is equal to

$$\begin{aligned}&\left|\left(\iint(S_M[h(x,y)]+g(x,y))f(x,y)dxdy\right)\right.\\&\left.\left(\iint(S_M[h(x,y)]-g(x,y))f(x,y)dxdy\right)\right|.\end{aligned}$$

By using the Cauchy-Schwarz inequality, it is bounded by

$$\begin{aligned}&\|f\|_2\|S_Mh+g\|_2\|f\|_2\|S_Mh-g\|_2\\&\leq\|f\|_2^2(\|h\|_2+\|g\|_2)(\|S_Mh-S_Mg\|_2+\|S_Mg-g\|_2)\\&\leq2\Delta_Y\|f\|_\infty^{3/2}\|\eta\|_\infty(\|h-g\|_2+\|S_Mg-g\|_2)\\&\leq2\Delta_Y\|f\|_\infty^{3/2}\|\eta\|_\infty\left(\Delta_Y\|\eta\|_\infty\|S_Mf-f\|_2+\|S_Mg-g\|_2\right)\end{aligned}$$

by using the previous calculations. Collecting the two inequalities gives (A2) for $i=j=3$. Finally, since by assumption $\forall t\in L^2(d\mu)$, $\|S_Mt-t\|_2\to0$ when $n\to\infty$, a direct

consequence of (A2) is that

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{Var}(L(X_1, Y_1)) \\
&= 4\left[\iint g(x,y)^2 f(x,y)dxdy - \left(\iint g(x,y)f(x,y)dxdy\right)^2\right] \\
&= \Lambda(f,\eta).
\end{aligned}$$

We then conclude by noting that $\mathrm{Var}(\sqrt{n}(P_nL)) = \mathrm{Var}(L(X_1, Y_1))$. □

We can now study the convergence of $\sqrt{n}(P_nL)$, which is given in the following lemma:

Lemma A.8: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$\sqrt{n}P_nL \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda(f,\eta)).$$

Proof: We first note that

$$\sqrt{n}\left(P_n(2g) - 2\iint g(x,y)f(x,y)dxdy\right) \to \mathcal{N}(0, \Lambda(f,\eta))$$

where $g(x,y) = \int \eta(x,y,u)f(x,u)du$.
It is then sufficient to show that the expectation of the square of

$$R = \sqrt{n}\left[P_nL - \left(P_n(2g) - 2\iint g(x,y)f(x,y)dxdy\right)\right]$$

converges to 0. We have

$$\begin{aligned}
\mathbb{E}(R^2) &= \mathrm{Var}(R) \\
&= n\mathrm{Var}(P_nL) + n\mathrm{Var}(P_n(2g)) - 2n\mathrm{Cov}(P_nL, P_n(2g))
\end{aligned}$$

We know that $n\mathrm{Var}(P_n(2g)) \to \Lambda(f,\eta)$ and Lemma A.7 shows that $n\mathrm{Var}(P_nL) \to \Lambda(f,\eta)$. Then, we just have to prove that

$$\lim_{n\to\infty} n\mathrm{Cov}(P_nL, P_n(2g)) = \Lambda(f,\eta).$$

We have

$$n\mathrm{Cov}(P_nL, P_n(2g)) = \mathbb{E}(2L(X_1, Y_1)g(X_1, Y_1))$$

because L is centered. Since

$$L(X_1, Y_1) = 2h(X_1, Y_1) + 2S_Mg(X_1, Y_1) - 2S_Mh(X_1, Y_1) - 4A'B + 2A'CA,$$

we get

$$n\mathrm{Cov}(P_nL, P_n(2g)) = 4\iint h(x,y)g(x,y)f(x,y)dxdy$$
$$+4\iint S_M g(x,y)g(x,y)f(x,y)dxdy$$
$$-4\iint S_M h(x,y)g(x,y)f(x,y)dxdy - 8\sum_i a_i b_i \iint g(x,y)f(x,y)dxdy$$
$$+4A'CA\iint g(x,y)f(x,y)dxdy$$

which converges to $4\left[\iint g(x,y)^2 f(x,y)dxdy - \left(\iint g(x,y)f(x,y)dudxdy\right)^2\right]$ which is equal to $\Lambda(f,\eta)$. We finally deduce that

$$\sqrt{n}P_nL \to \mathcal{N}(0, \Lambda(f,\eta))$$

in distribution. □

In order to prove the asymptotic normality of $\hat{\theta}_n$, the last step is to control the remainder term in the Hoeffding's decomposition:

Lemma A.9: *Assuming the hypotheses of Theorem 3.2 hold, we have*

$$\sqrt{n}(2A'B - A'CA - \theta) \to 0.$$

Proof: $\sqrt{n}(2A'B - A'CA - \theta) \to 0$ is equal to

$$\sqrt{n}\left[2\iint g(x,y)S_M f(x,y)dxdy\right.$$
$$-\iiint S_M f(x,y_1)S_M f(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2$$
$$\left.-\iiint f(x,y_1)f(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2\right].$$

By replacing g we get

$$\sqrt{n}\left[2\iiint S_M f(x,y_1)f(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2\right.$$
$$-\iiint S_M f(x,y_1)S_M f(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2$$
$$\left.-\iiint f(x,y_1)f(x,y_2)\eta(x,y_1,y_2)dxdy_1dy_2\right]$$

With integral manipulation, we show it is also equal to

$$\begin{aligned}
&\sqrt{n}\left[\iiint S_M f(x,y_1)(f(x,y_2)-S_M f(x,y_2))\eta(x,y_1,y_2)dxdy_1dy_2\right.\\
&\left.-\iiint f(x,y_2)(S_M f(x,y_1)-f(x,y_1))\eta(x,y_1,y_2)dxdy_1dy_2\right]\\
&\leq \sqrt{n}\Delta_Y\|\eta\|_\infty\left(\|S_M f\|_2\|S_M f-f\|_2+\|f\|_2\|S_M f-f\|_2\right)\\
&\leq 2\sqrt{n}\Delta_Y\|f\|_2\|\|\eta\|_\infty\|S_M f-f\|_2\\
&\leq 2\sqrt{n}\Delta_Y\|f\|_2\|\|\eta\|_\infty\left(\sup_{i\notin M}|c_i|^2\right)^{1/2}\\
&\approx 2\Delta_Y\|f\|_2\|\|\eta\|_\infty\sqrt{\frac{m}{n}},
\end{aligned}$$

which converges to 0 when $n\to\infty$ since $m/n\to 0$. □

Collecting now the results of Lemmas A.6, A.7 and A.9 we get (8) since

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right)\to\mathcal{N}(0,\Lambda(f,\eta))$$

in distribution. We finally have to prove (9). Remark that

$$\begin{aligned}
n\mathbb{E}\left(\hat{\theta}_n-\theta\right)^2 &= n\mathrm{Bias}^2(\hat{\theta}_n)+n\mathrm{Var}(\hat{\theta}_n)\\
&= n\mathrm{Bias}^2(\hat{\theta}_n)+n\mathrm{Var}(U_nK)+n\mathrm{Var}(P_nL)
\end{aligned}$$

We previously proved that

$$\begin{aligned}
n\mathrm{Bias}^2(\hat{\theta}_n)&\leq\lambda\Delta_Y^2\|\eta\|_\infty^2\frac{m}{n}\ \text{ for some }\lambda\in\mathbb{R},\\
n\mathrm{Var}(U_nK)&\leq\mu\Delta_Y^2\|f\|_\infty^2\|\eta\|_\infty^2\frac{m}{n}\ \text{ for some }\mu\in\mathbb{R}.
\end{aligned}$$

Moreover, (A2) imply

$$|n\mathrm{Var}(P_nL)-\Lambda(f,\eta)|\leq\gamma\left[\|S_Mf-f\|_2+\|S_Mg-g\|_2\right],$$

where γ is a increasing function of $\|f\|_\infty$,$\|\eta\|_\infty$ and Δ_Y. We then deduce (9) which ends the proof of Theorem 3.2.

A.3. *Proof of Theorem 3.3*

To prove the inequality we will use the work of Ibragimov and Khasḿinskii (1991) (see also chapter 25 of Van Der Vaart (1998)) on efficient estimation. The first step is the computation of the Fréchet derivative of $\theta(f)$ at a point f_0. Straightforward calculations

show that

$$\theta(f)-\theta(f_0)=\iint\left[2\int\psi(x,y,z)f_0(x,z)dz\right](f(x,y)-f_0(x,y))\,dxdy + O\left(\iint(f(x,y)-f_0(x,y))^2dxdy\right)$$

from which we deduce that the Fréchet derivative of $\theta(f)$ at f_0 is

$$\theta'(f_0)\cdot u=\left\langle 2\int\psi(x,y,z)f_0(x,z)dz,u\right\rangle \quad (u\in L^2(dxdy)),$$

where $\langle\cdot,\cdot\rangle$ is the scalar product in $L^2(dxdy)$. We can now use the results of Ibragimov and Khasḿinskii (1991). Denote $H(f_0)=H(f_0)=\left\{u\in L^2(dxdy),\iint u(x,y)\sqrt{f_0(x,y)}dxdy=0\right\}$ the set of functions in $L^2(dxdy)$ orthogonal to $\sqrt{f_0}$, $\mathrm{Proj}_{H(f_0)}$ the projection on $H(f_0)$, $A_n(t)=(\sqrt{f_0})t/\sqrt{n}$ and $P_{f_0}^{(n)}$ the joint distribution of $(X_1,\ldots,X_n)$ under f_0. Since here $X_1,\ldots,X_n$ are i.i.d., $\left\{P_f^{(n)},f\in\mathcal{E}\right\}$ is locally asymptotically normal at all points $f_0\in\mathcal{E}$ in the direction $H(f_0)$ with normalizing factor $A_n(f_0)$. Ibragimov and Khas'minskii result say that under these conditions, denoting $K_n=B_n\theta'(f_0)A_n\mathrm{Proj}_{H(f_0)}$ with $B_n(u)=\sqrt{n}u$, if $K_n\to K$ weakly and if $K(u)=\langle t,u\rangle$, then for every estimator $\hat\theta_n$ of $\theta(f)$ and every family $\mathcal{V}(f_0)$ of vicinities of f_0, we have

$$\inf_{\{\mathcal{V}(f_0)\}}\liminf_{n\to\infty}\sup_{f\in\mathcal{V}(f_0)}n\mathbb{E}(\hat\theta_n-\theta(f_0))^2\geq\|t\|^2_{L^2(dxdy)}.$$

Here,

$$K_n(u)=\sqrt{n}\theta'(f_0)\cdot\frac{1}{\sqrt{n}}\sqrt{f_0}\mathrm{Proj}_{H(f_0)}(u)=\theta'(f_0)\cdot\left(\sqrt{f_0}\left(u-\sqrt{f_0}\int u\sqrt{f}_0\right)\right)$$

does not depend on n and

$$\begin{aligned}K(u)&=\iint\left[2\int\psi(x,y,z)f_0(x,z)dz\right]\sqrt{f_0(x,y)}\\&\quad\left(u(x,y)-\sqrt{f_0(x,y)}\int u\sqrt{f}_0\right)dxdy\\&=\iint\left[2\int\psi(x,y,z)f_0(x,z)dz\right]\sqrt{f_0(x,y)}u(x,y)dxdy\\&\quad-\iint\left[2\int\psi(x,y,z)f_0(x,z)dz\right]f_0(x,y)dxdy\int u\sqrt{f}_0\\&=\langle t,u\rangle\end{aligned}$$

where

$$t(x,y) = \left[2\int \psi(x,y,z) f_0(x,z)dz\right] \sqrt{f_0(x,y)} - \left(\iint \left[2\int \psi(x,y,z) f_0(x,z)dz\right] f_0(x,y)dxdy\right) \sqrt{f_0(x,y)}.$$

The semiparametric Cramér-Rao bound for our problem is $\|t\|^2_{L^2(dxdy)}$:

$$\begin{aligned}\|t\|^2_{L^2(dxdy)} &= 4\iint \left[\int \psi(x,y,z) f_0(x,z)dz\right]^2 f_0(x,y)dxdy \\ &\quad -4\left(\iint \left[\int \psi(x,y,z) f_0(x,z)dz\right] f_0(x,y)dxdy\right)^2 \\ &= 4\iint g_0(x,y)^2 f_0(x,y)dxdy - 4\left(\iint g_0(x,y) f_0(x,y)\right)^2\end{aligned}$$

where $g_0(x,y) = \int \psi(x,y,z) f_0(x,z)dz$. Finally, we recognize the expression of $\Lambda(f_0,\psi)$ given in Theorem 3.2.

A.4. *Proof of Theorem 3.4*

We will first control the remainder term Γ_n :

$$\Gamma_n = \frac{1}{6}F'''(\xi)(1-\xi)^3.$$

Let us recall that

$$\begin{aligned}F'''(\xi) = \iiint \frac{\left(\int \hat{f}(x,y)dy\right)^2}{\left(\int \xi f(x,y) + (1-\xi)\hat{f}(x,y)dy\right)^5} \\ \big[\big(\hat{m}(x)-\varphi(y)\big)\big(\hat{m}(x)-\varphi(z)\big)\big(\hat{m}(x)-\varphi(t)\big) \\ \left(\int \hat{f}(x,y)dy\right)\dddot{\psi}\left(\hat{r}(\xi,x)\right) - 3\big(\hat{m}(x)-\varphi(y)\big)\big(\hat{m}(x)-\varphi(z)\big) \\ \left(\int [\xi f(x,y) + (1-\xi)\hat{f}(x,y)]dy\right)\ddot{\psi}\left(\hat{r}(\xi,x)\right)\Big] \\ \Big(f(x,y)-\hat{f}(x,y)\Big)\Big(f(x,z)-\hat{f}(x,z)\Big) \\ \Big(f(x,t)-\hat{f}(x,t)\Big)dxdydzdt\end{aligned}$$

Assumptions A2 and A3 ensure that the first part of the integrand is bounded by a constant μ :

$$\begin{aligned}\Gamma_n &\leq \frac{1}{6}\mu \iiiint |f(x,y)-\hat{f}(x,y)||f(x,z)-\hat{f}(x,z)| \\ &\quad |f(x,t)-\hat{f}(x,t)|dxdydzdt \\ &\leq \frac{1}{6}\mu \int \left(\int |f(x,y)-\hat{f}(x,y)|dy\right)^3 dx \\ &\leq \frac{1}{6}\mu\Delta_Y^2 \iint |f(x,y)-\hat{f}(x,y)|^3 dxdy\end{aligned}$$

by the Hölder inequality. Then $\mathbb{E}(\Gamma_n^2) = O(\mathbb{E}[(\int |f-\hat{f}|^3)^2]) = O(\mathbb{E}[\|f-\hat{f}\|_3^6])$. Since $\hat{f}$ verifies assumption A2, this quantity has order $O(n_1^{-6\lambda})$. If we further assume that $n_1 \approx n/\log(n)$ and $\lambda > 1/6$, we get $E(\Gamma_n^2) = o(\frac{1}{n})$, which proves that the remainder term Γ_n is negligible. We will now show that $\sqrt{n}\left(\hat{T}_n - T(f)\right)$ and $Z_n = \frac{1}{n_2}\sum_{j=1}^{n_2} H(f,X_j,Y_j) - \iint H(f,x,y)f(x,y)dxdy$ have the same asymptotic behavior. The idea is that we can easily get a central limit theorem for Z_n with asymptotic variance

$$C(f) = \iint H(f,x,y)^2 f(x,y)dxdy - \left(\iint H(f,x,y)f(x,y)dxdy\right)^2,$$

which imply both (10) and (11) (we will show at the end of the proof that $C(f)$ can be expressed such as in the theorem). In order to show that $\sqrt{n}\left(\hat{T}_n - T(f)\right)$ and Z_n have the same asymptotic behavior, we will prove that

$$R = \sqrt{n}\left[\hat{T}_n - T(f) - \left(\frac{1}{n_2}\sum_{j=1}^{n_2} H(f,X_j,Y_j) - \iint H(f,x,y)f(x,y)dxdy\right)\right]$$

has a second-order moment converging to 0. Let us note that $R = R_1 + R_2$ where

$$\begin{aligned}R_1 &= \sqrt{n}\left[\hat{T}_n - T(f)\right. \\ &\quad \left. - \left(\frac{1}{n_2}\sum_{j=1}^{n_2} H(\hat{f},X_j,Y_j) - \iint H(\hat{f},x,y)f(x,y)dxdy\right)\right], \\ R_2 &= \sqrt{n}\left[\frac{1}{n_2}\sum_{j=1}^{n_2}\left(H(\hat{f},X_j,Y_j) - \iint H(\hat{f},x,y)f(x,y)dxdy\right)\right] \\ &\quad -\sqrt{n}\left[\frac{1}{n_2}\sum_{j=1}^{n_2}\left(H(f,X_j,Y_j) - \iint H(f,x,y)f(x,y)dxdy\right)\right].\end{aligned}$$

We propose to show that both $\mathbb{E}(R_1^2)$ and $\mathbb{E}(R_2^2)$ converge to 0. We can write R_1 as follows :

$$R_1 = -\sqrt{n}\left[\hat{Q}' - Q' + \Gamma_n\right]$$

where

$$Q' = \iiint K(\hat{f}, x, y, z) f(x, y) f(x, z),$$
$$K(\hat{f}, x, y, z) = \frac{1}{2} \frac{\ddot{\psi}(\hat{m}(x))}{\left(\int \hat{f}(x, y) dy\right)} \big(\hat{m}(x) - \varphi(y)\big)\big(\hat{m}(x) - \varphi(z)\big)$$

and $\hat{Q}'$ is the corresponding estimator. Since $\mathbb{E}\left(\Gamma_n^2\right) = o(1/n)$, we just have to control the expectation of the square of $\sqrt{n}\left[\hat{Q}' - Q'\right]$:

Lemma A.10: *Assuming the hypotheses of Theorem 3.4 hold, we have*

$$\lim_{n\to\infty} n\mathbb{E}\left(\hat{Q}' - Q'\right)^2 = 0.$$

Proof: The bound given in (9) states that if $|M_n|/n \to 0$ we have

$$\begin{aligned}
&\left| n\mathbb{E}\left[\left(\hat{Q}' - Q'\right)^2 | \hat{f}\right]\right. \\
&\left. -4\left[\iint \hat{g}(x, y)^2 f(x, y) dx dy - \left(\iint \hat{g}(x, y) f(x, y) dx dy\right)^2\right]\right| \\
&\leq \gamma_1(\|f\|_\infty, \|\psi\|_\infty, \Delta_Y)\left[\frac{|M_n|}{n} + \|S_M f - f\|_2 + \|S_M \hat{g} - \hat{g}\|_2\right]
\end{aligned}$$

where $\hat{g}(x, y) = \int K(\hat{f}, x, y, z) f(x, z) dz$. By deconditioning, we get

$$\begin{aligned}
&\left| n\mathbb{E}\left[\left(\hat{Q}' - Q'\right)^2\right]\right. \\
&\left. -4\mathbb{E}\left[\iint \hat{g}(x, y)^2 f(x, y) dx dy - \left(\iint \hat{g}(x, y) f(x, y) dx dy\right)^2\right]\right| \\
&\leq \gamma_1(\|f\|_\infty, \|\psi\|_\infty, \Delta_Y)\left[\frac{|M_n|}{n} + \|S_M f - f\|_2 + \mathbb{E}\left(\|S_M \hat{g} - \hat{g}\|_2\right)\right].
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\left(\|S_M \hat{g} - \hat{g}\|_2\right) &\leq \mathbb{E}\left(\|S_M \hat{g} - S_M g\|_2\right) + \mathbb{E}\left(\|S_M g - g\|_2\right) \\
&\leq \mathbb{E}\left(\|\hat{g} - g\|_2\right) + \mathbb{E}\left(\|S_M g - g\|_2\right)
\end{aligned}$$

where $g(x,y) = \int K(f,x,y,z)f(x,z)dz$. The second term converges to 0 since $g \in L^2(dxdy)$ and $\forall t \in L^2(dxdy)$, $\int (S_M t - t)^2 d\mu \to 0$. Moreover

$$\begin{aligned}
\|\hat{g} - g\|_2^2 &= \iint \left[\hat{g}(x,y) - g(x,y)\right]^2 f(x,y)dxdy \\
&= \iint \left[\int \left(K(\hat{f},x,y,z) - K(f,x,y,z)\right) f(x,z)dz\right]^2 f(x,y)dxdy \\
&\leq \iint \left[\int \left(K(\hat{f},x,y,z) - K(f,x,y,z)\right)^2 dz\right] \\
&\quad \left[\int f(x,z)^2 dz\right] f(x,y)dxdy \\
&\leq \Delta_Y^2 \|f\|_\infty^3 \iiint \left(K(\hat{f},x,y,z) - K(f,x,y,z)\right)^2 dxdz \\
&\leq \delta \Delta_Y^3 \|f\|_\infty^3 \iint (f(x,y) - \hat{f}(x,y))^2 dxdy
\end{aligned}$$

for some constant δ by applying the mean value theorem to $K(f,x,y,z) - K(\hat{f},x,y,z)$. Of course, the bound δ is obtained here by considering assumptions A1, A2 and A3. Since $\mathbb{E}(\|f - \hat{f}\|_2) \to 0$, we get $\mathbb{E}\left(\|\hat{g} - g\|_2\right) \to 0$. Let us now show that the expectation of

$$\iint \hat{g}(x,y)^2 f(x,y)dxdy - \left(\iint \hat{g}(x,y)f(x,y)dxdy\right)^2$$

converges to 0. We will only develop the proof for the first term :

$$\begin{aligned}
&\left|\iint \hat{g}(x,y)^2 f(x,y)dxdy - \iint g(x,y)^2 f(x,y)dxdy\right| \\
&\leq \iint \left|\hat{g}(x,y)^2 - g(x,y)^2\right| f(x,y)dxdy \\
&\leq \lambda \iint \left(\hat{g}(x,y) - g(x,y)\right)^2 dxdy \\
&\leq \lambda \|\hat{g} - g\|_2^2
\end{aligned}$$

for some constant λ. By taking the expectation of both sides, we see it is enough to show

that $\mathbb{E}\left(\|\hat{g}-g\|_2^2\right) \to 0$, which is done exactly as above. Besides, we can verify that

$$\begin{aligned} g(x,y) &= \int K(f,x,y,z) f(x,z) dz \\ &= \frac{1}{2} \frac{\ddot{\psi}(m(x))}{\left(\int f(x,y)dy\right)} \big(m(x) - \varphi(y)\big) \\ &\quad \left(m(x) \int f(x,z)dz - \int \varphi(z) f(x,z) dz\right) \\ &= 0, \end{aligned}$$

which proves that the expectation of $\iint \hat{g}(x,y)^2 f(x,y)dxdy$ converges to 0. Similar considerations show that the expectation of the second term $\left(\iint \hat{g}(x,y) f(x,y) dxdy\right)^2$ also converges to 0. We finally have

$$\lim_{n\to\infty} n\mathbb{E}\left(\hat{Q}' - Q'\right)^2 = 0.$$

□

Lemma A.10 imply that $\mathbb{E}(R_1^2) \to 0$. We will now prove that $\mathbb{E}(R_2^2) \to 0$:

$$\begin{aligned} \mathbb{E}(R_2^2) = \frac{n}{n_2} \mathbb{E}\left[\iint \left(H(f,x,y) - H(\hat{f},x,y)\right)^2 f(x,y)dxdy\right] \\ -\frac{n}{n_2} \mathbb{E}\left[\iint H(f,x,y) f(x,y)dxdy - \iint H(\hat{f},x,y) f(x,y)dxdy\right]^2. \end{aligned}$$

The same arguments as before (mean value theorem and assumptions A2 and A3) show that $\mathbb{E}(R_2^2) \to 0$. At last, we can give another expression for the asymptotic variance :

$$C(f) = \iint H(f,x,y)^2 f(x,y)dxdy - \left(\iint H(f,x,y) f(x,y)dxdy\right)^2.$$

We will prove that

$$C(f) = \mathbb{E}\left(\mathrm{Var}(\varphi(Y)|X)\left[\dot{\psi}\left(\mathbb{E}(Y|X)\right)\right]^2\right) + \mathrm{Var}\left(\psi\left(\mathbb{E}(\varphi(Y)|X)\right)\right).$$

Remark that

$$\begin{aligned}\iint H(f,x,y)f(x,y)dxdy &= \iint \left([\varphi(y) - m(x)]\,\dot{\psi}(m(x)) + \psi(m(x)) \right) f(x,y)dxdy \\ &= \iint m(x)\dot{\psi}(m(x))f(x,y)dxdy - \iint m(x)\dot{\psi}(m(x))f(x,y)dxdy \\ &\quad + \iint \psi(m(x))f(x,y)dxdy \\ &= \mathbb{E}\left(\psi\left(\mathbb{E}(\varphi(Y)|X)\right)\right). \end{aligned} \tag{A3}$$

Moreover,

$$\begin{aligned} H(f,x,y)^2 &= [\varphi(y) - m(x)]^2\,\dot{\psi}(m(x))^2 + \psi(m(x))^2 + 2\,[\varphi(y) - m(x)]\,\dot{\psi}(m(x))\psi(m(x)) \\ &= \varphi(y)^2\dot{\psi}(m(x))^2 + m(x)^2\dot{\psi}(m(x))^2 - 2\varphi(y)m(x)\dot{\psi}(m(x))^2 \\ &\quad + \psi(m(x))^2 + 2\,[\varphi(y) - m(x)]\,\dot{\psi}(m(x))\psi(m(x)). \end{aligned}$$

We can then rewrite $\iint H(f,x,y)^2 f(x,y)dxdy$ as:

$$\begin{aligned} &\iint \varphi(y)^2\dot{\psi}(m(x))^2 f(x,y)dxdy + \iint m(x)^2\dot{\psi}(m(x))^2 f(x,y)dxdy \\ &-2\iint \varphi(y)m(x)\dot{\psi}(m(x))^2 f(x,y)dxdy + \iint \psi(m(x))^2 f(x,y)dxdy \\ &+2\iint \varphi(y)\dot{\psi}(m(x))\psi(m(x))f(x,y)dxdy - 2\iint m(x)\dot{\psi}(m(x))\psi(m(x))f(x,y)dxdy \\ &= \iint v(x)\dot{\psi}(m(x))^2 f(x,y)dxdy - \iint m(x)^2\dot{\psi}(m(x))f(x,y)dxdy + \iint \psi(m(x))^2 f(x,y)dxdy \\ &= \iint \left(\left[v(x) - m(x)^2\right]\dot{\psi}(m(x))^2 + \psi(m(x))^2 \right) f(x,y)dxdy \\ &= \mathbb{E}\left(\left[v(X) - m(X)^2\right]\dot{\psi}(m(X))^2 \right) + \mathbb{E}\left(\psi(m(X))^2\right) \\ &= \mathbb{E}\left(\left[\mathbb{E}(\varphi(Y)^2|X) - \mathbb{E}(\varphi(Y)|X)^2\right]\left[\dot{\psi}(\mathbb{E}(\varphi(Y)|X))\right]^2 \right) + \mathbb{E}\left(\psi(\mathbb{E}(\varphi(Y)|X))^2\right) \\ &= \mathbb{E}\left(\mathrm{Var}(\varphi(Y)|X)\left[\dot{\psi}\left(\mathbb{E}(Y|X)\right)\right]^2 \right) + \mathbb{E}\left(\psi(\mathbb{E}(\varphi(Y)|X))^2\right) \end{aligned}$$

where we have set $v(x) = \int \varphi(y)^2 f(x,y)dy / \int f(x,y)dy$. This result and (A3) give the desired form for $C(f)$ which ends the proof of Theorem 3.4.

A.5. *Proof of Theorem 3.5*

We follow the proof of Theorem 3.3. Assumptions A2 and A3 imply that

$$T(f) - T(f_0) = \iint \Big(\big[\varphi(y) - m_0(x)\big]\dot{\psi}(m_0(x)) + \psi(m_0(x)) \Big) \Big(f(x,y) - f_0(x,y) \Big) dxdy + O\left(\int (f - f_0)^2 \right)$$

where $m_0(x) = \int \varphi(y) f_0(x,y) dy / \int f_0(x,y) dy$. This result shows that the Fréchet derivative of $T(f)$ at f_0 is $T'(f_0) \cdot h = \langle H(f_0, \cdot), h \rangle$ where

$$H(f_0, x, y) = \Big(\big[\varphi(y) - m_0(x)\big]\dot{\psi}(m_0(x)) + \psi(m_0(x)) \Big).$$

We then deduce that

$$\begin{aligned} K(h) &= T'(f_0) \cdot \left(\sqrt{f_0} \left(h - \sqrt{f_0} \int h \sqrt{f}_0 \right) \right) \\ &= \int H(f_0, \cdot) \sqrt{f_0} h - \int H(f_0, \cdot) \sqrt{f_0} \int h \sqrt{f_0} \\ &= \langle t, h \rangle \end{aligned}$$

with

$$t = H(f_0, \cdot) \sqrt{f_0} - \left(\int H(f_0, \cdot) f_0 \right) \sqrt{f_0}.$$

The semiparametric Cramér-Rao bound for this problem is thus

$$\|t\|^2_{L^2(dxdy)} = \int H(f_0, \cdot)^2 f_0 - \left(\int H(f_0, \cdot) f_0 \right)^2 = C(f_0)$$

where we recognize the expression of $C(f_0)$ in Theorem 3.5.