



HAL
open science

New estimation of Sobol' indices using kernels

Sébastien Da Veiga, Fabrice Gamboa, Agnès Lagnoux, Thierry Klein,
Clémentine Prieur

► **To cite this version:**

Sébastien Da Veiga, Fabrice Gamboa, Agnès Lagnoux, Thierry Klein, Clémentine Prieur. New estimation of Sobol' indices using kernels. 2023. hal-04052837

HAL Id: hal-04052837

<https://hal.science/hal-04052837>

Preprint submitted on 30 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

New estimation of Sobol' indices using kernels

Sébastien da Veiga¹, Fabrice Gamboa², Thierry Klein³, Agnès Lagnoux⁴,
and Clémentine Prieur⁵

1

²Institut de Mathématiques de Toulouse and ANITI; UMR5219.
Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France.

³Institut de Mathématiques de Toulouse; UMR5219. Université de
Toulouse; ENAC - Ecole Nationale de l'Aviation Civile , Université de
Toulouse, France

⁴Institut de Mathématiques de Toulouse; UMR5219. Université de
Toulouse; CNRS. UT2J, F-31058 Toulouse, France.

⁵Laboratoire Jean Kuntzmann; CNRS, UMR 5224. Université Grenoble
Alpes, F-38400 Saint Martin d'Hères , France

March 30, 2023

Contents

1	Introduction	2
2	Setting	4
2.1	Model and purpose	4
2.2	Notation and assumptions	5
3	Central limit theorem	6
4	Efficiency	9
5	Bandwidth selection	11
6	Extension to an unknown density f_X	12
7	Numerical applications	16
A	Proof of the results	16
A.1	Proof of Theorem 3.1	16
A.2	Proof of the remaining results	22
A.3	Construction of a kernel of order k in dimension d	26

1 Introduction

The use of complex computer models for the simulation and analysis of natural systems from physics, engineering and other fields is by now routine. These models usually depend on many input variables. Thus, it is crucial to understand which input parameter or which set of input parameters have an influence on the output. This is the aim of sensitivity analysis which has become an essential tool for system modeling and policy support (see, e.g., [39]). Global sensitivity analysis methods considers the input vector as random and propose a measure of the influence of each subset of its components on the output of interest. We refer to the seminal book by Saltelli [40] for an overview on global sensitivity analysis or to [10] for a synthesis of recent trends in the field. Among the different measures of global sensitivity analysis, variance-based measures are probably the most commonly used. The definition of the so-called Sobol' indices, introduced in [34] and later revisited in the framework of sensitivity analysis in [41] (see also [42]), is based on the Hoeffding decomposition of the variance [26]. More precisely, for $Y = f(X_1, \dots, X_p)$ where the inputs X_i are assumed to be mutually independent, the Sobol' index with respect to X is defined by

$$S^X = \frac{\text{Var}(\mathbb{E}[Y|X])}{\text{Var}(Y)} = \frac{\mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2}{\text{Var}(Y)}$$

where X is a subvector of (X_1, \dots, X_p) , for instance $X = (X_1, \dots, X_d)$ with $1 \leq d \leq p$ or $X = X_i$ (first-order index) with $1 \leq i \leq p$. For $i = 1, \dots, p$, it is also possible to measure the total influence of X_i , by itself or in interaction with other inputs, with the total Sobol' index defined by $1 - S^{X_{-i}}$ with $X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p)$ for $i = 1, \dots, p$.

In general, computing explicitly the theoretical value of S^X is hopeless; therefore one of the main tasks is to provide estimators of S^X that have some nice convergence properties (consistency, explicit rate of convergence, asymptotic efficiency...). Many different estimation procedures were developed in the last decades to estimate S^X or $\mathbb{E}[\mathbb{E}[Y|X]^2]$ only (since one easily estimates $\mathbb{E}[Y]$ and $\text{Var}(Y)$ by the empirical mean and the empirical variance respectively). Some are based on Monte Carlo, quasi Monte Carlo or nested Monte Carlo designs of experiments (see, e.g., [28, 33, 23]). Others are based on spectral approaches. Among this last class of procedures, let us cite without exhaustivity Fourier Amplitude Sensitivity Test (FAST) [8], Random Balance Design (RBD) [46], Effective Algorithm for computing global Sensitivity Indices (EASI) [36] and polynomial chaos expansions [45] (see also [10, Chapter 4] for a more complete review). In the sequel, we focus on two classes of methods that have emerged since they provide nice theoretical convergence guarantees.

The first class is based on the so-called Pick Freeze (PF) design of experiments, the theoretical properties of which (consistency, central limit theorem, concentration inequalities and Berry-Esseen bounds) have been studied in [18, 27]. Besides having very nice convergence properties, this method is very general since the only assumption needed to prove a central limit theorem is that $\mathbb{E}[Y^4] < \infty$. However, it requires the evaluation of the model on a PF design thus it cannot be computed by recycling former evaluations of the model. Also, the required sample size increases linearly with the number of indices that one wants to estimate. For instance, to estimate at rate \sqrt{n} a single Sobol' index, one needs a design of experiment of size $2n$. Additionally, this size grows to $(p+1)n$ to estimate the full set of first-order Sobol' indices (still at the rate \sqrt{n}). By evaluating PF

estimates on replicated designs, this cost has been reduced to $2n$ to estimate the full set of first-order and second-order indices [47, 20]. However the cost for estimating total Sobol' indices with this approach remains linear in the input space dimension.

The second class of methods is based on nearest neighbors and has been studied by several authors. See e.g. [16, 29, 30, 14, 24, 15]. For instance, in [15], the authors built an estimator of $\mathbb{E}[\mathbb{E}[Y|X]^2]$ based on two independent n -samples. The first one allows to estimate the regression function $\mathbb{E}[Y|X = x]$ using the first nearest neighbor of x among the first sample and the second one is used as a plug-in estimator. They proved a central limit theorem for their estimator. Nevertheless the recentering factor is not the unknown quantity of interest $\mathbb{E}[\mathbb{E}[Y|X]^2]$ but the expected value of their estimator. In addition, the bias is negligible if and only if the dimension $d \leq 3$. In [6], the author introduced a new coefficient of correlation based on ranks for $d = 1$. It was proven that this coefficient converges to a limit that appears to be the Cramér-von Mises index with respect to X_i (for $i = 1, \dots, p$) introduced in [19] as the size n of the sample goes to infinity. Note that the coefficient introduced in [6] admits a representation in terms of nearest neighbor (on the right) (see [17, Section3]). A central limit theorem was proved in [6] under the assumption that Y and X_i are independent. This theorem can be exploited for statistical tests of independence. The central limit theorem was extended in [31] (see also references therein) to the framework where Y is not a measurable function of X_i . The definition of the Chatterjee's coefficient was then extended in [2] including the case $d \geq 2$ with nearly optimal bounds on the rate of convergence. Chatterjee's approach was then adapted for building estimators of first-order Sobol' indices in [17], with a central limit theorem in the framework $d = 1$ and X_i not necessarily independent from Y . Broto *et al.* [4] also considered a nearest neighbor approach to estimate Sobol' indices of any order. The estimators they introduced are consistent but no rate of convergence was provided. Although very powerful it is clear thanks to the bias study in [15] that it is not possible to obtain a central limit theorem for estimating Sobol' indices of order larger than $d = 3$ using this approach. In particular, if $d \geq 5$, one can not estimate total order Sobol' indices with this kind of approaches at rate \sqrt{n} . It is important to note that all the procedures based on nearest neighbors require additional regularity assumptions on the model to control the bias. Close to nearest neighbor approaches are kernel methods. A natural estimator is the Nadaraya-Watson plug-in estimator or its generalization with local polynomial kernel smoothing. Based on works on the estimation of conditional covariance matrix [49, 32], kernel estimates of first-order Sobol' indices have been studied in [11, 9, 37, 43, 25]. In particular, the authors in [9] proved the asymptotic efficiency of their estimator of first-order Sobol' indices for $d = 1$. The analysis of such estimators also require regularity assumptions on the model.

Hence, in the one hand, the Pick Freeze procedure allows to estimate any Sobol' index at rate \sqrt{n} with a minimal assumption on the model but requires a very specific design of experiment, the size of which growing linearly with the number of indices that one wants to estimate (except the procedure based on replicated designs to estimate first-order and second-order indices, as already mentioned). In the other hand, the nearest neighbor approach requires a single n -sample but needs some regularity assumptions on the model and for which the \sqrt{n} parametric rate of convergence is possible only for indices up to order 3.

In the present paper, we develop a new approach mentioned in [9, Remark 1]. It consists

in extending the very interesting point of view introduced in [22, 21] to estimate general nonlinear integral functionals of a density on the real line, by using empirically a kernel estimator erasing the diagonal terms. Relaxing the positiveness assumption on the kernel and choosing a kernel of order large enough, we are able to prove a central limit theorem for estimating Sobol' indices of any order (the bias is killed thanks to this signed kernel).

The paper is organized as follows. In Section 2, we introduce the setting and fix notation. Section 3 is devoted to the statement of our main results. First we introduce our kernel estimator in terms of a degenerate U-statistics. Then we state a central limit theorem. A comparison study on the efficiency is provided in Section 4. Section 5 proposes an heuristic inspired from [12] for bandwidth selection. Our methodology is extended in Section 6 to the case where the distribution of the input is unknown. Section 7 presents several numerical illustrations of our estimation procedure. Most of the proofs are postponed to Appendices A.1 and A.2. Furthermore, Appendix A.3 presents the practical construction of a kernel of fixed order.

2 Setting

2.1 Model and purpose

The quantity of interest (QoI) Y is obtained from the numerical code and is regarded as a function g of the vector of the random inputs $(V_i)_{i=1,\dots,p}$

$$Y = g(V_1, \dots, V_p),$$

where the function g is defined on a compact subset of \mathbb{R}^p and is real-valued. Classically, the V_i 's are assumed to be independent random variables and a sensitivity analysis is performed using the Hoeffding decomposition [1, 48] leading to the standard Sobol' indices [41].

Let $\mathbf{u} \subset \{1, \dots, p\}$. In the sequel, X stands for a group of inputs: $X = (V_i)_{i \in \mathbf{u}}$ and $W = (V_i)_{i \notin \mathbf{u}}$ so that

$$Y = g(X, W) \tag{1}$$

(with a little abuse of notation). When \mathbf{u} is reduced to a singleton, X is simply one of the inputs V_i , $i = 1, \dots, p$ and S^X is the first order Sobol' index with respect to this input. In the sequel, the cardinality of \mathbf{u} is denoted by d : $|\mathbf{u}| = d$ with $1 \leq d \leq p$ and for the sake of simplicity, we assume that $\mathbf{u} = \{1, \dots, d\}$ without loss of generality.

Assume that $Y \in \mathbb{L}^2(\mathbb{R})$. We aim at estimating the Sobol' index of the output Y with respect to the d -dimensional vector X ; namely

$$S^X = \frac{\text{Var}(\mathbb{E}[Y|X])}{\text{Var}(Y)} = \frac{\mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2}{\text{Var}(Y)}. \tag{2}$$

Here, d is the order of the Sobol' index while p is the total number of inputs. When \mathbf{u} is reduced to a singleton, X is simply one of the inputs V_i , $i = 1, \dots, p$: $X = V_{i_0}$ and S^X is the first order Sobol' index with respect to this input V_{i_0} .

To estimate $\mathbb{E}[Y]$ and $\text{Var}(Y)$ from the n -sample $(X_j, Y_j)_{j=1, \dots, n}$ of the pair (X, Y) , we will naturally use the classical empirical mean and variance respectively. Thus we focus on the estimation of $\mathbb{E}[\mathbb{E}[Y|X]^2]$ in (2). In the sequel, we build a modified kernel-based procedure for the estimation of S^X on a n -sample $(X_j, Y_j)_{j=1, \dots, n}$ of the pair (X, Y) . A natural estimator inspired from the nearest neighbor estimator as considered in [4] and the kernel-based plug-in estimator in [22] would be

$$\binom{n}{2}^{-1} \sum_{1 \leq j < j' \leq n} \frac{Y_j Y_{j'}}{2} \left(\frac{K_{h_n}(X_{j'} - X_j)}{f_X(X_j)} + \frac{K_{h_n}(X_j - X_{j'})}{f_X(X_{j'})} \right)$$

for a bandwidth $h_n > 0$ and a kernel K_{h_n} . Nevertheless, as pointed in [3], boundary issues appear when the input domain is compact. To bypass this issue, we consider the following kernel-based estimator

$$T_{n, h_n} := \binom{n}{2}^{-1} \sum_{1 \leq j < j' \leq n} \frac{Y_j Y_{j'}}{2} \left(\frac{K_{h_n} \circ A_{X_j}(X_{j'} - X_j)}{f_X(X_j)} + \frac{K_{h_n} \circ A_{X_{j'}}(X_j - X_{j'})}{f_X(X_{j'})} \right). \quad (3)$$

for a bandwidth $h_n > 0$, a mirror-type transformation A_x , and a kernel K_{h_n} as defined in the next section.

2.2 Notation and assumptions

In this section, we give the general assumptions made in the paper.

- (A1) The support of (V_1, \dots, V_p) is of the form $[B_1, C_1] \times \dots \times [B_p, C_p]$ where $B_i < C_i$ for all $1 \leq i \leq p$. We denote by \mathcal{D} the support of X .
- (A2) The distribution of the random vector $(V, Y) \in \mathbb{R}^p \times \mathbb{R}$ is absolutely continuous with respect to the Lebesgue measure. The marginal probability density function of (X, Y) , V , X , and W are denoted by $f_{X,Y}$, f_V , f_X , and f_W respectively.

The next definition is borrowed from [3, Section 2.2] and allows to circumvent the boundary issues.

- (D1) For $x \in \mathcal{D}$, we define

$$A_x: \begin{cases} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ u = (u_1, \dots, u_d) & \mapsto & (\sigma_1(x_1)u_1, \dots, \sigma_d(x_d)u_d) \end{cases} \quad (4)$$

with $\sigma_i(s) := 1 - 2\mathbb{1}\left(\frac{B_i + C_i}{2}, C_i\right)(s) \in \{-1, 1\}$. Observe that $\mathcal{A} = \{A_x, x \in \mathcal{D}\}$ is a finite subset of $GL_d(\mathbb{R})$, $\mathcal{A} = \{A_1, \dots, A_\kappa\}$, with cardinal $\kappa = 2^d$. Moreover, it satisfies

- (i) for any $\ell = 1, \dots, \kappa$, $|\det(A_\ell)| = 1$;
- (ii) **Mirror condition:** for any $x \in \mathcal{D}$, there exists $A_\ell \in \mathcal{A}$ such that $A_x = A_\ell$ and $x + A_x^{-1}([0, 1/2]^d) \subset \mathcal{D}$.

For any integer d , let us introduce the multi-index notation. For any $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}_+^d$, we define the integer part of β by

$$\lfloor \beta \rfloor = (\lfloor \beta_1 \rfloor, \dots, \lfloor \beta_d \rfloor) =: \gamma \in \mathbb{N}^d.$$

In addition, we introduce

$$|\gamma| = \gamma_1 + \dots + \gamma_d, \quad \gamma! = \gamma_1! \dots \gamma_d!, \quad \text{and} \quad v^\beta = v_1^{\beta_1} \dots v_d^{\beta_d}$$

for any $v \in \mathbb{R}^d$.

Let $\alpha > 0$. We define $\mathcal{C}^\alpha(\mathcal{D})$ as the set of functions $\phi: \mathcal{D} \rightarrow \mathbb{R}$ that have derivatives up to order $\lfloor \alpha \rfloor$ and which partial derivative of order $\lfloor \alpha \rfloor$ is $\alpha - \lfloor \alpha \rfloor$ -Hölder. Namely, there exists $C_\phi > 0$ such that, for any x and $x' \in \mathcal{D}$, one has

$$\left| \frac{\partial^\beta \phi}{\partial x^\beta}(x) - \frac{\partial^\beta \phi}{\partial x^\beta}(x') \right| \leq C_\phi \|x - x'\|_\infty^{\alpha - \lfloor \alpha \rfloor} \quad (5)$$

for any $\beta \in \mathbb{N}^d$ such that $|\beta| = \lfloor \alpha \rfloor$ and where $(\partial^\beta \cdot / \partial x^\beta)$ stands for the partial derivative of order β . Here, $\|\cdot\|_\infty$ stands for the supremum norm.

(A3) Let $K: \mathbb{R}^d \rightarrow \mathbb{R}$ be a kernel with support included in \mathcal{D} such that $\|K\|_\infty < +\infty$ and $\int_{\mathcal{D}} K(u) du = 1$. We assume that K is of order $\lfloor \alpha \rfloor$ which means that $\int_{\mathcal{D}} u^\beta K(u) du = 0$ for any $\beta \in \mathbb{N}^d$ such that $0 < |\beta| \leq \lfloor \alpha \rfloor$. Finally, we define $K_h(x) = K(x/h)/h^d$ for any $x \in \mathcal{D}$.

(A4) The sequence $(h_n)_{n \in \mathbb{N}}$ of bandwidths is positive and such that $h_n \rightarrow 0$ as n goes to ∞ .

In this article, we write C for a deterministic and finite constant which value is allowed to change between occurrences.

3 Central limit theorem

Recall the definition of our estimator given in (3):

$$T_{n,h_n} = \binom{n}{2}^{-1} \sum_{1 \leq j < j' \leq n} \frac{Y_j Y_{j'}}{2} \left(\frac{K_{h_n} \circ A_{X_j}(X_{j'} - X_j)}{f_X(X_j)} + \frac{K_{h_n} \circ A_{X_{j'}}(X_j - X_{j'})}{f_X(X_{j'})} \right).$$

The goal is to derive a central limit theorem for T_{n,h_n} . Since T_{n,h_n} is a U-statistics, our analysis is based on the Hoeffding decomposition of U-statistics [26]. More precisely, for all $j, j' = 1, \dots, n$, we introduce the symmetric function given by

$$R\left(\begin{pmatrix} x_j \\ y_j \end{pmatrix}, \begin{pmatrix} x_{j'} \\ y_{j'} \end{pmatrix}\right) = \frac{y_j y_{j'}}{2} \left(\frac{K_{h_n} \circ A_{x_j}(x_{j'} - x_j)}{f_X(x_j)} + \frac{K_{h_n} \circ A_{x_{j'}}(x_j - x_{j'})}{f_X(x_{j'})} \right).$$

Then the Hoeffding projections of R are given by

$$\begin{aligned} \pi_1 R\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) &= \mathbb{E}\left[R\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right)\right] - \mathbb{E}\left[R\left(\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right)\right] \\ \pi_2 R\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) &= R\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) - \mathbb{E}\left[R\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right)\right] - \mathbb{E}\left[R\left(\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right)\right] \\ &\quad + \mathbb{E}\left[R\left(\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right)\right] \end{aligned}$$

(see [35, page 137]) leading to the Hoeffding decomposition

$$T_{n,h_n} - \mathbb{E}[T_{n,h_n}] = 2U_n^{(1)}(\pi_1 R) + U_n^{(2)}(\pi_2 R)$$

with

$$U_n^{(1)}(\pi_1 R) = \frac{1}{n} \sum_{j=1}^n (\pi_1 R) \begin{pmatrix} X_j \\ Y_j \end{pmatrix}$$

$$U_n^{(2)}(\pi_2 R) = \frac{2}{n(n-1)} \sum_{1 \leq j < j' \leq n} (\pi_2 R) \left(\begin{pmatrix} X_j \\ Y_j \end{pmatrix}, \begin{pmatrix} X_{j'} \\ Y_{j'} \end{pmatrix} \right).$$

Then following [22] we write

$$T_{n,h_n} - \mathbb{E}[T_{n,h_n}] = \frac{1}{n} \sum_{j=1}^n 2(Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2])$$

$$+ \underbrace{2U_n^{(1)}(\pi_1 R) - \frac{2}{n} \sum_{j=1}^n (Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2])}_{S_1} + \underbrace{U_n^{(2)}(\pi_2 R)}_{S_2}, \quad (6)$$

where $g_1(x) = \mathbb{E}[Y|X = x]$. Finally, we define $g_2(x) = \mathbb{E}[Y^2|X = x]$.

Theorem 3.1. *Suppose that Assumptions \mathcal{A}_1 to \mathcal{A}_4 hold and $\mathbb{E}[Y^2] < \infty$. Let $\alpha > 0$.*

I.1. Assume that $g_1 \in \mathbb{L}^1(\mathcal{D})$, $g_1 f_X \in \mathcal{C}^\alpha(\mathcal{D})$, and

$$\int_{\mathcal{D}} \left| g_1(x) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) \right| dx < \infty$$

for any β such that $1 \leq |\beta| < \lfloor \alpha \rfloor$. Then we have

$$\left| \mathbb{E}[T_{n,h_n}] - \mathbb{E}[\mathbb{E}[Y|X]^2] \right| \leq C h_n^\alpha. \quad (7)$$

I.2. Assume in addition that $g_1 \in \mathcal{C}^\alpha(\mathcal{D})$, $g_2/f_X \in \mathbb{L}^1(\mathcal{D}) \cap \mathbb{L}^2(\mathcal{D})$ and $g_2 f_X \in \mathbb{L}^2(\mathcal{D})$, and

$$\int_{\mathcal{D}} g_2(x) \left| \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) \frac{\partial^{\beta'} (g_1 f_X)}{\partial x^{\beta'}}(x) \right| dx < \infty$$

for any β and β' such that $1 \leq |\beta|, |\beta'| < \lfloor \alpha \rfloor$. Then we have

$$\mathbb{E} \left[\left(T_{n,h_n} - \mathbb{E}[T_{n,h_n}] - \frac{1}{n} \sum_{j=1}^n Z_j \right)^2 \right] \leq C h_n^{2\alpha} + \frac{C}{h_n^d n^2}. \quad (8)$$

where, for $j = 1, \dots, n$,

$$Z_j = 2(Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2]). \quad (9)$$

II. Assuming in addition that $\mathbb{E}[Y^4] < \infty$, $\alpha > d/2$, $h_n \xrightarrow{n \rightarrow \infty} 0$, $nh_n^d \xrightarrow{n \rightarrow \infty} \infty$, and $nh_n^{2\alpha} \xrightarrow{n \rightarrow \infty} 0$, we get

$$\sqrt{n} \left(T_{n,h_n} - \mathbb{E}[\mathbb{E}[Y|X]^2] \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 4\tau^2) \quad (10)$$

with $\tau^2 = \text{Var}(Y g_1(X))$.

The proof of Theorem 3.1 is postponed to Appendix A.1.

Remark 3.2.

1. The assumptions of Theorem 3.1 are fulfilled as soon as g and $f_X \in \mathcal{C}^\alpha(\mathcal{D})$ and f_X is bounded from below by a strictly positive constant.
2. The $\mathcal{C}^\alpha(\mathcal{D})$ regularity assumption can be relaxed to piecewise regularity, as soon as the number of pieces m is finite. The cardinality of \mathcal{A} is then increased to $\kappa = (2m)^d$.
3. We still assume that g and $f_X \in \mathcal{C}^\alpha(\mathcal{D})$ but one may relax the condition of boundedness from below on f_X . Since the inputs V_i 's are independent, $f_X = \prod_{i=1}^p f_{V_i}$, it amounts to control the local behavior of each marginal f_{V_i} at any point $v_0 \in \mathcal{D}$ such that $f_{V_i}(v_0) = 0$. We indeed assume that for any such v_0 , $f_{V_i}(v) = (v - v_0)^r h_{V_i}(v)$ with $h_{V_i}(v_0) \neq 0$ and $0 < r < 1/2$. For instance, one may take all the V_i 's Beta-distributed:

$$f_{V_i}(v) = \frac{v^{a_i-1}(1-v)^{b_i-1}}{B(a_i, b_i)} \mathbb{1}_{[0,1]}(v) \quad (11)$$

with parameters a_i and $b_i \in (1, 3/2)$ for $1 \leq i \leq p$ and $B(a_i, b_i)$ the normalizing constant equal to the beta function.

4. An example of convenient bandwidths $(h_n)_{n \geq 0}$ is given by $h_n = n^{-\gamma}$ with $1/(2\alpha) < \gamma < 1/d$.

Using the delta method in [48, Theorem 3.1], we are now able to get the asymptotic behavior of the estimation of S^X .

Corollary 3.3. *Under all the assumptions of Theorem 3.1 (II included), one has*

$$\sqrt{n} \left(\frac{T_{n, h_n} - \left(\frac{1}{n} \sum_{j=1}^n Y_j \right)^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2 - \left(\frac{1}{n} \sum_{j=1}^n Y_j \right)^2} - S^X \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (12)$$

where the limit variance σ^2 has an explicit expression given in (34).

Using one more time the delta method in [48, Theorem 3.1], we deduce the asymptotic behavior of the vector containing the p first-order Sobol' indices. Let us denote S^i the first-order Sobol index associated to $X = V_i$ and its estimator \widehat{S}^i given by:

$$\widehat{S}^i := \frac{T_{n, h_n} - \left(\frac{1}{n} \sum_{j=1}^n Y_j \right)^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2 - \left(\frac{1}{n} \sum_{j=1}^n Y_j \right)^2}.$$

Corollary 3.4. *Under all the assumptions of Theorem 3.1 (II included), one has*

$$\sqrt{n} \left((\widehat{S}^1, \dots, \widehat{S}^p)^T - (S^1, \dots, S^p)^T \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma), \quad (13)$$

where the limit variance Σ has an explicit expression given in (35).

Remark 3.5. The proof of Corollary 3.4 does not rely on the fact that we are dealing with order one indices, hence if u_1, \dots, u_r are distinct subsets of $\{1, \dots, p\}$ we also have under the same assumptions that

$$\sqrt{n} \left((\widehat{S}^{u_1}, \dots, \widehat{S}^{u_r})^T - (S^{u_1}, \dots, S^{u_r})^T \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_r), \quad (14)$$

where the limit variance Σ_r has an explicit expression similar to (35).

4 Efficiency

In this section, we compare several estimators proposed in the literature to estimate $\mathbb{E}[\mathbb{E}[Y|X]^2]$ (or even S^X) in terms of limiting variances obtained in the central limit theorem.

Estimating $\mathbb{E}[\mathbb{E}[Y|X]^2]$

Using a unique n -sample, one may compare the kernel estimator introduced in this paper and the procedure involving the estimators \widehat{T}_n defined in [9, page 11]. Such estimators are particularly tricky to compute and not easily tractable in practice. More precisely, the initial n -sample is split into two samples of sizes $n_1 = \lfloor n/\log n \rfloor$ and $n_2 = n - n_1 \approx n$. The first sample is dedicated to the estimation of the joint density of (X, Y) while the second one is used to compute a Monte-Carlo estimation of the integral involved in the quantity of interest.

In the following, we give the limiting variance σ_T^2 associated to the sequence $(\widehat{T}_n)_n$. By [9, Theorems 3.4 and 3.5], this sequence is asymptotically efficient to estimate $\mathbb{E}[\mathbb{E}[Y|X]^2]$; so that the variance σ_T^2 is the minimal variance expected for an estimator of $\mathbb{E}[\mathbb{E}[Y|X]^2]$ based on the previous design of experiment (a unique n -sample $(X_i, Y_i)_{i=1, \dots, n}$ of the input-output pair (X, Y)). We refer the reader to [9, 17].

Proposition 4.1 (Theorems 3.4 and 3.5 in [9]). *The minimal variance σ_T^2 is given by*

$$\sigma_T^2 = \text{Var}(g_1(X)(2Y - g_1(X))) = 4\tau^2 - 3\text{Var}(g_1^2(X))$$

with $\tau^2 = \text{Var}(Yg_1(X))$.

One may also compare our results to the estimation procedure proposed in [15]. Here again the initial $2n$ -sample is split into two samples of sizes n . The first sample allows to estimate the regression function $\mathbb{E}[Y|X = x]$ using the first nearest neighbor of x among the points of the first sample while the second sample is used as a plug-in estimator. They proved a central limit theorem. Nevertheless, in their result the recentering factor is not the unknown quantity of interest $\mathbb{E}[\mathbb{E}[Y|X]^2]$ as required but the expectation of the estimator. In addition, the bias is negligible if and only if the dimension $d \leq 3$ so that one can not expect a central limit theorem for the difference between the estimator and the quantity of interest $\mathbb{E}[\mathbb{E}[Y|X]^2]$ as soon as $d \geq 4$. In the following proposition, we give the limiting variance obtained with this estimation procedure. The factor 2 takes into account that we have considered two samples of size n .

Proposition 4.2 (Theorem 2.1 in [15]). *Under the assumptions of Theorem 2.1 in [15] and for $d \leq 3$, the limiting variance σ_D^2 is given by*

$$\sigma_D^2 = 2\left(\mathbb{E}[g_2(X)^2] - \mathbb{E}[g_1(X)^2]^2 + \frac{1}{2}\left(\mathbb{E}[g_2(X)g_1(X)^2] - \mathbb{E}[g_1(X)^4]\right)\right).$$

When $d = 1$, one may use the estimation procedure based on ranks introduced in [6] and studied in [17].

Proposition 4.3 (Proof of Theorem 4.1 in [17]). *Under the assumptions of Theorem 4.1 in [17], the limiting variance σ_D^2 is given by*

$$\begin{aligned} \sigma_R^2 = & \mathbb{E}\left[\text{Var}(YY'|X)\right] + \mathbb{E}\left[\text{Cov}(YY', YY''|X)\right] - \mathbb{E}[(Y + Y')g_x(X, W)X]^2 \\ & + \mathbb{E}[(Y + Y')(\tilde{Y} + \tilde{Y}')g_x(X, W)g_x(\tilde{X}, \tilde{W})(X \wedge \tilde{X})], \end{aligned}$$

where $Y = g(V_1, V_2, \dots, V_p) = g(X, W)$ (see (1)), $Y' = g(X, W')$, $Y'' = g(X, W'')$, $\tilde{Y} = g(\tilde{X}, \tilde{W})$, and $\tilde{Y}' = g(\tilde{X}, \tilde{W}')$ with X and \tilde{X} i.i.d., W, \tilde{W}, W' , and W'' i.i.d. also independent of X and \tilde{X} . Note that YY' and Y'' (respectively \tilde{Y}') are Pick-Freeze versions of Y (resp. \tilde{Y}). Here g_x stands for the first derivative of g with respect to the first coordinate.

Estimating S^X

The sequence of estimators \hat{T}_n being efficient to estimate $\mathbb{E}[\mathbb{E}[Y|X]^2]$, we get naturally an asymptotically efficient sequence of estimators of S^X . The proof of the following proposition relies on a result on efficiency in product space [48, Theorem 25.50] and another result on efficiency and delta method [48, Theorem 25.47]. The details can be found in [17].

Proposition 4.4. *Consider the sequence of estimators \hat{T}_n introduced in [9, page 11]. Assume that the joint distribution P of (X, Y) is absolutely continuous with respect to the product probability $P_X \otimes P_Y$, namely $P(dx, dy) = f_{X,Y}(x, y)P_X(dx)P_Y(dy)$. Then the sequence $(\hat{R}_n)_n$*

$$\hat{R}_n = \frac{\hat{T}_n - \left(\frac{1}{n} \sum_{i=1}^n Y_i\right)^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2 - \left(\frac{1}{n} \sum_{j=1}^n Y_j\right)^2}$$

is asymptotically efficient in estimating S^X . In addition, its (minimal) variance σ_{\min}^2 is

$$\sigma_{\min}^2 = \frac{1}{\text{Var}(Y)^2} \text{Var}\left(2\mathbb{E}[Y](1 - S^X)Y + S^X Y^2 + \mathbb{E}[Y|X](\mathbb{E}[Y|X] - 2Y)\right).$$

The delta method can also be applied to compute the asymptotic variance for the estimation of S^X with the other estimation procedures, namely the procedure in [15] and the one in [17], with the empirical mean and empirical variance to estimate respectively $\mathbb{E}[Y]$ and $\text{Var}(Y)$.

5 Bandwidth selection

Recall that $X = (V_i)_{1 \leq i \leq d}$, $W = (V_i)_{d+1 \leq i \leq p}$ as done in Section 2 so that $Y = g(X, W)$ as in (1). In the following, we propose a bandwidth selection procedure inspired from [12] to estimate $\mathbb{E}[\mathbb{E}[Y|X]^2] = \mathbb{E}[g_1(X)^2]$ with $g_1(x) = \mathbb{E}[Y|X = x]$ by

$$T_{n, h_n} = \frac{2}{n(n-1)} \sum_{1 \leq j < j' \leq n} \frac{Y_j Y_{j'}}{2} \left(\frac{K_{h_n} \circ A_{X_j}(X_{j'} - X_j)}{f_X(X_j)} + \frac{K_{h_n} \circ A_{X_{j'}}(X_j - X_{j'})}{f_X(X_{j'})} \right)$$

using the sample $\left((V_{1,j}, \dots, V_{p,j}), Y_j \right)_{1 \leq j \leq n}$ where $X_j = (V_{i,j})_{1 \leq i \leq d}$.

Based on the whole sample $\left((V_{1,j}, \dots, V_{p,j}), Y_j \right)_{1 \leq j \leq n}$, we first build a kernel estimator of the function g defined as:

$$\tilde{g}(v_1, \dots, v_p) = \frac{1}{n} \frac{1}{f_V(v_1, \dots, v_p)} \sum_{j=1}^n Y_j \prod_{i=1}^p \tilde{K}_{h_{0,i}}(V_{i,j} - v_i), \quad (15)$$

where as usual, $\tilde{K}_h(\cdot) = \tilde{K}(\cdot)/h$, where \tilde{K} is a kernel defined on \mathbb{R} and where the bandwidths $h_{0,i}$ for $i = 1, \dots, d$ are chosen using a classical rule of thumb. The kernel \tilde{K} is different from K and can be chosen arbitrarily. In the following, we will consider a Gaussian kernel. This kernel estimator \tilde{g} will serve as a test function which is expected to be close to g , on which the bandwidths of our estimator will be calibrated with a simple procedure. First, we can get an analytical expression of the Sobol' index of \tilde{g} , which will be our target. Then, \tilde{g} will be used to generate virtual samples that can be plugged in our estimator. Finally, we will optimize the bandwidths in order to reach the target. In the sequel the variables (v_1, \dots, v_p) will be replaced by random variables $(\tilde{V}_1, \dots, \tilde{V}_p)$ and we define $\tilde{X} := (\tilde{V}_1, \dots, \tilde{V}_d)$, hence we will have to deal with two sources of randomness one coming from our original sample $((V_{1,j}, \dots, V_{p,j}), Y_j)_{1 \leq j \leq n}$ and the other one from $(\tilde{V}_1, \dots, \tilde{V}_p)$. We denote by $\tilde{\mathbb{E}}$ for the conditional expectation knowing the first source of randomness (the sequence $((V_{1,j}, \dots, V_{p,j}), Y_j)_{1 \leq j \leq n}$ is considered as fixed and we integrate with respect to the distribution of $(\tilde{V}_1, \dots, \tilde{V}_p)$).

Then we define an approximation \tilde{g}_1 of g_1 based on \tilde{g} as:

$$\tilde{g}_1(x) = \tilde{\mathbb{E}}[\tilde{g}(\tilde{V}_1, \dots, \tilde{V}_p) | \tilde{X} = x].$$

Finally, the virtual sample from \tilde{g} is given by $((V_{1,j}, \dots, V_{p,j}), \tilde{Y}_j = \tilde{g}(V_{1,j}, \dots, V_{p,j}))_{1 \leq j \leq n}$ and is used to build our estimator

$$\tilde{T}_{n, h} = \frac{2}{n(n-1)} \sum_{1 \leq j < j' \leq n} \frac{\tilde{Y}_j \tilde{Y}_{j'}}{2} \left(\frac{K_{h_n} \circ A_{X_j}(X_{j'} - X_j)}{f_X(X_j)} + \frac{K_{h_n} \circ A_{X_{j'}}(X_j - X_{j'})}{f_X(X_{j'})} \right).$$

The optimal bandwidth h^* is then given by :

$$h^* = \operatorname{argmin}_h |\tilde{T}_{n, h} - \tilde{\mathbb{E}}[\tilde{g}_1(\tilde{X})^2]|. \quad (16)$$

Computation of $\tilde{\mathbb{E}}[\tilde{g}_1(\tilde{X})^2]$ First, one has

$$\tilde{\mathbb{E}}[\tilde{g}_1(\tilde{X})^2] = \frac{1}{n^2} \sum_{1 \leq j \leq j' \leq n} Y_j Y_{j'} \left(\prod_{i=1}^d \tilde{\beta}_{i,j,j'} \right) \left(\prod_{i=d+1}^p \tilde{\beta}_{i,j} \tilde{\beta}_{i,j'} \right) \quad (17)$$

where, for $1 \leq j, j' \leq n$,

$$\begin{aligned} \tilde{\beta}_{i,j,j'} &:= \int_{B_i}^{C_i} \tilde{K}_{h_{0,i}}(V_{i,j} - v) \tilde{K}_{h_{0,i}}(V_{i,j'} - v) \frac{1}{f_{V_i}(v)} dv \quad \text{for } 1 \leq i \leq d, \\ \tilde{\beta}_{i,j} &:= \int_{B_i}^{C_i} \tilde{K}_{h_{0,i}}(V_{i,j} - v) dv \quad \text{for } d+1 \leq i \leq p. \end{aligned}$$

The details of the calculus leading to (17) are postponed to Appendix A.2.

Second, the point is to choose a kernel \tilde{K} so that $\tilde{\mathbb{E}}[\tilde{g}_1(\tilde{X})^2]$ has an explicit expression. Note that if we assume that the kernel \tilde{K} is Gaussian and that the inputs are uniformly distributed on $[0, 1]$, this is the case: for $1 \leq j, j' \leq n$, $\tilde{\beta}_{i,j}$ and $\tilde{\beta}_{i,j,j'}$ writes

$$\begin{aligned} \tilde{\beta}_{i,j,j'} &= \frac{1}{\sqrt{2}h_{0,i}} \varphi\left(\frac{V_{i,j} - V_{i,j'}}{\sqrt{2}h_{0,i}}\right) \left[\Phi_{\frac{V_{i,j} + V_{i,j'}}{2}, \frac{h_{0,i}}{\sqrt{2}}}(1) - \Phi_{\frac{V_{i,j} + V_{i,j'}}{2}, \frac{h_{0,i}}{\sqrt{2}}}(0) \right] \quad \text{for } 1 \leq i \leq d, \\ \tilde{\beta}_{i,j} &= \Phi_{V_{i,j}, h_{0,i}}(1) - \Phi_{V_{i,j}, h_{0,i}}(0) \quad \text{for } d+1 \leq i \leq p. \end{aligned}$$

with φ the probability density function of the standard normal distribution and $\Phi_{\mu,\sigma}$ the cumulative distribution function of a normal distribution with mean μ and standard deviation σ .

Algorithm 1 Bandwidth selection

Enter the total number n of calls to the computer code
Choose the kernel \tilde{K}
Choose the bandwidths $(h_{0,i})_{1 \leq i \leq p}$ using a classical rule of thumb
Simulate a n -sample $((V_{1,j}, \dots, V_{p,j}), Y_j)_{1 \leq j \leq n}$ of input/output
Compute the values of $(\tilde{\beta}_{i,j,j'})_{1 \leq i \leq d, 1 \leq j, j' \leq n}$ and $(\tilde{\beta}_{i,j})_{d+1 \leq i \leq p, 1 \leq j \leq n}$
Compute the value of $\tilde{\mathbb{E}}[\tilde{g}_1(\tilde{X})^2]$ given in (17)
Compute the n sample $(\tilde{Y}_j = \tilde{g}(V_{1,j}, \dots, V_{p,j}))_{1 \leq j \leq n}$ where \tilde{g} is defined in (15)
Solve the optimization problem given in (16)
return the optimal value h^*

6 Extension to an unknown density f_X

In this section, we assume that the probability density function f_X is unknown and our aim is to propose a generalization of T_{n,h_n} with a plug-in estimator \hat{f}_X :

$$\hat{T}_{n,h} := \frac{2}{n(n-1)} \sum_{1 \leq j < j' \leq n} \frac{Y_j Y_{j'}}{2} \left(\frac{K_h \circ A_{X_j}(X_{j'} - X_j)}{\hat{f}_X(X_j)} + \frac{K_h \circ A_{X_{j'}}(X_j - X_{j'})}{\hat{f}_X(X_{j'})} \right).$$

Corollary 6.1. *Assume that the assumptions of Theorem 3.1 are satisfied. Furthermore, assume that*

$$\mathbb{E}\left[\left(\frac{f_X(X_1) - \hat{f}_X(X_1)}{\hat{f}_X(X_1)}\right)^2\right] = o\left(\frac{h_n^d}{n}\right). \quad (18)$$

Then we get

$$\sqrt{n}\left(\hat{T}_{n,h_n} - \mathbb{E}[\mathbb{E}[Y|X]^2]\right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 4\tau^2) \quad (19)$$

where τ^2 has been defined in Theorem 3.1.

The proof of Corollary 6.1 has been postponed in Appendix A.2.

Parametric estimation of f_X Here we assume that the unknown probability density function f_X belongs to a parametric family \mathcal{F} :

$$\mathcal{F} := \{f_\theta, \theta \in \Theta \subset \mathbb{R}^m\}$$

where f_θ is a probability density function supported by \mathcal{D} for any $\theta \in \Theta$. We denote by θ_0 the true parameter: $f_X = f_{\theta_0}$. Let $\hat{\theta}$ be a consistent estimator of θ_0 and estimate f_X by $f_{\hat{\theta}}$.

This setting is natural. Indeed, in several industrial applications, the input variables are assumed to follow a certain distribution, the parameters of which are given by expert judgment. For instance, in [38], the authors model the fuel mass M_{fuel} needed to link two fixed countries with a commercial aircraft by the Bréguet formula:

$$M_{fuel} = \left(M_{empty} + M_{load}\right) \left(e^{\frac{SFC \cdot g \cdot Ra}{V \cdot F} 10^{-3}} - 1\right) \quad (20)$$

where the cruise speed V and the lift-to-drag ratio F are assumed to be uniformly distributed and beta-distributed respectively with unknown parameters. See [38] for the description of the model with more details. Another example is given by the Gaussian Plume Model (GPM) that concerns a point source emitting contaminant into a unidirectional wind in an infinite domain. Such a model is also applied, for instance, to volcanic eruptions, pollen and insect dispersals (see, e.g., [5, 44]). More precisely, the contaminant concentration at location $(x, y, 0)$ (ground floor) involves several parameters and writes as:

$$C(x, y, 0) = \frac{Q}{2\pi Kx} e^{\frac{-u(y^2 + H^2)}{4Kx}}. \quad (21)$$

Based on reality constraints and guided by the expert knowledge, the stochastic parameters Q , K , and u of the model are assumed to be independent with uniform distribution on $[0, M_Q]$, $[0, M_K]$, and $[0, M_u]$ respectively.

Examples 6.2.

1. **Uniform distribution.** For sake of clarity, we only detail the framework of first-order Sobol' indices, that is $X = V_i$ for $i \in \{1, \dots, p\}$, and we assume that X is uniformly distributed on $[0, \theta]$ where θ is unknown. We estimate θ by

$$\hat{\theta} = \max(X_1, \dots, X_n).$$

It is a well known fact that $n(\theta - \hat{\theta})$ converges in distribution towards an exponential distribution of parameter $1/\theta$. Then one has

$$\mathbb{E} \left[\left(\frac{f_X(X_j) - \hat{f}_X(X_j)}{\hat{f}_X(X_j)} \right)^2 \right] = \mathbb{E} \left[\left(\frac{\theta^{-1} - \hat{\theta}^{-1}}{\hat{\theta}^{-1}} \right)^2 \right] = \frac{1}{\theta^2} \mathbb{E}[(\hat{\theta} - \theta)^2] = \frac{2}{(n+1)(n+2)}.$$

So that Condition (18) is fulfilled since $nh_n \rightarrow \infty$ ($d = 1$). Note that these computations can be easily generalized to the estimation of Sobol' indices of any order with uniformly distributed inputs.

2. **Beta distribution with parameters a and b in $(1, 3/2)$ unknown.** Once again, we only detail the framework of first-order Sobol' index estimation. Assume thus that $X = V_i$, for some $i \in \{1, \dots, p\}$ and assume that X is beta-distribution with a first unknown parameter a and a second known parameter b both in $(1, 3/2)$. Let $\hat{a} = b\bar{X}_m/(1 - \bar{X}_m)$ be the moment estimator of a , built from a m -sample $(\tilde{X}_1, \dots, \tilde{X}_m)$ of X , independent from the original n -sample. Here, \bar{X}_m stands for the empirical mean of the m -sample. Since the model is not evaluated on this new sample, even for costly numerical models, the estimation cost is still the one of n model evaluations.

Now let

$$\hat{f}_{a,b}(x) = \begin{cases} f_{a,b}(x) & \text{if } \hat{a} \in [1, 3/2] \\ \eta & \text{otherwise,} \end{cases}$$

where $\eta = (2 \int_0^1 \sqrt{x(1-x)} dx)^{-1}$.

Firstly, on the event $\{\hat{a} \notin [1, 3/2]\}$, one may check that $\|f_{a,b}\|_\infty \leq 2\eta$ so that

$$\left| \frac{f_{a,b}(X_j) - \hat{f}_{a,b}(X_j)}{\hat{f}_{a,b}(X_j)} \right| = \left| \frac{f_{a,b}(X_j) - \eta}{\eta} \right| \leq 1.$$

Furthermore, let us denote by Λ the log Laplace transform of X and Λ^* its Fenchel-Legendre transform:

$$\Lambda(\lambda) = \log \mathbb{E}[e^{\lambda X}] \quad \text{and} \quad \Lambda^*(s) = \sup_s \{\lambda s - \Lambda(\lambda)\}.$$

See e.g. [13] for details on these functions and their properties. Now, for all $\delta > 0$, it yields, for m large enough,

$$\begin{aligned} \mathbb{P}(\hat{a} \notin [1, 3/2]) &= \mathbb{P}(\bar{X}_m \leq \frac{1}{1+b}) + \mathbb{P}(\bar{X}_m \geq \frac{3/2}{3/2+b}) \\ &= \mathbb{P}\left(\bar{X}_m - \mathbb{E}[X] \leq \frac{1}{1+b} - \frac{a}{a+b}\right) + \mathbb{P}\left(\bar{X}_m - \mathbb{E}[X] \geq \frac{3/2}{3/2+b} - \frac{a}{a+b}\right) \\ &\leq 2 \exp\left\{-m \min\left\{\Lambda^*\left(\frac{1}{1+b} - \frac{a}{a+b}\right), \Lambda^*\left(\frac{3/2}{3/2+b} - \frac{a}{a+b}\right)\right\} - \delta\right\} \end{aligned} \quad (22)$$

since $\mathbb{E}[X] = a/(a+b)$ and using Cramér's theorem [13, Theorem 2.2.3] as $\Lambda < \infty$. Secondly, on the event $\{\hat{a} \in [1, 3/2]\}$, we have $\hat{a} > 0$, $2a - \hat{a} > 0$, and $3a - \hat{a} > 0$ and we get

$$\mathbb{E}\left[\left(\frac{f_{a,b}(X_j) - f_{\hat{a},b}(X_j)}{f_{\hat{a},b}(X_j)}\right)^2 \middle| \hat{a}\right] = \frac{B(\hat{a}, b)^2}{B(a, b)^3} B(3a - 2\hat{a}, b) - 2\frac{B(\hat{a}, b)}{B(a, b)^2} B(2a - \hat{a}, b) + 1.$$

Observe that the previous expression is a function of \bar{X}_m that vanishes at $\mathbb{E}[X]$. Moreover $\{\hat{a} \in [1, 3/2]\} = \{\bar{X}_m \in [1/(1+b), 3/(3+2b)]\}$ as $\hat{a} = b\bar{X}_m/(1 - \bar{X}_m)$. Thus, the function being infinitely differentiable, by applying the mean value, inequality we obtain

$$\begin{aligned} \mathbb{E}\left[\left(\frac{f_{a,b}(X_j) - f_{\hat{a},b}(X_j)}{f_{\hat{a},b}(X_j)}\right)^2 \mathbb{1}_{\hat{a} \in [1/2, 3/2]}\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{f_{a,b}(X_j) - f_{\hat{a},b}(X_j)}{f_{\hat{a},b}(X_j)}\right)^2 \middle| \hat{a}\right] \mathbb{1}_{\hat{a} \in [1/2, 3/2]}\right] \\ &= \mathbb{E}\left[\left(\varphi(\bar{X}_m) - \varphi(\mathbb{E}[X])\right)^2 \mathbb{1}_{\bar{X}_m \in [1/(1+2b), 3/(3+2b)]}\right] \\ &\leq C\mathbb{E}[|\bar{X}_m - \mathbb{E}[X]|] \\ &\leq C\mathbb{E}[(\bar{X}_m - \mathbb{E}[X])^2]^{1/2} \\ &= \frac{C}{m^{1/2}} \text{Var}(X)^{1/2}. \end{aligned} \quad (23)$$

Finally, in view of (22) and (23), Condition (18) is then fulfilled as soon as $m^{-1/2} = o(h_n n^{-1})$.

Nonparametric estimation of f_X Here the unknown density function f_X satisfies the assumptions of Theorem 3.1 and is moreover bounded from below by a constant $\eta > 0$. Then we estimate $f_X(x)$ by $\hat{f}_X(x) \vee \frac{\eta}{2}$ with $\hat{f}_X(x) = \frac{1}{m} \sum_{j=1}^m K_h \circ A_x(\bar{X}_j - m)$ the adaptive mirror-type estimator introduced in [3] built from a m -sample $(\bar{X}_1, \dots, \bar{X}_m)$ independent from the initial n -sample (X_1, \dots, X_n) . Then

$$\begin{aligned} \left(\frac{f(X_j) - \hat{f}_X(X_j) \vee \frac{\eta}{2}}{\hat{f}_X(X_j) \vee \frac{\eta}{2}}\right)^2 &\leq \frac{(f(X_j) - \hat{f}_X(X_j) \vee \frac{\eta}{2})^2}{\frac{\eta^2}{4}} \\ &\leq \frac{4}{\eta^2} \left\{ \left(f(X_j) - \frac{\eta}{2}\right)^2 \mathbb{1}_{\hat{f}_X(X_j) \leq \frac{\eta}{2}} + (f(X_j) - \hat{f}_X(X_j))^2 \mathbb{1}_{\hat{f}_X(X_j) > \frac{\eta}{2}} \right\} \\ &\leq \frac{4}{\eta^2} \left\{ \left(\|f_X\|_\infty - \frac{\eta}{2}\right)^2 \mathbb{1}_{\hat{f}_X(X_j) \leq \frac{\eta}{2}} + (f(X_j) - \hat{f}_X(X_j))^2 \mathbb{1}_{\hat{f}_X(X_j) > \frac{\eta}{2}} \right\} \end{aligned}$$

where the last inequality holds true as f_X is continuous and thus bounded on its bounded support. Thus

$$\mathbb{E}\left[\left(\frac{f(X_j) - \hat{f}_X(X_j) \vee \frac{\eta}{2}}{\hat{f}_X(X_j) \vee \frac{\eta}{2}}\right)^2\right] \leq \frac{4}{\eta^2} \left\{ \left(\|f_X\|_\infty - \frac{\eta}{2}\right)^2 \mathbb{E}[\mathbb{1}_{\hat{f}_X(X_j) \leq \frac{\eta}{2}}] + \mathbb{E}[(f(X_j) - \hat{f}_X(X_j))^2] \right\}.$$

In addition, we have

$$\mathbb{1}_{\hat{f}_X(X_j) \leq \frac{\eta}{2}} = \mathbb{1}_{f_X(X_j) + \hat{f}_X(X_j) - f_X(X_j) \leq \frac{\eta}{2}} = \mathbb{1}_{f_X(X_j) - \frac{\eta}{2} \leq f_X(X_j) - \hat{f}_X(X_j)} \leq \mathbb{1}_{\frac{\eta}{2} \leq f_X(X_j) - \hat{f}_X(X_j)}$$

yielding, using Markov's inequality,

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{\widehat{f}_X(X_j) \leq \frac{\eta}{2}}] &\leq \mathbb{E}\left[\mathbb{P}\left(f_X(X_j) - \widehat{f}_X(X_j) \geq \frac{\eta}{2} \mid \widetilde{X}_1, \dots, \widetilde{X}_m\right)\right] \\
&\leq \frac{4}{\eta^2} \mathbb{E}\left[\mathbb{E}\left[\left(f_X(X_j) - \widehat{f}_X(X_j)\right)^2 \mid \widetilde{X}_1, \dots, \widetilde{X}_m\right]\right] \\
&\leq \frac{4}{\eta^2} \mathbb{E}\left[\int_{\mathcal{D}} (f_X(x) - \widehat{f}_X(x))^2 f_X(x) dx\right] \\
&\leq \frac{4}{\eta^2} \|f_X\|_{\infty} \cdot \mathbb{E}\left[\int_{\mathcal{D}} (f_X(x) - \widehat{f}_X(x))^2 dx\right].
\end{aligned}$$

In the same way we get

$$\mathbb{E}[(f(X_j) - \widehat{f}_X(X_j))^2] \leq \|f_X\|_{\infty} \mathbb{E}\left[\int_{\mathcal{D}} (f_X(x) - \widehat{f}_X(x))^2 dx\right].$$

From [3, Theorem 4] and as $f_X \in \mathcal{C}^{\alpha}(\mathcal{D})$, it comes

$$\mathbb{E}\left[\int_{\mathcal{D}} (f_X(x) - \widehat{f}_X(x))^2 dx\right] \leq Cm^{\frac{-2\alpha}{2\alpha+d}}.$$

Thus, choosing the sample size m such that

$$m^{\frac{-2\alpha}{2\alpha+d}} = o\left(\frac{h_n^d}{n}\right),$$

the central limit theorem stated in Corollary 6.1 holds true.

7 Numerical applications

A Proof of the results

In the following, h_n and L_n are simply denoted by h and L respectively. In addition, recall that, we write C for a deterministic and finite constant which value is allowed to change between occurrences.

A.1 Proof of Theorem 3.1

Proof of (7). First, since $Y \in \mathbb{L}^2(\mathbb{R})$, one has

$$\begin{aligned}
\mathbb{E}[T_{n,h}] &= \frac{1}{2} \iint_{\mathcal{D}^2} \left(\frac{K_h \circ A_{x_1}(x_2 - x_1)}{f_X(x_1)} + \frac{K_h \circ A_{x_2}(x_1 - x_2)}{f_X(x_2)} \right) f_X(x_1) f_X(x_2) g_1(x_1) g_1(x_2) dx_1 dx_2 \\
&= \iint_{\mathcal{D}^2} \frac{K_h \circ A_{x_1}(x_2 - x_1)}{f_X(x_1)} f_X(x_1) f_X(x_2) g_1(x_1) g_1(x_2) dx_1 dx_2 \\
&= \int_{\mathcal{D}} \int_{\mathcal{D}_x} K(u) g_1(x) g_1(x + hA_x^{-1}(u)) f_X(x + hA_x^{-1}(u)) du dx
\end{aligned}$$

where the last display is obtained after the variable change $(x, u) = \left(x_1, A_{x_1}(x_2 - x_1)/h\right)$, using the fact that $|\det(A_x)| = 1$ and where $\mathcal{D} \times \mathcal{D}_x$ is the integration domain after the

variable change. Now observe that, for h small enough, $\mathcal{D} \subset \mathcal{D}_x$ by assumption and since $K(u) = 0$ for all $u \notin \mathcal{D}$, the integration domain of the last integral is reduced to \mathcal{D}^2 . In addition, since $\int_{\mathcal{D}} K(u) du = 1$ and $g_1(x) = \mathbb{E}[Y|X = x]$,

$$\mathbb{E}[\mathbb{E}[Y|X]^2] = \mathbb{E}[g_1^2(X)] = \int_{\mathcal{D}} g_1(x)^2 f_X(x) dx = \iint_{\mathcal{D}^2} g_1(x)^2 f_X(x) K(u) dx du$$

leading to

$$\begin{aligned} & \mathbb{E}[T_{n,h}] - \mathbb{E}[\mathbb{E}[Y|X]^2] \\ &= \iint_{\mathcal{D}^2} K(u) g_1(x) \left(g_1(x + hA_x^{-1}(u)) f_X(x + hA_x^{-1}(u)) - g_1(x) f_X(x) \right) dx du. \end{aligned} \quad (24)$$

The term in (24) can be handled with a Taylor expansion of $g_1 f_X$ (see, e.g., [7, Theorem 5.4]). The assumptions we use are (mainly) $g_1 f_X \in \mathcal{C}^\alpha(\mathcal{D})$ and, for any $1 \leq \beta \leq \lfloor \alpha \rfloor$,

$$\int_{\mathcal{D}} \left| g_1(x) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta} \right| dx < \infty.$$

Then we get a bound in $\mathcal{O}(h^\alpha)$. More precisely, by a multivariate Taylor expansion, since $g_1 f_X$ is $\mathcal{C}^\alpha(\mathcal{D})$, we get:

$$\begin{aligned} & g_1(x + uh) f_X(x + uh) - g_1(x) f_X(x) \\ &= \sum_{1 \leq |\beta| < \lfloor \alpha \rfloor} \frac{h^{|\beta|}}{\beta!} u^\beta \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) + h^{\lfloor \alpha \rfloor} \sum_{|\beta| = \lfloor \alpha \rfloor} R_\beta(x + uh) u^\beta \end{aligned} \quad (25)$$

with

$$R_\beta(x + uh) = \frac{\lfloor \alpha \rfloor}{\beta!} \int_0^1 (1-t)^{\lfloor \alpha \rfloor - 1} \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x + tuh) dt.$$

Then, one has

$$\begin{aligned} & \mathbb{E}[T_{n,h}] - \mathbb{E}[\mathbb{E}[Y|X]^2] \\ &= \sum_{1 \leq |\beta| < \lfloor \alpha \rfloor} \frac{h^{|\beta|}}{\beta!} \left(\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du \right) \left(\int_{\mathcal{D}} g_1(x) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) dx \right) \\ & \quad + h^{\lfloor \alpha \rfloor} \sum_{|\beta| = \lfloor \alpha \rfloor} \int_{\mathcal{D}} g_1(x) \int_{\mathcal{D}} u^\beta K(u) R_\beta(x + hA_x^{-1}(u)) du dx \\ &= \lfloor \alpha \rfloor h^{\lfloor \alpha \rfloor} \sum_{|\beta| = \lfloor \alpha \rfloor} \frac{1}{\beta!} \int_{\mathcal{D}} g_1(x) \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) \int_0^1 (1-t)^{\lfloor \alpha \rfloor - 1} \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x + thA_x^{-1}(u)) dt du dx \end{aligned}$$

using the fact that $(A_x^{-1}(u))^\beta$ is polynomial in (u_1, \dots, u_d) of degree β and K is of order $\lfloor \alpha \rfloor$ and that

$$\int_{\mathcal{D}} \left| g_1(x) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta} \right| dx < \infty$$

for any β such that $1 \leq |\beta| < \lfloor \alpha \rfloor$. Now, since K is of order $\lfloor \alpha \rfloor$ and

$$\int_{\mathcal{D}} u^\beta K(u) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) du = \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) \int_{\mathcal{D}} u^\beta K(u) du = 0$$

for any β such that $|\beta| = \lfloor \alpha \rfloor$, we get

$$\begin{aligned} \mathbb{E}[T_{n,h}] - \mathbb{E}[\mathbb{E}[Y|X]^2] &= \lfloor \alpha \rfloor h^{\lfloor \alpha \rfloor} \sum_{|\beta|=\lfloor \alpha \rfloor} \frac{1}{\beta!} \int_{\mathcal{D}} g_1(x) \int_{\mathcal{D}} (A_x^{-1}(u))^{\beta} K(u) \\ &\quad \int_0^1 (1-t)^{\lfloor \alpha \rfloor - 1} \left(\frac{\partial^{\beta}(g_1 f_X)}{\partial x^{\beta}}(x + thA_x^{-1}(u)) - \frac{\partial^{\beta}(g_1 f_X)}{\partial x^{\beta}}(x) \right) dt du dx. \end{aligned}$$

Then using the fact that $g_1 f_X \in \mathcal{C}^{\alpha}(\mathcal{D})$ and (5), one gets

$$\left| \frac{\partial^{\beta}(g_1 f_X)}{\partial x^{\beta}}(x + thA_x^{-1}(u)) - \frac{\partial^{\beta}(g_1 f_X)}{\partial x^{\beta}}(x) \right| \leq C_{g_1 f_X} (ht|A_x^{-1}(u)|)^{\alpha - \lfloor \alpha \rfloor} \quad (26)$$

for all u and $x \in \mathcal{D}$, for all $t \in [0, 1]$, for all $h \in (0, \infty)$, and for any $\beta \in \mathbb{N}^d$ such that $|\beta| = \lfloor \alpha \rfloor$. Then,

$$\begin{aligned} |\mathbb{E}[T_{n,h}] - \mathbb{E}[\mathbb{E}[Y|X]^2]| &\leq C_{g_1 f_X} \lfloor \alpha \rfloor h^{\alpha} \int_{\mathcal{D}} |g_1(x)| dx \\ &\quad \sum_{|\beta|=\lfloor \alpha \rfloor} \frac{1}{\beta!} \int_{\mathcal{D}} |A_x^{-1}(u)|^{\beta} |A_x^{-1}(u)|^{\alpha - \lfloor \alpha \rfloor} |K(u)| du \int_0^1 (1-t)^{\lfloor \alpha \rfloor - 1} t^{\alpha - \lfloor \alpha \rfloor} dt \\ &= Ch^{\alpha} \int_{\mathcal{D}} |g_1(x)| dx \times \left(\sum_{|\beta|=\lfloor \alpha \rfloor} \frac{1}{\beta!} \right) \times \left(\int_{\mathcal{D}} \|u\|_{\infty}^{\alpha} |K(u)| du \right) \\ &\leq Ch^{\alpha} \end{aligned}$$

since $\int_{\mathcal{D}} |g_1(x)| dx < \infty$ and

$$|A_x^{-1}(u)|^{\beta} = |u|^{\beta} = |u_1|^{\beta_1} \dots |u_d|^{\beta_d} \leq \|u\|_{\infty}^{\beta_1 + \dots + \beta_d} = \|u\|_{\infty}^{|\beta|} = \|u\|_{\infty}^{\lfloor \alpha \rfloor}. \quad \square$$

Proof of (8). Recall that, by (6), one has

$$T_{n,h} - \mathbb{E}[T_{n,h}] = \frac{1}{n} \sum_{j=1}^n Z_j + S_1 + S_2$$

where $Z_j = Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2]$ has been introduced in Theorem 3.1.

(1) We consider the second moment of S_1 . Since the pairs (X_j, Y_j) are i.i.d. distributed

as (X, Y) and the variables $Z_j = 2(Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2])$ are centered, one has

$$\begin{aligned}
n\mathbb{E}[S_1^2] &= 4n\mathbb{E}\left[\left(U_n^{(1)}(\pi_1 R) - \frac{1}{n} \sum_{j=1}^n (Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2])\right)^2\right] \\
&= 4n\text{Var}\left(U_n^{(1)}(\pi_1 R) - \frac{1}{n} \sum_{j=1}^n (Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2])\right) \\
&= 4n\text{Var}\left(\frac{1}{n} \sum_{j=1}^n \left[(\pi_1 R) \begin{pmatrix} X_j \\ Y_j \end{pmatrix} - (Y_j g_1(X_j) - \mathbb{E}[\mathbb{E}[Y|X]^2])\right]\right) \\
&= 4\text{Var}\left(\pi_1 R \begin{pmatrix} X \\ Y \end{pmatrix} - Y g_1(X)\right) \quad \text{by orthogonality of Hoeffding's decomposition} \\
&= 4\text{Var}\left(\frac{Y}{2f_X(X)} \int_{\mathcal{D}} (K_h \circ A_X)(u - X)(g_1 f_X)(u) du + \frac{Y}{2} \int_{\mathcal{D}} (K_h \circ A_u)(X - u)g_1(u) du - Y g_1(X)\right) \\
&\leq \mathbb{E}\left[Y^2 \left(\frac{1}{f_X(X)} \int_{\mathcal{D}} (K_h \circ A_X)(u - X)(g_1 f_X)(u) du + \int_{\mathcal{D}} (K_h \circ A_u)(X - u)g_1(u) du - 2g_1(X)\right)^2\right] \\
&\leq 2 \int_{\mathcal{D}} \left(\int_{\mathcal{D}} (K_h \circ A_x)(u - x)(g_1 f_X)(u) du - g_1(x) f_X(x)\right)^2 \frac{g_2(x)}{f_X(x)} dx \\
&\quad + 2 \int_{\mathcal{D}} \left(\int_{\mathcal{D}} (K_h \circ A_u)(x - u)g_1(u) du - g_1(x)\right)^2 g_2(x) f_X(x) dx \\
&=: 2(T_1 + T_2) \tag{27}
\end{aligned}$$

using $(a + b)^2 \leq 2(a^2 + b^2)$.

• We proceed following the same lines as in the proof of (7): using the variable change $(x, u) = (x_1, A_{x_1}(x_2 - x_1)/h)$, the fact that $\int_{\mathcal{D}} K(u) du = 1$, and the Taylor expansion of $g_1 f_X$ in (25). Then the first term can be rewritten as

$$\begin{aligned}
T_1 &= \int_{\mathcal{D}} \left(\int_{\mathcal{D}} K(u) g_1(x + hA_x^{-1}(u)) f_X(x + hA_x^{-1}(u)) du - g_1(x) f_X(x)\right)^2 \frac{g_2(x)}{f_X(x)} dx \\
&= \int_{\mathcal{D}} \left(\int_{\mathcal{D}} K(u) [g_1(x + hA_x^{-1}(u)) f_X(x + hA_x^{-1}(u)) - g_1(x) f_X(x)] du\right)^2 \frac{g_2(x)}{f_X(x)} dx \\
&= \int_{\mathcal{D}} \left(\sum_{1 \leq |\beta| < [\alpha]} \frac{h^{|\beta|}}{\beta!} \left(\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du\right) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x)\right. \\
&\quad \left.+ h^{[\alpha]} \sum_{|\beta| = [\alpha]} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) R_\beta(x + hA_x^{-1}(u)) du\right)^2 \frac{g_2(x)}{f_X(x)} dx \\
&\leq 2 \int_{\mathcal{D}} \left(\sum_{1 \leq |\beta| < [\alpha]} \frac{h^{|\beta|}}{\beta!} \left(\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du\right) \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x)\right)^2 \frac{g_2(x)}{f_X(x)} dx \\
&\quad + 2h^{2[\alpha]} \int_{\mathcal{D}} \left(\sum_{|\beta| = [\alpha]} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) R_\beta(x + hA_x^{-1}(u)) du\right)^2 \frac{g_2(x)}{f_X(x)} dx. \tag{28}
\end{aligned}$$

Since K is a kernel of order α and $(A_x^{-1}(u))^\beta$ is a polynomial in (u_1, \dots, u_d) of degree β , $\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du = 0$ for any β such that $|\beta| < \alpha$ and the first term of the previous

inequality equals to

$$\begin{aligned} & \sum_{1 \leq |\beta|, |\beta'| < [\alpha]} \frac{h^{|\beta|} h^{|\beta'|}}{\beta! \beta'!} \left(\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du \right) \left(\int_{\mathcal{D}} (A_x^{-1}(u'))^{\beta'} K(u') du' \right) \\ & \quad \times \int_{\mathcal{D}} \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) \frac{\partial^{\beta'} (g_1 f_X)}{\partial x^{\beta'}}(x) \frac{g_2(x)}{f_X(x)} dx \end{aligned} \quad (29)$$

vanishes as soon as

$$\int_{\mathcal{D}} \frac{g_2(x)}{f_X(x)} \left| \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta} \frac{\partial^{\beta'} (g_1 f_X)}{\partial x^{\beta'}} \right| dx < \infty$$

for any β and β' such that $1 \leq |\beta|, |\beta'| < [\alpha]$. Let us turn to the remaining term in (28). Using the definition of $R_\beta(x + hA_x^{-1}(u))$, we upper bound $T_1/(2h^{2[\alpha]})$ by

$$\begin{aligned} & \int_{\mathcal{D}} \left(\sum_{|\beta|=[\alpha]} \frac{1}{\beta!} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) \int_0^1 (1-t)^{[\alpha]-1} \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x + thA_x^{-1}(u)) dt du \right)^2 \frac{g_2(x)}{f_X(x)} dx \\ & = \int_{\mathcal{D}} \left(\sum_{|\beta|=[\alpha]} \frac{1}{\beta!} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) \int_0^1 (1-t)^{[\alpha]-1} \right. \\ & \quad \left. \times \left(\frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x - tuh) - \frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x) \right) dt du \right)^2 \frac{g_2(x)}{f_X(x)} dx. \end{aligned}$$

It suffices to proceed as previously in (29) rewriting $(\sum_\beta z_\beta)^2$ as $\sum_{\beta, \beta'} z_\beta z_{\beta'}$. As for the first term on the right hand side of the previous inequality, we apply (26) to get

$$\begin{aligned} |T_1| & \leq Ch^{2\alpha} \left(\int_{\mathcal{D}} \frac{g_2(x)}{f_X(x)} dx \right) \times \left(\sum_{|\beta|=[\alpha]} \frac{1}{\beta!} \right)^2 \times \left(\int_{\mathcal{D}} |A_x^{-1}(u)|^\beta |A_x^{-1}(u)|^{\alpha-[\alpha]} |K(u)| du \right)^2 \\ & \quad \times \left(\int_0^1 (1-t)^{[\alpha]-1} t^{\alpha-[\alpha]} dt \right)^2 \\ & = Ch^{2\alpha} \left(\int_{\mathcal{D}} \frac{g_2(x)}{f_X(x)} dx \right) \times \left(\sum_{|\beta|=[\alpha]} \frac{1}{\beta!} \right)^2 \times \left(\int_{\mathcal{D}} \|u\|_\infty^\alpha |K(u)| du \right)^2 \\ & \leq Ch^{2\alpha} \end{aligned} \quad (30)$$

since $g_2/f_X \in \mathbb{L}^1(\mathcal{D})$.

- Let us turn to T_2 . We proceed as for T_1 but performing a Taylor expansion of g_1 rather

than $g_1 f_X$.

$$\begin{aligned}
T_2 &= \int_{\mathcal{D}} \left(\int_{\mathcal{D}} (K_h \circ A_u)(x-u) g_1(u) du - g_1(x) \right)^2 g_2(x) f_X(x) dx \\
&= \int_{\mathcal{D}} \left(\int_{\mathcal{D}} K(u) [g_1(x + hA_x^{-1}(u)) - g_1(x)] du \right)^2 g_2(x) f_X(x) dx \\
&= \int_{\mathcal{D}} \left(\sum_{1 \leq |\beta| < [\alpha]} \frac{h^{|\beta|}}{\beta!} \left(\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du \right) \frac{\partial^\beta g_1}{\partial x^\beta}(x) \right. \\
&\quad \left. + h^{[\alpha]} \sum_{|\beta|=[\alpha]} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) R_\beta(x + hA_x^{-1}(u)) du \right)^2 g_2(x) f_X(x) dx \\
&\leq 2 \int_{\mathcal{D}} \left(\sum_{1 \leq |\beta| < [\alpha]} \frac{h^{|\beta|}}{\beta!} \left(\int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) du \right) \frac{\partial^\beta g_1}{\partial x^\beta}(x) \right)^2 g_2(x) f_X(x) dx \\
&\quad + 2h^{2[\alpha]} \int_{\mathcal{D}} \left(\sum_{|\beta|=[\alpha]} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) R_\beta(x + hA_x^{-1}(u)) du \right)^2 g_2(x) f_X(x) dx.
\end{aligned}$$

The first integral vanishes as soon as

$$\int_{\mathcal{D}} g_2(x) f_X(x) \left| \frac{\partial^\beta g_1}{\partial x^\beta} \frac{\partial^{\beta'} g_1}{\partial x^{\beta'}} \right| dx < \infty$$

for any β and β' such that $1 \leq |\beta|, |\beta'| < [\alpha]$. As for the second, one has

$$\begin{aligned}
T_2 &\leq 2h^{2[\alpha]} \int_{\mathcal{D}} \left(\sum_{|\beta|=[\alpha]} \frac{1}{\beta!} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) \int_0^1 (1-t)^{[\alpha]-1} \frac{\partial^\beta g_1}{\partial x^\beta}(x + thA_x^{-1}(u)) dt du \right)^2 g_2(x) f_X(x) dx \\
&= 2h^{2[\alpha]} \int_{\mathcal{D}} \left(\sum_{|\beta|=[\alpha]} \frac{1}{\beta!} \int_{\mathcal{D}} (A_x^{-1}(u))^\beta K(u) \int_0^1 (1-t)^{[\alpha]-1} \right. \\
&\quad \left. \times \left(\frac{\partial^\beta (g_1 f_X)}{\partial x^\beta}(x - tuh) - \frac{\partial^\beta g_1}{\partial x^\beta}(x) \right) dt du \right)^2 g_2(x) f_X(x) dx.
\end{aligned}$$

It suffices to proceed as previously in (30) to get

$$T_2 \leq Ch^{2\alpha}. \tag{31}$$

since $\int_{\mathcal{D}} g_2(x) f_X(x) dx = \mathbb{E}[g_2(X)] = \mathbb{E}[Y^2] < \infty$ by assumption on Y .

Finally,

$$n\mathbb{E}[S_1^2] \leq Ch^{2\alpha}. \tag{32}$$

(2) As for the second moment of S_2 , since π_2 is a projection, we have

$$\begin{aligned}
\mathbb{E}[S_2^2] &\leq \frac{1}{n(n-1)} \iint_{\mathcal{D}^2} \left(\frac{K_h^2 \circ A_{x_1}(x_2 - x_1)}{f_X(x_1)^2} + \frac{K_h^2 \circ A_{x_2}(x_1 - x_2)}{f_X(x_2)^2} \right) f_X(x_1) f_X(x_2) g_2(x_1) g_2(x_2) dx_1 dx_2 \\
&= \frac{2}{n(n-1)} \iint_{\mathcal{D}^2} K_h^2 \circ A_{x_1}(x_2 - x_1) \frac{f_X(x_2)}{f_X(x_1)} g_2(x_1) g_2(x_2) dx_1 dx_2 \\
&\leq \frac{2\kappa}{n(n-1)} \sup_{\ell=1, \dots, \kappa} \int_{\mathcal{D}} \int_{\mathcal{D}_\ell} K_h^2 \circ A_\ell(x_2 - x_1) \frac{f_X(x_2)}{f_X(x_1)} g_2(x_1) g_2(x_2) dx_1 dx_2 \\
&\leq \frac{2\kappa}{n(n-1)} \sup_{\ell=1, \dots, \kappa} \iint_{\mathcal{D}} K_h^2 \circ A_\ell(x_2 - x_1) \frac{f_X(x_2)}{f_X(x_1)} g_2(x_1) g_2(x_2) dx_1 dx_2 \\
&= \frac{2\kappa}{n(n-1)} \sup_{\ell=1, \dots, \kappa} \int_{\mathcal{D}} (K_h^2 \circ A_\ell) * \frac{g_2}{f_X}(y) g_2(y) f_X(y) dy \\
&\leq \frac{2\kappa}{n(n-1)} \|g_2 f_X\|_2 \left\| \frac{g_2}{f_X} \right\|_2 \sup_{\ell=1, \dots, \kappa} \|K_h^2 \circ A_\ell\|_1 \\
&= \frac{2\kappa}{n(n-1)h^d} \|g_2 f_X\|_2 \left\| \frac{g_2}{f_X} \right\|_2 \|K\|_1^2
\end{aligned}$$

using Young's convolution inequality again with $r = q = 2$ and $p = 1$. Since g_2/f_X and $g_2 f_X \in \mathbb{L}^2(\mathcal{D})$, it yields

$$\mathbb{E}[S_2^2] \leq \frac{C}{n(n-1)h^d}. \quad (33)$$

We conclude to (8) using (32) and (33). \square

Proof of (10). It is straightforward using (7), (8), and the standard central limit theorem for the random variables Z_i . \square

A.2 Proof of the remaining results

Proof of Corollary 3.3. To prove the global CLT, we first prove a CLT for the following linear combination

$$S_{n,h} := aT_{n,h} + b\bar{Y}_n + c\bar{Y}_n^2 = aT_{n,h} + \frac{b}{n} \sum_{j=1}^n Y_j + \frac{c}{n} \sum_{j=1}^n Y_j^2$$

for any a, b , and $c \in \mathbb{R}$ from which we deduce that $(T_{n,h}, \bar{Y}_n, \bar{Y}_n^2)$ is asymptotically Gaussian. Second, it suffices to conclude applying the delta method to the later vector

$$(T_{n,h}, \bar{Y}_n, \bar{Y}_n^2) = \left(T_{n,h}, \frac{1}{n} \sum_{j=1}^n Y_j, \frac{1}{n} \sum_{j=1}^n Y_j^2 \right)$$

with

$$\Phi(x, y, z) = \frac{x - y^2}{z - y^2}.$$

- We proceed as in the proof of Theorem 3.1: we control the bias and the variance terms as in (7) and (8) respectively. Then we conclude to the central limit theorem as in the proof of (10).

Let us first study the bias term. Let $m := a\mathbb{E}[\mathbb{E}[Y|X]^2] + b\mathbb{E}[Y] + c\mathbb{E}[Y^2]$. Then,

$$\begin{aligned}\mathbb{E}[S_{n,h}] - m &= \mathbb{E}[S_{n,h}] - \left(c\mathbb{E}[\mathbb{E}[Y|X]^2] + b\mathbb{E}[Y] + c\mathbb{E}[Y^2] \right) \\ &= a(\mathbb{E}[T_{n,h}] - \mathbb{E}[\mathbb{E}[Y|X]^2]) + b(\mathbb{E}[\bar{Y}_n] - \mathbb{E}[Y]) + c(\mathbb{E}[\bar{Y}_n^2] - \mathbb{E}[Y^2]) \\ &= a(\mathbb{E}[T_{n,h}] - \mathbb{E}[\mathbb{E}[Y|X]^2])\end{aligned}$$

and, by (7) in Theorem 3.1, we conclude that

$$\left| \mathbb{E}[S_{n,h}] - m \right| \leq aCh^\alpha.$$

As for the variance, the Hoeffding decomposition given in (6) becomes

$$\begin{aligned}S_{n,h} - \mathbb{E}[S_{n,h}] &= \frac{1}{n} \sum_{j=1}^n \left(aZ_j + b(Y_j - \mathbb{E}[Y]) + c(Y_j^2 - \mathbb{E}[Y^2]) \right) + a(S_1 + S_2) \\ &=: \frac{1}{n} \sum_{j=1}^n \tilde{Z}_j + a(S_1 + S_2),\end{aligned}$$

where Z_j has been defined in (9), S_1 and S_2 in (6). Then we deduce from (32) and (33) of Appendix A.1:

$$\mathbb{E} \left[\left(S_{n,h} - \mathbb{E}[S_{n,h}] - \frac{1}{n} \sum_{j=1}^n \tilde{Z}_j \right)^2 \right] = a^2 \mathbb{E} \left[\left(S_1 + S_2 \right)^2 \right] \leq a^2 C \left(h^{2\alpha} + \frac{1}{n^2 h^d} \right).$$

Thus it remains to use a CLT for $\frac{1}{n} \sum_{j=1}^n \tilde{Z}_j$. By Theorem 3.1 II(b), one gets merely directly

$$\sqrt{n} \left(S_{n,h} - m \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(0, \text{Var} \left(2aYg_1(X) + bY + cY^2 \right) \right)$$

from which we conclude that $\left(T_{n,h}, \bar{Y}_n, \bar{Y}_n^2 \right)$ is asymptotically Gaussian; namely

$$\sqrt{n} \left(\begin{pmatrix} T_{n,h} \\ \bar{Y}_n \\ \bar{Y}_n^2 \end{pmatrix} - M \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_3 \left(0, \Gamma \right),$$

where $M := \left(\mathbb{E}[\mathbb{E}[Y|X]^2], \mathbb{E}[Y], \mathbb{E}[Y^2] \right)^\top$ and

$$\Gamma := \begin{pmatrix} 4\tau^2 & 2\text{Cov}(Yg_1(X), Y) & 2\text{Cov}(Yg_1(X), Y^2) \\ 2\text{Cov}(Yg_1(X), Y) & \text{Var}(Y) & \text{Cov}(Y, Y^2) \\ 2\text{Cov}(Yg_1(X), Y^2) & \text{Cov}(Y, Y^2) & \text{Var}(Y^2) \end{pmatrix}.$$

- Now, as announced previously, we apply the delta method to $\left(T_{n,h}, \bar{Y}_n, \bar{Y}_n^2 \right)$ and

$$\Phi(x, y, z) = \frac{x - y^2}{z - y^2}$$

to get

$$\sqrt{n} \left(\frac{T_{n,h} - \left(\frac{1}{n} \sum_{j=1}^n Y_j \right)^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2 - \left(\frac{1}{n} \sum_{j=1}^n Y_j \right)^2} - S^X \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \ell^T \Gamma \ell$ with $\ell = \nabla \Phi(M)$. By assumption $\text{Var}(Y) \neq 0$, Φ is differentiable at M and we will see in the sequel that $\ell^T \Gamma \ell \neq 0$, so that the application of the delta method is justified. By differentiation, we get that, for any x, y , and z so that $z \neq y^2$:

$$\nabla \Phi(x, y, z) = \left(\frac{y^2}{z - y^2}, -2y \frac{z - x}{(z - y^2)^2}, -\frac{x - y^2}{(z - y^2)^2} \right)^T$$

so that

$$\begin{aligned} \ell = \nabla \Phi(M) &= \left(\frac{1}{\text{Var}(Y)}, -2\mathbb{E}[Y] \frac{\mathbb{E}[Y^2] - \mathbb{E}[\mathbb{E}[Y|X]^2]}{\text{Var}(Y)^2}, -\frac{S^X}{\text{Var}(Y)} \right)^T \\ &= \frac{1}{\text{Var}(Y)} \left(1, 2\mathbb{E}[Y](S^X - 1), -S^X \right)^T. \end{aligned}$$

Hence the asymptotic variance σ^2 can be computed

$$\begin{aligned} \sigma^2 &= \ell^T \Gamma \ell \\ &= \frac{1}{\text{Var}(Y)^2} \left(4 \left(\text{Var}(Y g_1(X)) - 2\text{Cov}(Y g_1(X), Y) \mathbb{E}[Y] + \mathbb{E}[Y]^2 \text{Var}(Y) \right) \right. \\ &\quad + 4S^X \left(2\text{Cov}(Y g_1(X), Y) \mathbb{E}[Y] - \text{Cov}(Y g_1(X), Y^2) - 2\mathbb{E}[Y]^2 \text{Var}(Y) + \mathbb{E}[Y] \text{Cov}(Y, Y^2) \right) \\ &\quad \left. + (S^X)^2 \left(4\mathbb{E}[Y]^2 \text{Var}(Y) - 4\mathbb{E}[Y] \text{Cov}(Y, Y^2) + \text{Var}(Y^2) \right) \right). \end{aligned} \quad (34)$$

The proof of Corollary 3.3 is now complete. \square

Proof of Corollary 3.4. For any $i = 1, \dots, p$, define $T_{n,h,i}$ as the estimator of $\mathbb{E}[\mathbb{E}[Y|V_i]^2]$. We proceed as for the proof of Corollary 3.3 to get a vectorial central limit theorem for $\left(T_{n,h,1}, \dots, T_{n,h,p}, \bar{Y}_n, \bar{Y}_n^2 \right)^T$. Let $M = \left(S^1, \dots, S^p, \mathbb{E}[Y], \mathbb{E}[Y^2] \right)^T$. Then, using the delta method with

$$\Phi(x_1, \dots, x_p, y, z) = \left(\frac{x_1 - y^2}{z - y^2}, \dots, \frac{x_p - y^2}{z - y^2} \right)^T$$

we get the result with

$$\Sigma = J_\Phi(M)^T \Gamma J_\Phi(M) \quad (35)$$

with $J_\Phi(M)$ the Jacobian of Φ at point M and Γ the covariance matrix of

$$\left(Y \mathbb{E}[Y|V_1], \dots, Y \mathbb{E}[Y|V_p], Y, Y^2 \right)^T. \quad \square$$

Proof of (17). Let $\tilde{X} = (\tilde{V}_1, \dots, \tilde{V}_d)$ and $x = (x_1, \dots, x_d)$. First of all, we compute

$$\begin{aligned}
\tilde{g}_1(x) &= \tilde{\mathbb{E}}[\tilde{g}(\tilde{V}_1, \dots, \tilde{V}_p) | \tilde{X} = x] \\
&= \frac{1}{n} \sum_{j=1}^n \tilde{\mathbb{E}} \left[\frac{Y_j}{f_{\tilde{V}}(\tilde{V}_1, \dots, \tilde{V}_p)} \prod_{i=1}^p \tilde{K}_{h_{0,i}}(V_{i,j} - \tilde{V}_i) \middle| \tilde{X} = x \right] \\
&= \frac{1}{n} \sum_{j=1}^n Y_j \prod_{i=1}^d \frac{\tilde{K}_{h_{0,i}}(V_{i,j} - v_i)}{f_{\tilde{V}_i}(v_i)} \prod_{i=d+1}^p \tilde{\mathbb{E}} \left[\frac{\tilde{K}_{h_{0,i}}(V_{i,j} - \tilde{V}_i)}{f_{\tilde{V}_i}(\tilde{V}_i)} \right] \\
&= \frac{1}{n} \frac{1}{f_{\tilde{X}}(x)} \sum_{j=1}^n Y_j \prod_{i=1}^d \tilde{K}_{h_{0,i}}(V_{i,j} - v_i) \prod_{i=d+1}^p \int_{B_i}^{C_i} \tilde{K}_{h_{0,i}}(V_{i,j} - v) dv \\
&= \frac{1}{n} \frac{1}{f_{\tilde{X}}(x)} \sum_{j=1}^n Y_j \prod_{i=1}^d \tilde{K}_{h_{0,i}}(V_{i,j} - v_i) \prod_{i=d+1}^p \tilde{\beta}_{i,j}
\end{aligned}$$

recalling that the support of V_i is $[B_i, C_i]$, and since the inputs are independent. Then

$$\begin{aligned}
\tilde{\mathbb{E}}[\tilde{g}_1(\tilde{X})^2] &= \tilde{\mathbb{E}} \left[\left(\frac{1}{n} \frac{1}{f_{\tilde{X}}(\tilde{X})} \sum_{j=1}^n Y_j \prod_{i=1}^d \tilde{K}_{h_{0,i}}(V_{i,j} - \tilde{V}_i) \prod_{i=d+1}^p \tilde{\beta}_{i,j} \right)^2 \right] \\
&= \frac{1}{n^2} \mathbb{E} \left[\frac{1}{f_{\tilde{X}}(\tilde{X})^2} \sum_{1 \leq j \leq j' \leq n} Y_j Y_{j'} \prod_{i=1}^d \tilde{K}_{h_{0,i}}(V_{i,j} - \tilde{V}_i) \tilde{K}_{h_{0,i}}(V_{i,j'} - \tilde{V}_i) \prod_{i=d+1}^p \tilde{\beta}_{i,j} \tilde{\beta}_{i,j'} \right] \\
&= \frac{1}{n^2} \sum_{1 \leq j \leq j' \leq n} Y_j Y_{j'} \left(\prod_{i=d+1}^p \tilde{\beta}_{i,j} \tilde{\beta}_{i,j'} \right) \int_{\mathcal{D}} \frac{1}{f_{\tilde{X}}(x)^2} \prod_{i=1}^d \tilde{K}_{h_{0,i}}(V_{i,j} - v_i) \tilde{K}_{h_{0,i}}(V_{i,j'} - v_i) f_{V_i}(v_i) dv_i \\
&= \frac{1}{n^2} \sum_{1 \leq j \leq j' \leq n} Y_j Y_{j'} \left(\prod_{i=d+1}^p \tilde{\beta}_{i,j} \tilde{\beta}_{i,j'} \right) \int_{\mathcal{D}} \frac{1}{f_{\tilde{X}}(x)} \prod_{i=1}^d \tilde{K}_{h_{0,i}}(V_{i,j} - v_i) \tilde{K}_{h_{0,i}}(V_{i,j'} - v_i) dv_i
\end{aligned}$$

recalling that $\mathcal{D} = [B_1, C_1] \times \dots \times [B_d, C_d]$. \square

Proof of Corollary 6.1. First, note that

$$\begin{aligned}
\sqrt{n}(\hat{T}_{n,h} - \mathbb{E}[\mathbb{E}[Y|X]^2]) &= \sqrt{n}(T_{n,h} - \mathbb{E}[\mathbb{E}[Y|X]^2]) + \frac{2\sqrt{n}}{n(n-1)} \sum_{1 \leq j < j' \leq n} \frac{Y_j Y_{j'}}{2} \\
&\quad \times \left(K_h \circ A_{X_j}(X_{j'} - X_j) \frac{f_X(X_j) - \hat{f}_X(X_j)}{\hat{f}_X(X_j) f_X(X_j)} + K_h \circ A_{X_{j'}}(X_j - X_{j'}) \frac{f_X(X_{j'}) - \hat{f}_X(X_{j'})}{\hat{f}_X(X_{j'}) f_X(X_{j'})} \right).
\end{aligned}$$

The first term in the right-hand side of the previous display is handled using Theorem 3.1. Thus it remains to prove that $\sqrt{n} \mathbb{E}[|\hat{T}_{n,h} - T_{n,h}|] \xrightarrow{n \rightarrow \infty} 0$. To do so, let us study the following expectation:

$$\begin{aligned}
&\mathbb{E} \left[\left| Y_j Y_{j'} \frac{K_h \circ A_{X_{j'}}(X_j - X_{j'})}{f_X(X_j)} \frac{f_X(X_j) - \hat{f}_X(X_j)}{\hat{f}_X(X_j)} \right|^2 \right] \\
&\leq \mathbb{E} \left[\left(Y_j Y_{j'} \frac{K_h \circ A_{X_{j'}}(X_j - X_{j'})}{f_X(X_j)} \right)^2 \right] \mathbb{E} \left[\left(\frac{f_X(X_j) - \hat{f}_X(X_j)}{\hat{f}_X(X_j)} \right)^2 \right].
\end{aligned}$$

One deduces from (33) that the first in the right-hand side of the previous display is $\mathcal{O}(1/h^d)$ and the proof is concluded using (18). \square

A.3 Construction of a kernel of order k in dimension d

In view of the assumptions in Section 2 on the kernel K that is required to be of order $\lfloor \alpha \rfloor$, we give in this section a procedure to construct kernels of a given order.

Let f_0 be a probability density function on \mathbb{R} and let ψ_0, \dots, ψ_k be the associated orthonormal polynomials: for any $m = 0, \dots, k$, ψ_m is a polynomial of order m and

$$\int \psi_m(x)\psi_{m'}(x)f_0(x)dx = \delta_{mm'}$$

for any m and $m' = 0, \dots, k$. For any monomial x^m , we define $\lambda^m = (\lambda_0^m, \dots, \lambda_k^m)$ as the coefficients of its decomposition in the f_0 -orthonormal basis (ψ_0, \dots, ψ_k) that is $x^m = \sum_{i=0}^k \lambda_i^m \psi_i(x)$. Now let $c \in \mathbb{R}^{k+1}$ solving the following linear system:

$$\sum_{i=0}^k \lambda_i^m c_i = \delta_{m0} \quad \text{for } m = 0, \dots, k.$$

We define $K_1(x) = \left(\sum_{i=0}^k \psi_i(x)c_i \right) f_0(x)$. Then, for any $m = 0, \dots, k$, one has

$$\begin{aligned} \int_{\mathbb{R}} x^m K_1(x) dx &= \int_{\mathbb{R}} \sum_{i=0}^k \lambda_i^m \psi_i(x) \sum_{i'=0}^k c_{i'} \psi_{i'}(x) f_0(x) dx \\ &= \sum_{i,i'=0}^k \lambda_i^m c_{i'} \int_{\mathbb{R}} \psi_i(x) \psi_{i'}(x) f_0(x) dx \\ &= \sum_{i=0}^k \lambda_i^m c_i \int_{\mathbb{R}} \psi_i(x)^2 f_0(x) dx = \sum_{i=0}^k \lambda_i^m c_i = \delta_{m0}. \end{aligned}$$

Thus K_1 is of order k as required and $\int_{\mathbb{R}} K_1(u) du = 1$ (case $m = 0$). To obtain a kernel of order k in dimension d , we proceed by tensorization defining:

$$K(x_1, \dots, x_d) = K_1(x_1) \dots K_1(x_d).$$

The tensorization allows to transfer the properties of K_1 to K .

References

- [1] A. Antoniadis. Analysis of variance on function spaces. Statistics: A Journal of Theoretical and Applied Statistics, 15(1):59–71, 1984.
- [2] M. Azadkia and S. Chatterjee. A simple measure of conditional dependence. arXiv preprint arXiv:1910.12327, 2019.
- [3] K. Bertin, N. Klutchnikoff, J. R. Léon, and C. Prieur. Adaptive density estimation on bounded domains under mixing conditions. Electronic Journal of Statistics, 14(1):2198 – 2237, 2020.
- [4] B. Broto, F. Bachoc, and M. Depecker. Variance reduction for estimation of shapley effects and adaptation to unknown input distribution. SIAM/ASA Journal on Uncertainty Quantification, 8(2):693–716, 2020.

- [5] M. Carrascal, M. Puigcerver, and P. Puig. Sensitivity of gaussian plume model to dispersion specifications. Theoretical and Applied Climatology, 48(2-3):147–157, 1993.
- [6] S. Chatterjee. A new coefficient of correlation. Journal of the American Statistical Association, pages 1–26, 2020.
- [7] R. Coleman. Calculus on normed vector spaces. Springer Science & Business Media, 2012.
- [8] R. Cukier, H. Levine, and K. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. Journal of computational physics, 26(1):1–42, 1978.
- [9] S. Da Veiga and F. Gamboa. Efficient estimation of sensitivity indices. Journal of Nonparametric Statistics, 25(3):573–595, 2013.
- [10] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. Basics and Trends in Sensitivity Analysis: Theory and Practice in R. SIAM, 2021.
- [11] S. Da Veiga, F. Wahl, and F. Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. Technometrics, 51(4):452–463, 2009.
- [12] B. Delyon and F. Portier. Integral approximation by kernel smoothing. Bernoulli, 22(4):2177–2208, 2016.
- [13] A. Dembo and O. Zeitouni. Large deviations techniques and applications, volume 38 of Applications of Mathematics (New York). Springer-Verlag, New York, second edition, 1998.
- [14] L. Devroye, P. G. Ferrario, L. Györfi, and H. Walk. Strong universal consistent estimate of the minimum mean squared error. Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, pages 143–160, 2013.
- [15] L. Devroye, L. Györfi, G. Lugosi, and H. Walk. A nearest neighbor estimate of the residual variance. Electronic Journal of Statistics, 12(1):1752–1778, 2018.
- [16] L. Devroye, D. Schäfer, L. Györfi, and H. Walk. The estimation problem of minimum mean squared error. Statistics & Decisions, 21(1):15–28, 2003.
- [17] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. Bernoulli, 28(4):2345–2374, 2022.
- [18] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical inference for Sobol Pick-Freeze Monte Carlo method. Statistics, 50(4):881–902, 2016.
- [19] F. Gamboa, T. Klein, and A. Lagnoux. Sensitivity analysis based on Cramér von Mises distance. SIAM/ASA Journal on Uncertainty Quantification, 6(2):522–548, Apr. 2018.
- [20] L. Gilquin, E. Arnaud, C. Prieur, and A. Janon. Making the best use of permutations to compute sensitivity indices with replicated orthogonal arrays. Reliability Engineering & System Safety, 187:28–39, 2019.

- [21] E. Giné and D. M. Mason. Uniform in bandwidth estimation of integral functionals of the density function. Scandinavian Journal of Statistics, 35(4):739–761, 2008.
- [22] E. Giné, R. Nickl, et al. A simple adaptive estimator of the integrated square of a density. Bernoulli, 14(1):47–61, 2008.
- [23] T. Goda. Computing the variance of a conditional expectation via non-nested Monte Carlo. Operations Research Letters, 45(1):63 – 67, 2017.
- [24] L. Györfi and H. Walk. On the asymptotic normality of an estimate of a regression functional. J. Mach. Learn. Res., 16:1863–1877, 2015.
- [25] M. B. Heredia, C. Prieur, and N. Eckert. Nonparametric estimation of aggregated sobol’indices: application to a depth averaged snow avalanche model. Reliability Engineering & System Safety, 212:107422, 2021.
- [26] W. Hoeffding. A class of statistics with asymptotically normal distribution. Ann. Math. Statistics, 19:293–325, 1948.
- [27] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two Sobol index estimators. ESAIM: Probability and Statistics, 18:342–364, 1 2014.
- [28] S. Kucherenko and S. Song. Different numerical estimators for main effect global sensitivity indices. Reliability Engineering & System Safety, 165:222–238, 2017.
- [29] E. Liitiäinen, F. Corona, and A. Lendasse. On nonparametric residual variance estimation. Neural Processing Letters, 28:155–167, 2008.
- [30] E. Liitiäinen, F. Corona, and A. Lendasse. Residual variance estimation using a nearest neighbor statistic. Journal of Multivariate Analysis, 101(4):811–823, 2010.
- [31] Z. Lin and F. Han. Limit theorems of chatterjee’s rank correlation. arXiv preprint arXiv:2204.08031, 2022.
- [32] J.-M. Loubes, C. Marteau, and M. Solís. Rates of convergence in conditional covariance matrix with nonparametric entries estimation. Communications in Statistics-Theory and Methods, 49(18):4536–4558, 2020.
- [33] A. B. Owen. Better estimation of small sobol’ sensitivity indices. ACM Trans. Model. Comput. Simul., 23(2):11:1–11:17, May 2013.
- [34] K. Pearson. On the partial correlation ratio. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 91(632):492–498, 1915.
- [35] V. H. Peña and E. Giné. Decoupling: From dependence to independence. Springer, 1999.
- [36] E. Plischke. An effective algorithm for computing global sensitivity indices (easi). Reliability Engineering & System Safety, 95(4):354–360, 2010.

- [37] E. Plischke and E. Borgonovo. Fighting the curse of sparsity: Probabilistic sensitivity measures from cumulative distribution functions. Risk Analysis, 40(12):2639–2660, 2020.
- [38] N. Rachdi, J.-C. Fort, and T. Klein. Stochastic inverse problem with noisy simulator-application to aeronautical model. Annales de la Faculté des Sciences de Toulouse, 6, 21:593–622, 2012.
- [39] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. Lo Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H. Maier. The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. Environmental Modelling and Software, 137:104954, Mar. 2021.
- [40] A. Saltelli, K. Chan, and E. Scott. Sensitivity analysis. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000.
- [41] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. Math. Modeling Comput. Experiment, 1(4):407–414 (1995), 1993.
- [42] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1-3):271–280, 2001.
- [43] M. Solís. Non-parametric estimation of the first-order sobol indices with bootstrap bandwidth. Communications in Statistics-Simulation and Computation, 50(9):2497–2512, 2021.
- [44] J. M. Stockie. The mathematics of atmospheric dispersion modeling. Siam Review, 53(2):349–372, 2011.
- [45] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety, 93(7):964–979, 2008.
- [46] S. Tarantola, D. Gatelli, and T. Mara. Random balance designs for the estimation of first order global sensitivity indices. Reliability Engineering & System Safety, 91(6):717–727, 2006.
- [47] J.-Y. Tissot and C. Prieur. A randomized orthogonal array-based procedure for the estimation of first-and second-order sobol’indices. Journal of Statistical Computation and Simulation, 85(7):1358–1381, 2015.
- [48] A. W. van der Vaart. Asymptotic statistics, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.
- [49] L.-X. Zhu and K.-T. Fang. Asymptotics for kernel estimate of sliced inverse regression. The Annals of Statistics, 24(3):1053–1068, 1996.