



PROJECTION PURSUIT THROUGH RELATIVE ENTROPY MINIMISATION

Jacques Touboul

► To cite this version:

Jacques Touboul. PROJECTION PURSUIT THROUGH RELATIVE ENTROPY MINIMISATION.
2008. hal-00429675

HAL Id: hal-00429675

<https://hal.science/hal-00429675>

Preprint submitted on 15 Dec 2009

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.

PROJECTION PURSUIT THROUGH RELATIVE ENTROPY MINIMISATION

Jacques Touboul

Laboratoire de Statistique Théorique et Appliquée

Université Pierre et Marie Curie

jack_touboul@hotmail.com

Consider a defined density on a set of very large dimension. It is quite difficult to find an estimate of this density from a data set. However, it is possible through a projection pursuit methodology to solve this problem. In his seminal article, Huber (see "Projection pursuit", Annals of Statistics, 1985) demonstrates the interest of his method in a very simple given case. He considers the factorization of density through a Gaussian component and some residual density. Huber's work is based on maximizing relative entropy. Our proposal leads to a new algorithm. Furthermore, we consider the case when the density to be factorized is estimated from an i.i.d. sample. In this case, we will propose a test for the factorization of the estimated density.

Keywords: projection pursuit; minimum relative entropy; maximum relative entropy; elliptic distribution; rate of convergence.

MSC: 62H40 62F03 62F10 62F30.

1 Outline of the article

The objective of Projection Pursuit is to generate one or several projections providing as much information as possible about the structure of the data set regardless of its size:

Once a structure has been isolated, the corresponding data are eliminated from the data set. Through a recursive approach, this process is iterated to find another structure in the remaining data, until no further structure can be evidenced in the data left at the end.

Friedman [Frie8487] and Huber [HUB85] count among the first authors to have introduced this type of approaches for evidencing structures. They each describe, with many examples, how to evidence such a structure and consequently how to estimate the density of such data through two different methodologies each. For a very long time, the two methodologies exposed by each of the above authors were thought to be equivalent but Mu Zhu [ZMU04] showed it was in fact not the case when the number of iterations in the algorithms exceeds the dimension of the space containing the data. In the present article, we will therefore only focus on Huber's study while taking into account Mu Zhu remarks.

At present, let us briefly introduce Huber's methodologies. We will then expose our approach and objective.

1.1 Huber's analytic approach

Let f be a density on \mathbb{R}^d . We define an instrumental density g with same mean and variance as f . Huber's methodology requires to start with performing the $K(f, g) = 0$ test. Should this test turn out to be positive, then $f = g$ and the algorithm stops. If the test were not to be verified, the first

step of Huber's algorithm amounts to defining a vector a_1 and a density $f^{(1)}$ by

$$a_1 = \arg \inf_{a \in \mathbb{R}_*^d} K(f \frac{g_a}{f_a}, g) \text{ and } f^{(1)} = f \frac{g_{a_1}}{f_{a_1}}, \quad (1)$$

where \mathbb{R}_*^d is the set of non null vectors of \mathbb{R}^d , where f_a (resp. g_a) stands for the density of $a^\top X$ (resp. $a^\top Y$) when f (resp. g) is the density of X (resp. Y). More exactly, this results from the maximisation of $a \mapsto K(f_a, g_a)$ since $K(f, g) = K(f_a, g_a) + K(f \frac{g_a}{f_a}, g)$ and it is assumed that $K(f, g)$ is finite. In a second step, Huber replaces f with $f^{(1)}$ and goes through the first step again. By iterating this process, Huber thus obtains a sequence (a_1, a_2, \dots) of vectors of \mathbb{R}_*^d and a sequence of densities $f^{(i)}$.

Remark 1. Huber stops his algorithm when the relative entropy equals zero or when his algorithm reaches the d^{th} iteration, he then obtains an approximation of f from g :

When there exists an integer j such that $K(f^{(j)}, g) = 0$ with $j \leq d$, he obtains $f^{(j)} = g$, i.e. $f = g \prod_{i=1}^j \frac{f_{a_i}^{(i-1)}}{g_{a_i}}$ since by induction $f^{(j)} = f \prod_{i=1}^j \frac{g_{a_i}}{f_{a_i}^{(i-1)}}$. Similarly, when, for all j , Huber gets

$K(f^{(j)}, g) > 0$ with $j \leq d$, he assumes $g = f^{(d)}$ in order to derive $f = g \prod_{i=1}^d \frac{f_{a_i}^{(i-1)}}{g_{a_i}}$.

He can also stop his algorithm when the relative entropy equals zero without the condition $j \leq d$ is met. Therefore, since by induction we have $f^{(j)} = f \prod_{i=1}^j \frac{g_{a_i}}{f_{a_i}^{(i-1)}}$ with $f^{(0)} = f$, we obtain $g = f \prod_{i=1}^j \frac{g_{a_i}}{f_{a_i}^{(i-1)}}$. Consequently, we derive a representation of f as $f = g \prod_{i=1}^j \frac{f_{a_i}^{(i-1)}}{g_{a_i}}$.

Finally, he obtains $K(f^{(0)}, g) \geq K(f^{(1)}, g) \geq \dots \geq 0$ with $f^{(0)} = f$.

At present, let us illustrate this through the following example:

Example 1 (From Huber's article). Let (X_1, \dots, X_d) be a random vector with density function f . We want to approximate f with a product of Gaussian densities. Let g be a Gaussian density on \mathbb{R}^d with same mean and variance as f . In the sequel, for all $a \in \mathbb{R}_*^d$, we will call f_a the density of $a^\top X$ and f_1, f_2, \dots, f_d the marginals of f . Thus, let us consider the two following hypotheses:

- Let us assume that the X_1, \dots, X_d are independent, i.e. $f(x) = \prod_{j=1}^d f_j(x_j)$, and
- $K(f_1, g_1) \geq K(f_2, g_2) \geq \dots \geq K(f_d, g_d) \geq 0$.

We assume the relative entropy between the marginals of identical ranking are well-ordered. Consequently, g_d and f_d are the closest densities and g_1 and f_1 the least similar densities.

Thus, on page 458 of [HUB85], Huber first shows that, for all b in \mathbb{R}_*^d , $K(f_b, g_b) \leq K(f_1, g_1)$. He sets $f^{(1)}(x) = f(x) \frac{g_1(x_1)}{f_1(x_1)}$, he therefore gets $a_1 = (1, 0, \dots, 0)$, $f^{(1)}(x) = \prod_{j=1}^d f_j^{(1)}(x_j)$ - where $f_1^{(1)} = g_1$ and $f_j^{(1)} = f_j$ for $j = 2, \dots, d$ - and $K(f, g) = K(f_1, g_1) + K(f^{(1)}, g)$ through relationship (1).

In the second step, Huber shows that, for all b in \mathbb{R}_*^d , $K([f^{(1)}]_b, g_b) \leq K(f_2, g_2)$. Hence, he gets $a_2 = (0, 1, 0, 0, \dots, 0)$, $f^{(2)}(x) = \prod_{j=1}^d f_j^{(2)}(x_j)$ - where $f_1^{(2)} = g_1$, $f_2^{(2)} = g_2$ and $f_j^{(2)} = f_j$ for $j = 3, \dots, d$ - as well as $K(f^{(1)}, g) = K(f_2^{(1)}, g_2) + K(f^{(2)}, g)$, i.e. $K(f, g) = K(f_1, g_1) + K(f_2, g_2) + K(f^{(2)}, g)$, since $f_2^{(1)} = f_2$. And so on. Huber shows that, for all b in \mathbb{R}_*^d , $K([f^{(d-1)}]_b, g_b) \leq K(f_d, g_d)$.

Therefore, he obtains :

- $a_d = (0, 0, 0, 0, \dots, 1)$,
- $f^{(d)}(x) = \prod_{j=1}^d f_j^{(d)}(x_j)$, where $f_d^{(d)} = g_d$ and $f_j^{(d)} = g_j$ for $j = 1, \dots, d-1$, i.e. $f^{(d)} = g$, and
- $K(f^{(d-1)}, g) = K(f_d^{(d-1)}, g_d) + K(f^{(d)}, g) = K(f_d, g_d)$ since $f^{(d)} = g$ implies $K(f^{(d)}, g) = 0$ and since $f_d^{(d-1)} = f_d$, i.e. $K(f, g) = K(f_1, g_1) + K(f_2, g_2) + \dots + K(f_d, g_d)$.

We therefore obtain $f = g \prod_{j=1}^d \frac{f_j}{g_j}$, which is the representation of f we have been looking for.

1.2 Huber's synthetic approach

Keeping the notations of the above section, we start with performing the $K(f, g) = 0$ test; should this test turn out to be positive, then $f = g$ and the algorithm stops, otherwise, the first step of his algorithm would consist in defining a vector a_1 and a density $g^{(1)}$ by

$$a_1 = \arg \inf_{a \in \mathbb{R}_*^d} K(f, g \frac{f_a}{g_a}) \text{ and } g^{(1)} = g \frac{f_{a_1}}{g_{a_1}}. \quad (2)$$

More exactly, this optimisation results from the maximisation of $a \mapsto K(f_a, g_a)$ since $K(f, g) = K(f_a, g_a) + K(f, g \frac{f_a}{g_a})$ and it is assumed that $K(f, g)$ is finite. In a second step, Huber replaces g with $g^{(1)}$ and goes through the first step again. By iterating this process, Huber thus obtains a sequence (a_1, a_2, \dots) of vectors of \mathbb{R}_*^d and a sequence of densities $g^{(i)}$.

Remark 2. First, in a similar manner to the analytic approach, this methodology enables us to approximate and even to represent f from g :

To obtain an approximation of f , Huber either stops his algorithm when the relative entropy equals zero, i.e. $K(f, g^{(j)}) = 0$ implies $g^{(j)} = f$ with $j \leq d$, or when his algorithm reaches the d^{th} iteration, i.e. he approximates f with $g^{(d)}$.

To obtain a representation of f , Huber stops his algorithm when the relative entropy equals zero, since $K(f, g^{(j)}) = 0$ implies $g^{(j)} = f$. Therefore, since by induction we have $g^{(j)} = g \prod_{i=1}^j \frac{f_{a_i}}{g_{a_i}}$ with $g^{(0)} = g$, we then obtain $f = g \prod_{i=1}^j \frac{f_{a_i}}{g_{a_i}}$.

Second, he gets $K(f, g^{(0)}) \geq K(f, g^{(1)}) \geq \dots \geq 0$ with $g^{(0)} = g$.

Finally, in [ZMU04], Mu Zhu shows that, even if the above two algorithms are exclusively based on the maximisation of the relative entropy $K(f_a, g_a)$, beyond d iterations, the data processing of these methodologies evidences significant differences, i.e. that past d iterations, the two methodologies are no longer equivalent. We will therefore only consider Huber's synthetic approach since g is known and since we want to find a representation of f .

1.3 Proposal

We start with performing the $K(f, g) = 0$ test; should this test turn out to be positive, then $f = g$ and the algorithm stops, otherwise, the first step of our algorithm would consist in defining a vector a_1 and a density $g^{(1)}$ by

$$a_1 = \arg \inf_{a \in \mathbb{R}_*^d} K(g \frac{f_a}{g_a}, f) \text{ and } g^{(1)} = g \frac{f_{a_1}}{g_{a_1}}. \quad (3)$$

In the second step, we will replace g with $g^{(1)}$, and we will repeat the first step. And so on, by iterating this process, we will end up obtaining a sequence (a_1, a_2, \dots) of vectors in \mathbb{R}_*^d and a sequence of densities $g^{(i)}$. We will prove that a_1 simultaneously optimises (1), (2) and (3). We will also prove that the underlying structures of f evidenced through this method are identical to the ones obtained through Huber's methods.

Remark 3. As in Huber's algorithms, we perform an approximation and a representation of f :

To obtain an approximation of f , we stop the algorithm when the relative entropy equals zero, i.e. we approximate f with $g^{(j)}$, or when the algorithm reaches the d^{th} iteration, i.e. we approximate f

with $g^{(d)}$.

To obtain a representation of f , we stop the algorithm when the relative entropy equals zero. Therefore, since by induction we have $g^{(j)} = g \prod_{i=1}^j \frac{f_{a_i}}{g_{a_i}^{(i-1)}}$ with $g^{(0)} = g$, we then obtain $f = g \prod_{i=1}^j \frac{f_{a_i}}{g_{a_i}^{(i-1)}}$.

Finally, we have $K(g^{(0)}, f) \geq K(g^{(1)}, f) \geq \dots \geq 0$ with $g^{(0)} = g$.

Let us study the following example:

Example 2. Let f be a density defined on \mathbb{R}^3 by $f(x_1, x_2, x_3) = n(x_1, x_2)h(x_3)$, with n being a bi-dimensional Gaussian density, and h being a non Gaussian density. Let us also consider g , a Gaussian density with same mean and variance as f .

The function $a \mapsto K(g \frac{f_a}{g_a}, f)$ reaches zero for $e_3 = (0, 0, 1)'$.

Indeed, since $g(x_1, x_2/x_3) = n(x_1, x_2)$, we have $K(g \frac{f_3}{g_3}, f) = K(n.f_3, f) = K(f, f) = 0$ as $f_3 = h$. We therefore obtain that $g(x_1, x_2/x_3) = f(x_1, x_2/x_3)$.

To recapitulate our method, if $K(g, f) = 0$, we derive f from the relationship $f = g$; should a sequence $(a_i)_{i=1, \dots, j}$, $j < d$, of vectors in \mathbb{R}_*^d defining $g^{(j)}$ and such that $K(g^{(j)}, f) = 0$ exist, then $f(\cdot/a_i^\top x, 1 \leq i \leq j) = g(\cdot/a_i^\top x, 1 \leq i \leq j)$, i.e. f coincides with g on the complement of the vector subspace generated by the family $\{a_i\}_{i=1, \dots, j}$ - see also section 2.1.2 for more detailed explanations.

In this paper, after having clarified the choice of g , we will consider the statistical solution to the representation problem, assuming that f is unknown and X_1, X_2, \dots, X_m are i.i.d. with density f . We will provide asymptotic results pertaining to the family of optimizing vectors $a_{k,m}$ - that we will define more precisely below - as m goes to infinity. Our results also prove that the empirical representation scheme converges towards the theoretical one. As an application, section 3.4 permits a new test of fit pertaining to the copula of an unknown density f and section 3.5 gives us an estimate of a density deconvoluted with a Gaussian component. Finally, we will compare Huber's optimisation methods with ours and we will present simulations.

2 The algorithm

2.1 The model

As explained by Friedman [Frie8487], Diaconis [DIAFREE84] and Hwang [HLL94], the choice of g depends on the family of distribution one wants to find in f . Until now, the choice has been only to use the class of Gaussian distributions. This can be extended to the class of elliptic distributions.

2.1.1 Elliptic laws

The interest of this class is that it is such that conditional densities with elliptic distributions are also elliptic - see [CAMBANIS81], [LANDS03]. This very property allows us to use this class in our algorithm - and in Huber's algorithms.

Definition 1. X is said to abide by a multivariate elliptic distribution - noted $X \sim E_d(\mu, \Sigma, \xi_d)$ - if X presents the following density, for any x in \mathbb{R}^d :

$$f_X(x) = \frac{c_d}{|\Sigma|^{1/2}} \xi_d \left(\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right)$$

- with Σ , being a $d \times d$ positive-definite matrix and with μ , being an d -column vector,
- with ξ_d , being referred as the "density generator",

- with c_d , being a normalisation constant, such that $c_d = \frac{\Gamma(d/2)}{(2\pi)^{d/2}} \left(\int_0^\infty x^{d/2-1} \xi_d(x) dx \right)^{-1}$, with $\int_0^\infty x^{d/2-1} \xi_d(x) dx < \infty$.

Property 1. 1/ For any $X \sim E_d(\mu, \Sigma, \xi_d)$, for any A , being a $m \times d$ matrix with rank $m \leq d$, and for any b , being an m -dimensional vector, we have $AX + b \sim E_m(A\mu + b, A\Sigma A', \xi_m)$.

Therefore, any marginal density of multivariate elliptic distribution is elliptic, i.e.

$$X = (X_1, X_2, \dots, X_d) \sim E_d(\mu, \Sigma, \xi_d) \Rightarrow X_i \sim E_1(\mu_i, \sigma_i^2, \xi_1), f_{X_i}(x) = \frac{c_1}{\sigma_i} \xi_1\left(\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right), 1 \leq i \leq d.$$

2/ Corollary 5 of [CAMBANIS81] states that conditional densities with elliptic distributions are also elliptic. Indeed, if $X = (X_1, X_2)' \sim E_d(\mu, \Sigma, \xi_d)$, with X_1 (resp. X_2) being a size $d_1 < d$ (resp. $d_2 < d$), then $X_1/(X_2 = a) \sim E_{d_1}(\mu', \Sigma', \xi_{d_1})$ with $\mu' = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$ and $\Sigma' = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, with $\mu = (\mu_1, \mu_2)$ and $\Sigma = (\Sigma_{ij})_{1 \leq i, j \leq 2}$.

Remark 4. In [LANDS03], the authors show that the multivariate Gaussian distribution derives from $\xi_d(x) = e^{-x}$. They also show that if $X = (X_1, \dots, X_d)$ has an elliptic density such that its marginals verify $E(X_i) < \infty$ and $E(X_i^2) < \infty$ for $1 \leq i \leq d$, then μ is the mean of X and Σ is the covariance matrix of X . Consequently, from now on, we will assume that we are in this case.

Definition 2. Let t be an elliptic density on \mathbb{R}^k and let q be an elliptic density on $\mathbb{R}^{k'}$. The elliptic densities t and q are said to belong to the same family - or class - of elliptic densities, if their generating densities are ξ_k and $\xi_{k'}$ respectively, which belong to a common given family of densities.

Example 3. Consider two Gaussian densities $\mathcal{N}(0, 1)$ and $\mathcal{N}((0, 0), Id_2)$. They are said to belong to the same elliptic families as they both present $x \mapsto e^{-x}$ as generating density.

2.1.2 Choice of g

Let us begin with studying the following case:

Let f be a density on \mathbb{R}^d . Let us assume there exists d not null independent vectors a_j , with $1 \leq j \leq d$, of \mathbb{R}^d , such that

$$f(x) = n(a_{j+1}^\top x, \dots, a_d^\top x) h(a_1^\top x, \dots, a_j^\top x), \quad (4)$$

with $j < d$, with n being an elliptic density on \mathbb{R}^{d-j-1} and with h being a density on \mathbb{R}^j , which does not belong to the same family as n . Let $X = (X_1, \dots, X_d)$ be a vector presenting f as density. Define g as an Elliptic distribution with same mean and variance as f .

For simplicity, let us assume that the family $\{a_j\}_{1 \leq j \leq d}$ is the canonical basis of \mathbb{R}^d :

The very definition of f implies that (X_{j+1}, \dots, X_d) is independent from (X_1, \dots, X_j) . Hence, property 1 enables us to derive that the density of (X_{j+1}, \dots, X_d) given (X_1, \dots, X_j) is n .

Let us assume that $K(g^{(j)}, f) = 0$, for some $j \leq d$. We then get $\frac{f(x)}{f_{a_1} f_{a_2} \dots f_{a_j}} = \frac{g(x)}{g_{a_1}^{(1-1)} g_{a_2}^{(2-1)} \dots g_{a_j}^{(j-1)}}$, since,

$$\text{by induction, we have } g^{(j)}(x) = g(x) \frac{f_{a_1}}{g_{a_1}^{(1-1)}} \frac{f_{a_2}}{g_{a_2}^{(2-1)}} \dots \frac{f_{a_j}}{g_{a_j}^{(j-1)}}.$$

Consequently, the fact that conditional densities with elliptic distributions are also elliptic as well as the above relationship enable us to state that

$$n(a_{j+1}^\top x, \dots, a_d^\top x) = f(\cdot / a_i^\top x, 1 \leq i \leq j) = g(\cdot / a_i^\top x, 1 \leq i \leq j).$$

In other words, f coincides with g on the complement of the vector subspace generated by the family $\{a_i\}_{i=1, \dots, j}$.

At present, if the family $\{a_j\}_{1 \leq j \leq d}$ is no longer the canonical basis of \mathbb{R}^d , then this family is again a basis of \mathbb{R}^d . Hence, lemma 7 - page 23 - implies that

$$g(./a_1^\top x, \dots, a_j^\top x) = n(a_{j+1}^\top x, \dots, a_d^\top x) = f(./a_1^\top x, \dots, a_j^\top x), \quad (5)$$

which is equivalent to having $K(g^{(j)}, f) = 0$ - since by induction $g^{(j)} = g \frac{f_{a_1}}{g_{a_1}^{(1-1)}} \frac{f_{a_2}}{g_{a_2}^{(2-1)}} \dots \frac{f_{a_j}}{g_{a_j}^{(j-1)}}$.

The end of our algorithm implies that f coincides with g on the complement of the vector subspace generated by the family $\{a_i\}_{i=1, \dots, j}$. Therefore, the nullity of the relative entropy provides us with information on the density structure.

In summary, the following proposition clarifies the choice of g which depends on the family of distribution one wants to find in f :

Proposition 1. *With the above notations, $K(g^{(j)}, f) = 0$ is equivalent to*

$$g(./a_1^\top x, \dots, a_j^\top x) = f(./a_1^\top x, \dots, a_j^\top x)$$

More generally, the above proposition leads us to defining the co-support of f as the vector space generated from vectors a_1, \dots, a_j .

Definition 3. *Let f be a density on \mathbb{R}^d . We define the co-vectors of f as the sequence of vectors a_1, \dots, a_j which solves the problem $K(g^{(j)}, f) = 0$ where g is an Elliptic distribution with same mean and variance as f . We define the co-support of f as the vector space generated from vectors a_1, \dots, a_j .*

2.2 Stochastic outline of the algorithm

Let X_1, X_2, \dots, X_m (resp. Y_1, Y_2, \dots, Y_m) be a sequence of m independent random vectors with same density f (resp. g). As customary in nonparametric Kullback-Lieber optimizations, all estimates of f and f_a (resp. g and g_a) are being performed using a subsample X_1, X_2, \dots, X_n (resp. Y_1, Y_2, \dots, Y_n) extracted from X_1, X_2, \dots, X_m (resp. Y_1, Y_2, \dots, Y_m) such that the estimates are bounded below by some positive deterministic sequence θ_m - see Annex B. Let \mathbb{P}_n be the empirical measure of the subsample X_1, X_2, \dots, X_n . Let f_n (resp. $g_n, f_{a,n}, g_{a,n}$ for any a in \mathbb{R}_*^d) be the kernel estimate of f (resp. g, f_a, g_a) which is built from X_1, X_2, \dots, X_n (resp. Y_1, Y_2, \dots, Y_n).

As defined in section 1.3, we introduce the following sequences $(a_k)_{k \geq 1}$ and $(g^{(k)})_{k \geq 1}$:

- a_k is a non null vector of \mathbb{R}^d such that $a_k = \arg \min_{a \in \mathbb{R}_*^d} K(g^{(k-1)} \frac{f_a}{g_a^{(k-1)}}, f)$,
- $g^{(k)}$ is the density such that $g^{(k)} = g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}$ with $g^{(0)} = g$.

The stochastic setting up of the algorithm uses f_n and g_n instead of f and g .

Thus, at the first step, we build the vector \tilde{a}_1 which minimizes the relative entropy between f_n and $g_n \frac{f_{\tilde{a}_1, n}}{g_{\tilde{a}_1, n}}$ and which estimates a_1 :

Proposition 11 page 19 and lemma 11 page 24 enables us to minimize the relative entropy between f_n and $g_n \frac{f_{\tilde{a}_1, n}}{g_{\tilde{a}_1, n}}$. Defining \tilde{a}_1 as the argument of this minimization, proposition 4 page 8 shows us that this vector tends to a_1 in n . Finally, we define the density $\check{g}^{(1)}$ as $\check{g}^{(1)} = g_n \frac{f_{\tilde{a}_1, n}}{g_{\tilde{a}_1, n}}$ which estimates $g^{(1)}$ through theorem 1.

At the second step, we build the vector \tilde{a}_2 which minimizes the relative entropy between f_n and $\check{g}^{(1)} \frac{f_{\tilde{a}_2, n}}{\check{g}_{\tilde{a}_2, n}^{(1)}}$ and which estimates a_2 . Since $g_a^{(1)}$ can generally not be determined analytically, let us define the following procedure:

- We generate $Y_1^{(1)}, Y_2^{(1)}, \dots, Y_m^{(1)}$ with density $g^{(1)}$,

- We truncate this sample as we had truncated Y_1, Y_2, \dots, Y_m in Annex B,
- We generate the kernel estimate of the univariate density $\check{g}_a^{(1)}$.

Consequently, proposition 11 page 19 and lemma 11 page 24 enable us to minimize the relative entropy between f_n and $\check{g}^{(1)} \frac{f_{a,n}}{\check{g}_{a,n}^{(1)}}$. Defining \check{a}_2 as the argument of this minimization, proposition 4 page 8 shows us that this vector tends to a_2 as n tends to infinity. Finally, we define the density $\check{g}^{(2)}$ as $\check{g}^{(2)} = \check{g}^{(1)} \frac{f_{\check{a}_2,n}}{\check{g}_{\check{a}_2,n}^{(1)}}$ which estimates $g^{(2)}$ through theorem 1.

And so on, we will end up obtaining a sequence $(\check{a}_1, \check{a}_2, \dots)$ of vectors in \mathbb{R}_*^d estimating the co-vectors of f and a sequence of densities $(\check{g}^{(k)})_k$ such that $\check{g}^{(k)}$ estimates $g^{(k)}$ through theorem 1.

3 Results

3.1 Convergence results

3.1.1 Hypotheses on f

In this paragraph, we define the set of hypotheses on f which we can possibly use in our work. Discussion of several on these hypotheses can be found in Annex D.

In this section, to be more legible we replace g with $g^{(k-1)}$. Let

$$\Theta = \mathbb{R}_*^d, \quad M(b, a, x) = \int \ln\left(\frac{g(x)}{f(x)} \frac{f_b(b^\top x)}{g_b(b^\top x)}\right) g(x) \frac{f_a(a^\top x)}{g_a(a^\top x)} dx - \left(\frac{g(x)}{f(x)} \frac{f_b(b^\top x)}{g_b(b^\top x)} - 1\right),$$

$$\mathbb{P}_n M(b, a) = \int M(b, a, x) d\mathbb{P}_n, \quad \mathbf{P} M(b, a) = \int M(b, a, x) f(x) dx,$$

where \mathbf{P} being the probability measure of f .

Similarly as in chapter V of [VDW], we define :

(H'1) : For all $\varepsilon > 0$, there is $\eta > 0$, such that for all $c \in \Theta$ verifying $\|c - a_k\| \geq \varepsilon$, we have $\mathbf{P} M(c, a) < \mathbf{P} M(a_k, a) - \eta$, with $a \in \Theta$.

(H'2) : There is a neighborhood of a_k , V , and a positive function H , such that, for all $c \in V$ we have $|M(c, a_k, x)| \leq H(x)$ (\mathbf{P} - a.s.) with $\mathbf{P} H < \infty$,

(H'3) : There is a neighborhood V of a_k , such that for all ε , there is a η such that for all $c \in V$ and $a \in \Theta$, verifying $\|a - a_k\| \geq \varepsilon$, we have $\mathbf{P} M(c, a_k) < \mathbf{P} M(c, a) - \eta$.

Putting $I_{a_k} = \frac{\partial^2}{\partial a^2} K(g \frac{f_{a_k}}{g_{a_k}}, f)$, and $x \mapsto \rho(b, a, x) = \ln\left(\frac{g(x)}{f(x)} \frac{f_b(b^\top x)}{g_b(b^\top x)}\right) \frac{g(x) f_a(a^\top x)}{g_a(a^\top x)}$, we consider now three new hypotheses:

(H'4) : There is a neighborhood V'_k of (a_k, a_k) such that, for all (b, a) of V'_k , the gradient $\nabla\left(\frac{g(x) f_a(a^\top x)}{g_a(a^\top x)}\right)$ and the Hessian $\mathcal{H}\left(\frac{g(x) f_a(a^\top x)}{g_a(a^\top x)}\right)$ exist (λ - a.s.), and the first order partial derivative $\frac{g(x) f_a(a^\top x)}{g_a(a^\top x)}$ and the first and second order derivative of $(b, a) \mapsto \rho(b, a, x)$ are dominated (λ - a.s.) by integrable functions.

(H'5) : The function $(b, a) \mapsto M(b, a, x)$ is \mathcal{C}^3 in a neighborhood V'_k of (a_k, a_k) for all x ; and all the partial derivatives of order 3 of $(b, a) \mapsto M(b, a, x)$ are dominated in V'_k by a \mathbf{P} - integrable function $H(x)$.

(H'6) : $\mathbf{P} \left\| \frac{\partial}{\partial b} M(a_k, a_k) \right\|^2$ and $\mathbf{P} \left\| \frac{\partial}{\partial a} M(a_k, a_k) \right\|^2$ are finite and the expressions $\mathbf{P} \frac{\partial^2}{\partial b_i \partial b_j} M(a_k, a_k)$ and I_{a_k} exist and are invertible.

Finally, we define

(H'7) : There exists k such that $\mathbf{P} M(a_k, a_k) = 0$.

(H'8) : $(\text{Var}_{\mathbf{P}}(M(a_k, a_k)))^{1/2}$ exists and is invertible.

(H'0): f and g are assumed to be positive and bounded.

3.1.2 Estimation of the first co-vector of f

Let \mathcal{R} be the class of all positive functions r defined on \mathbb{R} and such that $g(x)r(a^\top x)$ is a density on \mathbb{R}^d for all a belonging to \mathbb{R}_*^d . The following proposition shows that there exists a vector a such that $\frac{f_a}{g_a}$ minimizes $K(gr, f)$ in r :

Proposition 2. *There exists a vector a belonging to \mathbb{R}_*^d such that*

$$\arg \min_{r \in \mathcal{R}} K(gr, f) = \frac{f_a}{g_a} \text{ and } r(a^\top x) = \frac{f_a(a^\top x)}{g_a(a^\top x)}.$$

In the following of sections 3.1 and 3.2, we will use the kernel estimate of f (resp. g , f_a and g_a for any a in \mathbb{R}_*^d) instead of f (resp. g , f_a and g_a for any a in \mathbb{R}_*^d). We will keep the notation " f " (resp. " g ", " f_a " and " g_a ") to designate this estimate.

Following [BROKEZ], let us introduce the estimate of $K(g\frac{f_a}{g_a}, f)$, through

$$\tilde{K}(g\frac{f_a}{g_a}, f) = \int M(a, a, x) d\mathbb{P}_n(x)$$

Proposition 3. *Let*

$$\tilde{a} := \arg \inf_{a \in \mathbb{R}_*^d} \tilde{K}(g\frac{f_a}{g_a}, f).$$

Then, \tilde{a} is a strongly convergent estimate of a as defined in proposition 2.

Let us also introduce the following sequences $(\tilde{a}_k)_{k \geq 1}$ and $(\check{g}^{(k)})_{k \geq 1}$ - for any given n :

- \tilde{a}_k is an estimate of a_k as defined in proposition 3 with $\check{g}^{(k-1)}$ instead of g - see section 2.2 -,
- $\check{g}^{(k)}$ is such that $\check{g}^{(0)} = g$, $\check{g}^{(k)}(x) = \check{g}^{(k-1)}(x) \frac{f_{\tilde{a}_k}(\tilde{a}_k^\top x)}{[\check{g}^{(k-1)}]_{\tilde{a}_k}(\tilde{a}_k^\top x)}$, i.e. $\check{g}^{(k)}(x) = g(x) \prod_{j=1}^k \frac{f_{\tilde{a}_j}(\tilde{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\tilde{a}_j}(\tilde{a}_j^\top x)}$.

We also note that $\check{g}^{(k)}$ is a density.

3.1.3 Convergence study at the k^{th} step of the algorithm:

In this paragraph, we will show that the sequence $(\tilde{a}_k)_n$ converges towards a_k and that the sequence $(\check{g}^{(k)})_n$ converges towards $g^{(k)}$.

Let $\check{c}_n(a) = \arg \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$, with $a \in \Theta$, and $\check{\gamma}_n = \arg \inf_{a \in \Theta} \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$. We state

Proposition 4. *It holds*

1/ $\sup_{a \in \Theta} \|\check{c}_n(a) - a_k\|$ tends to 0 a.s.

2/ $\check{\gamma}_n$ tends to a_k a.s. .

Finally, the following theorem shows that $\check{g}^{(k)}$ converges almost everywhere towards $g^{(k)}$:

Theorem 1. *It holds*

$$\check{g}^{(k)} \rightarrow_n g^{(k)} \text{ a.s.}$$

3.2 Asymptotic Inference at the k^{th} step of the algorithm

The following theorem shows that $\check{g}^{(k)}$ converges towards $g^{(k)}$ at the rate $O_{\mathbf{P}}(m^{-1_{d=1} - \frac{4}{4+d} \mathbf{1}_{d>1}})$ in three different cases, namely for any given x , with the L^1 distance and with the relative entropy:

Theorem 2. *It holds*

$$\begin{aligned} |\check{g}^{(k)}(x) - g^{(k)}(x)| &= O_{\mathbf{P}}(m^{-1_{d=1} - \frac{4}{4+d} \mathbf{1}_{d>1}}), \\ \int |\check{g}_n^{(k)}(x) - g^{(k)}(x)| dx &= O_{\mathbf{P}}(m^{-1_{d=1} - \frac{4}{4+d} \mathbf{1}_{d>1}}), \\ |K(\check{g}_n^{(k)}, f) - K(g^{(k)}, f)| &= O_{\mathbf{P}}(m^{-1_{d=1} - \frac{4}{4+d} \mathbf{1}_{d>1}}). \end{aligned}$$

Then, the following theorem shows that the laws of our estimators of a_k , namely $\check{c}_n(a_k)$ and $\check{\gamma}_n$, converge towards a linear combination of Gaussian variables.

Theorem 3. *It holds*

$$\begin{aligned} \sqrt{n} \mathcal{A} \cdot (\check{c}_n(a_k) - a_k) &\xrightarrow{\mathcal{L}^{aw}} \mathcal{B} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial b} M(a_k, a_k)\|^2) + \mathcal{C} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial a} M(a_k, a_k)\|^2) \text{ and} \\ \sqrt{n} \mathcal{A} \cdot (\check{\gamma}_n - a_k) &\xrightarrow{\mathcal{L}^{aw}} \mathcal{C} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial b} M(a_k, a_k)\|^2) + \mathcal{C} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial a} M(a_k, a_k)\|^2) \\ \text{where } \mathcal{A} &= \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) (\mathbf{P} \frac{\partial^2}{\partial a \partial a} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a \partial b} M(a_k, a_k)), \mathcal{C} = \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) \text{ and} \\ \mathcal{B} &= \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a \partial a} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a \partial b} M(a_k, a_k). \end{aligned}$$

3.3 A stopping rule for the procedure

In this paragraph, we will call $\check{g}_n^{(k)}$ (resp. $\check{g}_{a,n}^{(k)}$) the kernel estimator of $\check{g}^{(k)}$ (resp. $\check{g}_a^{(k)}$). We will first show that $\check{g}_n^{(k)}$ converges towards f in k and n . Then, we will provide a stopping rule for this identification procedure.

3.3.1 Estimation of f

Through remark 6 page 13 and as explained in section 14 of [HUB85], the following lemma shows that $K(\check{g}_n^{(k-1)} \frac{f_{a_k,n}}{g_{a_k,n}^{(k-1)}}, f_{a_k,n})$ converges almost everywhere towards zero as k and then as n go to infinity:

Lemma 1. *We have $\lim_n \lim_k K(\check{g}_n^{(k)} \frac{f_{a_k,n}}{[\check{g}^{(k)}]_{a_k,n}}, f_n) = 0$ a.s.*

Consequently, the following proposition provides us with an estimate of f :

Theorem 4. *We have $\lim_n \lim_k \check{g}_n^{(k)} = f$ a.s.*

3.3.2 Testing of the criteria

In this paragraph, through a test of the criteria, namely $a \mapsto K(\check{g}_n^{(k)} \frac{f_{a,n}}{[\check{g}^{(k)}]_{a,n}}, f_n)$, we will build a stopping rule for this identification procedure.

First, the next theorem enables us to derive the law of the criteria:

Theorem 5. *For a fixed k , we have*

$$\sqrt{n} (Var_{\mathbf{P}}(M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)))^{-1/2} (\mathbb{P}_n M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n) - \mathbb{P}_n M(a_k, a_k)) \xrightarrow{\mathcal{L}^{aw}} \mathcal{N}(0, I), \text{ as } n \text{ goes to infinity,}$$

where k represents the k^{th} step of the algorithm, where I is the identity matrix in \mathbb{R}^d .

Note that k is fixed in theorem 5 since $\check{\gamma}_n = \arg \inf_{a \in \Theta} \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$ where M is a known function of k - see section 3.1.1. Thus, in the case where $K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}, f) = 0$, we obtain

Corollary 1. *We have $\sqrt{n} (Var_{\mathbf{P}}(M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)))^{-1/2} (\mathbb{P}_n M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)) \xrightarrow{\mathcal{L}^{aw}} \mathcal{N}(0, I)$.*

Hence, we propose the test of the null hypothesis

$$(H_0) : K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}, f) = 0 \text{ versus the alternative } (H_1) : K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}, f) \neq 0.$$

Based on this result, we stop the algorithm, then, defining a_k as the last vector generated, we derive from corollary 1 a α -level confidence ellipsoid around a_k , namely

$$\mathcal{E}_k = \{b \in \mathbb{R}^d; \sqrt{n}(\text{Var}_{\mathbf{P}}(M(b, b)))^{-1/2} \mathbb{P}_n M(b, b) \leq q_{\alpha}^{\mathcal{N}(0,1)}\}$$

where $q_{\alpha}^{\mathcal{N}(0,1)}$ is the quantile of a α -level reduced centered normal distribution and where \mathbb{P}_n is the empirical measure arising from a realization of the sequences (X_1, \dots, X_n) and (Y_1, \dots, Y_n) .

Consequently, the following corollary provides us with a confidence region for the above test:

Corollary 2. \mathcal{E}_k is a confidence region for the test of the null hypothesis (H_0) versus (H_1) .

3.4 Goodness-of-fit test for copulas

Let us begin with studying the following case:

Let f be a density defined on \mathbb{R}^2 and let g be an Elliptic distribution with same mean and variance as f . Assuming first that the algorithm leads us to having $K(g^{(2)}, f) = 0$ where family (a_i) is the canonical basis of \mathbb{R}^2 . Hence, we have $g^{(2)}(x) = g(x) \frac{f_1}{g_1} \frac{f_2}{g_2} = g(x) \frac{f_1}{g_1} \frac{f_2}{g_2}$ - through lemma 14 page 25 - and $g^{(2)} = f$. Therefore, $f = g(x) \frac{f_1}{g_1} \frac{f_2}{g_2}$, i.e. $\frac{f}{f_1 f_2} = \frac{g}{g_1 g_2}$, and then $\frac{\partial^2}{\partial x \partial y} C_f = \frac{\partial^2}{\partial x \partial y} C_g$ where C_f (resp. C_g) is the copula of f (resp. g).

At present, let f be a density on \mathbb{R}^d and let g be the density defined in section 2.1.2.

Let us assume that the algorithm implies that $K(g^{(d)}, f) = 0$. Hence, we have, for any $x \in \mathbb{R}^d$, $g(x) \prod_{k=1}^d \frac{f_{a_k}(a_k^{\top} x)}{[g^{(k-1)}]_{a_k}(a_k^{\top} x)} = f(x)$, i.e. $\frac{g(x)}{\prod_{k=1}^d g_{a_k}(a_k^{\top} x)} = \frac{f(x)}{\prod_{k=1}^d f_{a_k}(a_k^{\top} x)}$ - through lemma 14.

Moreover, the family $(a_i)_{i=1 \dots d}$ is a basis of \mathbb{R}^d - see lemma 15 page 25. Hence, putting $A = (a_1, \dots, a_d)$ and defining vector y (resp. density \tilde{f} , copula \tilde{C}_f of \tilde{f} , density \tilde{g} , copula \tilde{C}_g of \tilde{g}) as the expression of vector x (resp. density f , copula C_f of f , density g , copula C_g of g) in basis A , the above equality implies $\frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f = \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_g$. Finally, we perform a statistical test of the null hypothesis

$$(H_0) : \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f = \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_g \text{ versus } (H_1) : \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f \neq \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_g.$$

Since under (H_0) we have $K(g^{(d)}, f) = 0$, then, as explained in section 3.3.2, corollary 2 provides us with a confidence region for the test.

Theorem 6. Keeping the notations of corollary 2, we infer that \mathcal{E}_d is a confidence region for the test of the null hypothesis (H_0) versus the alternative (H_1) .

3.5 Rewriting of the convolution product

In the present paper, we first elaborated an algorithm aiming at isolating several known structures from initial datas. Our objective was to verify if for a known density on \mathbb{R}^d , there exists a known density n on \mathbb{R}^{d-j-1} such that, for $d > 1$

$$f(x) = n(a_{j+1}^{\top} x, \dots, a_d^{\top} x) h(a_1^{\top} x, \dots, a_j^{\top} x), \quad (6)$$

with $j < d$, with (a_1, \dots, a_d) being a basis of \mathbb{R}^d and with h being a density on \mathbb{R}^j .

Secondly, our next step has been to build an estimate (resp. a representation) of f without necessarily assuming that f meets relationship (6) - see theorem 4 (resp. proposition 13 page 21).

Consequently, let us consider Z_1 and Z_2 , two random vectors with respective densities h_1 and h_2 - which is Elliptic - on \mathbb{R}^d . Let us consider a random vector X such that $X = Z_1 + Z_2$ and let f be its density. This density can then be written as :

$$f(x) = h_1 * h_2(x) = \int_{\mathbb{R}^d} h_1(x)h_2(t - x)dt.$$

Then, the following property enables us to represent f under the form of a product and without the integral sign

Proposition 5. *Let ϕ be a centered Elliptic density with $\sigma^2.I_d$, $\sigma^2 > 0$, as covariance matrix, such that it is a product density in all orthogonal coordinate systems and such that its characteristic function $s \mapsto \Psi(\frac{1}{2}|s|^2\sigma^2)$ is integrable - see [LANDS03].*

Let f be a density on \mathbb{R}^d which can be deconvoluted with ϕ , i.e.

$$f = \bar{f} * \phi = \int_{\mathbb{R}^d} \bar{f}(x)\phi(t - x)dt,$$

where \bar{f} is some density on \mathbb{R}^d .

Let $g^{(0)}$ be the Elliptic density belonging to the same Elliptic family as f and having same mean and variance as f .

Then, the sequence $(g^{(k)})_k$ converges uniformly a.s. and in L^1 towards f in k , i.e.

$$\lim_{k \rightarrow \infty} \sup_{x \in \mathbb{R}^d} |g^{(k)}(x) - f(x)| = 0, \text{ and } \lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} |g^{(k)}(x) - f(x)|dx = 0.$$

Finally, along with the notations of section 3.3 and of proposition 5, the following theorem enables us to estimate any convolution product of a multivariate Elliptic density ϕ with a continuous density \bar{f} :

Theorem 7. *It holds $\lim_n \lim_k \check{g}_n^{(k)} = \bar{f} * \phi$ a.s.*

4 Comparison of all the optimisation methods

In this section, we will study Huber's algorithm in a similar manner to sections 2 and 3. We will then be able to compare our methodologies.

Remark 5 (Huber's algorithm). *First, until now, the choice has only been to use the class of Gaussian distributions. Here and similarly to section 2.1, we extend this choice to the class of elliptic distributions.*

Moreover, using the subsample X_1, X_2, \dots, X_n - see Annex B - and using the protocol of section 2.2 with $K(g_a, f_a)$ - see section 4.2 - instead of $K(g_{f_a}^{g_a}, f)$, proposition 11 page 19, lemma 11 page 24 and remark 6 page 13 enable us to perform Huber's algorithm :

- *we define \hat{a}_1 and the density $\hat{g}^{(1)}$ such that $\hat{a}_1 = \arg \max_{a \in \mathbb{R}_*^d} K(g_{a,n}, f_{a,n})$ and $\hat{g}^{(1)} = g_{\frac{f_{\hat{a}_1,n}}{g_{\hat{a}_1,n}}}$,*
- *we define \hat{a}_2 and the density $\hat{g}^{(2)}$ such that $\hat{a}_2 = \arg \max_{a \in \mathbb{R}_*^d} K(\hat{g}_{a,n}^{(1)}, f_{a,n})$ and $\hat{g}^{(2)} = \hat{g}_{\frac{f_{\hat{a}_2,n}}{\hat{g}_{\hat{a}_2,n}^{(1)}}}^{(1)}$,*

and so on, we will end up obtaining a sequence $(\hat{a}_1, \hat{a}_2, \dots)$ of vectors in \mathbb{R}_^d and a sequence of densities $\hat{g}^{(k)}$.*

4.1 Hypotheses on f

In this paragraph, we define the set of hypotheses on f which we can possibly use in this work. First, denote g in lieu of $g^{(k-1)}$. Let

$$\Theta_a^1 = \{b \in \Theta \mid \int (\frac{g_b(b^\top x)}{f_b(b^\top x)} - 1) f_a(a^\top x) dx < \infty\},$$

$$m(b, a, x) = \int \ln(\frac{g_b(b^\top x)}{f_b(b^\top x)}) g_a(a^\top x) dx - (\frac{g_b(b^\top x)}{f_b(b^\top x)} - 1),$$

$$\mathbf{P}^a m(b, a) = \int m(b, a, x) f_a(a^\top x) dx \text{ and } \mathbb{P}_n m(b, a) = \int m(b, a, x) \frac{f_a(a^\top x)}{f(x)} d\mathbb{P}_n,$$

with \mathbf{P}^a being the probability measure with density f_a .

Similarly as *Van der Vaart*, in chapter *V* of [VDW], let us define :

(H1) : For all $\varepsilon > 0$, there is $\eta > 0$ such that, for all $b \in \Theta_a^1$ verifying $\|b - a_k\| \geq \varepsilon$ for all $a \in \Theta$, we have $\mathbf{P}^a m(b, a) < \mathbf{P}^a m(a_k, a) - \eta$,

(H2) : There is a neighborhood of a_k , V , and a positive function H , such that, for all $b \in V$, we have $|m(b, a_k, x)| \leq H(x)$ (\mathbf{P}^a - a.s.) with $\mathbf{P}^a H < \infty$,

(H3) : There is a neighborhood V of a_k , such that for all ε , there is a η such that for all $b \in V$ and $a \in \Theta$, verifying $\|a - a_k\| \geq \varepsilon$, we have $\mathbf{P}^{a_k} m(b, a_k) - \eta > \mathbf{P}^a m(b, a)$.

Moreover, defining $x \rightarrow v(b, a, x) = \ln(\frac{g_b(b^\top x)}{f_b(b^\top x)}) g_a(a^\top x)$, putting:

(H4) : There exists a neighborhood of (a_k, a_k) , that we will name V_k , such that, for all (b, a) of V_k , the gradient $\nabla(\frac{g_a(a^\top x)}{f_a(a^\top x)})$ and the Hessian $\mathcal{H}(\frac{g_a(a^\top x)}{f_a(a^\top x)})$ exist (λ - a.s.) and the first order partial derivative $\frac{g_a(a^\top x)}{f_a(a^\top x)}$ and the first and second order derivative of order 3 of $(b, a) \mapsto v(b, a, x)$ are dominated (λ - a.s.) by integrable functions.

(H5) : The function $(b, a) \mapsto m(b, a)$ is \mathcal{C}^3 in a neighborhood V_k of (a_k, a_k) for all x and all the partial derivatives of $(b, a) \mapsto m(b, a)$ are dominated in V_k by a \mathbf{P}_- integrable function $H(x)$.

(H6) : $\mathbf{P} \|\frac{\partial}{\partial b} m(a_k, a_k)\|^2$ and $\mathbf{P} \|\frac{\partial}{\partial a} m(a_k, a_k)\|^2$ are finite and the quantities $\mathbf{P} \frac{\partial^2}{\partial b_i \partial b_j} m(a_k, a_k)$ and $\mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} m(a_k, a_k)$ are invertible.

(H7) : there exists k such that $\mathbf{P} m(a_k, a_k) = 0$.

(H8) : $(\text{Var}_{\mathbf{P}}(m(a_k, a_k)))^{1/2}$ exists and is invertible.

4.2 The first co-vector of f simultaneously optimizes four problems

Let us first study Huber's analytic approach.

Let \mathcal{R}' be the class of all positive functions r defined on \mathbb{R} and such that $f(x)r^{-1}(a^\top x)$ is a density on \mathbb{R}^d for all a belonging to \mathbb{R}_*^d . The following proposition shows that there exists a vector a such that $\frac{f_a}{g_a}$ minimizes $K(fr^{-1}, g)$ in r :

Proposition 6 (Analytic Approach). *There exists a vector a belonging to \mathbb{R}_*^d such that*

$$\arg \min_{r \in \mathcal{R}'} K(fr^{-1}, g) = \frac{f_a}{g_a}, \text{ and } r(a^\top x) = \frac{f_a(a^\top x)}{g_a(a^\top x)}$$

and $K(f, g) = K(f_a, g_a) + K(f \frac{g_a}{f_a}, g)$.

Let us also study Huber's synthetic approach:

Let \mathcal{R} be the class of all positive functions r defined on \mathbb{R} and such that $g(x)r(a^\top x)$ is a density on \mathbb{R}^d for all a belonging to \mathbb{R}_*^d . The following proposition shows that there exists a vector a such that $\frac{f_a}{g_a}$ minimizes $K(gr, f)$ in r :

Proposition 7 (Synthetic Approach). *There exists a vector a belonging to \mathbb{R}_*^d such that*

$$\arg \min_{r \in \mathcal{R}} K(f, gr) = \frac{f_a}{g_a}, \text{ and } r(a^\top x) = \frac{f_a(a^\top x)}{g_a(a^\top x)}$$

and $K(f, g) = K(f_a, g_a) + K(f, g \frac{f_a}{g_a})$.

In the meanwhile, the following proposition shows that there exists a vector a such that $\frac{f_a}{g_a}$ minimizes $K(g, fr^{-1})$ in r .

Proposition 8. *There exists a vector a belonging to \mathbb{R}_*^d such that*

$$\arg \min_{r \in \mathcal{R}} K(g, fr^{-1}) = \frac{f_a}{g_a}, \text{ and } r(a^\top x) = \frac{f_a(a^\top x)}{g_a(a^\top x)}$$

Moreover, we have $K(g, f) = K(g_a, f_a) + K(g, f \frac{g_a}{f_a})$.

Remark 6. *First, through property 4 page 18, we get $K(f, g \frac{f_a}{g_a}) = K(g, f \frac{g_a}{f_a}) = K(f \frac{g_a}{f_a}, g)$ and $K(f_a, g_a) = K(g_a, f_a)$. Thus, proposition 8 implies that finding the argument of the maximum $K(g_a, f_a)$ amounts to finding the argument of the maximum $K(f_a, g_a)$. Consequently, the criteria of Huber's methodologies is $a \mapsto K(g_a, f_a)$. Second, our criteria is $a \mapsto K(g \frac{g_a}{f_a}, f)$ and property 4 implies $K(g, f \frac{g_a}{f_a}) = K(g \frac{f_a}{g_a}, f)$. Consequently, since [BROKEZ] takes into account the very form of the criteria, we are then in a position to compare Huber's methodologies with ours.*

To recapitulate, the choice of $r = \frac{f_a}{g_a}$ enables us to simultaneously solve the following four optimisation problems, for $a \in \mathbb{R}_*^d$:

First, find a such that $a = \arg \inf_{a \in \mathbb{R}_*^d} K(f \frac{g_a}{f_a}, g)$ - pertaining to the analytic approach -

Second, find a such that $a = \arg \inf_{a \in \mathbb{R}_*^d} K(f, g \frac{f_a}{g_a})$ - pertaining to the synthetic approach -

Third, find a such that $a = \arg \sup_{a \in \mathbb{R}_*^d} K(g_a, f_a)$ - to compare Huber's methods with ours -

Fourth, find a such that $a = \arg \inf_{a \in \mathbb{R}_*^d} K(g \frac{f_a}{g_a}, f)$ - pertaining to our method.

4.3 On the sequence of the transformed densities $(g^{(j)})$

As already explained in the introduction section, the Mu Zhu article leads us to only consider Huber's synthetic approach. Moreover, in this section, we will use the kernel estimate of f (resp. g , f_a and g_a for any a in \mathbb{R}_*^d) instead of f (resp. g , f_a and g_a for any a in \mathbb{R}_*^d). We will keep the notation " f " (resp. " g ", " f_a " and " g_a ") to designate this estimate.

4.3.1 Estimation of the first co-vector of f

Using the subsample X_1, X_2, \dots, X_n - see Annex B - and following [BROKEZ], let us introduce the estimate of $K(g_a, f_a)$, through $\hat{K}(g_a, f_a) = \int m(a, a, x) (\frac{f_a(a^\top x)}{f(x)}) d\mathbb{P}_n$

Proposition 9. *Let $\hat{a} := \arg \sup_{a \in \mathbb{R}_*^d} \hat{K}(g_a, f_a)$. Then, \hat{a} is a strongly convergent estimate of a as defined in proposition 8.*

Finally, let us define the following sequences $(\hat{a}_k)_{k \geq 1}$ and $(\hat{g}^{(k)})_{k \geq 1}$ - for any given n :

- \hat{a}_k is an estimate of a_k as defined in proposition 9 with $\hat{g}^{(k-1)}$ instead of g - see remark 5 -,
- $\hat{g}^{(k)}$ is such that $\hat{g}^{(0)} = g$ and $\hat{g}^{(k)}(x) = \hat{g}^{(k-1)}(x) \frac{f_{\hat{a}_k}(\hat{a}_k^\top x)}{[\hat{g}^{(k-1)}]_{\hat{a}_k}(\hat{a}_k^\top x)}$, i.e. $\hat{g}^{(k)}(x) = g(x) \prod_{j=1}^k \frac{f_{\hat{a}_j}(\hat{a}_j^\top x)}{[\hat{g}^{(j-1)}]_{\hat{a}_j}(\hat{a}_j^\top x)}$.

We also note that $\hat{g}^{(k)}$ is a density.

4.3.2 Convergence study at the k^{th} step of the algorithm

Let $\hat{b}_n(a) = \arg \sup_{b \in \Theta} \mathbb{P}_n^a m(b, a)$, with $a \in \Theta$, and $\hat{\beta}_n = \arg \sup_{a \in \Theta} \sup_{b \in \Theta} \mathbb{P}_n^a m(b, a)$. We state

Proposition 10. *Both $\sup_{a \in \Theta} \|\hat{b}_n(a) - a_k\|$ and $\hat{\beta}_n$ converge toward a_k a.s.*

Finally, the following theorem shows that $\hat{g}^{(k)}$ converges almost everywhere towards $g^{(k)}$:

Theorem 8. *For any given k , it holds $\hat{g}^{(k)} \rightarrow_n g^{(k)}$ a.s.*

4.3.3 Asymptotic Inference at the k^{th} step of the algorithm

The following theorem shows that $\hat{g}^{(k)}$ converges towards $g^{(k)}$ at the rate $O_{\mathbf{P}}(m^{-1_{d=1}-\frac{4}{4+d}1_{d>1}})$ in three different cases, namely for any given x , with the L^1 distance and with the relative entropy:

Theorem 9. *It holds $|\hat{g}^{(k)}(x) - g^{(k)}(x)| = O_{\mathbf{P}}(m^{-1_{d=1}-\frac{4}{4+d}1_{d>1}})$, $\int |\hat{g}^{(k)}(x) - g^{(k)}(x)| dx = O_{\mathbf{P}}(m^{-1_{d=1}-\frac{4}{4+d}1_{d>1}})$, and $|K(f, \hat{g}^{(k)}) - K(f, g^{(k)})| = O_{\mathbf{P}}(m^{-1_{d=1}-\frac{4}{4+d}1_{d>1}})$.*

The following theorem shows that the laws of Huber's estimators of a_k , namely $\hat{b}_n(a_k)$ and $\hat{\beta}_n$, converge towards a linear combination of Gaussian variables.

Theorem 10. *It holds*

$$\begin{aligned} \sqrt{n}\mathcal{D}(\hat{b}_n(a_k) - a_k) &\xrightarrow{\mathcal{L}aw} \mathcal{E}\mathcal{N}_d(0, \mathbf{P}\|\frac{\partial}{\partial b}m(a_k, a_k)\|^2) + \mathcal{F}\mathcal{N}_d(0, \mathbf{P}\|\frac{\partial}{\partial a}m(a_k, a_k)\|^2) \text{ and} \\ \sqrt{n}\mathcal{D}(\hat{\beta}_n - a_k) &\xrightarrow{\mathcal{L}aw} \mathcal{G}\mathcal{N}_d(0, \mathbf{P}\|\frac{\partial}{\partial a}m(a_k, a_k)\|^2) + \mathcal{F}\mathcal{N}_d(0, \mathbf{P}\|\frac{\partial}{\partial b}m(a_k, a_k)\|^2) \\ \text{where } \mathcal{E} &= \mathbf{P}\frac{\partial^2}{\partial a^2}m(a_k, a_k), \mathcal{F} = \mathbf{P}\frac{\partial^2}{\partial a\partial b}m(a_k, a_k), \mathcal{G} = \mathbf{P}\frac{\partial^2}{\partial b^2}m(a_k, a_k) \text{ and} \\ \mathcal{D} &= (\mathbf{P}\frac{\partial^2}{\partial b^2}m(a_k, a_k)\mathbf{P}\frac{\partial^2}{\partial a^2}m(a_k, a_k) - \mathbf{P}\frac{\partial^2}{\partial a\partial b}m(a_k, a_k)\mathbf{P}\frac{\partial^2}{\partial b\partial a}m(a_k, a_k)) > 0. \end{aligned}$$

4.4 A stopping rule for the procedure

We will first give an estimation of f . We will then provide a stopping rule for this identification procedure. We note that, in this section, we will call f_n (respectively $\hat{g}_n^{(k)}$, $\forall a \in \Theta$ $\hat{g}_{a,n}^{(k)}$, $f_{a,n}$) the kernel estimator of density f (respectively $\hat{g}^{(k)}$, $\forall a \in \Theta$ $\hat{g}_a^{(k)}$, f_a).

Remark 7. In the case where f is known, then, as explained in section 14 of [HUB85], the sequence $(K(g_{a_k}^{(k-1)}, f_{a_k}))_{k \geq 1}$ converges towards zero. Many authors have studied this hypothesis and its consequences. For example, Huber deducts that, if f can be deconvoluted with a Gaussian component, $(K(g_{a_k}^{(k-1)}, f_{a_k}))_{k \geq 1}$ converges toward 0. He then shows that $g^{(i)}$ uniformly converges in L^1 towards f - see propositions 14.2 and 14.3 page 461 of his article. Similarly, Friedman in [Frie8487], page 19, and Hwang, in [HLL94] pages 16 and 17, develop other typical examples.

4.4.1 Estimation of f

The following lemma shows that $\lim_k K(\hat{g}_{a_k,n}^{(k)}, f_{a_k,n})$ converges towards zero as k and then as n :

Lemma 2. *We have $\lim_n \lim_k K(\hat{g}_{a_k,n}^{(k)}, f_{a_k,n}) = 0$, a.s.*

Then, the following theorem enables us to provide simulations through an estimation of f

Theorem 11. *We have $\lim_n \lim_k \hat{g}_n^{(k)} = f$, a.s.*

4.4.2 Testing of the criteria

In this paragraph, through a test of Huber's criteria, namely $a \mapsto K(\hat{g}_{a,n}^{(k)}, f_{a,n})$, we will build a stopping rule for the procedure. First, the next theorem gives us the law of Huber's criteria.

Theorem 12. *For a fixed k , we have*

$$\sqrt{n}(\text{Var}_{\mathbf{P}}(m(\hat{b}_n(\hat{\beta}_n), \hat{\beta}_n)))^{-1/2}(\mathbb{P}_n m(\hat{b}_n(\hat{\beta}_n), \hat{\beta}_n) - \mathbb{P}_n m(a_k, a_k)) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, I), \text{ as } n \text{ goes to infinity,}$$

where k represents the k^{th} step of the algorithm, where I is the identity matrix in \mathbb{R}^d .

Note that k is fixed in theorem 12 since $\hat{\beta}_n = \arg \sup_{a \in \Theta} \sup_{b \in \Theta} \mathbb{P}_n^a m(b, a)$ where m is a known function of k - see section 4.1. Thus, in the case where $K(g_a^{(k)}, f_a) = 0$, we obtain

Corollary 3. We have $\sqrt{n}(\text{Var}_{\mathbf{P}}(m(\hat{b}_n(\hat{\beta}_n), \hat{\beta}_n)))^{-1/2}(\mathbb{P}_n m(\hat{b}_n(\hat{\beta}_n), \hat{\beta}_n)) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, I)$.

Hence, we propose the test of the null hypothesis $(H_0) : K(g_{a_k}^{(k-1)}, f_{a_k}) = 0$ versus the alternative $(H_1) : K(g_{a_k}^{(k-1)}, f_{a_k}) \neq 0$. Based on this result, we stop the algorithm, then, defining a_k as the last vector generated from Huber's algorithm, we derive from corollary 3 a α -level confidence ellipsoid around a_k , namely $\mathcal{E}'_k = \{b \in \mathbb{R}^d; \sqrt{n}(\text{Var}_{\mathbf{P}}(m(b, b)))^{-1/2} \mathbb{P}_n m(b, b) \leq q_{\alpha}^{\mathcal{N}(0,1)}\}$ where $q_{\alpha}^{\mathcal{N}(0,1)}$ is the quantile of a α -level reduced centered normal distribution and where \mathbb{P}_n is the empirical measure arising from a realization of the sequences (X_1, \dots, X_n) and (Y_1, \dots, Y_n) .

Consequently, the following corollary provides us with a confidence region for the above test:

Corollary 4. \mathcal{E}'_k is a confidence region for the test of the null hypothesis (H_0) versus (H_1) .

5 Simulations

We will illustrate this section by detailing several examples.

In each example, the first part of the program will follow our algorithm and will aim at creating a sequence of densities $(g^{(j)})$, $j = 1, \dots, k$, $k < d$, such that $g(0) = g$, $g^{(j)} = g^{(j-1)} f_{a_j} / [g^{(j-1)}]_{a_j}$ and $K(g^{(k)}, f) = 0$, where K is the relative entropy and $a_j = \arg \inf_b K(g^{(j-1)} f_b / [g^{(j-1)}]_b, f)$, for all $j = 1, \dots, k$.

Moreover, in a second step, the program will follow Huber's method and will create a sequence of densities $(g^{(j)})$, $j = 1, \dots, k$, $k < d$, such that $g(0) = g$, $g^{(j)} = g^{(j-1)} f_{a_j} / [g^{(j-1)}]_{a_j}$ and $K(f, g^{(k)}) = 0$, where K is the relative entropy and $a_j = \arg \sup_b K([g^{(j-1)}]_b, f_b)$, for all $j = 1, \dots, k$.

Example 1 : We are in dimension 3(=d), let us consider a sample of 50(=n) values of a random variable X with a density law f defined by,

$$f(x) = \text{Normal}(x_1 + x_2). \text{Gumbel}(x_0 + x_2). \text{Gumbel}(x_0 + x_1),$$

where the Gumbel law parameters are $(-3, 4)$ and $(1, 1)$ and where the normal distribution parameters are $(-5, 2)$. Let us then generate a Gaussian random variable Y - that we will name g - with a density which presents the same mean and variance as f .

In the first part of the program, we theoretically obtain $k = 2$, $a_1 = (1, 0, 1)$ and $a_2 = (1, 1, 0)$ (or $a_2 = (1, 0, 1)$ and $a_1 = (1, 1, 0)$ which leads us to the same conclusion). To get this result, we perform the following test

$$(H_0) : (a_1, a_2) = ((1, 0, 1), (1, 1, 0)) \text{ versus } (H_1) : (a_1, a_2) \neq ((1, 0, 1), (1, 1, 0)).$$

Moreover, if i represents the last iteration of the algorithm, then

$$\sqrt{n}(\text{Var}_{\mathbf{P}}(M(c_n(\gamma_n), \gamma_n)))^{(-1/2)} \mathbb{P}_n M(c_n(\gamma_n), \gamma_n) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, 1),$$

and then we estimate (a_1, a_2) with the following $0.9(=\alpha)$ level confidence ellipsoid

$$\mathcal{E}_i = \{b \in \mathbb{R}^3; (\text{Var}_{\mathbf{P}}(M(b, b)))^{(-1/2)} \mathbb{P}_n M(b, b) \leq q_{\alpha}^{\mathcal{N}(0,1)} / \sqrt{n} \simeq 0, 2533 / 7.0710678 = 0.03582203\}.$$

Indeed, if $i = 1$ represents the last iteration of the algorithm, then $a_1 \in \mathcal{E}_1$, and if $i = 2$ represents the last iteration of the algorithm, then $a_2 \in \mathcal{E}_2$, and so on, if i represents the last iteration of the algorithm, then $a_i \in \mathcal{E}_i$.

Now, if we follow Huber's method, we also theoretically obtain $k = 2$, $a_1 = (1, 0, 1)$ and $a_2 = (1, 1, 0)$ (or $a_2 = (1, 0, 1)$ and $a_1 = (1, 1, 0)$ which leads us to the same conclusion). To get this result, we perform the following test:

$$(H_0) : (a_1, a_2) = ((1, 0, 1), (1, 1, 0)) \text{ versus } (H_1) : (a_1, a_2) \neq ((1, 0, 1), (1, 1, 0)).$$

The fact that, if i represents the last iteration of the algorithm, then

$$\sqrt{n}(\text{Var}_{\mathbf{P}}(m(b_n(\beta_n), \beta_n)))^{(-1/2)} \mathbb{P}_n m(b_n(\beta_n), \beta_n) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, 1),$$

enables us to estimate our sequence of (a_i) , reduced to (a_1, a_2) , through the following $0.9(=\alpha)$

level confidence ellipsoid $\mathcal{E}'_i = \{b \in \mathbb{R}^3; (Var_{\mathbf{P}}(m(b, b)))^{-1/2} \mathbb{P}_n m(b, b) \leq q_{\alpha}^{N(0,1)} / \sqrt{n} \simeq 0.03582203\}$. Indeed, if $i = 1$ represents the last iteration of the algorithm, then $a_1 \in \mathcal{E}'_1$, and if $i = 2$ represents the last iteration of the algorithm, then $a_2 \in \mathcal{E}'_2$, and so on, if i represents the last iteration of the algorithm, then $a_i \in \mathcal{E}'_i$. Finally, we obtain

	Our Algorithm	Huber's Algorithm
Projection Study number 0 :	minimum : 0.317505	maximum : 0.715135
	at point : (1.0,1.0,0)	at point : (1.0,1.0,0)
	P-Value : 0.99851	P-Value : 0.999839
Test :	$H_0 : a_1 \in \mathcal{E}_1 : \text{False}$	$H_0 : a_1 \in \mathcal{E}'_1 : \text{False}$
Projection Study number 1 :	minimum : 0.0266514	maximum : 0.00727748
	at point : (1.0,0,1.0)	at point : (1.0,0,1.0)
	P-Value : 0.998852	P-Value : 0.999835
Test :	$H_0 : a_2 \in \mathcal{E}_2 : \text{True}$	$H_0 : a_2 \in \mathcal{E}'_2 : \text{True}$
K(Kernel Estimation of $g^{(2)}, g^{(2)}$)	0.444388	0.794124

Therefore, we conclude that $f = g^{(2)}$.

Example 2 :

We are in dimension 2(=d), let us consider a sample of 50(=n) values of a random variable X with a density law f defined by, $f(x) = Cauchy(x_0).Normal(x_1)$, where the Cauchy law parameters are -5 and 1 and where the normal distribution parameters are (0, 1).

Our reasoning is the same as in Example 1. In the first part of the program, we theoretically obtain $k = 1$ and $a_1 = (1, 0)$. To get this result, we perform the following test:

$$(H_0) : a_1 = (1, 0) \text{ versus } (H_1) : a_1 \neq (1, 0).$$

We estimate a_1 by the following 0.75(= α) level confidence ellipsoid

$$\mathcal{E}_i = \{b \in \mathbb{R}^2; (Var_{\mathbf{P}}(M(b, b)))^{-1/2} \mathbb{P}_n M(b, b) \leq q_{\alpha}^{N(0,1)} / \sqrt{n} \simeq 0.6745/7.0710678 = 0.095388\}.$$

Now, if we follow Huber's method, we also theoretically obtain $k = 1$ and $a_1 = (1, 0)$. To get this result, we perform the following test: $(H_0) : a_1 = (1, 0) \text{ versus } (H_1) : a_1 \neq (1, 0)$. Hence, using the same reasoning as in Example 1, we estimate a_1 through the following 0.75 (= α) level confidence ellipsoid $\mathcal{E}'_i = \{b \in \mathbb{R}^2; (Var_{\mathbf{P}}(m(b, b)))^{-1/2} \mathbb{P}_n m(b, b) \leq q_{\alpha}^{N(0,1)} / \sqrt{n} \simeq 0.095388\}$. And, we obtain

	Our Algorithm	Huber's Algorithm
Projection Study n° 0 :	minimum : 0.00263554	maximum : 0.00376235
	at point : (1.0001,0)	at point : (1.0,0.0)
	P-Value : 0.998683	P-Value : 0.998121
Test :	$H_0 : a_1 \in \mathcal{E}_1 : \text{True}$	$H_0 : a_1 \in \mathcal{E}'_1 : \text{True}$
K(Kernel Estimation of $g^{(1)}, g^{(1)}$)	2.44546	2.32331

Therefore, we conclude that $f = g^{(1)}$.

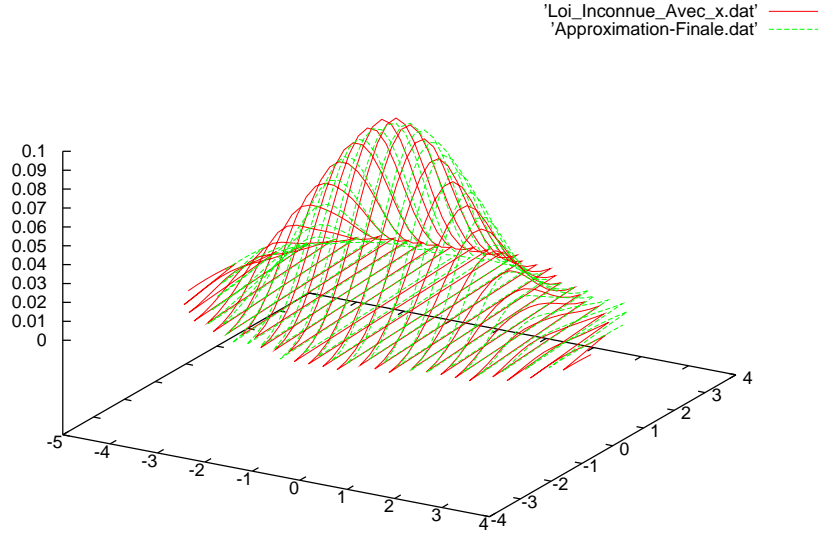


Figure 1: *Graph of the distribution to estimate (red) and of our own estimate (green).*

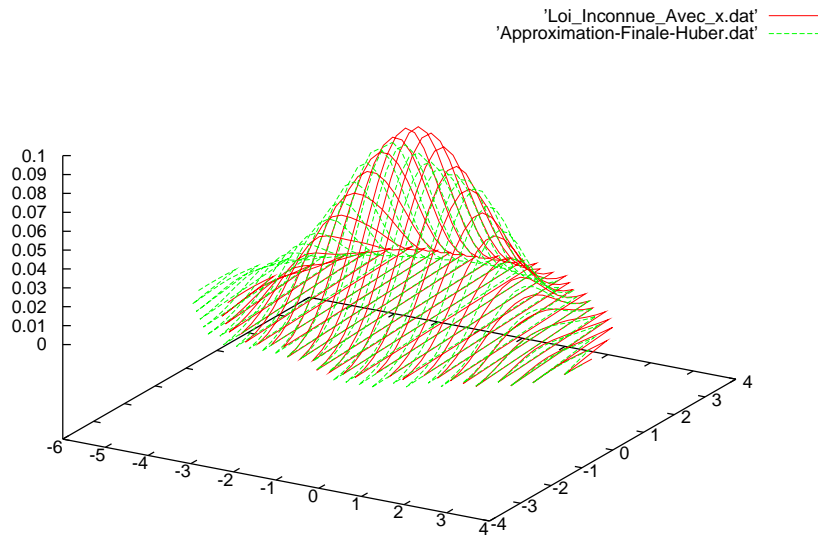


Figure 2: *Graph of the distribution to estimate (red) and of Huber's estimate (green).*

Critics of the simulations

We note that as the approximations accumulate and according to the power of the calculators used, we might obtain results above or below the value of the thresholds of the different tests. Moreover, in the case where f is unknown, we will never be sure to have reached the minimum or the maximum of the relative entropy: we have indeed used the simulated annealing method to solve our optimisation problem, and therefore it is only when the number of random jumps tends in theory towards infinity that the probability to get the minimum or the maximum tends towards 1. We also note that no theory on the optimal number of jumps to implement does exist, as this number depends on the specificities of each particular problem.

Finally, we choose the $50^{-\frac{4}{7}}$ (resp. $50^{-\frac{2}{3}}$) for the AMISE of example 1 (resp. example 2). This choice leads us to simulate 50 random variables - see [SCOTT92] page 151 -, none of which have been discarded to obtain the truncated sample.

Conclusion

Projection Pursuit is useful in evidencing characteristic structures as well as one-dimensional projections and their associated distributions in multivariate data sets.

Huber, in [HUB85], shows us how to achieve it through maximization of the relative entropy.

The present article demonstrates that our relative entropy minimisation method constitutes a good alternative to Huber's. Indeed, the convergence results and simulations we carried out convincingly fulfilled our expectations regarding our methodology.

A Reminders

Let us call h_a the density of $a^\top Z$ if h is the density of Z , and K the relative entropy or Kullback-Liebr distance. The function K is defined by - considering P and Q , two probabilities:

$$K(Q, P) = \int \varphi\left(\frac{\partial Q}{\partial P}\right) dP \text{ if } P \ll Q \text{ and}$$

$$K(Q, P) = +\infty \text{ otherwise,}$$

where $\varphi : x \mapsto x \ln(x) - x + 1$ is strictly convex.

Let us present some well-known properties of the relative entropy.

Property 2. *We have $K(P, Q) = 0 \Leftrightarrow P = Q$.*

Property 3. *The application $Q \mapsto K(Q, P)$ is :*

- *convex,*
- *lower semi-continuous (l.s.c.) for the topology that makes all the applications of the form $Q \mapsto \int f dQ$ continuous where f is bounded and continuous and*
- *l.s.c. for the topology of the uniform convergence, and greater than the L^1 distance.*

Moreover, corollary (1.29), page 19 of [LIVAJ], enables us to derive:

Property 4.

If $T : (X, A) \rightarrow (Y, B)$ is measurable and if $K(P, Q) < \infty$, then $K(P, Q) \geq K(PT^{-1}, QT^{-1})$, with equality being reached when T is surjective for (P, Q) .

And finally, according to theorem III.4 of [AZE97], we have

Theorem 13. *Let $f : I \rightarrow \mathbb{R}$ be a convex function. Then f is a Lipschitz function in all compact intervals $[a, b] \subset \text{int}\{I\}$. In particular, f is continuous on $\text{int}\{I\}$.*

A.1 Useful lemmas

Through a reductio ad absurdum argument, we derive lemmas 3 and 4 :

Lemma 3. *Let f be a density in \mathbb{R}^d bounded and positive. Then, any projection density of f - that we will name f_a , with $a \in \mathbb{R}_*^d$ - is also bounded and positive in \mathbb{R} .*

Lemma 4. *Let f be a density in \mathbb{R}^d bounded and positive. Then any density $f(./a^\top x)$, for any $a \in \mathbb{R}_*^d$, is also bounded and positive.*

By induction and from lemmas 3 and 4, we have

Lemma 5. *If f and g are positive and bounded densities, then $g^{(k)}$ is positive and bounded.*

Finally we introduce a last lemma

Lemma 6. *Let f be an absolutely continuous density, then, for all sequence (a_n) tending to a in \mathbb{R}_*^d , sequence f_{a_n} uniformly converges towards f_a .*

Proof :

For all a in \mathbb{R}_*^d , let F_a be the cumulative distribution function of $a^\top X$ and ψ_a be a complex function defined by $\psi_a(u, v) = F_a(\text{Re}(u + iv)) + iF_a(\text{Re}(v + iu))$, for all u and v in \mathbb{R} .

First, the function $\psi_a(u, v)$ is an analytic function, because $x \mapsto f_a(a^\top x)$ is continuous and as a result of the corollary of Dini's second theorem - according to which "A sequence of cumulative distribution functions which pointwise converges on \mathbb{R} towards a continuous cumulative distribution function F on \mathbb{R} , uniformly converges towards F on \mathbb{R} " - we deduct that, for all sequence (a_n) converging towards a , ψ_{a_n} uniformly converges towards ψ_a . Finally, the Weierstrass theorem, (see proposal (10.1) page 220 of the "Calcul infinitésimal" book of Jean Dieudonné), implies that all sequences $\psi'_{a,n}$ uniformly converge towards ψ'_a , for all a_n tending to a . We can therefore conclude. \square

B Study of the sample

Let X_1, X_2, \dots, X_m be a sequence of independent random vectors with same density f . Let Y_1, Y_2, \dots, Y_m be a sequence of independent random vectors with same density g . Then, the kernel estimators $f_m, g_m, f_{a,m}$ and $g_{a,m}$ of f, g, f_a and g_a , for all $a \in \mathbb{R}_*^d$, almost surely and uniformly converge since we assume that the bandwidth h_m of these estimators meets the following conditions (see [BOLE]):

(Hyp): $h_m \searrow_m 0$, $mh_m \nearrow_m \infty$, $mh_m/L(h_m^{-1}) \rightarrow_m \infty$ and $L(h_m^{-1})/LLm \rightarrow_m \infty$, with $L(u) = \ln(u \vee e)$.

Let us consider $A_1(m, a) = \frac{1}{m} \sum_{i=1}^m \ln \left\{ \frac{g_{a,m}(a^\top Y_i)}{f_{a,m}(a^\top Y_i)} \right\} \frac{g_{a,m}(a^\top Y_i)}{g_m(Y_i)}$, $A_2(m, a) = \frac{1}{m} \sum_{i=1}^m \left(\frac{g_{a,m}(a^\top X_i)}{f_{a,m}(a^\top X_i)} - 1 \right) \frac{f_{a,m}(a^\top X_i)}{f_m(X_i)}$, $B_1(m, a) = \frac{1}{m} \sum_{i=1}^m \ln \left\{ \frac{f_{a,m}(a^\top Y_i)}{g_{a,m}(a^\top Y_i)} \frac{g_m(Y_i)}{f_m(Y_i)} \right\} \frac{f_{a,m}(a^\top Y_i)}{g_{a,m}(a^\top Y_i)}$ and $B_2(m, a) = \frac{1}{m} \sum_{i=1}^m \left(1 - \frac{f_{a,m}(a^\top X_i)}{g_{a,m}(a^\top X_i)} \frac{g_m(X_i)}{f_m(X_i)} \right)$.

Our goal is to estimate the maximum of $K(g_a, f_a)$ and the minimum of $K(g_{\frac{f_a}{g_a}}, f)$. To achieve this, it is necessary for us to truncate the samples:

Let us consider now a sequence θ_m such that $\theta_m \rightarrow 0$, and $y_m/\theta_m^2 \rightarrow 0$, where y_m is the almost sure convergence rate of the kernel density estimator, i.e. $y_m = O_{\mathbf{P}}(m^{-\frac{2}{4+d}})$ - see lemma 12. We will generate f_m, g_m and $g_{b,m}$ from the starting sample and we will select the X_i and Y_i vectors and such that $f_m(X_i) \geq \theta_m$ and $g_m(Y_i) \geq \theta_m$, for all i and for all $b \in \mathbb{R}_*^d$ - for Huber's algorithm - and such that $f_m(X_i) \geq \theta_m$ and $g_{b,m}(b^\top Y_i) \geq \theta_m$, for all i and for all $b \in \mathbb{R}_*^d$ - for our algorithm.

The vectors meeting these conditions will be called X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_n .

Consequently, the next proposition provides us with the condition required to obtain our estimates

Proposition 11. *Using the notations introduced in [BROKEZ] and in sections 3.1.1 and 4.1, it holds*

$$\sup_{a \in \mathbb{R}_*^d} |(A_1(n, a) - A_2(n, a)) - K(g_a, f_a)| \rightarrow 0 \text{ a.s.}, \quad (7)$$

$$\sup_{a \in \mathbb{R}_*^d} |(B_1(n, a) - B_2(n, a)) - K(g \frac{f_a}{g_a}, f)| \rightarrow 0 \text{ a.s.} \quad (8)$$

Remark 8. We can take for θ_m the expression $m^{-\nu}$, with $0 < \nu < \frac{1}{4+d}$.

C Case study : f is known

In this Annex, we will study the case when f and g are known. We will then use the notations introduced in sections 3.1.1 and 3.1.2 with f and g , i.e. no longer with their kernel estimates.

C.1 Convergence study at the k^{th} step of the algorithm:

In this paragraph, when k is less than or equal to d , we will show that the sequence $(\check{a}_k)_n$ converges towards a_k and that the sequence $(\check{g}^{(k)})_n$ converges towards $g^{(k)}$.

Both $\check{\gamma}_n$ and $\check{c}_n(a)$ are M-estimators and estimate a_k - see [BROKEZ]. We state

Proposition 12. *Assuming (H'1) to (H'3) hold. Both $\sup_{a \in \Theta} \|\check{c}_n(a) - a_k\|$ and $\check{\gamma}_n$ tends to a_k a.s.*

Finally, the following theorem shows us that $\check{g}^{(k)}$ converges uniformly almost everywhere towards $g^{(k)}$, for any $k = 1..d$.

Theorem 14. *Assuming (H'1) to (H'3) hold. Then, $\check{g}^{(k)} \rightarrow_n g^{(k)}$ a.s. and uniformly a.e.*

C.2 Asymptotic Inference at the k^{th} step of the algorithm

The following theorem shows that $\check{g}^{(k)}$ converges at the rate $O_{\mathbf{P}}(n^{-1/2})$ in three different cases, namely for any given x , with the L^1 distance and with the relative entropy:

Theorem 15. *Assuming (H'0) to (H'3) hold, for any $k = 1, \dots, d$ and any $x \in \mathbb{R}^d$, we have*

$$|\check{g}^{(k)}(x) - g^{(k)}(x)| = O_{\mathbf{P}}(n^{-1/2}), \quad (9)$$

$$\int |\check{g}^{(k)}(x) - g^{(k)}(x)| dx = O_{\mathbf{P}}(n^{-1/2}), \quad (10)$$

$$|K(\check{g}^{(k)}, f) - K(g^{(k)}, f)| = O_{\mathbf{P}}(n^{-1/2}). \quad (11)$$

The following theorem shows that the laws of our estimators of a_k , namely $\check{c}_n(a_k)$ and $\check{\gamma}_n$, converge towards a linear combination of Gaussian variables.

Theorem 16. *Assuming that conditions (H'1) to (H'6) hold, then*

$$\sqrt{n} \mathcal{A}(\check{c}_n(a_k) - a_k) \xrightarrow{\mathcal{L}^{aw}} \mathcal{B} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial b} M(a_k, a_k)\|^2) + \mathcal{C} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial a} M(a_k, a_k)\|^2) \text{ and}$$

$$\sqrt{n} \mathcal{A}(\check{\gamma}_n - a_k) \xrightarrow{\mathcal{L}^{aw}} \mathcal{C} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial b} M(a_k, a_k)\|^2) + \mathcal{C} \cdot \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial a} M(a_k, a_k)\|^2)$$

where $\mathcal{A} = (\mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) (\mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a_i \partial b_j} M(a_k, a_k)))$, $\mathcal{C} = \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k)$ and

$$\mathcal{B} = \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a_i \partial b_j} M(a_k, a_k).$$

C.3 A stopping rule for the procedure

We now assume that the algorithm does not stop after d iterations. We then remark that, it still holds - for any $i > d$:

- $g^{(i)}(x) = g(x) \prod_{k=1}^i \frac{f_{a_k}(a_k^\top x)}{[g_n^{(k-1)}]_{a_k}(a_k^\top x)}$, with $g^{(0)} = g$.
- $K(g^{(0)}, f) \geq K(g^{(1)}, f) \geq K(g^{(2)}, f) \dots \geq 0$.
- Theorems 14, 15 and 16.

Moreover, through remark 6 page 13 and as explained in section 14 of [HUB85], the sequence $(K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}), f)_{k \geq 1}$ converges towards zero. Then, in this paragraph, we will show that $g^{(i)}$ converges towards f in i . And finally, we will provide a stopping rule for this identification procedure.

C.3.1 Representation of f

Under $(H'0)$, the following proposition shows us that the probability measure with density $g^{(k)}$ converges towards the probability measure with density f :

Proposition 13. *We have $\lim_k g^{(k)} = f$ a.s.*

C.3.2 Testing of the criteria

Through a test of the criteria, namely $a \mapsto K(g^{(k-1)} \frac{f_a}{g_a^{(k-1)}}), f)$, we will build a stopping rule for this procedure. First, the next theorem enables us to derive the law of the criteria.

Theorem 17. *Assuming that $(H'1)$ to $(H'3)$, $(H'6)$ and $(H'8)$ hold. Then,*

$$\sqrt{n}(\text{Var}_{\mathbf{P}}(M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)))^{-1/2}(\mathbb{P}_n M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n) - \mathbb{P}_n M(a_k, a_k)) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, I),$$

where k represents the k^{th} step of the algorithm and where I is the identity matrix in \mathbb{R}^d .

Note that k is fixed in theorem 17 since $\check{\gamma}_n = \arg \inf_{a \in \Theta} \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$ where M is a known function of k - see section 3.1.1. Thus, in the case where $K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}), f) = 0$, we obtain

Corollary 5. *Assuming that $(H'1)$ to $(H'3)$, $(H'6)$, $(H'7)$ and $(H'8)$ hold. Then,*

$$\sqrt{n}(\text{Var}_{\mathbf{P}}(M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)))^{-1/2}(\mathbb{P}_n M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, I).$$

Hence, we propose the test of the null hypothesis $(H_0) : K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}), f) = 0$ versus $(H_1) : K(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}), f) \neq 0$. Based on this result, we stop the algorithm, then, defining a_k as the last vector generated, we derive from corollary 5 a α -level confidence ellipsoid around a_k , namely $\mathcal{E}_k = \{b \in \mathbb{R}^d; \sqrt{n}(\text{Var}_{\mathbf{P}}(M(b, b)))^{-1/2} \mathbb{P}_n M(b, b) \leq q_\alpha^{\mathcal{N}(0,1)}\}$, where $q_\alpha^{\mathcal{N}(0,1)}$ is the quantile of a α -level reduced centered normal distribution.

Consequently, the following corollary provides us with a confidence region for the above test:

Corollary 6. *\mathcal{E}_k is a confidence region for the test of the null hypothesis (H_0) versus (H_1) .*

D Hypotheses' discussion

D.1 Discussion on $(H'2)$.

We verify this hypothesis in the case where :

- a_1 is the unique element of \mathbb{R}_*^d such that $f(\cdot/a_1^\top x) = g(\cdot/a_1^\top x)$, i.e. $K(g(\cdot/a_1^\top x) f_{a_1}(a_1^\top x), f) = 0, (1)$

- f and g are bounded and positive, (2)
 - there exists a neighborhood V of a_k such that, for all b in V and for all positive real A , there exists $\mathcal{S} > 0$ such that $g(./b'x) \leq \mathcal{S}.f(./b'x)$ with $\|x\| > A$ (3).
- Let us remark that we obtain the same proof with f , $g^{(k-1)}$ and a_k .

First, (1) implies that $g \frac{f_{a_1}}{g_{a_1}} = f$. Hence, $0 > \int \ln(\frac{g}{f} \frac{f_c}{g_c}) g \frac{f_{a_1}}{g_{a_1}} dx = -K(g \frac{f_c}{g_c}, f) > -K(g, f)$ by the very construction of $g \frac{f_c}{g_c}$. Besides, (2) and (3) imply that there exists a neighborhood V of a_k such that, for all c in V , there exists $\mathcal{S} > 0$ such that, for all x in \mathbb{R}^d , $g(./c'x) \leq \mathcal{S}.f(./c'x)$. Consequently, we get $|M(c, a_1, x)| \leq |-K(g, f)| + |-(\frac{g(./c'x)}{f(./c'x)} - 1)| \leq K(g, f) + \mathcal{S} + 1$. Finally, we obtain the existence a neighborhood V of a_k such that, for all c in V ,

$$|M(c, a_k, x)| \leq H(x) = K(g, f) + \mathcal{S} + 1 \text{ (P - a.s.) with } \mathbf{P}H < \infty.$$

D.2 Discussion on $(H'3)$.

We verify this hypothesis in the case where a_1 is the unique element of \mathbb{R}_*^d such that $f(./a_1^\top x) = g(./a_1^\top x)$, i.e. $K(g(./a_1^\top x)f_{a_1}(a_1^\top x), f) = 0$ - we obtain the same proof with f , $g^{(k-1)}$ and a_k .

Preliminary (A):

Shows that $A = \{(c, x) \in \mathbb{R}_^d \setminus \{a_1\} \times \mathbb{R}^d; \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} > \frac{f_c(c^\top x)}{g_c(c^\top x)} \text{ and } g(x) \frac{f_c(c^\top x)}{g_c(c^\top x)} > f(x)\} = \emptyset$ through a reductio ad absurdum, i.e. if we assume $A \neq \emptyset$.*

Thus, our hypothesis enables us to derive

$f(x) = f(./a_1^\top x)f_{a_1}(a_1^\top x) = g(./a_1^\top x)f_{a_1}(a_1^\top x) > g(./c^\top x)f_c(c^\top x) > f$
 since $\frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} \geq \frac{f_c(c^\top x)}{g_c(c^\top x)}$ implies $g(./a_1^\top x)f_{a_1}(a_1^\top x) = g(x) \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} \geq g(x) \frac{f_c(c^\top x)}{g_c(c^\top x)} = g(./c^\top x)f_c(c^\top x)$,
 i.e. $f > f$. We can therefore conclude.

Preliminary (B):

Shows that $B = \{(c, x) \in \mathbb{R}_^d \setminus \{a_1\} \times \mathbb{R}^d; \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} < \frac{f_c(c^\top x)}{g_c(c^\top x)} \text{ and } g(x) \frac{f_c(c^\top x)}{g_c(c^\top x)} < f(x)\} = \emptyset$ through a reductio ad absurdum, i.e. if we assume $B \neq \emptyset$.*

Thus, our hypothesis enables us to derive

$$f(x) = f(./a_1^\top x)f_{a_1}(a_1^\top x) = g(./a_1^\top x)f_{a_1}(a_1^\top x) < g(./c^\top x)f_c(c^\top x) < f$$

We can therefore conclude as above.

Let us now prove $(H'3)$:

We have $PM(c, a_1) - PM(c, a) = \int \ln(\frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)}) \{ \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} - \frac{f_c(c^\top x)}{g_c(c^\top x)} \} g(x) dx$. Moreover, the logarithm \ln is negative on $\{x \in \mathbb{R}_*^d; \frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)} < 1\}$ and is positive on $\{x \in \mathbb{R}_*^d; \frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)} \geq 1\}$.

Thus, the preliminary studies (A) and (B) show that $\ln(\frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)})$ and $\{ \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} - \frac{f_c(c^\top x)}{g_c(c^\top x)} \}$ always present a negative product. We can therefore conclude, since $(c, a) \mapsto PM(c, a_1) - PM(c, a)$ is not null for all c and for all $a \neq a_1$. \square

E Proofs

This last section includes the proofs of most of the lemmas, propositions, theorems and corollaries contained in the present article.

Remark 9. 1/ $(H'0)$ - according to which f and g are assumed to be positive and bounded - through lemma 5 (see page 19) implies that $\check{g}^{(k)}$ and $\hat{g}^{(k)}$ are positive and bounded.

2/ remark 4 implies that f_n , g_n , $\check{g}^{(k)}$ and $\hat{g}^{(k)}$ are positive and bounded since we consider a Gaussian kernel.

Proof of propositions 6 and 7. Let us first study proposition 7.

Without loss of generality, we will prove this proposition with x_1 in lieu of $a^\top X$.

Let us define $g^* = gr$. We remark that g and g^* present the same density conditionally to x_1 . Indeed,

$$g_1^*(x_1) = \int g^*(x) dx_2 \dots dx_d = \int r(x_1) g(x) dx_2 \dots dx_d = r(x_1) \int g(x) dx_2 \dots dx_d = r(x_1) g_1(x_1).$$

Thus, we can demonstrate this proposition.

We have $g(\cdot|x_1) = \frac{g(x_1, \dots, x_n)}{g_1(x_1)}$ and $g_1(x_1)r(x_1)$ is the marginal density of g^* . Hence,

$$\int g^* dx = \int g_1(x_1) r(x_1) g(\cdot|x_1) dx = \int g_1(x_1) \frac{f_1(x_1)}{g_1(x_1)} (\int g(\cdot|x_1) dx_2 \dots dx_d) dx_1 = \int f_1(x_1) dx_1 = 1 \text{ and since } g^* \text{ is positive, then } g^* \text{ is a density. Moreover,}$$

$$K(f, g^*) = \int f \{ \ln(f) - \ln(g^*) \} dx, \quad (12)$$

$$\begin{aligned} &= \int f \{ \ln(f(\cdot|x_1)) - \ln(g^*(\cdot|x_1)) + \ln(f_1(x_1)) - \ln(g_1(x_1)r(x_1)) \} dx, \\ &= \int f \{ \ln(f(\cdot|x_1)) - \ln(g(\cdot|x_1)) + \ln(f_1(x_1)) - \ln(g_1(x_1)r(x_1)) \} dx, \end{aligned} \quad (13)$$

as $g^*(\cdot|x_1) = g(\cdot|x_1)$. Since the minimum of this last equation (13) is reached through the minimization of $\int f \{ \ln(f_1(x_1)) - \ln(g_1(x_1)r(x_1)) \} dx = K(f_1, g_1 r)$, then property 2 necessarily implies that $f_1 = g_1 r$, hence $r = f_1/g_1$.

Finally, we have $K(f, g) - K(f, g^*) = \int f \{ \ln(f_1(x_1)) - \ln(g_1(x_1)) \} dx = K(f_1, g_1)$, which completes the demonstration of proposition 7.

Similarly, if we replace $f^* = fr^{-1}$ with f and g with g^* , we obtain the proof of proposition 6. \square

Proof of proposition 8. The demonstration is very similar to the one for proposition 7, save for the fact we now base our reasoning at row (12) on $\int g \{ \ln(g^*) - \ln(f) \} dx$ instead of $K(f, g^*) = \int f \{ \ln(f) - \ln(g^*) \} dx$. \square

Proof of proposition 2. The demonstration is also very similar to the one for proposition 7, save for the fact we now base our reasoning at row 12 on $K(g^*, f) = \int g^* \{ \ln(f) - \ln(g^*) \} dx$ instead of $K(f, g^*) = \int f \{ \ln(f) - \ln(g^*) \} dx$. \square

Proof of lemma 7.

Lemma 7. *If the family $(a_i)_{i=1\dots d}$ is a basis of \mathbb{R}^d then*

$$g(\cdot/a_1^\top x, \dots, a_j^\top x) = n(a_{j+1}^\top x, \dots, a_d^\top x) = f(\cdot/a_1^\top x, \dots, a_j^\top x).$$

Putting $A = (a_1, \dots, a_d)$, let us determine f in the A basis. Let us first study the function defined by $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $x \mapsto (a_1^\top x, \dots, a_d^\top x)$. We can immediately say that ψ is continuous and since A is a basis, its bijectivity is obvious. Moreover, let us study its Jacobian.

$$\text{By definition, it is } J_\psi(x_1, \dots, x_d) = \begin{vmatrix} \frac{\partial \psi_1}{\partial x_1} & \dots & \frac{\partial \psi_1}{\partial x_d} \\ \dots & \dots & \dots \\ \frac{\partial \psi_d}{\partial x_1} & \dots & \frac{\partial \psi_d}{\partial x_d} \end{vmatrix} = \begin{vmatrix} a_{1,1} & \dots & a_{1,d} \\ \dots & \dots & \dots \\ a_{d,1} & \dots & a_{d,d} \end{vmatrix} = |A| \neq 0 \text{ since } A \text{ is a}$$

basis. We can therefore infer : $\forall x \in \mathbb{R}^d$, $\exists! y \in \mathbb{R}^d$ such that $f(x) = |A|^{-1} \Psi(y)$, i.e. Ψ (resp. y) is the expression of f (resp of x) in basis A , namely $\Psi(y) = \tilde{n}(y_{j+1}, \dots, y_d) \tilde{h}(y_1, \dots, y_j)$, with \tilde{n} and \tilde{h} being the expressions of n and h in basis A . Consequently, our results in the case where the family $\{a_j\}_{1 \leq j \leq d}$ is the canonical basis of \mathbb{R}^d , still hold for Ψ in the A basis - see section 2.1.2. And then,

if \tilde{g} is the expression of g in basis A , we have $\tilde{g}(\cdot/y_1, \dots, y_j) = \tilde{n}(y_{j+1}, \dots, y_d) = \Psi(\cdot/y_1, \dots, y_j)$, i.e. $g(\cdot/a_1^\top x, \dots, a_j^\top x) = n(a_{j+1}^\top x, \dots, a_d^\top x) = f(\cdot/a_1^\top x, \dots, a_j^\top x)$. \square

Proof of lemma 8. By definition of the closure of a set, we have

Lemma 8. *The set Γ_c is closed in L^1 for the topology of the uniform convergence.*

Proof of lemma 9. Since K is greater than the L^1 distance, we have

Lemma 9. *For all $c > 0$, we have $\Gamma_c \subset \overline{B}_{L^1}(f, c)$, where $B_{L^1}(f, c) = \{p \in L^1; \|f - p\|_1 \leq c\}$.*

Proof of lemma 10. The definition of the closure of a set and lemma 6 (see page 19) imply

Lemma 10. *G is closed in L^1 for the topology of the uniform convergence.*

Proof of lemma 11.

Lemma 11. *$\inf_{a \in \mathbb{R}_*^d} K(g_{\frac{f_a}{g_a}}, f)$ is reached.*

Indeed, let G be $\{g_{\frac{f_a}{g_a}}; a \in \mathbb{R}_*^d\}$ and Γ_c be $\Gamma_c = \{p; K(p, f) \leq c\}$ for all $c > 0$. From lemmas 8, 9 and 10 (see page 24), we get $\Gamma_c \cap G$ is a compact for the topology of the uniform convergence, if $\Gamma_c \cap G$ is not empty. Hence, and since property 3 (see page 18) tells us that $Q \mapsto K(Q, P)$ is lower semi-continuous in L^1 for the topology of the uniform convergence, then the infimum is reached in L^1 . (Taking for example $c = K(g, f)$, Ω is necessarily not empty because we always have $K(g_{\frac{f_a}{g_a}}, f) \leq K(g, f)$). \square

Proof of lemma 12.

Lemma 12. *For any continuous density f , we have $y_m = |f_m(x) - f(x)| = O_{\mathbf{P}}(m^{-\frac{2}{4+d}})$.*

Defining $b_m(x)$ as $b_m(x) = |E(f_m(x)) - f(x)|$, we have $y_m \leq |f_m(x) - E(f_m(x))| + b_m(x)$. Moreover, from page 150 of [SCOTT92], we derive that $b_m(x) = O_{\mathbf{P}}(\sum_{j=1}^d h_j^2)$ where $h_j = O_{\mathbf{P}}(m^{-\frac{1}{4+d}})$. Then, we obtain $b_m(x) = O_{\mathbf{P}}(m^{-\frac{2}{4+d}})$. Finally, since the central limit theorem rate is $O_{\mathbf{P}}(m^{-\frac{1}{2}})$, we then obtain that $y_m \leq O_{\mathbf{P}}(m^{-\frac{1}{2}}) + O_{\mathbf{P}}(m^{-\frac{2}{4+d}}) = O_{\mathbf{P}}(m^{-\frac{2}{4+d}})$. \square

Proof of proposition 11. Let us first remark that we have $f(X_i) \geq \theta_n - y_n$, $g(Y_i) \geq \theta_n - y_n$ and $g_b(b^\top Y_i) \geq \theta_n - y_n$, for all i and for all $b \in \mathbb{R}_*^d$, thanks to the uniform convergence of the kernel estimators. Indeed, we have $f(X_i) = f(X_i) - f_n(X_i) + f_n(X_i) \geq -y_n + f_n(X_i)$, by definition of y_n , and then $f(X_i) \geq -y_n + \theta_n$, by hypothesis on $f_n(X_i)$. This is also true for g_n and $g_{b,n}$. This entails:

Lemma 13. $\sup_{b \in \mathbb{R}_*^d} |\frac{1}{n} \sum_{i=1}^n \{\frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1\} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \int \{\frac{g_b(b^\top x)}{f_b(b^\top x)} - 1\} f_b(b^\top x) dx| \rightarrow 0$ a.s.,
 $\sup_{b \in \mathbb{R}_*^d} |\frac{1}{n} \sum_{i=1}^n \ln \{\frac{g_{a,n}(a^\top Y_i)}{f_{a,n}(a^\top Y_i)}\} \frac{g_{a,n}(a^\top Y_i)}{g_n(Y_i)} - \int \ln(\frac{g_a(a^\top x)}{f_a(a^\top x)}) g_a(a^\top x) dx| \rightarrow 0$ a.s. and
 $\lim_{n \rightarrow \infty} \sup_{a \in \mathbb{R}_*^d} |(B_1(n, a) - B_2(n, a)) - K(g_{\frac{f_a}{g_a}}, f)| = 0$ a.s. with
 $B_1(n, a) = \frac{1}{n} \sum_{i=1}^n \ln \{\frac{f_{a,n}(a^\top Y_i)}{g_{a,n}(a^\top Y_i)} \frac{g_n(Y_i)}{f_n(Y_i)}\} \frac{f_{a,n}(a^\top Y_i)}{g_{a,n}(a^\top Y_i)}$ and $B_2(n, a) = \frac{1}{n} \sum_{i=1}^n (1 - \{\frac{f_{a,n}(a^\top X_i)}{g_{a,n}(a^\top X_i)} \frac{g_n(X_i)}{f_n(X_i)}\})$.

Proof :

Let us remark that

$$\begin{aligned} & |\frac{1}{n} \sum_{i=1}^n \{\frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1\} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \int \{\frac{g_b(b^\top x)}{f_b(b^\top x)} - 1\} f_b(b^\top x) dx| \\ &= |\frac{1}{n} \sum_{i=1}^n \{\frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1\} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \frac{1}{n} \sum_{i=1}^n \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1\} \frac{f_b(b^\top X_i)}{f(X_i)} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1\} \frac{f_b(b^\top X_i)}{f(X_i)} - \int \{\frac{g_b(b^\top x)}{f_b(b^\top x)} - 1\} f_b(b^\top x) dx| \\ &\leq |\frac{1}{n} \sum_{i=1}^n \{\frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1\} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \frac{1}{n} \sum_{i=1}^n \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1\} \frac{f_b(b^\top X_i)}{f(X_i)}| \end{aligned}$$

$$+ |\frac{1}{n} \sum_{i=1}^n \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1| \frac{f_b(b^\top X_i)}{f(X_i)} - \int \{ \frac{g_b(b^\top x)}{f_b(b^\top x)} - 1 \} f_b(b^\top x) dx|$$

Moreover, since $\int |\{ \frac{g_b(b^\top x)}{f_b(b^\top x)} - 1 \} f_b(b^\top x)| dx \leq 2$, the law of large numbers enables us to state:

$$|\frac{1}{n} \sum_{i=1}^n \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1| \frac{f_b(b^\top X_i)}{f(X_i)} - \int \{ \frac{g_b(b^\top x)}{f_b(b^\top x)} - 1 \} f_b(b^\top x) dx| \rightarrow 0 \text{ a.s.}$$

$$\text{Moreover, } |\frac{1}{n} \sum_{i=1}^n \{ \frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1 \} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \frac{1}{n} \sum_{i=1}^n \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1| \frac{f_b(b^\top X_i)}{f(X_i)}| \\ \leq \frac{1}{n} \sum_{i=1}^n |\{ \frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1 \} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \{ \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1 \} \frac{f_b(b^\top X_i)}{f(X_i)}|$$

and

$$|\{ \frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1 \} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \{ \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1 \} \frac{f_b(b^\top X_i)}{f(X_i)}| = |\frac{g_{b,n}(b^\top X_i) - f_{b,n}(b^\top X_i)}{f_n(X_i)} - \frac{g_b(b^\top X_i) - f_b(b^\top X_i)}{f(X_i)}| \\ \leq \frac{1}{|f(X_i)| \cdot |f_n(X_i)|} \{ |f(X_i)| \cdot |g_{b,n}(b^\top X_i) - g_b(b^\top X_i)| + |f(X_i) - f_n(X_i)| \cdot |g_b(b^\top X_i)| \\ + |f(X_i)| \cdot |f_{b,n}(b^\top X_i) - f_b(b^\top X_i)| + |f(X_i) - f_n(X_i)| \cdot |f_b(b^\top X_i)| \}, \text{ through the}$$

introduction of terms $g_b f - g_b f$ and $f f_b - f f_b$,

$$\leq \frac{O_{\mathbf{P}}(1) \cdot y_n}{\theta_n \cdot (\theta_n - y_n)} = O_{\mathbf{P}}(1) \frac{1}{\frac{y_n}{\theta_n^2} - \theta_n}, \text{ by definitions of } \theta_n \text{ and } y_n,$$

$\rightarrow 0$, a.s. because, $\frac{y_n}{\theta_n^2} \rightarrow 0$ a.s., by hypothesis on θ_n .

Consequently, $\frac{1}{n} \sum_{i=1}^n |\{ \frac{g_{b,n}(b^\top X_i)}{f_{b,n}(b^\top X_i)} - 1 \} \frac{f_{b,n}(b^\top X_i)}{f_n(X_i)} - \{ \frac{g_b(b^\top X_i)}{f_b(b^\top X_i)} - 1 \} \frac{f_b(b^\top X_i)}{f(X_i)}| \rightarrow 0$, as it is a Cesàro mean.

This enables us to conclude.

Similarly, we show that :

• $\sup_{b \in \mathbb{R}_*^d} |\frac{1}{n} \sum_{i=1}^n \ln \{ \frac{g_{a,n}(a^\top Y_i)}{f_{a,n}(a^\top Y_i)} \} \frac{g_{a,n}(a^\top Y_i)}{g_n(Y_i)} - \int \ln(\frac{g_a(a^\top x)}{f_a(a^\top x)}) g_a(a^\top x) dx| \rightarrow 0$ a.s. and then limit 7 page 20 holds,

• $\lim_{m \rightarrow \infty} \sup_{a \in \mathbb{R}_*^d} |(B_1(m, a) - B_2(m, a)) - K(g_{\frac{f_a}{g_a}}, f)| = 0$ a.s. and then limit 8 page 20 holds. \square

Proof of lemma 14.

Lemma 14. For any $p \leq d$, we have $f_{a_p}^{(p-1)} = f_{a_p}$ - see Huber's analytic method -, $g_{a_p}^{(p-1)} = g_{a_p}$ - see Huber's synthetic method - and $g_{a_p}^{(p-1)} = g_{a_p}$ - see our algorithm.

Proof :

As it is equivalent to prove either our algorithm or Huber's, we will only develop here the proof for our algorithm. Assuming, without any loss of generality, that the a_i , $i = 1, \dots, p$, are the vectors of the canonical basis, since $g^{(p-1)}(x) = g(x) \frac{f_1(x_1)}{g_1(x_1)} \frac{f_2(x_2)}{g_2(x_2)} \dots \frac{f_{p-1}(x_{p-1})}{g_{p-1}(x_{p-1})}$ we derive immediately that $g_p^{(p-1)} = g_p$. We note that it is sufficient to operate a change in basis on the a_i to obtain the general case. \square

Proof of lemma 15.

Lemma 15. If there exists p , $p \leq d$, such that $K(g^{(p)}, f) = 0$, then the family of $(a_i)_{i=1, \dots, p}$ - derived from the construction of $g^{(p)}$ - is free and orthogonal.

Proof :

Without any loss of generality, let us assume that $p = 2$ and that the a_i are the vectors of the canonical basis. Using a reductio ad absurdum with the hypotheses $a_1 = (1, 0, \dots, 0)$ and that $a_2 = (\alpha, 0, \dots, 0)$, where $\alpha \in \mathbb{R}$, we get $g^{(1)}(x) = g(x_2, \dots, x_d/x_1) f_1(x_1)$ and $f = g^{(2)}(x) = g(x_2, \dots, x_d/x_1) f_1(x_1) \frac{f_{\alpha a_1}(\alpha x_1)}{[g^{(1)}]_{\alpha a_1}(\alpha x_1)}$. Hence $f(x_2, \dots, x_d/x_1) = g(x_2, \dots, x_d/x_1) \frac{f_{\alpha a_1}(\alpha x_1)}{[g^{(1)}]_{\alpha a_1}(\alpha x_1)}$. It consequently implies that $f_{\alpha a_1}(\alpha x_1) = [g^{(1)}]_{\alpha a_1}(\alpha x_1)$ since

$$1 = \int f(x_2, \dots, x_d/x_1) dx_2 \dots dx_d = \int g(x_2, \dots, x_d/x_1) dx_2 \dots dx_d \frac{f_{\alpha a_1}(\alpha x_1)}{[g^{(1)}]_{\alpha a_1}(\alpha x_1)} = \frac{f_{\alpha a_1}(\alpha x_1)}{[g^{(1)}]_{\alpha a_1}(\alpha x_1)}.$$

Therefore, $g^{(2)} = g^{(1)}$, i.e. $p = 1$ which leads to a contradiction. Hence, the family is free.

Moreover, using a reductio ad absurdum we get the orthogonality. Indeed, we have

$$\int f(x) dx = 1 \neq +\infty = \int n(a_{j+1}^\top x, \dots, a_d^\top x) h(a_1^\top x, \dots, a_j^\top x) dx. \quad \square$$

Proof of lemma 16.

Lemma 16. We have $\Theta = \{b \in \Theta \mid \int (\frac{g(x)f_b(b^\top x)}{f(x)g_b(b^\top x)} - 1)f(x)dx < \infty\}$.

We get the result since $\int (\frac{g(x)f_b(b^\top x)}{f(x)g_b(b^\top x)} - 1)f(x) dx = \int (\frac{g(x)f_b(b^\top x)}{g_b(b^\top x)} - f(x)) dx = 0$. \square

Proof of propositions 12. In the same manner as in Proposition 3.4 of [BROKEZ], we prove this proposition through lemma 16. \square

Proof of propositions 4 and 10. Proposition 4 comes immediately from proposition 11 page 19 and lemma 12 page 20. Similarly, we prove proposition 10 since both $\sup_{a \in \Theta} \|\hat{b}_n(a) - a_k\|$ and $\hat{\beta}_n$ converge toward a_k a.s in the case where f is known - see also in annex C where the case when f is known has been carried out in our algorithm. \square

Proof of theorem 14. Using lemma 6 page 19 and since, for any k , $g^{(k)} = g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}$, we prove this theorem by induction. \square

Proof of theorems 1 and 8. We prove the theorem by induction. First, by the very definition of the kernel estimator $\check{g}_n^{(0)} = g_n$ converges towards g . Moreover, the continuity of $a \mapsto f_{a,n}$ and $a \mapsto g_{a,n}$ and proposition 4 imply that $\check{g}_n^{(1)} = \check{g}_n^{(0)} \frac{f_{a,n}}{g_{a,n}^{(0)}}$ converges towards $g^{(1)}$. Finally, since, for any k , $\check{g}_n^{(k)} = \check{g}_n^{(k-1)} \frac{f_{\check{a}_k,n}}{g_{\check{a}_k,n}^{(k-1)}}$, we conclude similarly as for $\check{g}_n^{(1)}$. In a similar manner, we prove theorem 8. \square

Proof of theorem 15.

relationship (9). Let us consider $\Psi_j = \{\frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} - \frac{f_{a_j}(a_j^\top x)}{[g^{(j-1)}]_{a_j}(a_j^\top x)}\}$. Since f and g are bounded, it is easy to prove that from a certain rank, we get, for any x given in \mathbb{R}^d

$$|\Psi_j| \leq \max(\frac{1}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)}, \frac{1}{[g^{(j-1)}]_{a_j}(a_j^\top x)}) |f_{\check{a}_j}(\check{a}_j^\top x) - f_{a_j}(a_j^\top x)|.$$

Remark 10. First, based on what we stated earlier, for any given x and from a certain rank, there is a constant $R > 0$ independent from n , such that $\max(\frac{1}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)}, \frac{1}{[g^{(j-1)}]_{a_j}(a_j^\top x)}) \leq R = R(x) = O(1)$. Second, since \check{a}_k is an M -estimator of a_k , its convergence rate is $O_{\mathbf{P}}(n^{-1/2})$.

Thus using simple functions, we obtain an upper and lower bound for $f_{\check{a}_j}$ and for f_{a_j} and we reach the following conclusion:

$$|\Psi_j| \leq O_{\mathbf{P}}(n^{-1/2}). \quad (14)$$

We finally obtain:

$$|\prod_{j=1}^k \frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} - \prod_{j=1}^k \frac{f_{a_j}(a_j^\top x)}{[g^{(j-1)}]_{a_j}(a_j^\top x)}| = \prod_{j=1}^k \frac{f_{a_j}(a_j^\top x)}{[g^{(j-1)}]_{a_j}(a_j^\top x)} |\prod_{j=1}^k \frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} \frac{[g^{(j-1)}]_{a_j}(a_j^\top x)}{f_{a_j}(a_j^\top x)} - 1|.$$

Based on relationship (14), the expression $\frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} \frac{[g^{(j-1)}]_{a_j}(a_j^\top x)}{f_{a_j}(a_j^\top x)}$ tends towards 1 at a rate of $O_{\mathbf{P}}(n^{-1/2})$ for all j . Consequently, $\prod_{j=1}^k \frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} \frac{[g^{(j-1)}]_{a_j}(a_j^\top x)}{f_{a_j}(a_j^\top x)}$ tends towards 1 at a rate of $O_{\mathbf{P}}(n^{-1/2})$. Thus from a certain rank, we get

$$|\prod_{j=1}^k \frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} - \prod_{j=1}^k \frac{f_{a_j}(a_j^\top x)}{[g^{(j-1)}]_{a_j}(a_j^\top x)}| = O_{\mathbf{P}}(n^{-1/2}) O_{\mathbf{P}}(1) = O_{\mathbf{P}}(n^{-1/2}).$$

In conclusion, we obtain $|\check{g}^{(k)}(x) - g^{(k)}(x)| = g(x) |\prod_{j=1}^k \frac{f_{\check{a}_j}(\check{a}_j^\top x)}{[\check{g}^{(j-1)}]_{\check{a}_j}(\check{a}_j^\top x)} - \prod_{j=1}^k \frac{f_{a_j}(a_j^\top x)}{[g^{(j-1)}]_{a_j}(a_j^\top x)}| \leq O_{\mathbf{P}}(n^{-1/2})$.

relationship (10). The relationship 9 of theorem 15 implies that $|\frac{\check{g}^{(k)}(x)}{g^{(k)}(x)} - 1| = O_{\mathbf{P}}(n^{-1/2})$ because, for any given x , $g^{(k)}(x) |\frac{\check{g}^{(k)}(x)}{g^{(k)}(x)} - 1| = |\check{g}^{(k)}(x) - g^{(k)}(x)|$. Consequently, there exists a smooth function C of \mathbb{R}^d in \mathbb{R}^+ such that $\lim_{n \rightarrow \infty} n^{-1/2} C(x) = 0$ and $|\frac{\check{g}^{(k)}(x)}{g^{(k)}(x)} - 1| \leq n^{-1/2} C(x)$, for any x . We then have $\int |\check{g}^{(k)}(x) - g^{(k)}(x)| dx = \int g^{(k)}(x) |\frac{\check{g}^{(k)}(x)}{g^{(k)}(x)} - 1| dx \leq \int g^{(k)}(x) C(x) n^{-1/2} dx$.

Moreover, $\sup_{x \in \mathbb{R}^d} |\check{g}^{(k)}(x) - g^{(k)}(x)| = \sup_{x \in \mathbb{R}^d} g^{(k)}(x) |\frac{\check{g}^{(k)}(x)}{g^{(k)}(x)} - 1|$
 $= \sup_{x \in \mathbb{R}^d} g^{(k)}(x) C(x) n^{-1/2} \rightarrow 0$ a.s., by theorem 14.

This implies that $\sup_{x \in \mathbb{R}^d} g^{(k)}(x) C(x) < \infty$ a.s., i.e. $\sup_{x \in \mathbb{R}^d} C(x) < \infty$ a.s. since $g^{(k)}$ has been

assumed to be positive and bounded - see remark 9.

Thus, $\int g^{(k)}(x)C(x)dx \leq \sup C \cdot \int g^{(k)}(x)dx = \sup C < \infty$ since $g^{(k)}$ is a density, therefore we can conclude $\int |\check{g}^{(k)}(x) - g^{(k)}(x)|dx \leq \sup C \cdot n^{-1/2} = O_{\mathbf{P}}(n^{-1/2})$. \square

relationship (11). We have

$K(\check{g}^{(k)}, f) - K(g^{(k)}, f) = \int f(\varphi(\frac{\check{g}^{(k)}}{f}) - \varphi(\frac{g^{(k)}}{f}))dx \leq \int f S |\frac{\check{g}^{(k)}}{f} - \frac{g^{(k)}}{f}|dx = S \int |\check{g}^{(k)} - g^{(k)}|dx$ with the line before last being derived from theorem 13 page 18 and where $\varphi : x \mapsto x \ln(x) - x + 1$ is a convex function and where $S > 0$. We get the same expression as the one found in our Proof of Relationship (10) section, we then obtain $K(\check{g}^{(k)}, f) - K(g^{(k)}, f) \leq O_{\mathbf{P}}(n^{-1/2})$. Similarly, we get $K(g^{(k)}, f) - K(\check{g}^{(k)}, f) \leq O_{\mathbf{P}}(n^{-1/2})$. We can therefore conclude. \square

Proof of lemma 17.

Lemma 17. *With the notations introduced in Appendix B. It holds $n = O(m^2)$.*

Proof :

Let us first study the Huber's case.

Let N be the random variable such that $N = \sum_{j=1}^m \mathbf{1}_{\{f_m(X_j) \geq \theta_m, g_m(Y_j) \geq \theta_m\}}$. The very definition of X_1, \dots, X_m and of Y_1, \dots, Y_m and the fact that the X_j the Y_j are independent, enable us to derive that $n = m \cdot \mathbf{P}(f_m(X_j) \geq \theta_m, g_m(Y_j) \geq \theta_m) = m \cdot \mathbf{P}(f_m(X_j) \geq \theta_m) \cdot \mathbf{P}(g_m(Y_j) \geq \theta_m)$.

Consequently, let us study $\mathbf{P}(f_m(X_i) \geq \theta_m)$.

Let $(\xi_i)_{i=1\dots m}$ be the sequence such that, for any i and any x in \mathbb{R}^d ,

$$\xi_i(x) = \prod_{l=1}^d \frac{1}{(2\pi)^{1/2}h_l} e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} - \int \prod_{l=1}^d \frac{1}{(2\pi)^{1/2}h_l} e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} f(x)dx.$$

Hence, for any given j and conditionally to $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_m$, the variables $(\xi_i(X_j))_{i=1\dots m}$ are i.i.d. and centered, have same second moment, and are such that

$$|\xi_i(X_j)| \leq \prod_{l=1}^d \frac{1}{(2\pi)^{1/2}h_l} + \prod_{l=1}^d \frac{1}{(2\pi)^{1/2}h_l} \int |f(x)|dx = 2 \cdot (2\pi)^{-d/2} \prod_{l=1}^d h_l^{-1}$$

since $\sup_x e^{-\frac{1}{2}x^2} \leq 1$.

Moreover, noting that $f_m(x) = \frac{1}{m} \sum_{i=1}^m \xi_i(x) + (2\pi)^{-d/2} \frac{1}{m} \sum_{i=1}^m \prod_{l=1}^d h_l^{-1} \int e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} f(x)dx$, we have

$$\begin{aligned} f_m(X_j) \geq \theta_m &\Leftrightarrow \frac{1}{m} \sum_{i=1}^m \xi_i(X_j) + (2\pi)^{-d/2} \frac{1}{m} \sum_{i=1}^m \prod_{l=1}^d h_l^{-1} \int e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} f(x)dx \geq \theta_m \\ &\Leftrightarrow \frac{1}{m-1} \sum_{i \neq j}^m \xi_i(X_j) \geq (\theta_m - (2\pi)^{-d/2} \frac{1}{m} \sum_{i=1}^m \prod_{l=1}^d h_l^{-1} \int e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} f(x)dx - \frac{1}{m} \xi_j(X_j)) \frac{m}{m-1} \end{aligned}$$

with $\xi_j(X_j) = 0$. Then, defining t (resp. ε) as $t = 2 \cdot (2\pi)^{-d/2} \prod_{l=1}^d h_l^{-1}$ (resp.

$\varepsilon = (\theta_m - (2\pi)^{-d/2} \prod_{l=1}^d h_l^{-1} \frac{1}{m} \sum_{i=1}^m \prod_{l=1}^d h_l^{-1} \int e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} f(x)dx) \frac{m}{m-1}$), the Bennet's inequality -[DEVG85] page 160- implies that $\mathbf{P}(\frac{1}{m-1} \sum_{i \neq j}^m \xi_i(X_j) \geq \varepsilon / X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_m) \leq 2 \cdot \exp(-\frac{(m-1)\varepsilon^2}{4t^2})$.

Finally, since the X_i are i.i.d. and since $\int (\prod_{l=1}^d e^{-\frac{1}{2}(\frac{x_l - y_l}{h_l})^2} f(x)dx) f(y)dy < 1$, then the law of large numbers says that $\frac{1}{m} \sum_{i=1}^m \int \prod_{l=1}^d e^{-\frac{1}{2}(\frac{x_l - X_{il}}{h_l})^2} f(x)dx \rightarrow_m \int \prod_{l=1}^d e^{-\frac{1}{2}(\frac{x_l - y_l}{h_l})^2} f(x)f(y)dxdy$ a.s. Consequently, since $0 < \nu < \frac{1}{4+d}$, we then obtain $-\frac{(m-1)\varepsilon^2}{4t^2} \sim_{\infty} -m \cdot \frac{1}{4} \int \prod_{l=1}^d e^{-\frac{1}{2}(\frac{x_l - y_l}{h_l})^2} f(x)f(y)dxdy$, i.e. the limit of $-\frac{(m-1)\varepsilon^2}{4t^2}$ is $-\infty$. Thus, from a certain rank, we have $e^{a_m} = O(1)$, i.e.

$\mathbf{P}(f_m(Y_j) \geq \theta_m) = O(1)$. Similarly, we get $\mathbf{P}(g_m(Y_j) \geq \theta_m) = O(1)$.

In conclusion, we can say that $n = m \cdot \mathbf{P}(f_m(X_j) \geq \theta_m) \cdot \mathbf{P}(g_m(Y_j) \geq \theta_m) = O(m^2)$.

In the same manner, we derive the same result as above for our method. \square

Proof of theorems 2 and 9. First, from lemma 12, we derive that, for any x , $\sup_{a \in \mathbb{R}_+^d} |f_{a,n}(a^\top x) - f_a(a^\top x)| = O_{\mathbf{P}}(n^{-\frac{2}{4+d}})$. Then, let us consider $\Psi_j = \frac{f_{\check{a}_j,n}(\check{a}_j^\top x)}{\check{g}_{\check{a}_j,n}^{(j-1)}(\check{a}_j^\top x)} - \frac{f_{a_j}(a_j^\top x)}{g_{a_j}^{(j-1)}(a_j^\top x)}$, we have

$$\Psi_j = \frac{1}{\check{g}_{\check{a}_j, n}^{(j-1)}(\check{a}_j^\top x) g_{a_j}^{(j-1)}(a_j^\top x)} ((f_{\check{a}_j, n}(\check{a}_j^\top x) - f_{a_j}(a_j^\top x)) g_{a_j}^{(j-1)}(a_j^\top x) + f_{a_j}(a_j^\top x) (g_{a_j}^{(j-1)}(a_j^\top x) - \check{g}_{\check{a}_j, n}^{(j-1)}(\check{a}_j^\top x))),$$

i.e. $|\Psi_j| = O_{\mathbf{P}}(n^{-\frac{1}{2}\mathbf{1}_{d=1} - \frac{2}{4+d}\mathbf{1}_{d>1}})$ since $f_{a_j}(a_j^\top x) = O(1)$ and $g_{a_j}^{(j-1)}(a_j^\top x) = O(1)$. We can therefore conclude similarly as in theorem 15 and through lemma 17. Similarly, we prove theorem 9. \square

Proof of theorem 16. First of all, let us remark that hypotheses $(H'1)$ to $(H'3)$ imply that $\check{\gamma}_n$ and $\check{c}_n(a_k)$ converge towards a_k in probability.

Hypothesis $(H'4)$ enables us to derive under the integrable sign after calculation,

$$\begin{aligned} \mathbf{P} \frac{\partial}{\partial b} M(a_k, a_k) &= \mathbf{P} \frac{\partial}{\partial a} M(a_k, a_k) = 0, \\ \mathbf{P} \frac{\partial^2}{\partial a_i \partial b_j} M(a_k, a_k) &= \mathbf{P} \frac{\partial^2}{\partial b_j \partial a_i} M(a_k, a_k) = \int \varphi'' \left(\frac{g f_{a_k}}{f g_{a_k}} \right) \frac{\partial}{\partial a_i} \frac{g f_{a_k}}{f g_{a_k}} \frac{\partial}{\partial b_j} \frac{g f_{a_k}}{f g_{a_k}} f \, dx, \\ \mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} M(a_k, a_k) &= \int \varphi' \left(\frac{g f_{a_k}}{f g_{a_k}} \right) \frac{\partial^2}{\partial a_i \partial a_j} \frac{g f_{a_k}}{f g_{a_k}} f \, dx, \quad \mathbf{P} \frac{\partial^2}{\partial b_i \partial b_j} M(a_k, a_k) = - \int \varphi'' \left(\frac{g f_{a_k}}{f g_{a_k}} \right) \frac{\partial}{\partial b_i} \frac{g f_{a_k}}{f g_{a_k}} \frac{\partial}{\partial b_j} \frac{g f_{a_k}}{f g_{a_k}} f \, dx, \\ \text{and consequently } \mathbf{P} \frac{\partial^2}{\partial b_i \partial b_j} M(a_k, a_k) &= - \mathbf{P} \frac{\partial^2}{\partial a_i \partial b_j} M(a_k, a_k) = - \mathbf{P} \frac{\partial^2}{\partial b_j \partial a_i} M(a_k, a_k), \text{ which implies,} \\ \frac{\partial^2}{\partial a_i \partial a_j} K(g_{\frac{f_{a_k}}{g_{a_k}}}, f) &= \mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} M(a_k, a_k) - \mathbf{P} \frac{\partial^2}{\partial b_i \partial b_j} M(a_k, a_k), \\ &= \mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial a_i \partial b_j} M(a_k, a_k), = \mathbf{P} \frac{\partial^2}{\partial a_i \partial a_j} M(a_k, a_k) + \mathbf{P} \frac{\partial^2}{\partial b_j \partial a_i} M(a_k, a_k). \end{aligned}$$

By definition of the estimators $\check{\gamma}_n$ and $\check{c}_n(a_k)$, we have $\begin{cases} \mathbb{P}_n \frac{\partial}{\partial b} M(b, a) = 0 \\ \mathbb{P}_n \frac{\partial}{\partial a} M(b(a), a) = 0 \end{cases}$

$$\text{ie } \begin{cases} \mathbb{P}_n \frac{\partial}{\partial b} M(\check{c}_n(a_k), \check{\gamma}_n) = 0 \\ \mathbb{P}_n \frac{\partial}{\partial a} M(\check{c}_n(a_k), \check{\gamma}_n) + \mathbb{P}_n \frac{\partial}{\partial b} M(\check{c}_n(a_k), \check{\gamma}_n) \frac{\partial}{\partial a} \check{c}_n(a_k) = 0, \end{cases} \quad \text{i.e. } \begin{cases} \mathbb{P}_n \frac{\partial}{\partial b} M(\check{c}_n(a_k), \check{\gamma}_n) = 0 \quad (E0) \\ \mathbb{P}_n \frac{\partial}{\partial a} M(\check{c}_n(a_k), \check{\gamma}_n) = 0 \quad (E1) \end{cases}$$

Under $(H'5)$ and $(H'6)$, and using a Taylor development of the $(E0)$ (resp. $(E1)$) equation, we infer there exists $(\bar{c}_n, \bar{\gamma}_n)$ (resp. $(\check{c}_n, \check{\gamma}_n)$) on the interval $[(\check{c}_n(a_k), \check{\gamma}_n), (a_k, a_k)]$ such that

$$\begin{aligned} -\mathbb{P}_n \frac{\partial}{\partial b} M(a_k, a_k) &= [(\mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k))^\top + o_{\mathbf{P}}(1), (\mathbf{P} \frac{\partial^2}{\partial a \partial b} M(a_k, a_k))^\top + o_{\mathbf{P}}(1)] a_n. \\ \text{(resp. } -\mathbb{P}_n \frac{\partial}{\partial a} M(a_k, a_k) &= [(\mathbf{P} \frac{\partial^2}{\partial b \partial a} M(a_k, a_k))^\top + o_{\mathbf{P}}(1), (\mathbf{P} \frac{\partial^2}{\partial a^2} M(a_k, a_k))^\top + o_{\mathbf{P}}(1)] a_n) \end{aligned}$$

with $a_n = ((\check{c}_n(a_k) - a_k)^\top, (\check{\gamma}_n - a_k)^\top)$. Thus we get

$$\begin{aligned} \sqrt{n} a_n &= \sqrt{n} \begin{bmatrix} \mathbf{P} \frac{\partial^2}{\partial b^2} M(a_k, a_k) & \mathbf{P} \frac{\partial^2}{\partial a \partial b} M(a_k, a_k) \\ \mathbf{P} \frac{\partial^2}{\partial b \partial a} M(a_k, a_k) & \mathbf{P} \frac{\partial^2}{\partial a^2} M(a_k, a_k) \end{bmatrix}^{-1} \begin{bmatrix} -\mathbb{P}_n \frac{\partial}{\partial b} M(a_k, a_k) \\ -\mathbb{P}_n \frac{\partial}{\partial a} M(a_k, a_k) \end{bmatrix} + o_{\mathbf{P}}(1) \\ &= \sqrt{n} (\mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) \frac{\partial^2}{\partial a \partial a} K(g_{\frac{f_{a_k}}{g_{a_k}}}, f))^{-1} \\ &\quad \cdot \begin{bmatrix} \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) + \frac{\partial^2}{\partial a \partial a} K(g_{\frac{f_{a_k}}{g_{a_k}}}, f) & \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) \\ \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) & \mathbf{P} \frac{\partial^2}{\partial b \partial b} M(a_k, a_k) \end{bmatrix} \cdot \begin{bmatrix} -\mathbb{P}_n \frac{\partial}{\partial b} M(a_k, a_k) \\ -\mathbb{P}_n \frac{\partial}{\partial a} M(a_k, a_k) \end{bmatrix} + o_{\mathbf{P}}(1) \end{aligned}$$

Moreover, the central limit theorem implies: $\mathbb{P}_n \frac{\partial}{\partial b} M(a_k, a_k) \xrightarrow{\mathcal{L}^{aw}} \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial b} M(a_k, a_k)\|^2)$,

$\mathbb{P}_n \frac{\partial}{\partial a} M(a_k, a_k) \xrightarrow{\mathcal{L}^{aw}} \mathcal{N}_d(0, \mathbf{P} \|\frac{\partial}{\partial a} M(a_k, a_k)\|^2)$, since $\mathbf{P} \frac{\partial}{\partial b} M(a_k, a_k) = \mathbf{P} \frac{\partial}{\partial a} M(a_k, a_k) = 0$, which leads us to the result. Finally, if f is known, we similarly prove theorem 10. \square

Proof of theorems 3 and 10. We immediately get the theorem through proposition 11 and theorem 16. \square

Proof of proposition 13. Let us consider $\psi, \psi_a, \psi^{(k)}, \psi_a^{(k)}$ the characteristic functions of densities $f, f_a, g^{(k-1)}$ and $[g^{(k-1)}]_a$.

We have $|\psi(ta) - \psi^{(k-1)}(ta)| = |\psi_a(t) - \psi_a^{(k-1)}(t)| \leq \int |f_a(a^\top x) - [g^{(k-1)}]_a(a^\top x)| dx$, and then $\sup_a |\psi_a(t) - \psi_a^{(k-1)}(t)| \leq \sup_a \int |f_a(a^\top x) - [g^{(k-1)}]_a(a^\top x)| dx \leq \sup_a K([g^{(k-1)}]_a, f_a)$ since $\psi(ta) = \mathbb{E}(e^{ita^\top x}) = \psi_a(t)$ - where $t \in \mathbb{R}$ and $a \in \mathbb{R}_*^d$ - and since the relative entropy is greater than the L^1 distance. Therefore, since, as explained in section 14 of Huber's article, we have $\lim_k K([g^{(k-1)}]_{a_k}, f_{a_k}) = 0$ we then get $\lim_k g^{(k)} = f$ - which is the Huber's representation of f . Moreover, we have $|\psi(t) - \psi^{(k)}(t)| \leq \int |f(x) - g^{(k)}(x)| dx \leq K(g^{(k)}, f)$. As explained in section 14 of Huber's article and through remark 6 page 13 and through the additive relation of proposition 6, we can say

that $\lim_k K(g^{(k-1)} \frac{f_{a_k}}{[g^{(k-1)}]_{a_k}}, f) = 0$. Consequently, we get $\lim_k g^{(k)} = f$ - which is our representation of f .

Proof of lemmas 1 and 2. Let us apply our algorithm between f and g . There exists a sequence of densities $(g^{(k)})_k$ such that $0 = K(g^{(\infty)}, f) \leq \dots \leq K(g^{(k)}, f) \leq \dots \leq K(g, f)$, (*)

where $g^{(\infty)} = \lim_k g^{(k)}$ which is a density by construction.

Moreover, let $(g_n^{(k)})_k$ be the sequence of densities such that $g_n^{(k)}$ is the kernel estimate of $g^{(k)}$. Since we derive from remark 9 page 22 an integrable upper bound of $g_n^{(k)}$, for all k , which is greater than f - see also the definition of φ in the proof of theorem 4 -, then the dominated convergence theorem implies that, for any k , $\lim_n K(g_n^{(k)}, f_n) = K(g^{(k)}, f)$, i.e., from a certain given rank n_0 , we have

$$0 \leq \dots \leq K(g_n^{(\infty)}, f_n) \leq \dots \leq K(g_n^{(k)}, f_n) \leq \dots \leq K(g_n, f_n), (**)$$

Consequently, through lemma 18 page 29, there exists a k such that

$$0 \leq \dots \leq K(\Psi_{n,k}^{(\infty)}, f_n) \leq \dots \leq K(g_n^{(\infty)}, f_n) \leq \dots \leq K(\Psi_{n,k-1}^{(\infty)}, f_n) \leq \dots \leq K(g_n, f_n), (***)$$

where $\Psi_{n,k}^{(\infty)}$ is a density such that $\Psi_{n,k}^{(\infty)} = \lim_k g_n^{(k)}$.

Finally, through the dominated convergence theorem and taking the limit as n in (***) we get

$$0 = K(g^{(\infty)}, f) = \lim_n K(g_n^{(\infty)}, f_n) \geq \lim_n K(\Psi_{n,k}^{(\infty)}, f_n) \geq 0.$$

The dominated convergence theorem enables us to conclude:

$$0 = \lim_n K(\Psi_{n,k}^{(\infty)}, f_n) = \lim_n \lim_k K(g_n^{(k)}, f_n). \text{ Similarly, we get lemma 2.} \quad \square$$

Proof of lemma 18.

Lemma 18. *With the notation of the proof of lemma 1, we have*

$$0 \leq \dots \leq K(\Psi_{n,k}^{(\infty)}, f_n) \leq \dots \leq K(g_n^{(\infty)}, f_n) \leq \dots \leq K(\Psi_{n,k-1}^{(\infty)}, f_n) \leq \dots \leq K(g_n, f_n), (***)$$

Proof :

First, as explained in section 4.2, we have $K(f^{(k)}, g) - K(f^{(k+1)}, g) = K(f_{a_{k+1}}^{(k)}, g_{a_{k+1}})$. Moreover, through remark 6 page 13, we also derive that $K(f^{(k)}, g) = K(g^{(k)}, f)$. Then, $K(f_{a_{k+1}}^{(k)}, g_{a_{k+1}})$ is the decreasing step of the relative entropies in (*) and leading to $0 = K(g^{(\infty)}, f)$. Similarly, through the construction of (**), we obtain that $K(f_{a_{k+1},n}^{(k)}, g_{a_{k+1},n})$ is the decreasing step of the relative entropies in (**) and leading to $K(g_n^{(\infty)}, f_n)$.

Second, through the conclusion of the section 4.2 and lemma 14.2 of Huber's article, we obtain that $K(f_{a_{k+1},n}^{(k)}, g_{a_{k+1},n})$ converges - in decreasing and in k - towards a positive function of n - that we will call ξ_n .

Third, the convergence of $(g^{(k)})_k$ - see proposition 13 - implies that, for any given n , the sequence $(K(g_n^{(k)}, f_n))_k$ is not finite. Then, through relationship (**), there exists a k such that $0 < K(g_n^{(k-1)}, f_n) - K(g_n^{(\infty)}, f_n) < \xi_n$.

Consequently, since $Q \mapsto K(Q, P)$ is l.s.c. - see property 3 page 18 - the relationship (**) implies (***). \square

Proof of theorems 4 and 11. We recall that $g_n^{(k)}$ is the kernel estimator of $\check{g}^{(k)}$. Since the relative entropy is greater than the L^1 -distance, we then have

$$\lim_n \lim_k K(g_n^{(k)}, f_n) \geq \lim_n \lim_k \int |g_n^{(k)}(x) - f_n(x)| dx$$

Moreover, the Fatou's lemma implies that

$$\lim_k \int |g_n^{(k)}(x) - f_n(x)| dx \geq \int \lim_k [|g_n^{(k)}(x) - f_n(x)|] dx = \int |[\lim_k g_n^{(k)}(x)] - f_n(x)| dx$$

$$\text{and } \lim_n \int |[\lim_k g_n^{(k)}(x)] - f_n(x)| dx \geq \int \lim_n [|[\lim_k g_n^{(k)}(x)] - f_n(x)|] dx$$

$$= \int |[\lim_n \lim_k g_n^{(k)}(x)] - \lim_n f_n(x)| dx.$$

We then obtain that $0 = \lim_n \lim_k K(g_n^{(k)}, f_n) \geq \int |[\lim_n \lim_k g_n^{(k)}(x)] - \lim_n f_n(x)| dx \geq 0$, i.e. that $\int |[\lim_n \lim_k g_n^{(k)}(x)] - \lim_n f_n(x)| dx = 0$.

Moreover, for any given k and any given n , the function $g_n^{(k)}$ is a convex combination of multivariate

Gaussian distributions. As derived at remark 4 of page 5, for all k , the determinant of the covariance of the random vector - with density $g^{(k)}$ - is greater than or equal to the product of a positive constant times the determinant of the covariance of the random vector with density f . Consequently, the form of the kernel estimate implies that there exists an integrable function φ such that, for any given k and any given n , we have $|g_n^{(k)}| \leq \varphi$.

Finally, the dominated convergence theorem enables us to say that $\lim_n \lim_k g_n^{(k)} = \lim_n f_n = f$, since f_n converges towards f and since $\int |[\lim_n \lim_k g_n^{(k)}(x)] - \lim_n f_n(x)| dx = 0$.

Similarly, we prove theorem 11. \square

Proof of theorem 17. Through a Taylor development of $\mathbb{P}_n M(\check{c}_n(a_k), \check{\gamma}_n)$ of rank 2, we get at point (a_k, a_k) :

$$\begin{aligned} \mathbb{P}_n M(\check{c}_n(a_k), \check{\gamma}_n) &= \mathbb{P}_n M(a_k, a_k) + \mathbb{P}_n \frac{\partial}{\partial a} M(a_k, a_k)(\check{\gamma}_n - a_k)^\top + \mathbb{P}_n \frac{\partial}{\partial b} M(a_k, a_k)(\check{c}_n(a_k) - a_k)^\top \\ &\quad + \frac{1}{2} \{ (\check{\gamma}_n - a_k)^\top \mathbb{P}_n \frac{\partial^2}{\partial a \partial a} M(a_k, a_k)(\check{\gamma}_n - a_k) + (\check{c}_n(a_k) - a_k)^\top \mathbb{P}_n \frac{\partial^2}{\partial b \partial b} M(a_k, a_k)(\check{c}_n(a_k) - a_k) \\ &\quad + (\check{\gamma}_n - a_k)^\top \mathbb{P}_n \frac{\partial^2}{\partial a \partial b} M(a_k, a_k)(\check{c}_n(a_k) - a_k) + (\check{c}_n(a_k) - a_k)^\top \mathbb{P}_n \frac{\partial^2}{\partial b \partial a} M(a_k, a_k)(\check{\gamma}_n - a_k) \} \end{aligned}$$

The lemma below enables us to conclude.

Lemma 19. *Let H be an integrable function and let $C = \int H d\mathbf{P}$ and $C_n = \int H d\mathbb{P}_n$, then, $C_n - C = O_{\mathbf{P}}(\frac{1}{\sqrt{n}})$.*

Thus we get $\mathbb{P}_n M(\check{c}_n(a_k), \check{\gamma}_n) = \mathbb{P}_n M(a_k, a_k) + O_{\mathbf{P}}(\frac{1}{\sqrt{n}})$,
i.e. $\sqrt{n}(\mathbb{P}_n M(\check{c}_n(a_k), \check{\gamma}_n) - \mathbf{P} M(a_k, a_k)) = \sqrt{n}(\mathbb{P}_n M(a_k, a_k) - \mathbf{P} M(a_k, a_k)) + o_{\mathbf{P}}(1)$.
Hence $\sqrt{n}(\mathbb{P}_n M(\check{c}_n(a_k), \check{\gamma}_n) - \mathbf{P} M(a_k, a_k))$ abides by the same limit distribution as $\sqrt{n}(\mathbb{P}_n M(a_k, a_k) - \mathbf{P} M(a_k, a_k))$, which is $\mathcal{N}(0, \text{Var}_{\mathbf{P}}(M(a_k, a_k)))$. \square

Proof of theorems 5 and 12. Through proposition 11 and theorem 17, we derive theorem 5. Similarly, we get theorem 12. \square

Proof of proposition 5. Let us first show by induction the following assertion

$$\mathcal{P}(k) = \{g^{(k)} \text{ allows a deconvolution } g^{(k)} = \bar{g}^{(k)} * \phi\}$$

Initialisation : For $k = 0$, we get the result since $g = g^{(0)}$ is elliptic.

Going from k to $k + 1$: Let us assume $\mathcal{P}(k)$ is true, we then show that $\mathcal{P}(k + 1)$.

Since the family of a_i , $i \leq k + 1$ is free - see lemma 15 - then, we define B as the basis of \mathbb{R}^d such that its $k + 1$ first vectors are the a_i , $i \leq k + 1$ - see the incomplete basis theorem for its existence.

Thus, in B and using the same procedure to prove lemma 7 page 23, we have

$\bar{g}^{(k)}(x) = \bar{g}^{(k)}(./x_{k+1}) \bar{g}_{k+1}^{(k)}(x_{k+1})$. Consequently, the very definition of the convolution product, the Fubini's theorem and the hypothesis on the Elliptical family used imply that

$g^{(k)}(x) = g^{(k)}(./x_{k+1}) g_{k+1}^{(k)}(x_{k+1})$ with $g^{(k)}(./x_{k+1}) = \bar{g}^{(k)}(./x_{k+1}) * E_{d-1}(0, \sigma^2 I_{d-1}, \xi_{d-1})$ and with $g_{k+1}^{(k)}(x_{k+1}) = \bar{g}_{k+1}^{(k)}(x_{k+1}) * E_1(0, \sigma^2, \xi_1)$. Finally, replacing $g_{k+1}^{(k)}$ with $f_{k+1} = \bar{f}_{k+1} * E_1(0, \sigma^2, \xi_1)$, we conclude this induction with $g^{(k+1)} = g^{(k)}(./x_{k+1}) f_{k+1}(x_{k+1})$.

Now, let us consider ψ (rep. $\bar{\psi}$, $\psi^{(k)}$, $\bar{\psi}^{(k)}$) the characteristic function of f (resp. \bar{f} , $g^{(k)}$, $\bar{g}^{(k)}$). We then have $\psi(s) = \bar{\psi}(s) \Psi(\frac{1}{2} \sigma^2 |s|^2)$ and $\psi^{(k)}(s) = \bar{\psi}^{(k)}(s) \Psi(\frac{1}{2} \sigma^2 |s|^2)$. Hence, ψ and $\psi^{(k)}$ are less or equal to $\Psi(\frac{1}{2} \sigma^2 |s|^2)$ which is integrable by hypothesis, i.e. ψ and $\psi^{(k)}$ are absolutely integrable. We then obtain $g^{(k)}(x) = (2\pi)^{-d} \int \psi^{(k)}(s) e^{-is^\top x} ds$ and $f(x) = (2\pi)^{-d} \int \psi(s) e^{-is^\top x} ds$.

Moreover, since the sequence $(\psi^{(k)})$ uniformly converges and since ψ and $\psi^{(k)}$ are less or equal to $\Psi(\frac{1}{2} \sigma^2 |s|^2)$, then the dominated convergence theorem implies that

$\lim_k |f(x) - g^{(k)}(x)| \leq (2\pi)^{-d} \int \lim_k |\psi(s) - \psi^{(k)}(s)| ds = 0$ a.s. i.e. $\lim_k \sup_x |f(x) - g^{(k)}(x)| = 0$ a.s.

Finally, since, by hypothesis, $(2\pi)^{-d} \int |\psi(s) - \psi^{(k)}(s)| ds \leq 2(2\pi)^{-d} \int \Psi(\frac{1}{2} \sigma^2 |s|^2) ds < \infty$, then the above limit and the dominated convergence theorem imply that $\lim_k \int |f(x) - g^{(k)}(x)| dx = 0$. \square

Proof of theorem 7. We immediately get the proof through theorem 4. \square

References

- [AZE97] AZE D., *Eléments d'analyse convexe et variationnelle*, Ellipse, 1997.
- [BOLE] Bosq D., Lecoutre J.-P. *Livre - Theorie De L'Estimation Fonctionnelle*, Economica, 1999.
- [BROKEZ] Broniatowski M., Keziou A. *Parametric estimation and tests through divergences and the duality technique. J. Multivariate Anal. 100 (2009), no. 1, 16–36.*
- [CAMBANIS81] Cambanis, Stamatis; Huang, Steel; Simons, Gordon. *On the theory of elliptically contoured distributions. J. Multivariate Anal. 11 (1981), no. 3, 368–385.*
- [DEVGY85] Devroye, Luc; Györfi, László. Distribution free exponential bound for the L_1 error of partitioning-estimates of a regression function. *Probability and statistical decision theory*, Vol. A (Bad Tatzmannsdorf, 1983), 67–76, Reidel, Dordrecht, 1985
- [DIAFREE84] Diaconis, Persi; Freedman, David. *Asymptotics of graphical projection pursuit. Ann. Statist. 12 (1984), no. 3, 793–815.*
- [DI80] Jean Dieudonné, *Calcul infinitésimal. 1980, Hermann.*
- [Frie8487] Friedman, Jerome H.; Stuetzle, Werner; Schroeder, Anne. *Projection pursuit density estimation. J. Amer. Statist. Assoc. 79 (1984), no. 387, 599–608.*
- [FR87] Friedman, Jerome H. *Exploratory projection pursuit. J. Amer. Statist. Assoc. 82 (1987), no. 397, 249–266.*
- [HUB85] Huber Peter J., *Projection pursuit, Ann. Statist., 13(2):435–525, 1985, With discussion.*
- [LANDS03] Landsman, Zinoviy M.; Valdez, Emiliano A. *Tail conditional expectations for elliptical distributions. N. Am. Actuar. J. 7 (2003), no. 4, 55–71.*
- [LIVAJ] Liese Friedrich and Vajda Igor, *Convex statistical distances, volume 95 of Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]. BSB B. G. Teubner Verlagsgesellschaft, 1987, with German, French and Russian summaries.*
- [HLL94] Jenq-Neng Hwang, Shyh-Rong Lay, and A. Lippman. *Nonparametric multivariate density estimation: a comparative study. Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on], 42(10):2795–2810, 1994.*
- [ROCKA] Rockafellar, R. Tyrrell., *Convex analysis. Princeton Mathematical Series, No. 28 Princeton University Press, Princeton, N.J. 1970 xviii+451 pp.*
- [SCOTT92] Scott, David W., *Multivariate density estimation. Theory, practice, and visualization. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1992. xiv+317 pp. ISBN: 0-471-54770-0.*
- [VDW] van der Vaart A. W., *Asymptotic statistics, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1998.*
- [ZMU04] Zhu, Mu. *On the forward and backward algorithms of projection pursuit. Ann. Statist. 32 (2004), no. 1, 233–244.*