FiME
LABORATOIRE COMMUN
DAUPHINE CREST EDF

LABORATOiRE DE FiNANCE
DES MARCHÉS DE L'ÉNERGiE

L2 Density Estimation Under Constraints

Christian Musso Nadia Oudjane

Rapport de Recherche
RR-FiME-09-03
Septembre 2009

DAUPHINE UNIVERSITÉ PARIS

CREST Centre de Recherche en Économie et Statistique

EDF

FiME Preprint
URL: http://www.fime-lab.org

L^2-DENSITY ESTIMATION UNDER CONSTRAINTS *

CHRISTIAN MUSSO.[1] AND NADIA OUDJANE[2]

Abstract. This paper presents a general methodology to estimate a probability density under linear constraints (on the support, bounded moments or quantiles,...). The proposed approximation is the projection of the free density estimation on the set of the probability densities satisfying the constraints. In some cases, the solution of this projection problem can be expressed in a simple parametric form as a function of the free density estimate.

Résumé. Ce papier présente une méthode d'estimation de densité sous des contraintes linéaires (contraintes portant sur le support, sur des moments bornés ou sur des quantiles,...). L'estimateur proposé est la projection de l'estimateur à noyaux non contraint sur l'espace des densités de probabilité satisfaisant les contraintes. Dans certains cas, cette projection s'exprime sous une forme paramétrique simple en fonction de l'estimateur non contraint.

1991 Mathematics Subject Classification. 62G07.

November 25, 2009.

1. Introduction

Kernel density estimation is one of the most popular approach to approximate the underlying density f of independent and identically distributed (i.i.d) observations $(X_1, \cdots, X_N)$. It consists in approximating the underlying density by a mixture of kernels equally weighted and centered on each point of the data set. In some statistical applications, one can have additionnal information about the underlying density f, for instance the support or some moments can be known. Different approaches have been proposed to modify the kernel estimate to take into account an a priori knowledge on the density. For example, in [8] a weighted bootstrap method is described in order to construct a density estimator for salary data which are necessarily non-negative. In [4], the author proposes empirical-likelihood density estimation which consists in weigthing the mixture of kernels with probability weights such that the weighted discrete distribution on the data points is the nearest to the original empirical distribution subject to the constraints to satisfy. In [10], the authors propose to modify locally the kernel density estimates around the support bound, then they apply their approach to the distribution of waves height which can take only non-negative values. Others kinds of constraints in density estimation are considered in the literature, like in [9], where the constraint is the unimodality of the density. In the present paper, we propose an alternative approach to modify properly the initial kernel density estimate. Our approach consists

Keywords and phrases: Kernel, Density estimation, Non-parametric estimation, Optimization under linear constraints.

* *Nadia Oudjane was supported by FiME Laboratoire de Finance des Marchs de l'Energie, (Dauphine University, the CREST and EDF R&D)*

[1] ONERA/DTIM/EVS, 29 av. de la Division Leclerc, 92322 cedex Châtillon, France.

[2] EDF R&D & University Paris 13, EDF, 1 av. du Général de Gaulle, 92140 cedex, Clamart, France.

in projecting the initial kernel density estimate on the set of L^2 probability densities satisfying the constraints. The method relies on a characterization theorem stated in Section 3 specifying a simple and explicit parametric form for the projection of a function on a closed convex subset of L^2. This general result of optimization is used to build our new density estimate. Then, we propose how to compute approximately the corresponding parameters and how to simulate our new density estimate in some simple examples.

The paper begins by considering kernel density estimation using negative kernels (which can take negative values). Such kernels can be useful because they allow bias reduction of the density approximation, but at the same time they can produce density estimates which take negative values, which is of course undesirable. To avoid this drawback, we propose to project the negative kernel density estimate on the subset of L^2-probability densities. This projection method has been introduced in [11]. It leads to a new density estimate and allows both bias reduction and positiveness of the approximation. A similar approach has been proposed in [7] but the authors have not seen the projection property of the new estimate, which allows useful generalizations. Then, to be able to apply this projection approach to a more general context of density estimation with several constraints, we need to generalize the result of Section 2 characterizing the projection of a function on the subset of L^2-probability densities. In section 3, we extend this approximation method in considering the projection on a more general closed and convex subset of L^2. In the fourth section, we consider the case of constrained density estimation. In two specific cases (support constraint and moment constraint), we provide some conditions under which the projection has a simple parametric form. Then, a method is proposed to compute and simulate the resulting constrained density estimate.

The last section is devoted to simulations. We observe the good behaviour of the projection density on simple examples with standard functions and on examples from density estimation. In particular, we analyse the performance of the projected estimate in the case of negative kernels and in the case of a support constraints.

Notice that the same approach can be used in other applications. For instance, when we develop a density with Edgeworth's serie, the resulting expansion can produce negative values. One could use this projection approach to obtain non-negative approximation.

2. Projection of a negative kernel density estimate on the set of probability densities

2.1. Kernel density estimation: a brief review

Kernel density estimators were introduced by Rosenblatt [13]. Initially the kernel was itself a density. In this paper, we focus on negative kernels which were introduced by Parzen [12] and Bartlett [2]. We briefly recall some classical results of density estimation theory. For simplicity, in this section, we assume that samples are one-dimensional, but the results presented here are generalized in a d-dimensional space (see [16] and [6]).
Let $s \geq 2$ be a positive integer, a function K defined on $\mathbb{R}$ is called kernel of order s if

$$\int_{\mathbb{R}} K(x)\, dx = 1 \ , \quad k_s = \int_{\mathbb{R}} |x|^s\, K(x)\, dx < \infty$$

$$\text{and} \quad \int_{\mathbb{R}} x^i\, K(x)\, dx = 0 \ , \quad \text{for all } i = 1, \cdots, s-1 \ .$$

A kernel of order $s \geq 3$ necessarily takes negative values. In view of the symmetry condition, we will only consider kernels of even order.
Let K be a kernel of order $s \geq 2$, and h be a positive real the function K_h defined by

$$K_h(x) = \frac{1}{h}\, K(\frac{x}{h}) \ , \quad \text{for all} \quad x \in \mathbb{R} \ , \tag{1}$$

will be called the scaled kernel associated to K with the bandwidth h.

2.1.1. *Error estimation*

The following proposition gives an approximation of kernel density estimators error.

Proposition 2.1. *Let K be a kernel of order $s \geq 2$ and p be a probability density defined on $\mathbb{R}$ having its derivatives well defined and continuous up to the order $s+1$. Let $(X_1, \cdots, X_N)$ be an i.i.d. sample from p. Let us consider $\hat{p}$ the function defined, for all $x \in \mathbb{R}$ by*

$$\hat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K(\frac{x - X_i}{h}) = \frac{1}{N} \sum_{i=1}^{N} K_h(x - X_i) \ . \tag{2}$$

Then the following approximation for the asymptotic mean integrated square error (MISE) holds,

$$\mathbb{E}\|\hat{p} - p\|_2^2 = \frac{k_s^2 h^{2s}}{(s!)^2} \int_{\mathbb{R}} (p^{(s)}(x))^2 \, dx + \frac{\int_{\mathbb{R}} K^2(x) \, dx}{Nh} + O(\frac{1}{N}) + o(h^{2s}) \ , \tag{3}$$

where the expectation is performed w.r.t. $(X_1, \cdots, X_N)$.

A proof of this expansion can be found in the monograph of Silverman [16]. For the treatment of the error in the L^1 sense see the monographs of Devroye and Györfi [5] and Devroye [6].
Notice that the bias term (the first term on the right-hand side of (3)) can be reduced in increasing order s. That is precisely the reason why we are interested in negative kernels. An alternative bias reduction method consists of transforming the data as proposed in [14].

2.1.2. *Kernel and bandwidth selection*

Classically, the kernel, K, and bandwidth, h, are chosen so as to minimize the mean integrated error $\mathbb{E}\|\hat{p}-p\|_1$, or the mean integrated square error (MISE), $\mathbb{E}\|\hat{p}-p\|_2^2$. For positive kernels of order $s = 2$, the density estimation theory, see [6, 16], provides the optimal choice for the kernel, in the MISE sense. The general expression of this optimal kernel is the Epanechnikov kernel, which is given by the following expression in the one dimensional case,

$$K_{\text{Epan}}(x) = \begin{cases} \frac{3}{4}(1 - |x|^2) & \text{if } |x| < 1 \\ 0 & \text{elsewhere,} \end{cases} \tag{4}$$

The optimal choice of the bandwith depends on the unknown underlying density p. Therefore, one way to choose the optimal bandwidth associated to the Epanechnikov kernel, h_{Epan}, is to assume, for instance, that the underlying density p is a centered standard Gaussian density. In this case, the optimal bandwidth, in the MISE sense, is easily computed and is given by [16]

$$h_{\text{Epan}} = [40\sqrt{\pi}]^{\frac{1}{5}} N^{-\frac{1}{5}} \ . \tag{5}$$

If the underlying density p is non standard (its standard deviation is different from one), h_{Epan} is multiplied by the standard deviation of p.

2.2. Characterization and properties of the L^2-projection on the convex set of L^2-probability densities

In order to reduce the bias of the kernel approximation $\hat{p}$ (first term on the right-hand side of (3)), one can use negative kernels with order $s \geq 3$. But, in this case, the appoximation $\hat{p}$ can take negative values which is of course undesirable. We propose here an optimal (in some sense) solution to this problem. Hereafter, results will be stated in the d-dimensional case.
More generally, if we are interested in approximating, by a probability density, a function (in our setting $f = \hat{p}$),

which can possibly take negative values and such that $\int f = 1$, one possibility, if $f \in \mathrm{L}^1$, is to use the following approximation,

$$g_1^*(x) = \frac{f^+(x)}{\int_{\mathbb{R}^d} f^+(x)dx} \ , \tag{6}$$

where $(a)^+ = \max(a, 0)$, for any $a \in \mathbb{R}$. Notice that the function g_1^* has the following property,

$$\|g_1^* - g\|_1 \leq \|f - g\|_1 \ , \tag{7}$$

for any probability density g. See Theorem 3 p. 269 in [5], for a proof. This property doesn't mean that g_1^* minimizes the distance between f and the set of probability densities. It means that replacing f by g_1^* can only improve the estimation of a probability density p on the base of a first estimate f. Hence, in the case of non-positive kernel density estimation where $f = \hat{p}$, this approach leads to a new estimate g_1^* which is always better (in the L^1 sense) than the original estimate $\hat{p}$:

$$\|g_1^* - p\|_1 \leq \|\hat{p} - p\|_1 \ .$$

Unfortunately, g_1^* is in general difficult to compute, because of the normalizing term $\int f^+$ which is difficult to evaluate when the dimension d is high. An alternative, if f is in L^2, is to build the projection of f on the subset of L^2-probability densities. The next proposition gives a characterization of this optimal approximation.

Proposition 2.2. *Let f be a bounded function defined on $\mathbb{R}^d$ with values in $\mathbb{R}$, such that*

$$\int_{\mathbb{R}^d} f(x)\, dx = 1 \ , \quad \textit{and} \quad \int_{\mathbb{R}^d} f^2(x)\, dx < \infty \ ,$$

then the projection g^* of f on the convex subset of L^2-probability densities is determined by

$$g^*(x) = (f(x) - \alpha^*)^+ \ , \quad \textit{for all} \quad x \in \mathbb{R}^d \ , \tag{8}$$

where α^* is the non-negative real such that

$$\int_{\mathbb{R}^d} (f(x) - \alpha^*)^+\, dx = 1 \ . \tag{9}$$

Consequently, g^* defined by (8) and (9) satisfies a property similar to (7) in the L^2 sense, i.e.

$$\|g^* - g\|_2 \leq \|f - g\|_2 \ , \tag{10}$$

for any L^2-probability density g.

Proof. First notice that $\int f^+ \geq \int f = 1$ implies the existence of α^* verifying (9). Indeed, the function $H(\alpha) = \int (f(x) - \alpha)^+ dx$ decreases continuously from $H(0) = \int f^+$ to $H(+\infty) = 0$ ($H(\alpha) = 0$ as soon as $\alpha \geq \sup f$). The proof of Proposition 2.2 is based on the characterization of the projection X^* of a point Y on a convex set S in an Hilbert space. X^* is the projection of Y on S if and only if

$$\langle Y - X^*, X - X^* \rangle \leq 0 \quad \text{for all} \quad X \in S \ .$$

In our case, S is the subset of L^2-probability densities (we easily check that S is convex). Then, it is sufficient to prove that the quantity,

$$\Delta = \int_{\mathbb{R}^d} [f(x) - (f(x) - \alpha^*)^+]\, [g(x) - (f(x) - \alpha^*)^+]\, dx$$

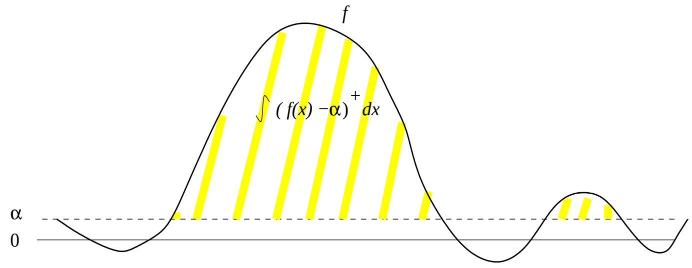


FIGURE 1. Illustration of the characterization of α^*.

is non-positive for any probability density $g \in L^2$, where α^* is defined in (9). Let us introduce the subsets of $\mathbb{R}^d$,

$$A^* = \{x \,|\, f(x) \geq \alpha^*\} \quad \text{and} \quad \bar{A}^* = \{x \,|\, f(x) < \alpha^*\} \ ,$$

we have

$$\begin{aligned}
\Delta &= \int_{A^*} [f(x) - f(x) + \alpha^*]\,[g(x) - f(x) + \alpha^*]\,dx + \int_{\bar{A}^*} f(x)g(x)\,dx \\
&= \alpha^* \left[\int_{A^*} g(x)\,dx - \int (f(x) - \alpha^*)^+\,dx\right] + \int_{\bar{A}^*} f(x)g(x)\,dx \\
&= \alpha^* \left[\int_{A^*} g(x)\,dx - 1\right] + \int_{\bar{A}^*} f(x)g(x)\,dx \\
&= -\alpha^* \int_{\bar{A}^*} g(x)\,dx + \int_{\bar{A}^*} f(x)g(x)\,dx \\
&= \int_{\bar{A}^*} g(x)(f(x) - \alpha^*)\,dx \ .
\end{aligned}$$

This quantity is non-positive since g is a non-negative function, which ends the first part of the proof. An illustration of the charactrization of α^* is shown on figure 1.
Property (10), is a direct consequence of the projection property. □

The following result shows that the L^2-projection characterized by (8) and (9) approximates f in the L^1 sense as well as approximation (6). Obviously, the L^2-projection outperforms approximation (6) in the L^2 sense.

Proposition 2.3. *Let f be a function in $L^2(\mathbb{R}^d, \mathbb{R})$, such that $\int f(x) = 1$. Then the L^2-projection of f on the subset of L^2-probability densities denoted by g^* (given by (8) and (9)) is as close to the function f as g_1^* given by (6), in the L^1 sense i.e.*

$$\|f - g_1^*\|_1 = \|f - g^*\|_1 \ .$$

Proof. Let us denote by,

$$E_1 = \|f - g_1^*\|_1 \ , \quad \text{and} \quad E_2 = \|f - g^*\|_1 \ .$$

Let us now evaluate the difference between the two errors E_1 and E_2. Scheffe's theorem yields

$$\begin{aligned}
E_1 - E_2 &= \|f - \frac{f^+}{\int f^+}\|_1 - \|f - (f - \alpha^*)^+\|_1 \\
&= 2\int (f - \frac{f^+}{\int f^+})^+ - (f - (f - \alpha^*)^+)^+
\end{aligned}$$

Then notice that

$$\begin{aligned}\int (f - \frac{f^+}{\int f^+})^+ &= \int_{f\geq 0} (f - \frac{f^+}{\int f^+})^+ + \int_{f<0} (f - \frac{f^+}{\int f^+})^+ \\ &= \int (f^+ - \frac{f^+}{\int f^+})^+ \ .\end{aligned}$$

Using the same argument for $\int (f - (f-\alpha^*)^+)^+$ yields

$$E_1 - E_2 = 2 \int (f^+ - \frac{f^+}{\int f^+})^+ - (f^+ - (f-\alpha^*)^+)^+ \ .$$

Since $\int f = 1$, then $\int f^+ \geq 1$ which yields

$$f^+(x) - \frac{f^+(x)}{\int f^+} \geq 0 \ , \quad \text{for all } x \in \mathbb{R}^d \ .$$

Since $\alpha^* \geq 0$ then $f(x) \geq f(x) - \alpha^*$ which yields

$$f^+(x) - (f(x) - \alpha^*)^+ \geq 0 \ , \quad \text{for all } x \in \mathbb{R}^d \ .$$

Hence,

$$\begin{aligned}E_1 - E_2 &= 2 \int (f^+ - \frac{f^+}{\int f^+}) - 2 \int (f^+ - (f-\alpha^*)^+) \ , \\ &= 2 \, [\int (f-\alpha^*)^+ - \frac{f^+}{\int f^+}] = 0 \ .\end{aligned}$$

□

2.3. **Computing and simulating g^***

Let us come back to the initial problem of estimating a density p on $\mathbb{R}^d$ starting from an i.i.d sample $(X_1, \cdots, X_N)$ according to p. Let $f = \hat{p}$ be a kernel density estimate like in (2) with, for all $x \in \mathbb{R}^d$,

$$K_h(x) = \frac{1}{h^d} K(\frac{x}{h}) \ ,$$

and with K being a negative kernel (a kernel which can take negative values). Assume that we know how to simulate the probability density $\frac{K^+}{\int K^+}$.

In this section, we are interested in computing and simulating the density estimate g^* obtained as the L²-projection of the possibly negative function f such that for all $x \in \mathbb{R}^d$,

$$g^*(x) = \left(\frac{1}{N} \sum_{i=1}^{N} K_h(x - X_i) - \alpha^* \right)^+ \ , \tag{11}$$

where α^* is the non-negative real such that

$$\int_{\mathbb{R}^d} \left(\frac{1}{N} \sum_{i=1}^{N} K_h(x - X_i) - \alpha^* \right)^+ dx = 1 \ . \tag{12}$$

In the first subsection, we propose a procedure to approximate the non-negative real α^* which determines the projection g^*. In the second subsection, we describe a procedure to generate random variables from g^*.

2.3.1. *Estimation of α^**

When the dimension d is high, estimating α^* by numerical integration becomes computationally expensive. So, we propose thereafter a fast method based on simulations. Let g denote the probability density defined as follows,

$$g(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{K_h^+(x - X_i)}{\int K^+} \ , \quad \text{for all} \quad x \in \mathbb{R}^d \ . \tag{13}$$

Given $(X_1, \cdots, X_N)$, let us consider the random variable

$$D = \frac{1}{N} \big[\sum_{i=1}^{N} K_h(Y - X_i) - U \sum_{i=1}^{N} K_h^+(Y - X_i) \big] \ , \tag{14}$$

where Y is a random variable following g and U is a uniform random variable on $[0, 1]$, each variable being independent of the others.

Proposition 2.4. *The non-negative real α^* characterizing the L^2-projection g^* given by (11) and (12) is the $(1/\int K^+)$-quantile of D i.e.*

$$\mathbb{P}(D \geq \alpha^*) = \frac{1}{\int K^+} \ . \tag{15}$$

Hence α^* can be estimated empirically in simulating an i.i.d. sample $(D_1, \cdots, D_M)$ according to the law of $D = D(Y, U)$ defined in (14) and in computing an approximation of the $(1/\int K^+)$-quantile of D using the corresponding ordered statistic.

Proof. Let α be a non-negative real, the following events are equivalent,

$$\begin{aligned}
\{D \geq \alpha\} \quad &\equiv \quad \left\{ \frac{\frac{1}{N} \sum_{i=1}^{N} K_h(Y - X_i) - \alpha}{\frac{1}{N} \sum_{i=1}^{N} K_h^+(Y - X_i)} \geq U \right\} \\
&\equiv \quad \left\{ \frac{\left(\frac{1}{N} \sum_{i=1}^{N} K_h(Y - X_i) - \alpha \right)^+}{\frac{1}{N} \sum_{i=1}^{N} K_h^+(Y - X_i)} \geq U \right\} \\
&\equiv \quad \left\{ \frac{H_\alpha(Y)}{g(Y)} \geq U \right\} \ ,
\end{aligned}$$

where

$$H_\alpha(Y) = \frac{1}{\int K^+} \left(\frac{1}{N} \sum_{i=1}^{N} K_h(Y - X_i) - \alpha \right)^+ .$$

This yields

$$\begin{aligned} \mathbb{P}(D \geq \alpha) &= \mathbb{P}\left[\frac{H_\alpha(Y)}{g(Y)} \geq U \right] , \\ &= \int_{\mathbb{R}^d} \min\left(1, \frac{H_\alpha(y)}{g(y)}\right) g(y)\ dy . \end{aligned}$$

Since $\alpha \geq 0$, then $H_\alpha(y) \leq g(y)$ for any $y \in \mathbb{R}^d$. Therefore, for any $\alpha \geq 0$,

$$\mathbb{P}(D \geq \alpha) = \int_{\mathbb{R}^d} H_\alpha(y)\ dy .$$

Notice that for $\alpha = \alpha^*$, we get

$$\int_{\mathbb{R}^d} H_\alpha(y) dy = \frac{1}{\int K^+} ,$$

which yields the desired result. □

2.3.2. *Simulation from g^**

The following proposition, proposes a method to simulate variables according to g^*.

Proposition 2.5. *Assume that we know how to simulate g defined by (13). Let h be the unnormalized function such that for all $x \in \mathbb{R}^d$,*

$$h(x) = \frac{1}{N} \sum_{i=1}^{N} K_h^+(x - X_i) = g(x) \int K^+ . \tag{16}$$

The following algorithm generates a random variable X according to g^* ,

Algorithm 2.6. *(L^2-projection generator)*

(1) *Simulate $Y \sim g$ and $U \sim [0, 1]$;*
(2) *If $U \leq \dfrac{g^*(Y)}{h(Y)}$ then set $X = Y$;*
(3) *Else go step 1.*

The acceptance probability of this algorithm is

$$P_a = \frac{1}{\int K^+} . \tag{17}$$

The proof is obvious using the well-known acceptance-rejection algorithm after observing that

$$g^*(x) \leq g(x) \int K^+ , \quad \text{for all} \quad x \in \mathbb{R}^d .$$

It is worth to note that the acceptance probability (17) is equal to one when K is a non-negative kernel. Indeed in this case $\alpha^* = 0$, $g^* = \hat{p}$ and the algorithm consists in simulating directly from $\hat{p}$. The acceptance probability P_a, decreases when the integral of $K^- = K - K^+$ (the negative part of K) increases.

Notice that the variables Y required to simulate g^* can be extracted from the variables Y used for the estimation of α^* by Proposition 2.4.

3. Projection of a function in the general case of n linear constraints

3.1. General formulation of the problem

In this section, we are interested in generalizing the result of Proposition 2.2, characterizing the general form of the projection on the subset of L^2-probability densities, to the case of a projection on a subset S of L^2 defined by n linear constraints. Let us consider an L^2 function f defined on $\mathbb{R}^d$ with values in $\mathbb{R}$. Let $(f_1, \cdots, f_n)$ be a collection of functions defined on $\mathbb{R}^d$ with values in $\mathbb{R}$ and $(d_1, \cdots, d_n)$ some reals. Let S be the convex set of L^2 defined as follows

$$S = \left\{ g \in L^2 \,|\, g \geq 0 \quad \text{and} \quad \int g f_k = d_k \ , \quad \text{for } k = 1 \cdots n \right\} . \tag{18}$$

Thereafter we will assume that the functions defining the linear constraints, $(f_1, \cdots, f_n)$, are such that S is closed. In this paper, we are specifically interested in the case where one of the constraints is such that $f_1 \equiv 1$ and $d_1 = 1$, i.e. where S is a subset of $L^2 \cap P(\mathbb{R}^d)$ with $P(\mathbb{R}^d)$ denoting the space of probability densities defined on $\mathbb{R}^d$. The aim of this section is to provide in most cases a simple expression for the projection g^* of f on the closed convex set S:

$$g^* = \underset{g \in S}{\arg\min} \int |f - g|^2 \ . \tag{19}$$

In the first subsection, we consider the projection as the solution of an optimization problem with linear constraints and a positivity constraint. The method of Lagrange multipliers provides the solution which is of the form $g^* = (f - \sum_{k=1}^n \alpha_k^* f_k)^+$, for some given reals α_k^*, in the special case where the functions defining the linear constraints f_k belong to L^2. Unfortunately, this approach is not valid if we are interested in probability densities defined on $\mathbb{R}^d$ since the *probability density constraint* will involve the constraint function $f_1 \equiv 1$ which is not in L^2, except in the special case where we consider functions $f, f_1, \cdots f_n$ defined on a bounded support $C \subset \mathbb{R}^d$. To overcome this difficulty, we consider, in the second subsection, the same optimization problem with constraint functions f_k that do not necessarily belong to L^2. We show that the projection on S has the same form as before as soon as there exist coefficients α_k^* such that $\int (f - \sum_{k=1}^n \alpha_k^* f_k)^+ f_k = d_k$ for $k = 1, \cdots, n$. The last subsection considers the special case of two contraints with contraints functions $f_1 \geq 0\,,\ f_2 \geq 0$ not necessarily in L^2. We give the conditions on (d_1, d_2) under which there exist non-negative coefficients α_k^* such that $\int (f - \alpha_1^* f_1 - \alpha_2^* f_2)^+ f_k = d_k$ for $k = 1, 2$. Therefore, under these these conditions, the solution can be expressed as $g^* = (f - \alpha_1^* f_1 - \alpha_2^* f_2)^+$. The main interest of these results is that we determine sufficient conditions under which the functional optimization problem (19) reduces to determine real valued coefficients α_k^*.

3.2. The case where the functions defining the constraints are in L^2

The projection problem (19) can be viewed as an optimization problem with several linear constraints and a positivity constraint. The following proposition gives the expression of the projection g^* of f on S defined in (18) using the method of Lagrange multipliers.

Proposition 3.1. *Let $f \in L^2(\mathbb{R}^d, \mathbb{R})$, consider the following optimization problem:*

$$\min_{g \in L^2} \int |f - g|^2 \quad \textit{with} \quad g \geq 0 \quad \textit{and} \quad \int g f_k = d_k \ \ \textit{for} \ \ k = 1, ..., n \ , \tag{20}$$

where $(f_1, \cdots, f_n)$ are linearly independent functions in $L^2(\mathbb{R}^d, \mathbb{R})$, then there exists $(\alpha_1^*, \cdots, \alpha_n^*) \in \mathbb{R}^n$ such that the solution of (20), g^* is given by

$$g^* = (f - \sum_{k=1}^n \alpha_k^* f_k)^+ \ . \tag{21}$$

Proof. The Lagrange function associated to the problem (20) is:

$$\mathrm{L} = \frac{1}{2}\|f-g\|^2 - \sum_{k=1}^{n}\lambda_k(d_k - \langle f_k, g\rangle) - \mu g\ , \tag{22}$$

where $\langle f_k, g\rangle = \int f_k g$ is the usual scalar product in L^2 and μ a real valued function defined on $\mathbb{R}^d$. With the assumption that the functions f_k belong to L^2, the theory states [1] that (20) has a unique solution g^* given by the necessary and sufficient following conditions :

$$\begin{cases} g^* - f - \sum_{k=1}^{n}\lambda_k f_k - \mu = 0\ , \\ \mu g^* = 0\ , \\ \mu \geq 0\ , \\ g^* \geq 0\ . \end{cases} \tag{23}$$

Let us introduce the subset of $\mathbb{R}^d$, $A = \{x \in \mathbb{R}^d \,|\, g^*(x) > 0\}$. The second equation of (23) implies $\mu = 0$ on A. Then, using the first equation of (23), we obtain the following equality on the subset A

$$g^*\mathbf{1}_A = (f + \sum_{k=1}^{n}\lambda_k f_k)\mathbf{1}_A = (f + \sum_{k=1}^{n}\lambda_k f_k)^+\mathbf{1}_A\ .$$

Since $g^* \geq 0$ on $\mathbb{R}^d$ then $g^* \equiv 0$ outside of A. On the other hand, the first equation of (23) leads to

$$(f + \sum_{k=1}^{n}\lambda_k f_k)\mathbf{1}_{\bar{A}} = -\mu\mathbf{1}_{\bar{A}} \leq 0\ .$$

where $\bar{A}$ denotes the complement of A, hence

$$g^*\mathbf{1}_{\bar{A}} = 0 = (f + \sum_{k=1}^{n}\lambda_k f_k)^+\mathbf{1}_{\bar{A}}\ .$$

This finally yields the following equality valid on the whole space $\mathbb{R}^d$,

$$g^* = (f + \sum_{k=1}^{n}\lambda_k f_k)^+\ .$$

Then, we end the proof by setting $\alpha_k^* = -\lambda_k$ for $k = 1, \cdots, n$. □

3.3. The case where the functions defining the constraints are not necessarily in L^2

Proposition 3.1 gives the form of the solution of the problem (19) when the constraint functions $f_1, \cdots, f_n$ are in L^2 without any other conditions. To be able to take into account the *probability density constraint* for densities defined on $\mathbb{R}^d$, which corresponds to $f_1 \equiv 1$ and $d_1 = 1$, we would like to relax the assumption that the functions defining the constraints $f_1, \cdots, f_n$ belong to L^2. The following proposition achieves this goal, but on the other hand requires the existence of coefficients $\alpha_1^*, \cdots \alpha_n^*$ satisfying equation (25).

Proposition 3.2. *Let $f \in \mathrm{L}^2(\mathbb{R}^d, \mathbb{R})$, and S be the closed convex set defined in (18). Assume that for $k = 1, \cdots, n$*

$$\int |f f_k| < \infty \ . \tag{24}$$

Assume moreover that there exist reals $(\alpha_1^*, \cdots, \alpha_n^*)$ such that for $k = 1, \cdots, n$

$$\int (f - \sum_{k=1}^{n} \alpha_k^* f_k)^+ f_k = d_k \ . \tag{25}$$

Then the projection g^* of f on the convex subset S of L^2 non-negative functions with linear constraints (18), is determined by

$$g^* = (f - \sum_{k=1}^{n} \alpha_k^* f_k)^+ = \arg\min_{g \in S} \int |f - g|^2 \ . \tag{26}$$

Proof. The proof is similar to the proof of Proposition 2.2 and is based on the characterization of the projection X^* of a point Y on a closed convex set S in a Hilbert space [3]. $X^* \in S$ is the projection of Y on S iff

$$\langle Y - X^*, X - X^* \rangle \leq 0 \quad \text{for all} \quad X \in S \ .$$

In our case, S is the subset of L^2 non-negative functions with linear constraints (18). First, we easily check that S is convex. Then, we notice that under conditions (24) and (25) $(f - \sum_{k=1}^{n} \alpha_k^* f_k)^+ \in S$. Hence, it is sufficient to prove that the quantity,

$$\Delta = \int [f - (f - \sum_{k=1}^{n} \alpha_k^* f_k)^+] \, [g - (f - \sum_{k=1}^{n} \alpha_k^* f_k)^+] \ , \tag{27}$$

is non-positive for any function $g \in S$, where $\alpha_1^*, \cdots, \alpha_n^*$ are defined in (25).
Let us introduce the subsets of $\mathbb{R}^d$,

$$A^* = \{x \in \mathbb{R}^d \,|\, f(x) \geq \sum_{k=1}^{n} \alpha_k^* f_k(x)\} \quad \text{and} \quad \bar{A}^* = \{x \in \mathbb{R}^d \,|\, f(x) < \sum_{k=1}^{n} \alpha_k^* f_k(x)\} \ .$$

Splitting expression (27) of Δ into two terms yields

$$\begin{aligned}
\Delta &= \int_{A^*} [f - f + \sum_{k=1}^{n} \alpha_k^* f_k] \, [g - f + \sum_{k=1}^{n} \alpha_k^* f_k] + \int_{\bar{A}^*} fg \\
&= \sum_{k=1}^{n} \alpha_k^* \left[\int_{A^*} g f_k - \int (f - \sum_{k=1}^{n} \alpha_k^* f_k)^+ f_k \right] + \int_{\bar{A}^*} fg \\
&= \sum_{k=1}^{n} \alpha_k^* \left[\int_{A^*} g f_k - d_k \right] + \int_{\bar{A}^*} fg \ .
\end{aligned}$$

Then using the fact that $g \in S$ and hence $\displaystyle\int g f_k = d_k$ yields

$$\begin{aligned} \Delta &= -\sum_{k=1}^{n} \alpha_k^* \int_{\bar{A}^*} g f_k + \int_{\bar{A}^*} f g \\ &= \int_{\bar{A}^*} g(f - \sum_{k=1}^{n} \alpha_k^* f_k) \ . \end{aligned}$$

Then notice that this quantity is non-positive since $g \in S$ is a non-negative function, which ends the proof. □

In general, it can be difficult to check whether the existence condition (25) is satisfied or not. The next proposition describes some conditions under which (25) is satisfied for non-negative coefficients and for two non-negative functions (not necessarily lying in L^2) defining the linear constraints. This result will be used later, in Section 4, to determine a practical way to compute coefficients α_1^* and α_2^* in the case of two linear constraints.

Proposition 3.3. *Let f be a bounded function in $\mathrm{L}^2(\mathbb{R}^d, \mathbb{R})$, consider the following constrained optimization problem:*

$$\min_{g \in \mathrm{L}^2} \int |f-g|^2 \quad \text{with} \quad g \geq 0 \quad \text{and} \quad \int g f_k = d_k \quad \text{for} \ \ k = 1, 2 \ , \tag{28}$$

where f_1 and f_2 are two non-negative measurable functions defined on $\mathbb{R}^d$ which are linearly independent (and not necessarily lying in $\mathrm{L}^2(\mathbb{R}^d, \mathbb{R})$) with $f_1 > 0$ and bounded. Assume that the following condition is verified:

$$0 \leq d_k \leq \int f^+ f_k < \infty \ , \quad \text{for} \quad k = 1, \, 2 \ . \tag{29}$$

Then there exist non-negative scalars (α_1^*, α_2^*) such that the solution of (28) has the following form

$$g^* = (f - \alpha_1^* f_1 - \alpha_2^* f_2)^+ \ , \quad \text{with} \quad (\alpha_1^*, \alpha_2^*) \in \mathbb{R}^+ \times \mathbb{R}^+ \ , \tag{30}$$

if and only if one of the following conditions is verified:

Condition 1: *: $\int\limits_{f_2=0} f^+ f_1 \leq d_1$ and d_2 is such that,*

$$\int (f - \bar{\alpha}_2 f_2)^+ f_2 \leq d_2 \leq \int (f - \bar{\alpha}_1 f_1)^+ f_2 \ . \tag{31}$$

where $(\bar{\alpha}_1, \bar{\alpha}_2)$ are the non-negative solutions of

$$\int (f - \bar{\alpha}_k f_k)^+ f_1 = d_1 \ , \quad \text{for} \quad k = 1, 2 \ . \tag{32}$$

In that case $\alpha_1^* \in [0, \bar{\alpha}_1]$ and $\alpha_2^* \in [0, \bar{\alpha}_2]$.

Condition 2: *: $\int\limits_{f_2=0} f^+ f_1 > d_1$ and d_2 is such that,*

$$0 \leq d_2 \leq \int (f - \bar{\alpha}_1 f_1)^+ f_2 \ . \tag{33}$$

In that case $\alpha_1^* \in [\bar{\bar{\alpha}}_1, \bar{\alpha}_1]$ *and* $\alpha_2^* \in [0, \infty)$*, where* $\bar{\alpha}_1$ *and* $\bar{\bar{\alpha}}_1$ *are the non-negative solution of*

$$\int (f - \bar{\alpha}_1 f_1)^+ f_1 = d_1 \quad , \quad \text{and} \quad \int_{f_2=0} (f - \bar{\bar{\alpha}}_1 f_1)^+ f_1 = d_1 \ . \tag{34}$$

Under one of these conditions, the solution points (α_1^*, α_2^*) *is either a unique point or a continuum of points of* $\mathbb{R}^+ \times \mathbb{R}^+$.

Notice that this proposition considers only the case of non-negative coefficients (α_1, α_2).

Proof. By Proposition 3.2, it is sufficient to prove the existence of non-negative coefficients α_1^* and α_2^* satisfying (25). Let us introduce the non-negative function, H defined on $\mathbb{R}^+ \times \mathbb{R}^+$, such that for any $(\alpha_1, \alpha_2) \in \mathbb{R}^+ \times \mathbb{R}^+$

$$H(\alpha_1, \alpha_2) = \int (f - \alpha_1 f_1 - \alpha_2 f_2)^+ f_1 \ . \tag{35}$$

This function is well defined for non-negative scalars (α_1, α_2) because of condition (29). f_1 and f_2 being non-negative functions, the maximum value of $H(\alpha_1, \alpha_2)$ on $\mathbb{R}^+ \times \mathbb{R}^+$ is $\int f^+ f_1$ and is obtained for $(\alpha_1, \alpha_2) = (0, 0)$. Now, let us consider the following equation,

$$H(\alpha_1, \alpha_2) = d_1 \ . \tag{36}$$

By Lemma 1 in Appendix A, $\alpha_2 \mapsto H(0, \alpha_2) = \int (f - \alpha_2 f_2)^+ f_1$ decreases continuously as α_2 increases, from $H(0, 0) = \int f^+ f_1$ to $H(0, \infty) = \int_{f_2=0} f^+ f_1$. Now, let us consider the two following cases,

- Case 1 : $\int_{f_2=0} f^+ f_1 \leq d_1$. From (29) we deduce that $H(0, 0) \geq d_1$ while $H(0, \infty) \leq d_1$. Therefore it exists $\bar{\alpha}_2$ such that $H(0, \bar{\alpha}_2) = d_1$. For a given $\alpha_2 \in \mathbb{R}^+$, there is a solution $\alpha_1 = \alpha_1(\alpha_2)$ of equation (36) iff $\alpha_2 \leq \bar{\alpha}_2$. Indeed, if $\alpha_2 \leq \bar{\alpha}_2$, $\alpha_1 \mapsto H(\alpha_1, \alpha_2)$ decreases from $H(0, \alpha_2) \geq d_1$ to $H(\infty, \alpha_2) = 0$ (using again Lemma 1 in Appendix A and the fact that $f_1 > 0$). Hence, for any $\alpha_2 \in [0, \bar{\alpha}_2]$, there exists $\alpha_1(\alpha_2) \in \mathbb{R}^+$ such that
$$H(\alpha_1(\alpha_2), \alpha_2) = d_1 \ , \tag{37}$$
where $\bar{\alpha}_2$ is given by equation (32). Moreover, notice that $\alpha_2 \mapsto \alpha_1(\alpha_2)$ decreases continuously from $\alpha_1(0) = \bar{\alpha}_1$ (where $\bar{\alpha}_1$ is given by the first equation of (32)) to $\alpha_1(\bar{\alpha}_2) = 0$. Hence, in case 1, $\alpha_1^* \in [0, \bar{\alpha}_1]$.

- Case 2 : $\int_{f_2=0} f^+ f_1 > d_1$. From (29) and Lemma 1 in Appendix A, we deduce that $H(0, 0) \geq d_1$ while $H(0, \infty) > d_1$. Hence, $H(0, \alpha_2) \geq d_1$ for any $\alpha_2 \geq 0$, $\alpha_1 \mapsto H(\alpha_1, \alpha_2)$ decreases from $H(0, \alpha_2) > d_1$ to $H(\infty, \alpha_2) = 0$. Hence, for any $\alpha_2 \in \mathbb{R}^+$, there exists $\alpha_1(\alpha_2) \in \mathbb{R}^+$ such that
$$H(\alpha_1(\alpha_2), \alpha_2) = d_1 \ .$$
Notice that by Lemma 1 in Appendix A, $\alpha_2 \mapsto \alpha_1(\alpha_2)$ decreases continuously from $\alpha_1(0) = \bar{\alpha}_1$ to $\alpha_1(\alpha_2 = \infty) = \bar{\bar{\alpha}}_1$ where $\bar{\bar{\alpha}}_1$ is solution of $\int_{f_2=0} (f - \bar{\bar{\alpha}}_1 f_1)^+ f_1 = d_1$. Hence in this case $\alpha_1^* \in [\bar{\bar{\alpha}}_1, \bar{\alpha}_1]$.

Now, let us deal with the second constraint. Let us introduce the following non-negative function G defined for any $\alpha_2 \in [0, \bar{\alpha}_2]$ in case 1 and for any $\alpha_2 \in [0, \infty]$ in case 2:

$$G(\alpha_2) = \int (f - \alpha_1(\alpha_2) f_1 - \alpha_2 f_2)^+ f_2 \ . \tag{38}$$

- Case 1 : G varies continuously from $G(0) = \int (f - \bar{\alpha}_1 f_1)^+ f_2$ to $G(\bar{\alpha}_2) = \int (f - \bar{\alpha}_2 f_2)^+ f_2$. That is, $G(\alpha_2) = d_2$ has a solution iff $G(\bar{\alpha}_2) \leq d_2 \leq G(0)$, which gives the condition (31) and then $\alpha_2^* \in [0, \bar{\alpha}_2]$.

- Case 2 : G varies continuously from $G(0) = \int (f - \bar{\alpha}_1 f_1)^+ f_2$ to $G(\alpha_2 = \infty) = 0$. The last equality results from Lemma 1 in Appendix A which gives:

$$\lim_{\alpha_2 \to +\infty} \int_{f_2>0} (f - \alpha_1(\alpha_2) f_1 - \alpha_2 f_2)^+ f_2 = 0 \ ,$$

  observing that $\lim_{\alpha_2 \to +\infty} \alpha_1(\alpha_2) = \bar{\bar{\alpha}}_1 < \infty$. We conclude that $G(\alpha_2) = d_2$ has a solution iff $d_2 \leq G(0)$, which gives the condition (33). Therefore, $\alpha_2^* \in [0, \infty)$.

By Lemma 2 in Appendix A, we know that G is a decreasing function which implies that $G(\alpha_2) = d_2$ has either a single or a continuum of solutions α_2^* under the conditions (31) - (33) while $\alpha_1^* = \alpha_1(\alpha_2^*)$ is the unique solution of (37), when α_2^* is given. Hence, the solutions (α_1^*, α_2^*) is under conditions (31) - (33) either a unique point or a continuum of points of $\mathbb{R}^+ \times \mathbb{R}^+$. □

4. Kernel density estimation under constraints

In this section, we consider the problem of estimating a density p on $\mathbb{R}^d$ based on a sample $(X_1, \cdots, X_N)$ i.i.d according to p and an additionnal information on p (e.g. the support, the mean or the variance, ...). Let $f = \hat{p}$ be a kernel density estimate like in (2) with, for all $x \in \mathbb{R}^d$,

$$K_h(x) = \frac{1}{h^d} K(\frac{x}{h}) \ ,$$

and with K being possibly a negative kernel (a kernel which can take negative values). Standard kernel density estimation cannot exploit extra information on the underlying density. Different approaches have been proposed to modify the kernel estimate to take into account an apriori knowledge on the density [4,8,10]. In this section, we propose an alternative approach based on the projection of the first kernel estimate $\hat{p}$ on the set probabilty densities statisfying the constraints using the general results of Section 3. More precisely, we are interested in computing the density estimate g^* obtained as the L^2-projection of the possibly negative function $f = \hat{p}$ on the set of probability densities satisfying the constraints, when it is of the following form

$$g^*(x) = \left(\frac{1}{N} \sum_{i=1}^{N} K_h(x - X_i) - \alpha_1^* f_1(x) - \alpha_2^* f_2(x) \right)^+ \ , \tag{39}$$

where $(f_1, d_1) = (\mathbf{1}_{\mathbb{R}^d}, 1)$ represents the probability density constraint, (f_2, d_2) the additionnal information on the underlying probability density p and α_1^*, α_2^* are the reals such that,

$$\int_{\mathbb{R}^d} \left(\frac{1}{N} \sum_{i=1}^{N} K_h(x - X_i) - \alpha_1^* f_1(x) - \alpha_2^* f_2(x) \right)^+ f_k(x) \ dx = d_k \ , \tag{40}$$

for $k = 1, 2$. In the first subsection, we consider the case of a constraint on the support of the density. In the second subsection, we consider the case where the functions f_k defining the constraints are positive and bounded.

4.1. **Kernel density estimation under support constraint**

Suppose we want to approximate a real valued function f defined on $\mathbb{R}^d$ by a probability density with support $C \subset \mathbb{R}^d$. Similarly as for non-positive density estimation in Section 2.2, one possibility, if $f \in L^1$, is to use the following approximation,

$$g_1^*(x) = \frac{f^+(x)\mathbf{1}_C(x)}{\int_{\mathbb{R}^d} f^+(x)\mathbf{1}_C(x)dx} \ , \quad \text{for all} \quad x \in \mathbb{R}^d \ , \tag{41}$$

Notice that, here again, the function g_1^* has the following property,

$$\|g_1^* - g\|_1 \leq \|f - g\|_1 \ , \tag{42}$$

for any probability density g, (see Theorem 3 p. 269 in [5], for a proof).

An alternative is to consider the L^2-projection of f on the set of probability densities with support C. The following Proposition gives the parametric form of this projection in a specific case.

Proposition 4.1. *Let f be a bounded function in $L^2(\mathbb{R}^d,\mathbb{R})$, consider the following constrained optimization problem:*

$$\min_{g\in L^2} \int |f-g|^2 \quad \textit{with} \quad g \geq 0 \ , \quad \int g = 1 \quad \textit{and} \quad \int_C g = 1 \ , \tag{43}$$

where C is a given subset of $\mathbb{R}^d$. Assume that the following condition is verified:

$$1 \leq \int_C f^+ \leq \int f^+ < \infty \ . \tag{44}$$

Then the solution of (43) has the following form

$$g^*(x) = (f(x) - \alpha_1^* - \alpha_2^* \mathbf{1}_C(x))^+ \tag{45}$$

with $\quad \alpha_1^* = \sup_{x\in\bar{C}} f(x)$ *and* α_2^* *such that* $\int g^* = 1$.

Proof. Notice that $\int g = 1$ and $\int_C g = 1$ implies that the set of constraints represents the densities with support included in C. Now, since $\int_C f^+ \geq 1$ and f is bounded then there exists $\alpha^* \geq 0$ such that

$$\int_C (f-\alpha^*)^+ = 1 \ .$$

Then, if we set $\alpha_2^* = \alpha^* - \alpha_1^*$, we easily check that $g^* = 0$ on $\bar{C}$ and that the constraints are satisfied

$$\int (f - \alpha_1^* - \alpha_2^*\mathbf{1}_C)^+ = \int_C (f-\alpha^*)^+ + \int_{\bar{C}} (f-\alpha_1^*)^+ = 1 \ .$$

By Proposition 3.2 we conclude that (45) is the L^2-projection of f on the set of constraints. Notice that α_2^* can be negative. □

Just as Proposition 2.3 in Section 2.2, the following proposition shows that the L^2-projection characterized by (26) and (9) approximates f in the L^1 sense as well as approximation (41). Of course, the L^2-projection outperforms approximation (41) in the L^2 sense.

Proposition 4.2. *Let f be a function in $L^2(\mathbb{R}^d,\mathbb{R})$, such that $\int f^+(x) \geq 1$. Then if the L^2-projection of f on the subset of L^2-probability densities with support $C \subset \mathbb{R}^d$ denoted by g^* has the following form*

$$g^* = (f - \alpha_1^* - \alpha_2^*\mathbf{1}_C)^+ \ , \quad \textit{with} \quad (\alpha_1^*, \alpha_2^*) \in \mathbb{R}^2 \ ,$$

then g^* is as close to the function f as g_1^*, in the L^1 sense i.e.

$$\|f - g_1^*\|_1 = \|f - g^*\|_1 \ , \quad \textit{with} \quad g_1^* = \frac{f^+\mathbf{1}_C}{\int_C f^+} \ .$$

Proof. The proof is similar to the proof of Proposition 2.3. □

Under assumptions of Proposition 4.1, where the L^2-projection has the form (45), one can use exactly the same procedure as described in Section 2.3 to compute $\alpha^* = \alpha_1^* + \alpha_2^* \geq 0$ as a quantile (assuming that we know how to simulate the probability density $K^+/\int K^+$). Then, to compute $\alpha_1^* = \sup_{\bar{C}} f$, in the specific case where $f = \hat{p}$, is a kernel density estimate i.e. $f(x) = \hat{p}(x) = \frac{1}{N}\sum_{i=1}^{N} K_h(x - X_i)$, it can be interesting to notice that if the support C is convex then

$$\sup_{x\in\bar{C}} f(x) = \sup_{x\in\partial C} f(x) \ ,$$

where ∂C is the frontier of C. Indeed, no sample X_i fall in $\bar{C}$ (and of course K is assumed unimodal).

4.2. Kernel density estimation with bounded moment constraint

In this section, we consider the case where we have an a priori knowledge on a moment of the underlying density p:

$$\mathbb{E}[f_2(X)] = d_2 \ , \quad \text{where } X \sim p \ , \tag{46}$$

and f_2 is a positive and bounded function defined on $\mathbb{R}^d$. In this case, we wish that our estimate will also satisfy the moment constraint (46). To achieve this goal, we propose to project the kernel density estimate $\hat{p}$ (2) on the set of probability densities satisfying (46). The following Corollary which derives from Proposition 3.3 will help us to characterize and to compute such estimate with bounded moment constraint.

Corollary 4.3. *Let f be a bounded function in $\mathrm{L}^2(\mathbb{R}^d, \mathbb{R})$, consider the following constrained optimization problem:*

$$\min_{g\in \mathrm{L}^2} \int |f-g|^2 \quad \textit{with} \quad g \geq 0 \ , \quad \int g f_1 = d_1 \quad \textit{and} \quad \int g f_2 = d_2 \ , \tag{47}$$

where f_1 and f_2 are positive and bounded functions defined on $\mathbb{R}^d$. Assume that the following condition is verified:

$$0 \leq d_k \leq \int f^+ f_k < \infty \ , \quad \textit{for} \quad k = 1,\, 2 \ . \tag{48}$$

Then there exist four non-negative scalars $(\bar{\alpha}_k^\ell)_{k,\ell=1,2}$ such that

$$\int (f - \bar{\alpha}_k^\ell f_k)^+ f_\ell = d_\ell \ , \quad \textit{for} \quad k = 1,2 \quad \textit{and} \quad \ell = 1,2 \ . \tag{49}$$

For any $k, \ell = 1, 2$, for any $\beta \in [0, \bar{\alpha}_{\bar{k}}^\ell]$, there exists a non-negative real $\alpha_k^\ell(\beta) \in [0, \bar{\alpha}_k^\ell]$ such that

$$\int (f - \alpha_k^\ell(\beta) f_k - \beta f_{\bar{k}})^+ f_\ell = d_\ell \ , \quad \textit{where} \quad \bar{k} = \begin{cases} 1 & \textit{if} \ \ k = 2 \\ 2 & \textit{if} \ \ k = 1 \end{cases} \ , \tag{50}$$

and $\beta \mapsto \alpha_k^\ell(\beta)$ is a decreasing function from $\alpha_k^\ell(0) = \bar{\alpha}_k^\ell$ to $\alpha_k^\ell(\bar{\alpha}_{\bar{k}}^\ell) = 0$. Then there exist non-negative scalars (α_1^*, α_2^*) such that the solution of (47) has the following form

$$g^* = (f - \alpha_1^* f_1 - \alpha_2^* f_2)^+ \ , \quad \textit{with} \quad (\alpha_1^*, \alpha_2^*) \in \mathbb{R}^+ \times \mathbb{R}^+ \ , \tag{51}$$

if and only if the following condition is satisfied:

$$(\bar{\alpha}_1^1 \leq \bar{\alpha}_1^2 \quad \textit{and} \quad \bar{\alpha}_2^1 \geq \bar{\alpha}_2^2) \quad \textit{or} \quad (\bar{\alpha}_1^1 \geq \bar{\alpha}_1^2 \quad \textit{and} \quad \bar{\alpha}_2^1 \leq \bar{\alpha}_2^2). \tag{52}$$

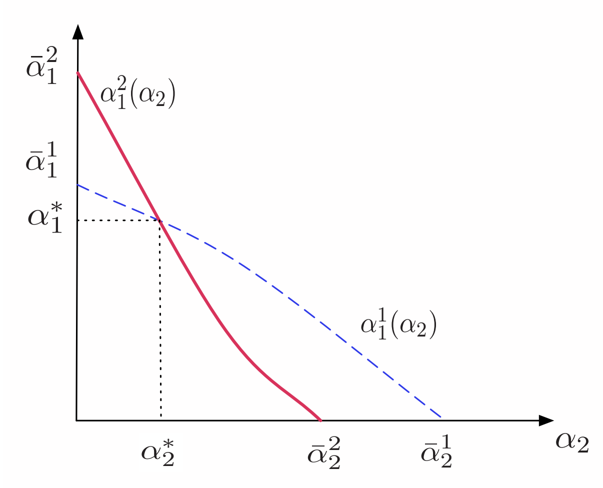


FIGURE 2. α_1^2 and α_1^1 as functions of α_2

If condition (52) is verified then $\alpha_1^* \in [0, \min(\bar{\alpha}_1^1, \bar{\alpha}_1^2)]$ *and* $\alpha_2^* \in [0, \min(\bar{\alpha}_2^2, \bar{\alpha}_2^1)]$ *are solutions of the following equivalent systems:*

$$\left\{ \begin{array}{lcl} \alpha_1^2(\alpha_2) & = & \alpha_1^1(\alpha_2) \ , \\ \alpha_1 & = & \alpha_1^2(\alpha_2) \ . \end{array} \right. \quad or \quad \left\{ \begin{array}{lcl} \alpha_2^1(\alpha_1) & = & \alpha_2^2(\alpha_1) \ , \\ \alpha_2 & = & \alpha_2^1(\alpha_1) \ . \end{array} \right. \tag{53}$$

Proof. The proof is a consequence of Proposition 3.3 noticing that since f_1 and f_2 are positive we only consider the Condition 1 of Proposition 3.3. The figure 2 represents (α_1^*, α_2^*) as the solution of the first system of (53) in the case where $\bar{\alpha}_1^1 \leq \bar{\alpha}_1^2$ and $\bar{\alpha}_2^1 \geq \bar{\alpha}_2^2$. □

Here again, one can use a similar approach as described in Section 2.3 to compute (α_1^*, α_2^*) as quantiles, when $f = \hat{p}$ is a kernel density estimate. Let us consider the random variable $D^{k,\ell}(\beta)$ defined for any integer $k, \ell = 1, 2$ and for any non-negative real β as follows:

$$D^{k,\ell}(\beta) = \frac{\frac{1}{N}\left[\sum_{i=1}^N K_h(Y - X_i) f_\ell(Y) - U \sum_{i=1}^N K_h^+(Y - X_i)\right] - \beta f_{\bar{k}}(Y) f_\ell(Y)}{f_k(Y) f_\ell(Y)} \ , \tag{54}$$

where U is a uniform random variable on $[0, 1]$ independent of Y which follows the density g (13) given $(X_1, ..., X_N)$,

$$g(x) = \frac{1}{N} \sum_{i=1}^N \frac{K_h^+(x - X_i)}{\int K^+} \ , \quad \text{for all} \quad x \in \mathbb{R}^d \ ,$$

The following Proposition allows us to compute $\alpha_k^\ell(\beta)$ as a quantile of $D^{k,\ell}(\beta)$.

Proposition 4.4. *Assume that the functions f_1 and f_2 defining the constraints are non-negative and bounded by one and that Condition (48) of Corollary 4.3 is satisfied. Then for any $k, \ell \in \{1, 2\}$ and for any $\beta \in [0, \bar{\alpha}_{\bar{k}}^\ell]$, the non-negative real $\alpha_k^\ell(\beta)$ defined by (50) is the $(d_\ell / \int K^+)$-quantile of $D^{k,\ell}(\beta)$ i.e.*

$$\mathbb{P}(D^{k,\ell}(\beta) \geq \alpha_k^\ell(\beta)) = \frac{d_\ell}{\int K^+} \leq 1 \ . \tag{55}$$

Notice first that since f_1 and f_2 are positive and bounded functions, then it is always possible to consider the case where the functions defining the constraints are bounded by one, in replacing (f_1, d_1) by $(f_1/\sup_{\mathbb{R}^d} f_1, d_1/\sup_{\mathbb{R}^d} f_1)$ and (f_2, d_2) by $(f_2/\sup_{\mathbb{R}^d} f_2, d_2/\sup_{\mathbb{R}^d} f_2)$.

Hence $\alpha_k^\ell(\beta)$ can be estimated empirically in simulating an i.i.d. sample $(D_1^{k,\ell}(\beta), \cdots, D_M^{k,\ell}(\beta))$ according to the law of $D^{k,\ell}(\beta)$ and in computing an approximation of the $(d_\ell/\int K^+)$-quantile of $D^{k,\ell}(\beta)$ using the corresponding ordered statistic.

First we compute $\bar{\alpha}_1^1$, $\bar{\alpha}_1^2$, $\bar{\alpha}_2^1$, $\bar{\alpha}_2^2$, (using definition (49)), to check wether Condition (52) is satisfied or not. If Condition (52) is satisfied then one can approximate (α_1^*, α_2^*) by dichotomy as follows. Let us consider a small threshold $\epsilon > 0$ of tolerated error.

Algorithm 4.5. *Computation of* (α_1^*, α_2^*).

(1) *Generate independently* $Y_1, \cdots Y_M$ *i.i.d. according to* g *and* $U_1, \cdots, U_M$ *i.i.d uniformly in* $[0, 1]$*;*
(2) *Set* $\beta_p = \beta_0 = \bar{\alpha}_2^2/2$*;*
(3) *Compute* $\hat{\alpha}_1^1(\beta_p)$ *and* $\hat{\alpha}_1^2(\beta_p)$ *by ordering* $(D_1^{1,1}(\beta_p), \cdots, D_M^{1,1}(\beta_p))$ *and* $(D^{1,2}(\beta_p), \cdots, D_M^{1,2}(\beta_p))$*;*
(4) *If* $|\hat{\alpha}_1^1(\beta_p) - \hat{\alpha}_1^2(\beta_p)| \le \epsilon$ *then set* $\hat{\alpha}_1^* = \hat{\alpha}_1^1(\beta_p)$ *and* $\hat{\alpha}_2^* = \beta_p$*;*
(5) *Else if* $\alpha_1^2(\beta_p) - \alpha_1^1(\beta_p) > \epsilon$ *then set* $\beta_{p+1} = \beta_p/2$ *else if* $\alpha_1^1(\beta_p) - \alpha_1^2(\beta_p) > \epsilon$ *then set* $\beta_{p+1} = (\beta_p + \bar{\alpha}_2^2)/2$ *and go to step 3.*

Noticing that α_1^2 is the inverse of α_2^2 and that similarly α_1^1 is the inverse of α_2^1, one can also use a fixed-point algorithm to estimate (α_1^*, α_2^*) as solutions of one of the following systems:

$$\left\{ \begin{array}{lcl} \alpha_2^1(\alpha_1^2(\alpha_2)) & = & \alpha_2 \ , \\ \\ \alpha_1 & = & \alpha_1^2(\alpha_2) \ . \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{lcl} \alpha_1^2(\alpha_2^1(\alpha_1)) & = & \alpha_1 \ , \\ \\ \alpha_2 & = & \alpha_2^1(\alpha_1) \ . \end{array} \right. \tag{56}$$

At each iteration of the fixed-point algorithm, one can approximate the evaluation of functions α_k^ℓ as a quantile by Proposition 4.4.

Proof. Let α be a given non-negative real, the following events are equivalent,

$$\begin{aligned} \left\{ D^{k,\ell}(\beta) \ge \alpha \right\} &\equiv \left\{ \frac{1}{N} \sum_{i=1}^N K_h(Y - X_i) f_\ell(Y) - \alpha f_k(Y) f_\ell(Y) - \beta f_{\bar{k}}(Y) f_\ell(Y) \right. \\ &\qquad \left. \ge \frac{U}{N} \sum_{i=1}^N K_h^+(Y - X_i) \right\} \\ &\equiv \left\{ \frac{\frac{1}{N} \sum_{i=1}^N K_h(Y - X_i) - \alpha f_k(Y) - \beta f_{\bar{k}}(Y)}{\frac{1}{N} \sum_{i=1}^N K_h^+(Y - X_i)} f_\ell(Y) \ge U \right\} \\ &\equiv \left\{ \frac{\left(\frac{1}{N} \sum_{i=1}^N K_h(Y - X_i) - \alpha f_k(Y) - \beta f_{\bar{k}}(Y) \right)^+ f_\ell(Y)}{\frac{1}{N} \sum_{i=1}^N K_h^+(Y - X_i)} \ge U \right\} \\ &\equiv \left\{ \frac{H_{\alpha,\beta}^{k,\ell}(Y)}{g(Y)} \ge U \right\} \ , \end{aligned}$$

where

$$H^{k,\ell}_{\alpha,\beta}(Y) = \frac{1}{\int K^+}\left(\frac{1}{N}\sum_{i=1}^{N} K_h(Y - X_i) - \alpha f_k(Y) - \beta f_{\bar{k}}(Y)\right)^+ f_\ell(Y) \ .$$

This yields

$$\begin{aligned}\mathbb{P}(D^{k,\ell}(\beta) \geq \alpha) &= \mathbb{P}\left[\frac{H^{k,\ell}_{\alpha,\beta}(Y)}{g(Y)} \geq U\right] \ , \\ &= \int_{\mathbb{R}^d} \min\left(1, \frac{H^{k,\ell}_{\alpha,\beta}(y)}{g(y)}\right) g(y)dy \ .\end{aligned}$$

Since the functions f_1 and f_2 defining the constraints are both non-negative and bounded by one, and since moreover the scalars α and β are also non-negative, then $H^{k,\ell}_{\alpha,\beta}(y) \leq g(y)$ for any $y \in \mathbb{R}^d$. Therefore, for any $\alpha \geq 0$,

$$\mathbb{P}(D^{k,\ell}(\beta) \geq \alpha) = \int_{\mathbb{R}^d} H^{k,\ell}_{\alpha,\beta}(y)dy \ .$$

Notice that for $\alpha = \alpha^\ell_k(\beta)$, we get

$$\int_{\mathbb{R}^d} H^{k,\ell}_{\alpha,\beta}(y)dy = \frac{d_\ell}{\int K^+} \ ,$$

which yields the desired result. □

5. Simulation results

5.1. Probability density constraint for negative kernel density estimate

In this section, we are interested in analyzing the performances of the L^2-projection, g^* (11), as a density estimate. We focus on the one dimensional case. Let us consider the negative kernel, introduced by Bartlett in [2] and defined, for all $x \in \mathbb{R}$, by

$$K_{\text{Bar}}(x) = \begin{cases} \frac{9}{8}(1 - \frac{5}{3}|x|^2) & \text{if } |x| < 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{57}$$

Epanechnikov (4) and Bartlett's kernels are represented on figure 3.

Notice that K_{Bar} is a four order kernel, hence this kernel allows a bias reduction of the density estimates (see Section 2.1). Moreover, notice that we know how to simulate the probability density $\frac{K^+_{\text{Bar}}}{\int K^+_{\text{Bar}}}$. Indeed, for all $x \in \mathbb{R}$,

$$\frac{K^+_{\text{Bar}}(x)}{\int_{\mathbb{R}} K^+_{\text{Bar}}(x)dx} = \sqrt{\frac{5}{3}} K_{\text{Epan}}\left(\sqrt{\frac{5}{3}}x\right) \ .$$

Hence, generating from $\frac{K^+_{\text{Bar}}}{\int K^+_{\text{Bar}}}$ reduces to generate from the Epanechnikov kernel [16]. Let $(X_1, \cdots, X_N)$ be a one dimensional sample from a density p. If we assume that the underlying density p is a standard centered Gaussian, then the optimal bandwidth associated to Bartlett's kernel can be explicitly computed as [16],

$$h_{\text{Bar}} = (3360\sqrt{\pi})^{1/9} N^{-1/9} \ . \tag{58}$$

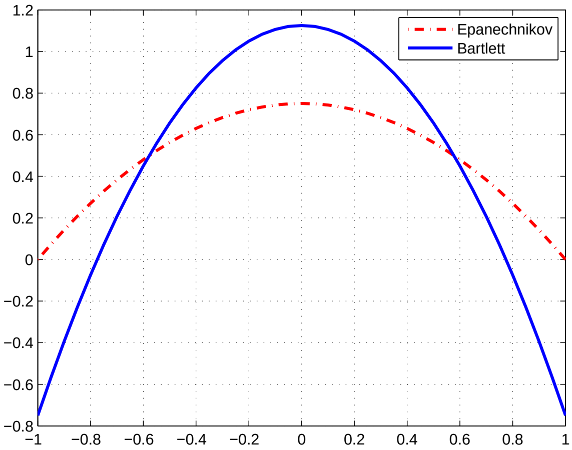


FIGURE 3. Epanechnikov and Bartlett kernel.

For a given sample size N, we simulate an N-sample according to a given probability density p on $\mathbb{R}$. Then, we compute:

(1) the kernel density estimate, $\hat{p}_{\text{Epan}}$, using the Epanechnikov kernel (4), with the bandwidth given by (5) and the mean integrated squared error w.r.t. the underlying density p,

$$\text{MISE}_{\text{Epan}} = \mathbb{E}\|\hat{p}_{\text{Epan}} - p\|_2^2 \ ;$$

(2) the kernel density estimate, $\hat{p}_{\text{Bar}}$, using the Bartlett kernel (57), with the bandwidth given by (58) and the mean integrated squared error w.r.t. the underlying density p,

$$\text{MISE}_{\text{Bar}} = \mathbb{E}\|\hat{p}_{\text{Bar}} - p\|_2^2 \ ;$$

(3) the L^2-projection of $\hat{p}_{\text{Bar}}$, $\hat{p}_{\text{Bar},2}$, defined by (11), using Proposition 2.4 to estimate α^* as a quantile of 10000 samples drawn from D defined in (14), and the mean integrated squared error w.r.t. the underlying density p,

$$\text{MISE}_{\text{Bar},2} = \mathbb{E}\|\hat{p}_{\text{Bar},2} - p\|_2^2 \ ;$$

(4) the "L^1-projection" of $\hat{p}_{\text{Bar}}$, $\hat{p}_{\text{Bar},1}$, defined by (6) (by means of 10000 grid points to estimate $\int \hat{p}_{\text{Bar}}^+$) and the mean integrated squared error w.r.t. the underlying density p,

$$\text{MISE}_{\text{Bar},1} = \mathbb{E}\|\hat{p}_{\text{Bar},1} - p\|_2^2 \ .$$

This procedure is repeated one hundred times such as to estimate empirically the MISE of each density estimate by the empirical mean of the MISE computed at each simulation.
The estimated MISE of the Epanechnikov estimation and the L^2-projection are reported for different sample sizes ($N = 25, 50, 100, \cdots, 4000$),

- on the left graph of figure 4, in the case where the underlying density p is a standard centered Gaussian,

$$p \ \sim \ \mathcal{N}(0,1) \ ;$$

- on the right graph of figure 4, in the case where the underlying density p is a Gaussian mixutre, $p \ \sim \ 0.5 * \mathcal{N}(0,1) + 0.5 * \mathcal{N}(5,1)$.

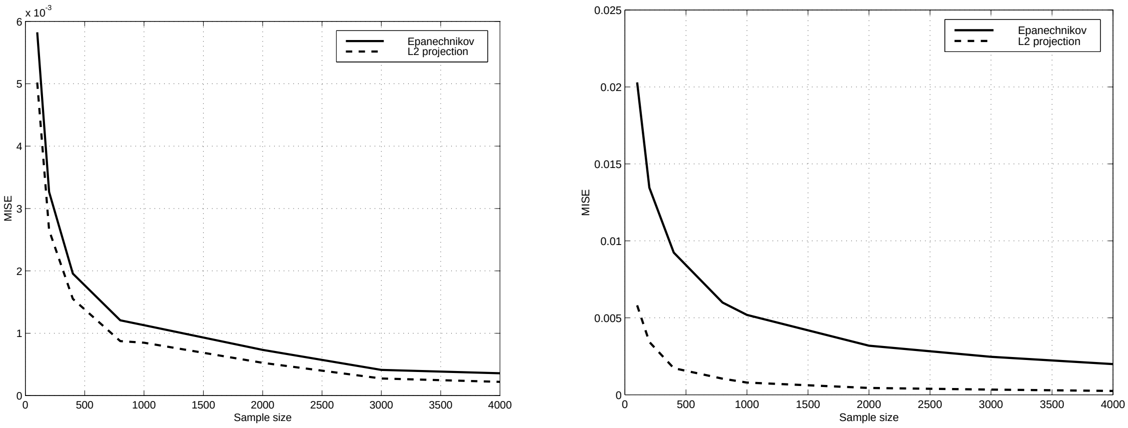


FIGURE 4. $\mathrm{MISE}_{\mathrm{Bar},2}$ and $\mathrm{MISE}_{\mathrm{Epan}}$ w.r.t. the sample size N when the underlying density is Gaussian (on the left) and a mixture of Gaussians (on the right).

The estimated MISE of Bartlett's estimate ($\mathrm{MISE}_{\mathrm{Bar}}$) and the "$L^1$-projection" ($\mathrm{MISE}_{\mathrm{Bar},1}$) are not reported on those figures because they are approximately equal to $\mathrm{MISE}_{\mathrm{Bar},2}$. Hence, on examples considered here, one can observe that

- the "L^1-projection" is approximately equivalent to the L^2-projection;
- the projection doesn't improve so much the estimation provided by Bartlett's estimate (which has the drawback provide negative values);
- the L^2-projection estimate is significantly better than Epanechnikov's estimate, especially for the multimodal case (see figures 4).

Hence, the approach developed in this paper leads to density estimates that are both probability densities and more precise, in many cases, than classical estimates (using positive kernels). Theoritically, Propositions 3.2 and 2.3 show that the L^2-projection is more precise than (at least as precise as) the "L^1-projection", in the L^1 sense or in the L^2 sense, (if p and $\hat{p}$ are in L^2). But practical examples considered here showed that these "projections" were equivalent for the special cases considered. It would be interesting to consider other practical cases where the L^2-projection is much better than the "L^1-projection".

5.2. **Probability density constraint with support constraint**

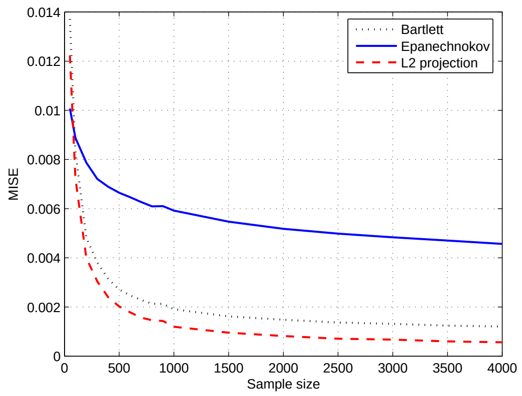
In this subsection, we consider two examples in dimension one of projection of a function on the set probability densities with a given support $C \subset \mathbb{R}$.

In the first example, we consider the framework of kernel density estimation just as in the above subsection 5.1 with an underlying density which is an exponential density with parameter $\lambda = 0.1$ i.e.

$$p(x) = \lambda \exp(-\lambda x)\, \mathbf{1}_{x \geq 0}(x) \ , \quad \text{for all } x \in \mathbb{R} \ . \tag{59}$$

The difference is that we use the a priori knowledge on the support, $C = \mathbb{R}^+$, to produce the new estimates $\hat{p}_{\mathrm{Bar},1}$ and $\hat{p}_{\mathrm{Bar},2}$ corresponding to the estimates g_1^* and g^* in (41, 45), on the base of the negative kernel estimate $\hat{p}_{\mathrm{Bar}}$. Notice that α_1^* and α_2^* characterizing $\hat{p}_{\mathrm{Bar},2}$ are estimated as suggested at the end of Section 4.1:

- $\alpha^* = \alpha_1^* + \alpha_2^*$ is estimated as the $1/\int K^+$ empirical quantile of D based on 10000 samples drawn from D (see (14));
- $\alpha_1^* = \sup\limits_{\bar{C}} \hat{p}(x) = \hat{p}_{\mathrm{Bar}}(0)$.



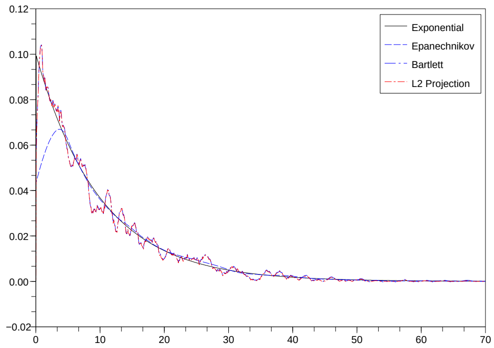


FIGURE 5. MISE of each density estimate (with or without support constraint) w.r.t. the sample size N (on the left) and example of density estimates for $N = 2000$ (on the right).

This approach is valid in the specific case where Condition (44) of Proposition 4.1, $\int_{\mathbb{R}+} \hat{p}^+_{\mathrm{Bar}} \geq 1$, is verified, which requires implicitly that the number of observations N is sufficiently large. Hence our approach is not valid for small samples. On figure 5, we can observe the estimated MISE (on 100 runs) of Epanechnikov's estimation $\hat{p}_{\mathrm{Epan}}$ and Bartlett's estimate $\hat{p}_{\mathrm{Bar}}$ (without taking into account the support constraint), and of the constrained estimate $\hat{p}_{\mathrm{Bar},2}$, for different sample sizes ($N = 100, \cdots, 4000$). We have not represented the performance of $\hat{p}_{\mathrm{Bar},1}$, the L^1-projection of Bartlett's kernel given by (41), since it's MISE is close to the MISE of $\hat{p}_{\mathrm{Bar},2}$. We observe that the projection improves Bartlett's estimate $\hat{p}_{\mathrm{Bar}}$ and produces a density estimate that is much better than Epanechnikov's kernel density estimate $\hat{p}_{\mathrm{Epan}}$.

In the second example, we consider the function f defined by $f(x) = \dfrac{1}{\pi}\dfrac{\sin(x)}{x}$ translated such that it is centered in $x = 10$. We intend to project f on the set of probability densities with a bounded support $C = [0, 25]$. We can check that this function satisfies Condition (44) of Proposition 4.1. On the figure 6, the L^2-projection g^* defined in (45) is compared with the so called L^1-projection $g^*_1(x) = \mathbf{1}_{x\in C} f^+(x)/\int_C f^+(x)$. Simulation results show that the L^1 error is identical for the 2 projections which agrees with Proposition 2.3 but the L^2-error is 0.18 for g^*_1 and 0.15 for g^*. Notice that, unlike the L^1-projection, the L^2-projection tends to preserve the main mode of the original function.

5.3. Probability density constraint with mean constraint

Here, we consider a function f with bounded support C, with $f_1 = 1$ and $d_1 = 1$ for the probability density constraint and $f_2 \equiv x$ and $d_2 = \mu \in \mathbb{R}$ for the mean constraint. The constraint functions are therefore in $\mathrm{L}^2(C)$ and the method of Lagrange multipliers of Proposition 3.1 gives the existence of the coefficients (α^*_1, α^*_2) (possibly negative) determined by the two constraints. The L^2-projection g^* of f is given by:

$$g^*(x) = (f(x) - \alpha^*_1 - \alpha^*_2 x))^+ \ . \tag{60}$$

As an illustrative example we set $f(x) = k(\varphi_1(x) + \varphi_2(x))$ where $\varphi_i = N(\beta_i, \sigma_i^2)$ are two Gaussian densities of mean and standard deviation ($\beta_1 = 1, \sigma_1 = 1$) and ($\beta_2 = 3, \sigma_2 = 0.2$). The support of f is given by $C = [\beta_1 - 3\sigma_1, \beta_1 + 3\sigma_1]$ and k is such that f is a probability density. The mean of f is 2 and the imposed mean is $\mu = 3$. On figure 7, g^* is compared with the original function.

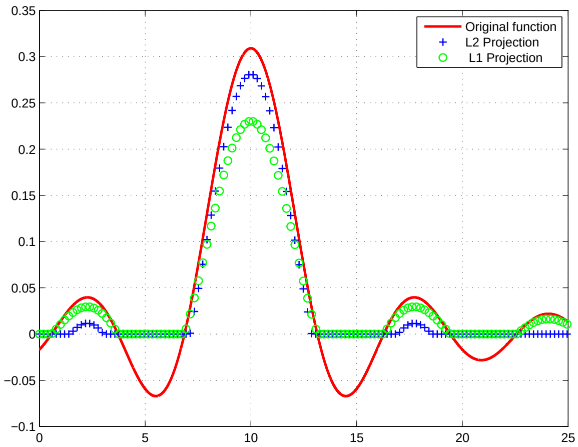


FIGURE 6. Original function compared with the 2 projections. Case of constraint functions not belonging to L^2.

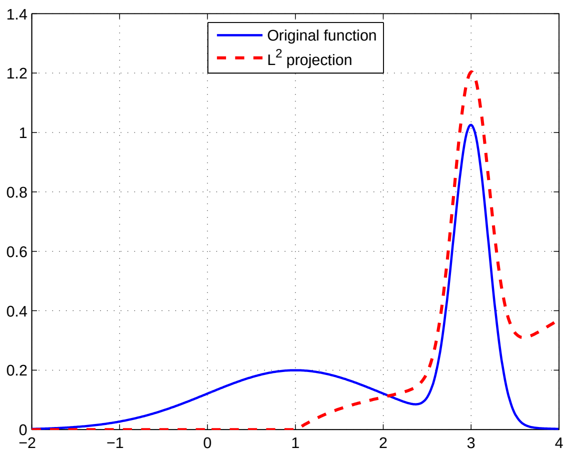


FIGURE 7. Original function f compared with the L^2 -projection g^*. Case of constraint functions belonging to L^2.

APPENDIX A

A.1. **Lemma 1**

Let f be a bounded real valued function defined on $\mathbb{R}^d$ and f_1, f_2 two non negative measurable functions defined on $(\mathbb{R}^d, \mathcal{R}^d)$ such that f_1 is bounded and $\int f^+ f_1 < \infty$. Then, we can define the nonnegative function k on $\mathbb{R}^+$ such that for any $\alpha_2 \in \mathbb{R}^+$,

$$k(\alpha_2) = \int (f - \alpha_2 f_2)^+ f_1 \ .$$

k is a non-increasing continuous function varying from $k(0) = \int f^+ f_1$ to $k(+\infty) = \int\limits_{\{x\,:\,f_2(x)=0\}} f^+ f_1$.

Proof. Clearly k is a non-increasing continuous function with $k(0) = \int f^+ f_1$. For any $\alpha_2 \in \mathbb{R}^+$, we can decompose $k(\alpha_2)$ as,

$$k(\alpha_2) = \int (f - \alpha_2 f_2)^+ f_1 = \int\limits_{f_2=0} f^+ f_1 + \int\limits_{f_2>0} (f - \alpha_2 f_2)^+ f_1 \ . \tag{61}$$

Then notice that

$$\int\limits_{f_2>0} (f - \alpha_2 f_2)^+ f_1 \leq M M_1 \int \mathbf{1}_{\{x\,:\,f_2(x)\in(0,M/\alpha_2)\}} \ , \quad \text{where} \quad M = \sup_{x\in\mathbb{R}^d} f(x) \ ,$$

and $M_1 = \sup_{x\in\mathbb{R}^d} f_1(x)$. Since f_2 is measurable then the right hand of the above inequality vanishes to zero when α_2 grows to infinity, which ends the proof. □

A.2. **Lemma 2**

Let f be a bounded function in $\mathrm{L}^2(\mathbb{R}^d, \mathbb{R})$. Consider f_1 and f_2, two non-negative measurable functions defined on $\mathbb{R}^d$ which are linearly independent (and not necessarily lying in $\mathrm{L}^2(\mathbb{R}^d, \mathbb{R})$) with $f_1 > 0$ and bounded. Assume that the following condition is verified:

$$0 \leq d_k \leq \int f^+ f_k < \infty \ , \quad \text{for} \quad k = 1, 2 \ . \tag{62}$$

Let us consider the two following cases,

Case 1: : $\int\limits_{f_2=0} f^+ f_1 \leq d_1$. Then there exists $\bar{\alpha}_2$ such that $\int (f - \bar{\alpha}_2 f_2)^+ f_1 = d_1$.

Case 2: : $\int\limits_{f_2=0} f^+ f_1 > d_1$.

Let $G(\alpha_2) = \int (f - \alpha_1(\alpha_2) f_1 - \alpha_2 f_2)^+ f_2$ defined in (38), with $\alpha_2 \in [0, \bar{\alpha}_2]$ in case 1 and with $\alpha_2 \in [0, \infty]$ in case 2. Then G is a decreasing function.

Proof. First, applying the implicit function theorem to $H(\alpha_1(\alpha_2), \alpha_2) = d_1$ defined in (37) gives :

$$\frac{d\alpha_1(\alpha_2)}{d\alpha_2} = -\left(\frac{\partial H}{\partial \alpha_1}\right)^{-1} \left.\frac{\partial H}{\partial \alpha_2}\right|_{\alpha_1 = cte} \ . \tag{63}$$

Formally, by approximating

$$\alpha_1(\alpha_2 + \varepsilon) \approx \alpha_1(\alpha_2) + \varepsilon \frac{d\alpha_1(\alpha_2)}{d\alpha_2} \ ,$$

for small ε , we develop $G(\alpha_2 + \varepsilon)$ around α_2 as follows

$$G(\alpha_2 + \varepsilon) - G(\alpha_2) \approx \int_{\psi \geq \varepsilon(\frac{d\alpha_1}{d\alpha_2} f_1 + f_2)} \psi f_2 - \int_{\psi \geq 0} \psi f_2 - \varepsilon \int_{\psi \geq 0} (\frac{d\alpha_1}{d\alpha_2} f_1 + f_2) f_2 \ ,$$

where $\psi = f - \alpha_1(\alpha_2)f_1 - \alpha_2 f_2$. One can prove that:

$$\int \psi f_2 \mathbf{1}_{\psi \geq \varepsilon(\frac{d\alpha_1}{d\alpha_2}f_1 + f_2)} - \int_{\psi \geq 0} \psi f_2 \approx \varepsilon(\frac{d\alpha_1}{d\alpha_2}f_1 + f_2)$$
$$\left(\int f_2 \mathbf{1}_{\varepsilon(\frac{d\alpha_1}{d\alpha_2}f_1+f_2) \leq \psi \leq 0} - \int f_2 \mathbf{1}_{0 \leq \psi \leq \varepsilon(\frac{d\alpha_1}{d\alpha_2}f_1+f_2)}\right) = \Delta(\varepsilon) \qquad (64)$$

Observing that $\lim_{\varepsilon \to \infty} \frac{\Delta(\varepsilon)}{\varepsilon} = 0$, we deduce by (64):

$$\frac{dG(\alpha_2)}{d\alpha_2} = -\int_{\Omega(\alpha)} (\frac{d\alpha_1}{d\alpha_2}f_1 + f_2)f_2 \ , \qquad (65)$$

where, for any $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$, $\Omega(\alpha)$ denotes the following subset of $\mathbb{R}^d$,

$$\Omega(\alpha) = \{x \,|\, f(x) - \alpha_1 f_1(x) - \alpha_2 f_2(x) \geq 0\} \ .$$

Similarly, we can prove that

$$\frac{\partial H}{\partial \alpha_1} = -\int_{\Omega(\alpha)} f_1^2 \quad \text{and} \quad \frac{\partial H}{\partial \alpha_2} = -\int_{\Omega(\alpha)} f_1 f_2 \ .$$

Injecting (63) in (65) we finally express the gradient of G as follows

$$\frac{\partial G}{\partial \alpha_2} = \frac{\left(\int_{\Omega(\alpha)} f_1 f_2\right)^2 - \int_{\Omega(\alpha)} f_1^2 \int_{\Omega(\alpha)} f_2^2}{\int_{\Omega(\alpha)} f_1^2} \ . \qquad (66)$$

Thanks to Schwarz inequality, we see that for all α_2, $\frac{\partial G}{\partial \alpha_2} \leq 0$ and G is strictly decreasing as soon as f_1 and f_2 are linearly independent on $\Omega(\alpha)$. □

References

[1] Alexeev, V. and Galeev, E. and Tikhomorov, V. (1987). *Recueil de problèmes d'optimisation.* Edition Mir.

[2] Bartlett, M. S. (1963). *Statistical estimation of density functions.* Sankhya Ser. A **2** 245–254.

[3] Chui, C. K. and Deutsch F. D. and Ward, J. D. (1990). *Constrained Best Approximation in Hilbert Space.* Constructive Approximation. **6** 35–64.

[4] Chen, S. X. (1997). *Empirical Likelihood-based kernel density estimation.* Australian Journal of Statistics. **1 (39)** 47–56.

[5] Devroye, L. and Gyorfi, L (1985). *Nonparametric Density Estimation : The L_1 View.* Wiley series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

[6] Devroye, L. (1987). *A Course on Density Estimation.* Progress in Probability and Statistics. **14**. Birkhuser, Boston.

[7] Glad, I. K. and Hjort, N. L. and Ushakov, N. G. (2003). *Correction of Density Estimators that are not Densities.* Scandinavian Journal of Statistics. **30** 415–427.

[8] Hall, P. and Presnell, B. (1999). *Density Estimation Under Constraints.* Journal of Computational and Graphical Statistics. **8** 259–277.

[9] Hall, P. and Huang, L. S. (2002). *Unimodal density estimation using kernel methods.* Statistica Sinica. **12** 965–990.

[10] Hominal, P. and Deheuvels, P. (1979). *Estimation non paramétrique de densité compte tenu d'informations sur le support.* Revue de statistique appliquée. **27 (3)** 47–68.

[11] Oudjane, N. and Musso, C (2005). L^2*-Density estimation with negative kernels.* Proceedings of ISPA.

[12] Parzen, E. (1962). *On estimation of a probability density and mode.* Annals of Mathematical Statistics. **33** 1065–1070.

[13] Rosenblatt, M. (1956). *Remarks on some non-parametric estimates of a density function.* Annals of Mathematical Statistics. **27** 642–669.

[14] Ruppert, D. and Cline, D. B. H. (1994). *Biais reduction in kernel density estimation by smoothed empirical transformations.* The Annals of Statistics. **22** 185–210.

[15] Serfling, R. J. (1982). *Asymptotic expansion*. Encyclopedia of Statistical Sciences. John Wiley & Sons, New York. 47–68.
[16] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. **26**. Chapman & Hall, London.

FiME
LABORATOIRE COMMUN
DAUPHINE CREST EDF

Laboratoire de Finance des Marchés de l'Énergie

Institut de Finance de Dauphine, Université Paris-Dauphine
1 place du Maréchal de Lattre de Tassigny
75775 PARIS Cedex 16
www.fime-lab.org