

Iterative Spline Algorithms for Estimating a k -monotone Density.

Fadoua Balabdaoui¹ and Jon A. Wellner²

University of Washington

June 23, 2004

Abstract

The *iterative $(2k-1)$ -st spline algorithm* is an extension of the *iterative cubic spline algorithm* developed and used by Groeneboom, Jongbloed and Wellner to compute an approximation of the “invelope” of the integrated two-sided Brownian motion $+t^4$ that is involved in the limiting distribution of the MLE or the LSE of a non-increasing and convex density on $(0, \infty)$ (Groeneboom, Jongbloed and Wellner (2001A, 2001B)). The *iterative $(2k-1)$ -st spline algorithm* was developed to compute the LSE of a k -monotone density on $(0, \infty)$ for any integer $k \geq 2$ and also to calculate an approximation of the envelopes (“invelopes”) of the $(k-1)$ -fold integral of two-sided Brownian motion $+ (k!/(2k!)t^{2k}$ when k is odd (even) on a finite interval $[-c, c]$ for some fixed $c > 0$. To compute the MLE, another variation is used and involves the computation of a spline of degree $k-1$ instead of a spline of degree $2k-1$. The principal of both algorithms is explained in details. We also show many examples of their applications.

¹ Research supported in part by National Science Foundation grant DMS-0203320

² Research supported in part by National Science Foundation grants DMS-0203320, and NIAID grant 2R01 AI291968-04

AMS 2000 subject classifications. Primary: 62G07; secondary 90C99.

Key words and phrases. optimization, cubic iterative spline algorithm, directional derivative, asymptotic distribution, convex, k -monotone, mixture model,

1 Introduction

The *iterative $(2k - 1)$ -st spline algorithm* is an extension of the *iterative cubic spline algorithm*, a term that was coined by GROENEBOOM, JONGBLOED, AND WELLNER (2001A). The latter was used to compute the “invelope” H of two-sided Brownian motion $+ t^4$ that is involved in the limiting distribution of the LSE and MLE of a non-increasing and convex density on $(0, \infty)$ (see GROENEBOOM, JONGBLOED, AND WELLNER (2001A)). The algorithm is described briefly in pages 1643 and 1644 of their article. However, more details about how this algorithm works can be found in GROENEBOOM, JONGBLOED, AND WELLNER (2003). Here, we try to give a full description about how the *iterative spline algorithms* are implemented to compute the LSE and MLE of a k -monotone density on $(0, \infty)$ for an arbitrary integer $k \geq 2$, and also to approximate the envelopes (“invelopes”) of the $(k - 1)$ -fold integral of two-sided Brownian motion $+ (k!/(2k)!) t^{2k}$ when k is odd (even) on a finite interval $[-c, c]$. These algorithms belong to the family of *vertex direction algorithms* (see GROENEBOOM, JONGBLOED, AND WELLNER (2003)). They were around for many decades and their development was motivated by problems in D-optimal design (see FEDOROV (1972), WYNN (1970), BÖHNING (1986)), estimation of random coefficients in regression models (see e.g. MALLETT (1986)), and nonparametric estimation in mixture models (see SIMAR (1976), BÖHNING (1982), LESPERANCE AND KALBFLEISCH (1992), GROENEBOOM, JONGBLOED, AND WELLNER (2003)), which will be the focus here. In mixture models, nonparametric estimation of the mixing distribution or the mixed density yields a constrained, infinite dimensional optimization (e.g. minimization) problem. Thus, an efficient computational method is needed. GROENEBOOM, JONGBLOED, AND WELLNER (2003) extended the algorithm that was implemented by SIMAR (1976) to compute the MLE of a compound (mixed) Poisson distribution. GROENEBOOM, JONGBLOED, AND WELLNER (2003) referred to this extension as the *support reduction algorithm*. The same authors developed and used the *iterative cubic spline algorithm* to compute the LSE of a non-increasing and convex density on $(0, \infty)$ and also to approximate the process H . However, the authors seem to reserve the term only for the second estimation problem.

In the *support reduction algorithms*, the support reduction step is very crucial and it is the only step where it is ensured that one “stays” in the class of functions considered in the optimization problem. In this manuscript, we explain in detail why in our estimation problems, such a step is always possible and we hope that this will shed more light on how the *iterative cubic spline algorithm* works. In the following, we present the general set-up. Let ϕ be a convex functional to be minimized over the class of functions

$$\mathcal{C} = \left\{ g = \int_{\Theta} f_{\theta} d\mu(\theta), \mu \text{ is a positive measure} \right\}.$$

The directional derivative of ϕ at the point g in the direction of f_{θ} is denoted by

$D_\phi(f_\theta, g)$ and defined by

$$D_\phi(f_\theta, g) = \lim_{\epsilon \searrow 0} \frac{\phi(g + \epsilon f_\theta) - \phi(g)}{\epsilon}.$$

Suppose that ϕ admits a unique minimizer, $\operatorname{argmin}_{g \in \mathcal{C}} \phi(g)$. Under the assumptions A1, A2' and A3, GROENEBOOM, JONGBLOED, AND WELLNER (2003) showed that the *support reduction algorithm* converges to $\operatorname{argmin}_{g \in \mathcal{C}} \phi(g)$. In the current estimation problems, these assumptions are satisfied. The manuscript will be organized as follows: In the first two sections, we describe the *iterative $(2k - 1)$ -st spline algorithm* and explain how it works for calculating the LSE of a k -monotone density and for approximating the stochastic process H_k . The last section is reserved for calculating the MLE of a k -monotone density. In this case, the algorithm is different as it involves a linearization step that is not required in the first two estimation problems. However, the algorithm shares with the *iterative $(2k - 1)$ -st spline algorithm* the same basic structure.

Based on two samples of size $n = 100$ and $n = 1000$, the MLE and LSE of the Exponential density, viewed respectively as a k -monotone density with $k = 3$ and $k = 6$, are computed. For the same values of k , approximations of the process H_k and some of its derivatives, on the interval $[-4, 4]$, are calculated.

2 Computing the LSE of a k -monotone density

Let X_1, \dots, X_n be n i.i.d. random variables from a k -monotone density g_0 on $(0, \infty)$ and let \mathbb{G}_n denote their empirical distribution function. We know from BALABDAOUI (2004B), Chapter 2 that the functional

$$\phi(g) = \frac{1}{2} \int_0^\infty g^2(t) dt - \int_0^\infty g(t) d\mathbb{G}_n(t)$$

defined on the space of square integrable k -monotone functions on $(0, \infty)$ admits a unique minimizer \tilde{g}_n . From BALABDAOUI (2004B), Chapter 2, Proposition 2.2.3 we know that \tilde{g}_n is a finite scale mixture of $Beta(1, k)$'s ; i.e., there exist an integer m , $\tilde{\theta}_1, \dots, \tilde{\theta}_m$ and $\tilde{w}_1, \dots, \tilde{w}_m$ such that for all $t > 0$

$$\tilde{g}_n(t) = \tilde{w}_1 \frac{k(\tilde{\theta}_1 - t)_+^{k-1}}{\tilde{\theta}_1^k} + \dots + \tilde{w}_m \frac{k(\tilde{\theta}_m - t)_+^{k-1}}{\tilde{\theta}_m^k}$$

where the weights $\tilde{w}_1, \dots, \tilde{w}_m$ do not necessarily sum up to one for $k > 2$ (see BALABDAOUI (2004A)). The directional derivative of the functional ϕ at a point g in the class

$$\mathcal{C} = \left\{ g : g(t) = \int_0^\infty \frac{k(\theta - t)_+^{k-1}}{\theta^k} d\mu(\theta), \mu \text{ is a positive measure} \right\}$$

in the direction of $f_\theta(t) = \frac{k(\theta-t)_+^{k-1}}{\theta^k}$, $\theta \in \Theta = (0, \infty)$ is given by

$$\begin{aligned} D_\phi(f_\theta, g) &= \int_0^\infty \frac{k(\theta-t)_+^{k-1}}{\theta^k} g(t) dt - \int_0^\infty \frac{k(\theta-t)_+^{k-1}}{\theta^k} d\mathbb{G}_n(t) \\ &= \frac{k}{\theta^k} (H(\theta, g) - \mathbb{Y}_n(\theta)) \end{aligned}$$

where $H(\cdot, g)$ and \mathbb{Y}_n are respectively the k -fold integral of g and $(k-1)$ -fold integral of the empirical distribution function \mathbb{G}_n . When $g = \tilde{g}_n$, then $H(\cdot, g)$ is nothing but \tilde{H}_n defined in BALABDAOUI (2004B), Chapter 2. It follows from the characterization of \tilde{g}_n that $D_\phi(f_\theta, \tilde{g}_n) \geq 0$ for all $\theta \in (0, \infty)$ and equal to zero if and only if θ belongs to the support of the mixing measure $\tilde{\mu}_n$ associated with the LSE \tilde{g}_n . The support reduction algorithm consists of the following steps:

1. Given the current iterate $g \in \mathcal{C}$ with support $S = \{\theta_1, \dots, \theta_p\}$, we find the minimizer of $\theta \mapsto D_\phi(f_\theta, g)$ over $(0, \infty)$. If $D_\phi(f_\theta, g) \geq 0$ for all $\theta \in (0, \infty)$, then we conclude that g is the LSE \tilde{g}_n . Otherwise, we denote the minimizer by θ_{p+1} . Since the rank of θ_{p+1} in the set $\{\theta_1, \dots, \theta_p\}$ is not important for the description of the algorithm, we can assume, without loss of generality, that $\theta_{p+1} \geq \max(S)$. Thus, the new set of support points is $S_{new} = \{\theta_1, \dots, \theta_p, \theta_{p+1}\}$.
2. We find the minimizer of ϕ over the class

$$\left\{ g : g(t) = \sum_{j=1}^{p+1} \sigma_j \frac{k(\theta_j - t)_+^{k-1}}{\theta_j^k}, \quad \sigma_j \in \mathbb{R}, \quad j = 0, \dots, p+1. \right\}.$$

This means that some of the weights $\sigma_1, \dots, \sigma_{p+1}$ can be negative. Let g_{min} denote this minimizer.

3. If all the weights σ_j are nonnegative, then we move to the first step. Otherwise, we need to “go back” to the original class of k -monotone functions and this is ensured by finding a coefficient $\lambda \in (0, 1)$ such that the function $(1-\lambda)g + \lambda g_{min}$ is k -monotone.

We will show that there exists always λ such that $(1-\lambda)g + \lambda g_{min}$ is k -monotone. This operation is actually equivalent to *deleting* one point from the new support S_{new} . We find the minimizer of ϕ over the class of k -monotone functions with the new *reduced* support. This reduction is carried on until the obtained minimizer is a k -monotone function; that is, the weights corresponding to its support points are all nonnegative.

Let $S = \{\theta_1, \dots, \theta_m\}$ be the current set of support points. The following lemma gives the characterization of the minimizer of ϕ in the class of functions g given by

$$g(t) = \sigma_1 \frac{k(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \dots + \sigma_m \frac{k(\theta_m - t)_+^{k-1}}{\theta_m^k}$$

where $0 < \theta_1 < \dots < \theta_m$ and $\sigma_1, \dots, \sigma_m \in \mathbb{R}$. This is also the class of polynomial splines s of degree $k - 1$ that are $(k - 2)$ -times continuously differentiable at the knots $\theta_1, \dots, \theta_m$ and satisfy the boundary conditions $s^{(j)}(\theta_m) = 0$ for $j = 0, \dots, k - 2$ (for a definition of polynomial splines, see e.g. NÜRNBERGER (1989), Definition 1.15, page 94). We denote this class by $\mathcal{C}'(\theta_1, \dots, \theta_m)$.

Lemma 2.1 *A function g is the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_m)$ if and only if g is the k -th derivative of the polynomial spline P of degree $2k - 1$ and knots $\theta_1, \dots, \theta_m$ that satisfies*

$$P(\theta_i) = \mathbb{Y}_n(\theta_i) \quad \text{for } i = 1, \dots, m, \quad (1)$$

$$P^{(j)}(0) = 0 \quad \text{for } j = 0, \dots, k - 1, \quad (2)$$

and

$$P^{(l)}(\theta_m) = 0 \quad \text{for } l = k, \dots, 2k - 2. \quad (3)$$

Proof. Let $\epsilon \in \mathbb{R}$ and suppose that g is the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_m)$. We have for all $j = 1, \dots, m$

$$D_\phi(f_{\theta_j}, g) = \lim_{\epsilon \rightarrow 0} \frac{\phi(g + \epsilon f_{\theta_j}) - \phi(g)}{\epsilon} = 0.$$

Conversely, suppose that $g \in \mathcal{C}'(\theta_1, \dots, \theta_m)$ satisfies $D_\phi(f_{\theta_j}, g) = 0$ for all $j = 1, \dots, m$. Let h be any arbitrary function in $\mathcal{C}(\theta_1, \dots, \theta_m)$. By convexity of ϕ , we have

$$\begin{aligned} \phi(h) - \phi(g) &\geq D_\phi(h - g, g) \\ &= D_\phi \left(\sum_{j=1}^m (\sigma_{j,h} - \sigma_{j,g}) f_{\theta_j}, g \right) \\ &= \sum_{j=1}^m (\sigma_{j,h} - \sigma_{j,g}) D(f_{\theta_j}, g) \\ &= 0 \end{aligned}$$

which implies that g is the minimizer.

Now, notice that $D_\phi(f_{\theta_j}, g) = 0$, $j = 1, \dots, m$, is equivalent to

$$H(\theta_j, g) = \mathbb{Y}_n(\theta_j), \quad j = 1, \dots, m,$$

where

$$H(\theta, g) = \int_0^\theta (\theta - t)^{k-1} g(t) dt.$$

By noticing that $H(\cdot, g)$ is a spline of degree $2k - 1$ and knots $\theta_1, \dots, \theta_m$ and satisfying the boundary conditions in (1, 2 and 3), the results follows. \blacksquare

The following lemma ensures that the *reduction step* is always possible.

Lemma 2.2 *Let $\{\theta_1, \dots, \theta_{m-1}\}$ be the set of support points of the current iterate g . Let $\theta_m = \operatorname{argmin}_{\theta \in (0, \infty)} D(f_\theta, g)$ and suppose without loss of generality that $\theta_m > \theta_{m-1}$. Let g_{\min} be the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_m)$. If g_{\min} is not k -monotone, then there exists $\lambda \in (0, 1)$ such that the function*

$$(1 - \lambda)g + \lambda g_{\min}$$

is k -monotone.

Proof. Since g_{\min} minimizes ϕ over a bigger class, it follows that

$$\phi(g_{\min}) < \phi(g).$$

The last inequality is strict because $g_{\min} \neq g$. Using convexity of ϕ , we can write for any $\epsilon > 0$,

$$\begin{aligned} \phi((1 - \epsilon)g + \epsilon g_{\min}) - \phi(g) &\leq (1 - \epsilon)\phi(g) + \epsilon\phi(g_{\min}) - \phi(g) \\ &= \epsilon(\phi(g_{\min}) - \phi(g)) \\ &< 0. \end{aligned}$$

Now, there exist $\sigma_{1,g}, \dots, \sigma_{m-1,g}$ such that $\sigma_{j,g} \geq 0$ for $j = 1, \dots, m - 1$ and $\sigma_{1,g_{\min}}, \dots, \sigma_{m,g_{\min}} \in \mathbb{R}$ such that g and g_{\min} can be written as

$$g(t) = \sigma_{1,g} k \frac{(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \dots + \sigma_{m-1,g} k \frac{(\theta_{m-1} - t)_+^{k-1}}{\theta_{m-1}^k}$$

and

$$g(t) = \sigma_{1,g_{\min}} k \frac{(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \dots + \sigma_{m,g_{\min}} k \frac{(\theta_m - t)_+^{k-1}}{\theta_m^k}.$$

By passing ϵ to the limit, we obtain

$$\begin{aligned} \lim_{\epsilon \searrow 0} \frac{\phi((1 - \epsilon)g + \epsilon g_{\min}) - \phi(g)}{\epsilon} &= D_\phi(g_{\min} - g, g) \\ &= \sigma_{m,g_{\min}} D_\phi(f_{\theta_m}, g) + \sum_{j=1}^{m-1} (\sigma_{j,g_{\min}} - \sigma_{j,g}) D_\phi(f_{\theta_j}, g) \\ &= \sigma_{m,g_{\min}} D_\phi(f_{\theta_m}, g) \end{aligned}$$

where in the last equality we used the fact that $D(f_{\theta_j}, g) = 0$ for $j = 1, \dots, m - 1$. Since by definition of θ_m , $D_\phi(f_{\theta_m}, g) < 0$ it follows that $\sigma_{m, g_{min}} > 0$. Let λ be in $[0, 1]$ and consider g_λ the weighted sum of g and g_{min} :

$$g_\lambda = (1 - \lambda)g + \lambda g_{min}.$$

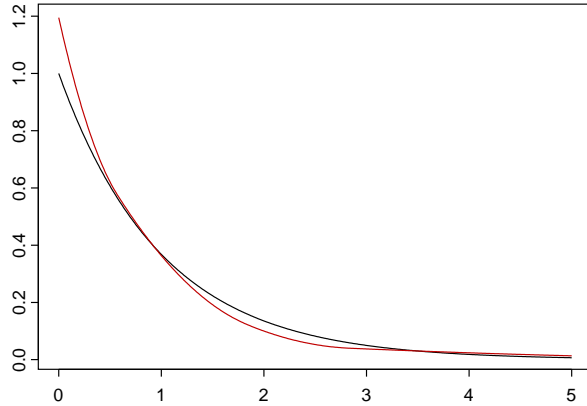


Figure 1: The exponential density (in black) and the Least Squares estimator of the (mixed) k -monotone density based on $n = 100$ and $k = 3$ (in red).

We want to find the largest λ such that g_λ is k -monotone. The parameter λ has to be chosen such that

$$\begin{aligned} (1 - \lambda)\sigma_{1,g} + \lambda\sigma_{1,g_{min}} &\geq 0 \\ &\vdots \\ (1 - \lambda)\sigma_{m-1,g} + \lambda\sigma_{m-1,g_{min}} &\geq 0 \\ (1 - \lambda)\sigma_{m,g} + \lambda\sigma_{m,g_{min}} &\geq 0. \end{aligned}$$

Note that the last inequality is automatically satisfied since $\sigma_{m, g_{min}} > 0$ and hence we only need to worry about the first $m - 1$ inequalities (it is implicitly assumed that $m \geq 2$). Let J be the set of integers $j \in \{1, \dots, m - 1\}$ such that

$$\sigma_{j, g_{min}} < 0.$$

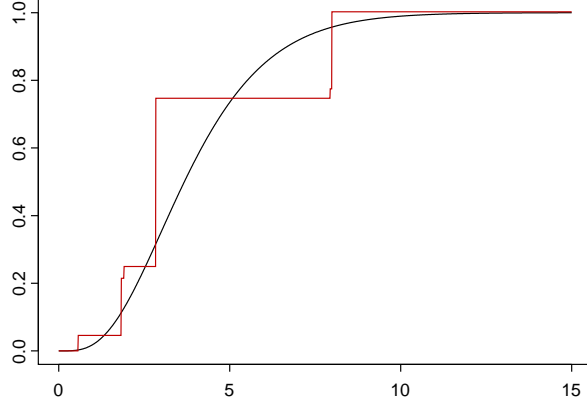


Figure 2: The cumulative distribution function of a $Gamma(4, 1)$ (in black) and the Least squares estimator of the mixing distribution based on $n = 100$ and $k = 3$ (in red).

For $j \in J$, define λ_j by

$$\lambda_j = \frac{\sigma_{j,g}}{\sigma_{j,g} - \sigma_{j,gmin}}.$$

Clearly, $\lambda_j \in (0, 1)$. Now, if we consider j_0 to be the index of the smallest λ_j ; i.e.,

$$j_0 = \operatorname{argmin}_{j \in J} \lambda_j,$$

then it is easy to verify that for all $j \in J$

$$(1 - \lambda_{j_0})\sigma_{j,g} + \lambda_{j_0}\sigma_{j,gmin} \geq 0$$

with equality if and only if $j = j_0$ (we assume here that j_0 is unique). To see that, notice that if $\lambda \in (0, 1)$ satisfies

$$(1 - \lambda)\sigma_{j,g} + \lambda\sigma_{j,gmin} \geq 0, \quad \text{for all } j \in J \tag{4}$$

then

$$\lambda \leq \lambda_j, \quad \text{for all } j \in J.$$

It follows that $\lambda \leq \min_{j \in J} \lambda_j = \lambda_{j_0}$ and that the maximal value of $\lambda \in (0, 1)$ satisfying the inequality in (4) is equal to λ_{j_0} .

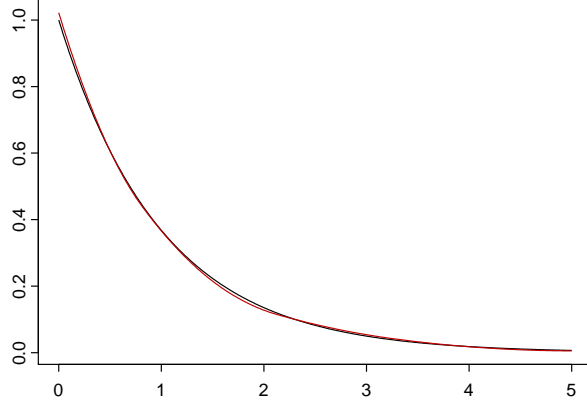


Figure 3: The exponential density (the true mixed density), in black and its Least Squares estimator based on $n = 1000$ and $k = 3$, in red.

Since $(1 - \lambda_{j_0})\sigma_{j_0, g} + \lambda_{j_0}\sigma_{j_0, g_{min}} = 0$, the knot θ_{j_0} is deleted from the set of knots $S = \{\theta_1, \dots, \theta_m\}$. The next step is to compute the $(2k - 1)$ -st spline with the new set of knots $S \setminus \{\theta_{j_0}\}$. Notice that by moving from the previous step to the new one, the monotonicity of the algorithm is maintained. Indeed, using again the convexity of ϕ , we have

$$\begin{aligned}
 \phi(g_{\lambda_{j_0}}) &= \phi((1 - \lambda_{j_0})g + \lambda_{j_0}g_{min}) \\
 &\leq (1 - \lambda_{j_0})\phi(g) + \lambda_{j_0}\phi(g_{min}) \\
 &< (1 - \lambda_{j_0})\phi(g_{min}) + \lambda_{j_0}\phi(g_{min}) \\
 &= \phi(g_{min}).
 \end{aligned}$$

Therefore, if g_{j_0} is the minimizer of ϕ over the class of functions $\mathcal{C}(S \setminus \{\theta_{j_0}\})$, we should have

$$\phi(g_{j_0}) \leq \phi(g_{\lambda_{j_0}})$$

which implies that $\phi(g_{j_0}) < \phi(g_{min})$. ■

To start the algorithm, we fix some initial value $\theta^{(0)} > X_{(n)}$ and minimize the functional ϕ over the cone

$$\mathcal{C}^{(0)} = \left\{ g : g(t) = C \frac{k(\theta^{(0)} - t)^{k-1}}{(\theta^{(0)})^k}, C > 0 \right\}.$$

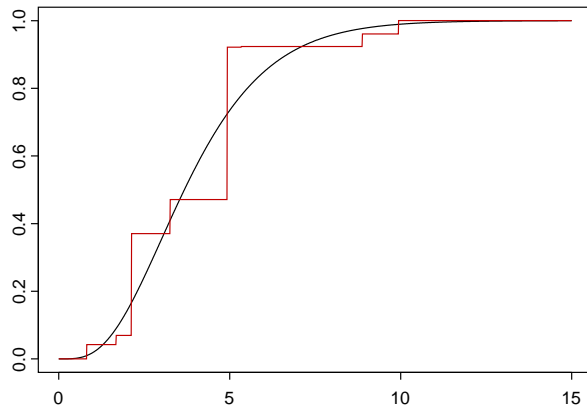


Figure 4: The cumulative distribution function of a $Gamma(4,1)$ (the true mixing distribution), in black and the Least Squares estimator of the mixing distribution based on $n = 1000$ and $k = 3$, in red.

For this purpose, we need to find the value $C^{(0)}$ that minimizes the quadratic function

$$C \mapsto \frac{k^2}{2(2k-1)\theta^{(0)}} C^2 - \frac{1}{n} \sum_{j=1}^n k \frac{(\theta^{(0)} - X_{(j)})^{k-1}}{(\theta^{(0)})^k} C$$

which yields

$$C^{(0)} = \left(\frac{2k-1}{k} \right) \frac{1}{n} \sum_{j=1}^n \frac{(\theta^{(0)} - X_{(j)})^{k-1}}{(\theta^{(0)})^{k-2}}.$$

As in GROENEBOOM, JONGBLOED, AND WELLNER (2003), we used an “alternative” directional derivative. Using their notation, the “usual” directional derivative at a point g in the direction of f_θ , denoted before by $D_\phi(f_\theta, g)$, is equal to $c_1(\theta)$, where

$$\phi(g + \epsilon f_\theta) = \phi(g) + \epsilon c_1(\theta) + \frac{\epsilon^2}{2} c_2(\theta)$$

with

$$c_2(\theta) = \int_0^\infty f_\theta^2(t) dt = \frac{k^2}{(2k-1)\theta}.$$

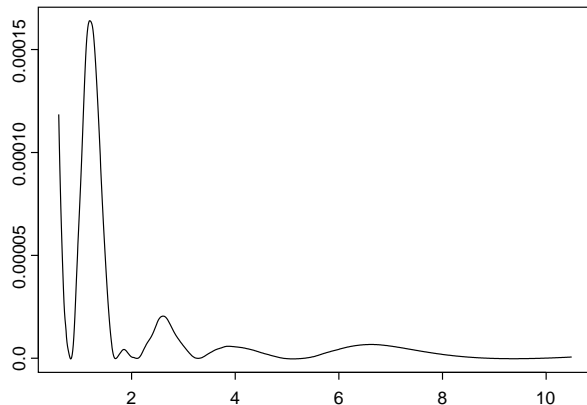


Figure 5: The directional derivative for the Least Squares estimator of the Exponential density based on $n = 1000$ and $k = 3$.

The “alternative” directional derivative is given by

$$\tilde{D}_\phi(f_\theta, g) = \frac{D_\phi(f_\theta, g)}{\sqrt{c_2(\theta)}} = k \frac{H(\theta, g) - \mathbb{Y}_n(\theta)}{\theta^{k-1/2}}.$$

Remark 2.1 *It should be mentioned here that the “gridless” step that was implemented by GROENEBOOM, JONGBLOED, AND WELLNER (2003) was not considered here. In practice, we only consider a finite grid over which we minimize the directional derivative. The obtained LSE is the minimizer of ϕ over the class of k -monotone functions whose support points belong to the finite grid. The purpose of the “gridless” implementation is to obtain a numerical solution that is closest to the theoretical one by perturbing the support points of the solution. By performing this fine tuning, one can run the algorithm once again considering the new grid and obtain a new minimizer. This step is repeated until the gradient of the functional ϕ is sufficiently small.*

Now we describe the preliminary simulations that we have performed. From a standard Exponential, we simulated two samples of respective sizes $n = 100$ and $n = 1000$. The Exponential density is completely monotone and therefore is k -monotone for all integers $k \geq 1$. This is actually the motivation behind considering nonparametric

estimation of k -monotone densities (see BALABDAOUI (2004B), Chapter 1 for more details). The code of the algorithm was written in S and can be found in, BALABDAOUI (2004B), Appendix C. To illustrate the asymptotic distribution theory developed in BALABDAOUI (2004B), Chapter 2 for any integer $k \geq 2$, we computed the LSE based on $n = 100$ and $n = 1000$ in two different cases: $k = 3$ and $k = 6$.

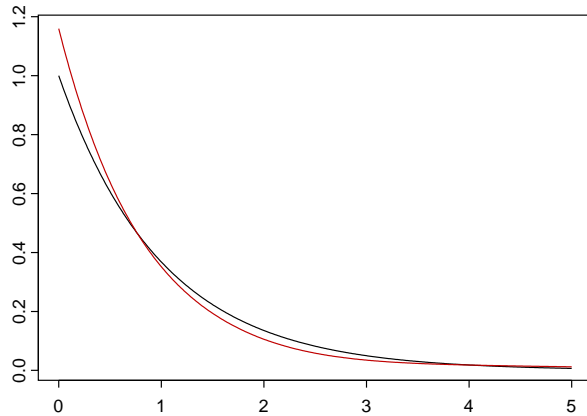


Figure 6: The exponential density (the true mixed density), in black and its Least Squares estimator based on $n = 100$ and $k = 6$, in red.

Note that if θ is a support point of the minimizing measure, then $\theta > X_{(1)}$. This follows from the simple fact that for all $\theta \in (0, X_{(1)})$, $(\theta - X_{(j)})_+^{k-1} = 0$ for $j = 1, \dots, n$. Therefore, adding $\theta \in (0, X_{(1)})$ to the set of support points does not effect the value of the sum $n^{-1} \sum_{j=1}^n g(X_j)$ whereas it increases the value of the integral $\int_0^\infty g^2(t)dt$. The minimization was performed on a finite grid such that, for given n and k , the maximal distance between its points is taken to be 10^{-2} . In practice, we found that it is enough to take $2kX_{(n)}$ as an upper bound for the largest support point as we obtained similar results with larger bounds. The obtained estimates can be found in Table 1.

For $k = 3$, the plots in Figure 1 and Figure 3 show the LSE of the Exponential density based on $n = 100$ and $n = 1000$ respectively. The “alternative” directional derivative $\tilde{D}_\phi(f_\theta, \tilde{g}_n)$, for $n = 1000$, is plotted in Figure 5. In the inverse problem, plots of the LSE of the true mixing distribution are shown in Figure 2 and Figure 4. In general, the true mixing distribution that corresponds to a standard Exponential

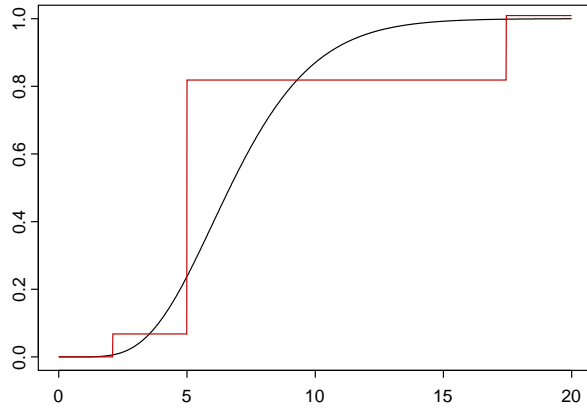


Figure 7: The cumulative distribution function of a $Gamma(7,1)$ (the true mixing distribution), in black and its Least squares estimator based on $n = 100$ and $k = 6$, in red.

when viewed as a k -monotone density is a $Gamma(k + 1, 1)$. Indeed, note that

$$\int_x^\infty \frac{1}{\Gamma(k)} (t-x)^{k-1} e^{-(t-x)} dt = 1$$

for all $x > 0$. It follows that,

$$\begin{aligned} \exp(-x) &= \int_x^\infty \frac{(t-x)^{k-1}}{(k-1)!} e^{-t} dt \\ &= \int_0^\infty \frac{(t-x)_+^{k-1}}{(k-1)!} e^{-t} dt \\ &= \int_0^\infty k \frac{(t-x)_+^{k-1}}{t^k} \frac{1}{k!} t^k e^{-t} dt \\ &= \int_0^\infty k \frac{(t-x)_+^{k-1}}{t^k} f_k(t) dt \end{aligned} \tag{5}$$

where f_k is the $Gamma(k + 1, 1)$ density.

For $k = 6$, similar plots were produced for $n = 100$ and $n = 1000$: for the direct problem, see Figure 6 and Figure 8, and for the inverse one, see Figure 7 and Figure 9.

Table 1: Table of the obtained LS estimates for $k = 3, 6$ and $n = 100, 1000$ and the corresponding numbers of iterations N_{it} . A support point is denoted by \tilde{a} and its mass by \tilde{w} .

k, n	N_{it}	(\tilde{a}, \tilde{w})
$k = 3, n = 100$	13	(0.569, 0.0459), (1.829, 0.168), (1.909, 0.0347), (2.839, 0.497), (7.939, 0.027), (7.989, 0.227)
$k = 3, n = 1000$	14	(0.814, 0.042), (1.674, 0.027), (2.124, 0.300), (3.254, 0.100), (4.924, 0.450), (5.334, 0.001), (8.874, 0.037), (9.934, 0.039)
$k = 6, n = 100$	4	(2.109, 0.067), (4.999, 0.750), (17.449, 0.190)
$k = 6, n = 1000$	6	(2.625, 0.017), (3.615, 0.478), (6.575, 0.478), (11.375, 0.262)

The figures show consistency of the LSE and it is clear that convergence for estimating the Exponential density is much faster than for estimating the Gamma distribution. This is expected since in the direct problem, the rate of convergence is $n^{-k/(2k+1)}$ whereas it is equal to $n^{-1/(2k+1)}$ in the inverse problem. Note also the rate $n^{-1/(2k+1)}$ is slower for larger k and therefore, one should expect to see fewer support points as $k \rightarrow \infty$. This fact is confirmed in the numerical examples above (for $n = 1000$, there are 8 support points for $k = 3$ and 4 for $k = 6$, see Table 1) and in many other simulations that we performed.

3 Approximation of the process H_k on $[-c, c]$

We will focus here on the case when k is even. When k is odd, the steps are very similar. The goal of the algorithm is to find the minimizer of the functional

$$\phi(g) = \frac{1}{2} \int_{-c}^c g^2(t) dt - \int_{-c}^c g(t) dX_k(t)$$

where

$$dX_k(t) = dW(t) + t^k dt$$

and W is two-sided Brownian motion starting at 0, over \mathcal{C} the class of functions g that are k -convex; i.e. $g^{(k-2)}$ exists and is convex, and satisfies the boundary conditions

$$\left(g^{(k-2)}(\pm c), \dots, g^{(2)}(\pm c), g(\pm c) \right) = \left(\frac{k!}{2!} c^2, \dots, \frac{k!}{(k-2)!} c^{k-2}, c^k \right). \quad (6)$$

Recall that if $H_{c,k}$ is the k -fold integral of $g_{c,k}$ determined by

$$H_{c,k}(c) = Y_k(c), H_{c,k}^{(2)}(c) = Y_k^{(2)}(c), \dots, H_{c,k}^{(k-2)}(c) = Y_k^{(k-2)}(c), \quad (7)$$

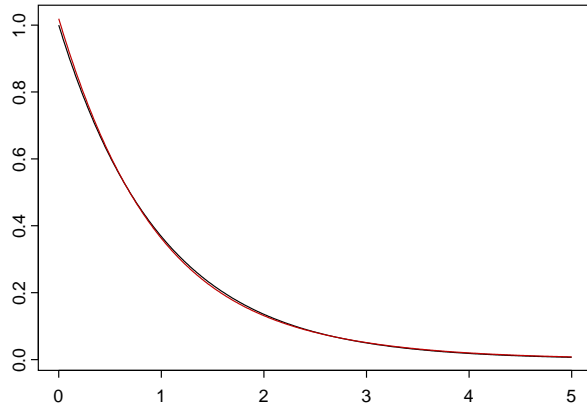


Figure 8: The exponential density (the true mixed density), in black and its Least Squares estimator based on $n = 1000$ and $k = 6$, in red.

then $g_{c,k}$ is the minimizer if and only if

$$H_{c,k}(t) \geq Y_k(t), \quad t \in [-c, c]$$

and

$$\int_{-c}^c (H_{c,k}(t) - Y_k(t)) dg_{c,k}^{(k-1)}(t) = 0,$$

where

$$Y_k(t) \stackrel{d}{=} \begin{cases} \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{k!}{(2k)!} t^{2k}, & t \geq 0 \\ \int_t^0 \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{k!}{(2k)!} t^{2k}, & t < 0 \end{cases}$$

The above characterization gives a necessary and sufficient condition for a function g in the considered class to be the solution for the minimization problem. But it also implies that this solution cannot have a strictly increasing $(k-1)$ -st derivative on a set with nontrivial interior. Indeed, if we assume that there exists an open interval $I \subseteq (-c, c)$ of positive length on which $g_{c,k}^{(k-1)}$ is strictly increasing, then this would imply that $Y_k = H_{c,k}$ on I and that the $(k-1)$ -fold integral of Brownian motion is in $C^{2k-2}(I)$. Therefore, the function $g_{c,k}^{(k-1)}$ has to increase on a set of Lebesgue measure zero. We conjecture that this set is finite and consists of the discontinuity points of the monotone

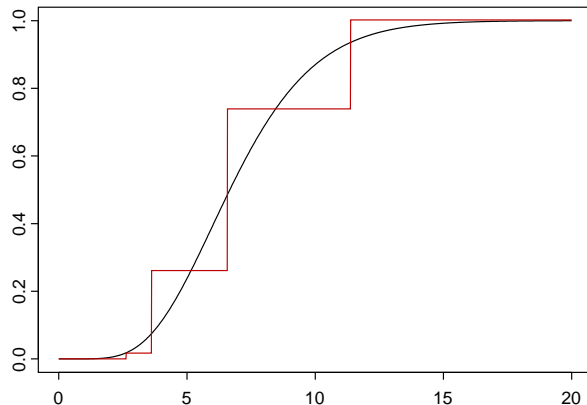


Figure 9: The cumulative distribution function of a $Gamma(7,1)$ (the true mixing distribution), in black and its Least squares estimator based on $n = 1000$ and $k = 6$, in red.

function $g_{c,k}^{(k-1)}$. For the particular case of $k = 2$, there is still no proof available for this conjecture (see GROENEBOOM, JONGBLOED, AND WELLNER (2001A), Section 4). The main difficulty of this problem lies in the fact that in principle, the monotone function $g_{c,k}^{(k-1)}$ could be a Cantor-type function in which case, the set on which it increases is Lebesgue measure zero and is uncountable (see e.g. GELBAUM AND OLMSTED (1964), example 15, page 96). Based on this conjecture, $H_{c,k}$ is a spline of degree $2k - 1$ that stays above Y_k and touches it at the discontinuity points of $g_{c,k}^{(k-1)}$; i.e., those points where $H_{c,k}^{(2k-2)} = g_{c,k}^{(k-2)}$ changes its slope. Therefore, in order to obtain the solution $g_{c,k}$ and its derivatives $g'_{c,k}, \dots, g_{c,k}^{(k-1)}$, we first find $H_{c,k}$ and then differentiate it $(k + j)$ -times for $j = 0, \dots, k - 1$.

The steps of the *support reduction algorithm* are very similar to those described in the previous section on calculating the LSE of a k -monotone density. In view of the conjecture, we can restrict ourselves to the class of functions

$$\mathcal{C} = \left\{ g : g(t) = \sum_{j=0}^{k-1} \lambda_j \frac{t^j}{j!} + \mu_1 (t - \theta_1)_+^{k-1} + \dots + \mu_p (t - \theta_p)_+^{k-1}, p \in \mathbb{N} \setminus \{0\} \right\}$$

where $\lambda_j \in \mathbb{R}$, $\mu_j \geq 0$ for $1 \leq j \leq p$ such that g satisfies the constraints in (6). Note that any element $g \in \mathcal{C}$ is a spline of degree $k - 1$ and simple knots $\theta_1, \dots, \theta_p$. This

means that g is $(k - 2)$ -times continuously differentiable at these knots. From each iterate $g \in \mathcal{C}$, we can move in the direction of the function

$$f_\theta(t) = \frac{(t - \theta)_+^{k-1}}{(k-1)!} + \alpha_{k-1}(\theta) \frac{(t+c)^{k-1}}{(k-1)!} + \alpha_{k-3}(\theta) + \frac{(t+c)^{k-3}}{(k-3)!} + \cdots + \alpha_1(\theta)(u+c)$$

where

$$\begin{aligned} \alpha_{k-1}(\theta) &= -\frac{(c-\theta)}{2c} \\ \alpha_{k-3}(\theta) &= -\alpha_{k-1}(\theta) \frac{(2c)^3}{3!} - \frac{(c-\theta)^3}{3!} \\ &\vdots \\ \alpha_1(\theta) &= -\alpha_{k-1}(\theta) \frac{(2c)^{k-1}}{(k-1)!} - \cdots - \alpha_3(\theta) \frac{(2c)^3}{3!} - \frac{(c-\theta)^{k-1}}{(k-1)!}. \end{aligned}$$

Indeed, for all $\theta \in [-c, c]$, the function f_θ is a spline of degree $k - 1$ with θ as its unique simple knot. Moreover, f_θ satisfies the boundary conditions

$$f_\theta^{(2j)}(\pm c) = 0, \quad \text{for } j = 0, \dots, (k-2)/2. \quad (8)$$

For an arbitrary $\epsilon > 0$, the function $g + \epsilon f_\theta$ belongs to the class \mathcal{C} and the directional derivative of ϕ at g in the direction of f_θ is given by

$$D_\phi(g, f_\theta) = H(\theta, g) - Y_k(\theta) \quad (9)$$

where $H(\cdot, g)$ is the k -fold integral of g determined by the boundary conditions

$$H^{(2j)}(\pm c, g) = Y_k^{(2j)}(\pm c), \quad \text{for } j = 0, \dots, (k-2)/2. \quad (10)$$

To see the equality in (9), note first that $D(f_\theta, g)$ is given by

$$\begin{aligned} D(f_\theta, g) &= \int_{-c}^c f_\theta(t)g(t)dt - \int_{-c}^c f_\theta(t)dX_k(t) \\ &= \int_{-c}^c f_\theta(t)d(H^{(k-1)}(t, g) - Y_k^{(k-1)}(t)) \end{aligned}$$

Thus, using successive integration by parts and the boundary conditions in (8) and (10), we can write

$$\begin{aligned} D_\phi(g, f_\theta) &= \left[\left(H^{(k-1)}(t, g) - Y_k^{(k-1)}(t) \right) f_\theta(t) \right]_{-c}^c - \int_{-c}^c \left(H^{(k-1)}(t, g) - Y_k^{(k-1)}(t) \right) f'_\theta(t)dt \\ &= - \int_{-c}^c \left(H^{(k-1)}(t, g) - Y_k^{(k-1)}(t) \right) f'_\theta(t)dt \end{aligned}$$

$$\begin{aligned}
&= - \left[\left(H^{(k-2)}(t, g) - Y_k^{(k-2)}(t) \right) f'_\theta(t) \right]_{-c}^c + \int_{-c}^c \left(H^{(k-2)}(t, g) - Y_k^{(k-2)}(t) \right) f''_\theta(t) dt \\
&= \int_{-c}^c \left(H^{(k-2)}(t, g) - Y_k^{(k-2)}(t) \right) f''_\theta(t) dt \\
&\quad \vdots \\
&= \int_{-c}^c \left(H(t, g) - Y_k(t) \right) f_\theta^{(k-1)}(t) dt \\
&= H(\theta, g) - Y_k(\theta).
\end{aligned}$$

Note that Y_k plays here a role that is similar to that of the process \mathbb{Y}_n . Let $S = \{\theta_1, \dots, \theta_m\}$ be the set of knots of the current iterate g . The function $H(\cdot, g)$ is a spline of degree $2k - 1$ with simple knots $-c, \theta_1, \dots, \theta_m, c$. If $H(\cdot, g) \geq Y_k$, then $g = H^{(k)}(\cdot, g)$ is the solution of the minimization problem. Otherwise, we add $\theta_{m+1} = \operatorname{argmin}_{\theta \in [-c, c]} (H(\cdot, g)(\theta) - Y_k(\theta))$ to the support S . Without loss of generality, we can assume that $\theta_1 < \dots < \theta_m < \theta_{m+1}$. Now, let $\mathcal{C}'(\theta_1, \dots, \theta_{m+1})$ be the class of polynomial splines of degree $k - 1$, with simple knots $\theta_1, \dots, \theta_{m+1}$ satisfying the boundary conditions in (6); i.e.,

$$\mathcal{C}'(\theta_1, \dots, \theta_{m+1}) = \left\{ g : g(t) = \sum_{j=0}^{k-1} \lambda_j \frac{t^j}{j!} + \sigma_1 (t - \theta_1)_+^{k-1} + \dots + \sigma_{m+1} (t - \theta_{m+1})_+^{k-1} \right\}$$

where $\sigma_j \in \mathbb{R}$ and the λ_j 's are different from the ones used in the definition of the class \mathcal{C} . Consider H_{min} to be the spline of degree $2k - 1$ and simple knots $\theta_1, \dots, \theta_{m+1}$ satisfying

$$H_{min}(\theta_j) = Y_k(\theta_j), \quad \text{for } j = 1, \dots, m + 1.$$

$$H_{min}^{(2j)}(\pm c) = Y_k^{(2j)}(\pm c), \quad \text{for } j = 0, \dots, (k - 2)/2$$

and

$$H_{min}^{(2j)}(\pm c) = \frac{k!}{(2k - 2j)!} c^{2k-2j}, \quad \text{for } j = k, \dots, (2k - 2)/2.$$

The following lemma gives the solution of minimizing ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_{m+1})$.

Lemma 3.1 *Let H_{min} be the spline defined above. The function $g_{min} = H_{min}^{(k)}$ is the minimizer of the functional ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_{m+1})$.*

Proof. The arguments are very similar to those used in the proof of Lemma 2.2. ■

There exist $\lambda_0, \dots, \lambda_{2k-1}$, and $\sigma_1, \dots, \sigma_{m+1}$ such that the spline H_{min} can be written as

$$H_{min} = H(t, g_{min}) = \sum_{j=0}^{2k-1} \lambda_j \frac{t^j}{j!} + \sigma_1 (t - \theta_1)_+^{2k-1} + \dots + \sigma_{m+1} (t - \theta_{m+1})_+^{2k-1}.$$

To find the parameters $\lambda_{2k-1}, \dots, \lambda_1, \lambda_0$ and $\sigma_1, \dots, \sigma_{m+1}$, we solve a linear system of dimension $(2k + m + 1) \times (2k + m + 1)$ using the $2k + m + 1$ boundary conditions satisfied by H_{min} .

The reduction step is given by the following lemma:

Lemma 3.2 *Let g be the current iterate in \mathcal{C} with knots $\theta_1, \dots, \theta_m$ and $g_{min} = H_{min}^{(k)}$ be new minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_{m+1})$. If g_{min} is not in the class \mathcal{C}' , then there exists $\lambda \in (0, 1)$ such that $(1 - \lambda)g + \lambda g_{min} \in \mathcal{C}'$.*

Proof. The arguments are very similar to those used in the proof of Lemma 2.2. ■

The steps of the algorithm can be summarized as follows:

1. Given the current iterate g with set of simple knots $S = \{\theta_1, \dots, \theta_m\}$, we calculate $\text{argmin}_{\theta \in [-c, c]} D_\theta(f_\theta, g) = \text{argmin}_{\theta \in [-c, c]} (H(\theta, g) - Y_k(\theta))$. If $D_\theta(f_\theta, g) \geq 0$ for all $\theta \in [-c, c]$, then g is the minimizer of ϕ over the class of splines \mathcal{C} and its k -fold integral $H(\cdot, g)$ is an approximation of the process H_k . Otherwise, we denote $\theta_{m+1} = \text{argmin}_{\theta \in [-c, c]} (H(\theta, g) - Y_k(\theta))$. If we assume without loss of generality that $\theta_{m+1} > \theta_m$, then $S_{new} = \{\theta_1, \dots, \theta_m, \theta_{m+1}\}$ is the new set of knots.
2. We find g_{min} the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \dots, \theta_{m+1})$.
3. If $g_{min} \in \mathcal{C}$, we move the Step 1. Otherwise, we find the maximal value of $\lambda \in (0, 1)$ such that $(1 - \lambda)g + \lambda g_{min} \in \mathcal{C}$. By finding such a λ , a point θ_j for some $j \in \{1, \dots, m\}$ will be deleted from the current support. We find the minimizer over $\mathcal{C}'(S_{new} \setminus \{\theta_j\})$. This will be repeated until the minimizer is in the class \mathcal{C} .

The algorithm has to start somewhere and the most natural starting spline is the polynomial $H_{c,k}^{(0)}$ that was used in BALABDAOUI (2004B), Chapter 3 to prove that $H_{c,k}$ and Y_k have at least a point of touch with probability converging to 1 as $c \rightarrow \infty$. Recall that $H_{c,k}^{(0)}$ is the unique polynomial P of degree $2k - 2$ that satisfies (6) and (7). To be conform with the notation used in BALABDAOUI (2004B), Chapter 2 we write the polynomial $H_{c,k}^{(0)}(t)$ as

$$\begin{aligned} H_{c,k}^{(0)}(t) &= \frac{\alpha_{2k-2}}{(2k-2)!} t^{2k-2} + \frac{\alpha_{2k-4}}{(2k-2)!} t^{2k-2} + \dots + \frac{\alpha_k}{k!} t^k + \frac{\alpha_{k-1}}{(k-1)!} t^{k-1} \\ &\quad + \frac{\alpha_{k-2}}{(k-2)!} t^{k-2} \dots + \alpha_0, \end{aligned}$$

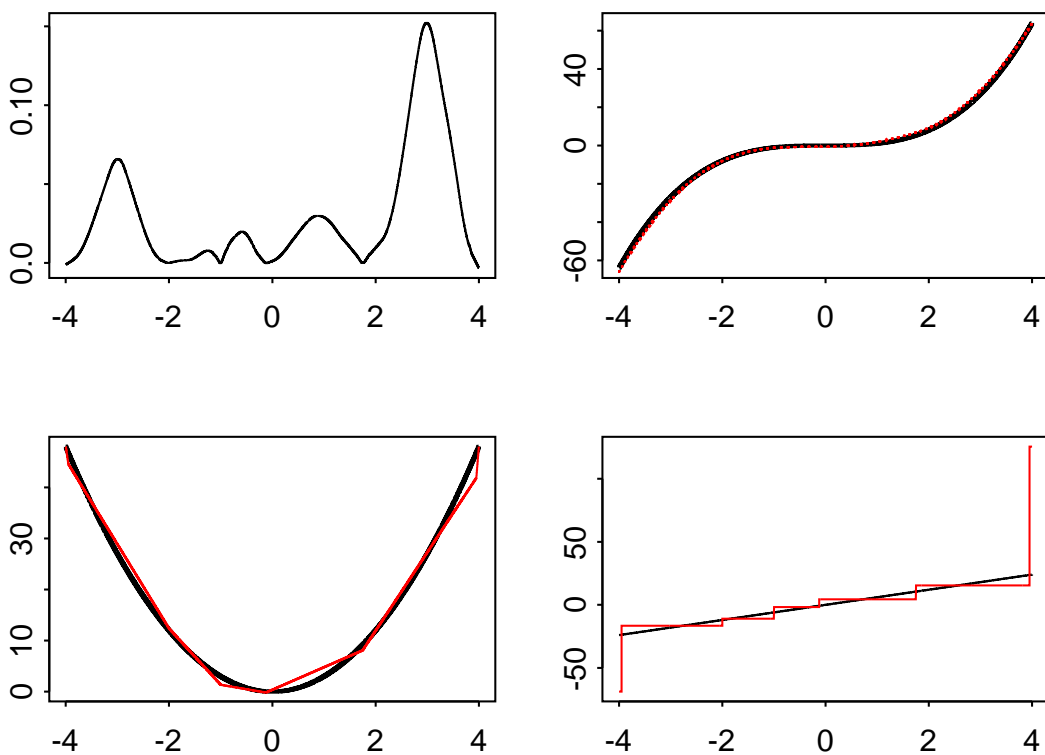


Figure 10: Plots of $-(H_{4,3} - Y_3)$, $g_{4,3} = H_{4,3}^{(3)}$ the LS solution (dashed red line) and t^3 (solid black line), $g'_{4,3} = H_{4,3}^{(4)}$ (solid red line) and $3t^2$ (solid black line), and $g''_{4,3} = H_{4,3}^{(5)}$ (solid red line) and $6t$ (solid black line).

where $\alpha_{2k-2}, \dots, \alpha_k$ are given by

$$\alpha_{2k-2} = \frac{k!}{2!}c^2,$$

$$\alpha_{2k-2j} = \frac{k!}{(2j)!}c^{2j} - \left(\frac{\alpha_{2k-2}}{(2j-2)!}c^{2j-2} + \dots + \frac{\alpha_{2k-2j+2}}{2!}c^2 \right)$$

for $j = 2, \dots, k/2$, whereas $\alpha_{k-1}, \alpha_{k-2}, \dots, \alpha_0$ are given by

$$\alpha_{k-1} = \frac{Y_k^{(k-2)}(c) - Y_k^{(k-2)}(-c)}{2c},$$

$$\alpha_{k-2} = \frac{Y_k^{(k-2)}(-c) + Y_k^{(k-2)}(c)}{2} - \left(\frac{\alpha_{2k-2}}{k!}c^k + \dots + \frac{\alpha_k}{2!}c^2 \right),$$

$$\alpha_{k-2j-1} = \frac{Y_k^{(k-2j-2)}(c) - Y_k^{(k-2j-2)}(-c)}{2c} - \left(\frac{\alpha_{k-1}}{(2j+1)!}c^{2j} + \dots + \frac{\alpha_{k-2j+1}}{3!}c^2 \right),$$

and

$$\alpha_{k-2j-2} = \frac{Y_k^{(k-2j-2)}(c) + Y_k^{(k-2j-2)}(-c)}{2} - \left(\frac{\alpha_{2k-2}}{(k+2j)!}c^{k+2j} + \dots + \frac{\alpha_{k-2j}}{2!}c^2 \right)$$

for $j = 1, \dots, (k-2)/2$.

Example 3.1 For $k = 2$, $H_{c,2}^{(0)}$ is given by

$$H_{c,2}^{(0)}(t) = \frac{\alpha_2}{2!}t^2 + \alpha_1 t + \alpha_0, \quad t \in [-c, c]$$

with

$$\alpha_2 = c^2, \quad \alpha_1 = \frac{Y_2(c) - Y_2(-c)}{2c}, \quad \alpha_0 = \frac{Y_2(-c) + Y_2(c)}{2} - c^2.$$

Example 3.2 For $k = 4$, $H_{c,4}^{(0)}$ is given by

$$H_{c,4}^{(0)}(t) = \frac{\alpha_6}{6!}t^6 + \frac{\alpha_4}{4!}t^4 + \frac{\alpha_3}{3!}t^3 + \frac{\alpha_2}{2!}t^2 + \alpha_1 t + \alpha_0, \quad t \in [-c, c]$$

with

$$\alpha_6 = \frac{4!}{2!}c^2, \quad \alpha_4 = \left(1 - \frac{4!}{(2!)^2} \right) c^4, \quad \alpha_3 = \frac{Y_4''(c) - Y_4''(-c)}{2c},$$

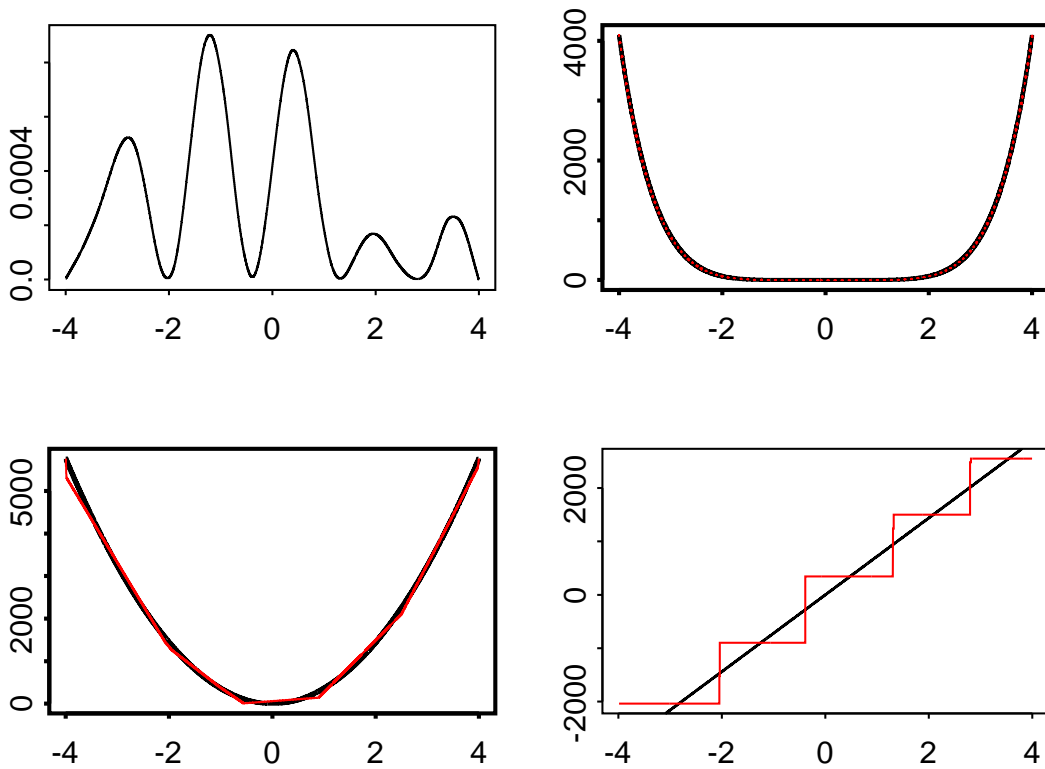


Figure 11: Plots of $(H_{4,6} - Y_6)$, $g_{4,6} = H_{4,6}^{(6)}$ the LS solution (dashed red line) and t^6 (solid black line), $g_{4,6}^{(4)} = H_{4,6}^{(10)}$ (solid red line) and $((6!)/2!) t^2$ (solid black line), and $g_{4,6}^{(5)} = H_{4,6}^{(11)}$ (solid red line) and $6! t$ (solid black line).

Table 2: Table of set of touch points S between the processes $H_{n,k}$ and Y_k for $k = 3, n = 4, 6, 8$ and $k = 6, n = 4$, the value of the LS solution at the origin $g_{n,k}(0)$ and the corresponding number of iterations N_{it} .

$k, [-n, n]$	N_{it}	S	$g_{n,k}(0)$
$k = 3, [-4, 4]$	19	$\{-3.9501, -2.0004, -2.0000, -1.0000, -0.1250, 1.7500, 3.9511\}$	-0.6016
$k = 3, [-6, 6]$	36	$\{-5.9501, -3.9238, -3.9213, -1.9995, -1.0000, -0.1250, 1.7500, 4.0097, 4.0107, 4.0112\}$	-0.5990
$k = 3, [-8, 8]$	42	$\{-6.9985, -5.9995, -4.7495, -4.2500, -3.9892, -3.9873, -1.9995, -1.7500, -1.0000, -0.1250, 1.7500, 4.0356, 4.0390, 6.3291, 6.6250\}$	-0.6004
$k = 6, [-4, 4]$	37	$\{-3.9941, -2.0478, -2.0385, -0.3886, 1.3056, 1.3208, 2.7983, 2.8149, 2.8271\}$	-0.8203

$$\begin{aligned}\alpha_2 &= \frac{Y_4''(-c) + Y_4''(c)}{2} - \left(\frac{\alpha_6}{4!} c^4 + \frac{\alpha_4}{2!} c^2 \right) \\ &= \frac{Y_4''(-c) + Y_4''(c)}{2} - \left(1 - \frac{4!}{(2!)^3} \right) c^6\end{aligned}$$

$$\begin{aligned}\alpha_1 &= \frac{Y_4(c) - Y_4(-c)}{2c} - \frac{\alpha_3}{3!} c^2 \\ &= \frac{Y_4(c) - Y_4(-c)}{2c} - \frac{1}{3!} \left(\frac{Y_4''(c) - Y_4''(-c)}{2c} \right)\end{aligned}$$

and

$$\begin{aligned}\alpha_0 &= \frac{Y_4(-c) + Y_4(c)}{2} - \left(\frac{\alpha_6}{6!} c^6 + \frac{\alpha_4}{4!} c^4 + \frac{\alpha_2}{2!} c^2 \right) \\ &= \frac{Y_4(-c) + Y_4(c)}{2} - \frac{1}{2!} \left(\frac{Y_4''(-c) + Y_4''(c)}{2} \right) c^2 - \left(\frac{4!}{2!6!} + \frac{1}{4!} \left(1 - \frac{4!}{(2!)^2} \right) \right. \\ &\quad \left. - \frac{1}{2!} \left(1 - \frac{4!}{(2!)^3} \right) \right) c^8.\end{aligned}$$

The algorithm was run to obtain an approximation to the process H_k and some of the derivatives $H_k^{(j)}$ for $k = 3$ and $k = 6$ on the interval $[-4, 4]$. Furthermore,

for $k = 3$ we obtained similar approximations but on the bigger intervals $[-6, 6]$ and $[-8, 8]$. The purpose of these additional computations was to look at the effect of letting $c \rightarrow \infty$ on the locations of the jump points and also on the heights of the jumps. A C program, implementing an approximation to the processes $Y_k, Y'_k, \dots, Y_k^{(k-1)}$ on any interval $[-n, n]$ for $n \in \mathbb{N} \setminus \{0\}$ was developed and can be found in, BALABDAOUI (2004B), Appendix C. The approximation to Brownian motion and its successive primitives on $[0, 1]$ was based on the Haar function construction (see e.g. ROGERS AND WILLIAMS (1994), Section 1.6). To obtain an approximation of these processes on $[-n, n]$, independent copies were generated on the intervals $[j, j+1]$ for $j = -n, \dots, n-1$ and pasted “smoothly” at the boundaries. A detailed description of the method and related formulas can be found in BALABDAOUI (2004B), Appendix B. For both $k = 3$ and $k = 6$, we took a finite grid with a mesh of size 2^{-11} . The iterative $2k - 1$ -th spline algorithm was written in S and the corresponding code can be found in BALABDAOUI (2004B), Appendix C. The C program was used offline and the obtained approximations to $Y_k, \dots, Y_k^{(k-1)}$ were stored in a matrix that was thereafter imported and used as an input for the iterative algorithm. For a given interval $[-n, n]$, the output is itself an approximation to the process $H_{n,k}$, the k -fold integral of the LS solution of the Gaussian problem $dX_k(t) = t^k dt + dW(t)$ on $[-n, n]$. An approximation to the derivatives $H'_{n,k}, \dots, H_{n,k}^{(2k-1)}$ can be also obtained on the same chosen grid.

For both $k = 3$ and $k = 6$, the upper left plot in Figure 10 and Figure 11 shows the difference $-(H_{n,k} - Y_k)$ and $H_{n,k} - Y_k$ on $[-4, 4]$ respectively. The sign of $H_{n,k} - Y_k$ is as expected: nonpositive (nonnegative) when k is odd (even). The curves touch the abscissa axis at the points where the derivative $H_{n,k}^{(2k-2)}$ changes its slope. In the upper right plots are the graphs of $g_{n,k} = H_{n,k}^{(k)}$ (in red) and $g_0(t) = t^k$ (in black). The difference between the graphs is not very visible but the motivation behind plotting the functions instead their difference was to show that the LS solution $g_{n,k}$ has the same “form” as the estimated function g_0 .

The lower right plots in Figure 10 and Figure 11 show the convex functions $H_{4,3}^{(4)}$ and $H_{4,6}^{(10)}$ (in red) on $[-4, 4]$ for $k = 3$ and $k = 6$ respectively. These derivatives estimate the “true” convex functions $3t^2$ and $(6!/2!)t^2$ (in black) respectively. The jump processes $H_{4,3}^{(5)}$ and $H_{4,6}^{(11)}$ (in red) are shown in the lower left part. They both estimate a linear function and are monotone since the slopes of $H_{4,3}^{(5)}$ and $H_{4,6}^{(11)}$ are increasing by convexity.

The set of points of touch between $H_{n,k}$ and Y_k for $k = 3, n = 4, 6, 8$ and $k = 6, n = 4$ are provided in Table 2. For $k = 3$, we generated first the process Y_3 and its derivatives Y'_3 and Y''_3 on the interval $[-8, 8]$. Then, we obtained the envelopes $H_{8,3}, H_{6,3}$ and $H_{4,3}$ using the appropriate boundary conditions at the points $-8, 8, -6, 6$ and $-4, 4$ (see Section 2 of Chapter 3 of BALABDAOUI (2004B) for more details on the construction of the envelope H_k when k is odd). It is clear that the obtained points of touch are different and this fact was already noticed by Groeneboom, Jongbloed and Wellner (2001A) in

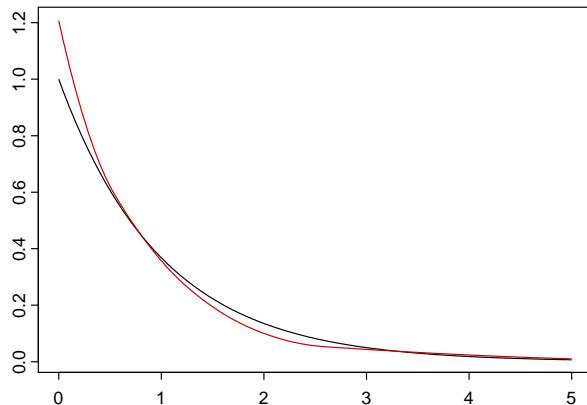


Figure 12: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 100$ and $k = 3$, in red.

the problem of estimating a convex function ($k = 2$). The authors also compared the value of the LS solution at the origin and found that it does not change very much as n increases. We notice the same fact for $k = 3$ (compare the values of $g_{n,3}(0)$ in Table 2). This stability is expected and follows from the fact that $\lim_{n \rightarrow \infty} g_{n,k}(0) = H_3^{(3)}(0)$.

4 Computing the MLE of a k -monotone density on $(0, \infty)$

Let X_1, \dots, X_n be n i.i.d random variables from a k -monotone density g_0 and \mathbb{G}_n be their empirical distribution function. Consider the functional

$$\phi(g) = - \int_0^\infty \log g(t) d\mathbb{G}_n(t) + \int_0^\infty g(t) dt$$

where g belongs to \mathcal{C} , the class of integrable k -monotone functions on $(0, \infty)$. In Section 2 of Chapter 2 in BALABDAOUI (2004B), it was established that ϕ admits a minimizer \hat{g}_n of the form

$$\hat{g}_n(t) = \hat{w}_1 \frac{k(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \dots + \hat{w}_m \frac{k(\theta_m - t)_+^{k-1}}{\theta_m^k}$$

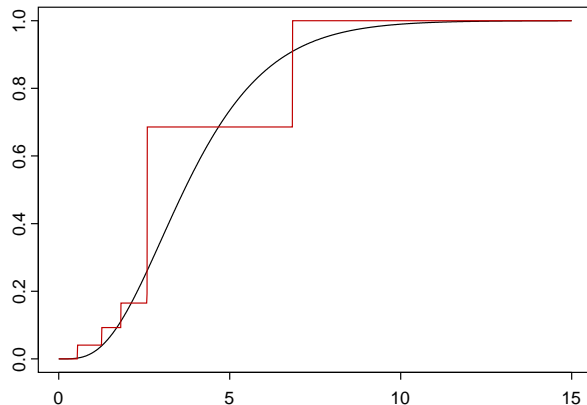


Figure 13: The cumulative distribution function of a $Gamma(4,1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 100$ and $k = 3$ (in red).

where $m \leq n$ and $\hat{w}_1 + \dots + \hat{w}_m = 1$, since this minimizer is nothing but the Maximum Likelihood estimator (\hat{g}_n maximizes $-\phi$). Note that in addition to the log-likelihood term, the functional ϕ is also composed of the “penalty” term $\int_0^\infty g(t)dt$. Without this term, the minimization problem will not be proper since for any nontrivial function $g \in \mathcal{C}$, we would have $\lim_{c \rightarrow \infty} \phi(cg) = -\lim_{c \rightarrow \infty} \log(c) = -\infty$. In the particular case of $k = 2$, GROENEBOOM, JONGBLOED, AND WELLNER (2001B) proved that the MLE is unique. For $k > 2$, we were able to prove the MLE is unique when $k = 3$ (see Lemma 2.2.5 in BALABDAOUI (2004B), Chapter 2) and we conjecture that this holds true for $k > 3$. GROENEBOOM, JONGBLOED, AND WELLNER (2003) noticed that the *support reduction algorithm* is more efficient when it is based on a Newton-type procedure instead of applying it directly to the objective function ϕ . This entails an additional linearization step based on the well-known approximation

$$\log(1+x) \simeq x - \frac{x^2}{2}$$

in the neighborhood of 0. Let \bar{g} be the current iterate and $g \in \mathcal{C}$ such that

$$\frac{g - \bar{g}}{\bar{g}}$$

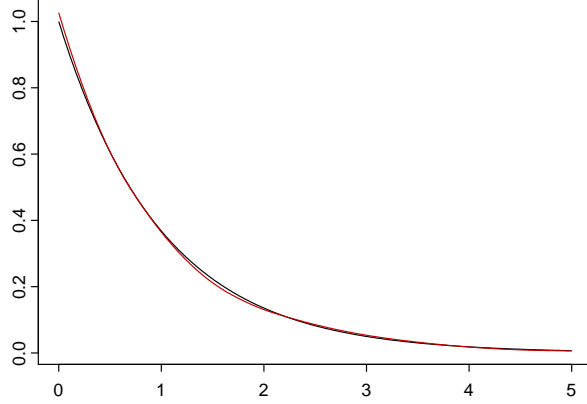


Figure 14: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 1000$ and $k = 3$, in red.

is very small. Then, we can write

$$\begin{aligned} \phi(g) &= \phi(\bar{g}) + \int_0^\infty -\frac{g(t) - \bar{g}(t)}{\bar{g}(t)} d\mathbb{G}_n(t) \\ &\quad + \int_0^\infty \frac{1}{2} \left(\frac{g(t) - \bar{g}(t)}{\bar{g}(t)} \right)^2 d\mathbb{G}_n(t) + \int_0^\infty (g(t) - \bar{g}(t)) dt. \end{aligned}$$

If we delete the terms that do not depend on f , we can define the following local objective function (see GROENEBOOM, JONGBLOED, AND WELLNER (2003))

$$\phi_q(g) = -2 \int_0^\infty \frac{g(t)}{\bar{g}(t)} d\mathbb{G}_n(t) + \int_0^\infty \frac{1}{2} \left(\frac{g(t)}{\bar{g}(t)} \right)^2 d\mathbb{G}_n(t) + \int_0^\infty g(t) dt.$$

Let $\epsilon > 0$ and $f_\theta(t) = k(t - \theta)_+^{k-1} / \theta^k, \theta > 0$. We have

$$\begin{aligned} \phi_q(g + \epsilon f_\theta) &= \phi_q(g) + \epsilon \left(\int_0^\infty -2 \frac{f_\theta(t)}{\bar{g}(t)} d\mathbb{G}_n(t) + \int_0^\infty \frac{g(t) f_\theta(t)}{(\bar{g}(t))^2} d\mathbb{G}_n(t) + \int_0^\infty f_\theta(t) dt \right) \\ &\quad + \frac{\epsilon^2}{2} \int_0^\infty \left(\frac{f_\theta(t)}{\bar{g}(t)} \right)^2 dt \\ &= \phi_q(g) + \epsilon c_1(\theta, g) + \frac{\epsilon^2}{2} c_2(\theta, g). \end{aligned}$$

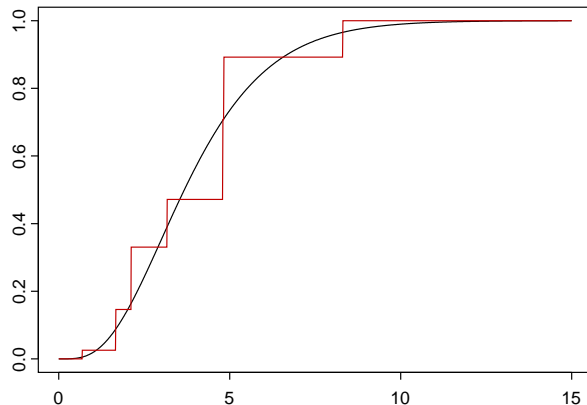


Figure 15: The cumulative distribution function of a $\text{Gamma}(4, 1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 1000$ and $k = 3$ (in red).

The “alternative” directional derivative of ϕ_q at the point g in the direction of f_θ is given by

$$\tilde{D}_{\phi_q}(f_\theta, g) = \frac{c_1(\theta, g)}{\sqrt{c_2(\theta, g)}}.$$

The algorithm consists of an *outer* and *inner* loops. Given a fixed finite grid Θ_f (note that the subscript f is for “finite” and that Θ_f corresponds to Θ_δ used in GROENEBOOM, JONGBLOED, AND WELLNER (2003)) and the current iterate \bar{g} , the *inner* loop is set up to find $\bar{g}_q = \operatorname{argmin}\{\phi_q(g) : g \in \operatorname{cone}(f_\theta, \theta \in \Theta_f)\}$. The next iterate is taken to be $(1 - \lambda)\bar{g} + \lambda\bar{g}_q$, where $\lambda \in (0, 1]$ is appropriately chosen to ensure monotonicity of the algorithm. A *reduction* step is needed to construct a starting value $g^{(0)}$ which will depend of course on the current iterate \bar{g} . To enter the *outer* loop, the minimal value $\min_{\theta \in \Theta_f} \tilde{D}_{\phi_q}(f_\theta, \bar{g})$ needs to be bigger than some fixed tolerance $-\eta$, otherwise we stop. Let $\bar{S} = \{\bar{\theta}_1, \dots, \bar{\theta}_p\}$ denote the set of support points of the current iterate \bar{g} . We proceed as follows:

1. We calculate $\min_{\theta \in \Theta_f} \tilde{D}_{\phi_q}(f_\theta, \bar{g})$. If it is smaller than $-\eta$, we stop. Otherwise, we move to the second step.

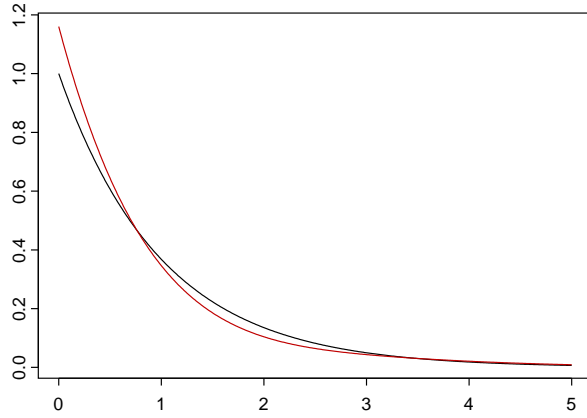


Figure 16: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 100$ and $k = 6$, in red.

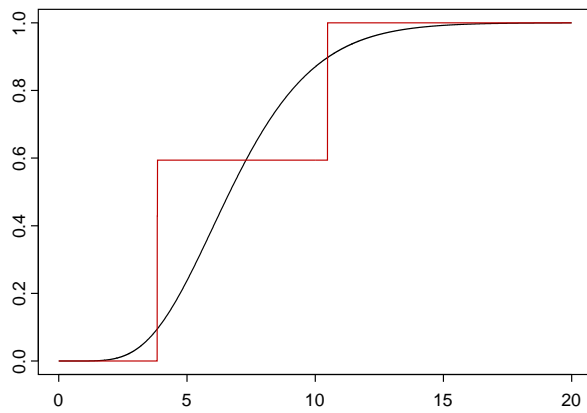


Figure 17: The cumulative distribution function of a $Gamma(7,1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 100$ and $k = 6$ (in red).

Table 3: Table of the obtained ML estimates for $k = 3, 6$ and $n = 100, 1000$. A support point is denoted by \hat{a} and its mass by \hat{w} .

k, n	(\hat{a}, \hat{w})
$k = 3, n = 100$	(0.549, 0.040), (1.259, 0.051), (1.819, 0.072), (2.579, 0.027), (2.589, 0.492), (6.839, 0.314)
$k = 3, n = 1000$	(0.684, 0.025), (1.664, 0.120), (2.114, 0.184), (3.164, 0.141) (4.794, 0.236), (4.824, 0.184), (8.304, 0.107)
$k = 6, n = 100$	(3.839, 0.428), (3.849, 0.165), (10.479, 0.405)
$k = 6, n = 1000$	(3.042, 0.186), (6.452, 0.300), (6.482, 0.267), (11.072, 0.018), (11.102, 0.226)

2. We minimize the local objective function ϕ_q (which depends on \bar{g}) over the cone

$$\mathcal{C}(\Theta_f) = \left\{ g : g(t) = \int_{\theta \in \Theta_f} f_\theta(t) d\mu(\theta), \text{ where } \mu \text{ is a positive measure on } \Theta_f \right\}.$$

For that, we need to find a starting function $g^{(0)}$. The current iterate \bar{g} is not necessarily a good choice and therefore we need to construct one. This can be done as follows: We first minimize the quadratic function

$$\psi(\alpha_1, \dots, \alpha_p) = \phi_q\left(\sum_{j=1}^p \alpha_j f_{\bar{\theta}_j}\right)$$

where $\alpha_1, \dots, \alpha_p \in \mathbb{R}$. Finding this minimum is achieved by finding the solution of the linear system

$$(DY)^t DY \underline{\alpha} = 2Y^t \underline{d} - \underline{n}_p \tag{1}$$

where $Y = (f_{\bar{\theta}_j}(X_i))_{i,j}$ is a $n \times p$ -matrix, D is the $n \times n$ diagonal matrix given by $D_{ii} = 1/\bar{g}(X_i)$, $\underline{d}^t = (1/\bar{g}(X_1), \dots, 1/\bar{g}(X_n))$, \underline{n}_p and $\underline{\alpha}$ are the $p \times 1$ vectors given by $\underline{n}_p^t = (n, \dots, n)$ and $\underline{\alpha}^t = (\alpha_1, \dots, \alpha_p)$ respectively.

Let $g_{min} = \sum_{j=1}^p \alpha_{j,min} f_{\bar{\theta}_j}$ be this minimum. Next, if g_{min} is k -monotone; i.e., $\alpha_{j,min} > 0$ for all $j = 1, \dots, p$, then we take $g^{(0)} = g_{min}$. Otherwise, we find $\lambda \in (0, 1)$ such that $(1 - \lambda)\bar{g} + \lambda g_{min}$ is k -monotone. Such a $\lambda \in (0, 1)$ will always exist and this follows from the same arguments of Lemma 2.2. We repeat the reduction and minimization steps till we find a minimizer that is k -monotone. We take this minimizer to be the starting function $g^{(0)}$. The support of $g^{(0)}$ is in general smaller than \bar{S} as a consequence of successive deletions of support points

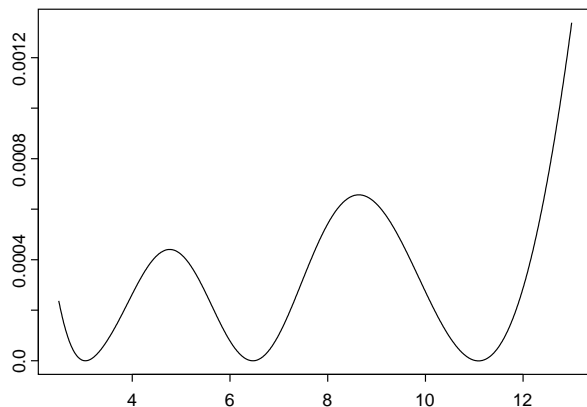


Figure 18: The directional derivative for the Maximum Likelihood estimator of the Exponential density based on $n = 1000$ and $k = 6$.

in the reduction steps.

In the *inner* loop, we proceed as we did for computing the LSE and the process $H_{n,k}$ (see the Section 1 and Section 2). Let m be an integer strictly smaller than p and let us denote the current iterate and its support by \bar{g}_{inner} and \bar{S}_{inner} . We assume without loss of generality that $\bar{S} = \{\bar{\theta}_1, \dots, \bar{\theta}_m\}$. Let $\bar{\theta}_{m+1} = \operatorname{argmin}_{\theta \in \Theta_f} D_{\phi_q}(f_{\theta}, \bar{g}_{inner})$. If $D_{\phi_q}(f_{\bar{\theta}_{m+1}}, \bar{g}_{inner}) \leq -\eta$, we stop. Otherwise, we assume without loss of generality that $\bar{\theta}_{m+1} > \bar{\theta}_m$ and find the minimizer of ϕ_q over the class

$$\mathcal{C}'(\bar{\theta}_1, \dots, \bar{\theta}_{m+1}) = \left\{ g : g = \sum_{j=1}^{m+1} \alpha_j f_{\bar{\theta}_j}, \alpha_j \in \mathbb{R}, j = 1, \dots, m+1 \right\}$$

by solving the linear system given in (1). If the minimizer, g_{min} , is k -monotone, then we take it as the next iterate. Otherwise, we find $\lambda \in (0, 1)$ such that $(1-\lambda)\bar{g}_{inner} + \lambda g_{min}$ is k -monotone and take the first minimizer that is k -monotone as the next iterate.

3. Let $g_{min} = \operatorname{argmin}\{\phi_q(g) : g \in \mathcal{C}(\Theta_f)\}$ obtained in the previous step. Since there is no guarantee that $\phi(g_{min}) \leq \phi(\bar{g})$, we apply the *Armijo* rule; that is, we find

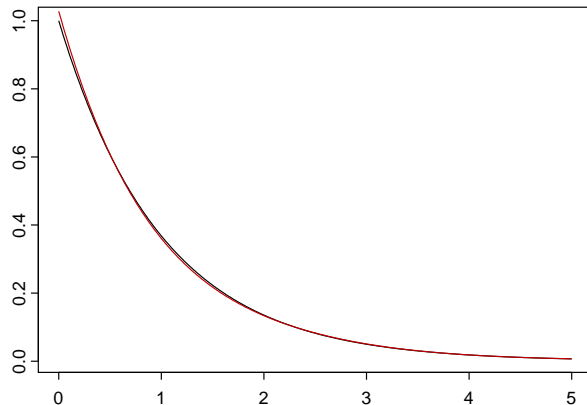


Figure 19: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 1000$ and $k = 6$, in red.

the smallest $\lambda \in (0, 1]$ such that

$$\phi((1 - \lambda)\bar{g} + \lambda g_{min}) \leq \phi(\bar{g}).$$

We take $(1 - \lambda)\bar{g} + \lambda g_{min}$ to be the new iterate for the *outer* loop.

For $k = 3$ and $k = 6$, we calculated the MLE of a standard Exponential based on the same samples of size $n = 100$ and $n = 1000$ used in the Least Squares estimation (see Section 2). The algorithm was coded in *S* and can be found in BALABDAOUI (2004B), Appendix C. To start the algorithm, we calculate $\theta^{(0)}$ the minimizer of the nonlinear function

$$\theta \mapsto -\frac{1}{n} \sum_{j=1}^n \log \left(\frac{k(\theta - X_j)^{k-1}}{\theta^k} \right)$$

for $\theta \geq X_{(n)} + a$, where a is some fixed positive number. This minimization can be performed using the *S* function *nlminb*. Different values of a yield different starting values but the numerical results remained unchanged for many different values which supports our conjecture about uniqueness of the MLE in the general case $k > 3$. As for we did for the LSE, we took a finite grid $\subseteq [X_{(1)}, 2kX_{(n)}]$ with a maximal mesh equal to 0.01. The ML estimation in the direct is illustrated by the plots in Figure 12 and Figure 14 for $k = 3$, and in Figure 16 and Figure 19 for $k = 6$. The “alternative” directional derivative $\tilde{D}_\phi(f_\theta, \hat{g}_n)$, for $n = 1000$ and $k = 6$, is plotted in Figure 18.

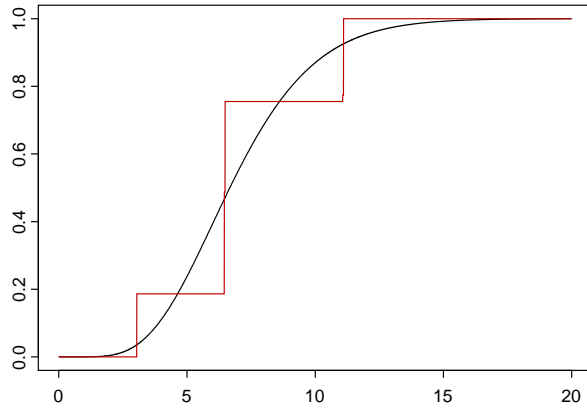


Figure 20: The cumulative distribution function of a $Gamma(7, 1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 1000$ and $k = 6$ (in red).

For the inverse problem, see Figure 13 and Figure 15 for $k = 3$, and Figure 17 and Figure 20 for $k = 6$. Consistency of the MLE is proved in BALABDAOUI (2004B), Chapter 2 and it can be clearly seen in these figures. As for the LSE, convergence in the inverse problem is much slower than in the direct one and the difference becomes more pronounced when k is large. Finally, it should be mentioned here that even if the MLE and LSE of the Exponential density show very small visible differences in the direct problem, it can be easily checked by comparing the locations of jump points or the heights of the jumps that these estimators are different (compare Table 1 and Table 3).

References

- Balabdaoui, F. (2004). A curious fact about k -monotone functions. Technical Report 426, Department of Statistics, University of Washington. Available at: <http://www.stat.washington.edu/www/research/reports/2004/>.
- Balabdaoui, F. (2004). Nonparametric estimation of a k -monotone density: A new asymptotic distribution theory. Ph.D. dissertation, University of Washington, Department of Statistics.
- Böhning, D. (1982). Convergence of Simar's algorithm for finding the maximum likelihood estimate of a compound Poisson process. *Ann. Statist.* **10**, 1006 - 1008.
- Böhning, D. (1986). A vertex exchange method in D -optimal design theory. *Metrika* **33**, 337 - 347.
- de Boor, C. and Fix G. J. (1973). Spline approximation by quasi-interpolants. *J. Approx. Theory* **8**, 19 - 45.
- de Boor, C. (1974). Bounding the error in spline interpolation. *SIAM Rev.* **16**, 531 - 544.
- de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York.
- DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. Springer-Verlag, Berlin.
- Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.
- Feller, W. (1939). Completely monotone functions and sequences. *Duke Math. J.* **5**, 662 - 674.
- Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*. Vol. 2, 2nd ed. Wiley, New York.
- Gelbaum, B. R. and Olmsted, J. M. (1964). *Counterexamples in Analysis*. Holden-Day, San Francisco.
- Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skand. Actuar.* **39**, 125 - 153.
- Groeneboom, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11**, 1016 - 1027.
- Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II. Lucien M. LeCam and Richard A. Olshen eds. Wadsworth, New York. 529 - 555.
- Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Th. Rel. Fields* **81**, 79 - 109.
- Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser, Boston.
- Groeneboom, P. (1996). Inverse problems in statistics. *Proceedings of the St. Flour Summer School in Probability. Lecture Notes in Math.* **1648**, 67 - 164. Springer,

- Berlin.
- Groeneboom, P. and Jongbloed, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist.* **23**, 1518 - 1542.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001a). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29**, 1620 - 1652.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001b). Estimation of convex functions: characterizations and asymptotic theory. *Ann. Statist.* **29**, 1653 - 1698.
- Groeneboom, P. and Wellner J.A. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics.* **10**, 388-400.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2003). The support reduction algorithm for computing nonparametric function estimates in mixture models. Available in Math. ArXiv. at: <http://front.math.ucdavis.edu/math.ST/0405511>.
- Hampel, F.R. (1987). Design, modelling and analysis of some biological datasets. In *Design, data and analysis, by some friends of Cuthbert Daniel*, C.L. Mallows, editor, 111 - 115. Wiley, New York.
- Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10**, 479 - 484.
- Jongbloed, G. (1995). *Three Statistical Inverse Problems; estimators-algorithms-asymptotics*. Ph.D. dissertation, Delft University of Technology, Department of Mathematics.
- Jongbloed, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett.* **50**, 279 - 284.
- Kulikov, V. N. (2002). *Direct and Indirect Use of Maximum Likelihood*. Ph.D. dissertation, Delft University of Technology.
- Lesperance, M. L. and Kalbfleisch, J. D. (1992). An algorithm for computing the nonparametric MLE of a mixing distribution. *Journal of the American Statistical Association* **87**, 120 - 126.
- Lindsay, B. G. (1983). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11**, 86 - 94.
- Lindsay, B. G. (1983). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11**, 86 - 94.
- Lindsay, B. G. (1983). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11**, 86 - 94.
- Mallet, A. (1986). A maximum likelihood estimation for random coefficient regression models. *Biometrika* **73**, 645 - 656.
- Miller, D. R. and Sofer, A. (1986). Least-squares regression under convexity and higher-order difference constraints with application to software reliability. In

- Advances in Order Restricted Inference. Lecture Notes in Statist.* **37**, 91 - 124.
Springer, New York.
- Nürnbergger, G. (1989). *Approximation by Spline Functions*. Springer-Verlag, New York.
- Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes and Martingales*. Wiley, New York.
- Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. John Wiley and Sons, New York.
- Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4**, 1200 - 1209.
- Sun, J. and Woodroffe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Plan. Infer.* **52**, 153 - 159.
- Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.* **23**, 189 - 207.
- Wynn, H. P. (1970). The sequential generation of D -optimum experimental designs. *Ann. Math. Statist.* **6**, 1286 - 1301.

UNIVERSITY OF WASHINGTON
STATISTICS
BOX 354322
SEATTLE, WASHINGTON 98195-4322
U.S.A.
e-mail: balabda@stat.washington.edu

UNIVERSITY OF WASHINGTON
STATISTICS
BOX 354322
SEATTLE, WASHINGTON 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu