Prior Density Estimation via Haar Deconvolution

Alexandre Leblanc   Jean-François Angers

CRM-2834

February 2002

Abstract

The problem of prior elicitation often arises when one is interested in doing inference in a Bayesian setting. Common solutions to this problem consist of hierarchical modelling, noninformative priors, different empirical Bayes techniques, etc. An alternative to these solutions might be to estimate the unknown prior directly from the available observations. In the case of the prior distribution of a location parameter, we approach the problem by deriving a nonparametric estimate of the unknown density. The proposed technique, called Haar deconvolution, is similar, in essence, to the well known Fourier deconvolution which can be used to obtain estimates of mixing densities. The resulting estimated density is a piecewise constant function which is easy to use to perform both numerical integration and statistical simulation.

Key words: Location parameter, nonparametric empirical Bayes methods, prior elicitation, density estimation, wavelets.

Résumé

Le problème du choix de la densité *a priori* se présente très souvent en inférence bayésienne. Les solutions les plus populaires à ce problème sont la modélisation hiérarchique, les lois *a priori* non informatives, diverses techniques bayésiennes empiriques, etc. Une alternative à ces approches consiste à estimer directement, à l'aide des observations, la densité *a priori* inconnue. Plus précisément, dans le cas de la densité *a priori* d'un paramètre de position, nous développons une technique permettant d'obtenir une estimation non paramétrique de la densité inconnue. La technique présentée, appelée déconvolution de Haar, est conceptuellement similaire à la déconvolution de Fourier déjà utilisée pour obtenir des estimateurs de densités mélangeantes. L'approximation résultante est une fonction constante par intervalles qui est facile d'utilisation pour les simulations et l'intégration numérique.

1 Motivation

Suppose that we have

$$X_i|\theta_i \sim f(x_i|\theta_i) \qquad \forall i \in \{1,\, 2,\, \ldots,\, n\}.$$

The Bayesian statistician interested in any type of inference on the θ_i's (estimation, hypothesis testing, confidence intervals, etc.) is confronted with having to identify one or more prior densities. Even if it is reasonable to assume that the θ_i's are independant and share a common prior density, say g, the statistician still has to determine this unique prior, which can sometimes be extremely difficult. Eventhough this process of prior elicitation can be problematic, it still is a necessary step since all Bayesian inference is based on posterior beliefs, namely, an update of the prior beliefs after having seen the available data. Usually, this means that one has to obtain some sort of posterior density for the θ_i's by combining the likelihood with the chosen prior, even before adressing the particular problem he or she is interested in.

Different approaches have already been proposed to assess this difficulty. Among others, hierarchical modelling, noninformative and reference priors and modelling with finite mixtures have been studied by many authors. Different empirical Bayes techniques have also been proposed, which have already led to important results. For instance, see Efron and Morris (1972) and Morris (1983) who have adopted the parametric empirical Bayes approach.

An alternative to these methods has caught our attention: trying to reconstruct the unknown prior using the available data. We propose to use the Haar wavelet basis to build a simple nonparametric estimate of the prior density based on the observed data $X_1,\, X_2,\, \ldots,\, X_n$. The proposed technique is, in essence, similar to the approach of Healy and Kim (1996), who performed Fourier deconvolution on the sphere to obtain a nonparametric estimate of the prior density for directional data.

Note that this problem is not exclusively Bayesian. It is essentially the same as the estimation of a mixing density in the more general context of mixture models. It is also similar to the estimation of a density under data contamination via additive errors. Many have adressed these problems using various techniques. Amongst others, Zhang (1990) and Fan (1991a, 1991b) have used kernel density estimators based on Fourier deconvolution, Goutis (1997) has used kernel density estimators constructed by conditionning on the observed data and Laird (1978) and Laird and Louis (1991) have used nonparametric maximum likelihood estimation.

2 Statement of the problem for location parameters

Consider the case where

$$X_i|\theta_i \sim f(x_i - \theta_i) \qquad \forall i \in \{1,\, 2,\, \ldots,\, n\},$$

and

$$\theta_i \sim g(\theta_i) \qquad \forall i \in \{1,\, 2,\, \ldots,\, n\}.$$

We suppose that $f \in \mathcal{L}^2(\mathbb{R})$ is specified and that $g \in \mathcal{L}^2(\mathbb{R})$ is unknown. Also, the observations are supposed to be conditionally independent, and the θ_i's are assumed independent.

As mentionned earlier, we are interested in estimating the unknown prior density g (of the unobservable $\theta_1, \theta_2, \ldots, \theta_n$) using the observed data $X_1, X_2, \ldots, X_n$. In order to accomplish this, we propose to use the Haar wavelet basis to first construct an approximation of

$$m(x) = \int_{\mathbb{R}} f(x-\theta) g(\theta)\, d\theta, \tag{1}$$

the marginal density common to all the observations. From this approximation of m, we will be able to construct an approximation of the unknown prior g.

3 Haar approximation of a density

In the present section, we discuss briefly the approximation of a known density using the Haar basis. For more details on this and other matters related to wavelet bases, we refer the reader to Daubechies (1992), Mallat (1998), Härdle *et al.* (1998) and Vidakovic (1999). (Note that the last two references given above also deal with many different applications of wavelets in statistics.)

Now, let

$$\phi(s) = \begin{cases} 1 & \text{if } s \in [0,\, 1), \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \psi(s) = \begin{cases} 1 & \text{if } s \in [0,\, 1/2), \\ -1 & \text{if } s \in [1/2,\, 1), \\ 0 & \text{otherwise,} \end{cases}$$

and define the dyadic dilates and translates of both ϕ and ψ as

$$\phi_{jk}(s) = 2^{j/2}\phi(2^j s - k) \quad \text{and} \quad \psi_{jk}(s) = 2^{j/2}\psi(2^j s - k),$$

for $j,\, k \in \mathbb{Z}$. Then, for any fixed $J \in \mathbb{Z}$,

$$\left\{ \phi_{Jk}, \psi_{jk} : j \geq J,\, k \in \mathbb{Z} \right\}$$

forms an orthonormal basis of $\mathcal{L}^2(\mathbb{R})$. Hence, any square integrable density h can be written as

$$h(s) = \sum_{k\in\mathbb{Z}} \alpha_{Jk}\phi_{Jk}(s) + \sum_{j\geq J}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(s), \tag{2}$$

where the coefficients α_{Jk} and β_{jk} are given by

$$\begin{aligned} \alpha_{Jk} &= < h,\, \phi_{Jk} > \\ &= \int_{\mathbb{R}} h(s)\phi_{Jk}(s)\, ds \\ &= 2^{J/2}\left[H\left(\frac{k+1}{2^J}\right) - H\left(\frac{k}{2^J}\right)\right], \end{aligned}$$

for $k \in \mathbb{Z}$, and

$$\begin{aligned} \beta_{jk} &= < h,\, \psi_{jk} > \\ &= 2^{j/2}\left[2H\left(\frac{k+1/2}{2^j}\right) - H\left(\frac{k+1}{2^j}\right) - H\left(\frac{k}{2^j}\right)\right], \end{aligned}$$

for j, $k \in \mathbb{Z}$. Here, H corresponds to the cdf associated with the density h.

Finally, if we define h_J by

$$h_J(s) = \sum_{k\in\mathbb{Z}} \alpha_{Jk}\phi_{Jk}(s), \tag{3}$$

then equation (2) becomes

$$h(s) = h_J(s) + \sum_{j\geq J}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(s).$$

This decomposition can be seen as a basic approximation of h, namely $h_J(s)$, to which details of higher levels are added. This makes it natural to use h_J to approximate h and hence, we will consider h_J to be the Haar approximation of h at level J. If J is chosen correctly, the approximation should do fairly well, as will be seen later. This is confirmed by Mallat's multiresolution analysis theory (*cf.* Mallat, 1989) from which we get the pointwise convergence of h_J, that is

$$h_J(s) \underset{J\to\infty}{\longrightarrow} h(s) \qquad \forall s \in \mathbb{R}, \tag{4}$$

and convergence in the $\mathcal{L}^2(\mathbb{R})$-norm of h_J, that is

$$\left\|h_J - h\right\|_2 = \left(\int_{\mathbb{R}} \left(h_J(s) - h(s)\right)^2 ds\right)^{1/2} \underset{J\to\infty}{\longrightarrow} 0. \tag{5}$$

4 Approximation of the marginal density

Our goal here is to construct the Haar approximation of m. As will be seen in section 5, this approximation will be useful to recover the unknown prior g. For this, we first note that the Haar approximation of g is given, according to equation (3), by

$$g_J(s) = \sum_{k\in\mathbb{Z}} b_{Jk}\phi_{Jk}(s), \tag{6}$$

where

$$b_{Jk} = 2^{J/2}\left[G\left(\frac{k+1}{2^J}\right) - G\left(\frac{k}{2^J}\right)\right],$$

for $k \in \mathbb{Z}$, and where G is the cdf associated with g.

Now, from equation (1), m can be approximated according to

$$\begin{aligned} m(x) &\simeq \int_{\mathbb{R}} f(x-s)g_J(s)\,ds \\ &= \sum_{k\in\mathbb{Z}} b_{Jk}\int_{\mathbb{R}} f(x-s)\phi_{Jk}(s)\,ds \\ &= \sum_{k\in\mathbb{Z}} b_{Jk}a_{Jk}(x), \end{aligned} \tag{7}$$

where

$$a_{Jk}(x) = f{*}\phi_{Jk}(x)$$
$$= 2^{J/2}\left[F\left(\frac{2^J x - k}{2^J}\right) - F\left(\frac{2^J x - k - 1}{2^J}\right)\right], \tag{8}$$

and where $f{*}\phi_{Jk}$ denotes the convolution of the functions f and ϕ_{Jk}, and F corresponds to the cdf associated with f. Hence, as an initial approximation of m, we define

$$\widetilde{m}_J(x) = \sum_{k=-N}^{N} b_{Jk} a_{Jk}(x),$$

which is nothing but a truncated version of the approximation given by equation (7). Note that the coefficients $a_{Jk}(x)$ can be computed for all $k \in \mathbb{Z}$ and $x \in \mathbb{R}$ since F is known. However, it is not the case for the coefficients b_{Jk} which depend on the unknown prior g through its cdf.

The following result is essentially taken from Leblanc and Angers (1999), with minor modifications that can be found in Leblanc (2002).

Theorem 1 Suppose that $f, g \in \mathcal{L}^2(\mathbb{R}) \cap C^3(\mathbb{R})$, and that $f^{(i)}, g^{(i)} \in \mathcal{L}^2(\mathbb{R})$ for $i = 1, 2, 3$. If there exists $\varepsilon > 0$, $c > 0$ and $M > 0$ such that

$$f(s)\, g(s) \leq \frac{c}{|s|^{3+\varepsilon}} \qquad \forall |s| > M,$$

and if $N \propto 2^{J(1+2/\varepsilon)}$, then

$$|\widetilde{m}_J(x) - m(x)| = O\big(2^{-2J}\big) \qquad \forall x \in \mathbb{R}.$$

□

Note that $\widetilde{m}_J$ is a very smooth approximation of m. In fact, since $f \in C^3(\mathbb{R})$, then $F \in C^4(\mathbb{R})$ and so is $\widetilde{m}_J$, depending on x only through the coefficients $a_{Jk}(x)$ given by equation (8).

Using the previous result, we can finally obtain an expression for m_J, the Haar approximation of m at level J. For this, first note that from equation (3), we know that m_J is given by

$$m_J(x) = \sum_{k\in\mathbb{Z}} d_{Jk} \phi_{Jk}(x),$$

where

$$d_{Jk} = < m, \phi_{Jk} >, \tag{9}$$

for $k \in \mathbb{Z}$. In the following lemma, we obtain closed form expressions for the coefficients d_{Jk}.

Lemma 1 Under the conditions of Theorem 1,

$$d_{Jk} = \frac{2^{-J/2}}{6} \sum_{l=-N}^{N} b_{Jl}\left[a_{J0}\left(\frac{k-l}{2^J}\right) + 4\, a_{J0}\left(\frac{2(k-l)+1}{2^{J+1}}\right) + a_{J0}\left(\frac{k-l+1}{2^J}\right)\right]$$
$$+ O\big(2^{-5J/2}\big) \qquad \forall k \in \mathbb{Z}.$$

□

The proof of Lemma 1 can be found in the Appendix.

This is a very interesting result since it links the coefficients d_{Jk} of the Haar approximation of the marginal density, to the coefficients b_{Jk} of the prior. More precisely, it enables us to construct a simple system of linear equations from which estimates for the b_{Jk}'s can be obtained. To see this, note that all the coefficients $a_{J0}(\cdot)$ are computable via equation (8), and that the unknown d_{Jk}'s are directly estimable from the available data $X_1, X_2, \ldots, X_n$.

5 Haar deconvolution

Throughout this section, we assume that the coefficients d_{Jk} are computable via equation (9). In this context, the expression of the d_{Jk}'s given by Lemma 1 is composed of known quantities, except for the coefficients b_{Jk} of the Haar approximation of the unknown prior g. Hence, it is here possible to construct a system of linear equations, satisfied by the b_{Jk}'s, which can be solved in order to recover the unknown coefficients.

First, let $\mathcal{N} = \{k : -N \leq k \leq N\}$ and define

$$\mathrm{b} = (b_{Jk})_{k\in\mathcal{N}},$$

and, for $k \in \mathbb{Z}$,

$$\mathrm{a}_k = \left(a_{J0}\left(\frac{k-l}{2^J}\right) + 4\, a_{J0}\left(\frac{2(k-l)+1}{2^{J+1}}\right) + a_{J0}\left(\frac{k-l+1}{2^J}\right) \right)_{l\in\mathcal{N}}.$$

Note that using these vectors and letting a_k^t denote transposition of the vector a_k, the result of Lemma 1 can be expressed as

$$d_{Jk} = \frac{2^{-J/2}}{6}\mathrm{a}_k^t\mathrm{b} + O\big(2^{-5J/2}\big),$$

for $k \in \mathbb{Z}$. Finally, let

$$\mathrm{d} = (d_{Jk})_{k\in\mathcal{N}},$$

and define the matrix

$$\mathrm{A} = \left(\mathrm{a}_k^t\right)_{k\in\mathcal{N}},$$

which is a $(2N+1)\times(2N+1)$ square matrix. We can now write the system of equations resulting from Lemma 1 in matrix form, that is

$$\mathrm{d} = \frac{2^{-J/2}}{6}\mathrm{A}\mathrm{b} + O\big(2^{-5J/2}\big)1_{2N+1}, \tag{10}$$

where 1_{2N+1} is a vector of length $2N+1$ having all its elements equal to unity.

The elements of A can be shown to be

$$\mathrm{A}_{ij} = 2^{J/2}\left[F\left(\frac{i-j}{2^J}\right) - F\left(\frac{i-j-1}{2^J}\right) + 4\,F\left(\frac{2(i-j)+1}{2^{J+1}}\right) - 4\,F\left(\frac{2(i-j)-1}{2^{J+1}}\right)\right],$$

for $i,\, j \in \{1,\, 2,\, \ldots,\, 2N+1\}$. Now, let $\mathrm{B} = 2^{-J/2}\mathrm{A}$, so that the elements of B are given by

$$\mathrm{B}_{ij} = F\left(\frac{i-j}{2^J}\right) - F\left(\frac{i-j-1}{2^J}\right) + 4\,F\left(\frac{2(i-j)+1}{2^{J+1}}\right) - 4\,F\left(\frac{2(i-j)-1}{2^{J+1}}\right),$$

for $i,\, j \in \{1,\, 2,\, \ldots,\, 2N+1\}$. Using this new matrix, equation (10) can be rewritten as

$$\mathrm{d} = \frac{1}{6}\mathrm{B}\mathrm{b} + O\big(2^{-5J/2}\big)1_{2N+1}. \tag{11}$$

Note that the elements of B satisfy

$$\mathrm{B}_{(i+1)(j+1)} = \mathrm{B}_{ij},$$

for $i,\, j \in \{1,\, 2,\, \ldots,\, 2N\}$. Note also that if f is a symmetric density function, then B will be a symmetric matrix.

Looking more closely at equation (11) (and remembering d is, for the moment, considered known), it seems natural to approximate b using

$$\mathrm{b}^* = 6\mathrm{B}^{-1}\mathrm{d}, \tag{12}$$

whenever B is nonsingular. This intuition is confirmed by the next theorem which is given without proof, following directly from equation (11).

Theorem 2 Under the conditions of Theorem 1, if B is nonsingular and if there exists $\delta < 5/2$ such that

$$\mathrm{B}^{-1}1_{2N+1} = O\big(2^{\delta J}\big)1_{2N+1},$$

then

$$\mathrm{b} = 6\mathrm{B}^{-1}\mathrm{d} + O\big(2^{-J(5/2-\delta)}\big)1_{2N+1}.$$

□

Note that the existence of a δ such as the one identified in Theorem 2 is empirically established, in different situations, in Leblanc (2002). Note also that this process of inversion is refered to as *Haar deconvolution*, since it is similar, in many points, to the Fourier deconvolution on the sphere studied by Healy and Kim (1996).

6 Nonparametric estimation of the prior

In the last section, we have assumed that d is known, although we have also mentionned it is usually not the case. This difficulty can be circumvented since d can be estimated from the sample $X_1,\, X_2,\, \ldots,\, X_n$. In fact, it can be shown that

$$\hat{d}_{Jk} = \frac{1}{n}\sum_{i=1}^{n} \phi_{Jk}(X_i) \tag{13}$$

is an unbiased estimate of d_{Jk}, for $k \in \mathcal{N}$. Hence,

$$\hat{\mathrm{d}} = \left(\hat{d}_{Jk}\right)_{k\in\mathcal{N}} \tag{14}$$

is a natural estimator of d to consider. Replacing d by $\hat{\mathrm{d}}$ in equation (12), we propose the following estimator for b which is well defined whenever B is nonsingular,

$$\hat{\mathrm{b}} = 6\mathrm{B}^{-1}\hat{\mathrm{d}}, \tag{15}$$

where

$$\hat{\mathrm{b}} = \left(\hat{b}_{Jk}\right)_{k\in\mathcal{N}}.$$

Note that the resulting estimator is biased since

$$\mathbb{E}\left[\hat{\mathrm{b}}\right] = 6\mathrm{B}^{-1}\mathrm{d} = \mathrm{b}^{*}.$$

Note also that the bias of $\hat{\mathrm{b}}$ depends on the quality of the approximation b^{*} of b.

It is now possible to define the Haar deconvolution estimator of the unknown prior g.

Definition 1 The Haar deconvolution estimator of level J of the prior g , which we denote $\hat{g}_J$, is given by

$$\hat{g}_J(s) = \sum_{k=-N}^{N} \hat{b}_{Jk}\phi_{Jk}(s) \qquad \forall s \in \mathbb{R}.$$

□

It is easily seen that $\hat{g}_J(s)$ is a biased estimator of $g(s)$ from the fact that $\hat{\mathrm{b}}$ is a biased estimator of b. However, this is of no great concern since it was established by Rosenblatt (1956) that it is not possible to construct a nonparametric density estimator that is unbiased for all densities. Of course, it is still of interest to show that the bias of the estimator given by Definition 1 is small (in some sense), and that this estimator also meets different convergence criterias. See Izenman (1991) for more details on the consistency of nonparametric density estimators.

Note that $\hat{g}_J$ usually does not lead to a *bona fide* estimate of g, that is, the estimated density may take on negative values and/or does not integrate to unity. To circumvent this problem, we define the truncated and normalized version of $\hat{g}_J$ as

$$\hat{g}_J^{+}(s) = \sum_{k=-N}^{N} \hat{B}_{Jk}\phi_{Jk}(s), \tag{16}$$

where

$$\hat{B}_{Jk} = \frac{2^{J/2}}{\sum_{l=-N}^{N}\left[\hat{b}_{Jl}\right]^{+}}\left[\hat{b}_{Jk}\right]^{+},$$

and where

$$\left[\hat{b}_{Jk}\right]^{+} = \begin{cases} \hat{b}_{Jk} & \text{if } \hat{b}_{Jk} \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that, thus defined, $\hat{g}_J^{+}$ is a *bona fide* estimator.

Now, before a detailed examination of the behavior of the Haar deconvolution estimator is done, it is clear, from equation (15) and from Definition 1, that we need to know more about $\hat{\mathrm{d}}$ and $\hat{\mathrm{b}}$. The next lemma, whose proof is outlined in the Appendix, gives relevant results pertaining to $\hat{\mathrm{d}}$.

Lemma 2 The estimator $\hat{\mathrm{d}}$ defined by equations (13) and (14) satisfies

$$n2^{-J/2}\hat{\mathrm{d}} \sim \mathcal{M}\left(n, 2N+1, 2^{-J/2}\mathrm{d}\right),$$

that is, $n2^{-J/2}\hat{\mathrm{d}}$ has a multinomial distribution with parameters n, $2N+1$ and $2^{-J/2}\mathrm{d}$. Consequently,

1. $\mathbb{E}\left[\hat{\mathrm{d}}\right] = \mathrm{d}$ and $\mathbb{C}\mathrm{ov}\left[\hat{\mathrm{d}}\right] = n^{-1}\Sigma$,

   where $\Sigma_{ij} = \begin{cases} d_{J(i-N-1)}(2^{J/2} - d_{J(i-N-1)}) & \text{if } i = j, \\ -d_{J(i-N-1)}d_{J(j-N-1)} & \text{otherwise,} \end{cases}$

   for $i, j \in \{1, 2, \ldots, 2N+1\}$.

2. $\hat{\mathrm{d}} \sim AN_{2N+1}\left(\mathrm{d}, n^{-1}\Sigma\right)$,

   that is, $\hat{\mathrm{d}}$ asymptotically follows a multivariate normal distribution with mean d and covariance matrix $n^{-1}\Sigma$.

3. $\hat{\mathrm{d}} \xrightarrow{P} \mathrm{d}$,

   that is, $\hat{\mathrm{d}}$ converges in probability to d.

□

We now give important results about the behavior of $\hat{\mathrm{b}}$. They are given without proof as they follow directly from Lemma 2.

Theorem 3 If B is nonsingular, the estimator $\hat{\mathrm{b}}$ defined by equation (15) has the following properties,

1. $\mathbb{E}\left[\hat{\mathrm{b}}\right] = 6\mathrm{B}^{-1}\mathrm{d} = \mathrm{b}^*$ and $\mathbb{C}\mathrm{ov}\left[\hat{\mathrm{b}}\right] = 36n^{-1}\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t$,

   where Σ is defined as in Lemma 2.

2. $\hat{\mathrm{b}} \sim AN_{2N+1}\left(6\mathrm{B}^{-1}\mathrm{d}, 36n^{-1}\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\right)$.

3. $\hat{\mathrm{b}} \xrightarrow{P} 6\mathrm{B}^{-1}\mathrm{d} = \mathrm{b}^*$.

□

At this point, we can now focus on the behavior of the Haar deconvolution estimator. For this, we first notice that letting

$$\Phi(s) = \left(\phi_{Jk}(s)\right)_{k\in\mathcal{N}},$$

$\hat{g}_J(s)$ can be written as

$$\hat{g}_J(s) = \hat{\mathrm{b}}^t\Phi(s). \tag{17}$$

The next corollary gives properties of $\hat{g}_J$ that follow directly from Theorem 3.

Corollary 1 When B is nonsingular, the Haar deconvolution estimator $\hat{g}_J$ (*cf.* Definition 1) satisfies

1. $\mathbb{E}\big[\hat{g}_J(s)\big] = 6\mathrm{d}^t(\mathrm{B}^{-1})^t\Phi(s)$ and $\mathbb{V}\mathrm{ar}\big[\hat{g}_J(s)\big] = 36n^{-1}\Phi^t(s)\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\Phi(s)$, where Σ is again defined as in Lemma 2.

2. $\hat{g}_J(s) \sim AN\left(6\mathrm{d}^t(\mathrm{B}^{-1})^t\Phi(s), 36n^{-1}\Phi^t(s)\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\Phi(s)\right) \qquad \forall s \in \bigcup_{k\in\mathcal{N}} A_{Jk}$, and $\mathbb{P}\big(\hat{g}_J(s) = 0\big) = 1 \qquad \forall s \notin \bigcup_{k\in\mathcal{N}} A_{Jk}$.

□

Here, we remind the reader that the Haar approximation of g, denoted by g_J, is defined by equation (6). We now define the truncated form of g_J, denoted by g_J^*, as

$$g_J^*(s) = \sum_{k=-N}^{N} b_{Jk}\phi_{Jk}(s) = \mathrm{b}^t\Phi(s). \tag{18}$$

Note that according to Theorem 1, we have that $N \propto 2^{J(1+2/\varepsilon)}$, where ε depends only on the nature of the densities f and g. Also, in Section 3, it was established that (*cf.* equations (4) and (5))

$$g_J(s) \underset{J\to\infty}{\longrightarrow} g(s) \qquad \forall s \in \mathbb{R},$$

and that

$$\big\|g_J - g\big\|_2 \underset{J\to\infty}{\longrightarrow} 0.$$

Hence, since $N = N(J) \underset{J\to\infty}{\longrightarrow} \infty$, we also have, for the truncated form of g_J, that

$$g_J^*(s) \underset{J\to\infty}{\longrightarrow} g(s) \qquad \forall s \in \mathbb{R},$$

and that

$$\big\|g_J^* - g\big\|_2 \underset{J\to\infty}{\longrightarrow} 0.$$

We now give three results, proven in the Appendix, dealing with the pointwise convergence of $\hat{g}_J$.

Theorem 4 Under the conditions of Theorems 1 and 2 with $\delta < 2$, the Haar deconvolution estimator of the prior g satisfies

1. $\mathbb{E}\big[\hat{g}_J(s)\big] = g_J^*(s) + O\big(2^{-J(2-\delta)}\big) \qquad \forall s \in \mathbb{R}$,

2. $\hat{g}_J(s) \xrightarrow{P} g_J^*(s) + O\big(2^{-J(2-\delta)}\big) \qquad \forall s \in \mathbb{R}$,

3. $\mathrm{MSE}\big[\hat{g}_J(s)\big] = \mathbb{E}\left[\big(\hat{g}_J(s) - g(s)\big)^2\right] \underset{n\to\infty}{\longrightarrow} \big(g_J^*(s) - g(s)\big)^2 + O\big(2^{-J(2-\delta)}\big) \qquad \forall s \in \mathbb{R}$.

□

Finally, we address the convergence of $\hat{g}_J$ in the $\mathcal{L}^2(\mathbb{R})$-norm. We particularly look at two popular criterias pertaining to this type of convergence, which are:

- the ISE (Integrated Squared Error) of $\hat{g}_J$, given by

$$\mathrm{ISE}\big[\hat{g}_J\big] = \|\hat{g}_J - g\|_2^2 = \int_{\mathbb{R}} \big(\hat{g}_J(s) - g(s)\big)^2 ds,$$

- the MISE (Mean Integrated Squared Error) of $\hat{g}_J$, given by

$$\text{MISE}\big[\hat{g}_J\big] = \mathbb{E}\left[\|\hat{g}_J - g\|_2^2\right],$$

  this last expectation taken with respect to the marginal density of each observation, defined by equation (1).

The proof of the following result can be found in the Appendix.

Theorem 5 Under the conditions of Theorems 1 and 2, if $\varepsilon > 1/2$ and $\delta < 2 - 1/\varepsilon$, the estimator $\hat{g}_J$ satisfies

1. $\text{ISE}\big[\hat{g}_J\big] \xrightarrow{P} \|g_J^* - g\|_2^2 + O\big(2^{-2J(2-\delta-1/\varepsilon)}\big)$.
2. $\text{MISE}\big[\hat{g}_J\big] \underset{n\to\infty}{\longrightarrow} \|g_J^* - g\|_2^2 + O\big(2^{-2J(2-\delta-1/\varepsilon)}\big)$.

□

We now turn to a simulated example which illustrates some features of the Haar deconvolution estimator.

7 Numerical example

In this last section, we present a short simulated example illustrating the use of the proposed methodology. But, before we do so, a few comments about the choices of J (the level of the Haar deconvolution estimator) and N (the parameter controlling the number of terms to include in the estimator) have to be made.

First, note that $\hat{g}_J^+$ is a normalized histogram estimator of g with bandwidth equal to 2^{-J}. Note also that $\hat{g}_J^+$ is based (*cf.* equations (15) and (16)) on the vector $\hat{\text{d}}$ which can be shown to correspond to the vector of coefficients of a normalized histogram estimate of the marginal density m, also with bandwidth 2^{-J}. For large values of J, this estimate of m could certainly be highly chaotic and, in these conditions, it seems it would be extremely difficult to recover any kind of satisfying estimate for the prior. Hence, if one hopes to get a sensible estimate for the prior, a small or moderate value of J should be used, eventhough the asymptotic results of the previous section tend to suggest the opposite.

For N, we propose to use $N = C2^J$ when the likelihood exhibits any type of exponential decay (which can be seen, in Theorem 1, as a limiting case where $\varepsilon = \infty$). Finally, we propose to choose the value of C so that $X_i \in \bigcup_{k\in\mathcal{N}} A_{Jk}$ for all $i \in \{1, 2, \ldots, n\}$, that is, to choose C so that the histogram estimate of m constructed using $\hat{\text{d}}$ is *bona fide*.

Now, consider the case where

$$X_i|\theta_i \sim N(\theta_i, 1) \qquad \forall i \in \{1, 2, \ldots, n\},$$

and where the observations $X_1, X_2, \ldots, X_n$ are conditionally independent. In addition, assume that $\theta_1, \theta_2, \ldots, \theta_n$ are independent and share a common prior density. Finally, consider the two following priors given by

$$g_j(\theta) \equiv \frac{1}{3}N(-\mu_j, 1) + \frac{2}{3}N(\mu_j, 1),$$

for $j = 1, 2$, with $\mu_1 = 3$ and $\mu_2 = 3/2$. Our goal here is to reconstruct the true priors g_1 and g_2 in two different ways. First, we want to get theoretical reconstructions of the two priors using the fact that the true marginal densities can be computed, and hence the true value of d can be obtained for each model using equation (9). We also want to reconstruct the priors from simulated data. In doing this, we hope to see which features of the two priors (*e.g.* location, height and separation of the modes) will be detected by the Haar deconvolution estimator.

Note that the marginal densities associated to each model are here given by

$$m_j \equiv \frac{1}{3}N(-\mu_j, 2) + \frac{2}{3}N(\mu_j, 2),$$

for $j = 1, 2$. Hence, we can indeed obtain the vectors of the Haar coefficients of each marginal density, namely d_1 and d_2, and generate data according to each model using the known marginal densities.

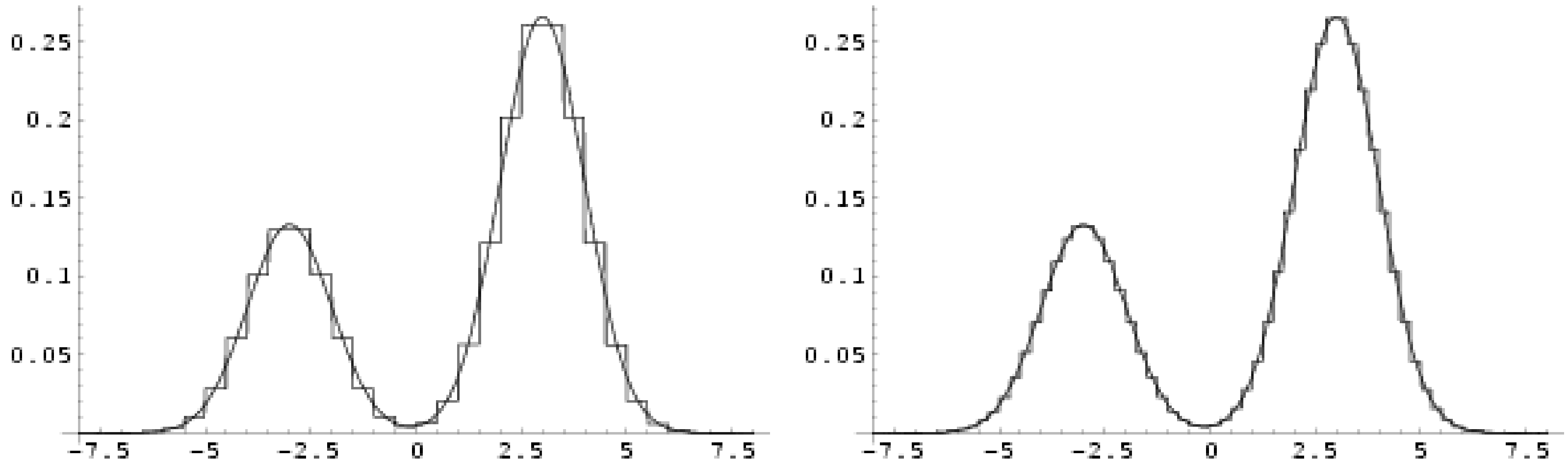
Figures 1 and 2 show the two priors reconstructed from the known values d_1 and d_2 at levels $J = 1$ and $J = 2$, using each time $N = 8 \times 2^J$. The technique is highly efficient in these cases as the reconstructed priors are all very satisfying. Note that truncation and renormalization was not necessary in these cases.

Figures 3 and 4 show normalized histogram estimates of m_1 and m_2 constructed using simulated samples of size $n = 600$ generated from the known marginal densities. Again, $N = 8 \times 2^J$ was used. Figures 5 and 6 show the two priors reconstructed using the same simulated samples. From these, we see that the special features of both priors seem to have been captured fairly well, except maybe for the height of the second mode of g_1. Looking more closely at Figure 3, we see that this was to be expected here since this particular region is under-represented by the data.

Finally, we address a phenomenon that can be witnessed in figures 5 and 6: there seems to be an oscillatory edge effect in the reconstructed priors. This phenomenon has also been witnessed by others while estimating a density via deconvolution. One solution that seems to be popular is to incorporate a smoothing kernel or damping factor that ensures a rapid decrease to 0 of the estimated density outside the sample range. This could certainly be done here, but the effects of this modification on the convergence of $\hat{g}_J$, given by theorems 4 and 5, would have to be studied. Also, this modification would not be done without affecting the simplicity of the proposed methodology, which is certainly, in the present form, it's biggest asset.

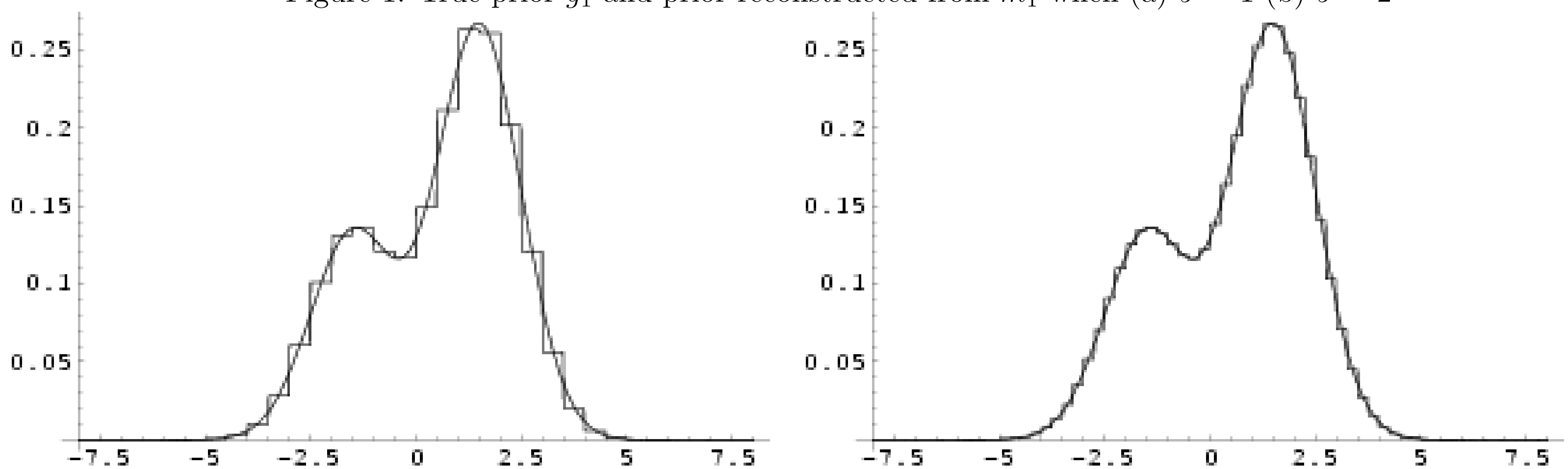
8 Conclusion

In conclusion, the proposed method is a simple answer to a difficult problem which seems to perform quite well. On the down side, the convergence of the Haar deconvolution estimator appears to be quite slow. This could be explained by the fact that deconvolution is a very difficult problem in itself, and that $\hat{\mathrm{d}}$ converges slowly to d (since it corresponds to the histogram estimate of the marginal density m, and that histogram estimators are well known to exhibit slow convergence, *cf.* Izenman, 1991). Hence, large sample sizes are necessary in order for the estimator to perform well.



(a) (b)

Figure 1: True prior g_1 and prior reconstructed from m_1 when (a) $J = 1$ (b) $J = 2$



(a) (b)

Figure 2: True prior g_2 and prior reconstructed from m_2 when (a) $J = 1$ (b) $J = 2$

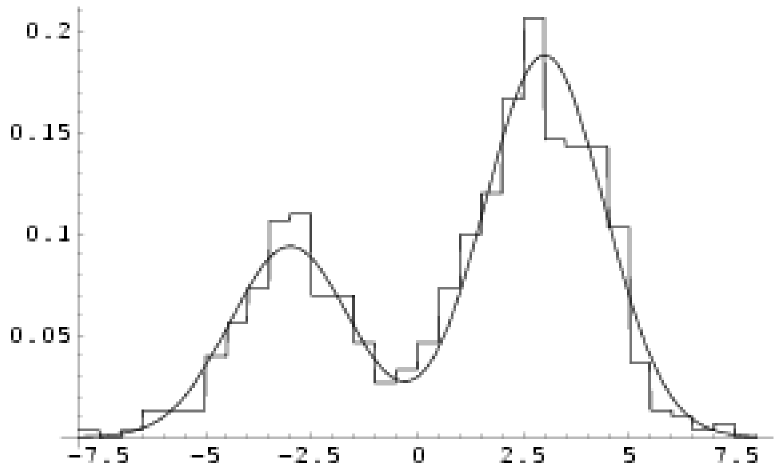


Figure 3: True marginal density m_1 and normalized histogram estimate of m_1 constructed from simulated data when $J = 1$

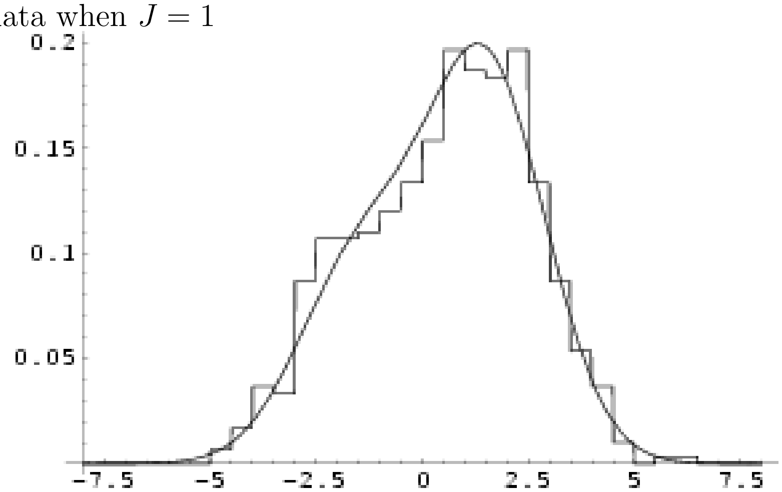
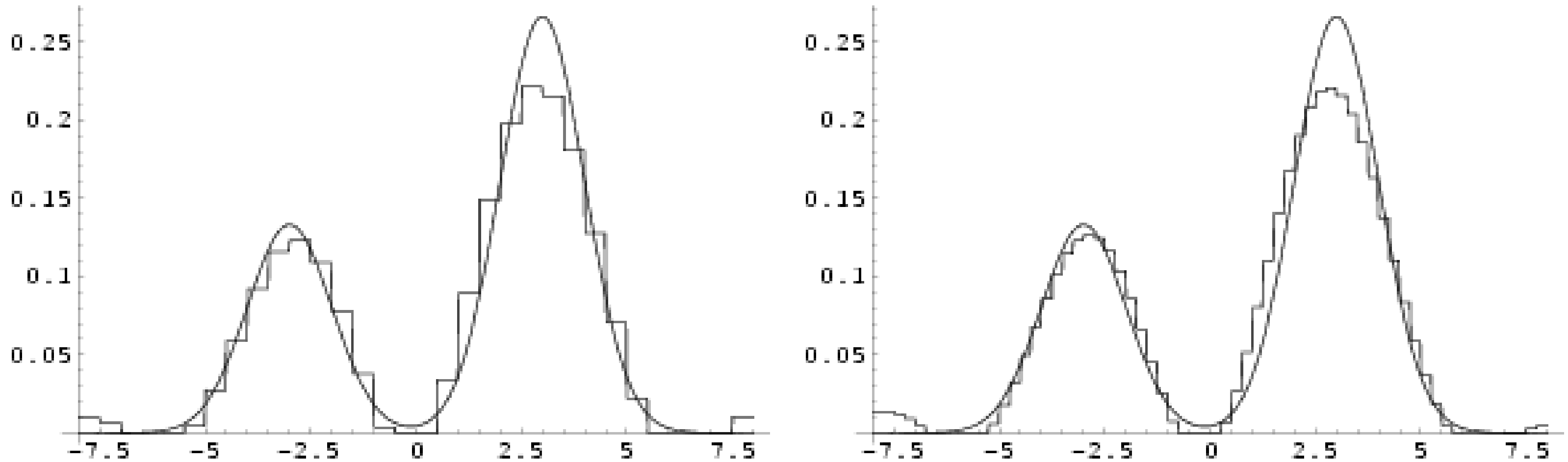
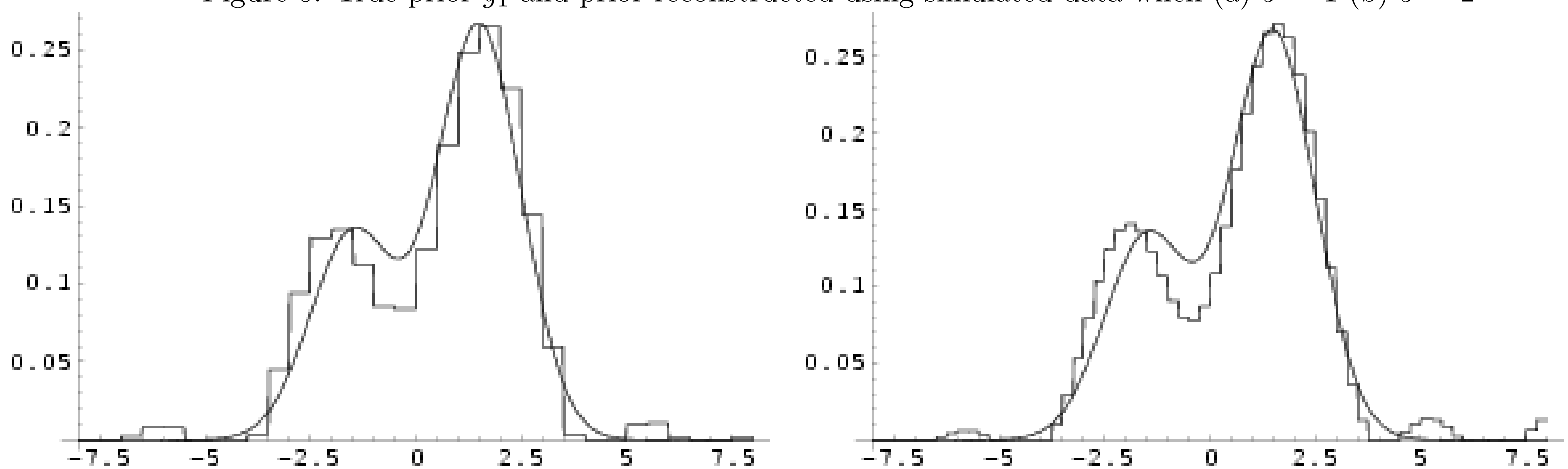


Figure 4: True marginal density m_2 and normalized histogram estimate of m_2 constructed from simulated data when $J = 1$



(a) (b)

Figure 5: True prior g_1 and prior reconstructed using simulated data when (a) $J = 1$ (b) $J = 2$



(a) (b)

Figure 6: True prior g_2 and prior reconstructed using simulated data when (a) $J = 1$ (b) $J = 2$

9 Appendix

Proof of Lemma 1 Applying Theorem 1 to equation (9), we first get

$$
\begin{aligned}
d_{Jk} &= \int_{\mathbb{R}} \left(\widetilde{m}_J(x) + O\big(2^{-2J}\big)\right) \phi_{Jk}(x)\,dx \\
&= < \widetilde{m}_J,\, \phi_{Jk} > + O\big(2^{-5J/2}\big),
\end{aligned}
\tag{19}
$$

since

$$
\int_{\mathbb{R}} \phi_{Jk}(x)\,dx = 2^{J/2} \int_{k/2^J}^{(k+1)/2^J} dx = 2^{-J/2}.
$$

The scalar product in equation (19) can be simplified using the fact that ϕ is a step function. Indeed, we have that

$$
\begin{aligned}
< \widetilde{m}_J,\, \phi_{Jk} > &= \int_{\mathbb{R}} \widetilde{m}_J(x)\phi_{Jk}(x)\,dx \\
&= 2^{J/2} \int_{k/2^J}^{(k+1)/2^J} \widetilde{m}_J(x)\,dx.
\end{aligned}
$$

Now, since $\widetilde{m}_J \in C^4(\mathbb{R})$, we can use Simpson's integration rule (*cf.* Davis and Rabinowitz, 1984, section 2.2) to get

$$
\int_{k/2^J}^{(k+1)/2^J} \widetilde{m}_J(x)\,dx = \frac{2^{-J}}{6}\left[\widetilde{m}_J\left(\frac{k}{2^J}\right) + 4\,\widetilde{m}_J\left(\frac{2k+1}{2^{J+1}}\right) + \widetilde{m}_J\left(\frac{k+1}{2^J}\right)\right] + O\big(2^{-5J}\big),
$$

and, consequently,

$$
< \widetilde{m}_J,\, \phi_{Jk} > = \frac{2^{-J/2}}{6}\left[\widetilde{m}_J\left(\frac{k}{2^J}\right) + 4\,\widetilde{m}_J\left(\frac{2k+1}{2^{J+1}}\right) + \widetilde{m}_J\left(\frac{k+1}{2^J}\right)\right] + O\big(2^{-9J/2}\big).
$$

Replacing the previous equality in equation (19) leads to

$$
d_{Jk} = \frac{2^{-J/2}}{6}\left[\widetilde{m}_J\left(\frac{k}{2^J}\right) + 4\,\widetilde{m}_J\left(\frac{2k+1}{2^{J+1}}\right) + \widetilde{m}_J\left(\frac{k+1}{2^J}\right)\right] + O\big(2^{-5J/2}\big),
$$

which can also be written, using equation (4), as

$$
d_{Jk} = \frac{2^{-J/2}}{6}\sum_{l=-N}^{N} b_{Jl}\left[a_{Jl}\left(\frac{k}{2^J}\right) + 4\,a_{Jl}\left(\frac{2k+1}{2^{J+1}}\right) + a_{Jl}\left(\frac{k+1}{2^J}\right)\right] + O\big(2^{-5J/2}\big).
$$

Finally, to obtain the wanted result, note that using equation (8), we can write

$$
a_{Jl}\left(\frac{k}{2^J}\right) = 2^{J/2}\left[F\left(\frac{k-l}{2^J}\right) - F\left(\frac{k-l-1}{2^J}\right)\right] = a_{J0}\left(\frac{k-l}{2^J}\right).
$$

□

Proof of Lemma 2 First, note that

$$\phi_{Jk}(X_i) = 2^{J/2}\, I\!I_{A_{Jk}}(X_i),$$

where $I\!I_{A_{Jk}}$ is the indicator function of the interval $A_{Jk} = [k/2^J,\, k+1/2^J)$. Thus, from equation (13), we have that

$$\hat{d}_{Jk} = \frac{2^{J/2}}{n} \sum_{i=1}^{n} I\!I_{A_{Jk}}(X_i),$$

and hence,

$$\frac{n}{2^{J/2}} \hat{\mathrm{d}} = \left(\sum_{i=1}^{n} I\!I_{A_{Jk}}(X_i) \right)_{k \in \mathcal{N}},$$

which corresponds to the vector of observed counts for each interval A_{Jk}, $k \in \mathcal{N}$. Hence, we have indeed that

$$n 2^{-J/2} \hat{\mathrm{d}} \sim \mathcal{M}\left(n, 2N+1, 2^{-J/2}\mathrm{d}\right).$$

The other results are a direct consequence of this fact. □

Proof of Theorem 4 To get the first result, note that from Corollary 1 and Theorem 2, we have

$$\begin{aligned}
\mathbb{E}\big[\hat{g}_J(s)\big] &= \big(6\mathrm{B}^{-1}\mathrm{d}\big)^t\, \Phi(s) \\
&= \big(\mathrm{b} + O\big(2^{-J(5/2-\delta)}\big) 1_{2N+1}\big)^t\, \Phi(s).
\end{aligned}$$

Hence, from equation (18) and noticing that, for $s \in \mathbb{R}$,

$$1_{2N+1}^t \Phi(s) = \sum_{k=-N}^{N} \phi_{Jk}(s) \le \sum_{k \in \mathbb{Z}} \phi_{Jk}(s) = 2^{J/2},$$

we can write

$$\begin{aligned}
\mathbb{E}\big[\hat{g}_J(s)\big] &= g_J^*(s) + O\big(2^{-J(5/2-\delta)}\big) 1_{2N+1}^t \Phi(s) \\
&= g_J^*(s) + O\big(2^{-J(2-\delta)}\big).
\end{aligned} \tag{20}$$

To prove the second result, note that Theorems 2 and 3 can be combined to give

$$\hat{\mathrm{b}} \xrightarrow{P} \mathrm{b} + O\big(2^{-J(5/2-\delta)}\big) 1_{2N+1}.$$

Using the same arguments as above, we then get that

$$\hat{g}_J(s) = \hat{\mathrm{b}}^t \Phi(s) \xrightarrow{P} \big(\mathrm{b} + O\big(2^{-J(5/2-\delta)}\big) 1_{2N+1}\big)^t\, \Phi(s) = g_J^*(s) + O\big(2^{-J(2-\delta)}\big).$$

Finally, to obtain the last result, recall that

$$\mathrm{MSE}\big[\hat{g}_J(s)\big] = \mathbb{V}\mathrm{ar}\big[\hat{g}_J(s)\big] + \big(\mathbb{E}\big[\hat{g}_J(s)\big] - g(s)\big)^2.$$

Using Corollary 1 and equation (20), we can then write

$$\begin{aligned}\text{MSE}\big[\hat{g}_J(s)\big] &= 36n^{-1}\Phi^t(s)\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\Phi(s) + \big(g_J^*(s) - g(s) + O\big(2^{-J(2-\delta)}\big)\big)^2\\ &= 36n^{-1}\Phi^t(s)\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\Phi(s) + \big(g_J^*(s) - g(s)\big)^2 + O\big(2^{-J(2-\delta)}\big),\end{aligned}$$

since g is a continuously differentiable density (and thus both g and g_J^* are bounded). Taking the limit when $n \to \infty$ of this last expression leads to the wanted result, that is

$$\text{MSE}\big[\hat{g}_J(s)\big] \underset{n\to\infty}{\longrightarrow} \big(g_J^*(s) - g(s)\big)^2 + O\big(2^{-J(2-\delta)}\big).$$

□

Proof of Theorem 5 We start by noticing that

$$\|\hat{g}_J - g\|_2^2 = \|\hat{g}_J - g_J^*\|_2^2 + \|g_J^* - g\|_2^2 - 2 < \hat{g}_J - g_J^*,\, g_J^* - g > .$$

But, from Definition 1 and equation (18), we have that

$$\hat{g}_J(s) - g_J^*(s) = \sum_{k\in\mathcal{N}} \big(\hat{b}_{Jk} - b_{Jk}\big)\, \phi_{Jk}(s), \tag{21}$$

and, from equations (2) and (18), we have that

$$g_J^*(s) - g(s) = \sum_{k\notin\mathcal{N}} b_{Jk}\phi_{Jk}(s) + \sum_{j\geq J}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(s), \tag{22}$$

where the coefficients β_{jk} are given by

$$\beta_{jk} = < g,\, \psi_{jk} > .$$

Also, since

$$\Big\{\phi_{Jk}, \psi_{jk} : j \geq J, k \in \mathbb{Z}\Big\}$$

forms an orthonormal basis of $\mathcal{L}^2(\mathbb{R})$, it is clear that for $k_1 \neq k_2$,

$$< \phi_{Jk_1},\, \phi_{Jk_2} >= 0,$$

and that

$$< \phi_{Jk_1},\, \psi_{jk_2} >= 0,$$

for all $j \geq J$ and $k_1,\, k_2 \in \mathbb{Z}$. When combined with equations (21) and (22), this implies that

$$< \hat{g}_J - g_J^*,\, g_J^* - g >= 0,$$

and that

$$\begin{aligned}\|\hat{g}_J - g_J^*\|_2^2 &= \sum_{k\in\mathcal{N}} \big(\hat{b}_{Jk} - b_{Jk}\big)^2\\ &= \big(\hat{\mathrm{b}} - \mathrm{b}\big)^t\big(\hat{\mathrm{b}} - \mathrm{b}\big).\end{aligned}$$

Hence, the following fundamental property can be obtained,

$$\|\hat{g}_J - g\|_2^2 = \|g_J^* - g\|_2^2 + \big(\hat{\mathrm{b}} - \mathrm{b}\big)^t\big(\hat{\mathrm{b}} - \mathrm{b}\big). \tag{23}$$

Now, using Theorem 3 we can write

$$\big(\hat{\mathrm{b}} - \mathrm{b}\big)^t\big(\hat{\mathrm{b}} - \mathrm{b}\big) \overset{P}{\longrightarrow} \big(\mathrm{b}^* - \mathrm{b}\big)^t\big(\mathrm{b}^* - \mathrm{b}\big).$$

On the other hand, from Theorem 2, we have that

$$\begin{aligned}\big(\mathrm{b}^* - \mathrm{b}\big)^t\big(\mathrm{b}^* - \mathrm{b}\big) &= 1_{2N+1}^t 1_{2N+1} O\big(2^{-2J(5/2-\delta)}\big) \\ &= (2N+1) O\big(2^{-2J(5/2-\delta)}\big) = O\big(2^{-2J(2-\delta-1/\varepsilon)}\big),\end{aligned} \tag{24}$$

since $N \propto 2^{-J(1+2/\varepsilon)}$. Using this along with equation (23), we see that

$$\mathrm{ISE}\big[\hat{g}_J\big] \overset{P}{\longrightarrow} \|g_J^* - g\|_2^2 + O\big(2^{-2J(2-\delta-1/\varepsilon)}\big),$$

which corresponds to the first result.

Also, note that from Theorem 3

$$\begin{aligned}\mathbb{E}\left[\big(\hat{\mathrm{b}} - \mathrm{b}\big)^t\big(\hat{\mathrm{b}} - \mathrm{b}\big)\right] &= \mathrm{tr}\left(\mathbb{C}\mathrm{ov}\big[\hat{\mathrm{b}}\big]\right) + \big(\mathrm{b}^* - \mathrm{b}\big)^t\big(\mathrm{b}^* - \mathrm{b}\big) \\ &= 36n^{-1}\mathrm{tr}\left(\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\right) + \big(\mathrm{b}^* - \mathrm{b}\big)^t\big(\mathrm{b}^* - \mathrm{b}\big)\end{aligned}$$

Hence, from equation (24), we get the following alternate expression for the previous expectation,

$$\mathbb{E}\left[\big(\hat{\mathrm{b}} - \mathrm{b}\big)^t\big(\hat{\mathrm{b}} - \mathrm{b}\big)\right] = 36n^{-1}\mathrm{tr}\left(\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\right) + O\big(2^{-2J(2-\delta-1/\varepsilon)}\big).$$

Combining this with equation (23), it becomes clear that

$$\mathrm{MISE}\big[\hat{g}_J\big] = \|g_J^* - g\|_2^2 + 36n^{-1}\mathrm{tr}\left(\mathrm{B}^{-1}\Sigma(\mathrm{B}^{-1})^t\right) + O\big(2^{-2J(2-\delta-1/\varepsilon)}\big),$$

so that, taking the limit when $n \to \infty$, we finally have

$$\mathrm{MISE}\big[\hat{g}_J\big] \underset{n\to\infty}{\longrightarrow} \|g_J^* - g\|_2^2 + O\big(2^{-2J(2-\delta-1/\varepsilon)}\big),$$

which establishes the second and last result. □

10 References

Daubechies, I. (1992), *Ten lectures on wavelets.* SIAM, Philadelphia.

Davis, P.J. et Rabinowitz, P. (1984), *Methods of numerical integration*, 2nd edition. Academic Press, Orlando.

Efron, B. et Morris, C.N. (1972), Limiting the risk of Bayes and empirical Bayes estimators - Part II: the empirical Bayes case. *Journal of the American Statistical Association*, Vol. 67, p. 130-139.

Fan, J. (1991a), Global behavior of deconvolution kernel estimates. *Statistica Sinica*, Vol. 1, p. 541-551.

Fan, J. (1991b), On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, Vol. 19, p. 1257-1272.

Goutis, C. (1997), Nonparametric estimation of a mixing density via the kernel method. *Journal of the American Statistical Association*, Vol. 92, p. 1445-1450.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A.B. (1998), *Wavelets, approximations, and statistical applications*. Springer-Verlag, New York.

Healy, D.M. and Kim, P.T. (1996), An empirical Bayes approach to directional data and efficient computation on the sphere. *Annals of Statistics*, Vol. 24, p. 232-254.

Izenman, A.J. (1991), Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, Vol. 86, p. 205-224.

Laird, N.M. (1978), Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, Vol. 73, p. 805-811.

Laird, N.M. et Louis, T.A. (1991), Smoothing the nonparametric estimate of a prior distribution by roughening: a computational study. *Computational Statistics and Data Analysis*, Vol. 12, p. 27-37.

Leblanc, A. (2002), *Utilisation de la base de Haar en estimation bayésienne*. Ph.D. Thesis, Université de Montréal.

Leblanc, A. and Angers, J.-F. (1999), Bayesian estimation of a location parameter using the Haar basis. *Bayesian Statistics* 6 (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith eds.), p. 803-812, Oxford University Press, Oxford.

Mallat, S. (1989), Multiresolution approximations and wavelet orthonormal bases of $\mathcal{L}^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, Vol. 315, p. 69-88.

Mallat, S. (1998), *A wavelet tour of signal processing*. Academic Press, San Diego.

Morris, C.N. (1983), Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, Vol. 78, p. 47-65.

Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, Vol. 27, p. 832-837.

Vidakovic, B. (1999), *Statistical modelling by wavelets*. Wiley, New York.

Zhang, C.H. (1990), Fourier methods for estimating mixing densities and distributions. *Annals of Statistics*, Vol. 18, p. 806-831.