Working Paper Series
Department of Economics
University of Verona

On Deconvolution of Dirichlet-Laplace Mixtures

Catia Scricciolo

WP Number: 6 May 2016

ISSN: 2036-2919 (paper), 2036-4679 (online)

On Deconvolution of Dirichlet-Laplace Mixtures

Sulla Deconvoluzione di Modelli Miscuglio Dirichlet-Laplace

Catia Scricciolo*

Abstract We consider the problem of Bayesian density deconvolution, when the mixing density is modelled as a Dirichlet-Laplace mixture. We give an assessment of the posterior accuracy in recovering the true mixing density, when it is itself a Laplace mixture. The results partially complement those in Donnet *et al.* (2014) and an application of them to empirical Bayes density deconvolution can be envisaged.

Abstract *Si considera il problema della stima bayesiana della densità misturante in modelli miscuglio di tipo Dirichlet-Laplace. Vengono presentati alcuni risultati sull'accuratezza della legge finale nell'approssimare la densità della "vera" distribuzione misturante, quando sia essa stessa un miscuglio di Laplace. Questi risultati completano parzialmente quelli di Donnet et al. (2014) e se ne può intravedere un'applicazione al problema della deconvoluzione empirico-bayesiana di densità.*

Key words: contraction rates, deconvolution, Dirichlet-Laplace mixtures

1 Introduction

We consider the *inverse* problem of making inference on the *mixing* density in location mixtures, given draws from the *mixed* density. Consider the model

$$Y = X + \varepsilon,$$

where X and ε are independent random variables (or rv's in short). Let f_{0X} denote the Lebesgue density on $\mathbb{R}$ of X and f_ε the Lebesgue density on $\mathbb{R}$ of the error ε. The density of Y, denoted by f_{0Y}, is then the convolution

* Department of Economics, University of Verona, Polo Universitario Santa Marta, Via Cantarane 24, I-37129 Verona, ITALY, e-mail: catia.scricciolo@univr.it

$$f_{0Y}(\cdot) = (f_{0X} * f_\varepsilon)(\cdot) = \int f_\varepsilon(\cdot - u) f_{0X}(u)\,\mathrm{d}u.$$

In what follows, integrals where no limits are indicated have to be understood to be over the entire domain of integration. The error density f_ε is assumed to be known, while f_{0X} is unknown. Suppose we have observations $Y^{(n)} := (Y_1, \ldots, Y_n)$ from f_{0Y}, rather than from f_{0X}, and we want to recover f_{0X}.

Inverse problems arise when the object of interest is only indirectly observed. The deconvolution problem arises in a wide variety of applications, where the error density is typically modelled using the Gaussian kernel, but also Laplace mixtures have found relevant applications in speech recognition and astronomy, cf. [4]. Full density deconvolution, together with the related many normal means estimation problem, has drawn attention in the literature since the late 1950's and different estimation methods have been developed since then, especially taking the frequentist approach, the most popular being based on nonparametric maximum likelihood and on deconvolution kernel density estimators. This problem has only recently been studied from a Bayesian perspective: the typical scheme considers the mixing distribution generated from a Dirichlet process or the mixing density generated from a Dirichlet process mixture of a fixed kernel density. Posterior contraction rates for the mixing distribution in Wasserstein metrics have been investigated in [3, 6]: convergence in Wasserstein metrics for discrete mixing distributions has a natural interpretation in terms of convergence of the single atoms supporting the probability measures. Adaptive recovery rates for deconvolving a density f_{0X} in a Sobolev space, which automatically rules out the possibility that the density is a Laplace mixture, in the cases when the error density is either ordinary or super-smooth, are derived in [2] for the fully Bayes and empirical Bayes approaches, the latter employing a data-driven choice of the prior hyper-parameters of the Dirichlet process base measure. Recently, [5] have proposed a nonparametric method for density deconvolution based on a two-step "bin and smooth" procedure: the first step consists in using the sample observations to construct a histogram with equally sized bins, the second step considers the number of observations falling into each bin to form a proxy Poisson likelihood and derive a maximum *a posteriori* estimate of the mixing density f_{0X} under a prior supported on a space of smooth densities. This procedure has proven to have an excellent performance with reduced computational costs when compared to fully Bayes methods.

In this note, we consider Bayesian density deconvolution, when the mixing density f_X is modelled as a Dirichlet-Laplace mixture. In Sect. 2, using some inversion inequalities, developed in Lemma 2 reported in the Appendix, which can also be of independent interest, we give an assessment of the posterior accuracy in recovering the true density f_{0X}, when it is itself a Laplace mixture. Some final remarks and comments are exposed in Sect. 3. Auxiliary results are reported in the Appendix.

2 Results

In this section, we present some results on Bayesian density deconvolution when the mixing density f_X is modelled as a Dirichlet-Laplace mixture. We consider densities f_X of the form $f_{F,\sigma_0}(\cdot) \equiv (F * \psi_{\sigma_0})(\cdot) = \int \psi_{\sigma_0}(\cdot - \theta)\,\mathrm{d}F(\theta)$, where $\psi_{\sigma_0}(x) = (2\sigma_0)^{-1}e^{-|x|/\sigma_0}$, $x \in \mathbb{R}$, is a *Laplace* with scale parameter $0 < \sigma_0 < +\infty$ and F is any distribution on $\mathbb{R}$. As a prior for F, we adopt a Dirichlet process with base measure γ, denoted by $\mathfrak{D}_\gamma$. A Dirichlet process on a measurable space $(\mathfrak{X}, \mathscr{A})$, with finite and positive base measure γ on $(\mathfrak{X}, \mathscr{A})$, is a random probability measure F on $(\mathfrak{X}, \mathscr{A})$ such that, for every finite partition $(A_1, \ldots, A_k)$ of $\mathfrak{X}$, the vector of probabilities $(F(A_1), \ldots, F(A_k))$ has Dirichlet distribution with parameters $(\gamma(A_1), \ldots, \gamma(A_k))$. A Dirichlet process mixture of Laplace densities can be described as follows:

- $F \sim \mathfrak{D}_\gamma$,
- given F, the rv's $\theta_1, \ldots, \theta_n$ are iid F,
- given $(F, \theta_1, \ldots, \theta_n)$, the rv's $e_1, \ldots, e_n$ are iid ψ_{σ_0},
- sampled values from f_X are defined as $X_i := \theta_i + e_i$, for $i = 1, \ldots, n$.

We are interested in recovering the density f_{0X} when the observations $Y^{(n)}$ are sampled from $f_{0Y} = f_{0X} * f_\varepsilon$, where f_ε is given. Assuming that f_{0X} is itself a location mixture of Laplace densities, $f_{0X} = f_{G_0,\sigma_0} = (G_0 * \psi_{\sigma_0})$, we approximate f_{0Y} with the random density $f_Y = f_X * f_\varepsilon = f_{F,\sigma_0} * f_\varepsilon$, where $f_X \equiv f_{F,\sigma_0}$ is modelled as a Dirichlet-Laplace mixture. In order to qualify the regularity of f_ε, we preliminarily recall some definitions that will be used in Proposition 2 and in Lemmas 1, 2 of the Appendix. For real $1 \le p < \infty$, let $L^p(\mathbb{R}) := \{f \,|\, f : \mathbb{R} \to \mathbb{R},\, f \text{ is Borel measurable}, \int |f|^p\,\mathrm{d}\lambda < \infty\}$, where λ denotes Lebesgue measure. For $f \in L^p(\mathbb{R})$, the integral $\int |f|^p\,\mathrm{d}\lambda$ will also be written as $\int |f(x)|^p\,\mathrm{d}x$.

Definition 1. For $f \in L^1(\mathbb{R})$, the complex-valued function $\hat{f}(t) := \int e^{itx} f(x)\,\mathrm{d}x$, $t \in \mathbb{R}$, is called the *Fourier transform* (or *Ft* in short) *of* f.

The relationship between Fourier transforms and characteristic functions (or cfs in short) is revised for later use. For a rv X with cdf F_X and cf $\phi_X(t) := \mathbb{E}[e^{itX}]$, $t \in \mathbb{R}$, if F_X is absolutely continuous, then the distribution of X has Lebesgue density, say f_X, and the cf of X coincides with the Ft of f_X, that is, $\phi_X(t) = \int e^{itx} f(x)\,\mathrm{d}x = \hat{f}_X(t)$, $t \in \mathbb{R}$. As in [7, 8], we adopt the following definition.

Definition 2. Let f be a Lebesgue probability density on $\mathbb{R}$. The Fourier transform of f is said to *decrease algebraically of degree* $\beta > 0$ if

$$\lim_{|t|\to+\infty} |t|^\beta |\hat{f}(t)| = B_f, \quad 0 < B_f < +\infty. \tag{1}$$

Recalling that for real-valued functions f and g, the notation $f \sim g$ means that $f/g \to 1$ in an asymptotic regime that is clear from the context, relationship (1) can be rewritten as $B_f^{-1}|\hat{f}(t)| \sim |t|^{-\beta}$ as $|t| \to \infty$, a notation adopted in what follows.

Remark 1. We now exhibit some examples of distributions satisfying condition (1):

- a gamma distribution with shape and scale parameters $\nu, \lambda > 0$, respectively, which has cf $(1 + it/\lambda)^{-\nu}$;

- the distribution with cf $(1+|t|^{\alpha})^{-1}$, $t \in \mathbb{R}$, for $0 < \alpha \le 2$, which is called α-*Laplace distribution*, also termed *Linnik's distribution*. The case $\alpha = 2$ renders the cf of the Laplace distribution;
- if S_α is a symmetric stable distribution with cf $e^{-|t|^{\alpha}}$, $0 < \alpha \le 2$, and V_β is an independent rv with density $\Gamma(1+1/\beta)e^{-v^{\beta}}$, $v > 0$, then the rv $S_\alpha V_\beta^{\beta/\alpha}$ has cf $(1+|t|^{\alpha})^{-1/\beta}$. For $\beta = 1$, this reduces to the cf of an α-Laplace distribution.

Let $\Pi(\cdot \mid Y^{(n)})$ denote the posterior distribution of a Dirichlet-Laplace mixture prior for f_X. Assuming that $Y^{(n)}$ is a sample of iid observations from $f_{0Y} = (G_0 * \psi_{\sigma_0}) * f_\varepsilon$, we study the capability of the posterior distribution to accurately recover f_{0X}. We write "$\lesssim$" and "$\gtrsim$" for inequalities valid up to a constant multiple that is universal or fixed within the context, but anyway inessential for our purposes.

Proposition 1. *Let $f_{0Y} = (G_0 * \psi_{\sigma_0}) * f_\varepsilon$, with G_0 and f_ε compactly supported on an interval $[-a, a]$. Consider a Dirichlet-Laplace mixture prior for f_X, with base measure γ compactly supported on $[-a, a]$ that possesses Lebesgue density bounded away from 0 and ∞. Then, for sufficiently large constant M_p, with $1 \le p \le 2$, we have $\Pi(\|f_X - f_{0X}\|_p > M_p(n/\log n)^{-3/8} \mid Y^{(n)}) \to 0$ in $P^{(n)}_{f_{0Y}}$-probability.*

The result follows from (2.2) for $p = 1$ and from (2.3) for $p = 2$ of Theorem 2 in [3]. The open question concerns recovery rates when f_ε is not compactly supported.

Proposition 2. *Let $f_{0Y} = (G_0 * \psi_{\sigma_0}) * f_\varepsilon$, with f_ε having cf that either (i) decreases algebraically of degree 2 or (ii) satisfies $|\hat{f}_\varepsilon(t)| \gtrsim e^{-\rho|t|^r}$, $t \in \mathbb{R}$. If for sequences $\varepsilon_{n,p} \downarrow 0$ such that $n\varepsilon_{n,p}^2 \to \infty$ and sufficiently large constants M_p, with $p = 1, 2$, we have $\Pi(\|f_Y - f_{0Y}\|_p > M_p\varepsilon_{n,p} \mid Y^{(n)}) \to 0$ in $P^{(n)}_{f_{0Y}}$-probability, then also $\Pi(\|f_X - f_{0X}\|_2 > \eta_{n,p} \mid Y^{(n)}) \to 0$ in $P^{(n)}_{f_{0Y}}$-probability, where $\eta_{n,2} \propto \varepsilon_{n,2}^{3/7}$ for $p = 2$ and $\hat{f}_\varepsilon$ as in (i), and $\eta_{n,1} \propto (-\log \varepsilon_{n,1})^{-3/2r}$ for $p = 1$ and $\hat{f}_\varepsilon$ as in (ii).*

The assertion follows from Lemma 2 in case (i) for $\alpha = \beta = 2$ and from Remark 4 in case (ii) for $\beta = 2$. The open problem remains the derivation of posterior contraction rates $\varepsilon_{n,p}$ for a general true mixing distribution G_0.

3 Final remarks

In this note, we have considered the density deconvolution problem, when the mixing density f_X is modelled as a Dirichlet-Laplace mixture. The problem arises in a variety of contexts and Laplace mixtures, as an alternative to Gaussian mixtures, have relevant applications. Propositions 1 and 2 give an assessment of the quality of the recovery of the true density f_{0X} when it is itself a Laplace mixture. The results partially complement that of Proposition 1 in [2] which, even if stated for an empirical Bayes context, is also valid for the fully Bayes density deconvolution problem: yet, only the case when f_{0X} lies in a Sobolev space is therein considered, see (3.9) in [2], which automatically rules out the case of a Laplace (mixture) density. An application of the present results to empirical Bayes density deconvolution, along the lines of Proposition 1 in [2], when the base measure of the Dirichlet process is data-driven selected, can be envisaged.

Appendix

In this section, we report some results that are used to establish the assertion of Proposition 2. The following lemma, which is essentially contained in the first theorem of section 3B in [8], is invoked in Lemma 2 to prove an inversion inequality.

Lemma 1. *Let $f \in \mathbb{L}^2(\mathbb{R})$ be a Lebesgue probability density with cf satisfying condition (1) for $\beta > 1/2$. Let $h \in \mathbb{L}^1(\mathbb{R})$ have Ft $\hat{h}$ so that*

$$I_\beta^2[\hat{h}] := \int \frac{|1-\hat{h}(t)|^2}{|t|^{2\beta}}\,\mathrm{d}t < +\infty. \tag{2}$$

Then, $\delta^{-2(\beta-1/2)}\|f - f * h_\delta\|_2^2 \to B_f^2 \times I_\beta^2[\hat{h}]$ as $\delta \to 0$.

Proof. Since $f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ and $h \in \mathbb{L}^1(\mathbb{R})$, then $\|f * h_\delta\|_p \le \|f\|_p \|h_\delta\|_1 < +\infty$, for $p = 1, 2$. Thus, $(f - f * h_\delta) \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Hence, $\|f - f * h_\delta\|_2^2 = \delta^{2(\beta-1/2)}\{B_f^2 \times I_\beta^2[\hat{h}] + \int |z|^{-2\beta}|1-\hat{h}(z)|^2[|z/\delta|^{2\beta}|\hat{f}(z/\delta)|^2 - B_f^2]\,\mathrm{d}z\}$, where the second integral tends to zero by the dominated convergence theorem (or DCT in short) due to assumption (2). The assertion then follows. ☐

In the following remark, which is due to [1], see Section 3, we consider a sufficient condition for $h \in L^1(\mathbb{R})$ to satisfy requirement (2).

Remark 2. Let $h \in L^1(\mathbb{R})$. Then, $\int_1^{+\infty} t^{-2\beta}|1-\hat{h}(t)|^2\,\mathrm{d}t < +\infty$ for $\beta > 1/2$. If there exists $r \in \mathbb{N}$ such that $\int x^m h(x)\,\mathrm{d}x = 1_{\{0\}}(m)$ for $m = 0, 1, \ldots, r-1$ and $\int x^r h(x)\,\mathrm{d}x \neq 0$, then $t^{-r}[1-\hat{h}(t)] = -t^{-r}[\hat{h}(t)-1] = -t^{-r}\int [e^{itx} - \sum_{j=0}^{r-1}(itx)^j/(j!)]h(x)\,\mathrm{d}x$ so that

$$\frac{[1-\hat{h}(t)]}{t^r} = -\frac{i^r}{(r-1)!}\int x^r h(x) \int_0^1 (1-u)^{r-1} e^{itxu}\,\mathrm{d}u\,\mathrm{d}x \to -\frac{i^r}{r!}\int x^r h(x)\,\mathrm{d}x$$

as $t \to 0$. For $r \ge \beta$, the integral $\int_0^1 t^{-2\beta}|1-\hat{h}(t)|^2\,\mathrm{d}t < +\infty$. Conversely, for $r < \beta$, the integral diverges. Thus, for $1/2 < \beta \le 2$, any symmetric probability density h on $\mathbb{R}$ with finite second moment is such that $I_\beta^2[\hat{h}] < +\infty$ and condition (2) is verified.

The inequality established in the next lemma relates the $\mathbb{L}_2$-distance between a mixing density f_{0X} and an approximating (possibly random) mixing density f_X to the $\mathbb{L}^2$-distance between the corresponding mixed densities $f_{0X} * f_\varepsilon$ and $f_X * f_\varepsilon$.

Lemma 2. *Let $f_{0X}, K \in \mathbb{L}^2(\mathbb{R})$ and f_ε be Lebesgue probability densities on $\mathbb{R}$ with cfs satisfying condition (1) for $1/2 < \beta \le 2$ and α such that $1/2 < \beta \le \alpha \le 2$, respectively. For $f_X \equiv f_{F,\sigma_0} \equiv F * K_{\sigma_0}$, with any cdf F and some fixed $0 < \sigma_0 < +\infty$, we have $\|f_X - f_{0X}\|_2 \lesssim \|(f_X - f_{0X}) * f_\varepsilon\|_2^{(\beta-1/2)/(\alpha+\beta-1/2)}$.*

Proof. We use a Lebesgue probability density h on $\mathbb{R}$ to characterize regular densities in terms of their approximation properties. For $\delta > 0$, let $h_\delta(\cdot) := \delta^{-1}h(\cdot/\delta)$ and define g_δ as the inverse Ft of $\hat{h}_\delta/\hat{f}_\varepsilon$, that is, $g_\delta(x) := (2\pi)^{-1}\int e^{-itx}[\hat{h}_\delta(t)/\hat{f}_\varepsilon(t)]\,\mathrm{d}t$, $x \in \mathbb{R}$. Therefore, the Ft of g_δ is $\hat{g}_\delta = \hat{h}_\delta/\hat{f}_\varepsilon$. Then, $h_\delta = f_\varepsilon * g_\delta$ and $f_X * h_\delta = (f_X * f_\varepsilon) * g_\delta$. We have $\|f_X - f_{0X}\|_2^2 \le \|f_X - f_X * h_\delta\|_2^2 + \|(f_X - f_{0X}) * h_\delta\|_2^2 +$

$\|f_{0X} - f_{0X} * h_\delta\|_2^2 =: T_1 + T_2 + T_3$, where $T_3 \lesssim \delta^{2(\beta-1/2)}$ because, by the assumption on f_{0X} and in virtue of Lemma 1, $(B_{f_{0X}} \times I_\beta[\hat{h}])^{-2}\|f_{0X} - f_{0X} * h_\delta\|_2^2 = (B_{f_{0X}} \times I_\beta[\hat{h}])^{-2}T_3 \sim \delta^{2(\beta-1/2)}$ for h being, for example, the density of any α-Laplace distribution with $1/2 < \beta \leq \alpha \leq 2$. Since $f_X \equiv f_{F,\sigma_0} = F * K_{\sigma_0}$, for every cdf F we have $T_1 \leq \sigma_0^{-2\beta}\delta^{2(\beta-1/2)}\{B_K^2 \times I_\beta^2[\hat{h}] + \int |z|^{-2\beta}|1-\hat{h}(z)|^2[|z\sigma_0/\delta|^{2\beta}|\hat{K}(z\sigma_0/\delta)|^2 - B_K^2]\,\mathrm{d}z\}$, where the second integral converges to zero by the DCT. Thus, $T_1 \lesssim \delta^{2(\beta-1/2)}$. Concerning the term T_2, we have $T_2 \leq \sup_{t\in\mathbb{R}}[|\hat{h}_\delta(t)|/|\hat{f}_\varepsilon(t)|]^2 \int |\hat{f}_X(t) - \hat{f}_{0X}(t)|^2|\hat{f}_\varepsilon(t)|^2\,\mathrm{d}t \lesssim \delta^{-2\alpha}\|(f_X - f_{0X}) * f_\varepsilon\|_2^2$. Combining partial results, we have $\|f_X - f_{0X}\|_2^2 \lesssim \delta^{2(\beta-1/2)} + \delta^{-2\alpha}\|(f_X - f_{0X}) * f_\varepsilon\|_2^2$. Choosing δ to minimize the right-hand side (or RHS in short) of the last inequality, we get $\delta^* = O(\|(f_X - f_{0X}) * f_\varepsilon\|_2^{1/(\alpha+\beta-1/2)})$, which yields the final assertion, thus completing the proof. □

Some remarks on Lemma 2 are in order.

Remark 3. Lemma 2 still holds if, as for the density f_X, it is assumed that f_{0X} is the convolution of a distribution G_0 and a kernel density $\tilde{K}$, that is, $f_{0X} = G_0 * \tilde{K}$, with the cf of $\tilde{K}$ satisfying condition (1) for $1/2 < \beta \leq 2$.

Remark 4. If the error density f_ε has cf such that $|\hat{f}_\varepsilon(t)| \gtrsim e^{-\rho|t|^r}$, $t \in \mathbb{R}$, for some constants $0 < \rho < +\infty$ and $0 < r \leq 2$, then the assertion of Lemma 2 becomes $\|f_X - f_{0X}\| \lesssim (-\log\|(f_X - f_{0X}) * f_\varepsilon\|_1)^{-(\beta-1/2)/r}$, where the L^1-distance is employed instead of the L^2-distance. The result can be proved by considering any symmetric density h, with finite second moment, such that its cf is compactly supported, say, on $[-1, 1]$. Then, $T_2 \leq \|(f_X - f_{0X}) * f_\varepsilon\|_1^2\|g_\delta\|_2^2$, where $\|g_\delta\|_2^2 \lesssim e^{\rho'\delta^r}$ for every $\rho' > 2\rho$. So, $\|f_X - f_{0X}\|_2 \lesssim \delta^{2(\beta-1/2)} + e^{\rho'\delta^r}\|(f_X - f_{0X}) * f_\varepsilon\|_1^2$. The optimal choice for δ yielding the result arises from minimizing the RHS of the last inequality.

References

[1] Davis, K.B.: Mean integrated square error properties of density estimates. Ann. Statist. **5**(3), 530–535 (1977)

[2] Donnet, S., Rivoirard, V., Rousseau, J., Scricciolo, C.: Posterior concentration rates for empirical Bayes procedures, with applications to Dirichlet process mixtures. http://arxiv.org/pdf/1406.4406v1.pdf (2014)

[3] Gao, F., van der Vaart, A.: Posterior contraction rates for deconvolution of Dirichlet-Laplace mixtures. Electron. J. Statist. **10**(1), 608–627 (2016)

[4] Kotz, S., Kozubowski, T.J., Podgorski, K.: The Laplace distribution and generalizations. A revisit with applications to communications, economics, engineering, and finance. Birkhäuser, Boston (2001)

[5] Madrid-Padilla, O.-H., Polson, N.G., Scott, J.G.: A decovolution path for mixtures. http://arxiv.org/pdf/1511.06750v1.pdf (2015)

[6] Nguyen, X.: Convergence of latent mixing measures in finite and infinite mixtures models. Ann. Statist. **41**(1), 370–400 (2013)

[7] Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist. **33**(3), 1065–1076 (1962)

[8] Watson, G.S., Leadbetter, M.R.: On the estimation of the probability density, I. Ann. Math. Statist. **34**(2), 480–491 (1963)