

Posterior rates of convergence for Dirichlet mixtures of exponential power densities

Catia Scricciolo

Department of Decision Sciences

Luigi Bocconi University

Via Röntgen 1, I-20136 Milan, Italy

e-mail: catia.scricciolo@unibocconi.it

Abstract: A Dirichlet mixture of exponential power distributions, as a prior on densities supported on the real line in the problem of Bayesian density estimation, is a natural generalization of a Dirichlet mixture of normals, which has been shown to possess good frequentist asymptotic properties in terms of posterior consistency and rates of convergence. In this article, we establish upper bounds on the rates of convergence for the posterior distribution of a Dirichlet mixture of exponential power densities, assuming that the true density has the same form as the model. When the kernel is analytic and the mixing distribution has either compact support or sub-exponential tails, a nearly parametric rate, up to a logarithmic factor whose exponent depends on the tail behaviour of the base measure of the Dirichlet process and the exponential decay rate at zero of the prior for the scale parameter, is obtained. The result covers the important special case where the true density is a location mixture of normals and shows that a nearly parametric rate arises also when the prior on the scale contains zero in its support, provided it has a sufficiently fast decay rate at zero. This improves on some recent results on density estimation with Dirichlet mixtures of normals by allowing the inverse-gamma distribution, which is a commonly used prior on the square of the bandwidth. When the kernel is not infinitely differentiable at zero, as the case may be depending on the shape parameter, the posterior distribution is shown to concentrate around the sampling density at a slower rate.

AMS 2000 subject classifications: Primary 62G07, 62G20.

Keywords and phrases: Bayesian density estimation, mixture of exponential power distributions, posterior consistency, posterior distribution, rate of convergence, sieve.

Received April 2010.

Contents

| | | |
|---|-------------------------|-----|
| 1 | Introduction | 271 |
| 2 | Preliminaries | 274 |

*This research was mainly supported by grants from Luigi Bocconi University and the Italian Ministry of Education, University and Research (MIUR Grant No. 2008MK3AFZ_003), research project entitled “Bayesian Methods: Theoretical Developments and Novel Applications”.

3 Finite mixtures 275
 4 Infinite mixtures 283
 4.1 Analytic kernels 285
 4.2 Non-analytic kernels 294
 5 Final remarks 298
 Appendix: Auxiliary results 298
 Acknowledgements 306
 References 306

1. Introduction

Mixtures of probability distributions naturally arise in some contexts as models for observations of phenomena with multiple underlying factors. In a Bayesian framework, mixture models provide convenient constructs for density estimation because a prior can be induced on a space of densities by putting a prior on the mixing distribution of a given collection of kernels. This approach, which has the virtue of combining conceptual simplicity of the scheme with flexibility of the model due to the wide range of possible kernel choices depending on the sample space, was initiated by Ferguson [5] and Lo [17], who used a Dirichlet process prior on the mixing distribution and derived the expressions for the posterior and the Bayes’ density estimator (or predictive distribution). The difficulty in defining a prior on a set of densities using directly a Dirichlet process lies in the fact that this process selects discrete distributions. A typical choice for the kernel when the sample space is the real line is the normal density, but an exponential power density can be more generally used. The motivation for considering mixtures of other distributions than the normal lies in the fact that the empirical distributions of many phenomena fail to conform to a normal. Exponential power distributions may represent a reasonable alternative when the discrepancy is in the tails. For example, the normal-Laplace distribution, which results from the convolution of independent normal and Laplace components, behaves like the normal in the middle of its range and like the Laplace in the tails. The interest in this distribution is due to its role in describing the stopped rate of a Brownian motion, when the starting value is generated by a normal and the hazard stopping rate is constant. Its use in the study of high frequency price data is pointed out in Reed [19]. The normal-Laplace distribution makes an interesting case for considering mixtures of exponential power distributions because, as it will be shown, available results on normal mixtures do not cover the case where the mixing distribution has fatter tails than those of the kernel.

An *Exponential Power* (EP) distribution has probability density function

$$f(x; \theta, \sigma, p) = \frac{1}{2\sigma p^{1/p}\Gamma(1 + 1/p)} \exp\{-(|x - \theta|/\sigma)^p/p\}, \quad x \in \mathbb{R}, \quad (1.1)$$

with $\theta \in \mathbb{R}$ and $\sigma, p > 0$. In the sequel, we shall use the symbol c_p to denote the normalizing constant, *i.e.*, $c_p := 2p^{1/p}\Gamma(1 + 1/p)$, and let $\psi_{\sigma,p}(\cdot)$ stand for the

density function of an EP distribution with $\theta = 0$,

$$\psi_{\sigma,p}(x) := f(x; 0, \sigma, p), \quad x \in \mathbb{R}.$$

An EP distribution is characterized by three parameters: the *location* parameter $\theta = \mathbb{E}[X]$, the *scale* parameter $\sigma \equiv \sigma_p = \{\mathbb{E}[|X - \theta|^p]\}^{1/p}$ and the *shape* parameter (or *exponent*) p that determines the thickness of the tails. Note that $\text{Var}(X) = \sigma^2 p^{2/p} \Gamma(3/p) / \Gamma(1/p)$. For $p = 2$, the expression in (1.1) reduces to the density of a normal,

$$\psi_{\sigma,2}(x - \theta) = \phi_{\sigma}(x - \theta) = \sigma^{-1} \phi((x - \theta)/\sigma), \quad x \in \mathbb{R},$$

where $\phi(\cdot)$ stands for the standard normal density. As the shape parameter varies, the EP density describes platikurtic distributions for $p > 2$, namely, distributions with lighter tails than those of a normal, and leptokurtic distributions for $0 < p < 2$. In particular, for $p = 1$, the *Laplace* or *double-exponential* density

$$\psi_{\sigma,1}(x - \theta) = (2\sigma)^{-1} e^{-|x - \theta|/\sigma}, \quad x \in \mathbb{R},$$

is obtained. EP distributions with $0 < p < 1$ are called *fractional* distributions. They have super-Laplace tails, *i.e.*, tails that are beyond those of the Laplace density. For $p \rightarrow 0^+$, the limiting distribution is degenerate at θ . For $p \rightarrow \infty$, the limiting distribution is uniform over the interval $(\theta - \sigma, \theta + \sigma)$,

$$\lim_{p \rightarrow \infty} \psi_{\sigma,p}(x - \theta) = (2\sigma)^{-1} I_{(-1,1]}((x - \theta)/\sigma), \quad x \in \mathbb{R}.$$

The derivation of EP distributions in the above form can be attributed to, among others, Lunetta [18] and Vianelli [26], but EP distributions were first obtained with a different parametrization by Subbotin [24] as a generalization of the Gaussian distribution to model the distribution of random errors. For a review of the main properties of Subbotin's EP distributions see, *e.g.*, Johnson, Kotz and Balakrishnan [12], Chap. 24. EP distributions have been considered as an alternative to the normal distribution in statistical modelling and used to study Bayesian robustness, see, *e.g.*, Choy and Smith [3]. West [30] showed that EP distributions with $p \in [1, 2]$ allow for a scale mixture of normals representation, the normal ($p = 2$) being a degenerate mixture, while Walker and Gutiérrez-Peña [28] proved that EP distributions with $p \geq 1$ allow for a scale mixture of uniforms representation:

$$\psi_{\sigma,p}(x - \theta) = \int_0^{\infty} f_{X|U}(x|u) f_U(u) du, \quad x \in \mathbb{R},$$

where

$$X|(U = u) \sim \text{Uniform} \left(\theta - \frac{\sigma p^{1/p}}{\sqrt{2}} u^{1/p}, \theta + \frac{\sigma p^{1/p}}{\sqrt{2}} u^{1/p} \right)$$

and

$$U \sim \text{Gamma}(1 + 1/p, 2^{-p/2}).$$

In this article, we consider location mixtures of EP distributions. The scale parameter is assumed to be distributed independently of the mixing measure. Thus, a prior on a set of densities is induced by the product measure of the prior for the mixing distribution and the prior for the scale. In what follows, we shall mainly focus on the case where the standard choice of a Dirichlet process prior for the mixing distribution is considered. We are interested in establishing whether and, if so, how fast the posterior distribution concentrates its mass around the true density, as the amount of data increases. Weak and strong posterior consistency, as well as rates of convergence, for Dirichlet location and location-scale mixtures of normals have been studied in the recent literature. Ghosal, Ghosh and Ramamoorthi [7] showed that, under general conditions, a Dirichlet location mixture prior with Gaussian kernel gives rise to a consistent posterior in the total variation distance. Lijoi, Prünster and Walker [16] weakened the condition on the tails of the base measure of the Dirichlet process from an exponential to a power decay law. Tokdar [25] proved strong consistency for Dirichlet location-scale mixtures of normals and weakened a moment condition on the true density (Theorem 3.3, page 96). Ghosal and van der Vaart [8] obtained (upper bounds on the) rates of convergence of the posteriors for Dirichlet location and location-scale normal mixture priors in the case where the true density is itself a location or a location-scale mixture of normals, the mixing distribution having either compact support or sub-Gaussian tails, the scale being restricted to a known bounded interval. They showed that, under regularity conditions on the prior, the posterior converges at a nearly parametric rate, up to a logarithmic term whose exponent depends on the tail behaviour of the base measure. For Dirichlet location mixtures of normals, assuming the same set-up, Walker, Lijoi and Prünster [29] slightly improved the power of the logarithmic term. This fast rate seems to be due to a combination of factors: the form of the true density, which is exactly of the type selected by the prior, together with the infinite degree of smoothness of the Gaussian kernel and the fact that a mixture of normals, with mixing distribution having exponentially decaying tails, can be approximated by a finite mixture of normals with a restricted number of components. This yields a high prior concentration rate and, when the scale is forced to stay bounded away from zero and infinity, a small entropy number, leading to a nearly parametric rate. It may be argued whether the assumption on the scale parameter plays a crucial role in determining the posterior rate or may be removed without affecting the rate, except possibly for a logarithmic factor. If the prior on the scale has full support on $(0, \infty)$, one has to consider a richer approximating class of densities, a sieve, with scale bounded below and above by sequences respectively decreasing to zero and increasing to infinity at proper rates. Employing a fully supported prior for the scale amounts for regularity conditions on the tails, the requirement on the decay rate at zero being expectedly more restrictive than that at infinity, because the most important values are those included in a neighborhood of zero. In fact, as the bandwidth tends to zero, normal mixtures approximate any density and the complexity of the sieve increases.

The focus of the article is on rates of convergence for posterior distributions of Dirichlet EP mixtures. We begin to show that a finite EP mixture can be

estimated at a nearly parametric rate, whatever the value of the kernel shape parameter, using a finite EP mixture prior. Then we consider true densities that may be infinite EP mixtures. First the case where the kernel is analytic and the mixing distribution has either compact support or sub-exponential tails is considered. No constraint is assumed on the scale. We show that a nearly parametric rate arises also when the prior on the scale has full support on $(0, \infty)$. The main result states that, under regularity conditions on the tails of the base measure of the Dirichlet process and the prior for the scale, the posterior of a Dirichlet mixture of EP distributions converges at the target rate. The conditions we impose on the tails of the prior for the scale are satisfied by the common choice of an inverse-gamma prior on σ^2 . This result covers the important special case of Dirichlet normal mixtures, complementing the findings of Ghosal and van der Vaart [8], extending those of Walker, Lijoi and Prünster [29] and improving on those of Scricciolo [21], where a slower rate, heavily depending on the tail decay rate at zero of the prior for the bandwidth, was obtained. We also consider the case where the kernel is not infinitely differentiable at zero. A slower rate is obtained even assuming the scale to lie in a compact interval.

The article is organized as follows. Section 2 describes the set-up and introduces some notation. Section 3 analyzes the case of finite mixtures. Section 4 presents the main results on infinite mixtures. It is split into two subsections: Subsection 4.1 deals with the case where the kernel is analytic, Subsection 4.2 investigates the case of kernels that are not infinitely differentiable at zero. Section 5 is devoted to final remarks and discussion. Auxiliary results invoked in the proofs of the theorems are reported in the Appendix.

2. Preliminaries

Suppose we have observations X_1, \dots, X_n from an unknown density $f_{F, \sigma}$ on \mathbb{R} which is a *location* mixture of EP distributions,

$$f_{F, \sigma}(x) := \int \psi_{\sigma, p}(x - \theta) dF(\theta), \quad x \in \mathbb{R},$$

where F denotes the mixing distribution. In what follows, F will be also used to indicate the corresponding probability measure. Note that $f_{F, \sigma}$ is just the convolution of $\psi_{\sigma, p}$ and F ,

$$f_{F, \sigma} = F * \psi_{\sigma, p}.$$

The scale parameter σ is assumed to be distributed independently of F according to some distribution G on $(0, \infty)$. Let π denote the overall prior on $\mathfrak{M}(\mathbb{R}) \times (0, \infty)$, where $\mathfrak{M}(\mathbb{R})$ is the set of all probability measures on \mathbb{R} . The prior π induces a prior on the class of densities $\mathcal{F} := \{f_{F, \sigma} : (F, \sigma) \in \mathfrak{M}(\mathbb{R}) \times (0, \infty)\}$ via the mapping $(F, \sigma) \mapsto f_{F, \sigma}$. We shall use the same symbol π to denote either measure, the correct interpretation being clear from the context. We assume that \mathcal{F} is equipped with a metric d , which may be either the Hellinger

$d_H(f, g) := (\int (f^{1/2} - g^{1/2})^2 d\lambda)^{1/2}$, where λ is the Lebesgue measure on \mathbb{R} , or the one induced by the L_1 -norm, $\|f - g\|_1 := \int |f - g| d\lambda$.

We are interested in assessing the rate of convergence for the posterior distribution corresponding to a prior π , under the non-Bayesian assumption that X_1, \dots, X_n are i.i.d. observations from a density f_0 . In the following, it will be explicitly stated whether f_0 is itself a location mixture of EP distributions, *i.e.*, $f_0 \equiv f_{F_0, \sigma_0}$, where F_0 and σ_0 denote the true values of F and σ , respectively. A sequence $\varepsilon_n \rightarrow 0$, as $n \rightarrow \infty$, is said to be an upper bound on the posterior rate of convergence relative to a metric d on \mathcal{F} if, for some constant $M > 0$, the posterior probability

$$\pi(\{f_{F, \sigma} : d(f_{F, \sigma}, f_0) > M\varepsilon_n\} | X_1, \dots, X_n) \rightarrow 0,$$

P_0^∞ -almost surely or in P_0^n -probability, where P_0 stands for the probability measure corresponding to f_0 . In order to derive the rates of convergence for Dirichlet mixtures of EP distributions, we shall appeal to a theorem of Ghosal and van der Vaart [8], page 1239, reported as Theorem A.1 in the Appendix. Condition (A.1) involves the packing number of a class of densities, say \mathcal{F}' , equipped with a (semi-)metric d , denoted by $D(\varepsilon, \mathcal{F}', d)$, which is defined as the maximum number of points in \mathcal{F}' such that the distance between each pair is at least ε . This number is related to the ε -covering number $N(\varepsilon, \mathcal{F}', d)$, the minimum number of balls of radius ε needed to cover \mathcal{F}' , by the inequalities

$$N(\varepsilon, \mathcal{F}', d) \leq D(\varepsilon, \mathcal{F}', d) \leq N(\varepsilon/2, \mathcal{F}', d). \tag{2.1}$$

The logarithm of the packing or covering number is referred to as the (metric) entropy. The symbols “ \lesssim ” and “ \gtrsim ” will be used throughout to indicate inequalities valid up to constants that may be universal or depend on P_0 . Fixed constants within the present set-up will be inessential to our purposes.

3. Finite mixtures

The analysis of data generated by mixture models may reasonably begin from the case where we have observations of phenomena with multiple underlying factors whose cardinality is finite. If the responsible factors are known to be finitely many, then a discrete EP mixture prior can be conveniently used to estimate the sampling density. We begin to consider the case where all the components of the mixture have the same value of the shape parameter. The generic density takes the form

$$f_{F, \sigma}(\cdot) = \sum_{j=1}^k w_j \psi_{\sigma, p}(\cdot - \theta_j),$$

with F a discrete distribution having a finite number of atoms which can be represented as $\sum_{j=1}^k w_j \delta_{\theta_j}$, where δ_{θ_j} denotes a point mass at θ_j . The number of mixture components is a random variable (r.v.) K with probability mass function $\rho(\cdot)$ on the positive integers. Given $K = k$, the vector $\mathbf{w}_k := (w_1, \dots, w_k)$

of mixing weights has conditional distribution π_k on the $(k - 1)$ -dimensional simplex $\Delta_k := \{\mathbf{w}_k \in \mathbb{R}^k : 0 \leq w_j \leq 1, j = 1, \dots, k, \sum_{j=1}^k w_j = 1\}$ and the locations $\theta_1, \dots, \theta_k$ are independent r.v.'s with common distribution Π . We shall also write $\boldsymbol{\theta}_k$ for $(\theta_1, \dots, \theta_k)$ to ease the notation. The hierarchical model can be thus described:

- $K \sim \rho$ and $\sigma \sim G$ independently;
- given (k, σ) , the r.v.'s $\theta_1, \dots, \theta_k \stackrel{\text{i.i.d.}}{\sim} \Pi$ and $\mathbf{w}_k \sim \pi_k$ independently;
- given $(k, \sigma, \boldsymbol{\theta}_k, \mathbf{w}_k)$, the r.v.'s X_1, \dots, X_n are conditionally independent and identically distributed with density $f_{F, \sigma}$.

We shall show that, if the true density is itself a finite mixture of EP densities with the same shape, under regularity conditions on the prior, the posterior converges at nearly parametric rate $n^{-1/2}(\log n)$. This rate agrees with those found for other discrete kernel mixture priors, when the true density is of the same form as the assumed model, see, for instance, Ghosal [6] for Bernstein polynomials, Kruijer and van der Vaart [15] for beta mixtures, Scricciolo [22] for histograms and polygons. The proof exploits the approximation properties of the mixtures under consideration to find a good fitting distribution of a given density in a proper subclass. In fact, any finite EP mixture can be approximated arbitrarily closely (in the distance induced by the L_1 -norm) by mixtures having exactly the same number of components, with locations and mixing weights taking values in suitable neighborhoods of the corresponding true elements. This technique is used to provide an estimate of the prior concentration rate as well as an upper bound on the metric entropy of a sieve set. In the first case, the number of mixture components is constant, this leading to a high prior concentration rate; in the second case, it can be chosen to increase quite slowly, at a logarithmic rate, as the approximation error goes to zero. In the metric entropy estimate, we shall also need the bandwidth to approach zero at a logarithmic rate, thus a suitable condition on the decay rate of the prior G for σ at zero is required. We shall use the following assumption.

- (A) G has a continuous and positive Lebesgue density g on an interval containing σ_0 and, for constants $d > 0, \gamma, \varpi \in (0, \infty]$, satisfies

$$G(s) \lesssim e^{-ds^{-\gamma}} \quad \text{as } s \rightarrow 0 \quad \text{and} \quad 1 - G(s) \lesssim s^{-\varpi} \quad \text{as } s \rightarrow \infty. \quad (3.1)$$

The common choice of an inverse-gamma prior on σ^2 when the kernel is Gaussian meets requirement (A). Let $\sigma^2 \sim \text{IG}(\alpha, \beta)$, with parameters $\alpha, \beta > 0$. The condition on the tail behaviour at zero is satisfied with $\gamma = 2$. In fact, for some constant $d > 0$,

$$G(s) = \Pr(\sigma \leq s) = \Pr(\sigma^2 \leq s^2) \lesssim e^{-ds^{-2}}, \quad s > 0.$$

The condition on the tail behaviour at infinity is satisfied with $\varpi = 2\alpha$,

$$1 - G(s) = \Pr(\sigma > s) = \Pr(\sigma^2 > s^2) \leq \frac{\beta^\alpha}{\alpha\Gamma(\alpha)} s^{-2\alpha}, \quad s > 0.$$

The next theorem states the first result of the section.

Theorem 3.1. *Let $p > 0$ be fixed. Suppose that $f_0 = F_0 * \psi_{\sigma_0, p}$ is a finite EP mixture. Assume that*

- (i) *the prior G for σ satisfies condition (A);*
- (ii) *the prior ρ for the number of components is such that, for constants $b, B > 0$, $0 < \rho(k) \leq Be^{-bk}$ for all $k \in \mathbb{N}$;*
- (iii) *for each $k \in \mathbb{N}$, the prior π_k for the mixing weights is a Dirichlet distribution $\text{Dir}(\alpha_1, \dots, \alpha_k)$, with parameters $\alpha_1, \dots, \alpha_k$ such that, for constants $a, A > 0$, $D \geq 1$ and for $0 < \varepsilon \leq 1/(Dk)$,*

$$A\varepsilon^a \leq \alpha_j \leq D, \quad j = 1, \dots, k;$$

- (iv) *the common prior Π for the locations has a continuous and positive density on an interval containing the support of F_0 and, for constants $c, \vartheta > 0$, satisfies*

$$\Pi(\{\theta : |\theta| > t\}) \lesssim e^{-ct^\vartheta}, \quad \text{for large } t > 0.$$

Then, the posterior rate of convergence relative to d_H is $\varepsilon_n = n^{-1/2}(\log n)$.

Proof. We appeal to Theorem A.1 and show that conditions (A.1)–(A.3) are satisfied with sequences $\bar{\varepsilon}_n = n^{-1/2}(\log n)$ and $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^{1/2}$, so that (an upper bound on) the rate is given by $\varepsilon_n := (\bar{\varepsilon}_n \vee \tilde{\varepsilon}_n) = \bar{\varepsilon}_n$. For sequences k_n of positive integers, a_n, s_n and t_n of positive real numbers specified below, let $\mathcal{F}_n := \bigcup_{k=1}^{k_n} \mathcal{F}_{a_n, s_n, t_n}^{(k)}$, where $\mathcal{F}_{a_n, s_n, t_n}^{(k)}$ is the class of EP mixtures with at most k components defined as

$$\mathcal{F}_{a_n, s_n, t_n}^{(k)} := \left\{ f_{F, \sigma} : F = \sum_{j=1}^k w_j \delta_{\theta_j}, \quad |\theta_j| \leq a_n, \quad j = 1, \dots, k, \quad s_n \leq \sigma \leq t_n \right\}.$$

Along the lines of Lemma 3 of Ghosal and van der Vaart [10], pages 705–707, it can be shown that

$$N(\bar{\varepsilon}_n, \mathcal{F}_{a_n, s_n, t_n}^{(k)}, \|\cdot\|_1) \lesssim \left(\frac{t_n - s_n}{s_n \bar{\varepsilon}_n} \vee 1 \right) \times \left[\left(\frac{2a_n}{s_n \bar{\varepsilon}_n} + 1 \right) \times \left(\frac{5}{\bar{\varepsilon}_n \wedge 1} \right) \right]^k.$$

Thus,

$$\begin{aligned} D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) &\leq \sum_{k=1}^{k_n} D(\bar{\varepsilon}_n, \mathcal{F}_{a_n, s_n, t_n}^{(k)}, d_H) \\ &\leq \sum_{k=1}^{k_n} N(\bar{\varepsilon}_n^2/4, \mathcal{F}_{a_n, s_n, t_n}^{(k)}, \|\cdot\|_1) \\ &\lesssim k_n \times \frac{t_n}{s_n \bar{\varepsilon}_n^2/4} \times \left(\frac{a_n}{s_n \bar{\varepsilon}_n^4/16} \right)^{k_n}. \end{aligned}$$

For t_n such that $(\log t_n) = O(\log n)$,

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim k_n \log \left(\frac{a_n}{s_n \bar{\varepsilon}_n} \right).$$

For constants $D_2 \geq D_1 > 0$ and $E, F, L > 0$ to be chosen as below indicated, let $D_1(\log n) \leq k_n \leq D_2(\log n)$, $a_n = L(\log \bar{\varepsilon}_n^{-1})^{1/\vartheta}$, $s_n = E(\log \bar{\varepsilon}_n^{-1})^{-1/\gamma}$ and $t_n = \bar{\varepsilon}_n^{-F}$. Then, $\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim n\bar{\varepsilon}_n^2$ and condition (A.1) is verified.

We check that condition (A.2) is satisfied with $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^{1/2}$. For a constant $c_2 > 0$ that will arise from condition (A.3), let $D_1 \geq (c_2 + 4)/b$, $E \leq [d/4(c_2 + 4)]^{1/\gamma}$, $F \geq [4(c_2 + 4)/\varpi]$ and $L \geq [4(c_2 + 4)/c]^{1/\vartheta}$. By assumptions (i), (ii) and (iv), we have

$$\begin{aligned} \pi(\mathcal{F}_n^c) &\leq \sum_{k=k_n+1}^{\infty} \rho(k) + \mathbb{E}[K]\Pi([-a_n, a_n]^c) + [G(s_n) + 1 - G(t_n)] \\ &\lesssim e^{-bk_n} + e^{-ca_n^\vartheta} + e^{-ds_n^{-\gamma}} + t_n^{-\varpi} \\ &\lesssim e^{-(c_2+4)n\bar{\varepsilon}_n^2}, \end{aligned}$$

where $\mathbb{E}[K] < \infty$ by (ii).

Now, we consider condition (A.3). Recall that, by assumption, f_0 is a finite EP mixture, *i.e.*, for some $k_0 \in \mathbb{N}$, $\mathbf{w}_{k_0}^0 := (w_1^0, \dots, w_{k_0}^0) \in \Delta_{k_0}$ and $\boldsymbol{\theta}_{k_0}^0 := (\theta_1^0, \dots, \theta_{k_0}^0) \in \mathbb{R}^{k_0}$,

$$f_0(\cdot) = \sum_{j=1}^{k_0} w_j^0 \psi_{\sigma_0, p}(\cdot - \theta_j^0).$$

For any $\sigma > 0$ and any discrete distribution F on \mathbb{R} with k_0 support points $\boldsymbol{\theta}_{k_0} := (\theta_1, \dots, \theta_{k_0}) \in \mathbb{R}^{k_0}$ and mixing weights $\mathbf{w}_{k_0} := (w_1, \dots, w_{k_0}) \in \Delta_{k_0}$, by inequalities (A.6) and (A.7),

$$\begin{aligned} \|f_{F, \sigma} - f_0\|_1 &\leq \|f_{F, \sigma} - f_{F, \sigma_0}\|_1 + \|f_{F, \sigma_0} - f_0\|_1 \\ &\lesssim \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} + \sum_{j=1}^{k_0} w_j^0 \|\psi_{\sigma_0, p}(\cdot - \theta_j) - \psi_{\sigma_0, p}(\cdot - \theta_j^0)\|_1 \\ &\quad + \sum_{j=1}^{k_0} |w_j - w_j^0| \|\psi_{\sigma_0, p}(\cdot - \theta_j)\|_1 \\ &\lesssim \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} + \|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} + \|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1}, \end{aligned}$$

where, for any pair $\mathbf{x}_k, \mathbf{y}_k \in \mathbb{R}^k$, $\|\mathbf{x}_k - \mathbf{y}_k\|_{\ell_1}$ denotes the ℓ_1 -distance $\sum_{j=1}^k |x_j - y_j|$. Let $0 < \varepsilon \leq [(\sigma_0/2) \wedge (1 - e^{-1})/\sqrt{2}]$ be fixed. For any density $f_{F, \sigma}$ with $\sigma > 0$ such that $|\sigma - \sigma_0| \leq \varepsilon$ and $F = \sum_{j=1}^{k_0} w_j \delta_{\theta_j}$, where $\|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1} \leq \varepsilon$ and $\|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} \leq \varepsilon$, we have $d_H^2(f_{F, \sigma}, f_0) \leq \|f_{F, \sigma} - f_0\|_1 \lesssim \varepsilon$. Thus, by

Lemma A.10, for a suitable constant $c_1 > 0$,

$$\left\{ f_{F, \sigma} : F = \sum_{j=1}^{k_0} w_j \delta_{\theta_j}, \|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1} \leq \varepsilon, \|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} \leq \varepsilon, |\sigma - \sigma_0| \leq \varepsilon \right\} \subseteq B_{\text{KL}} \left(f_0; c_1 \varepsilon \left(\log \frac{1}{\varepsilon} \right)^2 \right).$$

Defined the set $N(\mathbf{w}_{k_0}^0; \varepsilon) := \{\mathbf{w}_{k_0} \in \Delta_{k_0} : \|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1} \leq \varepsilon\}$, for small enough $\varepsilon > 0$, using Lemma A.1 of Ghosal [6], pages 1278–1279, and assumption (iii), we have $\pi_{k_0}(N(\mathbf{w}_{k_0}^0; \varepsilon)) \gtrsim \exp\{-c_0 k_0 \log(1/\varepsilon)\}$, where c_0 is a positive constant depending only on a, A, D and $\sum_{j=1}^{k_0} \alpha_j$. Denoted by $B(\boldsymbol{\theta}_{k_0}^0; \varepsilon)$ the $\boldsymbol{\theta}_{k_0}^0$ -centered ℓ_1 -ball of radius ε ,

$$B(\boldsymbol{\theta}_{k_0}^0; \varepsilon) := \{\boldsymbol{\theta}_{k_0} \in \mathbb{R}^{k_0} : \|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} \leq \varepsilon\},$$

by condition (iv), the prior probability of $B(\boldsymbol{\theta}_{k_0}^0; \varepsilon)$ under the k_0 -fold product measure $\Pi^{\otimes k_0}$ can be bounded below as follows:

$$\Pi^{\otimes k_0}(B(\boldsymbol{\theta}_{k_0}^0; \varepsilon)) \geq \prod_{j=1}^{k_0} \Pi \left(\left[\theta_j^0 - \frac{\varepsilon}{k_0}, \theta_j^0 + \frac{\varepsilon}{k_0} \right] \right) \gtrsim \exp \left\{ -d_1 k_0 \log \frac{1}{\varepsilon} \right\},$$

where d_1 is a positive constant. Therefore, for $c_1, d_2 > 0$,

$$\begin{aligned} \pi(B_{\text{KL}}(f_0; c_1 \varepsilon (\log \varepsilon^{-1})^2)) &\gtrsim \left(\int_{\sigma_0 - \varepsilon}^{\sigma_0 + \varepsilon} g \, d\lambda \right) \\ &\quad \times \rho(k_0) \pi_{k_0}(N(\mathbf{w}_{k_0}^0; \varepsilon)) \Pi^{\otimes k_0}(B(\boldsymbol{\theta}_{k_0}^0; \varepsilon)) \\ &\gtrsim \exp \left\{ -d_2 k_0 \log \frac{1}{\varepsilon} \right\}. \end{aligned}$$

Set $\xi := (c_1 \varepsilon)^{1/2} \log(1/\varepsilon)$, since $\log(1/\varepsilon) \lesssim \log(1/\xi)$, we have $\pi(B_{\text{KL}}(f_0; \xi^2)) \gtrsim \exp\{-c_2 \log(1/\xi)\}$, where c_2 is a positive constant (possibly depending on f_0). Replacing ξ with $\tilde{\varepsilon}_n$, we have $\pi(B_{\text{KL}}(f_0; \tilde{\varepsilon}_n^2)) \gtrsim \exp\{-c_2 n \tilde{\varepsilon}_n^2\}$ for sufficiently large n , and the proof is complete. \square

Remark 3.1. The condition on the parameters of the Dirichlet distribution contained in assumption (iii) is satisfied if, for example, a noninformative specification is considered, as when all the α_j 's are taken to be equal to some constant $D \geq 1$.

Now, we consider the case where the shape parameter is allowed to mix both in the prior and in the true density. The generic density has the form

$$f_{F, \mathbf{p}, \sigma}(\cdot) := \sum_{j=1}^k w_j \psi_{\sigma, p_j}(\cdot - \theta_j), \quad F = \sum_{j=1}^k w_j \delta_{\theta_j},$$

where $\mathbf{p}_k := (p_1, \dots, p_k)$ is the vector of shape parameters that are assumed to be independent r.v.'s with common distribution H . The hierarchical model can be described as follows:

- $K \sim \rho$ and $\sigma \sim G$ independently;
- given (k, σ) , $\mathbf{w}_k \sim \pi_k$, $\theta_1, \dots, \theta_k \stackrel{\text{i.i.d.}}{\sim} \Pi$ and $p_1, \dots, p_k \stackrel{\text{i.i.d.}}{\sim} H$ independently;
- given $(k, \sigma, \mathbf{w}_k, \boldsymbol{\theta}_k, \mathbf{p}_k)$, the r.v.'s X_1, \dots, X_n are conditionally independent and identically distributed with density $f_{F, \mathbf{p}_k, \sigma}$.

Theorem 3.2. *Suppose that f_0 is a finite mixture of EP densities of the type*

$$f_0(\cdot) = \sum_{j=1}^{k_0} w_j^0 \psi_{\sigma_0, p_j^0}(\cdot - \theta_j^0), \quad (3.2)$$

with $p_j^0 > 0$ for all $j = 1, \dots, k_0$. Assume that conditions (i)–(iv) of Theorem 3.1 are satisfied and

- (v) *the common prior H for the shape parameters has a continuous and positive Lebesgue density h on $(0, \infty)$ and, for constants $l > 0$, $\beta \in [1, \infty]$, $v \in (0, \infty]$, satisfies*

$$H(p) \lesssim e^{-lp^{-\beta}} \quad \text{as } p \rightarrow 0 \quad \text{and} \quad 1 - H(p) \lesssim p^{-v} \quad \text{as } p \rightarrow \infty. \quad (3.3)$$

Then, the posterior rate of convergence relative to d_H is $\varepsilon_n = n^{-1/2}(\log n)$.

Proof. The main steps in the proof of Theorem 3.1 go through in this case. Conditions (A.1)–(A.3) are satisfied with the same sequences $\bar{\varepsilon}_n = n^{-1/2}(\log n)$ and $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^{1/2}$. We begin to consider condition (A.1). For sequences k_n, a_n, s_n, t_n specified as in Theorem 3.1, and p_n, q_n of positive real numbers to be chosen below, let $\mathcal{F}_n := \bigcup_{k=1}^{k_n} \mathcal{F}_n^{(k)}$, where $\mathcal{F}_n^{(k)}$ is the class of mixtures with at most k components defined as

$$\mathcal{F}_n^{(k)} := \left\{ f_{F, \mathbf{p}_k, \sigma} : F = \sum_{j=1}^k w_j \delta_{\theta_j}, \quad |\theta_j| \leq a_n, \quad j = 1, \dots, k, \right. \\ \left. p_n \leq p_j \leq q_n, \quad j = 1, \dots, k, \quad s_n \leq \sigma \leq t_n \right\}.$$

Similarly to Theorem 3.1 and using Lemma A.4, it can be proved that for large enough n ,

$$N(\bar{\varepsilon}_n, \mathcal{F}_n^{(k)}, \|\cdot\|_1) \lesssim \left(\frac{t_n - s_n}{s_n \bar{\varepsilon}_n} \vee 1 \right) \times \left(\frac{2a_n e^{1/p_n}}{s_n \bar{\varepsilon}_n} + 1 \right)^k \\ \times \left(\frac{q_n - p_n}{p_n \bar{\varepsilon}_n} \vee 1 \right)^k \times \left(\frac{5}{\bar{\varepsilon}_n \wedge 1} \right)^k.$$

Consequently,

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim k_n \left[\log \left(\frac{a_n q_n t_n}{p_n s_n \bar{\varepsilon}_n} \right) + \frac{1}{p_n} \right].$$

For constants $0 < C_1 \leq [l/4(c_2 + 4)]^{1/\beta}$ and $C_2 \geq [4(c_2 + 4)/v]$, where $c_2 > 0$ is the constant arising from condition (A.3), let $p_n = C_1 (\log \bar{\varepsilon}_n^{-1})^{-1/\beta}$ and $q_n = \bar{\varepsilon}_n^{-C_2}$. Since $\beta \geq 1$, $\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim n \bar{\varepsilon}_n^2$ and condition (A.1) is satisfied.

Using assumptions (i), (ii), (iv) and (v), condition (A.2) is seen to be satisfied with $\tilde{\varepsilon}_n = n^{-1/2} (\log n)^{1/2}$. In fact,

$$\begin{aligned} \pi(\mathcal{F}_n^c) &\leq \sum_{k=k_n+1}^{\infty} \rho(k) + \mathbb{E}[K] \Pi([-a_n, a_n]^c) \\ &\quad + \mathbb{E}[K] [H(p_n) + 1 - H(q_n)] + [G(s_n) + 1 - G(t_n)] \\ &\lesssim e^{-bk_n} + e^{-ca_n^\vartheta} + e^{-lp_n^{-\beta}} + q_n^{-v} + e^{-ds_n^{-\gamma}} + t_n^{-\varpi} \\ &\lesssim e^{-(c_2+4)n\tilde{\varepsilon}_n^2}, \end{aligned}$$

where $\mathbb{E}[K] < \infty$ by assumption (ii).

To check that condition (A.3) is satisfied with $\tilde{\varepsilon}_n$, recall that, by assumption, f_0 is a finite EP mixture of the type $f_{F_0, \mathbf{p}_{k_0}^0, \sigma_0}$, with $\mathbf{p}_{k_0}^0 := (p_1^0, \dots, p_{k_0}^0)$ having all components $p_j^0 > 0$. For any $\sigma > 0$, any $\mathbf{p}_{k_0} := (p_1, \dots, p_{k_0})$ with positive p_j 's, and any discrete distribution F on \mathbb{R} with k_0 support points $\boldsymbol{\theta}_{k_0} := (\theta_1, \dots, \theta_{k_0}) \in \mathbb{R}^{k_0}$ and weights $\mathbf{w}_{k_0} := (w_1, \dots, w_{k_0}) \in \Delta_{k_0}$, by inequalities (A.6), (A.7) and Lemma A.4,

$$\begin{aligned} \|f_{F, \mathbf{p}_{k_0}, \sigma} - f_0\|_1 &\leq \|f_{F, \mathbf{p}_{k_0}, \sigma} - f_{F, \mathbf{p}_{k_0}, \sigma_0}\|_1 + \|f_{F, \mathbf{p}_{k_0}, \sigma_0} - f_{F, \mathbf{p}_{k_0}^0, \sigma_0}\|_1 \\ &\quad + \|f_{F, \mathbf{p}_{k_0}^0, \sigma_0} - f_0\|_1 \\ &\leq 2 \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} + \sum_{j=1}^{k_0} w_j \|\psi_{\sigma_0, p_j}(\cdot - \theta_j) - \psi_{\sigma_0, p_j^0}(\cdot - \theta_j)\|_1 \\ &\quad + \sum_{j=1}^{k_0} w_j^0 \|\psi_{\sigma_0, p_j^0}(\cdot - \theta_j) - \psi_{\sigma_0, p_j^0}(\cdot - \theta_j^0)\|_1 \\ &\quad + \sum_{j=1}^{k_0} |w_j - w_j^0| \|\psi_{\sigma_0, p_j^0}(\cdot - \theta_j)\|_1 \\ &\leq 2 \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} \\ &\quad + \frac{1}{\min_{1 \leq j \leq k_0} p_j^0} \sum_{j=1}^{k_0} \frac{|p_j - p_j^0|}{p_j} \|g_{(\cdot - \theta_j)/\sigma_0}^{(1)}(z_j^*)\|_1 \\ &\quad + \frac{2}{\sigma_0 \min_{1 \leq j \leq k_0} c_{p_j^0}} \|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} + \|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1} \end{aligned}$$

$$\begin{aligned} &\lesssim \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} + \sum_{j=1}^{k_0} \frac{|p_j - p_j^0|}{p_j} \|g_{(\cdot - \theta_j)/\sigma_0}^{(1)}(z_j^*)\|_1 \\ &\quad + \|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} + \|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1}, \end{aligned}$$

where $\min_{1 \leq j \leq k_0} p_j^0 > 0$ by assumption and $\min_{1 \leq j \leq k_0} c_{p_j^0} > 0$ because, for any $p > 0$, c_p is positive with $\lim_{p \rightarrow 0^+} c_p = 0$ and $\lim_{p \rightarrow \infty} c_p = 2$. For each $j = 1, \dots, k_0$, z_j^* is a point lying in between $1/(p_j \vee p_j^0)$ and $1/(p_j \wedge p_j^0)$. Since the value of $\|g_{(\cdot - \theta_j)/\sigma_0}^{(1)}(z_j^*)\|_1$ does not depend on θ_j , the symbol θ_j may be suppressed. For ease of notation, we shall write $\|g_{(\cdot/\sigma_0)}^{(1)}(z_j^*)\|_1$. If, for $0 < \varepsilon \leq \min_{1 \leq j \leq k_0} p_j^0/2$, we have $\|\mathbf{p}_{k_0} - \mathbf{p}_{k_0}^0\|_{\ell_1} \leq \varepsilon$, then

$$\begin{aligned} \|f_{F, \mathbf{p}_{k_0}, \sigma_0} - f_{F, \mathbf{p}_{k_0}^0, \sigma_0}\|_1 &\leq \frac{1}{\min_{1 \leq j \leq k_0} p_j^0} \sum_{j=1}^{k_0} \frac{|p_j - p_j^0|}{p_j} \|g_{(\cdot/\sigma_0)}^{(1)}(z_j^*)\|_1 \\ &\leq \frac{2}{(\min_{1 \leq j \leq k_0} p_j^0)^2} \|g_{(\cdot/\sigma_0)}^{(1)}(z_0)\|_1 \|\mathbf{p}_{k_0} - \mathbf{p}_{k_0}^0\|_{\ell_1} \\ &\lesssim \|\mathbf{p}_{k_0} - \mathbf{p}_{k_0}^0\|_{\ell_1} \lesssim \varepsilon, \end{aligned}$$

where the second inequality descends from the fact that, since $\|g_{(\cdot/\sigma_0)}^{(1)}(z)\|_1$ is continuous in z on $(0, \infty)$, there exists a point z_0 (depending on f_0) such that $\|g_{(\cdot/\sigma_0)}^{(1)}(z_j^*)\|_1 \leq \|g_{(\cdot/\sigma_0)}^{(1)}(z_0)\|_1 < \infty$ for all $j = 1, \dots, k_0$. Fix $0 < \varepsilon \leq \min\{(\sigma_0/2), (\min_{1 \leq j \leq k_0} p_j^0/2), (1 - e^{-1})/\sqrt{2}\}$. For any density $f_{F, \mathbf{p}_{k_0}, \sigma}$ in the set

$$\left\{ \begin{aligned} f_{F, \mathbf{p}_{k_0}, \sigma} : F &= \sum_{j=1}^{k_0} w_j \delta_{\theta_j}, \quad \|\mathbf{w}_{k_0} - \mathbf{w}_{k_0}^0\|_{\ell_1} \leq \varepsilon, \quad w_j \geq \frac{\varepsilon^2}{8}, \quad j = 1, \dots, k_0, \\ \|\boldsymbol{\theta}_{k_0} - \boldsymbol{\theta}_{k_0}^0\|_{\ell_1} &\leq \varepsilon, \quad \|\mathbf{p}_{k_0} - \mathbf{p}_{k_0}^0\|_{\ell_1} \leq \varepsilon, \quad 0 < p_j \leq p_j^0, \quad j = 1, \dots, k_0, \\ |\sigma - \sigma_0| &\leq \varepsilon \end{aligned} \right\},$$

we have $d_{\text{H}}^2(f_{F, \mathbf{p}_{k_0}, \sigma}, f_0) \leq \|f_{F, \mathbf{p}_{k_0}, \sigma} - f_0\|_1 \lesssim \varepsilon$ and, for a suitable $\delta \in (0, 1]$, the quantity M_δ^2 defined in (A.16) is such that $M_\delta^2 = O(1/\varepsilon^2)$. Thus, by Theorem 5 of Wong and Shen [31], pages 357–358, for a suitable constant $c_1 > 0$, the above set is contained in $B_{\text{KL}}(f_0; c_1 \varepsilon (\log \varepsilon^{-1})^2)$. Using arguments similar to those of Theorem 3.1 and the fact that, for small enough $\varepsilon > 0$, by condition (v),

$$\begin{aligned} &\Pr(\{\mathbf{p}_{k_0} : \|\mathbf{p}_{k_0} - \mathbf{p}_{k_0}^0\|_{\ell_1} \leq \varepsilon, \quad 0 < p_j \leq p_j^0, \quad j = 1, \dots, k_0\}) \\ &\geq \prod_{j=1}^{k_0} \Pr\left(\left[p_j^0 - \frac{\varepsilon}{k_0}, p_j^0\right]\right) = \prod_{j=1}^{k_0} \left(\int_{p_j^0 - \varepsilon/k_0}^{p_j^0} h(p) \, dp\right) \gtrsim \exp\left\{-d_2 k_0 \log \frac{1}{\varepsilon}\right\}, \end{aligned}$$

we have $\pi(B_{\text{KL}}(f_0; c_1\varepsilon(\log \varepsilon^{-1})^2)) \gtrsim \exp\{-d_3k_0 \log(1/\varepsilon)\}$, for some $d_3 > 0$. Therefore, for a suitable constant $c_2 > 0$ (possibly depending on f_0), we have $\pi(B_{\text{KL}}(f_0; \tilde{\varepsilon}_n^2)) \gtrsim \exp\{-c_2n\tilde{\varepsilon}_n^2\}$ for sufficiently large n . \square

Remark 3.2. An example of distribution with density of the form (3.2) is given by the *Laplace-normal* (or *Gauss-Laplace*) mixture

$$f_0(\cdot) = w_0\psi_{\sigma_0,1}(\cdot - \theta_0) + (1 - w_0)\phi_{\sigma_0}(\cdot - \theta_0)$$

with equal locations, which was used by Kanji [14], Jones and McLachlan [13] to fit wind shear data. Haas, Mittnik and Paoletta [11] used the Gauss-Laplace mixture for modelling and predicting financial risk.

Remark 3.3. Suppose the investigator wants to assign positive prior probability to EP densities having tails equal to or lighter than those of a Laplace. A prior distribution for p on $[1, \infty)$ can be specified by considering a beta distribution for $1/p$ as in Box and Tiao [2], page 167,

$$1/p \sim \text{Beta}(a, a), \quad a \geq 1.$$

Thus, $1/p$ has a symmetric distribution around the normal theory value $1/2$. Deviations from the normality can be taken into account by “adjusting” the value of the parameter a . When $a = 1$, the prior distribution is uniform over $(0, 1]$. When $a > 1$, the prior distribution is symmetric, with mode $a/2$, and assigns high probability to EP distributions in a neighborhood of the normal. As $a \rightarrow \infty$, the prior density becomes more and more peaked at $1/2$ and converges to a delta function, representing an assumption of exact normality. Since the prior distribution for p corresponding to a beta for $1/p$ is left-truncated, the lower tail condition in (3.3) is trivially satisfied with $\beta = \infty$, while the upper tail condition is satisfied with $v = a$ because

$$1 - H(p) \leq \frac{\Gamma(2a)}{a[\Gamma(a)]^2} p^{-a}, \quad p \geq 1.$$

Remark 3.4. As a consequence of Theorem 3.1 or Theorem 3.2, the Bayes’ estimator,

$$\hat{f}_n(\cdot) := \int f(\cdot)\pi(df|X_1, \dots, X_n),$$

converges to f_0 in the Hellinger distance, in P_0^n -probability, at a rate at least as fast as $n^{-1/2}(\log n)$,

$$d_H(\hat{f}_n, f_0) = O_P(n^{-1/2}(\log n)),$$

see, e.g., Theorem 5 of Shen and Wasserman [23], page 694.

4. Infinite mixtures

In this section, we analyze the case where the underlying factors may be infinitely many. We consider densities of the form $f_{F,\sigma} = F * \psi_{\sigma,p}$, where F can

be any distribution on \mathbb{R} . As a prior for F , we adopt a Dirichlet process \mathfrak{D}_α with base measure α . We recall that a Dirichlet process on a measurable space $(\mathfrak{X}, \mathcal{A})$, with a finite and positive base measure α on $(\mathfrak{X}, \mathcal{A})$, is a random probability measure F on $(\mathfrak{X}, \mathcal{A})$ such that, for every finite partition (A_1, \dots, A_k) of \mathfrak{X} , the vector of probabilities $(F(A_1), \dots, F(A_k))$ has a Dirichlet distribution with parameters $(\alpha(A_1), \dots, \alpha(A_k))$. Observations from a Dirichlet process mixture of EP distributions can be structurally described as follows:

- $F \sim \mathfrak{D}_\alpha$ and $\sigma \sim G$ independently;
- given (F, σ) , the r.v.'s $\theta_1, \dots, \theta_n \stackrel{\text{i.i.d.}}{\sim} F$;
- given $(F, \sigma, \theta_1, \dots, \theta_n)$, the r.v.'s $e_1, \dots, e_n \stackrel{\text{i.i.d.}}{\sim} \psi_{\sigma, p}$;
- the observations are defined as $X_i := \theta_i + e_i$, for $i = 1, \dots, n$.

Let $\pi = \mathfrak{D}_\alpha \times G$ denote the prior distribution for (F, σ) , with the proviso that the symbol π will also be used for the prior induced on \mathcal{F} by the mapping $(F, \sigma) \mapsto f_{F, \sigma}$, the ambiguity being resolved by the context. We recall that π is said to be *strongly consistent at f_0* if, for every $\varepsilon > 0$, $\pi(B(f_0; \varepsilon) | X_1, \dots, X_n) \rightarrow 1$ a.s. $[P_0^\infty]$, with $B(f_0; \varepsilon) := \{f_{F, \sigma} : d(f_{F, \sigma}, f_0) \leq \varepsilon\}$, where d can be either the Hellinger or the L_1 -metric. Ascertainment of posterior consistency for Dirichlet EP mixture priors can proceed as for Dirichlet normal mixture priors, for which sufficient conditions were derived by Ghosal, Ghosh and Ramamoorthi [7] in Theorem 7, page 152, using the sieve approach. Lijoi, Prünster and Walker [16], Theorem 1, page 1293, weakened their conditions adopting an alternative approach due to Walker [27]. In both approaches, the main idea behind the proof of consistency is to show that the prior satisfies Schwartz's [20] condition on the positivity of the probability of Kullback-Leibler neighborhoods of f_0 , known as the *Kullback-Leibler property* and indicated by the notation $f_0 \in \text{KL}(\pi)$. The following proposition provides sufficient conditions for f_0 to be in the Kullback-Leibler support of π .

Proposition 4.1. *Assume that $\pi = \mathfrak{D}_\alpha \times G$, with the base measure α supported on \mathbb{R} and the prior G supported on $(0, \infty)$. Let f_0 be a continuous density on \mathbb{R} satisfying the following conditions:*

- (a) for some constant $0 < M_0 < \infty$, $0 < f_0(x) \leq M_0$ for all $x \in \mathbb{R}$;
- (b) $|\int f_0(x) \log f_0(x) dx| < \infty$;
- (c) for some $\delta_0 > 0$, $\int f_0(x) \log[f_0(x)/\varphi_{\delta_0}(x)] dx < \infty$, where $\varphi_{\delta_0}(x) := \inf_{|t-x| < \delta_0} f_0(t)$;
- (d) given $p > 0$, there exists $\eta_0 > 0$ such that $\int |x|^{p(1+\eta_0)} f_0(x) dx < \infty$.

Then, $f_0 \in \text{KL}(\pi)$.

Proof. First, note that $\pi = \mathfrak{D}_\alpha \times G$ is a Type II mixture prior or a Prior 2, in the terminology of Wu and Ghosal [32], page 299. Thus, we can appeal to their Theorem 3, page 310. Once observed that the weak support of \mathfrak{D}_α is $\mathfrak{M}(\mathbb{R})$, the proof parallels that of Theorem 6, page 312, which deals with the case where $p = 1$, namely, the kernel is a Laplace density. Note that conditions (a)–(c) coincide with conditions B4–B6 of Theorem 2, pages 305–309. See also Wu and

Ghosal [33]. It is, therefore, sufficient to check that conditions B3 and B7 are satisfied. We begin with condition B3. Since, for $x \neq 0$,

$$\frac{\psi_{1,p}^{(1)}(x)}{\psi_{1,p}(x)} = -|x|^{p-1} \text{sgn}(x) = \begin{cases} -x^{p-1}, & x > 0, \\ (-x)^{p-1}, & x < 0, \end{cases}$$

we have

$$x \frac{\psi_{1,p}^{(1)}(x)}{\psi_{1,p}(x)} = -|x|^p \leq -1, \quad \text{for } |x| \geq 1,$$

and the condition is satisfied. As for condition B7, we need to check that, for any $a \in \mathbb{R}$ and $b > 0$,

$$\int f_0(x) |\log \psi_{1,p}((x-a)/b)| \, dx < \infty,$$

and, for some $\eta_0 > 0$,

$$\left| \int f_0(x) \log \psi_{1,p}(2x|x|^{\eta_0}) \, dx \right| < \infty.$$

The linearity of $\log \psi_{1,p}(x)$ as a function of $|x|^p$, together with assumption (d), imply that the above integrals are finite. \square

4.1. Analytic kernels

In this subsection, we consider the case where the true density $f_0 = F_0 * \psi_{\sigma_0,p}$ is a location mixture of EP densities with shape parameter p that is an even integer. If $p = 2m$, $m \in \mathbb{N}$, then, for any value σ_0 of the scale, the kernel $\psi_{\sigma_0,p}$ is an analytic function. It is infinitely differentiable on \mathbb{R} , and, in particular, at the origin. We establish that, if the true mixing distribution F_0 has either compact support or sub-exponential tails, then the sequence of posterior distributions (weakly) converges to a point mass at the true probability measure P_0 with a nearly parametric rate, up to a logarithmic factor. The key idea of the proof is the following. A density f_0 of the stated form can be uniformly approximated by a *finite* mixture of EP distributions with a number of components that increases at a logarithmic rate, as the approximation error goes to zero. Because of the analyticity of the kernel, such a finitely supported mixing distribution can be found by matching a relatively small number of moments of F_0 or of its (re-normalized) restriction to a compact set, see Lemma A.5 (cf. Lemma 3.1 of Ghosal and van der Vaart [8], pages 1240–1241). This result is used to provide an exponential lower bound on the prior probability of Kullback-Leibler type neighborhoods of f_0 as well as an exponential upper bound on the covering number of a sieve set.

Theorem 4.1. *Let p be a fixed even integer. Suppose that $f_0 = F_0 * \psi_{\sigma_0,p}$, with the true mixing distribution F_0 having compact support. If the base measure α*

has a continuous and positive density on an interval containing the support of F_0 and, for constants $b, \delta > 0$, satisfies the tail condition

$$\alpha(\{\theta : |\theta| > t\}) \lesssim e^{-bt^\delta}, \quad \text{for large } t > 0, \tag{4.1}$$

the prior G for σ satisfies condition (A), then the posterior rate of convergence relative to d_H is $\varepsilon_n = n^{-1/2}(\log n)^\kappa$, where

$$\kappa = \frac{1}{2} + \left[\frac{1}{2} \vee p \left(\frac{1}{\delta} + \frac{1}{\gamma} \right) \right]. \tag{4.2}$$

Proof. We show that conditions (A.1) through (A.3) of Theorem A.1 are satisfied by $\bar{\varepsilon}_n = n^{-1/2}(\log n)^\kappa$, with κ as in (4.2), and $\tilde{\varepsilon}_n = n^{-1/2}(\log n)$. Given $\eta \in (0, 1/5)$, for positive constants E, F and L to be suitably chosen later on, let $s = E(\log \eta^{-1})^{-2/\gamma}$, $0 < t \leq \exp\{F(\log \eta^{-1})^2\}$ and $0 < a \leq L(\log \eta^{-1})^{2/\delta}$. Define

$$\mathcal{F}_{a, \eta, s, t} := \{f_{F, \sigma} : F([-a, a]) \geq 1 - \eta, \quad s \leq \sigma \leq t\}$$

and

$$\mathcal{F}_{a, s, t} := \{f_{F, \sigma} : F([-a, a]) = 1, \quad s \leq \sigma \leq t\}.$$

An estimate of the η -metric entropy of $\mathcal{F}_{a, \eta^2/16, s, t}$ is provided. Using the second inequality in (2.1) and $d_H^2(f, g) \leq \|f - g\|_1$, we have

$$\begin{aligned} \log D(\eta, \mathcal{F}_{a, \eta^2/16, s, t}, d_H) &\leq \log N(\eta/2, \mathcal{F}_{a, \eta^2/16, s, t}, d_H) \\ &\leq \log N(\eta^2/4, \mathcal{F}_{a, \eta^2/16, s, t}, \|\cdot\|_1) \\ &\leq \log N(\eta^2/8, \mathcal{F}_{a, s, t}, \|\cdot\|_1) \\ &\lesssim \left(\log \frac{1}{\eta} \right)^{2\kappa}, \end{aligned}$$

where the third inequality descends from Lemma A.3 of Ghosal and van der Vaart [8], page 1261, and the last one from Lemma A.9. The choice $\eta_n = \bar{\varepsilon}_n = n^{-1/2}(\log n)^\kappa$, $s_n = E(\log \eta_n^{-1})^{-2/\gamma}$, $t_n = \exp\{F(\log \eta_n^{-1})^2\}$ and $a_n = L(\log \eta_n^{-1})^{2/\delta}$ leads to conclude that, for $\mathcal{F}_n := \mathcal{F}_{a_n, \eta_n^2/16, s_n, t_n}$, condition (A.1) is verified because

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim (\log n)^{2\kappa} = n\bar{\varepsilon}_n^2.$$

Now we show that condition (A.2) is also satisfied. By the independence of F and σ ,

$$\begin{aligned} \pi(\mathcal{F}_{a, \eta^2/16, s, t}^c) &= \pi(\{(F, \sigma) : F([-a, a]) \geq 1 - \eta^2/16, \quad \sigma \in [s, t]^c\}) \\ &\quad + \Pi(\{F : F([-a, a]) < 1 - \eta^2/16\}) \\ &\leq [G(s) + 1 - G(t)] + \Pi(\{F : F([-a, a]^c) > \eta^2/16\}) \\ &\leq [G(s) + 1 - G(t)] + 16 \frac{\alpha([-a, a]^c)}{\eta^2 \alpha(\mathbb{R})} \\ &\lesssim e^{-ds^{-\gamma}} + t^{-\varpi} + \frac{e^{-ba^\delta}}{\eta^2}, \end{aligned} \tag{4.3}$$

where in (4.3) we have used Markov's inequality and the fact that $F([-a, a]^c)$ has a beta distribution with parameters $\alpha([-a, a]^c)$ and $\alpha([-a, a])$. The last line follows from (3.1) of assumption (A) and from condition (4.1). Take η_n, s_n, t_n, a_n and \mathcal{F}_n as before. For a positive constant c_2 which, as shown below, arises from condition (A.3), let $0 < E \leq [d/16(c_2 + 4)]^{1/\gamma}$, $F \geq [16(c_2 + 4)/\varpi]$ and $L \geq [(16(c_2 + 4) + 1)/b]^{1/\delta}$. Then,

$$\begin{aligned} \pi(\mathcal{F}_n^c) &\lesssim e^{-ds_n^{-\gamma}} + t_n^{-\varpi} + \frac{e^{-ba_n^\delta}}{\eta_n^2} \\ &\lesssim \exp\{-(c_2 + 4)(\log n)^2\} \\ &\quad + \exp\left\{\left(\log \frac{1}{\eta_n^2}\right) - [16(c_2 + 4) + 1] \left(\log \frac{1}{\eta_n}\right)^2\right\} \\ &\lesssim \exp\{-(c_2 + 4)n\tilde{\varepsilon}_n^2\}. \end{aligned}$$

It remains to be checked that condition (A.3) is satisfied. We show that, for small enough $\varepsilon > 0$, there exist constants $c_1, c_2 > 0$ so that

$$\pi(B_{\text{KL}}(f_0; \varepsilon^2)) \geq c_1 \exp\left\{-c_2 \left(\log \frac{1}{\varepsilon}\right)^2\right\}. \quad (4.4)$$

The proof is in the same spirit as that of Theorem 5.1 by Ghosal and van der Vaart [8], pages 1251–1253. Let $0 < \varepsilon \leq [(\sigma_0/2) \wedge (1 - e^{-1})/\sqrt{2}]$ be fixed. Let $[-a_0, a_0]$ be the support of F_0 . By Lemma A.5, there exists a discrete distribution F'_0 (depending on ε), supported on (at most) $N \lesssim \log(1/\varepsilon)$ points $\theta_1, \dots, \theta_N$ in $[-a_0, a_0]$ that are at least 2ε -separated, such that

$$\|f_{F'_0, \sigma_0} - f_0\|_\infty \lesssim \varepsilon.$$

By Lemma A.8, $\|f_{F'_0, \sigma_0} - f_0\|_1 \lesssim \varepsilon(\log \varepsilon^{-1})^{1/p}$. Represent F'_0 as $\sum_{j=1}^N w_j \delta_{\theta_j}$, where $|\theta_j - \theta_k| \geq 2\varepsilon$ for all $j \neq k$. For any probability measure F on \mathbb{R} such that

$$\sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon]) - w_j| \leq \varepsilon$$

and any $\sigma > 0$ such that $|\sigma - \sigma_0| \leq \varepsilon$, we have

$$\begin{aligned} \|f_{F, \sigma} - f_{F'_0, \sigma_0}\|_1 &\lesssim \|\psi_{\sigma, p} - \psi_{\sigma_0, p}\|_1 + \frac{\varepsilon}{\sigma \wedge \sigma_0} \\ &\quad + 2 \sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon]) - w_j| \\ &\lesssim \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} + \frac{\varepsilon}{\sigma \wedge \sigma_0} + 2\varepsilon \lesssim \varepsilon, \end{aligned} \quad (4.5)$$

where in (4.5) we have used inequality (A.6). It follows that

$$d_{\mathbb{H}}^2(f_{F,\sigma}, f_0) \leq \|f_{F,\sigma} - f_{F'_0,\sigma_0}\|_1 + \|f_{F'_0,\sigma_0} - f_0\|_1 \lesssim \varepsilon \left(\log \frac{1}{\varepsilon}\right)^{1/p}.$$

Since $[\theta_j - \varepsilon, \theta_j + \varepsilon] \subset [-a_0 - 1, a_0 + 1]$ for all $j = 1, \dots, N$, we have $F([-a_0 - 1, a_0 + 1]) > 1/2$. By Lemma A.10,

$$\left\{ f_{F,\sigma} : \sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon]) - w_j| \leq \varepsilon, \quad |\sigma - \sigma_0| \leq \varepsilon \right\} \subseteq B_{\text{KL}} \left(f_0; \varepsilon \left(\log \frac{1}{\varepsilon}\right)^{2+1/p} \right). \quad (4.6)$$

Using the independence of F and σ , together with the assumptions on the base measure α and the prior G for σ , we obtain the bound in (4.4). Thus, condition (A.3) is satisfied with $\tilde{\varepsilon}_n = n^{-1/2}(\log n)$ and the proof is complete. \square

Remark 4.1. Given $p = 2m$, $m \in \mathbb{N}$, the best rate is obtained when $\kappa = 1$. This implies having $\delta \geq 2p$, namely, the base measure α should have quite rapidly decaying tails. Note that the rate depends on the tail behaviour of the prior G for σ only through its exponential decay rate γ at zero.

Remark 4.2. If $p = 2$, *i.e.*, the kernel is Gaussian, and $\gamma = \infty$, *i.e.*, the scale parameter σ is bounded below away from zero, $\sigma \in [\underline{\sigma}, \infty)$ for some known $\underline{\sigma} > 0$, then (an upper bound on) the rate is

$$\varepsilon_n = n^{-1/2}(\log n)^\kappa, \quad \kappa = \frac{1}{2} + \left(\frac{1}{2} \vee \frac{2}{\delta}\right),$$

which is the same found by Ghosal and van der Vaart [8] in Theorem 5.1, page 1250, wherein, except for assumption (A), the same conditions as in Theorem 4.1 are postulated.

The assumption of Theorem 4.1 that the true mixing distribution F_0 has compact support can be relaxed to allow for an unbounded set of locations, without affecting the nearly parametric rate, by requiring F_0 to have sub-exponential tails and the base measure α to have an EP density with suitably constrained shape parameter.

Theorem 4.2. Let p be a fixed even integer. Suppose that $f_0 = F_0 * \psi_{\sigma_0,p}$, with the true mixing distribution F_0 satisfying the tail condition

$$F_0(\{\theta : |\theta| > t\}) \lesssim e^{-c_0 t^p}, \quad \text{for large } t > 0, \quad (4.7)$$

for some constant $c_0 > 0$. If the base measure α has a density α' such that, for constants $b > 0$ and $0 < \delta \leq p$, satisfies

$$\alpha'(\theta) \propto e^{-b|\theta|^\delta}, \quad \theta \in \mathbb{R}, \quad (4.8)$$

the prior distribution G for σ satisfies assumption (A), then the posterior rate of convergence relative to $d_{\mathbb{H}}$ is $\varepsilon_n = n^{-1/2}(\log n)^\kappa$, where

$$\kappa = \frac{1}{2} + p \left(\frac{1}{\delta} + \frac{1}{\gamma} \right). \quad (4.9)$$

Proof. First, note that a density α' , as assumed in (4.8), is continuous and positive on \mathbb{R} and that, for $\delta \geq 1$, the corresponding base measure α satisfies the tail condition (4.1) in virtue of inequality (A.14). For $\delta \in (0, 1)$, inequality (A.15) incurs an additional factor of $a^{1-\delta}$ in the upper bound for $\alpha([-a, a]^c)$ in (4.3), which can be absorbed into e^{-ba^δ} for large enough b . Thus, condition (A.2) is satisfied. Furthermore, since $(p/\delta) \geq 1$, the value of κ in (4.2) reduces to that in (4.9).

The proof differs from that of Theorem 4.1 only in the derivation of the bound $\pi(B_{\text{KL}}(f_0; \varepsilon^2)) \gtrsim \exp\{-c_2(\log \varepsilon^{-1})^2\}$, as $\varepsilon \rightarrow 0$. For given $0 < \varepsilon \leq [(\sigma_0/2) \wedge (1 - e^{-1})/\sqrt{2}]$, let $a_\varepsilon := (c_0^{-1} \log \varepsilon^{-1})^{1/p}$. Let F_0^* be the re-normalized restriction of F_0 to $[-a_\varepsilon, a_\varepsilon]$. Using Lemma A.3 of Ghosal and van der Vaart [8], page 1261, and assumption (4.7), we have $\|f_{F_0^*, \sigma_0} - f_0\|_1 \lesssim \varepsilon$. By Lemma A.5, there exists a discrete distribution F'_0 , which matches the (finite) moments of F_0^* up to the order $N \lesssim \log(1/\varepsilon)$ and has at most N support points in $[-a_\varepsilon, a_\varepsilon]$ that are (at least) 2ε -separated, such that $\|f_{F'_0, \sigma_0} - f_{F_0^*, \sigma_0}\|_\infty \lesssim \varepsilon$. By Lemma A.8, $\|f_{F'_0, \sigma_0} - f_{F_0^*, \sigma_0}\|_1 \lesssim \varepsilon(\log \varepsilon^{-1})^{1/p}$. Represented F'_0 as $\sum_{j=1}^N w_j \delta_{\theta_j}$, with $|\theta_j - \theta_k| \geq 2\varepsilon$ for all $j \neq k$, for any distribution F on \mathbb{R} such that

$$\sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon]) - w_j| \leq \varepsilon \quad (4.10)$$

and any $\sigma > 0$ such that $|\sigma - \sigma_0| \leq \varepsilon$, using the same chain of inequalities as in (4.5), we have $\|f_{F, \sigma} - f_{F'_0, \sigma_0}\|_1 \lesssim \varepsilon$. Thus,

$$\begin{aligned} d_{\mathbb{H}}^2(f_{F, \sigma}, f_0) &\leq \|f_{F, \sigma} - f_{F'_0, \sigma_0}\|_1 + \|f_{F'_0, \sigma_0} - f_{F_0^*, \sigma_0}\|_1 + \|f_{F_0^*, \sigma_0} - f_0\|_1 \\ &\lesssim \varepsilon \left(\log \frac{1}{\varepsilon} \right)^{1/p}. \end{aligned}$$

Now we can invoke Lemma A.10, taking into account Remark A.3. To this end, note that, if F satisfies relationship (4.10), then $F([-a_\varepsilon - 1, a_\varepsilon + 1]) > 1/2$. Thus, inclusion (4.6) holds true. To apply Lemma A.2 of Ghosal and van der Vaart [8], pages 1260–1261, observe that, for each $|\theta_j| \leq a_\varepsilon$, we have $\alpha([\theta_j - \varepsilon, \theta_j + \varepsilon]) \gtrsim \varepsilon e^{-ba_\varepsilon^\delta} \gtrsim \varepsilon^{b'}$ for some constant $b' > 0$, because, by assumption, $(\delta/p) \leq 1$. Then, the proof can be completed by laying out the same arguments as in Theorem 4.1. \square

Remark 4.3. Given $p = 2m$, $m \in \mathbb{N}$, the best rate is obtained for $\delta = p$, namely, when the tail decay rate of the base measure α equals the shape parameter of the kernel. In such a case,

$$\kappa = \frac{3}{2} + \frac{p}{\gamma}.$$

If, in addition, $\gamma = \infty$, *i.e.*, the scale parameter σ is bounded below away from zero, $\sigma \in [\underline{\sigma}, \infty)$ for some known $\underline{\sigma} > 0$, then $\varepsilon_n = n^{-1/2}(\log n)^{3/2}$. Note that this rate does not depend on p . Also, it agrees with the one obtained by Ghosal and van der Vaart [8] in Theorem 5.2, pages 1254–1255, for mixtures of normal densities with standard deviation σ lying in a bounded interval $[\underline{\sigma}, \bar{\sigma}] \subset (0, \infty)$.

Remark 4.4. Given $p > 0$, if the true density $f_0 = F_0 * \psi_{\sigma_0, p}$, then

$$\begin{aligned} \psi_{\sigma_0, p}(2x) \lesssim f_0(x) &\leq \psi_{\sigma_0, p}(x/2) \\ &+ \frac{1}{\sigma_0 c_p} F_0(\{\theta : |\theta| \geq |x|/2\}), \quad \text{for large } |x|. \end{aligned} \quad (4.11)$$

If the true mixing distribution F_0 has sub-exponential tails, *i.e.*, for constants $c_0, q > 0$,

$$F_0(\{\theta : |\theta| > t\}) \lesssim e^{-c_0 t^q}, \quad \text{for large } t > 0,$$

then it is the term with the slower decay rate that dominates in the upper bound in (4.11),

$$\psi_{\sigma_0, p}(2x) \lesssim f_0(x) \lesssim e^{-c_1 |x|^{p \wedge q}}, \quad \text{for large } |x|.$$

Therefore, under assumption (4.7) of Theorem 4.2,

$$f_0(x) \sim e^{-C_0 |x|^p},$$

this meaning that $f_0(x)/e^{-C_0 |x|^p} \rightarrow c$, as $|x| \rightarrow \infty$, where c is a positive constant. Theorem 4.2 contemplates only the case where F_0 has tail decay rate at least as fast as that of the kernel, as when, for example, f_0 is the convolution of two normals,

$$f_0 = \phi_{\tau_0} * \phi_{\sigma_0} = \phi_{(\tau_0^2 + \sigma_0^2)^{1/2}}.$$

If, instead, f_0 is the convolution of a Gaussian kernel and a Laplace distribution, $f_0 = \psi_{\tau_0, 1} * \phi_{\sigma_0}$, then Theorem 4.2 does not apply.

The result of Theorem 4.2 for the important special case of a Gaussian kernel with an inverse-gamma prior on σ^2 is separately stated in the following corollary.

Corollary 4.1. *Suppose that f_0 is a mixture of normals, $f_0 = F_0 * \phi_{\sigma_0}$, with the true mixing distribution F_0 having sub-Gaussian tails,*

$$F_0(\{\theta : |\theta| > t\}) \lesssim e^{-c_0 t^2}, \quad \text{for large } t > 0,$$

for some constant $c_0 > 0$. If the base measure α is normal and the prior on σ^2 is an inverse-gamma, then the posterior rate of convergence relative to $d_{\mathbb{H}}$ is $\varepsilon_n = n^{-1/2}(\log n)^{5/2}$.

In the next corollary, we state a result on the posterior expected density \hat{f}_n , for which an explicit expression can be found in Lo [17], Theorem 2, pages 353–354. The “in probability” statement is understood to be with respect to P_0^n .

Corollary 4.2. *Under the conditions of Theorem 4.1 or Theorem 4.2,*

$$d_H(\hat{f}_n, f_0) = O_P(n^{-1/2}(\log n)^\kappa),$$

where κ is as in (4.2) or (4.9), respectively.

In Theorem 4.1 and Theorem 4.2, verification of the remaining mass condition (A.2) for \mathcal{F}_n^c has led us to ask for the base measure α to have either sub-exponential tails or a density α' of a prescribed form. In the next theorem, the use of Lemma A.11, which provides an upper bound on the posterior probability of \mathcal{F}_n^c by exploiting the properties of the Dirichlet process, allows us to impose a less restrictive condition on α . In fact, condition (4.12) requires α' to have tails only bounded below by those of an EP density. On the other hand, however, we consider a stronger condition on the upper tail of the prior G for σ , a request that might be due to the method of proof. The requirement on G is formalized hereafter as condition (A') for easy reference.

(A') G has a continuous and positive Lebesgue density g on an interval containing σ_0 and, for constants $d_1, d_2 > 0, \gamma_1, \gamma_2 \in (0, \infty]$, satisfies

$$G(s) \lesssim e^{-d_1 s^{-\gamma_1}} \quad \text{as } s \rightarrow 0 \quad \text{and} \quad 1 - G(s) \lesssim e^{-d_2 s^{\gamma_2}} \quad \text{as } s \rightarrow \infty.$$

Assumption (A') differs from assumption (A) because it requires G to have an exponentially decaying tail also at infinity. This rules out the possibility of using an inverse-gamma prior on σ^2 , unless σ is known to lie in some interval $(0, \bar{\sigma}]$, with $0 < \bar{\sigma} < \infty$: in such a case, in fact, a right-truncated inverse-gamma distribution trivially satisfies the tail condition at infinity. For example, a prior distribution for σ verifying condition (A') may have density of the form

$$g(\sigma) = \frac{\nu e^\beta}{1 + \nu} [\beta \sigma^{-2} e^{-\beta/\sigma} I_{(0, 1]}(\sigma) + \beta \sigma^{\nu-1} e^{-\beta \sigma^\nu} I_{(1, \infty)}(\sigma)], \quad \sigma > 0,$$

with parameters $\beta, \nu > 0$. Note that g is a continuous and positive density proportional to an inverse-gamma $IG(1; \beta)$ on $(0, 1]$ and to a Weibull $W(\nu; \beta)$ on $(1, \infty)$. Condition (A') is satisfied with $\gamma_1 = 1$ and $\gamma_2 = \nu$, *i.e.*,

$$G(s) = \frac{\nu e^\beta}{1 + \nu} e^{-\beta s^{-1}}, \quad 0 < s \leq 1,$$

and

$$1 - G(s) = \frac{e^\beta}{1 + \nu} e^{-\beta s^\nu}, \quad s > 1.$$

The conditions on G appearing in (A') were also postulated by Ghosal and van der Vaart [10], page 699, to assess posterior rates of convergence for Dirichlet normal mixtures at smooth densities. However, these authors consider a sample-size-dependent prior on σ : in their set-up, in fact, G is the distribution of σ/σ_n , with σ_n a sequence of positive real numbers such that $n^{-a_1} \lesssim \sigma_n \lesssim n^{-a_2}$, for some $0 < a_2 < a_1 < 1$.

Theorem 4.3. *Let p be a fixed even integer. Suppose that $f_0 = F_0 * \psi_{\sigma_0, p}$, with the true mixing distribution F_0 having either compact support or sub-exponential tails as in (4.7). If the base measure α has a continuous and positive density α' such that, for constants $b > 0$ and $0 < \delta < p$, satisfies*

$$\alpha'(\theta) \gtrsim e^{-b|\theta|^\delta}, \quad \text{for large } |\theta|, \tag{4.12}$$

the prior distribution G for σ satisfies condition (A'), then the posterior rate of convergence relative to d_H is $\varepsilon_n = n^{-1/2}(\log n)^\kappa$, where

$$\kappa = \frac{1}{2} + p \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2} \right) + \left(\frac{1}{2} \vee \frac{p\delta}{\gamma_2(p - \delta)} \right). \tag{4.13}$$

Proof. For sequences a_n, η_n, s_n and t_n to be suitably chosen later on, let \mathcal{F}_n be defined as in Theorem 4.1. The proof differs from that of Theorem 4.1 for the arguments laid out to show that the convergence $\pi(\mathcal{F}_n^c | X_1, \dots, X_n) \rightarrow 0$, in P_0^n -probability, as $n \rightarrow \infty$, holds true. Note that

$$\begin{aligned} \pi(\mathcal{F}_n^c | X_1, \dots, X_n) &\leq \Pr(\sigma < s_n | X_1, \dots, X_n) + \Pr(\sigma > t_n | X_1, \dots, X_n) \\ &\quad + \Pr(\{F : F([-a_n, a_n]^c) > \eta_n^2/16\} | X_1, \dots, X_n) \\ &=: J_n^{(1)} + J_n^{(2)} + J_n^{(3)}. \end{aligned}$$

Let $c_2 > 0$ be the constant arising from condition (A.3). For constants $E \leq [d_1/16(c_2 + 4)]^{1/\gamma_1}$ and $F \geq [16(c_2 + 4)/d_2]^{1/\gamma_2}$, let $s_n = E(\log \eta_n^{-1})^{-2/\gamma_1}$ and $t_n = F(\log \eta_n^{-1})^{2/\gamma_2}$. For $k = 1, 2$, the sequence $J_n^{(k)} \xrightarrow{P} 0$, because $\mathbb{E}_0^n[J_n^{(k)}] \rightarrow 0$ in virtue of Lemma 1 in Ghosal and van der Vaart [9], page 195, (see also Lemma 5 of Barron, Schervish and Wasserman [1], pages 543–544). To deal with $J_n^{(3)}$, note that

$$\begin{aligned} \mathbb{E}_0^n[J_n^{(3)}] &\leq \mathbb{E}_0^n[\Pr(\{F : F([-a_n, a_n]^c) > \eta_n^2/16\} | X_1, \dots, X_n) I_{\{\max_{1 \leq i \leq n} |X_i| \leq a_n/2\}}] \\ &\quad + \mathbb{E}_0^n[I_{\{\max_{1 \leq i \leq n} |X_i| > a_n/2\}}]. \end{aligned} \tag{4.14}$$

To show that

$$\mathbb{E}_0^n[I_{\{\max_{1 \leq i \leq n} |X_i| > a_n/2\}}] = 1 - [1 - P_0(|X_1| > a_n/2)]^n \rightarrow 0,$$

we distinguish the two cases where F_0 has compact support or sub-exponential tails:

- (i) if F_0 has compact support, i.e., $F_0([-a_0, a_0]) = 1$ for some $a_0 > 0$, then for sufficiently large n so that $(a_n/4) \geq a_0$,

$$P_0(|X_1| > a_n/2) \leq (4\sigma_0)^p a_n^{-(p-1)} \psi_{\sigma_0, p}(a_n/4) =: U_n;$$

(ii) if F_0 has sub-exponential tails as in (4.7), then

$$P_0(|X_1| > a_n/2) \lesssim U_n + \frac{4^p}{p\sigma_0 c_p c_0} a_n^{-(p-1)} \exp\left\{-c_0 \left(\frac{a_n}{4}\right)^p\right\}.$$

For a_n such that, for some constant $\varrho \geq 0$,

- in case (i),

$$\frac{1}{p\sigma_0^p} \left(\frac{a_n}{4}\right)^p \geq (1 + \varrho)(\log n), \tag{4.15}$$

- in case (ii),

$$\left(\frac{1}{p\sigma_0^p} \wedge c_0\right) \left(\frac{a_n}{4}\right)^p \geq (1 + \varrho)(\log n), \tag{4.15'}$$

we have $nP_0(|X_1| > a_n/2) \rightarrow 0$, hence $\mathbb{E}_0^n[I_{\{\max_{1 \leq i \leq n} |X_i| > a_n/2\}}] \rightarrow 0$. Note that, since $p > 1$ and $a_n \rightarrow \infty$, it suffices to take $\varrho = 0$ in (4.15) or (4.15').

The first term of (4.14) can be handled by appealing to Lemma A.11, with $\eta^2 = \eta_n^2/16$, $T = t_n$ and $a = a_n = L(\log \eta_n^{-1})^\xi$, where

$$\xi = \frac{2}{\gamma_2} + \left(\frac{1}{p} \vee \frac{2\delta}{\gamma_2(p - \delta)}\right)$$

and L is a positive constant chosen so that, not only is condition (4.15) or (4.15') satisfied, but also for some $\omega > 0$,

$$n \exp\left\{-\frac{(a_n/t_n)^p}{p2^{p+1}} + ba_n^\delta + 2 \log \frac{1}{\eta_n}\right\} \leq n \exp\left\{\log \frac{1}{n^{1+\omega}}\right\} = \frac{1}{n^\omega} \rightarrow 0.$$

If $n\eta_n^2 \rightarrow \infty$, then also the first term in (A.17) converges to zero. To complete the proof, it remains to choose η_n so that condition (A.1) is also satisfied. From Theorem 4.1, which appeals to Lemma A.9, it is known that

$$\log D(\eta_n, \mathcal{F}_n, d_H) \lesssim \left[\left(\frac{a_n}{s_n}\right)^p \vee \left(\log \frac{1}{\eta_n}\right)\right] \times \left(\log \frac{1}{\eta_n}\right).$$

For $\eta_n = \bar{\varepsilon}_n = n^{-1/2}(\log n)^\kappa$, with κ as in (4.13), we have $\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim n\bar{\varepsilon}_n^2$. Also, $n\eta_n^2 = n\bar{\varepsilon}_n^2 = (\log n)^{2\kappa} \rightarrow \infty$. Recalling that $\tilde{\varepsilon}_n = n^{-1/2}(\log n)$, an upper bound on the rate is given by $\varepsilon_n := (\bar{\varepsilon}_n \vee \tilde{\varepsilon}_n) = \bar{\varepsilon}_n$. \square

Remark 4.5. Let $\text{supp}(g)$ denote the support of the prior density g for σ . An inspection of the proof of Theorem 4.3 shows that, when $\text{supp}(g) \subseteq (0, \bar{\sigma}]$, for some $0 < \bar{\sigma} < \infty$, then δ is also allowed to take the value p , i.e., $0 < \delta \leq p$. Furthermore, it turns out that $\kappa = 1 + p/\gamma$, with γ the tail decay rate of G at zero. This rate is better than the one we would get using Theorem 4.2: in fact, since $(p/\delta) \geq 1$, it results $\frac{1}{2} + p/\delta + p/\gamma > 1 + p/\gamma$.

Remark 4.6. If G has compact support $[\underline{\sigma}, \bar{\sigma}] \subset (0, \infty)$, which corresponds to having $\gamma_1 = \gamma_2 = \infty$, then $\kappa = 1$ and, for any value of $p = 2m$, $m \in \mathbb{N}$, an upper bound on the rate is $n^{-1/2}(\log n)$, which is the same obtained by Walker,

Lijoi and Prünster [29], pages 742–744, for the special case where a normal base measure α is used and a normal mixture f_0 , with mixing distribution F_0 having sub-Gaussian tails, is considered. Theorem 4.3 extends their result to EP kernels using a different approach and improves it on by relaxing the condition on the tail behaviour of α' from a sub-exponential to an over-exponential decay rate.

Corollary 4.3. *Under the conditions of Theorem 4.3,*

$$d_H(\hat{f}_n, f_0) = O_P(n^{-1/2}(\log n)^\kappa),$$

where κ is as in (4.13).

4.2. Non-analytic kernels

In this subsection, we study the case where the kernel is not infinitely differentiable at zero. This case corresponds to values of $p \neq 2m$, $m \in \mathbb{N}$. Let \underline{p} denote the greatest integer strictly smaller than p . If p is not an even integer, then the kernel is only \underline{p} -times differentiable at the origin. Precisely, for $p \in (0, 1]$, the kernel is not differentiable at zero. For $p \in (1, 2)$, the kernel is only one time differentiable at zero, with

$$\psi_{\sigma,p}^{(1)}(0) = 0.$$

In fact,

$$\psi_{\sigma,p}^{(1)}(x) = -\frac{1}{\sigma^p} \psi_{\sigma,p}(x) |x|^{p-1} \operatorname{sgn}(x), \quad \forall x \in \mathbb{R}, \tag{4.16}$$

with $0 < (p - 1) < 1$ in $|x|^{p-1}$. When $p > 2$ and $p \neq 2m$, $m \in \mathbb{N}$, for every $1 \leq k \leq (\underline{p} - 1)$, by Leibniz formula for the k th derivative of a product of functions applied to (4.16), for any $x \in \mathbb{R}$,

$$\begin{aligned} \psi_{\sigma,p}^{(k+1)}(x) &:= D^k \psi_{\sigma,p}^{(1)}(x) \\ &= -\frac{1}{\sigma^p} \sum_{l=0}^k \binom{k}{l} D^l (|x|^{p-1} \operatorname{sgn}(x)) D^{k-l} \psi_{\sigma,p}(x) \\ &= -\frac{1}{\sigma^p} |x|^{p-1} \operatorname{sgn}(x) \psi_{\sigma,p}^{(k)}(x) \\ &\quad -\frac{1}{\sigma^p} \sum_{l=1}^k \binom{k}{l} \left[\prod_{h=0}^{l-1} (p-1-h) \right] |x|^{(p-1)-l} [\operatorname{sgn}(x)]^{1+l} \psi_{\sigma,p}^{(k-l)}(x), \end{aligned}$$

where $D^k f$ denotes the k th derivative of a function f , with the convention that $D^0 f = f$. Thus, for each $j = 1, \dots, \underline{p}$,

$$\psi_{\sigma,p}^{(j)}(0) = 0.$$

Note that, for $j = \underline{p}$, because of the term $|x|^{p-\underline{p}} [\operatorname{sgn}(x)]^{\underline{p}}$, where $0 < (p - \underline{p}) \leq 1$, the function $\psi_{\sigma,p}^{(\underline{p})}$ cannot be further differentiated at zero.

The fact that $\psi_{\sigma,p}$ is not infinitely differentiable at the origin seems to play a key role in the search for a finitely supported approximating mixing distribution F'_0 of a given F_0 with a restricted number of atoms. When $p \neq 2m$, $m \in \mathbb{N}$, the arguments of Lemma 3.1 in Ghosal and van der Vaart [8], pages 1240–1241, cannot be used to find F'_0 by the moment matching condition, because θ and x do not separate by factorization (see their equation (3.8), page 1241). An approximating mixing distribution can be found using the matching condition combined with a preliminary partitioning argument, but this incurs an additional factor of $\varepsilon^{-(1 \vee 1/p)}$ in the number of support points (cf. Lemma A.5), leading to a slower than nearly parametric rate.

Theorem 4.4. *Let $p > 1/2$ be such that $p \neq 2m$, $m \in \mathbb{N}$. Suppose that $f_0 = F_0 * \psi_{\sigma_0,p}$, with the true mixing distribution F_0 having either compact support or sub-exponential tails as in (4.7) and the true scale σ_0 lying in a compact interval $[\underline{\sigma}, \bar{\sigma}] \subset (0, \infty)$. If the base measure α has a continuous and positive density α' such that, for constants $b > 0$ and $0 < \delta \leq p$, satisfies (4.12), the prior G for σ is supported on $[\underline{\sigma}, \bar{\sigma}]$ and has a continuous and positive density on an interval containing σ_0 , then the posterior rate of convergence relative to d_H is*

$$\varepsilon_n = n^{-1/2[1+(1 \vee 1/p)]} (\log n)^\kappa,$$

where $\kappa > 0$ depends on p .

Proof. We show that conditions (A.1) and (A.3) of Theorem A.1 are satisfied by sequences $\bar{\varepsilon}_n = n^{-1/2[1+(1 \vee 1/p)]} (\log n)^\varsigma$ and $\tilde{\varepsilon}_n = n^{-1/2[1+(1 \vee 1/p)]} (\log n)^\kappa$, with $0 < \varsigma < \kappa$, so that $\varepsilon_n := (\bar{\varepsilon}_n \vee \tilde{\varepsilon}_n) = \tilde{\varepsilon}_n$. We begin to consider condition (A.1). Given $\eta \in (0, 1/5)$, for a constant $L > 0$ to be suitably chosen, let $0 < a \leq L(\log \eta^{-1})^{1/p}$. Let $\mathcal{F}_{a,\eta,\underline{\sigma},\bar{\sigma}}$ and $\mathcal{F}_{a,\underline{\sigma},\bar{\sigma}}$ be defined as in the proof of Theorem 4.1. Reasoning as in Lemma 3 of Ghosal and van der Vaart [10], pages 705–707, and using the result of Lemma A.5, it can be proved that, for any $p > 0$ such that $p \neq 2m$, $m \in \mathbb{N}$, if $(a/\underline{\sigma}) \lesssim (\log \eta^{-1})^{1/p}$, then for small enough $\eta > 0$,

$$\begin{aligned} \log N(\eta(\log \eta^{-1})^{1/p}, \mathcal{F}_{a,\underline{\sigma},\bar{\sigma}}, \|\cdot\|_1) &\lesssim \log \left(\frac{\bar{\sigma}}{\underline{\sigma}\eta} \right) \\ &+ \left[\eta^{-(1 \vee 1/p)} \left(\log \frac{1}{\eta} \right)^{1/p} \right] \times \left[\log \left(\frac{2a}{\underline{\sigma}\eta^{1 \vee 1/p}} + 1 \right) + \log \frac{1}{\eta} \right]. \end{aligned} \quad (4.17)$$

Thus,

$$\begin{aligned} \log D(\eta^{1/2}(\log \eta^{-1})^{1/2p}, \mathcal{F}_{a,\eta/16,\underline{\sigma},\bar{\sigma}}, d_H) &\leq \log N(\eta^{1/2}(\log \eta^{-1})^{1/2p}/2, \mathcal{F}_{a,\eta/16,\underline{\sigma},\bar{\sigma}}, d_H) \\ &\leq \log N(\eta(\log \eta^{-1})^{1/p}/4, \mathcal{F}_{a,\eta/16,\underline{\sigma},\bar{\sigma}}, \|\cdot\|_1) \\ &\leq \log N(\eta(\log \eta^{-1})^{1/p}/8, \mathcal{F}_{a,\underline{\sigma},\bar{\sigma}}, \|\cdot\|_1) \\ &\lesssim \eta^{-(1 \vee 1/p)} \left(\log \frac{1}{\eta} \right)^{1+1/p}, \end{aligned}$$

where the third inequality descends from Lemma A.3 of Ghosal and van der Vaart [8], page 1261, and the last one from (4.17). Let

$$\varsigma = \frac{(1 + 1/p) + (1/p)(1 \vee 1/p)}{2[1 + (1 \vee 1/p)]}.$$

Choosing

$$\eta_n^{1/2}(\log \eta_n^{-1})^{1/2p} = n^{-1/2[1+(1 \vee 1/p)]}(\log n)^\varsigma = \bar{\varepsilon}_n$$

and $a_n = L(\log \eta_n^{-1})^{1/p}$, for $\mathcal{F}_n := \mathcal{F}_{a_n, \eta_n/16, \underline{\sigma}, \bar{\sigma}}$, condition (A.1) is seen to be verified,

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim n\bar{\varepsilon}_n^2.$$

Now we show that $\pi(\mathcal{F}_n^c | X_1, \dots, X_n) \rightarrow 0$, in P_0^n -probability, as $n \rightarrow \infty$. We have

$$\pi(\mathcal{F}_n^c | X_1, \dots, X_n) \leq \Pr(\{F : F([-a_n, a_n]^c) > \eta_n/16\} | X_1, \dots, X_n) =: J_n,$$

where J_n plays the role of $J_n^{(3)}$ in the proof of Theorem 4.3. Note that, since the prior G for σ is compactly supported, for each $n \in \mathbb{N}$, the terms $J_n^{(1)}$ and $J_n^{(2)}$ are both equal to zero, P_0^n -almost surely. Now,

$$\begin{aligned} \mathbb{E}_0^n[J_n] &\leq \mathbb{E}_0^n[\Pr(\{F : F([-a_n, a_n]^c) > \eta_n/16\} | X_1, \dots, X_n) I_{\{\max_{1 \leq i \leq n} |X_i| \leq a_n/2\}}] \\ &\quad + \mathbb{E}_0^n[I_{\{\max_{1 \leq i \leq n} |X_i| > a_n/2\}}]. \end{aligned}$$

Both terms on the right-hand side can be seen to converge to zero as in the proof of Theorem 4.3. In particular, the first term is handled by appealing to Lemma A.11, with $\eta^2 = \eta_n/16$, $T = \bar{\sigma}$ and $a = a_n$. In this case, $\xi = 1/p$ because $\gamma_2 = \infty$. Also, $n\eta_n \rightarrow \infty$, which guarantees that the first addendum in (A.17) converges to zero.

To show that condition (A.3) is satisfied with $\tilde{\varepsilon}_n$, we provide a lower bound on $\pi(B_{\text{KL}}(f_0; \varepsilon^2))$, as $\varepsilon \rightarrow 0$. Let $0 < \varepsilon \leq [(\sigma_0/2) \wedge (1 - e^{-1})/\sqrt{2}]$ be fixed. We begin to consider the case where F_0 has compact support, say $[-a_0, a_0]$ for some $a_0 > 0$. By Lemma A.5, there exists a discrete distribution F'_0 (depending on ε), supported on at most

$$N \lesssim \varepsilon^{-(1 \vee 1/p)}(\log \varepsilon^{-1})^{1/p}$$

points $\theta_1, \dots, \theta_N \in [-a_0, a_0]$, such that

$$\|f_{F'_0, \sigma_0} - f_0\|_\infty \lesssim \varepsilon.$$

For small enough $\varepsilon > 0$, by Lemma A.8,

$$\|f_{F'_0, \sigma_0} - f_0\|_1 \lesssim \varepsilon \left(\log \frac{1}{\varepsilon} \right)^{1/p}.$$

For $p > 1/2$, without loss of generality, we may assume that the support points of F'_0 are at least $2\varepsilon^2$ -separated (cf. Ghosal and van der Vaart [8], page 1252). Represented F'_0 as $\sum_{j=1}^N w_j \delta_{\theta_j}$, with $|\theta_j - \theta_k| \geq 2\varepsilon^2$ for all $j \neq k$, for any probability measure F on \mathbb{R} such that

$$\sum_{j=1}^N |F([\theta_j - \varepsilon^2, \theta_j + \varepsilon^2]) - w_j| \leq \varepsilon^2$$

and any $\sigma > 0$ such that $|\sigma - \sigma_0| \leq \varepsilon$,

$$\begin{aligned} \|f_{F, \sigma} - f_{F'_0, \sigma_0}\|_1 &\lesssim \|\psi_{\sigma, p} - \psi_{\sigma_0, p}\|_1 + \frac{\varepsilon^2}{\sigma \wedge \sigma_0} \\ &\quad + 2 \sum_{j=1}^N |F([\theta_j - \varepsilon^2, \theta_j + \varepsilon^2]) - w_j| \\ &\lesssim \frac{|\sigma - \sigma_0|}{\sigma \wedge \sigma_0} + \frac{\varepsilon^2}{\sigma \wedge \sigma_0} + 2\varepsilon^2 \lesssim \varepsilon. \end{aligned}$$

Thus,

$$d_{\mathbb{H}}^2(f_{F, \sigma}, f_0) \leq \|f_{F, \sigma} - f_{F'_0, \sigma_0}\|_1 + \|f_{F'_0, \sigma_0} - f_0\|_1 \lesssim \varepsilon \left(\log \frac{1}{\varepsilon} \right)^{1/p}.$$

Reasoning as in the proof of Theorem 4.1, for a constant $c > 0$,

$$\begin{aligned} \left\{ f_{F, \sigma} : \sum_{j=1}^N |F([\theta_j - \varepsilon^2, \theta_j + \varepsilon^2]) - w_j| \leq \varepsilon^2, \quad |\sigma - \sigma_0| \leq \varepsilon \right\} \\ \subseteq B_{\text{KL}} \left(f_0; c\varepsilon \left(\log \frac{1}{\varepsilon} \right)^{2+1/p} \right). \end{aligned}$$

Note that, for $p > 1/2$, the condition $\varepsilon^2 \leq (1/N)$ of Lemma A.2 in Ghosal and van der Vaart [8], page 1260, is satisfied. Hence, for constants $c', c'' > 0$,

$$\begin{aligned} \pi(B_{\text{KL}}(f_0; c\varepsilon(\log \varepsilon^{-1})^{2+1/p})) &\gtrsim \exp\{-c'N \log(1/\varepsilon)\} \\ &\gtrsim \exp\{-c''\varepsilon^{-(1 \vee 1/p)}(\log \varepsilon^{-1})^{1+1/p}\}. \end{aligned}$$

Let $\tilde{\varepsilon}_n = n^{-1/2[1+(1 \vee 1/p)]}(\log n)^\kappa$ with

$$\kappa = \frac{(1 + 1/p) + (2 + 1/p)(1 \vee 1/p)}{2[1 + (1 \vee 1/p)]}.$$

Then, for a suitable constant $c_2 > 0$, we have $\pi(B_{\text{KL}}(f_0; \tilde{\varepsilon}_n^2)) \gtrsim \exp\{-c_2 n \tilde{\varepsilon}_n^2\}$.

Following the arguments in the proof of Theorem 4.2, the same bound can be obtained for the case where F_0 has sub-exponential tails. To complete the proof, note that $\kappa > \varsigma$, thus an upper bound on the rate is given by $\varepsilon_n = \tilde{\varepsilon}_n$. \square

Remark 4.7. The result covers the case of a normal-Laplace density f_0 , for which a rate of $n^{-1/4}(\log n)^{5/4}$ is obtained.

Some remarks are in order. Differently from Theorem 4.1, Theorem 4.2 and Theorem 4.3, the posterior rate has been derived here assuming that the scale lies in a compact interval. The result is based on inequalities that are explicit in the lower and upper bounds on the scale so that, in principle, a rate could be obtained also when σ is unconstrained. Yet, as the rate would be slower, the aim is not pursued. As far as we are aware, the (minimax) optimal rate of convergence, relative to the Hellinger or the L_1 -metric, for this density estimation problem is unknown. Admittedly, we have found only an upper bound on the posterior rate and are not able to say whether, except possibly for a logarithmic factor, this bound is sharp or not.

5. Final remarks

In this article, we have studied frequentist asymptotic properties of posterior distributions of Dirichlet EP mixture priors, focussing on rates of convergence. Some theoretical properties usually discussed for specific densities nested in the family of EP distributions, like the normal, have been extended to more general kernels with non-trivial implications on posterior rates. The discrepancy observed in the rates for analytic and non-analytic kernels seems to suggest that, for infinite mixtures, the accuracy of the posterior in quantifying the uncertainty on the true density may heavily depend on the regularity of the kernel. It should be mentioned that a similar behaviour has been recently noted also by de Jonge and van Zanten [4] in a regression setting, insofar that the regularity of the kernel influences the posterior contraction rate, leading to a slower rate when the kernel is not analytic. This feature deserves further investigation to clarify whether and, if so, why the lack of regularity of the kernel causes a loss in the rate.

The results on posterior rates are of interest not only in density estimation, as previously pointed out, but also in the context of linear regression with unknown error distribution, see, *e.g.*, Ghosal and van der Vaart [9], pages 205–207. In both settings, however, the cases where the true mixing distribution has tails decaying not exponentially fast or the true density is a scale mixture of normals, like the Cauchy or Student’s t distribution, are not covered. This suggests that a potential future direction to pursue is an extension of previous results to Dirichlet *scale* mixtures of EP distributions.

Appendix: Auxiliary results

This Appendix reports the statement of a theorem in the recent literature (cf. Theorem A.1), which is instrumental to derive the main results of the article, establishes some facts about the uniform approximation of EP mixtures by finite mixtures (cf. Lemma A.5) and provides an upper bound on the metric entropy

of a sieve set of EP mixtures (cf. Lemma A.9). Some lemmas are straightforward extensions to EP mixtures of results valid for normal mixtures, thus only those proofs requiring further arguments are presented.

Theorem A.1. [THEOREM 2.1 IN GHOSAL AND VAN DER VAART [8]]. *Let π_n be a sequence of priors on a class of densities \mathcal{F} equipped with a metric d that can be either the Hellinger or the one induced by the L_1 -norm. Suppose that for positive sequences $\bar{\varepsilon}_n, \tilde{\varepsilon}_n \rightarrow 0$ such that $n(\bar{\varepsilon}_n^2 \wedge \tilde{\varepsilon}_n^2) \rightarrow \infty$, constants $c_1, c_2, c_3, c_4 > 0$ and sets $\mathcal{F}_n \subseteq \mathcal{F}$, we have*

$$\log D(\bar{\varepsilon}_n, \mathcal{F}_n, d) \leq c_1 n \bar{\varepsilon}_n^2, \quad (\text{A.1})$$

$$\pi_n(\mathcal{F} \setminus \mathcal{F}_n) \leq c_3 e^{-(c_2+4)n\bar{\varepsilon}_n^2}, \quad (\text{A.2})$$

$$\pi_n(B_{\text{KL}}(f_0; \tilde{\varepsilon}_n^2)) \geq c_4 e^{-c_2 n \tilde{\varepsilon}_n^2}, \quad (\text{A.3})$$

where $B_{\text{KL}}(f_0; \tilde{\varepsilon}_n^2) := \{f : \int \log(f_0/f) dP_0 \leq \tilde{\varepsilon}_n^2, \int (\log(f_0/f))^2 dP_0 \leq \tilde{\varepsilon}_n^2\}$. Then, for $\varepsilon_n := (\bar{\varepsilon}_n \vee \tilde{\varepsilon}_n)$ and a sufficiently large constant $M > 0$, the posterior probability

$$\pi_n(\{f : d(f, f_0) > M\varepsilon_n\} | X_1, \dots, X_n) \rightarrow 0$$

in P_0^n -probability, as $n \rightarrow \infty$.

The following four lemmas present some useful inequalities.

Lemma A.1. *Let $p > 0$ be fixed. For any $\sigma > 0$ and $x, y \in \mathbb{R}$,*

$$\sup_{\theta \in \mathbb{R}} |\psi_{\sigma,p}(x-\theta) - \psi_{\sigma,p}(y-\theta)| \lesssim \frac{1}{\sigma} \left(\frac{|x-y|}{\sigma} \right)^{1 \wedge p}. \quad (\text{A.4})$$

Proof. Let $\sigma > 0$ and $x, y \in \mathbb{R}$ be fixed. We begin to consider the case where $p \in (0, 1)$. For any $\theta \in \mathbb{R}$,

$$|\psi_{\sigma,p}(x-\theta) - \psi_{\sigma,p}(y-\theta)| \leq \frac{1}{\sigma c_p} \left[1 - \exp \left\{ -\frac{1}{p} \left(\frac{|x-y|}{\sigma} \right)^p \right\} \right] \leq \frac{1}{p c_p} \frac{|x-y|^p}{\sigma^{1+p}},$$

where we have used the inequalities

$$|u-v|^p \leq |u-z|^p + |z-v|^p, \quad \forall u, v, z \in \mathbb{R}, \quad p \in (0, 1),$$

and $1 - e^{-t} \leq t$, for $t \geq 0$. Now, we deal with the case where $p \geq 1$. For any $\theta \in \mathbb{R}$,

$$\begin{aligned} |\psi_{\sigma,p}(x-\theta) - \psi_{\sigma,p}(y-\theta)| &= \left| \int_{|x-\theta| \wedge |y-\theta|}^{|x-\theta| \vee |y-\theta|} \psi_{\sigma,p}^{(1)}(z) dz \right| \\ &\leq \sup_{z \geq 0} |\psi_{\sigma,p}^{(1)}(z)| ||x-\theta| - |y-\theta|| \lesssim \frac{|x-y|}{\sigma^2}, \end{aligned}$$

because

$$\sup_{z \geq 0} |\psi_{\sigma,p}^{(1)}(z)| = \frac{1}{\sigma^2 c_p} \left(\frac{p-1}{e} \right)^{1-1/p}.$$

Thus, for any $p > 0$,

$$|\psi_{\sigma,p}(x - \theta) - \psi_{\sigma,p}(y - \theta)| \lesssim \frac{1}{\sigma} \left(\frac{|x - y|}{\sigma} \right)^{1 \wedge p}, \quad \forall \theta \in \mathbb{R},$$

and the result follows. \square

The following two bounds on the supremum and L_1 -norm of the difference between two EP mixtures are the same as those for normal mixtures (cf. the proof of Lemma 3 in Ghosal and van der Vaart [10], pages 705–707).

Lemma A.2. *Let $p > 0$ be fixed. For any probability measure F on \mathbb{R} and any pair $\sigma, \sigma' > 0$,*

$$\begin{aligned} \|F * \psi_{\sigma,p} - F * \psi_{\sigma',p}\|_{\infty} &\leq \|\psi_{\sigma,p} - \psi_{\sigma',p}\|_{\infty} \\ &\lesssim \frac{|\sigma - \sigma'|}{\sigma\sigma'} \lesssim \frac{|\sigma - \sigma'|}{(\sigma \wedge \sigma')^2} \end{aligned} \quad (\text{A.5})$$

and

$$\|F * \psi_{\sigma,p} - F * \psi_{\sigma',p}\|_1 \leq \|\psi_{\sigma,p} - \psi_{\sigma',p}\|_1 \leq 2 \frac{|\sigma - \sigma'|}{\sigma \wedge \sigma'}. \quad (\text{A.6})$$

Note that the constant in the upper bound (A.6) does not depend on p . The proof of the following inequality for the case where $p = 2$, i.e., the kernel is Gaussian, can be found within the proof of Lemma 1 of Ghosal, Ghosh and Ramamoorthi [7], pages 156–157.

Lemma A.3. *Let $p > 0$ be fixed. For any $\sigma > 0$ and any pair $\theta_j, \theta_k \in \mathbb{R}$,*

$$\|\psi_{\sigma,p}(\cdot - \theta_j) - \psi_{\sigma,p}(\cdot - \theta_k)\|_1 \leq 2 \|\psi_{1,p}\|_{\infty} \frac{|\theta_j - \theta_k|}{\sigma} \leq \frac{2}{c_p} \frac{|\theta_j - \theta_k|}{\sigma}. \quad (\text{A.7})$$

Lemma A.4. *Let $\sigma > 0$ be given. For any pair $p, p' > 0$,*

$$\|\psi_{\sigma,p} - \psi_{\sigma,p'}\|_1 = \frac{|p - p'|}{pp'} \|g_{(\cdot/\sigma)}^{(1)}(z^*)\|_1,$$

where $g_{(\cdot/\sigma)}(z) := \psi_{\sigma, z^{-1}}(\cdot)$ and $1/(p \vee p') < z^* < 1/(p \wedge p')$. Furthermore,

$$\|\psi_{\sigma,p} - \psi_{\sigma,p'}\|_1 \lesssim \frac{|p - p'|}{p \wedge p'} \times \begin{cases} 1, & \text{as } z^* \rightarrow 0, \\ \frac{1}{p \vee p'}, & \text{as } z^* \rightarrow \infty. \end{cases} \quad (\text{A.8})$$

Proof. Let $z := 1/p$ and

$$g_{(x/\sigma)}(z) := \psi_{\sigma, z^{-1}}(x) = \frac{z^z}{2\sigma\Gamma(1+z)} \exp\{-z(|x|/\sigma)^{1/z}\}, \quad z > 0.$$

By the Mean Value Theorem,

$$\|\psi_{\sigma,p} - \psi_{\sigma,p'}\|_1 = \|g_{(\cdot/\sigma)}(z) - g_{(\cdot/\sigma)}(z')\|_1 = \frac{|p-p'|}{pp'} \|g_{(\cdot/\sigma)}^{(1)}(z^*)\|_1,$$

where $1/(p \vee p') < z^* < 1/(p \wedge p')$ and

$$g_{(x/\sigma)}^{(1)}(z^*) = g_{(x/\sigma)}(z^*) [1 + \log z^* - \Psi(1 + z^*)] \\ + g_{(x/\sigma)}(z^*) (|x/\sigma|)^{1/z^*} [\log(|x/\sigma|)^{1/z^*} - 1], \quad x \in \mathbb{R},$$

with $\Psi(z) := D^1 \log \Gamma(z)$ the *Digamma* function. Clearly,

$$\|g_{(\cdot/\sigma)}^{(1)}(z^*)\|_1 \leq |1 + \log z^* - \Psi(1 + z^*)| \\ + \int g_{(x/\sigma)}(z^*) (|x/\sigma|)^{1/z^*} |\log(|x/\sigma|)^{1/z^*} - 1| dx \\ =: T_1 + T_2.$$

As $z^* \rightarrow 0$,

$$T_1 \lesssim (z^*)^{-1} + \gamma + O(z^*) \lesssim (z^*)^{-1},$$

where γ is the Euler-Mascheroni constant, and $T_2 \lesssim \sigma(z^*)^{-1}$. The first bound in (A.8) follows the fact that $(z^*)^{-1} < (p \vee p')$. As $z^* \rightarrow \infty$, $T_1 = O(1)$ and $T_2 = O(1)$, whence the second bound. \square

The next lemma establishes that, at least when the kernel is analytic, EP mixtures can be uniformly approximated by finite EP mixtures with a relatively small number of components.

Lemma A.5. *Let $p > 0$ be fixed. Let $0 < \varepsilon < 1$ and $a, \sigma > 0$ be given. Define $D := [(a/\sigma) \vee (p \log \varepsilon^{-1})^{1/p}]$. For any probability measure F on $[-a, a]$, there exists a discrete probability measure F' on $[-a, a]$, with at most*

$$N \lesssim \begin{cases} D^p, & \text{if } p = 2m, \\ D\varepsilon^{-(1 \vee 1/p)}, & \text{if } p \neq 2m, \end{cases} \quad m \in \mathbb{N},$$

support points, such that

$$\|F * \psi_{\sigma,p} - F' * \psi_{\sigma,p}\|_\infty \lesssim \frac{\varepsilon}{\sigma}. \tag{A.9}$$

Proof. In the case where p is an even integer the result can be proved as in Lemma 3.1 of Ghosal and van der Vaart [8], pages 1240–1241. In the case where $p \neq 2m$, $m \in \mathbb{N}$, for any constant $M > 0$ and any probability measure F' on $[-a, a]$,

$$\|F * \psi_{\sigma,p} - F' * \psi_{\sigma,p}\|_\infty \leq \sup_{|x| < M} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)| \tag{A.10}$$

$$+ \sup_{|x| \geq M} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)|. \tag{A.11}$$

For $M = 2[a \vee \sigma(p \log \varepsilon^{-1})^{1/p}] = 2\sigma D$, the term in (A.11) is bounded above, up to a constant, by (ε/σ) ,

$$\sup_{|x| \geq M} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)| \leq 2\psi_{\sigma,p}(M-a) \lesssim \frac{\varepsilon}{\sigma}.$$

To bound above the term in (A.10), we consider a finite partition of $(-M, M)$ into $k = \lceil (2M/\sigma)\varepsilon^{-(1 \vee 1/p)} \rceil = O(D\varepsilon^{-(1 \vee 1/p)})$ disjoint, consecutive intervals I_1, \dots, I_k each of length $l \leq \sigma\varepsilon^{1 \vee 1/p}$ and, possibly, a final interval I_{k+1} of length $l_{k+1} < \sigma\varepsilon^{1 \vee 1/p}$. Let J stand for the total number of intervals in the partition. It is either $J = k$ or $J = k + 1$. Then,

$$\begin{aligned} \sup_{|x| < M} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)| \\ = \max_{1 \leq j \leq J} \sup_{x \in I_j} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)|. \end{aligned}$$

For each $j = 1, \dots, J$, let x_j denote the left endpoint of the interval I_j . In fact, any point in I_j would serve the aim. Then,

$$\begin{aligned} \sup_{x \in I_j} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)| \\ = \sup_{x \in I_j} |(F * \psi_{\sigma,p})(x) \mp (F * \psi_{\sigma,p})(x_j) \mp (F' * \psi_{\sigma,p})(x_j) - (F' * \psi_{\sigma,p})(x)| \\ \leq 2 \sup_{x \in I_j} \sup_{|\theta| \leq a} |\psi_{\sigma,p}(x - \theta) - \psi_{\sigma,p}(x_j - \theta)| \\ + \left| \int_{-a}^a \psi_{\sigma,p}(x_j - \theta) d(F - F')(\theta) \right|, \quad (\text{A.12}) \end{aligned}$$

where, by inequality (A.4),

$$\sup_{|\theta| \leq a} |\psi_{\sigma,p}(x - \theta) - \psi_{\sigma,p}(x_j - \theta)| \lesssim \frac{1}{\sigma} (|x - x_j|/\sigma)^{1 \wedge p} \lesssim \frac{\varepsilon}{\sigma}, \quad \forall x \in I_j.$$

Furthermore, if for each $j = 1, \dots, J$,

$$\int_{-a}^a \psi_{\sigma,p}(x_j - \theta) dF(\theta) = \int_{-a}^a \psi_{\sigma,p}(x_j - \theta) dF'(\theta), \quad (\text{A.13})$$

then the term in (A.12) vanishes and $\sup_{x \in I_j} |(F * \psi_{\sigma,p})(x) - (F' * \psi_{\sigma,p})(x)| \lesssim (\varepsilon/\sigma)$ for all $j = 1, \dots, J$. Inequality (A.9) follows by combining the bounds on the terms in (A.10) and (A.11). It remains to choose F' so that all the matching conditions in (A.13) are satisfied. Since the functions $\psi_{\sigma,p}(x_j - \cdot)$ are continuous on $[-a, a]$, by Lemma A.1 of Ghosal and van der Vaart [8], page 1260, F' can be chosen to be a discrete probability measure on $[-a, a]$, with at most $N = J + 1$ support points, such that all the above conditions are satisfied. The proof is thus complete. \square

Remark A.1. Combining inequality (A.5) with Lemma A.5, for any pair $\sigma, \sigma' > 0$,

$$\|F * \psi_{\sigma,p} - F' * \psi_{\sigma',p}\|_\infty \lesssim \frac{|\sigma - \sigma'|}{(\sigma \wedge \sigma')^2} + \frac{\varepsilon}{\sigma \wedge \sigma'},$$

where F' is a discrete probability measure on $[-a, a]$, with at most

$$N \lesssim \begin{cases} ([a/(\sigma \wedge \sigma')] \vee [p \log(1/\varepsilon)]^{1/p})^p, & \text{if } p = 2m, \\ ([a/(\sigma \wedge \sigma')] \vee [p \log(1/\varepsilon)]^{1/p}) \varepsilon^{-(1 \vee 1/p)}, & \text{if } p \neq 2m, \end{cases} \quad m \in \mathbb{N},$$

support points.

The next lemma provides an upper bound on the tail probability of an EP distribution with $p \geq 1$.

Lemma A.6. *Let $p \geq 1$ be fixed and let $\sigma > 0$ be given. For any $B > 0$,*

$$\int_B^\infty \psi_{\sigma,p}(x) \, dx \leq \frac{\sigma^p}{B^{p-1}} \psi_{\sigma,p}(B). \tag{A.14}$$

Proof. Write

$$\begin{aligned} \psi_{\sigma,p}(B) &= - \int_B^\infty \psi_{\sigma,p}^{(1)}(x) \, dx = \int_B^\infty \frac{x^{p-1}}{\sigma^p} \psi_{\sigma,p}(x) \, dx \\ &\geq \frac{B^{p-1}}{\sigma^p} \int_B^\infty \psi_{\sigma,p}(x) \, dx. \end{aligned}$$

Inequality (A.14) follows immediately. □

Remark A.2. For $\sigma = 1$ and $p = 2$, inequality (A.14) reduces to

$$\int_B^\infty \phi(x) \, dx \leq \frac{\phi(B)}{B},$$

which, rewritten in terms of the Mill's ratio, takes the form

$$\frac{\phi(B)}{1 - \Phi(B)} \geq B,$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Next, a similar bound valid for any $p > 0$ is derived which, in particular, holds for $p \in (0, 1)$.

Lemma A.7. *Let $p > 0$ be fixed and let $\sigma > 0$ be given. For any $B > \sigma$,*

$$\int_B^\infty \psi_{\sigma,p}(x) \, dx \leq \frac{B}{(B/\sigma)^p - 1} \psi_{\sigma,p}(B). \tag{A.15}$$

Proof. For each $B > 0$,

$$\begin{aligned} \int_B^\infty \psi_{\sigma,p}(x) \, dx &= -B\psi_{\sigma,p}(B) + \int_B^\infty \left(\frac{x}{\sigma}\right)^p \psi_{\sigma,p}(x) \, dx \\ &\geq -B\psi_{\sigma,p}(B) + \left(\frac{B}{\sigma}\right)^p \int_B^\infty \psi_{\sigma,p}(x) \, dx, \end{aligned}$$

where the first equality follows from integration by parts. If $(B/\sigma) > 1$, then inequality (A.15) holds. \square

The following lemma provides an upper bound on the L_1 -distance between two EP mixtures by relating it to the corresponding L_∞ -distance. It is based on Lemma A.7 and extends an analogous result valid for normal mixtures (cf. Lemma 3.2 of Ghosal and van der Vaart [8], pages 1242–1243).

Lemma A.8. *Let $p > 0$ be fixed. Given $a > 0$, let F and F' be probability measures on $[-a, a]$. For given $\sigma > 0$, define*

$$d_\infty := \|F * \psi_{\sigma,p} - F' * \psi_{\sigma,p}\|_\infty.$$

If $d_\infty \leq (e^{-2/p}/\sigma)$, then

$$\|F * \psi_{\sigma,p} - F' * \psi_{\sigma,p}\|_1 \lesssim d_\infty \left[a \vee \sigma \left(p \log \frac{1}{\sigma d_\infty} \right)^{1/p} \right].$$

The next lemma gives an upper bound on the L_1 -metric entropy of a sieve set of EP mixtures. It is based on Lemma A.5 and Lemma A.8 and can be proved similarly to Lemma 3 of Ghosal and van der Vaart [10], pages 705–707, which deals with normal mixtures.

Lemma A.9. *Let $p = 2m$, $m \in \mathbb{N}$, be fixed. Let $0 < \varepsilon < 1/5$. Let $0 < s < t$ and $a > 0$ be such that, for some $\nu > 0$,*

$$\frac{a}{s} \lesssim \left(\log \frac{1}{\varepsilon} \right)^\nu.$$

*Define $\mathcal{F}_{a,s,t} := \{F * \psi_{\sigma,p} : F([-a, a]) = 1, s \leq \sigma \leq t\}$. Then,*

$$\begin{aligned} \log N(\varepsilon, \mathcal{F}_{a,s,t}, \|\cdot\|_1) &\lesssim \log \left(\frac{t}{s\varepsilon} \right) \\ &\quad + \left[\left(\frac{a}{s} \right)^p \vee \left(\log \frac{1}{\varepsilon} \right) \right] \times \left[\log \left(\frac{2a}{s\varepsilon} + 1 \right) + \log \frac{1}{\varepsilon} \right]. \end{aligned}$$

The following assertion is useful to estimate the prior probability of Kullback-Leibler type neighborhoods of f_0 when checking condition (A.3). Ghosal and van der Vaart [8] present this result for $p = 2$ (cf. Lemma 4.1, pages 1248–1249).

Lemma A.10. *Let $p > 0$ be fixed. Let F' be a probability measure on \mathbb{R} such that, for some constant $b' > 0$,*

$$F'(\{\theta : |\theta| > t\}) \lesssim e^{-b't^p}, \quad \text{for large } t > 0.$$

*Let F be another probability measure on \mathbb{R} such that, for some constant $B > 0$, $F([-B, B]) \geq 1/2$. For $\sigma, \sigma' > 0$, let $q := F * \psi_{\sigma, p}$ and $q' := F' * \psi_{\sigma', p}$. For $0 < \varepsilon \leq [(\sigma'/2) \wedge (1 - e^{-1})/\sqrt{2}]$, if $|\sigma - \sigma'| \leq \varepsilon$ and $d_H(q, q') \leq \varepsilon$, then*

$$\begin{aligned} \int q' \left(\log \frac{q'}{q} \right) d\lambda &\lesssim \varepsilon^2 \log \frac{1}{\varepsilon}, \\ \int q' \left(\log \frac{q'}{q} \right)^2 d\lambda &\lesssim \varepsilon^2 \left(\log \frac{1}{\varepsilon} \right)^2. \end{aligned}$$

Remark A.3. The assertion of Lemma A.10 still holds true if

$$B \equiv B(\varepsilon) = O((\log \varepsilon^{-1})^{1/p}).$$

In such a case, in fact, for a suitable $\delta \in (0, 1]$ (depending on b', σ', p) and some $\xi > 0$,

$$M_\delta^2 := \int_{\{(q'/q) \geq e^{1/\delta}\}} q' \left(\frac{q'}{q} \right)^\delta d\lambda \tag{A.16}$$

is such that $M_\delta^2 = O(1/\varepsilon^\xi)$ and Theorem 5 of Wong and Shen [31], pages 357–358, goes through.

The next lemma is a version of Lemma 11 of Ghosal and van der Vaart [10], pages 715–717, adapted to EP mixtures.

Lemma A.11. *Let $p > 0$ be fixed. Let X_1, \dots, X_n be i.i.d. observations from a probability measure P_0 with density $f_0 = F_0 * \psi_{\sigma_0, p}$. Suppose that the model is a location mixture of EP densities, i.e., $f_{F, \sigma} = F * \psi_{\sigma, p}$, with the scale parameter σ distributed independently of the mixing distribution F . If the base measure α of the Dirichlet process prior for F has a continuous and positive density α' on $[-a, a]$, with $a \geq 1$, then, for any $0 < T \leq (a/2)$ and $\eta > 0$, there exists a constant $K > 0$ (depending only p) such that*

$$\begin{aligned} &\mathbb{E}_0^n [\Pr(\{F : F([-a, a]^c) > \eta^2\} | X_1, \dots, X_n) I_{\{\max_{1 \leq i \leq n} |X_i| \leq a/2\}}] \\ &\lesssim \frac{\alpha([-a, a]^c)}{\eta^2 [\alpha(\mathbb{R}) + n]} + \frac{K n e^{-(a/T)^p / (p2^{p+1})}}{\eta^2 \lambda_a} \\ &\quad + \mathbb{E}_0^n [\Pr(\sigma > T | X_1, \dots, X_n)], \end{aligned} \tag{A.17}$$

where $\lambda_a := \inf_{|\theta| \leq a} \alpha'(\theta) > 0$.

Acknowledgements

The author would like to thank two anonymous referees, an Associate Editor and the Editor for their comments and suggestions. She is also grateful to Judith Rousseau and Willem Kruijer for stimulating discussions and kind hospitality while visiting the ENSAE (École Nationale de la Statistique et de l'Administration Économique) in Paris, where part of this work was written.

References

- [1] BARRON, A., SCHERVISH, M. J. AND WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536–561. [MR1714718](#)
- [2] BOX, G. E. P. AND TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Company, Reading, Massachusetts. [MR0418321](#)
- [3] CHOY, S. T. B. AND SMITH, A. F. M. (1997). On robust analysis of a normal location parameter. *J. Roy. Statist. Soc. Ser. B* **59** 463–474. [MR1440591](#)
- [4] DE JONGE, R. AND VAN ZANTEN, J. H. (2010). Adaptive nonparametric Bayesian inference using location-scale mixture priors. *Ann. Statist.* **38** 3300–3320. [MR2766853](#)
- [5] FERGUSON, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, Eds.) 287–302. Academic Press, New York. [MR0736538](#)
- [6] GHOSAL, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29** 1264–1280. [MR1873330](#)
- [7] GHOSAL, S., GHOSH, J. K. AND RAMAMOORTHY, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158. [MR1701105](#)
- [8] GHOSAL, S. AND VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. [MR1873329](#)
- [9] GHOSAL, S. AND VAN DER VAART, A. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35** 192–223. [MR2332274](#)
- [10] GHOSAL, S. AND VAN DER VAART, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* **35** 697–723. [MR2336864](#)
- [11] HAAS, M., MITTNIK, S. AND PAOLELLA, M. S. (2006). Modelling and predicting market risk with Laplace-Gaussian mixture distributions. *Appl. Finan. Econ.* **16** 1145–1162.
- [12] JOHNSON, N. L., KOTZ, S. AND BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*. Vol. **2**, 2nd ed., John Wiley & Sons, Inc., New York. [MR1326603](#)

- [13] JONES, P. N. AND MCLACHLAN, G. J. (1990). Laplace-normal mixtures fitted to wind shear data. *J. Appl. Statist.* **17** 271–276.
- [14] KANJI, G. K. (1985). A mixture model for wind shear data. *J. Appl. Statist.* **12** 49–58.
- [15] KRUIJER, W. AND VAN DER VAART, A. (2008). Posterior convergence rates for Dirichlet mixtures of beta densities. *J. Statist. Plann. Inference* **138** 1981–1992. [MR2406419](#)
- [16] LIJOI, A., PRÜNSTER, I. AND WALKER, S. G. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *J. Amer. Statist. Assoc.* **100** 1292–1296. [MR2236442](#)
- [17] LO, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12** 351–357. [MR0733519](#)
- [18] LUNETTA, G. (1963). Di una generalizzazione dello schema della curva normale. *Annali della Facoltà di Economia e Commercio di Palermo* **17** 237–244.
- [19] REED, W. J. (2006). The normal-Laplace distribution and its relatives. In *Advances in Distribution Theory, Order Statistics, and Inference* (N. Balakrishnan, E. Castillo and J. M. Sarabia, Eds.) 61–74. Birkhäuser Boston. [MR2226215](#)
- [20] SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26. [MR0184378](#)
- [21] SCRICCILO, C. (2001). Convergence rates of posterior distributions for Dirichlet mixtures of normal densities. *Working Paper 2001.21, Dipartimento di Scienze Statistiche, Università degli Studi di Padova*.
- [22] SCRICCILO, C. (2007). On rates of convergence for Bayesian density estimation. *Scand. J. Statist.* **34** 626–642. [MR2368802](#)
- [23] SHEN, X. AND WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714. [MR1865337](#)
- [24] SUBBOTIN, M. T. (1923). On the law of frequency of error. *Mat. Sb.* **31** 296–301.
- [25] TOKDAR, S. T. (2006). Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhyā* **68** 90–110. [MR2301566](#)
- [26] VIANELLI, S. (1963). La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza. *Statistica* **23** 447–474.
- [27] WALKER, S. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32** 2028–2043. [MR2102501](#)
- [28] WALKER, S. G. AND GUTIÉRREZ-PEÑA, E. (1999). Robustifying Bayesian procedures (with discussion). In *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds.) 685–710. Oxford Univ. Press, New York. [MR1724876](#)
- [29] WALKER, S. G., LIJOI, A. AND PRÜNSTER, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Statist.* **35** 738–746. [MR2336866](#)
- [30] WEST, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74** 646–648. [MR0909372](#)

- [31] WONG, W. H. AND SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. [MR1332570](#)
- [32] WU, Y. AND GHOSAL, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electron. J. Statist.* **2** 298–331. [MR2399197](#)
- [33] WU, Y. AND GHOSAL, S. (2009). Correction to: “Kullback Leibler property of kernel mixture priors in Bayesian density estimation”. *Electron. J. Statist.* **3** 316–317. [MR2497160](#)