Rates for Bayesian Estimation of Location-Scale Mixtures of Super-Smooth Densities

Catia Scricciolo

Abstract

We consider Bayesian nonparametric density estimation with a Dirichlet process kernel mixture as a prior on the class of Lebesgue univariate densities, the emphasis being on the achievability of the error rate $n^{-1/2}$, up to a logarithmic factor, depending upon the kernel. We derive rates of convergence for the Bayes' estimator of *super-smooth* densities that are location-scale mixtures of densities whose Fourier transforms have sub-exponential tails. We show that a nearly parametric rate is attainable in the L^1-norm, under weak assumptions on the tail decay of the true mixing distribution and the overall Dirichlet process base measure.

1 Introduction

Consider the estimation of a density f_0 on $\mathbb{R}$ from observations $X_1, \ldots, X_n$ taking a Bayesian nonparametric approach. A prior is defined on a metric space of probability measures with Lebesgue density and a summary of the posterior, e.g., the posterior expected density, is employed. The so-called what if approach, which consists in investigating frequentist asymptotic properties of the posterior, under the non-Bayesian assumption that the data are generated from a *fixed* density, provides a way to validate priors on infinite-dimensional spaces. Desirable asymptotic properties of posterior distributions are consistency, minimax-optimal concentration rate of the posterior mass around the "truth" as the sample size grows, possibly with full adaptation to the regularity level of f_0, if unknown, and distributional convergence.

C. Scricciolo (✉)
Bocconi University, Via Roentgen n. 1, 20136 Milan, Italy
e-mail: catia.scricciolo@unibocconi.it

G. Alleva, A. Giommi (eds.), *Topics in Theoretical and Applied Statistics*,
Studies in Theoretical and Applied Statistics,
DOI 10.1007/978-3-319-27274-0_5

For bounded and convex distances, posterior contraction rates yield upper bounds on convergence rates of the Bayes' estimator, thus motivating the interest in their study. Since the seminal articles of Ferguson [2] and Lo [4], the idea of constructing priors on spaces of densities by convoluting a fixed kernel with a random distribution has been successfully exploited in density estimation. Even if much progress has been done during the last decade in understanding frequentist asymptotic properties of mixture models, the choice of the kernel is a topic largely ignored in the literature, except for the article of Wu and Ghosal [9], mainly focussed on consistency. Posterior contraction rates for Dirichlet process kernel mixture priors have been investigated by Ghosal and van der Vaart [3] and Scricciolo [5]. One key message is that some constraints on the regularity of the kernel and on the tail decay of the true mixing distribution are necessary to accurately estimate a density. Most of the literature has dealt with the estimation of mixtures, with normal (or generalized normal) kernel and mixing distribution having either compact support or sub-exponential tails, finding a *nearly parametric* rate, up to a logarithmic factor, in the L^1-distance, but there are almost no results beyond the Gaussian kernel. The aim of this work is to contribute to the understanding of the role of the kernel choice in density estimation with a Dirichlet process mixture prior. The main result states that a nearly parametric rate can be attained to estimate mixtures of super-smooth densities having Fourier transforms that decay exponentially, whatever the kernel tail decay, heavy tailed distributions, like Student's-t or Cauchy, being included, which have been proved to be extremely useful in accurately modeling different kinds of financial data. For example, individual stock indices can be modeled as stable laws. Multivariate stable laws have been fruitfully used in computer networks, see Bickson and Guestrin [1]. The assumption on the exponential tail decay of the true mixing distribution seems unavoidable in order to find a finite approximating mixture with a sufficiently restricted number of points. This step is a delicate mathematical point in the proof, see Lemma 1. Such an approximation result, which is reported in the Appendix, may be of autonomous interest as well. In Sect. 2, we fix the notation and present the result.

2 Main Result

We derive rates for location-scale mixtures of super-smooth densities. The model is $f_{F,G}(x) := \int_0^\infty (F * K_\sigma)(x)\,\mathrm{d}G(\sigma)$, $x \in \mathbb{R}$, where K is a kernel density, $F \sim D_\alpha$ is a Dirichlet process with base measure $\alpha := \alpha(\mathbb{R})\bar{\alpha}$, for $0 < \alpha(\mathbb{R}) < \infty$ and $\bar{\alpha}$ a probability measure on $\mathbb{R}$, and $G \sim D_\beta$, with finite and positive base measure β on $(0, \infty)$. We assume that $f_0 = f_{F_0, G_0}$, with F_0 and G_0 denoting the true mixing distributions for the location and scale parameters, respectively. We use the following assumptions.

(A) The true mixing distribution G_0 for the scale parameter satisfies

$$\int_0^\infty \sigma \, \mathrm{d}G_0(\sigma) < \infty \qquad \text{and} \qquad \int_0^\infty \frac{1}{\sigma} \, \mathrm{d}G_0(\sigma) < \infty. \tag{1}$$

Also, for constants $d_1, d_2 > 0$ and $0 < \gamma_1^0, \gamma_2^0 \le \infty$,

$$G_0(s) \lesssim e^{-d_1 s^{-\gamma_1^0}} \quad \text{as } s \to 0 \quad \text{and} \quad 1 - G_0(s) \lesssim e^{-d_2 s^{\gamma_2^0}} \quad \text{as } s \to \infty.$$

(B) The base measure β of the Dirichlet process prior for G has a continuous and positive Lebesgue density β' on $(0, \infty)$ such that, for constants $C_j, D_j > 0$, $j = 1, \ldots, 4$, $q_1, q_2, r_1, r_2 \ge 0$ and $0 < \gamma_1, \gamma_2 \le \infty$,

$$C_1 \sigma^{-q_1} e^{-C_2 \sigma^{-\gamma_1} (\log(1/\sigma))^{r_1}} \le \beta'(\sigma) \le C_3 \sigma^{-q_1} e^{-C_4 \sigma^{-\gamma_1} (\log(1/\sigma))^{r_1}} \tag{2}$$

for all σ in a neighborhood of 0, and

$$D_1 \sigma^{q_2} e^{-D_2 \sigma^{\gamma_2} (\log \sigma)^{r_2}} \le \beta'(\sigma) \le D_3 \sigma^{q_2} e^{-D_4 \sigma^{\gamma_2} (\log \sigma)^{r_2}} \tag{3}$$

for all σ large enough.

Remark 1 The right-hand side requirement in (1) has also been postulated by Tokdar [7], see condition 3 of Lemma 5.1 and condition 4 of Theorem 5.2, pp. 102–103. If, for example, G_0 is an IG(ν, λ), with shape parameter $\nu > 0$ and scale parameter $\lambda > 0$, then $\int_0^\infty \sigma^{-1} \, \mathrm{d}G_0(\sigma) = (\nu/\lambda) < \infty$. If G_0 is a right-truncated distribution, then the requirement on the upper tail is satisfied with $\gamma_2^0 = \infty$. A right-truncated Inverse-Gamma distribution meets all the requirements of assumption (A).

Remark 2 Condition (2) is satisfied (with $r_1 = 0$) if β' is an Inverse-Gamma distribution. It can be seen that (2) implies that

$$\beta((0, s]) \le \exp\left\{ -\frac{C_4}{2} s^{-\gamma_1} \left(\log \frac{1}{s} \right)^{r_1} \right\} \lesssim e^{-\frac{1}{2} C_4 s^{-\gamma_1}} \qquad \text{as } s \to 0.$$

Condition (3) has been considered by van der Vaart and van Zanten [8], p. 2660, and implies that $\beta((s, \infty)) \lesssim \exp\{-D_4 s^{\gamma_2}/2\}$ as $s \to \infty$, see Lemma 4.9, p. 2669.

We assess rates for location-scale mixtures of symmetric stable laws. The result goes through to location-scale mixtures of Student's-t distributions.

Theorem 1 *Let K be the density of a symmetric stable law of index $0 < r \le 2$. Suppose that $f_0 = \int_0^\infty (F_0 * K_\sigma) \, \mathrm{d}G_0(\sigma)$, with the true mixing distribution F_0 for the* *location parameter satisfying the tail condition*

$$F_0(\{\theta : |\theta| > t\}) \lesssim \exp\{-c_0 t^{1+I_{(1,2]}(r)/(r-1)}\} \qquad \text{for large } t > 0, \tag{4}$$

for some constant $c_0 > 0$*, and the true mixing distribution* G_0 *for the scale parameter satisfying assumption* (A)*, with* $\gamma_2^0 = \infty$*. If the base measure* α *has a density* α' *such that, for constants* $b > 0$ *and* $0 < \delta \le 1 + I_{(1,2]}(r)/(r-1)$*, satisfies*

$$\alpha'(\theta) \propto e^{-b|\theta|^\delta}, \qquad \theta \in \mathbb{R}, \tag{5}$$

the base measure β *satisfies assumption* (B)*, with* $0 < \gamma_j \le \gamma_j^0 \le \infty$ *and* $\gamma_j < \gamma_j^0$ *if* $r_j > 0$*,* $j = 1, 2$*, then the posterior rate of convergence relative to the Hellinger distance is* $\varepsilon_n = n^{-1/2}(\log n)^\kappa$*, with* $\kappa > 0$ *depending on* γ_1^0*,* γ_1*,* γ_2*, and* r*.*

Proof The proof is in the same spirit as that of Theorem 4.1 in Scricciolo [6], which, for space limitations, cannot be reported here. Let $\bar{\varepsilon}_n = n^{-1/2}(\log n)^\kappa$ and $\tilde{\varepsilon}_n = n^{-1/2}(\log n)^\tau$, with $\kappa > \tau > 0$ whose rather lengthy expressions we refrain from writing down. Let $0 < s_n \le E(\log(1/\bar{\varepsilon}_n))^{-2\tau/\gamma_1}$, $0 < S_n \le F(\log(1/\bar{\varepsilon}_n))^{2\tau/\gamma_2}$, and $0 < a_n \le L(\log(1/\bar{\varepsilon}_n))^{2\tau/\delta}$, with $E, F, L > 0$ suitable constants. Replacing the expression of N in (A.19) of Lemma A.7 of Scricciolo [6], with that in Lemma 1, we can estimate the covering number of the sieve set

$$\mathscr{F}_n := \{f_{F,G} : F([-a_n, a_n]) \ge 1 - \bar{\varepsilon}_n/2, \quad G([s_n, S_n]) \ge 1 - \bar{\varepsilon}_n/2\}$$

and show that $\log D(\bar{\varepsilon}_n, \mathscr{F}_n, d_{\mathrm{H}}) \lesssim (\log n)^{2\kappa} = n\bar{\varepsilon}_n^2$. Verification of the remaining mass condition $\pi(\mathscr{F}_n^c) \lesssim \exp\{-(c_2+4)n\tilde{\varepsilon}_n^2\}$ can proceed as in the aforementioned theorem using, among others, the fact that $2\tau > 1$.

We now turn to consider the small ball probability condition. For $0 < \varepsilon < 1/4$, let $a_\varepsilon := (c_0^{-1}\log(1/(s_\varepsilon\varepsilon)))^{1/(1+I_{(1,2]}(r)/(r-1))}$ and $s_\varepsilon := (d_1^{-1}\log(1/\varepsilon))^{-1/\gamma_1^0}$. Let G_0^* be the re-normalized restriction of G_0 to $[s_\varepsilon, S_0]$, with S_0 the upper endpoint of the support of G_0, and F_0^* the re-normalized restriction of F_0 to $[-a_\varepsilon, a_\varepsilon]$. Then, $\|f_{F_0^*, G_0^*} - f_0\|_1 \lesssim \varepsilon$. By Lemma 1, there exist discrete distributions $F_0' := \sum_{j=1}^N p_j\delta_{\theta_j}$ on $[-a_\varepsilon, a_\varepsilon]$ and $G_0' := \sum_{k=1}^N q_k\delta_{\sigma_k}$ on $[s_\varepsilon, S_0]$, with at most $N \lesssim (\log(1/\varepsilon))^{2\tau-1}$ support points, such that $\|f_{F_0', G_0'} - f_{F_0^*, G_0^*}\|_\infty \lesssim \varepsilon$. For $T_\varepsilon := (2a_\varepsilon \vee \varepsilon^{-1/(r+I_{(0,1]}(r))})$,

$$\|f_{F_0', G_0'} - f_{F_0^*, G_0^*}\|_1 \lesssim T_\varepsilon \|f_{F_0', G_0'} - f_{F_0^*, G_0^*}\|_\infty + T_\varepsilon^{-r} \lesssim \varepsilon^{1-1/(r+I_{(0,1]}(r))}.$$

Without loss of generality, the θ_j's and σ_k's can be taken to be at least 2ε-separated. For any distribution F on $\mathbb{R}$ and G on $(0, \infty)$ such that

$$\sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon]) - p_j| \le \varepsilon \quad \text{and} \quad \sum_{k=1}^N |G([\sigma_k - \varepsilon, \sigma_k + \varepsilon]) - q_k| \le \varepsilon,$$

by the same arguments as in the proof of Theorem 4.1 in Scricciolo [6],

$$\|f_{F,G} - f_{F_0', G_0'}\|_1 \lesssim \varepsilon.$$

Consequently,

$$\begin{aligned} d_{\mathrm{H}}^2(f_{F,G}, f_0) &\le \|f_{F,G} - f_{F_0', G_0'}\|_1 + \|f_{F_0', G_0'} - f_{F_0^*, G_0^*}\|_1 + \|f_{F_0^*, G_0^*} - f_0\|_1 \\ &\lesssim \varepsilon^{1-1/(r+I_{(0,1]}(r))}. \end{aligned}$$

By an analogue of the last part of the same proof, we get that $\pi(B_{\mathrm{KL}}(f_0;\, \tilde{\varepsilon}_n^2)) \gtrsim \exp\{-c_2 n \tilde{\varepsilon}_n^2\}$.

Remark 3 Assumptions (4) on F_0 and (5) on α' imply that $\mathrm{supp}(F_0) \subseteq \mathrm{supp}(\alpha)$, thus, F_0 is in the *weak* support of D_α. Analogously, assumptions (A) on G_0 and (B) on β', together with the restrictions on γ_j^0, γ_j, $j = 1, 2$, imply that $\mathrm{supp}(G_0) \subseteq \mathrm{supp}(\beta)$, thus, G_0 is in the *weak* support of D_β.

Remark 4 If $\gamma_1 = \gamma_2 = \infty$, then also $\gamma_1^0 = \gamma_2^0 = \infty$, i.e., the true mixing distribution G_0 for σ is compactly supported on an interval $[s_0, S_0]$, for some $0 < s_0 \le S_0 < \infty$, and (an upper bound on) the rate is given by $\varepsilon_n = n^{-1/2}(\log n)^\kappa$, with κ whose value for Gaussian mixtures ($r = 2$) reduces to the same found by Ghosal and van der Vaart [3] in Theorem 6.1, p. 1255.

Appendix

The following lemma provides an upper bound on the number of mixing components of finite location-scale mixtures of symmetric stable laws that uniformly approximate densities of the same type with compactly supported mixing distributions. We use $\mathbb{E}$ and $\mathbb{E}'$ to denote expectations corresponding to priors G and G' for the scale parameter Σ, respectively.

Lemma 1 *Let K be a density with Fourier transform such that, for constants A, $\rho > 0$ and $0 < r < 2$, $\Phi_K(t) = Ae^{-\rho|t|^r}$, $t \in \mathbb{R}$. Let $0 < \epsilon < 1$, $0 < a < \infty$ and $0 < s \le S < \infty$ be given, with $(a/s) \ge 1$. For any pair of probability measures F on $[-a, a]$ and G on $[s, S]$, there exist discrete probability measures F' on $[-a, a]$ and G' on $[s, S]$, with at most*

$$N \lesssim \begin{cases} \frac{a}{s} \times \left(\frac{S}{s}\right)^r \left(\log \frac{1}{s\epsilon}\right)^{1+1/r}, & \text{if} \quad 0 < r \le 1, \\ \max\left\{\left(\frac{a}{s}\right)^{r/(r-1)}, \left(\frac{S}{s}\right)^{r/(r-1)} \left(\log \frac{1}{s^r\epsilon}\right)^{1/(r-1)}\right\}, & \text{if} \quad 1 < r < 2, \end{cases}$$

support points, such that $\|\mathbb{E}[F * K_\Sigma] - \mathbb{E}'[F' * K_\Sigma]\|_\infty \lesssim \epsilon$.

Proof We consider first the case where $1 < r < 2$ because, since $(a/s) \geq 1$ by assumption, we can appeal to Lemma A.1 of Scricciolo [6]. The arguments of the first part of the proof can be then used to deal also with the case where $0 < r \leq 1$. For each $s \leq \sigma \leq S$, since $\int_{-\infty}^{\infty} |\Phi_K(\sigma t)|\,\mathrm{d}t < \infty$, the inversion formula can be applied to recover both $F * K_\sigma$ and $F' * K_\sigma$. For any $M > 0$ and $x \in \mathbb{R}$,

$$
\begin{aligned}
&|\mathbb{E}[(F * K_\Sigma)(x)] - \mathbb{E}'[(F' * K_\Sigma)(x)]| \\
&= \frac{1}{2\pi}\bigg| \int_s^S \int_{-\infty}^{\infty} e^{-itx} \Phi_K(\sigma t)\Phi_F(t)\,\mathrm{d}t\,\mathrm{d}G(\sigma) \\
&\quad - \int_s^S \int_{-\infty}^{\infty} e^{-itx} \Phi_K(\sigma t)\Phi_{F'}(t)\,\mathrm{d}t\,\mathrm{d}G'(\sigma) \bigg| \\
&= \frac{1}{2\pi}\left| \left(\int_{|t|\leq M} + \int_{|t|>M}\right) e^{-itx}\left[\Phi_F(t)\mathbb{E}[\Phi_K(\Sigma t)] - \Phi_{F'}(t)\mathbb{E}'[\Phi_K(\Sigma t)]\right]\mathrm{d}t\right|.
\end{aligned}
$$

Let

$$
U := \frac{1}{2\pi}\left| \int_{|t|\leq M} e^{-itx}\left[\Phi_F(t)\mathbb{E}[\Phi_K(\Sigma t)] - \Phi_{F'}(t)\mathbb{E}'[\Phi_K(\Sigma t)]\right]\mathrm{d}t\right|
$$

and

$$
V := \frac{1}{2\pi}\left| \int_{|t|> M} e^{-itx}\left[\Phi_F(t)\mathbb{E}[\Phi_K(\Sigma t)] - \Phi_{F'}(t)\mathbb{E}'[\Phi_K(\Sigma t)]\right]\mathrm{d}t\right|.
$$

For $M \geq (\rho^{1/r}s)^{-1}(\log(1/(s^r\varepsilon)))^{1/r}$,

$$
V \leq \frac{A}{2\pi}\int_{|t|>M}\int_s^S e^{-\rho(\sigma|t|)^r}\,\mathrm{d}(G+G')(\sigma)\,\mathrm{d}t \lesssim \varepsilon.
$$

In order to find an upper bound on U, we apply Lemma A.1 of Ghosal and van der Vaart [3], p. 1260, to both F and G. There exists a discrete probability measure F' on $[-a, a]$, with at most $N_1 + 1$ support points, where N_1 is a positive integer to be suitably chosen later on, such that it matches the (finite) moments of F up to the order N_1, i.e., $\mathbb{E}'[\Theta^j] = \mathbb{E}[\Theta^j]$ for all $j = 1, \ldots, N_1$. Analogously, there exists a discrete probability measure G' on $[s, S]$, with at most N_2 support points, where N_2 is a positive integer to be suitably chosen later on, such that

$$
\mathbb{E}'[\Sigma^{r\ell}] := \int_s^S \sigma^{r\ell}\,\mathrm{d}G'(\sigma) = \int_s^S \sigma^{r\ell}\,\mathrm{d}G(\sigma) =: \mathbb{E}[\Sigma^{r\ell}], \qquad \ell = 1, \ldots, N_2 - 1.
$$

Both N_1 and N_2 will be chosen to be increasing functions of ε. In virtue of the latter matching conditions,

$$\begin{aligned}\left|\mathbb{E}[\Phi_K(\Sigma t)]-\mathbb{E}'[\Phi_K(\Sigma t)]\right| &\le \left|\mathbb{E}\left[\Phi_K(\Sigma t)-A\sum_{\ell=0}^{N_2-1}\frac{(\rho(\Sigma|t|)^r)^\ell}{\ell!}\right]\right|\\ &\quad+\left|\mathbb{E}'\left[\Phi_K(\Sigma t)-A\sum_{\ell=0}^{N_2-1}\frac{(\rho(\Sigma|t|)^r)^\ell}{\ell!}\right]\right|\\ &\le \frac{2A}{(N_2)!}(\rho(S|t|)^r)^{N_2}, \qquad t\in\mathbb{R}. \end{aligned} \tag{6}$$

Using arguments of Lemma A.1 in Scricciolo [6] and inequality (6),

$$\begin{aligned} U &\le \frac{1}{2\pi}\int_{|t|\le M}\left|\Phi_F(t)\mathbb{E}[\Phi_K(\Sigma t)]-\Phi_{F'}(t)\mathbb{E}'[\Phi_K(\Sigma t)]\right|\,\mathrm{d}t\\ &\le \frac{1}{2\pi}\int_{|t|\le M}\left|\Phi_F(t)-\sum_{j=0}^{N_1}\frac{(it)^j}{j!}\mathbb{E}[\Theta^j]\right|\mathbb{E}[|\Phi_K(\Sigma t)|]\,\mathrm{d}t\\ &\quad+\frac{1}{2\pi}\int_{|t|\le M}\left|\Phi_{F'}(t)-\sum_{j=0}^{N_1}\frac{(it)^j}{j!}\mathbb{E}'[\Theta^j]\right|\mathbb{E}'[|\Phi_K(\Sigma t)|]\,\mathrm{d}t\\ &\quad+\frac{1}{2\pi}\sum_{j=0}^{N_1}\frac{|\mathbb{E}[\Theta^j]|}{j!}\int_{|t|\le M}|t^j|\left|\mathbb{E}[\Phi_K(\Sigma t)]-\mathbb{E}'[\Phi_K(\Sigma t)]\right|\,\mathrm{d}t\\ &\le \frac{4Aa^{N_1}}{\pi r(\rho s^r)^{(N_1+1)/r}}\frac{\Gamma((N_1+1)/r)}{\Gamma(N_1+1)}+\frac{2A}{\pi(N_2)!}(1+aM)^{N_1}(\rho(SM)^r)^{N_2}M\\ &\sim \frac{4Aa^{N_1}}{\pi r(\rho s^r)^{(N_1+1)/r}}\frac{\Gamma((N_1+1)/r)}{\Gamma(N_1+1)}+\frac{\sqrt{2}A(1+aM)^{N_1}(\rho(SM)^r)^{N_2}M}{\pi^{3/2}e^{-N_2}N_2^{N_2-1/2}}, \end{aligned}$$

where, in the last line, we have used Stirling's approximation for $(N_2)!$, assuming N_2 is large enough. For $N_1 \lesssim \max\{\log(1/(s\varepsilon)),\ (a/s)^{r/(r-1)}\}$,

$$U_1 := \frac{4Aa^{N_1}}{\pi r(\rho s^r)^{(N_1+1)/r}}\frac{\Gamma((N_1+1)/r)}{\Gamma(N_1+1)} \lesssim \varepsilon.$$

Let M be such that $aM \ge 1$ and $(\rho^{1/r}SM) \ge 2a$. Then, for $N_2 \ge \max\{(2N_1+1)(r-1)/(r(2-r)),\ e^3(\rho^{1/r}SM)^{r/(r-1)},\ \log(1/\varepsilon)\}$,

$$(1+aM)^{N_1}(\rho(SM)^r)^{N_2}M \le (\rho^{1/r}SM)^{rN_2+2N_1+1} \le (\rho^{1/r}SM)^{rN_2/(r-1)}$$

and

$$U_2 := \frac{\sqrt{2}A(1 + aM)^{N_1}(\rho(SM)^r)^{N_2}M}{\pi^{3/2}e^{-N_2}N_2^{N_2-1/2}} \lesssim \varepsilon.$$

Hence, $N_2 \lesssim \max\{(a/s)^{r/(r-1)},\ ((S/s))^{r/(r-1)}(\log(1/s^r\varepsilon))^{1/(r-1)}\}$.

In the case where $0 < r \le 1$, since $(a/s) \ge 1$, we need to restrict the support of the mixing distribution F. To the aim, we consider a partition of $[-a, a]$ into $k = \lceil (a/s)(\log(1/(s\varepsilon)))^{1/r-1} \rceil$ subintervals $I_1, \ldots, I_k$ of equal length $0 < l \le 2s(\log(1/(s\varepsilon)))^{-(1-r)/r}$ and, possibly, a final interval I_{k+1} of length $0 \le l_{k+1} < l$. Let J be the number of intervals in the partition, which can be either k or $k+1$. Write $F = \sum_{j=1}^J F(I_j)F_j$, where F_j denotes the re-normalized restriction of F to I_j. Then, for each $s \le \sigma \le S$, we have $(F * K_\sigma)(x) = \sum_{j=1}^J F(I_j)(F_j * K_\sigma)(x)$, $x \in \mathbb{R}$. For any probability measure F' such that $F'(I_j) = F(I_j)$, $j = 1, \ldots, J$,

$$\begin{aligned} &|\mathbb{E}[(F * K_\Sigma)(x)] - \mathbb{E}'[(F' * K_\Sigma)(x)]| \\ &\quad \le \sum_{j=1}^J F(I_j)|\mathbb{E}[(F_j * K_\Sigma)(x)] - \mathbb{E}'[(F_j' * K_\Sigma)(x)]|, \qquad x \in \mathbb{R}. \end{aligned}$$

Reasoning as in the case where $1 < r < 2$, with a to be understood as $l/2$ and N_1 as the number of support points of the generic F_j, for $M \ge ((\rho/2)^{1/r}s)^{-1}(\log(1/\varepsilon))^{1/r}$,

$$|\mathbb{E}[(F_j * K_\Sigma)(x)] - \mathbb{E}'[(F_j' * K_\Sigma)(x)]| \lesssim U + V \lesssim (U_1 + U_2) + \varepsilon, \qquad x \in \mathbb{R}.$$

Since $(a/s) \lesssim (\log(1/(s\varepsilon)))^{-(1-r)/r}$ by construction, for $N_1 = \log(1/(s\varepsilon))$, it turns out that $U_1 \lesssim \varepsilon$. For $N_2 \ge \max\{N_1,\ 2e^4\rho(SM)^r\log(1/(s\varepsilon)),\ \log(1/\varepsilon)\}$,

$$(1 + aM)^{N_1}(\rho(SM)^r)^{N_2}M \le M(2\rho(SM)^r\log(1/(s\varepsilon)))^{N_2}$$

and $U_2 \lesssim \varepsilon$. Then, $N_2 \lesssim (S/s)^r(\log(1/(s\varepsilon)))^2$ and the total number N^{T} of support points of F' is bounded above by

$$J \times N_1 \lesssim J \times N_2 \lesssim \frac{a}{s} \times \left(\frac{S}{s}\right)^r \left(\log\frac{1}{s\varepsilon}\right)^{1+1/r}.$$

The proof is thus complete.

Remark 5 Lemma 1 does not cover the case where $r = 2$, i.e., the kernel is Gaussian: this might possibly be due to the arguments laid out in the proof. This case can be retrieved from Lemma A.2 in Scricciolo [6] when $p = 2$.

References

1. Bickson, D., Guestrin, C.: Inference with multivariate heavy-tails in linear models. In: Lafferty, J.D., Williams, C.K.I., Shawe-Taylor, J., Zemel, R.S., Culotta, A. (eds.) Advances in Neural Information Processing Systems, vol. 23, pp. 208–216. Curran Associates, Inc., Red Hook (2010). http://papers.nips.cc/paper/3949-inference-with-multivariate-heavy-tails-in-linear-models.pdf
2. Ferguson, T.S.: Bayesian density estimation by mixtures of normal distributions. In: Rizvi, M.H., Rustagi, J.S., Siegmund, D. (eds.) Recent Advances in Statistics, pp. 287–302. Academic, New York (1983)
3. Ghosal, S., van der Vaart, A.W.: Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. Ann. Stat. **29**, 1233–1263 (2001)
4. Lo, A.Y.: On a class of Bayesian nonparametric estimates: I. Density estimates. Ann. Stat. **12**, 351–357 (1984)
5. Scricciolo, C.: Posterior rates of convergence for Dirichlet mixtures of exponential power densities. Electron. J. Stat. **5**, 270–308 (2011)
6. Scricciolo, C.: Rates of convergence for Bayesian density estimation with Dirichlet process mixtures of super-smooth kernels. Working Paper No.1. DEC, Bocconi University (2011)
7. Tokdar, S.T.: Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. Sankhyā **68**, 90–110 (2006)
8. van der Vaart, A.W., van Zanten, J.H.: Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. Ann. Stat. **37**, 2655–2675 (2009)
9. Wu, Y., Ghosal, S.: Kullback Leibler property of kernel mixture priors in Bayesian density estimation. Electron. J. Stat. **2**, 298–331 (2008)