Supplementary material for Posterior concentration rates for empirical Bayes procedures with applications to Dirichlet process mixtures

SOPHIE DONNET$^1, *$ VINCENT RIVOIRARD$^1,**$ JUDITH ROUSSEAU$^2$ and CATIA SCRICCILO$^3$

$^1$CEREMADE, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France. E-mail: *donnet@ceremade.dauphine; **rivoirard@ceremade.dauphine.fr

$^2$CEREMADE, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France and CREST-ENSAE, 3 Avenue Pierre Larousse, 92240 Malakoff, France. E-mail: rousseau@ceremade.dauphine.fr

$^3$Department of Decision Sciences, Bocconi University, Via Röntgen 1, 20136 Milano, Italy. E-mail: catia.scricchiolo@unibocconi.it

This supplementary material provides the proof of Theorem 2, Proposition 1 and Theorem 3. Equation numbers refer to the main paper. We denote by $C$ a positive constant that may change from line to line.

**Appendix A: Proof of Theorem 2**

It is enough to check that assumptions [A1] and [A2] of Theorem 1 are satisfied. We begin by defining the parameter transformation. Under a DPM prior law with base measure proportional to a Gaussian distribution with parameter $\gamma = (m, s^2)$, we have $p_{F,\sigma}(\cdot) = \sum_{j \geq 1} p_j \phi_{\sigma}(\cdot - \theta_j)$ almost surely, with independent sequences $(\theta_j)_{j \geq 1}$ and $(p_j)_{j \geq 1}$, the random variables $(\theta_j)_{j \geq 1}$ being independent and identically distributed according to $N(m, s^2)$. Hereafter, we use the notation $N_\gamma$ as shorthand for $N(m, s^2)$. We consider a set $K_n = [m_1, m_2] \times [s^2_1, s^2_n]$, with constants $-\infty < m_1 \leq m_2 < \infty$, $s^2_1 > 0$ and a positive sequence $s^2_n \to \infty$ as a power of $\log n$, such that $P_{\gamma_0}(\hat{\gamma}_n \in \overline{K}_n) = o(1)$. For a positive sequence $u_n \to 0$ to be suitably chosen, consider a $u_n$-covering of $[m_1, m_2]$ with intervals $I_k = [m_k, m_k + u_n)$, for $m_k = m_1 + (k-1)u_n$, $k = 1, \ldots, L_{mn}$, where $L_{mn} = \lceil 1 + (m_2 - m_1)/u_n \rceil$, and a covering of $[s^2_1, s^2_n]$ with intervals $J_l = [s^2_l, s^2_{l+1}] = [s^2_l(1 + u_n)^{l-1}, s^2_l(1 + u_n)^l]$, for $l = 1, \ldots, L_{sn}$, where $L_{sn} = \lceil 2u_n^{-1}\log(s_n/s_1) \rceil$.

For $s^2 \in J_l$, $l = 1, \ldots, L_{sn}$, let $p_l = (s^2_l/s^2_1)^{1/2}$. Let $m \in I_k$, $k = 1, \ldots, L_{mn}$. For any $\gamma' = (m_k, s^2_l)$ and $\gamma = (m, s^2)$, if $\theta_j \sim N_{\gamma'}$, then $\theta_j = [p_l(\theta_j - m_k) + m] \sim N_{\gamma}$, $j \in \mathbb{N}$.
Therefore, conditionally on $\sigma$, for $F \sim \text{DP}(\alpha_{\mathbb{R}} N_{\gamma})$,

$$\psi_{\gamma', \gamma}(p_{F, \sigma})(\cdot) = \sum_{j \geq 1} p_j \phi_\sigma(\cdot - \theta'_j - [(\rho_l - 1)\theta'_j - \rho_l m_k + m])$$

is distributed according to a DPM of Gaussian densities with base measure $\alpha_{\mathbb{R}} N_{\gamma}$. With abuse of notation, we shall also write $\psi_{\gamma', \gamma}(\theta'_j)$ to intend the parameter transformation $\theta'_j \mapsto \rho_l (\theta'_j - m_k) + m$.

In the sequel, we shall repeatedly use the following inequalities:

$$1 \leq \rho_l < (1 + u_n)^{1/2} \quad \text{and} \quad -m_k u_n < m - \rho_l m_k \leq u_n. \quad (A.2)$$

We first deal with the ordinary smooth case. To check that condition [A1] is satisfied, let $\sigma \in (\sigma_n/2, 2\sigma_n)$, with $\sigma_n = \epsilon_n^{1/\beta}$, and let $F^{*} = \sum_{j=1}^{N_\sigma} p'_j \delta_{\theta'_j}$ be a mixing distribution such that the Gaussian mixture $p_{F^{*}, \sigma}$ satisfies both requirements in (3.4) and the minimal distance between any pair of contiguous location points $\theta'_j$'s is bounded below by $\delta = \sigma_n^{2b}$, for some $b > \max\{1, (2\beta)^{-1}\}$. A partition $(U_j)_{\gamma=1}^{M}$ of $\mathbb{R}$ can be constructed following the proof of Theorem 4 in Shen et al. [6] so that $(U_j)_{\gamma=1}^{K_{n}}$ is a partition of $[-a_{\sigma}, a_{\sigma}]$, with $a_{\sigma} = a_{0} |\log \sigma|^{1/\gamma}$, composed of intervals $[\theta'_j - \delta/2, \theta'_j + \delta/2]$, for $j = 1, \ldots, N_{\sigma}$, and of intervals with diameter smaller than or equal to $\sigma$ to complete $[-a_{\sigma}, a_{\sigma}]$. Then, a partition of $(-\infty, -a_{\sigma}] \cup [a_{\sigma}, \infty)$ can be constructed with intervals $U_j$, for $j = K + 1, \ldots, M$, such that $a_1 \sigma n^{2b} \leq \alpha_{\mathbb{R}} N_{\gamma}(U_j) \leq 1$ for some constant $a_1 > 0$. Note that, as in Shen et al. [6], $M \leq \sigma^{-1}(\log n)^{1+1/\tau}$ and, for every $1 \leq j \leq K$, we have $N_{\sigma}(U_j) \geq (\delta/s)e^{-2(a_{\sigma}/s)^2} \geq \sigma_n \epsilon_n^{2b}$ uniformly in $\gamma \in K_n$. As in Shen et al. [6], define $B_n$ as the set of all $(F, \sigma)$ such that $\sigma \in (\sigma_n/2, 2\sigma_n)$ and

$$\sum_{j=1}^{M} |F(U_j) - p'_j| \leq 2\epsilon_n^{2b}, \quad \min_{1 \leq \gamma \leq M} F(U_j) \geq \epsilon_2^{4b}/2.$$ 

Following Lemma 10 of Ghosal and van der Vaart [2], for some constant $c > 0$,

$$\inf_{\gamma \in K_n} \pi(B_n \mid \gamma) \asymp \exp (-c \sigma_n^{-1}(\log n)^{2+1/\gamma}). \quad (A.3)$$

For every $(F, \sigma) \in B_n$, for $\gamma' = (m_k, s^2_{\sigma})$ and any $\gamma \in I_k \times J_l$, by the parameter transformation in (A.1) and the inequalities in (A.2),

$$\psi_{\gamma', \gamma}(p_{F, \sigma})(x) = \sum_{j \geq 1} p_j \phi_\sigma(x - \psi_{\gamma', \gamma}(\theta'_j))$$

$$\geq \sum_{j: \theta'_j \leq a_{\sigma}} p_j \phi_\sigma(x - \psi_{\gamma', \gamma}(\theta'_j))$$

$$> \sum_{j: \theta'_j \leq a_{\sigma}} p_j \phi_\sigma(x - \theta'_j)$$

$$\times \exp (-4|x - \theta'_j|/(\sigma_{\sigma} + 1) u_n + (a_{\sigma}^2 + 1) u_n^2)/\sigma_{\sigma}^2), \quad x \in \mathbb{R}.$$
Note that \((n\sigma_n)^{-1} = \epsilon_n^2\). Choose \(u_n \lesssim n^{-1}\sigma_n((\log n)^{-2/\tau} = \epsilon_n^2\sigma_n^2((\log n)^{-2/\tau}\). On the event \(A_n = \{\sum_{i=1}^n |X_i - m_0| \leq \tau_0^2 nk_n\}\), with \(k_n = O((\log n)^{1/\tau})\), using the inequality \(\log x \geq (x - 1)/x\) valid for every \(x > 0\), we have

\[
\ell_n(\psi_{\gamma', \gamma}(p_F, \sigma)) - \ell_n(p_0) > \ell_n(p_{F_n, \sigma}) - \ell_n(p_0) + n \log c_{\sigma} - 4n\sigma_n^{-2}(a_{\sigma} + 1)u_n(2\epsilon_n^2 + \tau_0^2 k_n) + (a_{\sigma}^2 + 1)u_n^2
\]

\[
\geq \ell_n(p_{F_n, \sigma}) - \ell_n(p_0) + n \log c_{\sigma} - C' n\epsilon_n^2
\]

\[
> \ell_n(p_{F_n, \sigma}) - \ell_n(p_0) - n\epsilon_n^2 - C' n\epsilon_n^2,
\]

where \(C' > 0\) is a large enough constant and \(p_{F_n, \sigma}(\cdot) := c_{\sigma}^{-1} \sum_{j:|\theta_j| \leq a_n} p_j \phi_\sigma(\cdot - \theta_j')\), with normalizing constant \(c_{\sigma} := \sum_{j:|\theta_j| \leq a_n} p_j > 1 - 2\epsilon_n^2 > 1 - \epsilon_n^2\) because \(b > 1\). The proof of Theorem 4 of Shen et al. [6], together with condition (3.4), implies that condition [A1] is satisfied for \(k\) as in part (i) of the statement of Theorem 2.

We now check that condition [A2] is satisfied. Let \(\mathcal{F}\) denote the set of all distribution functions on \(\mathbb{R}\) and

\[
\mathcal{F}_n := \left\{ F \in \mathcal{F} : F = \sum_{j \geq 1} p_j \delta_{\theta_j}, \ |\theta_j| \leq \sqrt{n} \ \forall 1 \leq j \leq H_n, \ \sum_{j > H_n} p_j \leq \epsilon_n \right\}.
\]

We consider the sieve set

\[
\mathcal{S}_n := \left\{ (F, \sigma) : (F, \sigma) \in \mathcal{F}_n \times [\sigma_n, \bar{\sigma}_n] \right\}, \quad (A.4)
\]

with \(\sigma_n = \sigma_n = \epsilon_n^{1/\beta}, \ \bar{\sigma}_n = \exp(t\epsilon_n^2)\) for some constant \(t > 0\) depending on the parameters \(\nu_1, \nu_2 > 0\) of the inverse-gamma prior distribution on \(\sigma\), and \(H_n = [nc_n^2/(\log n)]\).

For some constant \(x_0 > 0\), let \(a_n := 2x_0(\log n)^{1/\tau}\). For \(\gamma' = (m_k, s_k^2)\) and any \(\gamma \in I_k \times J_l\), if \(|\theta| \geq a_n\) and \(|x| \leq a_n/2\), then \(|x - \theta| \geq |\theta|/2\) and we can bound above \(\psi_{\gamma', \gamma}(p_F, \sigma)\) as
follows:

\[ \psi_{\gamma', \gamma}(p_{F, \sigma})(x) = \int_{-\infty}^{\infty} \phi_\sigma(x - \psi_{\gamma', \gamma}(\theta))dF(\theta) \leq \int_{-\infty}^{\infty} \phi_\sigma(x - \theta) \exp(u_n|x - \theta|(|\theta| + 1)/\sigma^2)dF(\theta) \]

\[ < \exp(3a_n^2u_n/\sigma^2) \int_{|\theta| < a_n} \phi_\sigma(x - \theta)dF(\theta) \]

\[ + \int_{|\theta| \geq a_n} \phi_\sigma(x - \theta) \exp(4u_n(x - \theta)^2/\sigma^2)dF(\theta) \]

\[ < \exp(3a_n^2u_n/\sigma^2) \int_{|\theta| < a_n} \phi_\sigma(x - \theta)dF(\theta) \]

\[ + \int_{|\theta| \geq a_n} \phi_\sigma((x - \theta)(1 - 8u_n)^{1/2})dF(\theta) \]

\[ \leq \max \left\{ \exp(3a_n^2u_n/\sigma^2), (1 - 8u_n)^{-1/2} \right\} \]

\[ \times \left( \int_{|\theta| < a_n} \phi_\sigma(x - \theta)dF(\theta) + \int_{|\theta| \geq a_n} \phi_\sigma(x - \theta)dF(\theta) \right), \quad x \in \mathbb{R}, \]

where \( F \sim \text{DP}(\alpha \mathcal{N}, \gamma) \) and \( \tilde{\sigma}_n := \sigma(1 - 8u_n)^{-1/2} \). Now, define the event

\[ \Omega_n := \left\{ -a_n/2 \leq \min_{1 \leq i \leq n} X_i \leq \max_{1 \leq i \leq n} X_i \leq a_n/2 \right\}. \]

Since by Condition (3.3), \( \mathbb{P}(\Omega_n^{(n)}) \leq e^{-cn\alpha_n^{\gamma}} \), we can replace the support \( \mathbb{R} \) of the density \( \psi_{\gamma', \gamma}(p_{F, \sigma}) \) with \( \Omega_n \) and, with abuse of the notation introduced in (2.4), define, for all \( (F, \sigma) \in \mathcal{S}_n \), the density \( q_{(F, \sigma), \gamma'} \) supported on \([-a_n/2, a_n/2]\) obtained from the re-normalized restriction to \([-a_n/2, a_n/2]\) of the function in the last line of (A.5). Replacing \( p_{F, \sigma} \) with \( p_{F, \sigma} \mathbb{1}_{[-a_n/2, a_n/2]} \) we then have that \( \|q_{(F, \sigma), \gamma'} - p_{F, \sigma}\|_1 = o(\epsilon_n) \). We can therefore consider the same tests as in Corollary 1 of Ghosal and van der Vaart [2] and condition (2.6) is verified, together with (2.7), using Proposition 2 of Shen et al. [6]. This implies that also condition (2.8) is satisfied. Since (A.3) implies condition (2.5), there only remains to verify assumption (2.4). The difficulty is to control \( q_{(F, \sigma), \gamma'} \) as \( \sigma \to 0 \).

For every \( (F, \sigma) \) with \( \sigma > \mathcal{G}_n \), the previously defined density \( q_{(F, \sigma), \gamma'} \) can be used as an upper bound on

\[ \sup_{\gamma: \|\gamma - \gamma'\| \leq u_n} \psi_{\gamma', \gamma}(p_{F, \sigma})(\cdot)\mathbb{1}_{[-a_n/2, a_n/2]}(\cdot), \]

with \( \|\gamma - \gamma'\| := (|m - m_k|^2 + |s - s_l|^2)^{1/2} \). Then, for some finite constant \( C_2 > 0 \), both integrals

\[ \int_{F} \int_{\sigma > \sigma_n} Q_{(F, \sigma), \gamma'}([-a_n/2, a_n/2]^n)\pi(dF \mid \gamma')d\pi(\sigma) \]
Supplementary material for Empirical Bayes posterior concentration rates

and

\[ \int_{F_n} \int_{\mathbb{R}_n} Q_n^{(a)}(\cdot,F,\gamma)([-a_n/2, a_n/2]^n)\pi(d\gamma) \]

are \( o(e^{-Cn^2}\pi(B_n \mid \gamma')) \) uniformly in \( \gamma' \in K_n \). We now study

\[ \sup_{\gamma' \in K_n} \int_{F_n} \int_{\mathbb{R}_n} Q_n^{(a)}(\cdot,F,\gamma)([-a_n/2, a_n/2]^n)\pi(d\gamma) \]

Consider the partition \( \bigcup_{j=0}^{\infty}(\sigma_n 2^{-(j+1)}, \sigma_n 2^{-j}) \) of \( (0, \sigma_n) \). For every \( j \geq 0 \), let \( u_{n,j} := n^{-1}e_n(\sigma_n 2^{-j})^2 \), with \( e_n = o(1) \). For every \( \gamma' \in K_n \), we consider a \( u_{n,j} \)-covering of \( \{ \gamma : \| \gamma - \gamma' \| \leq u_{n,j} \} \) with centering points \( \gamma_i, i = 1, \ldots, N_j \), where \( N_j \lesssim (u_n/u_{n,j})^2 \). When \( \sigma \in [\sigma_n 2^{-(j+1)}, \sigma_n 2^{-j}) \), for \( |x| \leq a_n/2 \),

\[ \sup_{\gamma : \| \gamma - \gamma' \| \leq u_{n,j}} \psi_{\gamma',\gamma}(p_{F,\gamma})(x) \leq \max_{1 \leq i \leq N_j} \sup_{\| \gamma - \gamma_i \| \leq u_{n,j}} \int_{-\infty}^{\infty} \phi_{\sigma}(x - \psi_{\gamma_i,\gamma}(\theta))dF(\theta) \]

\[ \leq \max_{1 \leq i \leq N_j} c_{n,i} g_{\sigma,i}(x), \]

where \( g_{\sigma,i} \) is the probability density on \([-a_n/2, a_n/2]\) proportional to

\[ \int_{|\theta| < a_n} \phi_{\sigma}(x - \theta)dF(\theta) + \int_{|\theta| \geq a_n} \phi_{\sigma}\left(\frac{x - \theta}{(1 + o(1/n))^{-1/2}}\right)dF(\theta), \]

with \( F \sim \text{DP}(\alpha_{\theta} N_n) \), and the normalizing constant

\[ c_{n,i} \leq F([-a_n, a_n]) \exp(12\sigma_n^2 \pi_n^{-1} e_n (\log n)^{2/\tau}) + [1 - F([-a_n, a_n])]O(1) = 1 + O(1). \]

This implies that, for \( e_n = O((\log n)^{-2/\tau}) \),

\[ \int_{F_n} \int_{[\sigma_n 2^{-(j+1)}, \sigma_n 2^{-j})} Q_n^{(a)}(\cdot,F,\gamma)([-a_n/2, a_n/2]^n)\pi(d\gamma) \]

\[ \lesssim \frac{u_n^2}{u_{n,j}^2} \pi([\sigma_n 2^{-(j+1)}, \sigma_n 2^{-j})] \]

\[ \lesssim \frac{u_n^2}{u_{n,j}^2} e_n^{-2} (\sigma_n 2^{-j})^{-4} e^{-2^{j+1}/\sigma_n}, \]

whence, for a suitable constant \( C > 0 \),

\[ \sup_{\gamma' \in K_n} \int_{F_n} \int_{[\sigma_n 2^{-(j+1)}, \sigma_n 2^{-j})} Q_n^{(a)}(\cdot,F,\gamma)([-a_n/2, a_n/2]^n)\pi(d\gamma) \]

\[ \lesssim \exp(-Cn^{-2}) \lesssim \exp\left(-Cn^2 e_n^{-2}\right) \]

and (2.4) is verified, which completes the proof for the ordinary smooth case.

We now consider the super-smooth case. The main difference with the ordinary smooth case lies in the fact that, since the rate \( e_n \) is nearly parametric, that is, \( ne_n^2 = O((\log n)^\kappa) \)
for some \( \kappa > 0 \), in order for condition (2.3) to be satisfied, in view of the constraint in (2.2), it suffices that, over some set \( B_n \), for suitable constants \( C, C_1 > 0 \),

\[
\sup_{\gamma \in K_n} \sup_{(F, \sigma) \in B_n} \mathbb{P}_{p_0}^{(n)} \left( \inf_{\psi \in \mathcal{K}_n} \left( \ell_n(\psi_{\gamma}, \gamma(p_{F, \sigma})) - \ell_n(p_0) \right) < -C_1 n \epsilon_n^2 \right) \lesssim e^{-Cn \epsilon_n^2}. \tag{A.6}
\]

It is known from Lemma 2 of Shen and Wasserman [7] that if, for any fixed \( \alpha \in (0, 1] \), the pair \((F, \sigma) \in S_n := \{(F, \sigma) : \rho_\alpha(p_0; p_{F, \sigma}) \leq \epsilon_n^2\}\), then, for every constant \( D > 0 \),

\[
\mathbb{P}_{p_0}^{(n)} \left( \ell_n(p_{F, \sigma}) - \ell_n(p_0) < -(1 + D) n \epsilon_n^2 \right) \lesssim e^{-\alpha D n \epsilon_n^2}.
\]

Let \( \sigma_n = O((\log n)^{-1/r}) \). It is known from Lemma 8 of Scricciolo [5] that, for \( \sigma \in (\sigma_n, \sigma_n + e^{-d_1(1/\sigma_n)^r}) \) with \( d_1 \) a positive constant, there exists a distribution \( F^* = \sum_{j=1}^{N_\sigma} p_j^* \delta_{\theta_j^*} \), with \( N_\sigma = O((a_\sigma/\sigma)^2) \) support points in \([-a_\sigma, a_\sigma]\), where \( a_\sigma = O(\sigma^{-r/(\tau n^2)}) \), such that, for some constant \( c > 0 \),

\[
\max \{ \mathbb{P}_{p_0} \log(p_0/p_{F^*, \sigma}), \mathbb{P}_{p_0} \log^2(p_0/p_{F^*, \sigma}) \} \lesssim e^{-c(1/\sigma)^r}.
\]

Inspection of the proof of the above mentioned Lemma 8 reveals that all arguments remain valid to bound above any \( \rho_\alpha(p_0; p_{F^*, \sigma}) \) divergence for \( \alpha \in (0, 1] \). In fact, using the inequality \(|a^\alpha - b^\alpha| \leq |a^\beta - b^\beta| |a/\alpha - b/\beta| \) valid for all \( a, b > 0 \) and \( 0 \leq \alpha \leq \beta \), if we set \( \beta = 1 \) in our case, then

\[
\rho_\alpha(p_0; p_{F^*, \sigma}) \leq \alpha^{-1} \mathbb{P}_{p_0} \left| (p_0/p_{F^*, \sigma}) - 1 \right|^\alpha.
\]

All bounds used in the proof of Lemma 8 for the various pieces in which the Kullback-Leibler divergence \( \mathbb{P}_{p_0} \log(p_0/p_{F^*, \sigma}) \) is split can be used here to bound above \( \rho_\alpha(p_0; p_{F^*, \sigma}) \). Thus,

\[
\rho_\alpha(p_0; p_{F^*, \sigma}) \lesssim e^{-c(1/\sigma)^r}
\]

for some constant \( c > 0 \) not depending on \( \alpha \). Construct a partition \((U_j)_{j=0}^{N_\sigma} \) of \( \mathbb{R} \), with \( U_0 := (\bigcup_{j=1}^{N_\sigma} U_j)^c \), \( U_j \ni \theta_j^* \) and \( \lambda(U_j) = O(\epsilon_n^{1/(\sigma_n^r)}) \), \( j = 1, \ldots, N_\sigma \), where here \( \lambda \) denotes Lebesgue measure. Then, \( \inf_{\gamma \in \mathcal{K}_n} \min_{1 \leq j \leq N_\sigma} \lambda(U_j) \gtrsim \epsilon_n^{1/(\sigma_n^r)} \). Defined the set

\[
B_n := \left\{ (F, \sigma) : \sigma \in (\sigma_n, \sigma_n + e^{-d_1(1/\sigma_n)^r}), \sum_{j=1}^{N_\sigma} |F(U_j) - p_j^*| \leq e^{-c_1(1/\sigma_n)^r} \right\},
\]

for some constants \( c_2, c_3 > 0 \) we have

\[
\inf_{\gamma \in \mathcal{K}_n} \pi(B_n | \gamma) \gtrsim \exp \left( -c_2 N_\sigma (1/\sigma_n)^r \right) = \exp \left( -c_3 n \epsilon_n^2 \right)
\]

and, for every \( (F, \sigma) \in B_n \),

\[
\rho_\alpha(p_0; p_{F, \sigma}) \lesssim \exp \left( -(1/\sigma_n)^r \right) \lesssim \epsilon_n^2.
\]
Supplementary material for Empirical Bayes posterior concentration rates

Reasoning as in the ordinary smooth case, for \( u_n = O(k_n^{-1} \sigma_n^2 (\log n)^{-2/(\tau_n^2)}) \), on the event \( A_n := \{ \sum_{i=1}^{n} |X_i - m_0| \leq \tau_0 n k_n \} \), with \( k_n \lesssim (\log n)^{1/\tau} \),
\[
\ell_n(\psi_{\gamma', \gamma}(p_{F, \sigma})) - \ell_n(p_0) \geq \ell_n(p_{F, \sigma}) - \ell_n(p_0) + n(c_\sigma - 1) - 4n\sigma_n^{-2}[(a_\sigma^2 + 1)u_n^2 + (a_\sigma + 1)u_n(2a_\sigma + \tau_0^2 k_n)] \\
\geq \ell_n(p_{F, \sigma}) - \ell_n(p_0) - nc_n^2 - C'nc_n^2
\]
for some constant \( C' > 0 \), with \( p_{F, \sigma}(\cdot) := c_\sigma^{-1} \sum_{j: |\theta_j' | \leq a_\sigma} p_j \phi \sigma(\cdot - \theta_j') \), where the normalizing constant
\[
c_\sigma := \sum_{j: |\theta_j' | \leq a_\sigma} p_j \geq 1 - e^{-c_1(1/\sigma_n)^r} \geq 1 - c_n^2.
\]

Lemma 2 of Shen and Wasserman [7] then implies that (A.6) is satisfied. The other parts of the proof of Theorem 2 for the ordinary smooth case go through to this case with modifications. We need to check that the probabilities \( \mathbb{P}_{p_0}^{(n)}(A_n^c) \) and \( \mathbb{P}_{p_0}^{(n)}(\Omega_n^c) \) converge to zero at appropriate rates. By a standard concentration inequality for sums of independent random variables, we have that, for a suitable constant \( c_2 > 0 \), \( \mathbb{P}_{p_0}^{(n)}(A_n^c) \lesssim e^{-c_2 n k_n^2} \). Also, by assumption (3.3) that \( p_0 \) has exponentially small tails, \( \mathbb{P}_{p_0}^{(n)}(\Omega_n^c) \lesssim e^{-c_3 n a_\sigma^r} \).

□

Appendix B: Proof of Proposition 1

The result relies on the following inversion inequalities that relate the \( L_2 \)-distance between the true mixing density and the (random) approximating mixing density in the sieve set \( S_n \), as defined in (A.4) in the proof of Theorem 2, to the \( L_2^* \) or the \( L_1 \)-distance between the corresponding mixed densities:
\[
\| p_Y - p_{0Y} \|_2 \lesssim \begin{cases} \\
\| K * p_Y - K * p_{0Y} \|_2^{\beta_1/(\beta_1 + \eta)}, & \text{ordinary smooth case}, \\
( - \log \| K * p_Y - K * p_{0Y} \|_1 )^{-\beta_1/r_1}, & \text{super-smooth case}.
\end{cases}
\]

To our knowledge, the first inequality, which concerns the ordinary smooth case, is new and of potential independent interest; the second one, concerning the super-smooth case, is similar to the one in Theorem 2 of Nguyen [3], which relates the Wasserstein distance between the mixing distributions to the \( L_1 \)-distance between the mixed densities. In what follows, we use “os” and “ss” as short-hands for “ordinary smooth” and “super-smooth”, respectively. To prove these inequalities, we instrumentally use the \( \text{sinc} \) kernel to characterize regular densities in terms of their approximation properties. We recall that the \( \text{sinc} \) kernel
\[
\text{sinc}(x) = \begin{cases} \\
(\sin x)/(\pi x), & \text{if } x \neq 0, \\
1/\pi, & \text{if } x = 0,
\end{cases}
\]
has Fourier transform \( \widehat{\text{sinc}} \) identically equal to 1 on \([-1, 1]\) and vanishing outside it. For \( \delta > 0 \), let \( \text{sinc}_\delta(\cdot) = \delta^{-1} \text{sinc}(\cdot/\delta) \) and define \( g_\delta \) as the inverse Fourier transform of
so that the optimal choice for $\delta$ turns out to be

$$
\delta = \begin{cases}
O(\|K \ast p_Y - K \ast p_{0Y}\|_2^{1/(\beta_1 + \eta)}), & \text{os case}, \\
O\left((-\log \|K \ast p_Y - K \ast p_{0Y}\|_1)^{-1/r_1}\right), & \text{ss case}.
\end{cases}
$$
For any $1 \leq q \leq \infty$,
\[
\|K * p_Y - K * p_{0Y}\|_q = \|(K * F) * \phi_{\sigma} - K * p_{0Y}\|_q = \|p_{F * K, \sigma} - K * p_{0Y}\|_q.
\]

Then,
\[
\|p_Y - p_{0Y}\|_2 = \|p_{F, \sigma} - p_{0Y}\|_2 \lesssim \min\{\|p_{F, \sigma} - K * p_{0Y}\|_2^{\beta_1/(\beta_1 + \eta)}, \quad \text{os case},
\]
\[
(- \log \|p_{F, \sigma} - K * p_{0Y}\|_1)^{-\beta_1/\tau_1}, \quad \text{ss case}.
\]

For suitable constants $\tau_1, \kappa_1 > 0$, let
\[
\psi_n = \begin{cases}
  n^{-(\beta_1 + \eta)/2(\beta_1 + \eta)}(\log n)^{\kappa_1}, & \text{os case}, \\
  n^{-1/2}(\log n)^{\tau_1}, & \text{ss case}.
\end{cases}
\]

Let $v_n$ be as in the statement of Proposition 1. Then, for all $(F, \sigma) \in \mathcal{S}_n$, the following inclusions hold:
\[
\begin{cases}
  \{ (F, \sigma) : \|p_{F, \sigma} - K * p_{0Y}\|_2 \lesssim \psi_n \} \subseteq \{ (F, \sigma) : \|p_Y - p_{0Y}\|_2 \lesssim v_n \}, & \text{os}, \\
  \{ (F, \sigma) : \|p_{F, \sigma} - K * p_{0Y}\|_1 \lesssim \psi_n \} \subseteq \{ (F, \sigma) : \|p_Y - p_{0Y}\|_2 \lesssim v_n \}, & \text{ss}.
\end{cases}
\]

For $q = 2$ in the ordinary smooth case and $q = 1$ in the super-smooth case, by virtue of Theorem 2, we have $\pi(\mathcal{S}_n : \|p_{F, \sigma} - K * p_{0Y}\|_q \lesssim \psi_n) \rightarrow \hat{\gamma}_n, X^{(n)} \rightarrow 1$ in $\mathbb{P}_{p_{0Y}}$-probability, which implies that $\pi(\mathcal{S}_n : \|p_Y - p_{0Y}\|_2 \lesssim v_n) \rightarrow \hat{\gamma}_n, X^{(n)} \rightarrow 1$ in the same mode of convergence, and the proof is complete. \hfill \Box

**Appendix C: Proof of Theorem 3**

For any intensity $\lambda$, we still denote $M_\lambda = \int_\Omega \lambda(t)dt$ and $\bar{\lambda} = M_\lambda^{-1} \times \lambda \in \mathcal{F}_1$. Without loss of generality, we can assume that $\Omega = [0, T]$. To apply Theorem 1, we must first define the transformation $\psi_{\gamma, \gamma'}$. Note that the parameter $\gamma$ only influences the prior on $\bar{\lambda}$ and has no impact on $M_\lambda$. As explained in Section 2, we can consider the following transformation: for all $\gamma, \gamma' \in \mathbb{R}_+^*$, we set, for any $t$,
\[
\bar{\lambda}(t) = \sum_{j \geq 1} p_j \frac{1_{(0, \theta_j)}(t)}{\theta_j}, \quad \psi_{\gamma, \gamma'}(\bar{\lambda})(t) = \sum_{j \geq 1} p_j \frac{1_{(0, G^{-1}_{\gamma'}(G_{\gamma}(\theta_j)))}(t)}{G^{-1}_{\gamma'}(G_{\gamma}(\theta_j))},
\]

with
\[
p_j = V_j \prod_{l < j} (1 - V_l), \quad V_j \sim \text{Beta}(1, A), \quad \theta_j \sim G_{\gamma} \quad \text{independently}.
\]

So, if $\bar{\lambda}$ is distributed according to a DPM of uniform distributions with base measure indexed by $\gamma$, then $\psi_{\gamma, \gamma'}(\bar{\lambda})$ is distributed according to a DPM of uniform distributions with base measure indexed by $\gamma'$. We prove Theorem 3 for both types of base measure introduced in (4.7). Let $G$ denote the cdf of a Gamma($a$, 1) random variable and $g$ its density.
For the first type of base measure we have $G_{\gamma'}^{-1}(G_\gamma(\theta)) = G^{-1}((G(\gamma\theta)G(\gamma'T)/G(\gamma T))/\gamma'$ if $\theta \leq T$ and $G_{\gamma'}^{-1}(G_\gamma(\theta)) = T$ if $\theta \geq T$. For any $\theta \in [0, T]$, if $\gamma' \geq \gamma$ then

$$G_{\gamma'}^{-1}(G_\gamma(\theta)) \leq \theta \quad \text{and} \quad G_{\gamma'}^{-1}(G_\gamma(\theta)) \geq \frac{\gamma\theta}{\gamma}. \quad (C.1)$$

The second inequality in (C.1) is straightforward. The first inequality in (C.1) is equivalent to $G(\gamma\theta)G(\gamma'T) \leq G(\gamma'T)G(\gamma T)$ and is deduced from the following argument. Let $\Delta(\theta) = G(\gamma\theta)G(\gamma'T) - G(\gamma'T)G(\gamma T)$. Then, $\Delta(0) = 0$ and $\Delta(T) = 0$. By Rolle’s Theorem, there exists $c \in (0, T)$ such that $\Delta'(c) = 0$. We have $\Delta'(\theta) = \gamma g(\gamma\theta)G(\gamma'T) - \gamma' g(\gamma'T)G(\gamma T)$ which is proportional to $\theta_{\theta - \gamma\theta}[\gamma e^{(\gamma' - \gamma)t}G(\gamma'T) - (\gamma')^2 G(\gamma T)]$. The function inside brackets is increasing so that $\Delta'(c) = 0$ for $\theta = c$ and $\Delta'(\theta) \geq 0$ for $\theta > c$. Therefore, $\Delta$ is first decreasing and then increasing. Since $\Delta(0) = \Delta(T) = 0$, $\Delta$ is negative on $(0, T)$, which achieves the proof of (C.1). For the second type of base measure, for $\theta \leq T$, we have that, for every $\gamma$, $\gamma' > 0$, $G_{\gamma'}^{-1}(G_\gamma(\theta)) = T\gamma\theta/[\gamma'(T - \theta\gamma'/\gamma')]$ and (C.1) is straightforward.

We first verify assumption [A1]. At several places, by using (4.1) and (4.4), we use that, under $P^{(n)}$ to $\Gamma_n$, for any interval $I$, the number of points of $N$ falling in $I$ is controlled by the number of points of a Poisson process with intensity $n(1 + \alpha)n_{2L}$ falling in $I$. Let $u_n = (n \log n)^{-1}$ so that $u_n = o(\epsilon_n^2)$ and choose $k \geq 6$ so that $u_n^{-1} = o((n\epsilon_n^2)^k/2)$ and (2.2) holds (note that $N_n(u_n)$ is the same order as $u_n^{-1}$). Using the proof of Corollary 4.1 of Donnet et al. [1], we construct $B_n = B_n^\gamma$ (since in [A1] $B_n$ may depend on $\gamma$) as the set of $\lambda = M_\lambda\lambda$ such that $|M_\lambda - M_{\lambda0}| \leq \epsilon_n$ and $\psi_{\gamma, \gamma + u_n}(\lambda) \in B_n$, with $B_n = \{\lambda, P^\gamma(x) = \int_x^\infty \theta e^{-\lambda}d\theta' : \theta' \in \mathcal{N}'\}$, where $\mathcal{N}'$ is as defined in the proof of Lemma 8 in Appendix A of Salomond [4]. Note that from Lemma 8 in Appendix A of Salomond [4], if $\lambda \in B_n$ and $|M_\lambda - M_{\lambda0}| \leq \epsilon_n$,

$$P^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_0) \leq - (\kappa_0 + 1)\epsilon_n^2 | \Gamma_n) = O((n\epsilon_n^2)^{-k/2}) \log n^k) \quad (C.2)$$

for $\kappa_0$ a constant. To prove (2.3), it is enough to control $\inf_{\gamma' \in [\gamma, \gamma + u_n]} \ell_n(M_\lambda \psi_{\gamma, \gamma'}(\lambda))$. Using (C.1), we have that for any $\gamma' \in [\gamma, \gamma + u_n]$, on $\Gamma_n$,

$$G_{\gamma + u_n}^{-1}(G_\gamma(\theta)) \leq G_{\gamma'}^{-1}(G_\gamma(\theta))$$

and

$$G_{\gamma'}^{-1}(G_\gamma(\theta)) \leq \frac{\gamma + u_n}{\gamma} G_{\gamma + u_n}^{-1}(G_\gamma(\theta)) \leq \frac{\gamma + u_n}{\gamma} G_{\gamma + u_n}^{-1}(G_\gamma(\theta)).$$

Therefore,

$$\frac{\gamma}{\gamma + u_n} \psi_{\gamma, \gamma + u_n}(\lambda)(t) \leq \psi_{\gamma, \gamma'}(\lambda)(t)$$

$$\leq \psi_{\gamma, \gamma + u_n}(\lambda)(t) + \sum_{j \geq 1} p_j \frac{G_{\gamma + u_n}^{-1}(G_\gamma(\theta)) - u_n G_{\gamma + u_n}^{-1}(G_\gamma(\theta))}{G_{\gamma + u_n}^{-1}(G_\gamma(\theta))}(t) \quad (C.3)$$
so that, for \( n \) large enough,
\[
\inf_{\gamma' \in [\gamma, \gamma + u_n]} \ell_n(M_{\lambda} \psi_{\gamma, \gamma'}(\bar{\lambda})) = \inf_{\gamma' \in [\gamma, \gamma + u_n]} \left\{ \int_0^T \log(M_{\lambda} \psi_{\gamma, \gamma'}(\bar{\lambda}))(t) \, dN_t - \int_0^T M_{\lambda} \psi_{\gamma, \gamma'}(\bar{\lambda})(t) \, Y_t \, dt \right\} \\
\geq \int_0^T \log \left( M_{\lambda} \psi_{\gamma, \gamma + u_n}(\bar{\lambda})(t) \right) \, dN_t + N[0, T] \log \left( \frac{\gamma}{\gamma + u_n} \right) \\
- M_{\lambda} \int_0^T \psi_{\gamma, \gamma + u_n}(\bar{\lambda})(t) \, Y_t \, dt - \frac{u_n}{\gamma} \left( M_{\lambda} n(1 + \alpha) m_2 \right) \\
\geq \ell_n(M_{\lambda} \psi_{\gamma, \gamma + u_n}(\bar{\lambda})) - \frac{u_n}{\gamma} \left( M_{\lambda} n(1 + \alpha) m_2 + 2N[0, T] \right),
\]
where the last line uses \( \log(1-x) \geq -2x \) for \( x > 0 \) small enough. By using the Bienaymé-Chebyshev inequality, if \( Z \) is a Poisson variable with parameter \( n(1 + \alpha)m_2M_{\lambda_0} \), we have
\[
P(|Z - n(1 + \alpha)m_2M_{\lambda_0}| > n(1 - \alpha)m_2M_{\lambda_0}) = o(1).
\]
Then the event \( \{N[0, T] \leq 2M_{\lambda_0}nm_2n\} \) has probability going to 1 and, on this event,
\[
\inf_{\gamma' \in [\gamma, \gamma + u_n]} \ell_n(M_{\lambda} \psi_{\gamma, \gamma'}(\bar{\lambda})) \geq \ell_n(M_{\lambda} \psi_{\gamma, \gamma + u_n}(\bar{\lambda})) - n\gamma^{-1}m_2[M_{\lambda}(1 + \alpha) + 4M_{\lambda_0}]u_n. \tag{C.4}
\]
Combining this lower bound with (C.2), for all \( \lambda = M_{\lambda} \psi_{\gamma, \gamma + u_n}(\bar{\lambda}) \), with \( \psi_{\gamma, \gamma + u_n}(\bar{\lambda}) \in \bar{B}_n \) and \( |M_{\lambda} - M_{\lambda_0}| \leq \epsilon_n \),
\[
P_{\lambda_0}^{(n)} \left( \inf_{\gamma' \in [\gamma, \gamma + u_n]} \ell_n(M_{\lambda} \psi_{\gamma, \gamma'}(\bar{\lambda})) - \ell_n(\lambda_0) \leq -(\kappa_0 + 2) n \epsilon_n^2 \mid \Gamma_n \right) = O((n\epsilon_n^2)^{-k/2}).
\]
The left hand inequality is then negligible with respect to \( u_n \) which is the same order as \( N_n(u_n)^{-1} \) and assumption \([A1]\) is satisfied if \( C_1 \geq \kappa_0 + 2 \).

We now verify assumption \([A2]\). First, note that using (C.3), (2.8) is obviously satisfied. Mimicking the proof of Lemma 8 of Salomond [4], we have that over any compact subset \( K' \) of \( (0, \infty) \),
\[
\inf_{\gamma \in K'} \pi_1(\bar{B}_n \mid \gamma) \geq e^{-C_k n \epsilon_n^2} \tag{C.5}
\]
for some \( C_k > 0 \), when \( n \) is large enough. By definition of \( \bar{B}_n^\gamma \), \( \pi(\bar{B}_n^\gamma \mid \gamma) = \pi_1(\bar{B}_n \mid \gamma + u_n) \pi_M([M_{\lambda_0} - \epsilon_n, M_{\lambda_0} + \epsilon_n]) \) which, together with (C.5), implies that \( \inf_{\gamma \in K} \pi(\bar{B}_n^\gamma \mid \gamma) \geq e^{-2C_k n \epsilon_n^2} \) when \( n \) is large enough, so that (2.5) is satisfied as soon as \( j \) is large enough. We now define the measure \( Q_{\lambda, \gamma}^{(n)} \) with
\[
dQ_{\lambda, \gamma}^{(n)} = 1_{\Gamma_n} \times \sup_{\gamma' \in [\gamma, \gamma + u_n]} \exp(\ell_n(M_{\lambda} \psi_{\gamma, \gamma'}(\bar{\lambda}))) \, d\mu
\]
and \( \mu \) the measure such that under \( \mu \) the process is an homogeneous Poisson process with intensity 1. Using (C.1) and similarly to (C.4), we obtain that, for all \( \gamma' \in [\gamma, \gamma + u_n] \),
\[
\bar{\lambda}(t) - \sum_{j \geq 1} p_j \frac{1_{(\gamma_j / \gamma) > \theta_j}}{\theta_j} \leq \psi_{\gamma, \gamma'}(\bar{\lambda})(t) \leq \frac{\gamma + u_n}{\gamma} \bar{\lambda}(t)
\]
and we have

$$Q^{(n)}_{\lambda, \gamma}(\chi^{(n)}) = E^{(n)}_{\lambda} \left[ 1_{\Gamma_n} \sup_{\gamma' \in [\gamma, \gamma + u_n]} \exp \left( -M_{\lambda} \int_0^T \psi_{\gamma, \gamma'}(\bar{\lambda})(t)Y_t dt \right) \right]$$

$$\leq E^{(n)}_{\lambda} \left[ 1_{\Gamma_n} \exp(nm_2(1 + \alpha)M_{\lambda}\gamma^{-1} u_n + \log(1 + u_n/\gamma)N[0, T]) \right]$$

$$\leq E^{(n)}_{\lambda} \left[ 1_{\Gamma_n} \exp(nm_2(1 + \alpha)M_{\lambda}\gamma^{-1} u_n + u_n\gamma^{-1} N[0, T]) \right]$$

$$\leq \exp(nm_2(1 + \alpha)M_{\lambda}\gamma^{-1} u_n + (1 + \alpha)nm_2M_{\lambda}(e^{u_n/\gamma} - 1))$$

$$\leq \exp(3nm_2(1 + \alpha)\gamma^{-1} M_{\lambda} u_n)$$

when $n$ is large enough since $\exp(x) - 1 \leq 2x$ for $x > 0$ small enough. Let $\phi_{n,j}$ be the tests defined in Proposition 6.2 of Donnet et al. [1] over $S_{n,j}(\tilde{e}_n)$. Using the previous computations, we have

$$Q^{(n)}_{\lambda, \gamma}[1 - \phi_{n,j}] \leq E^{(n)}_{\lambda} \left[ (1 - \phi_{n,j}) \exp(nm_2(1 + \alpha)M_{\lambda}\gamma^{-1} u_n + u_n\gamma^{-1} N[0, T])1_{\Gamma_n} \right]$$

$$\leq e^{\exp(nm_2(1 + \alpha)M_{\lambda}\gamma^{-1} u_n + \log(1 + u_n/\gamma)N[0, T])1_{\Gamma_n}}$$

$$\leq e^{4nm_2(1 + \alpha)\gamma^{-1} M_{\lambda} u_n} \sup \left\{ e^{-cn_3\tilde{e}_n^2/2}, e^{-cn_3\tilde{e}_n/2} \right\}.$$

As in Salomond [4], we set $S_n = \{ \tilde{\lambda} : \tilde{\lambda}(0) \leq M_n \}$, with $M_n = \exp(c_1 n\tilde{e}_n^2)$ and $c_1 > 0$ is a constant. From Lemma 9 of Salomond [4], there exists $a > 0$ such that $\sup_{\gamma \in K'} \pi_1(S_n^c | \gamma) \leq e^{-c_1(a+1)n\tilde{e}_n^2}$, so that when $n$ is large enough,

$$\sup_{\gamma \in K'} \int_{R_+} \int_{S_n} Q^{(n)}_{\lambda, \gamma}(\chi^{(n)}) d\pi_1(\tilde{\lambda} | \gamma) \pi_M(M_{\lambda}) dM_{\lambda}$$

$$\lesssim e^{-c_1(a+1)n\tilde{e}_n^2} \int_{R_+} e^{\delta M_{\lambda} \pi_M(M_{\lambda}) dM_{\lambda}} \lesssim e^{-c_1(a+1)n\tilde{e}_n^2},$$

with $\delta$ that can be chosen as small as needed since $nu_n = o(1)$. This proves (2.4) by conveniently choosing $c_1$. Combining the above upper bound with Proposition 6.2 of Donnet et al. [1], together with Remark 1, achieves the proof of Theorem 3. □

References


