Submitted to Bernoulli
arXiv: 1406.4406v1

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 SCRICCIOLO[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.scricciolo@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 $\mathrm{N}(m,\, s^2)$. Hereafter, we use the notation N_γ as shorthand for $\mathrm{N}(m,\, s^2)$. We consider a set $\mathcal{K}_n = [m_1,\, m_2] \times [s_1^2,\, s_n^2]$, with constants $-\infty < m_1 \leq m_2 < \infty$, $s_1^2 > 0$ and a positive sequence $s_n^2 \to \infty$ as a power of $\log n$, such that $\mathbb{P}_{p_0}^{(n)}(\hat\gamma_n \in \mathcal{K}_n^c) = o(1)$. 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 $[s_1^2,\, s_n^2]$ with intervals $J_l = [s_l^2,\, s_{l+1}^2) = [s_1^2(1+u_n)^{l-1},\, s_1^2(1+u_n)^l)$, 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 $\rho_l = (s^2/s_l^2)^{1/2}$. Let $m \in I_k$, $k = 1,\, \ldots,\, L_{mn}$. For any $\gamma' = (m_k,\, s_l^2)$ and $\gamma = (m,\, s^2)$, if $\theta_j' \sim \mathrm{N}_{\gamma'}$, then $\theta_j = [\rho_l(\theta_j' - m_k) + m] \sim \mathrm{N}_\gamma$, $j \in \mathbb{N}$.

Therefore, conditionally on σ, for $F \sim \mathrm{DP}(\alpha_{\mathbb{R}}\mathrm{N}_{\gamma'})$,

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

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

$$\psi_{\gamma',\gamma}(\theta'_j) = \rho_l(\theta'_j - m_k) + m. \tag{A.1}$$

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

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

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 θ_j^*'s is bounded below by $\delta = \sigma\epsilon_n^{2b}$, for some $b > \max\{1,\, (2\beta)^{-1}\}$. A partition $(U_j)_{j=1}^M$ of $\mathbb{R}$ can be constructed following the proof of Theorem 4 in Shen et al. [6] so that $(U_j)_{j=1}^K$ is a partition of $[-a_\sigma, a_\sigma]$, with $a_\sigma = a_0|\log\sigma|^{1/\tau}$, 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 σ 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 $a_1\sigma\epsilon_n^{2b} \leq \alpha_{\mathbb{R}}\mathrm{N}_\gamma(U_j) \leq 1$ for some constant $a_1 > 0$. Note that, as in Shen et al. [6], $M \lesssim \sigma^{-1}(\log n)^{1+1/\tau}$ and, for every $1 \leq j \leq K$, we have $\mathrm{N}_\gamma(U_j) \gtrsim (\delta/s)e^{-2(a_\sigma/s)^2} \gtrsim \sigma_n\epsilon_n^{2b}$ uniformly in $\gamma \in \mathcal{K}_n$. 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

$$\sum_{j=1}^{M} |F(U_j) - p_j^*| \leq 2\epsilon_n^{2b}, \qquad \min_{1\leq j\leq M} F(U_j) \geq \epsilon_n^{4b}/2.$$

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

$$\inf_{\gamma\in\mathcal{K}_n} \pi\left(B_n \mid \gamma\right) \gtrsim \exp\left(-c\sigma_n^{-1}(\log n)^{2+1/\tau}\right). \tag{A.3}$$

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

$$\begin{aligned}
\psi_{\gamma',\gamma}(p_{F,\sigma})(x) &= \sum_{j\geq 1} p_j \phi_\sigma(x - \psi_{\gamma',\gamma}(\theta'_j)) \\
&\geq \sum_{j:\,|\theta'_j|\leq a_\sigma} p_j \phi_\sigma(x - \psi_{\gamma',\gamma}(\theta'_j)) \\
&> \sum_{j:\,|\theta'_j|\leq a_\sigma} p_j \phi_\sigma(x - \theta'_j) \\
&\qquad\qquad \times \exp\left(-4[|x - \theta'_j|(a_\sigma + 1)u_n + (a_\sigma^2 + 1)u_n^2]/\sigma_n^2\right), \quad x \in \mathbb{R}.
\end{aligned}$$

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 $A_n = \{\sum_{i=1}^n |X_i - m_0| \leq \tau_0^2 n k_n\}$, with $k_n = O((\log n)^{1/\tau})$, using the inequality $\log x \geq (x-1)/x$ valid for every $x > 0$, we have

$$\begin{aligned}\ell_n(\psi_{\gamma',\gamma}(p_{F,\sigma})) - \ell_n(p_0) &> \ell_n(p_{F_n,\sigma}) - \ell_n(p_0) + n\log c_\sigma \\ &\qquad - 4n\sigma_n^{-2}[(a_\sigma+1)u_n(2a_\sigma + \tau_0^2 k_n) + (a_\sigma^2+1)u_n^2] \\ &\geq \ell_n(p_{F_n,\sigma}) - \ell_n(p_0) + n\log c_\sigma - C' n\epsilon_n^2 \\ &> \ell_n(p_{F_n,\sigma}) - \ell_n(p_0) - n\epsilon_n^2 - C' n\epsilon_n^2,\end{aligned}$$

where $C' > 0$ is a large enough constant and $p_{F_n,\sigma}(\cdot) := c_\sigma^{-1}\sum_{j:\,|\theta_j'|\leq a_\sigma} p_j\phi_\sigma(\cdot - \theta_j')$, with normalizing constant

$$c_\sigma := \sum_{j:\,|\theta_j'|\leq a_\sigma} p_j > 1 - 2\epsilon_n^{2b} > 1 - \epsilon_n^2$$

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

$$\mathcal{S}_n := \{(F,\,\sigma) :\ (F,\,\sigma) \in \mathcal{F}_n \times [\underline{\sigma}_n,\,\bar{\sigma}_n]\}, \tag{A.4}$$

with $\underline{\sigma}_n = \sigma_n = \epsilon_n^{1/\beta}$, $\bar{\sigma}_n = \exp(tn\epsilon_n^2)$ for some constant $t > 0$ depending on the parameters $\nu_1,\,\nu_2 > 0$ of the inverse-gamma prior distribution on σ, and $H_n = \lfloor n\epsilon_n^2/(\log n)\rfloor$. For some constant $x_0 > 0$, let $a_n := 2x_0(\log n)^{1/\tau}$. For $\gamma' = (m_k,\, s_l^2)$ 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:

$$
\begin{aligned}
\psi_{\gamma',\gamma}(p_{F,\sigma})(x) &= \int_{-\infty}^{\infty} \phi_\sigma(x-\psi_{\gamma',\gamma}(\theta))\mathrm{d}F(\theta) \\
&\le \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)\exp\left(u_n|x-\theta|(|\theta|+1)/\sigma^2\right)\mathrm{d}F(\theta) \\
&< \exp\left(3a_n^2 u_n/\sigma^2\right)\int_{|\theta|<a_n} \phi_\sigma(x-\theta)\mathrm{d}F(\theta) \\
&\qquad + \int_{|\theta|\ge a_n} \phi_\sigma(x-\theta)\exp\left(4u_n(x-\theta)^2/\sigma^2\right)\mathrm{d}F(\theta) \\
&< \exp\left(3a_n^2 u_n/\sigma^2\right)\int_{|\theta|<a_n} \phi_\sigma(x-\theta)\mathrm{d}F(\theta) \\
&\qquad + \int_{|\theta|\ge a_n} \phi_\sigma((x-\theta)(1-8u_n)^{1/2})\mathrm{d}F(\theta) \\
&\le \max\left\{\exp\left(3a_n^2 u_n/\sigma^2\right),\, (1-8u_n)^{-1/2}\right\} \\
&\qquad \times \left(\int_{|\theta|<a_n} \phi_\sigma(x-\theta)\,\mathrm{d}F(\theta) + \int_{|\theta|\ge a_n} \phi_{\tilde\sigma_n}(x-\theta)\mathrm{d}F(\theta)\right), \quad x\in\mathbb{R},
\end{aligned}
\tag{A.5}
$$

where $F \sim \mathrm{DP}(\alpha_{\mathbb{R}}\mathrm{N}_{\gamma'})$ and $\tilde\sigma_n := \sigma(1-8u_n)^{-1/2}$. Now, define the event

$$
\Omega_n := \left\{ -a_n/2 \le \min_{1\le i\le n} X_i \le \max_{1\le i\le n} X_i \le a_n/2 \right\}.
$$

Since by Condition (3.3), $\mathbb{P}_{p_0}^{(n)}(\Omega_n^c) \lesssim e^{-cna_n^\tau}$, we can replace the support $\mathbb{R}$ of the density $\psi_{\gamma',\gamma}(p_{F,\sigma})$ with Ω_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 $p_{F,\sigma}\mathbf{1}_{[-a_n/2,\,a_n/2]}$ 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,σ) with $\sigma > \underline{\sigma}_n$, the previously defined density $q_{(F,\sigma),\gamma'}$ can be used as an upper bound on

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

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

$$
\int_{\mathcal{F}}\int_{\sigma>\bar\sigma_n} Q_{(F,\sigma),\gamma'}^{(n)}([-a_n/2,\, a_n/2]^n)\pi(\mathrm{d}F\mid\gamma')\mathrm{d}\pi(\sigma)
$$

and

$$\int_{\mathcal{F}_n^c}\int_{\underline{\sigma}_n}^{\bar{\sigma}_n} Q^{(n)}_{(F,\sigma),\gamma'}([-a_n/2,\, a_n/2]^n)\pi(\mathrm{d}F \mid \gamma')\mathrm{d}\pi(\sigma)$$

are $o(e^{-C_2 n\epsilon_n^2}\pi(B_n \mid \gamma'))$ uniformly in $\gamma' \in \mathcal{K}_n$. We now study

$$\sup_{\gamma'\in\mathcal{K}_n}\int_{\mathcal{F}}\int_0^{\underline{\sigma}_n} Q^{(n)}_{(F,\sigma),\gamma'}([-a_n/2,\, a_n/2]^n)\pi(\mathrm{d}F \mid \gamma')\mathrm{d}\pi(\sigma).$$

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

$$\begin{aligned}\sup_{\gamma:\,\|\gamma-\gamma'\|\leq u_n}\psi_{\gamma',\gamma}(p_{F,\sigma})(x) &\leq \max_{1\leq i\leq N_j}\sup_{\gamma:\,\|\gamma-\gamma_i\|\leq u_{n,j}}\int_{-\infty}^{\infty}\phi_\sigma(x-\psi_{\gamma_i,\gamma}(\theta))\mathrm{d}F(\theta)\\ &\leq \max_{1\leq i\leq N_j} c_{n,i}g_{\sigma,i}(x),\end{aligned}$$

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

$$\int_{|\theta|<a_n}\phi_\sigma(x-\theta)\mathrm{d}F(\theta) + \int_{|\theta|\geq a_n}\phi_\sigma\left(\frac{x-\theta}{(1+o(1/n))^{-1/2}}\right)\mathrm{d}F(\theta),$$

with $F \sim \mathrm{DP}(\alpha_{\mathbb{R}}\mathrm{N}_{\gamma_i})$, and the normalizing constant

$$c_{n,i} \leq F((-a_n,\, a_n))\exp\left(12x_0^2 n^{-1}e_n(\log n)^{2/\tau}\right) + [1-F((-a_n,\, a_n))]O(1) = 1+O(1).$$

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

$$\begin{aligned}\int_{\mathcal{F}}\int_{\underline{\sigma}_n 2^{-(j+1)}}^{\underline{\sigma}_n 2^{-j}} Q^{(n)}_{(F,\sigma),\gamma'}([-a_n/2,\, a_n/2]^n)\pi(\mathrm{d}F \mid \gamma')\mathrm{d}\pi(\sigma) &\lesssim N_j\pi([\underline{\sigma}_n 2^{-(j+1)},\, \underline{\sigma}_n 2^{-j}))\\ &\lesssim \frac{u_n^2}{u_{n,j}^2}\pi([\underline{\sigma}_n 2^{-(j+1)},\, \underline{\sigma}_n 2^{-j}))\\ &\lesssim u_n^2 n^2 e_n^{-2}(\underline{\sigma}_n 2^{-j})^{-4}e^{-2^{j-1}/\underline{\sigma}_n},\end{aligned}$$

whence, for a suitable constant $C > 0$,

$$\sup_{\gamma'\in\mathcal{K}_n}\int_{\mathcal{F}}\int_0^{\underline{\sigma}_n} Q^{(n)}_{(F,\sigma),\gamma'}([-a_n/2,\, a_n/2]^n)\pi(\mathrm{d}F \mid \gamma')\mathrm{d}\pi(\sigma) \lesssim \exp\left(-C\underline{\sigma}_n^{-1}\right) \lesssim \exp\left(-Cn\epsilon_n^2\right)$$

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 ϵ_n is nearly parametric, that is, $n\epsilon_n^2 = O((\log n)^\kappa)$

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, $C_1 > 0$,

$$\sup_{\gamma' \in \mathcal{K}_n} \sup_{(F,\sigma)\in B_n} \mathbb{P}_{p_0}^{(n)}\left(\inf_{\gamma:\, \|\gamma-\gamma'\| \le u_n} \ell_n(\psi_{\gamma',\gamma}(p_{F,\sigma})) - \ell_n(p_0) < -C_1 n\epsilon_n^2 \right) \lesssim e^{-Cn\epsilon_n^2}. \quad \text{(A.6)}$$

It is known from Lemma 2 of Shen and Wasserman [7] that if, for any fixed $\alpha \in (0,\, 1]$, the pair $(F,\, \sigma) \in S_n := \{(F,\, \sigma) :\ \rho_\alpha(p_0;\, p_{F,\sigma}) \le \epsilon_n^2\}$, 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 $a_\sigma = O(\sigma^{-r/(\tau\wedge 2)})$, 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 $|a^\alpha - b^\alpha| \le |a^\beta - b^\beta|^{\alpha/\beta}$ valid for all a, $b > 0$ and $0 \le \alpha \le \beta$, if we set $\beta = 1$ in our case, then

$$\rho_\alpha(p_0;\, p_{F^*,\sigma}) \le \alpha^{-1}\mathbb{P}_{p_0}|(p_0/p_{F^*,\sigma}) - 1|^\alpha.$$

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 α. 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 $\lambda(U_j) = O(e^{-c_1(1/\sigma)^r})$, $j = 1, \ldots, N_\sigma$, where here λ denotes Lebesgue measure. Then, $\inf_{\gamma\in\mathcal{K}_n} \min_{1\le j\le N_\sigma} \mathrm{N}_\gamma(U_j) \gtrsim e^{-c_1(1/\sigma)^r}$. Defined the set

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

for some constants c_2, $c_3 > 0$ we have

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

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

$$\rho_\alpha(p_0;\, p_{F,\sigma}) \lesssim \exp\left(-(1/\sigma_n)^r\right) \lesssim \epsilon_n^2.$$

Reasoning as in the ordinary smooth case, for $u_n = O(k_n^{-1}\sigma_n^2\epsilon_n^2(\log n)^{-2/(\tau\wedge 2)})$, on the event $A_n := \{\sum_{i=1}^n |X_i - m_0| \le \tau_0^2 n k_n\}$, with $k_n \lesssim (\log n)^{1/\tau}$,

$$\begin{aligned}\ell_n(\psi_{\gamma',\gamma}(p_{F,\sigma})) - \ell_n(p_0) &\ge \ell_n(p_{F_n,\sigma}) - \ell_n(p_0) + n(c_\sigma - 1)\\ &\qquad - 4n\sigma_n^{-2}[(a_\sigma^2+1)u_n^2 + (a_\sigma+1)u_n(2a_\sigma + \tau_0^2 k_n)]\\ &\ge \ell_n(p_{F_n,\sigma}) - \ell_n(p_0) - n\epsilon_n^2 - C' n\epsilon_n^2\end{aligned}$$

for some constant $C' > 0$, with $p_{F_n,\sigma}(\cdot) := c_\sigma^{-1}\sum_{j:\,|\theta_j'|\le a_\sigma} p_j\phi_\sigma(\cdot - \theta_j')$, where the normalizing constant

$$c_\sigma := \sum_{j:\,|\theta_j'|\le a_\sigma} p_j \ge 1 - e^{-c_1(1/\sigma_n)^r} \ge 1 - \epsilon_n^2.$$

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, $\mathbb{P}_{p_0}^{(n)}(\Omega_n^c) \lesssim e^{-c_3 n a_n^\tau}$. $\square$

Appendix B: Proof of Proposition 1

The result relies on the following inversion inequalities that relate the $\mathbb{L}_2$-distance between the true mixing density and the (random) approximating mixing density in the sieve set $\mathcal{S}_n$, as defined in (A.4) in the proof of Theorem 2, to the $\mathbb{L}_2$- or the $\mathbb{L}_1$-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 $\mathbb{L}_1$-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 *sinc* kernel to characterize regular densities in terms of their approximation properties. We recall that the *sinc* kernel

$$\operatorname{sinc}(x) = \begin{cases} (\sin x)/(\pi x), & \text{if } x \ne 0,\\ 1/\pi, & \text{if } x = 0,\end{cases}$$

has Fourier transform $\widehat{\operatorname{sinc}}$ identically equal to 1 on $[-1, 1]$ and vanishing outside it. For $\delta > 0$, let $\operatorname{sinc}_\delta(\cdot) = \delta^{-1}\operatorname{sinc}(\cdot/\delta)$ and define g_δ as the inverse Fourier transform of

$\widehat{\text{sinc}}_\delta/\hat{K}$, that is

$$g_\delta(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\frac{\widehat{\text{sinc}}_\delta(t)}{\hat{K}(t)}\mathrm{d}t,\quad x\in\mathbb{R}.$$

Let $\hat{g}_\delta=\widehat{\text{sinc}}_\delta/\hat{K}$ be the Fourier transform of g_δ. So, $\text{sinc}_\delta=K*g_\delta$ and $p_Y*\text{sinc}_\delta=(p_Y*K)*g_\delta=(K*p_Y)*g_\delta$. Then,

$$\begin{aligned}\|p_Y-p_{0Y}\|_2^2&\le\|p_Y*\text{sinc}_\delta-p_{0Y}*\text{sinc}_\delta\|_2^2+\|p_Y-p_Y*\text{sinc}_\delta\|_2^2+\|p_{0Y}-p_{0Y}*\text{sinc}_\delta\|_2^2\\&\lesssim\|p_Y*\text{sinc}_\delta-p_{0Y}*\text{sinc}_\delta\|_2^2+\|p_Y-p_Y*\text{sinc}_\delta\|_2^2+\delta^{2\beta_1}\end{aligned}$$

because, by assumption (3.9),

$$\begin{aligned}\|p_{0Y}-p_{0Y}*\text{sinc}_\delta\|_2^2&=\int_{-\infty}^{\infty}|\hat{p}_{0Y}(t)|^2|1-\widehat{\text{sinc}}_\delta(t)|^2\mathrm{d}t\\&<\delta^{2\beta_1}\int_{|t|>1/\delta}(1+t^2)^{\beta_1}|\hat{p}_{0Y}(t)|^2\mathrm{d}t\lesssim\delta^{2\beta_1}.\end{aligned}$$

Now, recall that $p_Y=p_{F,\sigma}=F*\phi_\sigma$. For $(F,\sigma)\in\mathcal{F}_n$, $\sigma\ge\underline{\sigma}_n\propto C\delta(\log n)^{\kappa_2}$, where $2\kappa_2\ge1$, $C^2/2\ge(2\beta_1+1)/[2(\beta_1+\eta)+1]$ and $\delta\equiv\delta_n\gtrsim n^{-1/[2(\beta_1+\eta)+1]}$,

$$\begin{aligned}\|p_Y-p_Y*\text{sinc}_\delta\|_2^2&\le\int_{|t|>1/\delta}|\hat{\phi}_\sigma(t)|^2\mathrm{d}t\\&\lesssim(\sigma^2/\delta)^{-1}e^{-(\sigma/\delta)^2/2}\\&\lesssim[\delta(\log n)^{2\kappa_2}]^{-1}e^{-C^2(\log n)^{2\kappa_2}/2}\lesssim\delta^{2\beta_1}.\end{aligned}$$

In the ordinary smooth case,

$$\begin{aligned}\|p_Y*\text{sinc}_\delta-p_{0Y}*\text{sinc}_\delta\|_2^2&=\|(K*p_Y)*g_\delta-(K*p_{0Y})*g_\delta\|_2^2\\&\le\delta^{-2\eta}\int_{|t|\le1/\delta}|\hat{K}(t)|^2|\hat{p}_Y(t)-\hat{p}_{0Y}(t)|^2\mathrm{d}t\\&\le\delta^{-2\eta}\|K*p_Y-K*p_{0Y}\|_2^2.\end{aligned}$$

In the super-smooth case, $\|p_Y*\text{sinc}_\delta-p_{0Y}*\text{sinc}_\delta\|_2^2\le\|K*p_Y-K*p_{0Y}\|_1^2\|g_\delta\|_2^2$, where

$$\|g_\delta\|_2^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\widehat{\text{sinc}}_\delta(t)|^2}{|\hat{K}(t)|^2}\mathrm{d}t=\frac{1}{2\pi}\int_{\delta|t|\le1}|\hat{K}(t)|^{-2}\mathrm{d}t\lesssim e^{2\varrho\delta^{-r_1}}.$$

Combining partial results, for $(F,\sigma)\in\mathcal{S}_n$,

$$\|p_Y-p_{0Y}\|_2^2\lesssim\delta^{2\beta_1}+\begin{cases}\|K*p_Y-K*p_{0Y}\|_2^2\times\delta^{-2\eta}, & \text{os case,}\\ \|K*p_Y-K*p_{0Y}\|_1^2\times e^{2\varrho\delta^{-r_1}}, & \text{ss case,}\end{cases}$$

so that the optimal choice for δ turns out to be

$$\delta=\begin{cases}O(\|K*p_Y-K*p_{0Y}\|_2^{1/(\beta_1+\eta)}), & \text{os case,}\\ O\left((-\log\|K*p_Y-K*p_{0Y}\|_1)^{-1/r_1}\right), & \text{ss case.}\end{cases}$$

For any $1 \le q \le \infty$,

$$\|K * p_Y - K * p_{0Y}\|_q = \|(K * F) * \phi_\sigma - K * p_{0Y}\|_q = \|p_{F*K,\sigma} - K * p_{0Y}\|_q.$$

Then,

$$\|p_Y - p_{0Y}\|_2 = \|p_{F,\sigma} - p_{0Y}\|_2 \lesssim \begin{cases} \|p_{F*K,\sigma} - K * p_{0Y}\|_2^{\beta_1/(\beta_1+\eta)}, & \text{os case,} \\ (-\log \|p_{F*K,\sigma} - K * p_{0Y}\|_1)^{-\beta_1/r_1}, & \text{ss case.} \end{cases}$$

For suitable constants τ_1, $\kappa_1 > 0$, let

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

Let v_n be as in the statement of Proposition 1. Then, for all $(F, \sigma) \in \mathcal{S}_n$, the following inclusions hold:

$$\begin{cases} \{(F, \sigma) : \|p_{F*K,\sigma} - K * p_{0Y}\|_2 \lesssim \psi_n\} \subseteq \{(F, \sigma) : \|p_Y - p_{0Y}\|_2 \lesssim v_n\}, & \text{os,} \\ \{(F, \sigma) : \|p_{F*K,\sigma} - K * p_{0Y}\|_1 \lesssim \psi_n\} \subseteq \{(F, \sigma) : \|p_Y - p_{0Y}\|_2 \lesssim v_n\}, & \text{ss.} \end{cases}$$

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

Appendix C: Proof of Theorem 3

For any intensity λ, we still denote $M_\lambda = \int_\Omega \lambda(t)\mathrm{d}t$ 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 γ only influences the prior on $\bar\lambda$ and has no impact on M_λ. As explained in Section 2, we can consider the following transformation: for all γ, $\gamma' \in \mathbb{R}_+^*$, we set, for any t,

$$\bar\lambda(t) = \sum_{j\ge1} p_j \frac{\mathbf{1}_{(0,\,\theta_j)}(t)}{\theta_j}, \qquad \psi_{\gamma,\gamma'}(\bar\lambda)(t) = \sum_{j\ge1} p_j \frac{\mathbf{1}_{(0,\,G_{\gamma'}^{-1}(G_\gamma(\theta_j)))}(t)}{G_{\gamma'}^{-1}(G_\gamma(\theta_j))},$$

with

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

So, if $\bar\lambda$ is distributed according to a DPM of uniform distributions with base measure indexed by γ, then $\psi_{\gamma,\gamma'}(\bar\lambda)$ is distributed according to a DPM of uniform distributions with base measure indexed by γ'. 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 $G_{\gamma'}^{-1}(G_\gamma(\theta)) = G^{-1}\left(G(\gamma\theta)G(\gamma' T)/G(\gamma T)\right)/\gamma'$ if $\theta \le T$ and $G_{\gamma'}^{-1}(G_\gamma(\theta)) = T$ if $\theta \ge T$. For any $\theta \in [0, T]$, if $\gamma' \ge \gamma$ then

$$G_{\gamma'}^{-1}(G_\gamma(\theta)) \le \theta \quad \text{and} \quad G_{\gamma'}^{-1}(G_\gamma(\theta)) \ge \frac{\gamma\theta}{\gamma'}. \tag{C.1}$$

The second inequality in (C.1) is straightforward. The first inequality in (C.1) is equivalent to $G(\gamma\theta)G(\gamma' T) \le G(\gamma'\theta)G(\gamma T)$ and is deduced from the following argument. Let $\Delta(\theta) = G(\gamma\theta)G(\gamma' T) - G(\gamma'\theta)G(\gamma T)$. 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 $\Delta'(\theta) = \gamma g(\gamma\theta)G(\gamma' T) - \gamma' g(\gamma'\theta)G(\gamma T)$ which is proportional to $\theta^{a-1}e^{-\gamma'\theta}[\gamma^a e^{(\gamma'-\gamma)\theta}G(\gamma' T) - (\gamma')^a G(\gamma T)]$. The function inside brackets is increasing so that $\Delta'(\theta) \le 0$ for $\theta \le c$ and $\Delta'(\theta) \ge 0$ for $\theta \ge c$. Therefore, Δ is first decreasing and then increasing. Since $\Delta(0) = \Delta(T) = 0$, Δ is negative on $(0, T)$, which achieves the proof of (C.1). For the second type of base measure, for $\theta \le T$, we have that, for every $\gamma, \gamma' > 0$, $G_{\gamma'}^{-1}(G_\gamma(\theta)) = T\gamma\theta/[\gamma'(T - \theta + \theta\gamma/\gamma')]$ and (C.1) is straightforward.

We first verify assumption [A1]. At several places, by using (4.1) and (4.4), we use that, under $\mathbb{P}_\lambda^{(n)}(\cdot \mid \Gamma_n)$, 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 $n(1+\alpha)m_2\lambda$ falling in I. Let $u_n = (n\log n)^{-1}$ so that $u_n = o(\bar\epsilon_n^2)$ and choose $k \ge 6$ so that $u_n^{-1} = o((n\bar\epsilon_n^2)^{k/2})$ 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 $\tilde B_n = \tilde B_n^\gamma$ (since in [A1] $\tilde B_n$ may depend on γ) as the set of $\lambda = M_\lambda \bar\lambda$ such that $|M_\lambda - M_{\lambda_0}| \le \bar\epsilon_n$ and $\psi_{\gamma,\gamma+u_n}(\bar\lambda) \in \bar B_n$, with $\bar B_n = \{\bar\lambda_{P'}(x) = \int_x^\infty \theta^{-1}\mathrm{d}P'(\theta) : P' \in \mathcal{N}\}$, 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 $\bar\lambda \in \bar B_n$ and $|M_\lambda - M_{\lambda_0}| \le \bar\epsilon_n$,

$$\mathbb{P}_{\lambda_0}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_0) \le -(\kappa_0+1)n\bar\epsilon_n^2 \mid \Gamma_n\right) = O((n\bar\epsilon_n^2)^{-k/2}(\log n)^k) \tag{C.2}$$

for κ_0 a constant. To prove (2.3), it is enough to control $\inf_{\gamma' \in [\gamma,\, \gamma+u_n]} \ell_n(M_\lambda \psi_{\gamma,\gamma'}(\bar\lambda))$. Using (C.1), we have that for any $\gamma' \in [\gamma, \gamma+u_n]$, on Γ_n,

$$G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j)) \le G_{\gamma'}^{-1}(G_\gamma(\theta_j))$$

and

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

Therefore,

$$\begin{aligned} \frac{\gamma}{\gamma+u_n}\psi_{\gamma,\gamma+u_n}(\bar\lambda)(t) &\le \psi_{\gamma,\gamma'}(\bar\lambda)(t) \\ &\le \psi_{\gamma,\gamma+u_n}(\bar\lambda)(t) + \sum_{j\ge 1} p_j \frac{\mathbf{1}_{\left(G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j)),\, \frac{\gamma+u_n}{\gamma}G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j))\right)}(t)}{G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j))} \end{aligned} \tag{C.3}$$

so that, for n large enough,

$$\begin{aligned}
\inf_{\gamma'\in[\gamma,\,\gamma+u_n]} \ell_n(M_\lambda\psi_{\gamma,\gamma'}(\bar\lambda)) &= \inf_{\gamma'\in[\gamma,\,\gamma+u_n]}\left\{\int_0^T \log(M_\lambda\psi_{\gamma,\gamma'}(\bar\lambda)(t))\mathrm{d}N_t - \int_0^T M_\lambda\psi_{\gamma,\gamma'}(\bar\lambda)(t)Y_t\mathrm{d}t\right\}\\
&\geq \int_0^T \log\left(M_\lambda\psi_{\gamma,\gamma+u_n}(\bar\lambda)(t)\right)\mathrm{d}N_t + N[0,\,T]\log\left(\frac{\gamma}{\gamma+u_n}\right)\\
&\qquad - M_\lambda\int_0^T \psi_{\gamma,\gamma+u_n}(\bar\lambda)(t)Y_t\mathrm{d}t - \frac{M_\lambda u_n n(1+\alpha)m_2}{\gamma}\\
&\geq \ell_n(M_\lambda\psi_{\gamma,\gamma+u_n}(\bar\lambda)) - \frac{u_n}{\gamma}[M_\lambda n(1+\alpha)m_2 + 2N[0,\,T]],
\end{aligned}$$

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

$$\mathbb{P}(|Z-n(1+\alpha)m_2M_{\lambda_0}|>n(1-\alpha)m_2M_{\lambda_0}) = o(1).$$

Then the event $\{N[0,T]\leq 2M_{\lambda_0}m_2n\}$ has probability going to 1 and, on this event,

$$\begin{aligned}
&\inf_{\gamma'\in[\gamma,\,\gamma+u_n]} \ell_n(M_\lambda\psi_{\gamma,\gamma'}(\bar\lambda))\\
&\qquad\geq \ell_n(M_\lambda\psi_{\gamma,\gamma+u_n}(\bar\lambda)) - n\gamma^{-1}m_2[M_\lambda(1+\alpha)+4M_{\lambda_0}]u_n.
\end{aligned}\tag{C.4}$$

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 $|M_\lambda - M_{\lambda_0}|\leq\bar\epsilon_n$,

$$\mathbb{P}^{(n)}_{\lambda_0}\left(\inf_{\gamma'\in[\gamma,\,\gamma+u_n]}\ell_n(M_\lambda\psi_{\gamma,\gamma'}(\bar\lambda)) - \ell_n(\lambda_0)\leq -(\kappa_0+2)n\bar\epsilon_n^2 \mid \Gamma_n\right) = O((n\bar\epsilon_n^2)^{-k/2}).$$

The left hand side of the previous 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 $\mathcal{K}'$ of $(0,\,\infty)$,

$$\inf_{\gamma\in\mathcal{K}'}\pi_1\left(\bar B_n\mid\gamma\right)\geq e^{-C_kn\bar\epsilon_n^2}\tag{C.5}$$

for some $C_k>0$, when n is large enough. By definition of $\tilde B_n^\gamma$, $\pi(\tilde B_n^\gamma\mid\gamma) = \pi_1(\bar B_n\mid\gamma+u_n)\pi_M([M_{\lambda_0}-\bar\epsilon_n,\,M_{\lambda_0}+\bar\epsilon_n])$ which, together with (C.5), implies that $\inf_{\gamma\in\mathcal{K}}\pi(B_n^\gamma\mid\gamma)\geq e^{-2C_kn\bar\epsilon_n^2}$ when n is large enough, so that (2.5) is satisfied as soon as j is large enough. We now define the measure $Q^{(n)}_{\lambda,\gamma}$, with

$$\mathrm{d}Q^{(n)}_{\lambda,\gamma} = \mathbf{1}_{\Gamma_n}\times\sup_{\gamma'\in[\gamma,\,\gamma+u_n]}\exp(\ell_n(M_\lambda\psi_{\gamma,\gamma'}(\bar\lambda)))\mathrm{d}\mu$$

and μ the measure such that under μ the process is an homogeneous Poisson process with intensity 1. Using (C.1) and similarly to (C.4), we obtain that, for all $\gamma'\in[\gamma,\,\gamma+u_n]$,

$$\bar\lambda(t) - \sum_{j\geq1}p_j\frac{\mathbf{1}_{(\gamma\theta_j/\gamma',\,\theta_j)}(t)}{\theta_j}\leq\psi_{\gamma,\gamma'}(\bar\lambda)(t)\leq\frac{\gamma+u_n}{\gamma}\bar\lambda(t)$$

and we have

$$
\begin{aligned}
Q_{\lambda,\gamma}^{(n)}(\mathcal{X}^{(n)}) = \mathbb{E}_\lambda^{(n)} \Bigg[\mathbf{1}_{\Gamma_n} \sup_{\gamma' \in [\gamma,\, \gamma+u_n]} \exp \Bigg(-M_\lambda \int_0^T \psi_{\gamma,\gamma'}(\bar\lambda)(t) Y_t \mathrm{d}t \\
+ \int_0^T \log\left(M_\lambda \psi_{\gamma,\gamma'}(\bar\lambda)(t)\right) \mathrm{d}N_t \Bigg) \Bigg] \\
\le \mathbb{E}_\lambda^{(n)} \left[\mathbf{1}_{\Gamma_n} \exp(nm_2(1+\alpha)M_\lambda \gamma^{-1} u_n + \log(1+u_n/\gamma) N[0,\, T])\right] \\
\le \mathbb{E}_\lambda^{(n)} \left[\mathbf{1}_{\Gamma_n} \exp(nm_2(1+\alpha)M_\lambda \gamma^{-1} u_n + u_n\gamma^{-1} N[0,\, T])\right] \\
\le \exp\left(nm_2(1+\alpha) M_\lambda \gamma^{-1} u_n + (1+\alpha) nm_2 M_\lambda (e^{u_n/\gamma} - 1)\right) \\
\le \exp\left(3nm_2(1+\alpha)\gamma^{-1} M_\lambda u_n\right)
\end{aligned}
$$

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

$$
\begin{aligned}
Q_{\lambda,\gamma}^{(n)}[1-\phi_{n,j}] &\le \mathbb{E}_\lambda^{(n)}[(1-\phi_{n,j}) \exp(nm_2(1+\alpha)M_\lambda\gamma^{-1}u_n + u_n\gamma^{-1}N[0,\, T])\mathbf{1}_{\Gamma_n}] \\
&\le e^{nm_2(1+\alpha)M_\lambda\gamma^{-1}u_n} (\mathbb{E}_\lambda^{(n)}\left[(1-\phi_{n,j})\mathbf{1}_{\Gamma_n}\right] \mathbb{E}_\lambda^{(n)}[e^{2\gamma^{-1}u_n N[0,\, T]}\mathbf{1}_{\Gamma_n}])^{1/2} \\
&\le e^{4nm_2(1+\alpha)\gamma^{-1}M_\lambda u_n} \max\{e^{-cnj^2\bar\epsilon_n^2/2},\, e^{-cnj\bar\epsilon_n/2}\}.
\end{aligned}
$$

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

$$
\begin{aligned}
\sup_{\gamma\in\mathcal{K}'} \int_{\mathbb{R}_+} \int_{\mathcal{S}_n^c} Q_{\lambda,\gamma}^{(n)}(\mathcal{X}^{(n)}) \mathrm{d}\pi_1(\bar\lambda \mid \gamma) \pi_M(M_\lambda) \mathrm{d}M_\lambda \\
\lesssim e^{-c_1(a+1)n\bar\epsilon_n^2} \int_{\mathbb{R}_+} e^{\delta M_\lambda} \pi_M(M_\lambda)\mathrm{d}M_\lambda \lesssim e^{-c_1(a+1)n\bar\epsilon_n^2},
\end{aligned}
$$

with δ 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

[1] Donnet, S., Rivoirard, V., Rousseau, J., and Scricciolo, C. (2015). Posterior concentration rates for counting processes with Aalen multiplicative intensities. *arXiv:1407.6033v1, to appear in Bayesian Analysis.*

[2] Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.*, 35(2):697–723.

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

[4] Salomond, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Statist.*, 8(1):1380–1404.

[5] Scricciolo, C. (2014). Adaptive Bayesian density estimation in L^p-metrics with Pitman-Yor or normalized inverse-Gaussian process kernel mixtures. *Bayesian Analysis*, 9:475–520.

[6] Shen, W., Tokdar, S., and Ghosal, S. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika*, 100:623–640.

[7] Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.*, 29:687–714.