Posterior concentration rates for empirical Bayes procedures, with applications to Dirichlet Process mixtures

Sophie Donnet, Vincent Rivoirard, Judith Rousseau, Catia Scricciolo

Posterior concentration rates for empirical Bayes procedures, with applications to Dirichlet Process mixtures

Sophie Donnet[*], Vincent Rivoirard[†], Judith Rousseau[‡] and Catia Scricciolo[§]

June 16, 2014

Abstract

In this paper we provide general conditions to check on the model and the prior to derive posterior concentration rates for data-dependent priors (or empirical Bayes approaches). We aim at providing conditions that are close to the conditions provided in the seminal paper by Ghosal and van der Vaart (2007a). We then apply the general theorem to two different settings: the estimation of a density using Dirichlet process mixtures of Gaussian random variables with base measure depending on some empirical quantities and the estimation of the intensity of a counting process under the Aalen model. A simulation study for inhomogeneous Poisson processes also illustrates our results. In the former case we also derive some results on the estimation of the mixing density and on the deconvolution problem. In the latter, we provide a general theorem on posterior concentration rates for counting processes with Aalen multiplicative intensity with priors not depending on the data.

Keywords: Empirical Bayes, posterior concentration rates, Dirichlet process mixtures, counting processes, Aalen model.

Short title Empirical Bayes posterior concentration rates.

1 Introduction

In a Bayesian approach to inference, the prior distribution should, in principle, be chosen independently of the data; however, it is not always an easy task to elicit the values of the prior hyperparameters and a common practice is to estimate them by reasonable empirical quantities. The prior is then data-dependent and the approach falls under the umbrella of empirical Bayes methods, as opposed to fully Bayesian methods. More formally, consider a statistical model $(\mathbb{P}_\theta^{(n)},\, \theta \in \Theta)$ over a sample space $\mathcal{X}^{(n)}$, together with a family of prior distributions $\pi(\cdot \mid \gamma)$, $\gamma \in \Gamma$, on the parameter space Θ. A Bayesian statistician would either set γ to a specific value or integrate it out using a probability distribution in a hierarchical specification of the prior for θ. Both approaches would lead to a prior distribution for θ that does not depend on the data, say π, resulting in a posterior distribution $\pi(\cdot \mid X^{(n)})$. However, it is often the case that knowledge is not *a priori* available to either fix a value for γ or elicit a prior distribution for it, so that these hyperparameters are more easily estimated from the data. Throughout the paper, we will denote by $\hat{\gamma}_n$ a data-driven selection of the prior hyperparameters. There are many instances in the literature where empirical Bayes selection of the prior hyperparameters is performed, sometimes

[*]Université de Paris Dauphine, France
[†]Université de Paris Dauphine, France
[‡]CREST, France
[§]Bocconi University, Italy

without explicitly mentioning it. Some examples concerning the parametric case can be found in Casella (1985), Berger (1985) and Ahmed and Reid (2001). Regarding the nonparametric case, Richardson and Green (1997) propose a default empirical Bayes approach to deal with parametric and nonparametric mixtures of Gaussian random variables; McAuliffe et al. (2006) propose another empirical Bayes approach for Dirichlet process mixtures of Gaussian random variables, while in Knapik et al. (2012) and Szabó et al. (2013) an empirical Bayes procedure is proposed in the context of the white noise model to obtain adaptive posterior distributions. There are many other instances of empirical Bayes methods in the literature, especially in applied problems. Quoting Hjort et al. (2010), "Efron (2003) argues that the brightest statistical future may be reserved for *empirical Bayes* methods".

In this paper, our aim is not to claim that empirical Bayes methods are better than fully Bayesian methods in some way, but rather to provide tools to study frequentist asymptotic properties of empirical Bayes posterior distributions, given their wide use in practice. Surprisingly, very little is known, in a general framework, on the asymptotic behavior of such empirical Bayes posterior distributions. It is common belief that, if $\hat{\gamma}_n$ asymptotically converges to some value γ^*, then the empirical Bayes posterior distribution associated with $\hat{\gamma}_n$ is eventually close to the Bayesian posterior associated with γ^*. Results have been obtained in explicit specific frameworks in Knapik et al. (2012) and Szabó et al. (2013) for the white noise model; in Clyde and George (2000) and Cui and George (2008) for wavelets or variable selection. Recently, Petrone et al. (2014) have investigated asymptotic properties of empirical Bayes posterior distributions: they have obtained general conditions for frequentist consistency of empirical Bayes posteriors and, in the parametric case, studied when strong merging between Bayesian and maximum marginal likelihood empirical Bayes posterior distributions takes place.

In this work, we are interested in studying the frequentist asymptotic behavior of empirical Bayes posterior distributions $\pi(\cdot \mid X^{(n)}, \hat{\gamma}_n)$ in terms of contraction rates. Let $d(\cdot, \cdot)$ be a loss function, say a metric or a pseudo-metric, on Θ and, for $\theta_0 \in \Theta$, let U_ϵ be a neighborhood of θ_0, *i.e.*, a set of the form $\{\theta : d(\theta, \theta_0) < \epsilon\}$ with $\epsilon > 0$. The empirical Bayes posterior distribution is said to concentrate at θ_0, with rate ϵ_n relative to d, where ϵ_n is a positive sequence converging to 0, if the empirical Bayes posterior probability of U_{ϵ_n} tends to 1 in $\mathbb{P}_{\theta_0}^{(n)}$-probability. In the case of fully Bayesian procedures, there has been so far a vast literature on posterior consistency and contraction rates since the seminal articles of Barron et al. (1999) and Ghosal et al. (2000). Following ideas of Schwartz (1965), Ghosal et al. (2000) in the case of independent and identically distributed (iid) observations and Ghosal and van der Vaart (2007a) in the case of non-iid observations have developed an elegant and powerful methodology to assess posterior contraction rates which boils down to lower bound the prior mass of Kullback-Leibler type neighborhoods of $\mathbb{P}_{\theta_0}^{(n)}$ and to construct exponentially powerful tests for the testing problem $H_0 : \theta = \theta_0$ versus $H_1 : \theta \in \{\theta : d(\theta, \theta_0) > \epsilon_n\}$. However, this approach cannot be taken to deal with posterior distributions corresponding to data-dependent priors. Therefore, in this paper, we develop a similar methodology for deriving posterior contraction rates in the case where the prior distribution depends on the data through a data-driven choice $\hat{\gamma}_n$ of the hyperparameters.

In Section 2, we provide a general theorem on posterior contraction rates for empirical Bayes posterior distributions in the spirit of those presented in Ghosal and van der Vaart (2007a), which is then applied to nonparametric mixture models. Two main applications are considered: Dirichlet mixtures of Gaussian distributions for the problem of density estimation in Section 3 and Dirichlet mixtures of uniform random variables for estimating the intensity function of counting processes obeying the Aalen model in Section 4. Dirichlet process mixtures have been introduced by Ferguson (1974) and have proved to be a major tool in Bayesian nonparametrics, see for instance Hjort et al. (2010).

Rates of convergence for fully Bayesian posterior distributions corresponding to Dirichlet process mixtures of Gaussian distributions have been widely studied and it has been proved that they lead to minimax-optimal procedures over a wide collection of density functional classes, see Ghosal and van der Vaart (2001), Ghosal and van der Vaart (2007b), Kruijer et al. (2010), Shen et al. (2013) and Scricciolo (2014). Here, we extend existing results to the case of a Gaussian base measure with data-dependent mean and variance, as advocated for instance by Richardson and Green (1997). Furthermore, due to some inversion inequalities, we get, as a by-product, empirical Bayes posterior recovery rates for the problem of density deconvolution when the errors are ordinary or super-smooth and the mixing density is modeled as a Dirichlet process mixture of normal densities with a Gaussian base measure having data-driven chosen mean and variance. The problem of Bayesian density deconvolution when the mixing density is modeled as a Dirichlet process mixture of Gaussian densities and the errors are super-smooth has been recently studied by Sarkar et al. (2013).

In Section 4, we focus on Aalen multiplicative intensity models, which constitutes a major class of counting processes that have been extensively used in the analysis of data arising from various fields like medicine, biology, finance, insurance and social sciences. General statistical and probabilistic literature on such processes is very huge and we refer the reader to Karr (1991), Andersen et al. (1993), Daley and Vere-Jones (2003) and Daley and Vere-Jones (2008) for a good introduction. In the specific setting of nonparametric Bayesian statistics, practical or methodological contributions have been obtained by Lo (1992), Adams et al. (2009) or Cheng and Yuan (2013). First quite general theoretical results have been obtained by Belitser et al. (2013) who established the frequentist asymptotic behavior of posterior distributions for intensities of Poisson processes. We extend their results by considering Aalen multiplicative intensity models instead of simple Poisson models (see Theorem 5). In Theorem 4, we derive rates of convergence for empirical Bayes estimation of monotone non-increasing intensity functions in counting processes satisfying the Aalen multiplicative intensity model using Dirichlet process mixtures of uniform distributions with a truncated gamma base measure whose scale parameter is estimated from the data. Numerical illustrations are also presented, in this context.

Proofs and technical derivations are postponed to Section 5. Instrumental and auxiliary results are reported in the Appendix in Section 6.

Notations and context. Let $X^{(n)} \in \mathcal{X}^{(n)}$ be the observations with $(\mathcal{X}^{(n)}, \mathcal{A}_n, \mathbb{P}_\theta^{(n)}, \theta \in \Theta)$ a sequence of experiments, where $\mathcal{X}^{(n)}$ and Θ are Polish spaces endowed with their Borel σ-fields $\mathcal{A}_n$ and $\mathcal{B}$ respectively. We assume that there exists a σ-finite measure $\mu^{(n)}$ on $\mathcal{X}^{(n)}$ (for convenience of notation, we suppress dependence on n in $\mu^{(n)}$ in what follows) dominating all probability measures $\mathbb{P}_\theta^{(n)}$, $\theta \in \Theta$. For each $\theta \in \Theta$, we denote $\ell_n(\theta)$ the associated log-likelihood. We also denote by $\mathbb{E}_\theta^{(n)}$ the expectation with respect to $\mathbb{P}_\theta^{(n)}$. We consider a family of prior distributions $(\pi(\cdot \mid \gamma))_{\gamma \in \Gamma}$ on Θ, where $\Gamma \subset \mathbb{R}^d$ with $d \geq 1$, and we denote by $\pi(\cdot \mid X^{(n)}, \gamma)$ the posterior corresponding to the prior $\pi(\cdot \mid \gamma)$ which is given by

$$\pi(B \mid X^{(n)}, \gamma) = \frac{\int_B e^{\ell_n(\theta)} \mathrm{d}\pi(\theta \mid \gamma)}{\int_\Theta e^{\ell_n(\theta)} \mathrm{d}\pi(\theta \mid \gamma)}, \quad B \in \mathcal{B}.$$

We denote by $\mathrm{KL}(\cdot;\, \cdot)$ the Kullback-Leibler divergence. Given $\theta_1, \theta_2 \in \Theta_n$, we denote by $V_k(\theta_1;\, \theta_2)$ the recentered k-th moment of the log-likelihood difference associated to θ_1 and θ_2. So, we have:

$$\mathrm{KL}(\theta_1;\, \theta_2) = \mathbb{E}_{\theta_1}^{(n)}[\ell_n(\theta_1) - \ell_n(\theta_2)],$$

$$V_k(\theta_1;\, \theta_2) = \mathbb{E}_{\theta_1}^{(n)}[|\ell_n(\theta_1) - \ell_n(\theta_2) - \mathbb{E}_{\theta_1}^{(n)}[\ell_n(\theta_1) - \ell_n(\theta_2)]|^k].$$

We denote by $h(f_1, f_2)$ the Hellinger metric between two densities f_1 and f_2, *i.e.*, $h^2(f_1,\, f_2) = \int(\sqrt{f_1}(x) - \sqrt{f_2}(x))^2 \mathrm{d}x$. Throughout the text, we denote by $D(\zeta,\, B,\, d(\cdot,\, \cdot))$ the ζ-covering number of B by d-balls of radius ζ, for any set B, any positive constant ζ and any pseudo-metric $d(\cdot,\, \cdot)$. We denote θ_0 the true parameter.

2 General result on posterior concentration rates for empirical Bayes

Let $\hat{\gamma}_n$ be a measurable function of the observations. The associated empirical Bayes posterior distribution is then $\pi(\cdot \mid X^{(n)},\, \hat{\gamma}_n)$. In this section, we present a general theorem to obtain posterior concentration rates for the empirical Bayes posterior $\pi(\cdot \mid X^{(n)},\, \hat{\gamma}_n)$. Our aim is to give conditions resembling those usually considered in fully Bayesian approaches. For this purpose, we first define the usual quantities. We assume that, with probability going to 1, $\hat{\gamma}_n$ belongs to a subset $\mathcal{K}_n$ of Γ:

$$\mathbb{P}_{\theta_0}^{(n)}\left(\hat{\gamma}_n \in \mathcal{K}_n^c\right) = o(1). \tag{2.1}$$

For any positive sequence $(u_n)_n$, let $N_n(u_n)$ stand for the u_n-covering number of $\mathcal{K}_n$ relative to the Euclidean distance which is denoted by $\|\cdot\|$. For instance, if $\mathcal{K}_n$ is included in a ball of radius R_n then $N_n(u_n) = O((R_n/u_n)^d)$.

As in Ghosal and van der Vaart (2007a), for any $\epsilon_n > 0$, we introduce the ϵ_n-Kullback-Leibler neighborhood of θ_0 defined, for $k \geq 1$, by

$$\bar{B}_{k,n} = \{\theta :\, \mathrm{KL}(\theta_0;\, \theta) \leq n\epsilon_n^2, V_k(\theta_0;\, \theta) \leq n^{k/2}\epsilon_n^k\}.$$

Let $d(\cdot,\, \cdot)$ be a loss function on Θ. We consider the posterior concentration rates in terms of $d(\cdot,\, \cdot)$ using the following neighborhoods:

$$U_{M\epsilon_n} = \{\theta \in \Theta :\, d(\theta_0,\, \theta) \leq M\epsilon_n\}$$

with $M > 0$. For any integer j, we define

$$S_{n,j} = \{\theta \in \Theta :\, d(\theta_0,\, \theta) \in (j\epsilon_n,\, (j+1)\epsilon_n]\}.$$

In order to obtain posterior concentration rates with data-dependent priors, we express the impact of $\hat{\gamma}_n$ on the prior in the following way: for all $\gamma,\, \gamma' \in \Gamma$, we construct a measurable transformation $\psi_{\gamma,\gamma'} : \Theta \to \Theta$ such that if $\theta \sim \pi(\cdot \mid \gamma)$ then $\psi_{\gamma,\gamma'}(\theta) \sim \pi(\cdot \mid \gamma')$. Let $e_n(\cdot,\, \cdot)$ be another semi-metric on $\Theta \times \Theta$. We consider the following set of assumptions. Let $\Theta_n \subset \Theta$ and k be fixed.

[A1] There exist a sequence $(u_n)_n$ and $\tilde{B}_n \subset \bar{B}_{k,n}$ such that

$$N_n(u_n) = o((n\epsilon_n^2)^{k/2}) \tag{2.2}$$

and

$$\sup_{\gamma \in \mathcal{K}_n} \sup_{\theta \in \tilde{B}_n} \mathbb{P}_{\theta_0}^{(n)} \left\{ \inf_{\gamma' :\, \|\gamma'-\gamma\| \leq u_n} \ell_n(\psi_{\gamma,\gamma'}(\theta)) - \ell_n(\theta_0) < -n\epsilon_n^2 \right\} = o(N_n(u_n)^{-1}). \tag{2.3}$$

[A2] Defining for all $\gamma \in \mathcal{K}_n$, $\mathrm{d}Q_{\gamma,n}^{\theta}(x^{(n)}) = \sup_{\|\gamma'-\gamma\| \leq u_n} e^{\ell_n(\psi_{\gamma,\gamma'}(\theta))(x^{(n)})} \mathrm{d}\mu(x^{(n)})$, we have

$$\sup_{\gamma \in \mathcal{K}_n} \frac{\int_{\Theta \setminus \Theta_n} Q_{\gamma,n}^{\theta}(\mathcal{X}^{(n)}) \mathrm{d}\pi(\theta \mid \gamma)}{\pi(\tilde{B}_n \mid \gamma)} = o(N_n(u_n)^{-1} e^{-2n\epsilon_n^2}) \tag{2.4}$$

and there exist constants $\zeta,\, K > 0$ such that

- for all j large enough,

$$\sup_{\gamma \in \mathcal{K}_n} \frac{\pi(S_{n,j} \cap \Theta_n \mid \gamma)}{\pi(\tilde{B}_n \mid \gamma)} \leq e^{Knj^2\epsilon_n^2/2}, \tag{2.5}$$

- for all $\epsilon > 0$, for all $\theta \in \Theta_n$ with $d(\theta_0, \theta) > \epsilon$ there exist tests $\phi_n(\theta)$ satisfying

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_n(\theta)] \leq e^{-Kn\epsilon^2}, \quad \sup_{\gamma \in \mathcal{K}_n} \sup_{e_n(\theta', \theta) \leq \zeta\epsilon} \int_{\mathcal{X}^{(n)}} [1 - \phi_n(\theta)] \mathrm{d}Q_{\gamma,n}^{\theta'} \leq e^{-Kn\epsilon^2}, \tag{2.6}$$

- for all j large enough,

$$\log D(\zeta j \epsilon_n, S_{n,j}, e_n(\cdot, \cdot)) \leq K(j+1)^2 n\epsilon_n^2/2. \tag{2.7}$$

We then have the following theorem.

Theorem 1 *Let $\theta_0 \in \Theta$. Assume that the estimator $\hat{\gamma}_n$ satisfies* (2.1) *and the prior satisfies assumptions [A1] and [A2], with $n\epsilon_n^2 \to \infty$ and $\epsilon_n \to 0$. Then, for J_1 large enough,*

$$\mathbb{E}_{\theta_0}^{(n)}[\pi(U_{J_1\epsilon_n}^c \mid X^{(n)}, \hat{\gamma}_n)] = o(1),$$

where $U_{J_1\epsilon_n}^c$ is the complementary of $U_{J_1\epsilon_n}$ in Θ.

The proof of Theorem 1 shows that the posterior concentration rate of the empirical Bayes posterior based on $\hat{\gamma}_n \in \mathcal{K}_n$ is bounded from above by the worst concentration rate over the classes of posterior distributions corresponding to the priors $(\pi(\cdot \mid \gamma), \gamma \in \mathcal{K}_n)$. In particular, assume that each posterior $\pi(\cdot \mid X^{(n)}, \gamma)$ converges at rate $\epsilon_n(\gamma) = (n/\log n)^{-\alpha(\gamma)}$, where $\gamma \to \alpha(\gamma)$ is Lipschitzian, and $\hat{\gamma}_n$ converges to γ^* at rate v_n, with $v_n = o((\log n)^{-1})$ namely $\mathcal{K}_n = [\gamma^* - v_n, \gamma^* + v_n]$, then the posterior concentration rate of the empirical Bayes posterior is of the order $O(\epsilon_n(\gamma^*))$. Indeed, in this case, $\sup_{\gamma \in \mathcal{K}_n}\{\epsilon_n(\gamma)/\epsilon_n(\gamma^*)\} = \sup_{\gamma \in \mathcal{K}_n}(n/\log n)^{|\alpha(\gamma) - \alpha(\gamma^*|)} = O(1)$. This is of special interest when $\epsilon_n(\gamma^*)$ is optimal. Proving that the posterior distribution has optimal posterior concentration rate then boils down to proving that $\hat{\gamma}_n$ converges to the oracle value γ^*.

Remark 1 *As in Ghosal and van der Vaart (2007a), we can replace conditions* (2.6) *and* (2.7) *by the existence of a global test ϕ_n over $S_{n,j}$ similar to equation (2.7) of Ghosal and van der Vaart (2007a) satisfying*

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_n] = o(N_n(u_n)^{-1}), \quad \sup_{\gamma \in \mathcal{K}_n} \sup_{\theta \in S_{n,j}} \int_{\mathcal{X}^{(n)}} (1 - \phi_n) \mathrm{d}Q_{\gamma,n}^{\theta} \leq e^{-Knj^2\epsilon_n^2}$$

without modifying the conclusion. Note also that when the loss function $d(\theta, \theta_0)$ is not bounded it is often the case that obtaining exponential control on the error rates in the form $e^{-Kn\epsilon^2}$ or $e^{-Knj^2\epsilon_n^2}$ is not possible for large values of j. It is enough in that case to consider a modification $\tilde{d}(\theta, \theta_0)$ of the loss which affects only the values of θ for which $d(\theta, \theta_0)$ is large and to prove (2.6) *and* (2.7) *for $\tilde{d}(\theta, \theta_0)$, defining $S_{n,j}$ and the covering number $D(\cdot)$ with respect to $\tilde{d}(\cdot, \cdot)$. As an illustration of this remark, see the proof of Theorem 4.*

The assumptions of [A2] are very similar to those for establishing posterior concentration rates proposed for instance by Ghosal and van der Vaart (2007a) (see their Theorem 1 and the associated proof). We need to strengthen some conditions to take into account that we only know that $\hat{\gamma}_n$ lies in the compact set $\mathcal{K}_n$ with high probability.

The key idea here is to construct a transformation $\psi_{\gamma,\gamma'}$ which allows to transfer the dependence on the data from the prior to the likelihood, similarly to what is considered in Petrone et al. (2014). The only difference with the general theorems of Ghosal and van der Vaart (2007a) lies in the control of the log-likelihood difference $\ell_n(\psi_{\gamma,\gamma'}(\theta)) - \ell_n(\theta_0)$ when $\|\gamma - \gamma'\| \le u_n$. In nonparametric cases where $n\epsilon_n^2$ is a power of n, u_n can be chosen very small as soon as k can be chosen large enough so that controlling this difference uniformly is not such a drastic condition. In parametric models, where at best $n\epsilon_n^2$ is a power of $\log n$, this become more involved and u_n needs to be large or $\mathcal{K}_n$ needs to be small enough so that $N_n(u_n)$ can be chosen of order $O(1)$. Note that in parametric models it is typically easier to use a more direct control of $\pi(\theta \mid \gamma)/\pi(\theta \mid \gamma')$, the ratio of the prior densities with respect to a common dominating measure. In nonparametric prior models this is usually not possible since no such dominating measure exists in most cases. In Sections 3 and 4, we apply Theorem 1 to two different types of Dirichlet process mixture models: Dirichlet process mixtures of Gaussian distributions used to model smooth densities and Dirichlet process mixture model of uniforms to model monotone non-increasing intensities in the context of Aalen point processes. It is interesting to note that in the case of general nonparametric mixture models there exists a general construction of $\psi_{\gamma,\gamma'}$. More precisely, consider a mixture model in the form

$$f(\cdot) = \sum_{j=1}^{K} p_j h_{\theta_j}(\cdot), \quad K \sim \pi_K, \tag{2.8}$$

and, conditionally on K, $p = (p_j)_{j=1}^K \sim \pi_p$ and $\theta_1, \ldots, \theta_K$ are iid with cdf G_γ. The Dirichlet process mixture corresponds to $\pi_K = \delta_{(+\infty)}$ and π_p is the GEM distribution obtained from the stick-breaking construction, see for instance Ghosh and Ramamoorthi (2003). Models in the form (2.8) also cover priors on curves if the $(p_j)_{j=1}^K$ are not restricted to the simplex. Denote by $\pi(\cdot \mid \gamma)$ the prior probability on f induced by (2.8). Then, for all $\gamma, \gamma' \in \Gamma$, if f is represented as (2.8) and is thus distributed according to $\pi(\cdot \mid \gamma)$, then

$$f'(\cdot) = \sum_{j=1}^{K} p_j h_{\theta_j'}(\cdot), \qquad \text{with} \quad \theta_j' = G_{\gamma'}^{-1}(G_\gamma(\theta_j)),$$

is distributed according to $\pi(\cdot \mid \gamma')$, where $G_{\gamma'}^{-1}(\cdot)$ denotes the generalized inverse of the cumulative distribution function.

We now give the proof of Theorem 1.

Proof. We consider a chaining argument and we split $\mathcal{K}_n$ into $N_n(u_n)$ balls of radius u_n and we denote by $(\gamma_i)_{i=1,\ldots,N_n(u_n)}$ the centers of these balls. We have

$$\mathbb{E}_{\theta_0}^{(n)}[\pi(U_{J_1\epsilon_n}^c \mid X^{(n)}, \hat{\gamma}_n)] \le \mathbb{E}_{\theta_0}^{(n)}[\max_i \rho_n(\gamma_i)] \le N_n(u_n) \max_{i=1,\ldots,N_n(u_n)} \mathbb{E}_{\theta_0}^{(n)}[\rho_n(\gamma_i)], \tag{2.9}$$

with

$$\rho_n(\gamma_i) := \sup_{\|\gamma-\gamma_i\|\le u_n} \pi(U_{J_1\epsilon_n}^c \mid X^{(n)}, \gamma) = \sup_{\|\gamma-\gamma_i\|\le u_n} \frac{\int_{U_{J_1\epsilon_n}^c} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta))-\ell_n(\theta_0)} \mathrm{d}\pi(\theta \mid \gamma_i)}{\int_\Theta e^{\ell_n(\psi(\gamma_i,\gamma)(\theta))-\ell_n(\theta_0)} \mathrm{d}\pi(\theta \mid \gamma_i)}.$$

We now study, for any i, $\mathbb{E}_{\theta_0}^{(n)}[\rho_n(\gamma_i)]$. We mimic the proof of Lemma 9 of Ghosal and van der Vaart (2007a). For every j large enough, by (2.7), there exist $L_{j,n}$ balls of centers $\theta_{j,1}, \ldots, \theta_{j,L_{j,n}}$, with radius $\zeta j\epsilon_n$ relative to the e_n-distance, covering $S_{n,j} \cap \Theta_n$, with

$L_{j,n} \leq \exp(K(j+1)^2 n\epsilon_n^2/2)$. We then consider tests $\phi_n(\theta_{j,\ell})$ satisfying (2.6) with $\epsilon = j\epsilon_n$. By setting

$$\phi_n = \max_{j \geq J_1} \max_{\ell \in \{1,\dots,L_{j,n}\}} \phi_n(\theta_{j,\ell}),$$

we obtain

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_n] \leq \sum_{j \geq J_1} L_{j,n} e^{-Knj^2\epsilon_n^2} = O(e^{-KJ_1^2 n\epsilon_n^2/2}). \tag{2.10}$$

Moreover, for any $j \geq J_1$, any $\theta \in S_{n,j} \cap \Theta_n$ and any i,

$$\int_{\mathcal{X}^{(n)}} (1-\phi_n) \mathrm{d}Q_{\gamma_i,n}^{\theta} \leq e^{-Knj^2\epsilon_n^2}. \tag{2.11}$$

Since for all i we have $\rho_n(\gamma_i) \leq 1$, using the usual decomposition,

$$\mathbb{E}_{\theta_0}^{(n)}[\rho_n(\gamma_i)] \leq \mathbb{E}_{\theta_0}^{(n)}[\phi_n] + \mathbb{P}_{\theta_0}^{(n)}(A_{n,i}^c) + \frac{e^{2n\epsilon_n^2}}{\pi(\tilde{B}_n \mid \gamma_i)} C_{n,i}, \tag{2.12}$$

with

$$A_{n,i} = \left\{ \inf_{\|\gamma-\gamma_i\| \leq u_n} \int_{\Theta} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \mathrm{d}\pi(\theta \mid \gamma_i) \geq e^{-2n\epsilon_n^2} \pi(\tilde{B}_n \mid \gamma_i) \right\}$$

and

$$C_{n,i} = \mathbb{E}_{\theta_0}^{(n)} \left[(1-\phi_n) \sup_{\|\gamma-\gamma_i\| \leq u_n} \int_{U_{J_1\epsilon_n}^c} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \mathrm{d}\pi(\theta \mid \gamma_i) \right].$$

We study each term of (2.12). First, by using (2.2) and (2.10),

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_n] = o(N_n(u_n)^{-1}).$$

Secondly, since

$$e^{n\epsilon_n^2} \inf_{\|\gamma-\gamma_i\| \leq u_n} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \geq 1_{\{\inf_{\|\gamma-\gamma_i\| \leq u_n} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \geq e^{-n\epsilon_n^2}\}},$$

we have

$$\begin{aligned}
\mathbb{P}_{\theta_0}^{(n)}(A_{n,i}^c) &\leq \mathbb{P}_{\theta_0}^{(n)} \left\{ \int_{\tilde{B}_n} \inf_{\|\gamma-\gamma_i\| \leq u_n} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \frac{\mathrm{d}\pi(\theta \mid \gamma_i)}{\pi(\tilde{B}_n \mid \gamma_i)} < e^{-2n\epsilon_n^2} \right\} \\
&\leq \mathbb{P}_{\theta_0}^{(n)} \left\{ \int_{\tilde{B}_n} 1_{\{\inf_{\|\gamma-\gamma_i\| \leq u_n} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \geq e^{-n\epsilon_n^2}\}} \frac{\mathrm{d}\pi(\theta \mid \gamma_i)}{\pi(\tilde{B}_n \mid \gamma_i)} < e^{-n\epsilon_n^2} \right\} \\
&\leq \mathbb{P}_{\theta_0}^{(n)} \left\{ \int_{\tilde{B}_n} 1_{\{\inf_{\|\gamma-\gamma_i\| \leq u_n} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} < e^{-n\epsilon_n^2}\}} \frac{\mathrm{d}\pi(\theta \mid \gamma_i)}{\pi(\tilde{B}_n \mid \gamma_i)} > 1 - e^{-n\epsilon_n^2} \right\} \\
&\leq (1-e^{-n\epsilon_n^2})^{-1} \int_{\tilde{B}_n} \mathbb{P}_{\theta_0}^{(n)} \left\{ \inf_{\|\gamma-\gamma_i\| \leq u_n} \ell_n(\psi_{\gamma_i,\gamma}(\theta)) - \ell_n(\theta_0) \leq -n\epsilon_n^2 \right\} \frac{\mathrm{d}\pi(\theta \mid \gamma_i)}{\pi(\tilde{B}_n \mid \gamma_i)} \\
&= o(N_n(u_n)^{-1}),
\end{aligned}$$

by (2.3). Also, we have

$$\begin{aligned}
C_{n,i} &\leq \mathbb{E}_{\theta_0}^{(n)} \left[(1-\phi_n) \int_{U_{J_1\epsilon_n}^c} \sup_{\|\gamma-\gamma_i\| \leq u_n} e^{\ell_n(\psi(\gamma_i,\gamma)(\theta)) - \ell_n(\theta_0)} \mathrm{d}\pi(\theta \mid \gamma_i) \right] \\
&\leq \int_{U_{J_1\epsilon_n}^c} \int_{\mathcal{X}^{(n)}} (1-\phi_n) \mathrm{d}Q_{\gamma_i,n}^{\theta} \mathrm{d}\pi(\theta \mid \gamma_i) \\
&\leq \int_{\Theta \setminus \Theta_n} Q_{\gamma_i,n}^{\theta}(\mathcal{X}^{(n)}) \mathrm{d}\pi(\theta \mid \gamma_i) + \sum_{j \geq J_1} \int_{S_{n,j} \cap \Theta_n} \int_{\mathcal{X}^{(n)}} (1-\phi_n) \mathrm{d}Q_{\gamma_i,n}^{\theta} \mathrm{d}\pi(\theta \mid \gamma_i).
\end{aligned}$$

By using (2.4), (2.5) and (2.11),

$$
\begin{aligned}
C_{n,i} &\leq \sum_{j\geq J_1} e^{-Knj^2\epsilon_n^2}\pi(S_{n,j}\cap\Theta_n \mid \gamma_i) + o(N_n(u_n)^{-1}e^{-2n\epsilon_n^2}\pi(\tilde{B}_n \mid \gamma_i))\\
&\leq \sum_{j\geq J_1} e^{-Knj^2\epsilon_n^2/2}\pi(\tilde{B}_n \mid \gamma_i) + o(N_n(u_n)^{-1}e^{-2n\epsilon_n^2}\pi(\tilde{B}_n \mid \gamma_i))\\
&= o(N_n(u_n)^{-1}e^{-2n\epsilon_n^2}\pi(\tilde{B}_n \mid \gamma_i))
\end{aligned}
$$

and

$$
\max_{i=1,\,\ldots,\,N_n(u_n)} \frac{e^{2n\epsilon_n^2}}{\pi(\tilde{B}_n \mid \gamma_i)} C_{n,i} = o(N_n(u_n)^{-1}).
$$

Having controlled each term in (2.12), by (2.9), the proof of Theorem 1 is achieved. □

3 Adaptive posterior contraction rates for empirical Bayes Dirichlet process mixtures of Gaussian densities

Let $X^{(n)} = (X_1,\,\ldots,\,X_n)$ be n iid observations from an unknown Lebesgue density p on $\mathbb{R}$. Consider the following prior distribution on the class of densities $\mathcal{P} = \{p : \mathbb{R} \to \mathbb{R}_+ \mid \|p\|_1 = 1\}$:

$$
\begin{gathered}
p(\cdot) = p_{F,\sigma}(\cdot) := \int_{-\infty}^{\infty} \phi_\sigma(\cdot - \theta)\,\mathrm{d}F(\theta),\\
F \sim \mathrm{DP}(\alpha_{\mathbb{R}}\mathrm{N}_{(m,s^2)}), \quad \sigma \sim \mathrm{IG}(\nu_1,\,\nu_2), \quad \nu_1,\,\nu_2 > 0,
\end{gathered}
\tag{3.1}
$$

where $\alpha_{\mathbb{R}}$ is a positive constant, $\phi_\sigma(\cdot) = \sigma^{-1}\phi(\cdot/\sigma)$, with $\phi(\cdot)$ the density of a standard Gaussian distribution, and $\mathrm{N}_{(m,s^2)}$ denotes the Gaussian distribution with mean m and variance s^2. Set $\gamma = (m,\,s^2) \in \Gamma \subseteq \mathbb{R}\times\mathbb{R}_+^*$, let $\hat{\gamma}_n : \mathbb{R}^n \to \Gamma$ be some measurable function of the observations. Typical choices are $\hat{\gamma}_n = (\bar{X}_n,\,S_n^2)$, with the sample mean $\bar{X}_n = \sum_{i=1}^n X_i/n$ and the sample variance $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/n$, and $\hat{\gamma}_n = (\bar{X}_n,\,R_n)$, with the range $R_n = \max_i X_i - \min_i X_i$ as in Richardson and Green (1997). Let $\mathcal{K}_n \subset \mathbb{R}\times\mathbb{R}_+^*$ be compact, independent of the data $X^{(n)}$ and such that

$$
\mathbb{P}_{p_0}^{(n)}(\hat{\gamma}_n \in \mathcal{K}_n) = 1 + o(1). \tag{3.2}
$$

Throughout this section, we assume that the true density p_0 satisfies the following tail condition:

$$
p_0(x) \lesssim e^{-c_0|x|^\tau} \quad \text{for } |x| \text{ large enough}, \tag{3.3}
$$

with finite constants $c_0,\,\tau > 0$. Let $\mathbb{E}_{p_0}[X_1] = m_0 \in \mathbb{R}$ and $\mathbb{V}\mathrm{ar}_{p_0}[X_1] = \tau_0^2 \in \mathbb{R}_+^*$. If $\hat{\gamma}_n = (\bar{X}_n,\,S_n^2)$, then condition (3.2) is satisfied with $\mathcal{K}_n = [m_0 - (\log n)/\sqrt{n},\,m_0 + (\log n)/\sqrt{n}] \times [\tau_0^2 - (\log n)/\sqrt{n},\,\tau_0^2 + (\log n)/\sqrt{n}]$, while, if $\hat{\gamma}_n = (\bar{X}_n,\,R_n)$, then $\mathcal{K}_n = [m_0 - (\log n)/\sqrt{n},\,m_0 + (\log n)/\sqrt{n}] \times [a,\,2(2c_0^{-1}\log n)^{1/\tau}]$.

Mixtures of Gaussian densities have been extensively used and studied in Bayesian nonparametric literature. Posterior contraction rates have been first investigated by Ghosal and van der Vaart (2001) and Ghosal and van der Vaart (2007b). Following an idea of Rousseau (2010), Kruijer et al. (2010) have proved that nonparametric location mixtures of Gaussian densities lead to adaptive posterior contraction rates over the full scale of locally Hölder log-densities on $\mathbb{R}$. This result has been extended to the multivariate set-up by Shen et al. (2013) and to supersmooth densities by Scricciolo (2014). The key idea behind these results is that, for an ordinary β-smooth density p_0, given $\sigma > 0$ small enough, there exists a finite mixing distribution F^*, with $N_\sigma = O(\sigma^{-1}|\log\sigma|^{\rho_2})$ support points in $[-a_\sigma,\, a_\sigma]$, for $a_\sigma = O(|\log\sigma|^{1/\tau})$, such that the corresponding Gaussian mixture density $p_{F^*,\sigma}$ satisfies

$$n^{-1}\mathrm{KL}(p_0;\, p_{F^*,\sigma}) \lesssim \sigma^{2\beta}, \qquad n^{-k/2}V_k(p_0;\, p_{F^*,\sigma}) \lesssim \sigma^{k\beta}, \quad k \geq 2, \tag{3.4}$$

see, for instance, Lemma 4 in Kruijer et al. (2010). In all these articles, only the case where $k = 2$ has been treated for the inequality on the right-hand side (RHS) of (3.4), but the extension to any $k > 2$ is straightforward. The regularity assumptions considered in Kruijer et al. (2010), Shen et al. (2013) and Scricciolo (2014) to verify (3.4) are slightly different. For instance, Kruijer et al. (2010) assume that $\log p_0$ satisfies some locally Hölder conditions, while Shen et al. (2013) consider Hölder-type conditions on p_0 and Scricciolo (2014) Sobolev-type assumptions. To avoid taking into account all these special cases, we state (3.4) as a condition. We then have the following theorem, where the distance d defining the ball $U_{J_1\epsilon_n}$ with center at p_0 and radius $J_1\epsilon_n$ can equivalently be the Hellinger or the $\mathbb{L}_1$-distance. Note that the constant J_1 may be different for each one of the results stated below.

Theorem 2 *Suppose that p_0 satisfies the tail condition* (3.3) *and that the inequality on the RHS of* (3.4) *holds with $k > 8(2\beta+1)$. Consider a prior distribution of the form* (3.1)*, with an empirical Bayes selection $\hat\gamma_n$ for γ, and assume that $\hat\gamma_n$ and $\mathcal{K}_n$ satisfy condition* (3.2)*, where $\mathcal{K}_n \subseteq [m_1,\, m_2]\times[a_1,\, a_2(\log n)^{b_1}]$ for some constants $m_1,\, m_2 \in \mathbb{R}$, $a_1,\, a_2 > 0$ and $b_1 \geq 0$. Then, for a sufficiently large constant $J_1 > 0$,*

$$\mathbb{E}_{p_0}^{(n)}[\pi(U_{J_1\epsilon_n}^c \mid X^{(n)},\, \hat\gamma_n)] = o(1), \qquad \textit{with } \epsilon_n = n^{-\beta/(2\beta+1)}(\log n)^{a_3},$$

for some constant $a_3 > 0$.

In Theorem 2, the constant a_3 is the same as that appearing in the rate of convergence for the posterior distribution corresponding to a non data-dependent prior with fixed γ.

The crucial step for assessing posterior contraction rates in the case where p_0 is super-smooth, which, as in the ordinary smooth case, consists in proving the existence of a finitely supported Gaussian mixture density that approximates p_0 with an error of the appropriate order, requires some refinements. We suppose that p_0 has Fourier transform $\widehat{p_0}(t) = \int_{-\infty}^{\infty} e^{itx}p_0(x)\mathrm{d}x$, $t\in\mathbb{R}$, that satisfies for some finite constants $\rho,\, L > 0$ the integrability condition

$$\int_{-\infty}^{\infty} |\widehat{p_0}(t)|^2 e^{-2(\rho|t|)^r}\mathrm{d}t \leq 2\pi L^2, \qquad \text{with } r \in [1,\, 2], \tag{3.5}$$

where the regularity of p_0, which is measured through a scale of integrated tail bounds on the Fourier transform $\widehat{p_0}$, is related to the characteristic exponent r. Densities satisfying condition (3.5) are analytic on $\mathbb{R}$ and increasingly "smooth" as r increases. They form a larger class than that of analytic densities, including relevant statistical examples like the Gaussian distribution which corresponds to $r = 2$, the Cauchy distribution which corresponds to $r = 1$, all symmetric stable laws, Student's-t distribution, distributions with characteristic functions vanishing outside a compact set as well as their mixtures and convolutions. In order to state a counter-part of requirement (3.4) for the super-smooth case, we consider, for $\alpha \in (0,\, 1]$, the ρ_α-divergence of a density p from p_0 which is defined as $\rho_\alpha(p_0;\, p) = \alpha^{-1}\{\mathbb{E}_{p_0}[(p_0/p)^\alpha(X_1)] - 1\}$. Following the trail of Lemma 8 in Scricciolo (2014), it can be proved that, for any density p_0 satisfying condition (3.5), together with the monotonicity and tail conditions (b) and (c), respectively, of Section 4.2 in Scricciolo (2014), for $\sigma > 0$ small enough, there exists a finite mixing distribution F^*, with N_σ support points in $[-a_\sigma,\, a_\sigma]$, so that

$$\rho_\alpha(p_0;\, p_{F^*,\sigma}) \lesssim e^{-c(1/\sigma)^r} \qquad \text{for every } \alpha \in (0,\, 1], \tag{3.6}$$

where $a_\sigma = O(\sigma^{-r/(\tau\wedge 2)})$ and $N_\sigma = O((a_\sigma/\sigma)^2)$. Since

$$n^{-1}\mathrm{KL}(p_0;\, p_{F^*,\sigma}) = \lim_{\beta\to 0^+} \rho_\beta(p_0;\, p_{F^*,\sigma}) \le \rho_\alpha(p_0;\, p_{F^*,\sigma}) \quad \text{for every } \alpha \in (0,\, 1],$$

inequality (3.6) is stronger than that on the LHS of (3.4) and allows to obtain an almost sure lower bound on the denominator of the ratio defining the empirical Bayes posterior probability of the set $U^c_{J_1\epsilon_n}$, see Lemma 2 of Shen and Wasserman (2001).

Theorem 3 *Suppose that p_0 satisfies condition* (3.5) *and that the tail condition* (3.3) *holds with $\tau > 1$ such that $(\tau - 1)r \le \tau$. Suppose also that the monotonicity condition* (*b*) *of Section 4.2 in Scricciolo (2014) is satisfied. Consider a prior distribution of the form* (3.1)*, with an empirical Bayes selection $\hat{\gamma}_n$ for γ, and assume that $\hat{\gamma}_n$ and $\mathcal{K}_n$ satisfy condition* (3.2)*, where $\mathcal{K}_n \subseteq [m_1,\, m_2] \times [a_1,\, a_2(\log n)^{b_1}]$ for some constants $m_1,\, m_2 \in \mathbb{R}$, $a_1,\, a_2 > 0$ and $b_1 \ge 0$. Then, for a sufficiently large constant $J_1 > 0$,*

$$\mathbb{E}^{(n)}_{p_0}[\pi(U^c_{J_1\epsilon_n} \mid X^{(n)},\, \hat{\gamma}_n)] = o(1), \quad \text{with } \epsilon_n = n^{-1/2}(\log n)^{5/2+3/r}.$$

We now present some results on empirical Bayes posterior recovery rates for mixing distributions. These results are derived from Theorem 2 and Theorem 3 via some inversion inequalities. We first consider the case where the sampling density p_0 is itself a mixture of Gaussian densities and derive rates for recovering the true mixing distribution relative to any Wasserstein metric of order $1 \le q < \infty$. We then assess empirical Bayes posterior recovery rates relative to the $\mathbb{L}_2$-distance for the density deconvolution problem in the ordinary and super-smooth cases.

We begin by recalling the definition of *Wasserstein distance of order q*. For any $1 \le q < \infty$, define the *Wasserstein distance of order q* between any two Borel probability measures ν and ν' on Θ with finite qth-moment, *i.e.*, $\int_\Theta d^q(\theta,\, \theta_0)\nu(\mathrm{d}\theta) < \infty$ for some (hence any θ_0) in Θ, as $W_q(\nu,\, \nu') := (\inf_{\mu\in\Pi(\nu,\,\nu')} \int_{\Theta\times\Theta} d^q(\theta,\, \theta')\mu(\mathrm{d}\theta,\, \mathrm{d}\theta'))^{1/q}$, where μ runs over the set $\Pi(\nu,\, \nu')$ of all joint probability measures on $\Theta\times\Theta$ with marginal distributions ν and ν'. When $q = 2$, we take d to be the Euclidean distance $\|\cdot\|$. Posterior rates, relative to Wasserstein metrics, for recovering mixing distributions have been recently investigated by Nguyen (2013). In the following result, the prior probability measure corresponds to the product of a Dirichlet process, with a data-dependent base measure $\alpha_\mathbb{R}\mathrm{N}_{\hat{\gamma}_n}$, and a point mass at a given σ_0, in symbols, $\mathrm{DP}(\alpha_\mathbb{R}\mathrm{N}_{\hat{\gamma}_n}) \times \delta_{\sigma_0}$.

Corollary 1 *Suppose that $p_0 = F_0 * \phi_{\sigma_0}$, where the true mixing distribution F_0 satisfies the tail condition $F_0(\theta : |\theta| > t) \lesssim e^{-c_0t^2}$ for t large enough and σ_0 denotes the true value for the scale. Consider a prior distribution of the form $\mathrm{DP}(\alpha_\mathbb{R}\mathrm{N}_{\hat{\gamma}_n}) \times \delta_{\sigma_0}$ and assume that $\hat{\gamma}_n$ and $\mathcal{K}_n$ satisfy condition* (3.2)*, where $\mathcal{K}_n \subseteq [m_1,\, m_2] \times [a_1,\, a_2(\log n)^{b_1}]$ for some constants $m_1,\, m_2 \in \mathbb{R}$, $a_1,\, a_2 > 0$ and $b_1 \ge 0$. Then, for every $1 \le q < \infty$, there exists a sufficiently large constant $J_1 > 0$ so that*

$$\mathbb{E}^{(n)}_{p_0}[\pi(F : W_q(F,\, F_0) \ge J_1(\log n)^{-1/2} \mid X^{(n)},\, \hat{\gamma}_n)] = o(1).$$

The result implies that optimal recovery of mixing distributions is possible using Dirichlet process mixtures of Gaussian densities with an empirical Bayes selection for the prior hyper-parameters of the base measure. Dedecker and Michel (2013) have shown that, for the deconvolution problem with super-smooth errors, the rate $(\log n)^{-1/2}$ is minimax-optimal over a slightly larger class of probability measures than the one herein considered.

We now assess adaptive recovery rates for empirical Bayes density deconvolution when the errors are ordinary or super-smooth and the mixing density is modeled as a Dirichlet process mixture of Gaussian densities with a data-driven choice for the prior hyper-parameters of the base measure. The problem of deconvoluting a density when the mixing density is modeled as a Dirichlet process mixture of Gaussian densities and the errors are super-smooth has been recently investigated by Sarkar et al. (2013). In a frequentist set-up, rates for density deconvolution have been studied by Carroll and Hall (1988), Fan (1992), Fan (1991b), Fan (1991a). Consider the following model

$$X = Y + \varepsilon,$$

where Y and ε are independent random variables. Let p_Y denote the density of Y and K the density of the error measurement ε. The density of X is the convolution $p_X(\cdot) = (K * p_Y)(\cdot) = \int K(\cdot - y)p_Y(y)\mathrm{d}y$. The density K is assumed to be completely known and its Fourier transform $\widehat{K}$ to satisfy either

$$|\widehat{K}(t)| \gtrsim (1+t^2)^{-\eta/2}, \quad t \in \mathbb{R}, \qquad \text{(ordinary smooth case)} \tag{3.7}$$

for some $\eta > 0$, or

$$|\widehat{K}(t)| \gtrsim e^{-\varrho|t|^r}, \quad t \in \mathbb{R}, \qquad \text{(super-smooth case)} \tag{3.8}$$

for some ϱ, $r > 0$. The density p_Y is modeled as a Dirichlet process mixture of Gaussian densities as in (3.1) with an empirical Bayes choice $\hat{\gamma}_n$ of γ. Assuming data $X^{(n)} = (X_1, \ldots, X_n)$ are iid observations from a density $p_{0X} = K * p_{0Y}$ such that the Fourier transform $\widehat{p_{0Y}}$ of the mixing density p_{0Y} satisfies

$$\int_{-\infty}^{\infty} (1+t^2)^{\beta} |\widehat{p_{0Y}}(t)|^2 \, \mathrm{d}t < \infty \quad \text{for some } \beta > 1/2, \tag{3.9}$$

we derive adaptive empirical Bayes posterior convergence rates for recovering p_{0Y}.

Corollary 2 *Suppose that $\hat{K}$ satisfies either condition* (3.7) *(ordinary smooth case) or condition* (3.8) *(super-smooth case) and that $\widehat{p_{0Y}}$ satisfies the integrability condition* (3.9)*. Consider a prior for p_Y of the form* (3.1)*, with an empirical Bayes selection $\hat{\gamma}_n$ for γ. Suppose that $p_{0X} = K * p_{0Y}$ satisfies the conditions of Theorem 2 in the ordinary smooth case, with d being the $\mathbb{L}_2$-distance, or those of Theorem 3 in the super-smooth case. Then, there exists a sufficiently large constant $J_1 > 0$ so that*

$$\mathbb{E}_{p_{0X}}^{(n)}[\pi(\|p_Y - p_{0Y}\|_2 \geq J_1\epsilon_n \mid X^{(n)}, \hat{\gamma}_n)] = o(1),$$

where, for some constant $\kappa_1 > 0$,

$$\epsilon_n = \begin{cases} n^{-\beta/[2(\beta+\eta)+1]}(\log n)^{\kappa_1}, & \text{if } \hat{K} \text{ satisfies } (3.7), \\ (\log n)^{-\beta/r}, & \text{if } \hat{K} \text{ satisfies } (3.8). \end{cases}$$

The obtained rates are minimax-optimal, up to a logarithmic factor, in the ordinary smooth case and minimax-optimal in the super-smooth case. Inspection of the proof of Corollary 2 shows that, since the result is based on inversion inequalities that relate the $\mathbb{L}_2$-distance between the true mixing density and the random approximating mixing density to the $\mathbb{L}_2/\mathbb{L}_1$-distance between the corresponding mixed densities, once adaptive rates are known for the direct problem of Bayes or empirical Bayes estimation of the sampling density p_{0X}, the proof can be applied to yield adaptive recovery rates for either the Bayes or the empirical Bayes density deconvolution problem. If compared to the approach followed by Sarkar et al. (2013), the present strategy simplifies the derivation of adaptive recovery rates in a Bayesian density deconvolution problem. The ordinary smooth case seems to be first treated here even for the fully Bayesian adaptive density deconvolution problem.

4 Applications to counting processes with Aalen multiplicative intensity

In this section, we illustrate our results on counting processes with Aalen multiplicative intensity. Bayesian nonparametric methods have so far been mainly adopted to explore possible prior distributions on intensity functions with the aim of showing that Bayesian nonparametric inference for inhomogeneous Poisson processes can give satisfactory results in applications, see, *e.g.*, Kottas and Sansó (2007). First frequentist asymptotic behaviors of posteriors like consistency or computations of rates of convergence have been obtained by Belitser et al. (2013) still for inhomogeneous Poisson processes. As explained in introduction, Theorems 4 and 5 in Section 4.2 extend these results. Section 4.3 illustrates our procedures on artificial data.

4.1 Notations and setup

Let N be a counting process adapted to a filtration $(\mathcal{G}_t)_t$ with compensator Λ so that $(N_t - \Lambda_t)_t$ is a zero-mean $(\mathcal{G}_t)_t$-martingale. A counting process satisfies the *Aalen multiplicative intensity model* if $\mathrm{d}\Lambda_t = Y_t\lambda(t)\mathrm{d}t$, where λ is a non-negative deterministic function called, with a slight abuse, the intensity function in the sequel and $(Y_t)_t$ is a non-negative predictable process. Informally,

$$\mathbb{E}[N[t,\, t+\mathrm{d}t] \mid \mathcal{G}_{t^-}] = Y_t\lambda(t)\mathrm{d}t, \tag{4.1}$$

see Andersen et al. (1993), Chapter III. We assume that $\Lambda_t < \infty$ almost surely for every t. We also assume that the processes N and Y both depend on an integer n and we consider estimation of λ (not depending on n) in the asymptotic perspective $n \to \infty$, while T is kept fixed. The following cases illustrate the interest for this model.

- **Inhomogeneous Poisson process.** We observe n independent Poisson processes with common intensity λ. This model is equivalent to the model where we observe a Poisson process with intensity $n \times \lambda$, so it corresponds to the case $Y_t \equiv n$.

- **Survival analysis with right-censoring.** This model is popular in biomedical problems. We have n patients and, for each patient i, we observe $(Z_i,\, \delta_i)$ with $Z_i = \min\{X_i,\, C_i\}$, where X_i represents the lifetime of the patient, C_i is the independent censoring time and $\delta_i = 1_{X_i \le C_i}$. In this case, we set $N_t^i = \delta_i \times 1_{Z_i \le t}$, $Y_t^i = 1_{Z_i \ge t}$ and λ is the hazard rate of the X_i's: if f is the density of X_1, $\lambda(t) = f(t)/\mathbb{P}(X_1 \ge t)$. Then, N (respectively Y) is obtained by aggregating the n independent processes N^i's (respectively the Y^i's): for any $t \in [0,\, T]$, $N_t = \sum_{i=1}^n N_t^i$ and $Y_t = \sum_{i=1}^n Y_t^i$.

- **Finite state Markov process.** Let $X = (X(t))_t$ be a Markov process with finite state space $\mathbb{S}$ and right-continuous sample paths. We assume the existence of integrable *transition intensities* λ_{hj} from state h to state j for $h \neq j$. We assume we are given n independent copies of the process X, denoted by $X^1, \ldots, X^n$. For any $i \in \{1, \ldots, n\}$, let N_t^{ihj} be the number of direct transitions for X^i from h to j in $[0,\, t]$, for $h \neq j$. Then, the intensity of the multivariate counting process $\mathfrak{N}^i = (N^{ihj})_{h \neq j}$ is $(\lambda_{hj}Y^{ih})_{h\neq j}$, with $Y_t^{ih} = 1_{\{X^i(t^-)=h\}}$. As previously, we can consider $\mathfrak{N}$ (respectively Y^h) by aggregating the processes $\mathfrak{N}^i$ (respectively the Y^{ih}'s): $\mathfrak{N}_t = \sum_{i=1}^n \mathfrak{N}_t^i$, $Y_t^h = \sum_{i=1}^n Y_t^{ih}$ and $t \in [0,\, T]$. The intensity of each component $(N_t^{hj})_t$ of $(\mathfrak{N}_t)_t$ is then $(\lambda_{hj}(t)Y_t^h)_t$. We refer the reader to Andersen et al. (1993), p. 126, for more details. In this case, N is either one of the N^{hj}'s or the aggregation of some processes for which the λ_{hj}'s are equal.

We denote λ_0 the true intensity and we define $\mu_n(t) := \mathbb{E}^{(n)}_{\lambda_0}[Y_t]$ and $\tilde{\mu}_n(t) := n^{-1}\mu_n(t)$. We assume the existence of a non-random set $\Omega \subset [0, T]$ such that there are two constants m_1, m_2 satisfying for any n,

$$m_1 \le \inf_{t\in\Omega} \tilde{\mu}_n(t) \le \sup_{t\in\Omega} \tilde{\mu}_n(t) \le m_2 \tag{4.2}$$

and $\alpha \in (0, 1)$ such that if $\Gamma_n := \{\sup_{t\in\Omega} |n^{-1}Y_t - \tilde{\mu}_n(t)| \le \alpha m_1\} \cap \{\sup_{t\in\Omega^c} Y_t = 0\}$, where Ω^c is the complementary set of Ω in $[0, T]$, then

$$\lim_n \mathbb{P}^{(n)}_{\lambda_0}(\Gamma_n) = 1. \tag{4.3}$$

Assumption (4.2) implies that on Γ_n,

$$\forall\, t \in \Omega, \quad (1-\alpha)\tilde{\mu}_n(t) \le \frac{Y_t}{n} \le (1+\alpha)\tilde{\mu}_n(t). \tag{4.4}$$

Remark 2 *Since our results are asymptotic in nature, we can assume, without loss of generality, that* (4.2) *is true only for n large enough.*

Remark 3 *Our assumptions are satisfied, for instance, for the first two illustrative models introduced above. For inhomogeneous Poisson processes,* (4.2) *and* (4.3) *are obviously satisfied with $\Omega = [0, T]$ since $Y_t \equiv \mu_n(t) \equiv n$. For right-censoring models, with $Y^i_t = 1_{Z_i \ge t}$, $i = 1, \ldots, n$, we denote by Ω the support of the Z_i's and by $M_\Omega = \max \Omega \in \overline{\mathbb{R}}_+$. Then,* (4.2) *and* (4.3) *are satisfied if $M_\Omega > T$ or $M_\Omega \le T$ and $\mathbb{P}(Z_1 = M_\Omega) > 0$ (the concentration inequality is implied by using the DKW inequality).*

Recall that the log-likelihood for Aalen processes is given by, see Andersen et al. (1993),

$$\ell_n(\lambda) = \int_0^T \log(\lambda(t))\mathrm{d}N_t - \int_0^T \lambda(t)Y_t\mathrm{d}t. \tag{4.5}$$

Since N is empty on Ω^c almost surely, we only consider estimation over Ω. So, we set

$$\mathcal{F} = \left\{\lambda : \Omega \to \mathbb{R}_+ :\ \int_\Omega \lambda(t)\mathrm{d}t < \infty\right\}$$

endowed with the classical $\mathbb{L}_1$-norm: for all λ, $\lambda' \in \mathcal{F}$, let $\|\lambda - \lambda'\|_1 = \int_\Omega |\lambda(t) - \lambda'(t)|\mathrm{d}t$. We assume that the true intensity λ_0 satisfies $\lambda_0 \in \mathcal{F}$ and, for any $\lambda \in \mathcal{F}$, we write $\lambda = M_\lambda \bar{\lambda}$, where $M_\lambda = \int_\Omega \lambda(t)\mathrm{d}t$ and $\bar{\lambda} \in \mathcal{F}_1$, with $\mathcal{F}_1 = \{f \in \mathcal{F} :\ \int_\Omega f(t)\mathrm{d}t = 1\}$. Note that a prior probability measure π on $\mathcal{F}$ can be written as $\pi_1 \otimes \pi_M$, where π_1 is a probability distribution on $\mathcal{F}_1$ and π_M is a probability distribution on $\mathbb{R}_+$. This representation will be used in next section.

4.2 Empirical Bayes and Bayes posterior concentration rates for monotone non-increasing intensities

In this section, we concentrate on estimation of monotone non-increasing intensities, which is equivalent to considering $\bar{\lambda}$ monotone non-increasing in the above described parameterization. To construct a prior on the set of monotone non-increasing densities on $[0, T]$, we use their representation as mixtures of uniform densities as provided by Williamson (1956) and we consider a Dirichlet process prior on the mixing distribution:

$$\bar{\lambda}(x) = \int_0^\infty \frac{\mathbb{1}_{(0,\theta)}(x)}{\theta}\mathrm{d}P(\theta), \quad P|G_\gamma, A \sim \mathrm{DP}(AG_\gamma), \tag{4.6}$$

where G_γ is a distribution on $[0, T]$. This prior has been studied by Salomond (2013) for estimating monotone non-increasing densities. Here, we extend his results to the case of a monotone non-increasing intensity of an Aalen process with a data-dependent γ. We denote by $\pi(\cdot \mid N)$ the posterior distribution given the observations of the process N.

We study the family of distributions G_γ, with density denoted g_γ, belonging to one of the following families of densities with respect to the Lebesgue measure: for $\gamma > 0$, $a > 1$,

$$g_\gamma(\theta) = \frac{1}{G(T\gamma)} \gamma^a \theta^{a-1} e^{-\gamma\theta} 1\!\!1_{\{0 \le \theta \le T\}} \quad \text{or} \quad \left(\frac{1}{\theta} - \frac{1}{T}\right)^{-1} \sim \Gamma(a, \gamma), \tag{4.7}$$

where G is the cumulative distribution function of a $\Gamma(a, 1)$ random variable. Assume that, with probability going to 1, $\hat{\gamma}_n$ belongs to a fixed compact subset of $(1, \infty)$ denoted by $\mathcal{K}$. We then have the following theorem, which is an application of Theorem 1.

Theorem 4 *Let $\bar{\epsilon}_n = (n/\log n)^{-1/3}$. Assume that the prior π_M on the mass M is absolutely continuous with respect to the Lebesgue measure with positive and continuous density on $\mathbb{R}_+$ and that it has a finite Laplace transform in a neighbourhood of 0. Assume that the prior $\pi_1(\cdot \mid \gamma)$ on $\bar{\lambda}$ is a Dirichlet process mixture of uniform distributions defined by (4.6), with $A > 0$ and the base measure G_γ defined by (4.7). Let $\hat{\gamma}_n$ be a measurable function of the observations satisfying $\mathbb{P}^{(n)}_{\lambda_0}(\hat{\gamma}_n \in \mathcal{K}) = 1 + o(1)$ for some fixed compact subset $\mathcal{K} \subset (1, \infty)$. Assume also that (4.2) and (4.3) are satisfied and for any $k \ge 1$ there exists $C_{1k} > 0$ such that*

$$\mathbb{E}^{(n)}_{\lambda_0}\left[\left(\int_\Omega (Y_t - \mu_n(t))^2 \mathrm{d}t\right)^k\right] \le C_{1k} n^k \tag{4.8}$$

Then, there exists $J_1 > 0$ such that

$$\mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 \bar{\epsilon}_n \mid N, \hat{\gamma}_n)] = o(1)$$

and

$$\sup_{\gamma \in \mathcal{K}} \mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 \bar{\epsilon}_n \mid N, \gamma)] = o(1).$$

The proof of Theorem 4 is given in Section 5.5.1. It consists in verifying conditions [A1] and [A2] of Theorem 1 and is based on a general theorem on posterior concentration rates for Aalen processes which is presented below since it is of interest on its own. Let v_n be a positive sequence going to 0 such that $nv_n^2 \to \infty$. For all $j \ge 1$, we define

$$\bar{S}_{n,j} = \left\{\bar{\lambda} \in \mathcal{F}_1 :\ \|\bar{\lambda} - \bar{\lambda}_0\|_1 \le \frac{2(j+1)v_n}{M_{\lambda_0}}\right\},$$

where $M_{\lambda_0} = \int_\Omega \lambda_0(x)\mathrm{d}x$, $\bar{\lambda}_0 = M_{\lambda_0}^{-1}\lambda_0$. For $H > 0$ and $k \ge 2$, if $k_{[2]} = \min\{2^\ell :\ \ell \in \mathbb{N},\ 2^\ell \ge k\}$, we define

$$\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n) = \left\{\bar{\lambda} \in \mathcal{F}_1 :\ \ h^2(\bar{\lambda}_0, \bar{\lambda}) \le \frac{v_n^2}{1 + \log \|\frac{\bar{\lambda}_0}{\bar{\lambda}}\|_\infty},\ \max_{2 \le j \le k_{[2]}} E_j(\bar{\lambda}_0, \bar{\lambda}) \le v_n^2,\ \left\|\frac{\bar{\lambda}_0}{\bar{\lambda}}\right\|_\infty \le n^H,\ \|\bar{\lambda}\|_\infty \le H\right\},$$

where, for every integer j, $E_j(\bar{\lambda}_0, \bar{\lambda}) = \int_\Omega \bar{\lambda}_0(x) \left|\log \bar{\lambda}_0(x) - \log \bar{\lambda}(x)\right|^j \mathrm{d}x$ and $\|\cdot\|_\infty$ stands for the sup-norm.

Theorem 5 *Let v_n be a positive sequence satisfying $v_n^2 \ge \log n/n$ and $v_n = o(1)$. Assume that (4.2) and (4.3) are satisfied and that the prior π_M on the mass M is absolutely continuous with respect to the Lebesgue measure with positive and continuous density on $\mathbb{R}_+$. Assume also that (4.8) is satisfied for some $k \ge 2$. Finally, assume that the prior π_1 on $\bar{\lambda}$ satisfies the following assumptions for some $H > 0$.*

(*i*) *There exists* $\mathcal{F}_n \subset \mathcal{F}_1$ *such that*

$$\pi_1\left(\mathcal{F}_n^c\right) \leq e^{-(\kappa_0+2)nv_n^2}\pi_1(\bar{B}_{k,n}(H,\,\bar{\lambda}_0,\,v_n)),$$

with κ_0 *as defined in* (6.5)*, and for all* $\xi,\,\delta > 0$*,*

$$\log D(\xi,\,\mathcal{F}_n,\,\|\cdot\|_1) \leq n\delta \quad \textit{for all } n \textit{ large enough.}$$

(*ii*) *For all* $\zeta,\,\delta,\,\beta > 0$*, there exists* $J_0 > 0$ *such that, for all* j *such that* $J_0 \leq j \leq \beta/v_n$*,*

$$\frac{\pi_1(\bar{S}_{n,j})}{\pi_1(\bar{B}_{k,n}(H,\,\bar{\lambda}_0,\,v_n))} \leq e^{\delta(j+1)^2nv_n^2}$$

and

$$\log D(\zeta j v_n,\,\bar{S}_{n,j}\cap\mathcal{F}_n,\,\|\cdot\|_1) \leq \delta(j+1)^2nv_n^2.$$

Then, there exists $J_1 > 0$ *such that, when* $n \to \infty$*,*

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda:\ \|\lambda-\lambda_0\|_1 > J_1v_n \mid N)] = O((nv_n^2)^{-k/2}).$$

To the best of our knowledge, the only other paper dealing with posterior concentration rates in related models is Belitser et al. (2013) which considers inhomogeneous Poisson processes. Theorem 5 differs from their general Theorem 1 in two aspects. First, we do not restrict ourselves to inhomogeneous Poisson processes. Second, and more importantly, our set of conditions is quite different. More specifically, we do not need to assume that λ_0 is bounded from below and we do not need to bound from below the prior mass of neighborhoods of λ_0 for the sup-norm, but merely the prior mass of neighborhoods for the Hellinger distance, as in Theorem 2.2 of Ghosal et al. (2000). Our aim, in Theorem 5, is to propose a set of conditions to derive posterior concentration rates on the intensity which are as close as possible to the conditions used in the density model by parameterizing λ as $\lambda = M_\lambda\bar{\lambda}$, where $\bar{\lambda}$ is a probability density on Ω.

Remark 4 *Note that if* $\bar{\lambda} \in \bar{B}_{2,n}(H,\,\bar{\lambda}_0,\,v_n)$*, then, for any* $j > 2$*,* $E_j(\bar{\lambda}_0,\,\bar{\lambda}) \leq H^{j-2}v_n^2(\log n)^{j-2}$ *so that, using Proposition 1, if we replace* $\bar{B}_{k,n}(H,\,\bar{\lambda}_0,\,v_n)$ *with* $\bar{\lambda} \in \bar{B}_{2,n}(H,\,\bar{\lambda}_0,\,v_n)$ *in the assumptions of Theorem 5, we obtain the same type of conclusion: for any* $k \geq 2$*, such that condition* (4.8) *is satisfied,*

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda:\ \|\lambda-\lambda_0\|_1 \geq J_1v_n \mid N)] = O((nv_n^2)^{-k/2}(\log n)^{k(k_{[2]}-2)/2})$$

with an extra $(\log n)$*-term on the right-hand side above.*

Remark 5 *We now prove that Assumption* (4.8) *is reasonable by considering the previous examples. Obviously,* (4.8) *is satisfied in the case of inhomogeneous Poisson processes since* $Y_t = n$ *for every* t*. For the censoring model,* $Y_t = \sum_{i=1}^n 1_{Z_i\geq t}$*. We set, for any* i*,* $V_i = 1_{Z_i\geq t} - \mathbb{P}(Z_1 \geq t)$*. Then, for* $k \geq 1$*,*

$$\begin{aligned}
\mathbb{E}_{\lambda_0}^{(n)}\left[\left(\int_\Omega (Y_t-\mu_n(t))^2\mathrm{d}t\right)^k\right] &\leq \mathbb{E}_{\lambda_0}^{(n)}\left[\left(\int_0^T\left(\sum_{i=1}^n V_i\right)^2\mathrm{d}t\right)^k\right] \\
&\leq T^{k-1}\int_0^T \mathbb{E}_{\lambda_0}^{(n)}\left[\left(\sum_{i=1}^n V_i\right)^{2k}\right]\mathrm{d}t \\
&\leq C(k,\,T)\int_0^T\left(\sum_{i=1}^n \mathbb{E}_{\lambda_0}^{(n)}[V_i^{2k}] + \left(\sum_{i=1}^n \mathbb{E}[V_i^2]\right)^k\right)\mathrm{d}t \\
&\leq C_{1k}n^k,
\end{aligned}$$

where $C(k,T)$ and C_{1k} depend on k and T by using the Hölder and Rosenthal inequalities (see, for instance, Theorem C.2 of Härdle et al. (1998)). Under some mild conditions, similar computations can be derived for finite state Markov processes.

4.3 Numerical illustration

We now present a numerical experiment to highlight the impact of an empirical prior distribution for finite sample sizes n, in the case of an inhomogeneous Poisson process. Let $(W_i)_{i=1\dots N(T)}$ be the observed points of the process N over $[0,T]$, $N(T)$ being the observed number of jumps over $[0,T]$. We recall that the intensity function has the form $n\lambda_0(t) = nM_{\lambda_0}\overline{\lambda_0}(t)$ (n being known) where $\int_0^T \overline{\lambda}(t)dt = 1$.

The estimations of M_{λ_0} and $\overline{\lambda_0}$ can be done separately, given the factorisation in (4.5). Provided the use of a Gamma prior distribution on M_{λ_0} ($M_{\lambda_0} \sim \Gamma(a_M, b_M)$), we have

$$M_{\lambda_0}|N \sim \Gamma(a_M + N(T), b_M + n).$$

The non-parametric Bayesian estimation of $\overline{\lambda}$ is more involved. However in the case of Dirichlet process mixtures of uniforms as a prior model on $\bar{\lambda}$ we can use the same algorithms as those considered for density estimation. In this Section we restrict ourselves to the case where the base measure in the Dirichlet process is the second possibility in (4.7), namely $G_\gamma =_{\mathcal{D}} \left[\frac{1}{T} + \frac{1}{\Gamma(a,\gamma)}\right]^{-1}$. It satisfies the assumptions of Theorem 4 and also presents computational advantages. Hence, three hyperparameters are involved in this prior, namely A the mass of the Dirichlet process, a and γ. The hyperparameter A strongly conditions the number of classes in the posterior distribution of $\overline{\lambda}$. In order to mitigate its influence on the posterior, we propose to consider a hierarchical approach and set a Gamma prior distribution on A: $A \sim \Gamma(a_A, b_A)$. In the absence of additional information, we set $a_A = b_A = 1/10$, corresponding to a weakly informative prior. Theorem 4 applies for any $a > 1$. We arbitrarily set $a = 2$, the influence of a is not studied in this paper. We compare three strategies for the determination of γ in our simulation study.

Strategy 1: Empirical Bayes - We propose the following empirical estimator:

$$\widehat{\gamma}_n = \Psi^{-1}\left[\overline{W}_{N(T)}\right] \tag{4.9}$$

which comes from the following equality:

$$\begin{aligned}\mathbb{E}\left[\overline{W}_{N(T)}|N(T)\right] = \mathbb{E}\left[\frac{1}{N(T)}\sum_{i=1}^{N(T)} W_i \,\middle|\, N(T)\right] &= \int_0^T t\int_\theta \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(t)dG_\gamma(\theta)\,dt \\ &= \frac{\gamma^a}{2\Gamma(a)}\int_{\frac{1}{T}}^\infty \frac{e^{-\gamma/(\nu-\frac{1}{T})}}{(\nu-\frac{1}{T})^{(a+1)}}\frac{1}{\nu}d\nu := \Psi(\gamma)\end{aligned}$$

Strategy 2: Fixed γ - In order to avoid the empirical prior, one can fix $\gamma = \gamma_0$. In order to study the impact of a bad choice of γ on the behaviour of the posterior distribution, we propose to choose γ_0 far from a well calibrated value for γ, which would be $\gamma^* = \Psi^{-1}(E_{theo})$ with $E_{theo} = \mathbb{E}\left[W|N(T)\right]$. Since, in the simulation study, we know the true λ_0, γ^* can be computed and we consider

$$\gamma_0 = \rho\cdot\Psi^{-1}(E_{theo}), \quad \text{where} \quad E_{theo} = \int_0^T t\overline{\lambda}(t)dt, \quad \rho \in \{0.01, 30, 100\} \tag{4.10}$$

Strategy 3: Hierarchical Bayes - We consider a hierarchical prior on γ given by:

$$\gamma \sim \Gamma(a_\gamma, b_\gamma)$$

In order to make a fair comparison with the empirical posterior, we center the prior distribution on $\widehat{\gamma}$. Besides, in the simulation study, we set two different hierarchical hyperparameters (a_γ, b_γ), corresponding to two prior variances. More precisely, (a_γ, b_γ) are such that the prior expectation is equal to $\widehat{\gamma}$ and the prior variance is σ_γ^2, the values of σ_γ being specified in Table 1.

Samples from the posterior distribution of $(\overline{\lambda}, A, \gamma)|N$ are generated using a Gibbs algorithm, decomposed into two or three steps:

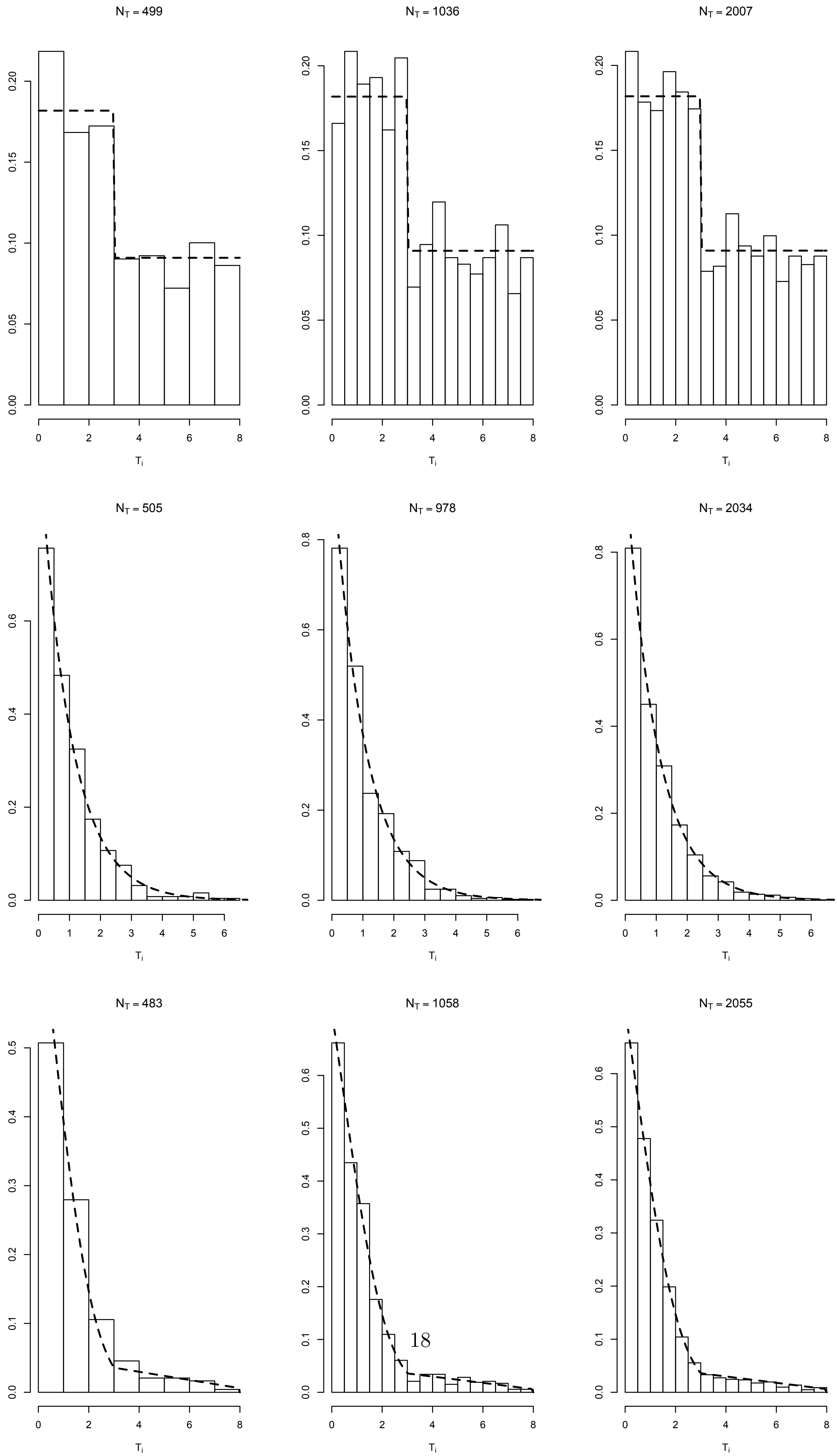
$$[1] \quad \overline{\lambda}|A, \gamma, N \qquad [2] \quad A|\overline{\lambda}, \gamma, N \qquad [3]^\dagger \quad \gamma|A, \overline{\lambda}, N,$$

step $[3]^\dagger$ only existing if a fully Bayesian strategy is adopted on the hyperparameter γ (strategy 3). We use the algorithm developed by Fall and Barat (2012).
The various strategies for calibrating the hyperparameter γ are tested on 3 different intensity functions (non null over $[0, T]$, with $T = 8$):

$$\begin{aligned}
\lambda_{0,1}(t) &= \left[4 \ \mathbb{1}_{[0,3]}(t) + 2 \ \mathbb{1}_{[3,8]}(t)\right] \\
\lambda_{0,2}(t) &= e^{-0.4t} \\
\lambda_{0,3}(t) &= \left[\cos^{-1}\Phi(t)\mathbb{1}_{[0,3]}(t) - (\frac{1}{6}\cos^{-1}\Phi(3)t - \frac{3}{2}\cos^{-1}\Phi(3))\mathbb{1}_{[3,8]}(t)\right]
\end{aligned}$$

where Φ is the probability function of the standard normal distribution. For each intensity $\lambda_{0,1}$, $\lambda_{0,2}$ and $\lambda_{0,3}$, we simulate 3 datasets corresponding to $n = 500, 1000, 2000$ respectively. The histograms of the 9 datasets with the true corresponding normalized intensities $\overline{\lambda_{0,i}}$ are plotted on Figure 1. In the following we denote by $\mathbf{D}_n^i$ the dataset associated with n and intensity $\lambda_{0,i}$. For each dataset, we adopt the 3 exposed strategies to calibrate γ. The posterior distributions are sampled using 30000 iterations of which are removed a burn-in period of 15000 iterations.

To compare the three different strategies used to calibrate γ, several criteria are taken into account: tuning of the hyperparameters, quality of the estimation, convergence of the MCMC, computational time. In terms of tuning effort on γ, the empirical Bayes (strategy 1) together with the fixed γ approach are comparable and significantly simpler than the hierarchical one, which increases the space to be explored by the MCMC algorithm and consequently slows down its convergence. Moreover, setting an hyper-prior distribution on γ requires to choose the parameters of this additional distribution (a_γ and b_γ) and thus postpone the problem, even though these second order hyperparameters are supposed to be less influential. In our particular example, computational time, for a fixed number of iterations here equal to $N_{iter} = 30000$, turned out to be also a key point. Indeed, the simulation of the $\bar{\lambda}$ conditionally to the other variables involves an accept-reject (AR) step. For some particular values of γ (small values of γ), we observe that the acceptance rate of the AR step can be dramatically low, automatically inflating the execution time of the algorithm, this phenomenon occurring randomly. The computational times (CpT) are summarized in Table 1, which also provides for each of the 9 datasets the number of points (N_T), the $\hat{\gamma}$ computed using formula (4.9) and to be compared with the targeted value $\Psi^{-1}(E_{theo})$, the perturbation factor ρ used in the fixed γ strategy and the standard deviation of the prior distribution of γ σ_γ (the prior mean being $\hat{\gamma}$) used in the two hierarchical approaches. The second hierarchical prior distribution (last columns of Table 1) corresponds to a prior distribution more concentrated around the empirical value $\hat{\gamma}$.
On Figure 2 (respectively 3 and 4), we plot for each strategy and each dataset the posterior


N_T = 499
N_T = 1036
N_T = 2007
N_T = 505
N_T = 978
N_T = 2034
N_T = 483
N_T = 1058
N_T = 2055
T_i

		N_T	Empirical $\widehat{\gamma}$	CpT	γ fixed $\rho\Psi^{-1}(E_{theo})$	CpT	Hierarchical σ_γ	CpT	Hierarchical 2 σ_γ	CpT
$\lambda_{0,1}$	D^1_{500}	499	0.0386	523.57	0.01×0.0323	<u>2085.03</u>	0.005	<u>12051.22</u>	0.001	447.75
	D^1_{1000}	1036	0.0372	783.53		<u>1009.58</u>		791.28		773.33
	D^1_{2000}	2007	0.0372	1457.40		1561.64		1477.50		1456.03
$\lambda_{0,2}$	D^2_{500}	505	0.6605	1021.73	100×0.6667	1022.59	0.1	663.54	0.01	1047.42
	D^2_{1000}	978	0.6857	1873.05		1416.40		1207.07		2018.89
	D^2_{2000}	2034	0.6827	4849.80		2236.02		2533.62		4644.55
$\lambda_{0,3}$	D^3_{500}	483	0.4094	782.19	30×0.4302	822.12	0.1	788.14	0.01	788.00
	D^3_{1000}	1058	0.4398	1610.47		2012.96		1559.17		1494.75
	D^3_{2000}	2055	0.4677	3546.57		<u>9256.71</u>		3179.96		2770.83

Table 1: Values of γ (or the hyperparameters of its prior distribution) and computational time required for the Gibbs algorithm with 30000 iterations. N_T: number of observations for each dataset D^i_n. CpT: Computational Time. σ_γ: standard deviation of the prior distribution on γ

median of $\overline{\lambda}_1$ (respectively $\overline{\lambda}_2$ and $\overline{\lambda}_3$) together with a pointwise credible interval using the 0.1% and 0.9% empirical quantiles obtained from the posterior simulation. Table 2 gives the distances between the estimated normalized intensities $\widehat{\overline{\lambda}}_i$ and the true $\overline{\lambda}_i$, for each dataset and each prior setting.

On $\lambda_{0,1}$ (which a simple 2-steps function), the 4 strategies lead to the same quality of estimation (in term of losses / distances between the functions of interest). In this case, it is thus interesting to have a look at the computational time in Table 1. We notice that for a small γ_0 or for a diffuse prior distribution on γ (so possibly generating small values of γ) the computational time explodes. In practice, this phenomenon can be so critical that the user may have to stop the execution and re-launch the algorithm. Moreover, interestingly the posterior mean of the number of non empty components in the mixture (computed over the last 10000 iterations) is equal in the case $n = 500$ to 4.21 in the empirical strategy, 11.42 when γ is fixed arbitrarily, 6.98 under the hierarchical large prior and 3.77 with the narrow hierarchical prior. In this case choosing a small value of γ leads to a posterior distribution on mixtures with too many non empty components. This phenomena tends to disappear when n increases.

For $\lambda_{0,2}$ and $\lambda_{0,3}$, a bad choice of γ - here γ too large in strategy 2 - or a not enough informative prior on γ (hierarchical prior with large variance) has a significant negative impact on the behavior of the posterior distribution. Contrariwise, the medians of the empirical and the informative hierarchical posterior distribution of λ have similar losses, as seen in Table 2.

5 Proofs

The notation $\lesssim$ will be used to denote inequality up to a constant that is fixed throughout. We denote by C a constant depending on m_1, m_2, k and so forth, which may change from line to line.

5.1 Proof of Theorem 2 on empirical Bayes Dirichlet process mixtures of Gaussian densities for the ordinary smooth case

To prove Theorem 2, we must verify assumptions [A1] and [A2]. We first define the change of parameter $\psi_{\gamma,\gamma'}(p_{F,\sigma})$. Under the Dirichlet process mixture prior we can write, when the base measure is associated to the parameter $\gamma = (m, s^2)$, that $p_{F,\sigma}(\cdot) = \sum_{j\geq 1} p_j\phi_\sigma(\cdot - \theta_j)$ almost

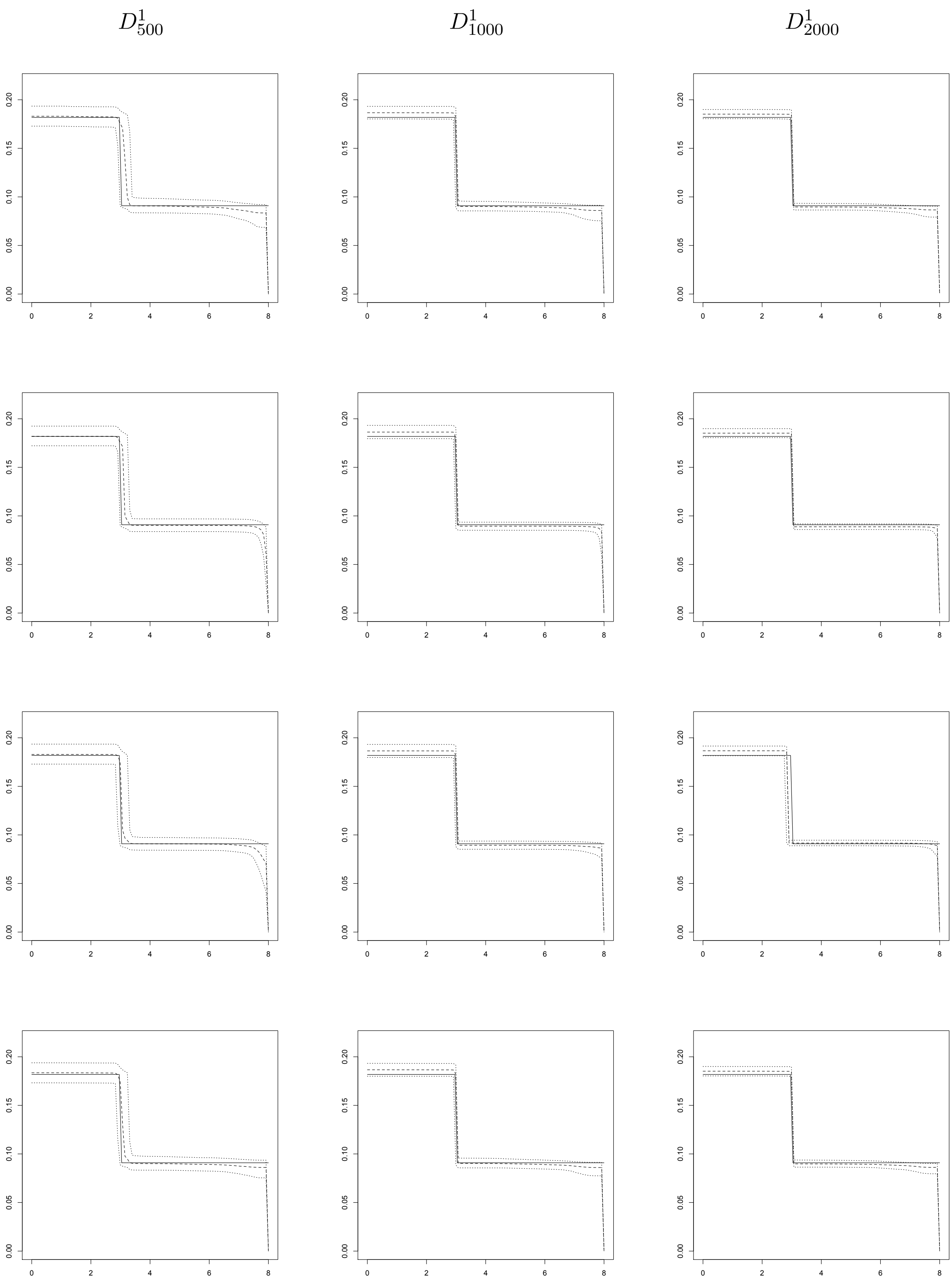


Figure 2: Estimation of $\overline{\lambda}_1$ from D^1_{500} (first column), D^1_{1000} (second column) and D^1_{2000} (third column) using the four strategies: empirical prior (line 1), fixed γ (line 2), hierarchical empirical prior (line 3), concentrated hierarchical empirical prior (line 4) . True density (plain line), estimation (dashed line) and confidence band (dotted lines)

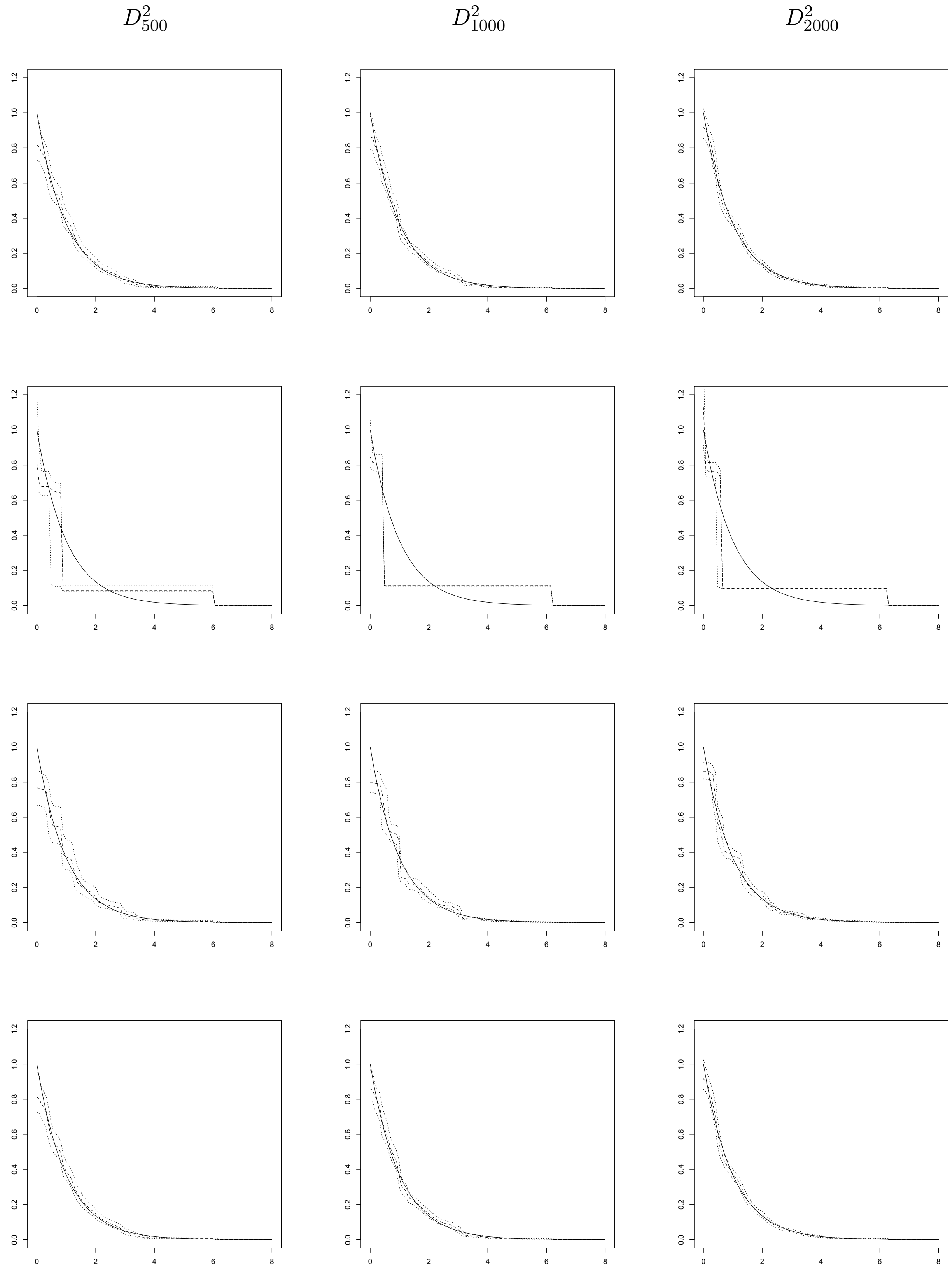


Figure 3: Estimation of $\overline{\lambda}_2$ from D^2_{500} (first column), D^2_{1000} (second column) and D^2_{2000} (third column) using the four strategies: empirical prior (line 1), fixe γ (line 2), hierarchical empirical prior (line 3), concentrated hierarchical empirical prior (line 4) . True density (plain line), estimation (dashed line) and confidence band (dotted lines)

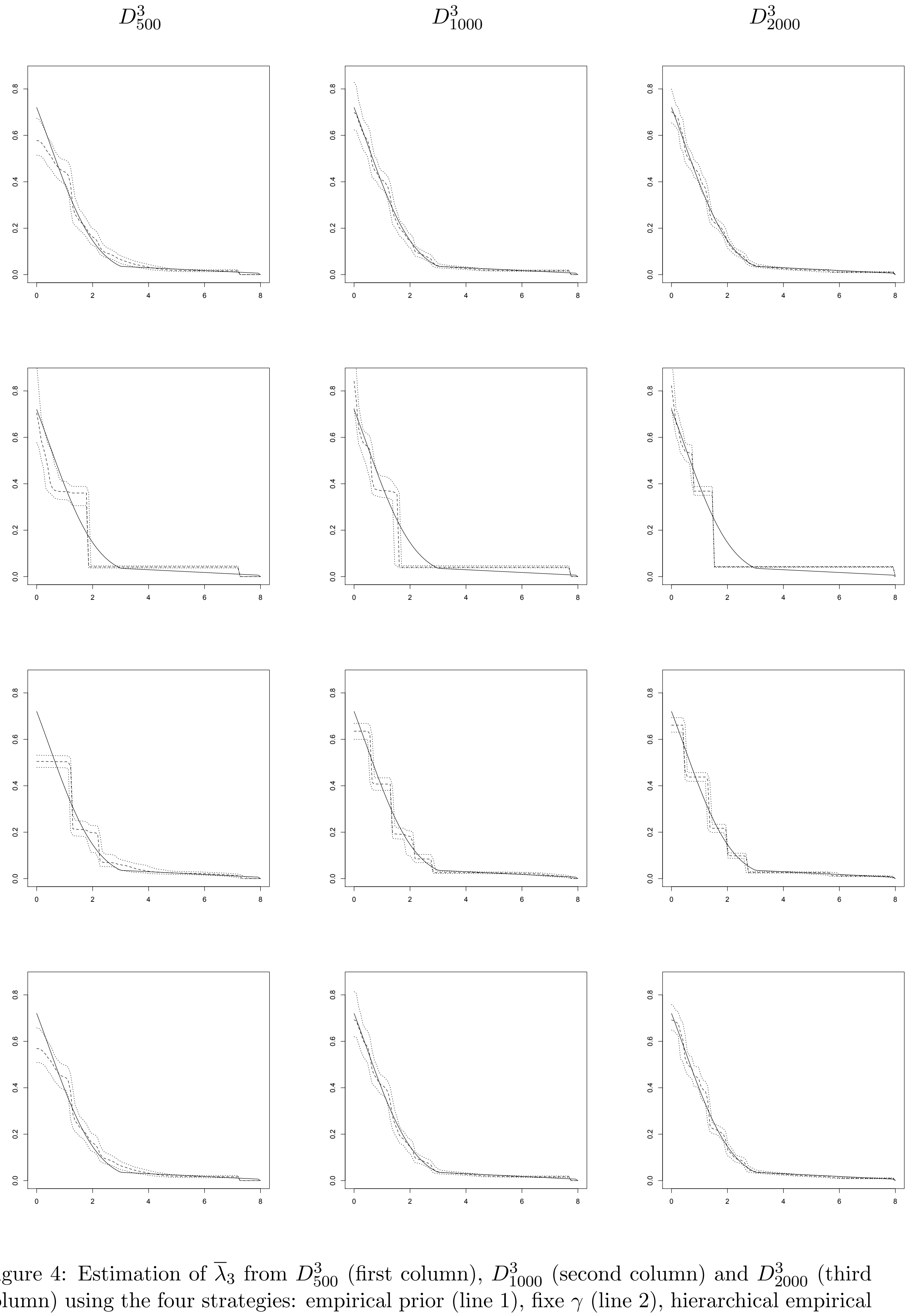


Figure 4: Estimation of $\overline{\lambda}_3$ from D^3_{500} (first column), D^3_{1000} (second column) and D^3_{2000} (third column) using the four strategies: empirical prior (line 1), fixe γ (line 2), hierarchical empirical prior (line 3), concentrated hierarchical empirical prior (line 4) . True density (plain line), estimation (dashed line) and confidence band (dotted lines)

		$\lambda_{0,1}$			$\lambda_{0,2}$			$\lambda_{0,3}$		
		D^1_{500}	D^1_{1000}	D^1_{2000}	D^2_{500}	D^2_{1000}	D^2_{2000}	D^3_{500}	D^3_{1000}	D^3_{2000}
d_{L_1}	Empir	0.0246	0.0238	0.0207	0.0921*	0.0817*	0.0549*	0.1382	0.0596	0.0606
	Fixed	0.0161	0.0219	0.0211	0.5381	0.7221	0.6356	0.3114	0.2852	0.2885
	Hierar	0.0132	0.0233	0.0317	0.1082	0.1280	0.0969	0.2154	0.1378	0.1405
	Hiera 2	0.0191	0.0240	0.0208	0.0925*	0.0815*	0.055*2	0.1383	0.0607	0.0724
d_{L_2}	Empir	0.0014	0.0006	0.0006	0.0010	0.0010	0.0008	0.0008	0.0005	0.0006
	Fixed	0.0014	0.0006	0.0006	0.0251	0.0376	0.0378	0.0084	0.0095	0.0104
	Hierar	0.0005	0.0006	0.0020	0.0033	0.0050	0.0034	0.0066	0.0070	0.0072
	Hiera 2	0.0009	0.0006	0.0006	0.0011	0.0011	0.0009	0.0009	0.0005	0.0013
$\lVert\cdot\rVert_\infty$	Empir	0.0909	0.0909	0.0909	0.1828	0.1369	0.0833	0.1421	0.0462	0.0428
	Fixed	0.0909	0.0909	0.0909	0.3261	0.5022	0.4271	0.1707	0.1818	0.2020
	Hierar	0.0909	0.0909	0.0909	0.2330	0.2007	0.1389	0.2154	0.1000	0.1190
	Hierar2	0.0909	0.0909	0.0909	0.1886	0.1413	0.0835	0.1515	0.0513	0.0555

Table 2: Distances between the estimated and the true density for all datasets and all strategies. Distance L_1 in horizontal block one, distance L_2 in horizontal block two, distance $\|\cdot\|_\infty$ in horizontal block three.

surely, with the $\theta_j \sim \mathrm{N}_{(m,s^2)}$ independently and independently of the $(p_j)_{j\geq 1}$. Without loss of generality, we assume that $\mathcal{K}_n = [m_1,\, m_2] \times [s_1^2,\, s_n^2]$, with $-\infty < m_1 \leq m_2 < \infty$, $s_1^2 > 0$ and s_n^2 possibly going to infinity as a power of $\log n$. Consider a u_n-covering of $[m_1,\, m_2]$ with intervals I_k, for $k = 1,\, \ldots,\, L_{mn}$, where $L_{mn} = \lfloor (m_2 - m_1)/u_n \rfloor$, and a covering of $[s_1^2,\, s_n^2]$ with intervals of the form $J_l = [s_1^2(1+u_n)^{l-1},\, s_1^2(1+u_n)^l]$, for $l = 1,\, \ldots,\, L_{sn}$, where $L_{sn} \leq 2u_n^{-1}\log(s_n/s_1)$ and such that $L_{sn} \to \infty$ as $n \to \infty$. We suppose that $u_n \to 0$.

For $s^2 \in J_l$, $l = 1,\, \ldots,\, L_{sn}$, let $\rho_l = (s^2/s_l^2)^{1/2} \leq (1+u_n)^{1/2}$, with $s_l^2 = s_1^2(1+u_n)^{l-1}$, and, for $m \in I_k$, $k = 1,\, \ldots,\, L_{mn}$, writing $m_k = m_1 + (k-1)u_n$, if, for every $j \in \mathbb{N}$, $\theta'_j \sim \mathrm{N}_{(m_k,\, s_l^2)} = \mathrm{N}_{\gamma'}$ and $\theta_j = \rho_\ell(\theta'_j - m_k) + m$, then $\theta_j \sim \mathrm{N}_{(m,\, s^2)} = \mathrm{N}_\gamma$. Therefore, conditionally on σ, for $F \sim \mathrm{DP}(\alpha_{\mathbb{R}}\mathrm{N}_\gamma)$,

$$\psi_{\gamma,\gamma'}(p_{F,\sigma})(x) = \sum_{j\geq 1} p_j \phi_\sigma(x - \theta'_j - [\theta'_j(\rho_l - 1) + m - m_k\rho_l]) \tag{5.1}$$

is distributed according to a Dirichlet process location mixture of Gaussian distributions, with base measure $\mathrm{N}_{\gamma'}$. The following inequalities are used repeatedly in the sequel,

$$(1+u_n)^{1/2} \geq \rho_l \geq 1, \qquad -m_k u_n \leq m - \rho_l m_k \leq u_n.$$

With abuse of notation, we also denote by $\psi_{\gamma,\gamma'}(\theta_j) = \rho_\ell(\theta'_j - m_k) + m$ for any $\gamma' = (m_k,\, s_l^2)$ and $\gamma = (m,\, s^2)$. We first verify assumption [A1]. From condition (3.4), let $\sigma \in (\sigma_n/2,\, 2\sigma_n)$, with $\sigma_n = \epsilon_n^{1/\beta}$, and F^* be as defined in (3.4), so that $F^* = \sum_{j=1}^{N_\sigma} p_j^* \delta_{\theta_j^*}$ with the minimal distance between any two θ_j's bounded from below by $\delta_j = \sigma\epsilon_n^{2b}$ for some $b > 0$. Construct a partition $(U_j)_{j=1}^M$ of $\mathbb{R}$ following the proof of Theorem 4 of Shen et al. (2013). Let $a_\sigma = a_0|\log\sigma|^{1/\tau}$. The partition is such that $(U_j)_{j=1}^K$ is a partition of $[-a_\sigma, a_\sigma]$ composed of intervals in the form $[\theta_j^* - \delta_j/2,\, \theta_j^* + \delta_j/2]$, $j = 1,\, \ldots,\, N_\sigma$, and of intervals with diameter smaller than or equal to σ to complete $[-a_\sigma,\, a_\sigma]$. Then, construct a partition of $(-\infty,\, -a_\sigma)$ and $(a_\sigma,\, \infty)$ with intervals U_j, $j > K$, such that $1 \geq \mathrm{N}_\gamma(U_j) \geq \sigma\epsilon_n^{2b}$. Note that, as in Shen et al. (2013), $M \lesssim \sigma^{-1}(\log n)^{1+1/\tau}$ and that, for all $j \leq K$, $\mathrm{N}_\gamma(U_j) \gtrsim \frac{\delta_j}{s} e^{-2a_\sigma^2/s^2} \gtrsim \epsilon_n^{b'}$ for some $b' > b$ uniformly in $\gamma \in \mathcal{K}_n$. As in

Shen et al. (2013), define B_n^γ as the set of (F, σ) such that $\sigma \in (\sigma_n/2, 2\sigma_n)$ and

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

Following Lemma 10 of Ghosal and van der Vaart (2007b), we obtain that for some constant c independent of $\gamma \in \mathcal{K}_n$,

$$\inf_{\gamma\in\mathcal{K}_n} \pi\left(B_n^\gamma \mid \gamma\right) \gtrsim e^{-c\sigma^{-1}(\log n)^{2+1/\tau}}. \tag{5.2}$$

Moreover, for all $(F, \sigma) \in B_n^\gamma$, with $\gamma' = (m_k, s_l^2)$ and any $\gamma \in I_k \times J_l$,

$$\begin{aligned} p_{F,\sigma}(x) &= \sum_{j\geq 1} p_j \phi_\sigma(x - \psi_{\gamma,\gamma'}(\theta_j)) \geq \sum_{j\geq 1} \mathbb{1}_{|\theta_j'|\leq a_\sigma} p_j \phi_\sigma(x - \psi_{\gamma,\gamma'}(\theta_j)) \\ &\geq \sum_{j\geq 1} \mathbb{1}_{|\theta_j'|\leq a_\sigma} p_j \phi_\sigma(x - \theta_j') e^{-4\frac{|x-\theta_j'|(a_\sigma+1)u_n + (a_\sigma^2+1)u_n^2}{\sigma_n^2}}. \end{aligned}$$

Note that $n^{-1}\sigma_n^{-1} = \epsilon_n^2$. Choosing $u_n \lesssim n^{-2}\sigma_n^{-1}(\log n)^{-1/\tau} = n^{-1}\epsilon_n^2(\log n)^{-1/\tau}$, we obtain that on the event $A_n = \{\sum_{i=1}^n |X_i - m_0| \leq n\tau_0^2 k_n\}$, for $k_n \geq (\log n)^{1/\tau}$, using the inequality $\log x \geq (x-1)/x$ valid for $x > 0$,

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

with a constant C' large enough, where $p_{F_n,\sigma}(x) = c_\sigma^{-1}\sum_{j\geq 1} \mathbb{1}_{|\theta_j'|\leq a_\sigma} p_j \phi_\sigma(x - \theta_j')$ and $c_\sigma = \sum_{j:\,|\theta_j'|\leq a_\sigma} p_j \geq (1 - 2\epsilon_n^{2b'}) \geq (1 - \epsilon_n^2)$ because $b' > 1$. The proof of Theorem 4 of Shen et al. (2013), together with condition (3.4), implies that [A1] is satisfied. We now prove [A2]. As in Proposition 2 of Shen et al. (2013), consider

$$\mathcal{F}_n = \left\{(F, \sigma) : \underline{\sigma}_n \leq \sigma \leq \bar{\sigma}_n,\ \ F = \sum_{j=1}^{\infty} p_j \delta_{\theta_j},\ \ |\theta_j| \leq n^{1/2}\ \ \forall\, j \leq H_n,\ \ \sum_{j>H_n} p_j \leq \epsilon_n \right\}, \tag{5.3}$$

with $\underline{\sigma}_n = \epsilon_n^{1/\beta}$, $\bar{\sigma}_n = \exp(tn\epsilon_n^2)$ for some $t > 0$ depending on the parameters (ν_1, ν_2) of the inverse-gamma distribution for σ, and $H_n = \lfloor n\epsilon_n^2/\log n \rfloor$. For some $x_0 > 0$, let $a_n = 2x_0(\log n)^{1/\tau}$, $\gamma' = (m_k, s_l^2) \in \mathcal{K}_n$, $\gamma \in I_k \times J_l$ and $|x| \leq a_n/2$. If $|\theta| \geq a_n$ then $|x - \theta| \geq |\theta|/2$ and, for $u_n \leq n^{-2}$, we can bound $p_{F,\sigma}$ as follows:

$$\begin{aligned} p_{F,\sigma}(x) &= \int_{\mathbb{R}} \phi_\sigma(x - \psi_{\gamma,\gamma'}(\theta))\mathrm{d}F'(\theta) \leq \int_{\mathbb{R}} \phi_\sigma(x - \theta) e^{\frac{|x-\theta|u_n(|\theta|+1)}{\sigma^2}} \mathrm{d}F'(\theta) \\ &\leq e^{3a_n(a_n+1)u_n\sigma^{-2}/2} \int_{|\theta|\leq a_n} \phi_\sigma(x-\theta)\mathrm{d}F'(\theta) + \int_{|\theta|>a_n} \phi_\sigma((x-\theta)(1-8u_n))\mathrm{d}F'(\theta) \\ &\leq e^{o(1/n)} \int_{|\theta|\leq a_n} \phi_\sigma(x-\theta)\mathrm{d}F'(\theta) + (1 + O(1/n)) \int_{|\theta|>a_n} \phi_{\tilde{\sigma}_n}(x-\theta)\mathrm{d}F'(\theta) \\ &= e^{o(1/n)} \left(\int_{|\theta|\leq a_n} \phi_\sigma(x-\theta)\,\mathrm{d}F'(\theta) + \int_{|\theta|>a_n} \phi_{\tilde{\sigma}_n}(x-\theta)\mathrm{d}F'(\theta) \right), \end{aligned}$$

where $F' \sim \mathrm{DP}(\alpha_{\mathbb{R}} \mathrm{N}_{\gamma'})$, $\tilde{\sigma}_n = \sigma(1 - 1/n^2)^{-1/2}$. Now, since $\mathbb{P}_{p_0}^{(n)}(\Omega_n) = 1 + o(1)$, with $\Omega_n = \{-a_n/2 \le \min_i X_i \le \max_i X_i \le a_n/2\}$, we can replace $\mathbb{R}^n$ with Ω_n and define for all $(F, \sigma) \in \mathcal{F}_n$,

$$q_\gamma^{F,\sigma}(x) = 1\!\!1_{[-a_n/2,\, a_n/2]}(x) e^{o(1/n)} \left(\int_{|\theta| \le a_n} \phi_\sigma(x - \theta) \mathrm{d}F(\theta) + \int_{|\theta| > a_n} \phi_{\tilde{\sigma}_n}(x - \theta) \mathrm{d}F(\theta) \right).$$

Similarly, we can replace $p_{F,\sigma}$ by $p_{F,\sigma} 1\!\!1_{[-a_n/2,\, a_n/2]}$. We then have that $q_\gamma^{F,\sigma}$ and $p_{F,\sigma}$ are contiguous and $\|q_\gamma^{F,\sigma} - p_{F,\sigma}\|_1 = o(1/n)$. Therefore, we can consider the same tests as in Lemma 1 of Ghosal and van der Vaart (2007b) and (2.6) is verified, together with (2.7) using Proposition 2 of Shen et al. (2013). Since (5.2) implies condition (2.5), there only remains to verify assumption (2.4). The difficulty here is to control $q_\gamma^{F,\sigma}$ as $\sigma \to 0$. Indeed, $q_\gamma^{F,\sigma}$ can be used as an upper bound on $\psi_{\gamma,\gamma'}(p_{F,\sigma})$ for all (F, σ) with $\sigma > \underline{\sigma}_n$, when $|x| \le a_n/2$, which we are allowed to consider since we can restrict ourselves to the event Ω_n. Thus, $\int_{\sigma > \bar{\sigma}_n} Q_\gamma^{F,\sigma}([-a_n/2,\, a_n/2]^n)\, \mathrm{d}\pi(F \mid \gamma) \mathrm{d}\pi(\sigma) = o(e^{-2n\epsilon_n^2}) \pi(B_n^\gamma \mid \gamma)$ uniformly in $\gamma \in \mathcal{K}_n$. We now study

$$\int_{\sigma < \underline{\sigma}_n} Q_\gamma^{F,\sigma}([-a_n/2,\, a_n/2]^n)\, \mathrm{d}\pi(F \mid \gamma) \mathrm{d}\pi(\sigma).$$

We split $(0, \underline{\sigma}_n) = \cup_{j=0}^\infty [\underline{\sigma}_n 2^{-(j+1)}, \underline{\sigma}_n 2^{-j})$. Then, for all $\sigma \in [\underline{\sigma}_n 2^{-(j+1)}, \underline{\sigma}_n 2^{-j})$, $j \ge 0$, define $u_{n,j} = n^{-1} e_n (\underline{\sigma}_n 2^{-j})^2$, with $e_n = o(1)$ and $\beta > 1/2$ to guarantee that $u_{n,j} \le n^{-2}$, so that, similarly to before, for all $\gamma \in \mathcal{K}_n$,

$$\sup_{\|\gamma - \gamma'\| \le u_{n,j}} \psi_{\gamma,\gamma'}(p_{F,\sigma}(x)) 1\!\!1_{[-a_n/2,\, a_n/2]}(x) \le q_\gamma^{F,\sigma}(x),$$

with the distance $\|\gamma - \gamma'\| = |m - m_k| + |s/s_\ell - 1|$. For all $\gamma \in \mathcal{K}_n$, using a $u_{n,j}$-covering of $\{\gamma' : \|\gamma - \gamma'\| \le u_n\}$ with centering points γ_i, $i = 1, \ldots, N_j$, where $N_j \le (u_n/u_{n,j})^2$, we have

$$\begin{aligned} q_\gamma^{F,\sigma}(x) &\le \max_{1 \le i \le N_j} \int_{\mathbb{R}} \sup_{\|\gamma' - \gamma_i\| \le u_{n,j}} \phi_\sigma(x - \psi_{\gamma,\gamma'}(\theta)) \mathrm{d}F(\theta) \\ &\le \max_{1 \le i \le N_j} \int_{\mathbb{R}} \phi_\sigma(x - \psi_{\gamma,\gamma_i}(\theta)) e^{\frac{|x - \psi_{\gamma,\gamma_i}(\theta)| u_{n,j}(1 + |\psi_{\gamma,\gamma_i}(\theta)|)}{\sigma^2}} \mathrm{d}F(\theta) \le \max_{1 \le i \le N_j} c_{n,i} g_{\sigma,i}(x), \end{aligned}$$

with $g_{\sigma,i}$ the probability density over $[-a_n/2,\, a_n/2]$ proportional to

$$\int_{|\theta| \le a_n} \phi_\sigma(x - \psi_{\gamma,\gamma_i}(\theta)) e^{\frac{2a_n u_{n,j}(1 + a_n)}{\sigma^2}} \mathrm{d}F(\theta) + \int_{|\theta| > a_n} \phi_\sigma((x - \psi_{\gamma,\gamma_j}(\theta))(1 - o(1/n))) \mathrm{d}F(\theta),$$

so that

$$c_{n,i} \le F([-a_n,\, a_n]) e^{16 x_0 e_n (\log n)^{1/\tau}/n} + (1 - F([-a_n,\, a_n]))(1 + o(1/n^{1/2})) = 1 + O(e_n (\log n)^{1/\tau}/n).$$

This implies that, for $e_n \le (\log n)^{-1/\tau}$,

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

whence $\int_{\sigma < \underline{\sigma}_n} Q_\gamma^{F,\sigma}([-a_n/2,\, a_n/2]^n) \mathrm{d}\pi(F \mid \gamma) \mathrm{d}\pi(\sigma) \lesssim e^{-\underline{\sigma}_n^{-1}/2}$ and (2.4) is verified, which completes the proof of Theorem 2.

5.2 Proof of Theorem 3 on empirical Bayes Dirichlet process mixtures of Gaussian densities for the super-smooth case

The main difference with the ordinary smooth case lies in the fact that, since the rate ϵ_n is almost parametric, we have $n\epsilon_n^2 = O((\log n)^\kappa)$ for a suitable finite constant $\kappa > 0$ so that, in order to compensate the number $N_n(u_n)$ of points, we need that, for some set $\tilde{B}_n$,

$$\sup_{\gamma\in\mathcal{K}_n}\sup_{p_{F,\sigma}\in\tilde{B}_n}\mathbb{P}_{p_0}^{(n)}\left(\inf_{\|\gamma-\gamma'\|\le u_n}\ell_n(\psi_{\gamma,\gamma'}(p_{F,\sigma}))-\ell_n(p_0)<-n\epsilon_n^2\right)\lesssim e^{-Cn\epsilon_n^2} \tag{5.4}$$

for some constant $C > 0$. It is known from Lemma 2 of Shen and Wasserman (2001) that if $p_{F,\sigma} \in S_n = \{p_{F,\sigma} : \rho_\alpha(p_0; p_{F,\sigma}) \le \epsilon_n^2\}$ then

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

Consider the same set $\mathcal{K}_n$ and the same u_n-covering as in the proof of Theorem 2. Take $\sigma_n = O((\log n)^{-1/r})$. Let $\sigma \in (\sigma_n,\, \sigma_n + e^{-d_1(1/\sigma_n)^r})$ for some positive constant d_1. It is known from Lemma 6.3 of Scricciolo (2014) that there exists a distribution $F^* = \sum_{j=1}^{N_\sigma} p_j^*\delta_{\theta_j^*}$, with $N_\sigma = O((a_\sigma/\sigma)^2)$ points in $[-a_\sigma,\, a_\sigma]$, where $a_\sigma = O(\sigma^{-r/(\tau\wedge 2)})$, such that, for some constant $c > 0$, $n^{-1}\max\{\mathrm{KL}(p_0; p_{F^*,\sigma}),\, V_2(p_0; p_{F^*,\sigma})\} \lesssim e^{-c(1/\sigma)^r}$. Inspection of the proof of Lemma 6.3 reveals that all arguments remain valid to bound above any ρ_α-divergence $\rho_\alpha(p_0; p_{F^*,\sigma})$ for $\alpha \in (0,\, 1]$. In fact, using the inequality valid for all $a,\, b > 0$, $|a^\alpha - b^\alpha| \le |a^\beta - b^\beta|^{\alpha/\beta}$, with $0 \le \alpha \le \beta$, having set in our case $\beta = 1$, we have $\rho_\alpha(p_0; p_{F^*,\sigma}) \le \alpha^{-1}\mathbb{E}_{p_0}\{[|p_0(X_1) - p_{F^*,\sigma}(X_1)|/p_{F^*,\sigma}(X_1)]^\alpha\}$. All bounds used in the proof of Lemma 6.3 for the various pieces in which $n^{-1}\mathrm{KL}(p_0; p_{F^*,\sigma})$ is split can be used here to bound above $\rho_\alpha(p_0; p_{F^*,\sigma})$. Thus, $\rho_\alpha(p_0; p_{F^*,\sigma}) \lesssim e^{-c(1/\sigma)^r}$. 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 $\lambda(U_j) = O(e^{-c_1(1/\sigma)^r})$, $j = 1, \dots, N_\sigma$. 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

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

we have $\inf_{\gamma\in\mathcal{K}_n}\pi(B_n \mid \gamma) \gtrsim \exp\{-c_2N_{\sigma_n}(1/\sigma_n)^r\} = \exp\{-c_3(\log n)^{5+6/r}\} = e^{-c_3n\epsilon_n^2}$ and, for every $(F,\, \sigma) \in B_n$, it results $\rho_\alpha(p_0; p_{F,\sigma}) \lesssim e^{-c_2(1/\sigma_n)^r} \lesssim \epsilon_n^2$. Moreover, reasoning as in the proof of Theorem 2, for $u_n \lesssim k_n^{-1}\sigma_n^2\epsilon_n^2(\log n)^{-1/(\tau\wedge 2)}$, on the event $A_n = \{\sum_{i=1}^n |X_i - m_0| \le n\tau_0^2k_n\}$, for $k_n = O(n)$,

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

for some positive constant C', with $p_{F_n,\sigma}(\cdot) = c_\sigma^{-1}\sum_{j\ge1}\mathbb{1}_{|\theta'_j|\le a_\sigma}p_j\phi_\sigma(x - \theta'_j)$, where $c_\sigma = \sum_{j:\,|\theta'_j|\le a_\sigma}p_j \ge (1 - e^{-c_1(1/\sigma_n)^r}) \ge 1 - \epsilon_n^2$. Hence, the proof of Lemma 2 of Shen and Wasserman (2001) implies that (5.4) is satisfied. The other parts of the proof of Theorem 2 go through to this case. We only need to verify that both $\mathbb{P}_{p_0}^{(n)}(A_n^c)$ and $\mathbb{P}_{p_0}^{(n)}(\Omega_n^c)$ go to 0. Indeed, by Markov's inequality, $\mathbb{P}_{p_0}^{(n)}(A_n^c) \le n^{-1} \le e^{-c_3n\epsilon_n^2}$. Also, $\mathbb{P}_{p_0}^{(n)}(\Omega_n^c) \lesssim e^{-c_4n\epsilon_n^2}$ because of the assumption (3.3) that p_0 has exponentially small tails.

5.3 Proof of Corollary 1 on empirical Bayes posterior contraction rates for mixing distributions in Wasserstein metrics

We appeal to Corollary 1 in Scricciolo (2014) and the following remark which gives an indication on how to remove the condition that Θ is bounded. In particular, we need that, for every $1 \le q < \infty$, there exist $q < u < \infty$ and $0 < B < \infty$ such that $\mathbb{E}_F[|X|^u] < B$ with $[\mathrm{DP}(\alpha_{\mathbb{R}}\mathrm{N}_{\hat{\gamma}_n})]$-probability one, for almost every sample path when sampling from $\mathbb{P}_{p_0}^{(\infty)}$. This can be proved appealing to the properties of the tails of the distribution functions sampled from a Dirichlet process as in Doss and Sellke (1982).

5.4 Proof of Corollary 2 on adaptive empirical Bayes density deconvolution

The result is based on the following inversion inequalities which relate the $\mathbb{L}_2$-distance between the true mixing density and the random approximating mixing density to the $\mathbb{L}_2/\mathbb{L}_1$-distance between the corresponding mixed densities:

$$\|p_Y - p_{0Y}\|_2 \lesssim \begin{cases} \|K * p_Y - K * p_{0Y}\|_2^{\beta/(\beta+\eta)}, & \text{ordinary smooth case,}\\ (-\log \|K * p_Y - K * p_{0Y}\|_1)^{-\beta/r}, & \text{super-smooth case.}\end{cases}$$

In what follows, we use "os" and "ss" as short-hands for "ordinary smooth" and "super-smooth", respectively. To prove the preceding inequalities, we instrumentally use the *sinc* kernel to characterize regular densities in terms of their approximation properties. We recall that the *sinc* kernel

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

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

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

Let $\widehat{g_\delta} = \widehat{\mathrm{sinc}_\delta}/\widehat{K}$ be the Fourier transform of g_δ. So, $\mathrm{sinc}_\delta = K * g_\delta$ and $p_Y * \mathrm{sinc}_\delta = (p_Y * K) * g_\delta = (K * p_Y) * g_\delta$. We then have

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

because $\|p_{0Y} - p_{0Y} * \mathrm{sinc}_\delta\|_2^2 = \int_{-\infty}^{\infty}|\widehat{p_{0Y}}(t)|^2|1 - \mathrm{sinc}_\delta(t)|^2\mathrm{d}t < \delta^{2\beta}\int_{|t|>1/\delta}(1+t^2)^\beta|\widehat{p_{0Y}}(t)|^2\mathrm{d}t \lesssim \delta^{2\beta}$ by assumption (3.9). Now, recall that $p_Y = p_{F,\sigma} = F * \phi_\sigma$. For $(F,\,\sigma) \in \mathcal{F}_n$ defined as in (5.3), with $\sigma \ge \underline{\sigma}_n \propto C\delta(\log n)^{\kappa_2}$, where $2\kappa_2 \ge 1$, $C^2/2 \ge (2\beta+1)/[2(\beta+\eta)+1]$ and $\delta \gtrsim n^{-1/[2(\beta+\eta)+1]}$, we have

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

In the ordinary smooth case,

$$\begin{aligned}\|p_Y * \mathrm{sinc}_\delta - p_{0Y} * \mathrm{sinc}_\delta\|_2^2 &= \|(K * p_Y) * g_\delta - (K * p_{0Y}) * g_\delta\|_2^2\\ &\le \delta^{-2\eta}\int_{|t|\le 1/\delta}|\widehat{K}(t)|^2|\widehat{p_Y}(t) - \widehat{p_{0Y}}(t)|^2\mathrm{d}t\\ &\le \delta^{-2\eta}\|K * p_Y - K * p_{0Y}\|_2^2.\end{aligned}$$

In the super-smooth case,

$$\|p_Y * \operatorname{sinc}_\delta - p_{0Y} * \operatorname{sinc}_\delta\|_2^2 = \|(K * p_Y) * g_\delta - (K * p_{0Y}) * g_\delta\|_2^2 \le \|K * p_Y - K * p_{0Y}\|_1^2 \|g_\delta\|_2^2,$$

where

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

Combining pieces, for $(F,\, \sigma) \in \mathcal{F}_n$,

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

so that the optimal choice for δ is

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

Taking into account that $p_Y = p_{F,\sigma} = F * \phi_\sigma$, for $1 \le q \le \infty$, we have

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

Then,

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

For suitable constants $\tau_1,\, \kappa_1 > 0$, let

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

and let ϵ_n be as in the statement. Hence, for all $(F,\, \sigma) \in \mathcal{F}_n$, the following inclusions hold:

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

For $q = 2$ in the ordinary smooth case and $q = 1$ in the super-smooth case, if $\pi(\{(F,\, \sigma) \in \mathcal{F}_n : \|p_{F*K,\sigma} - K * p_{0Y}\|_q \lesssim \psi_n\} \mid X^{(n)},\, \hat\gamma_n) \to 1$ in $\mathbb{P}_{p_{0X}}^{(n)}$-probability or $\mathbb{P}_{p_{0X}}^{(\infty)}$-almost surely, then also $\pi(\{(F,\, \sigma) : \|p_Y - p_{0Y}\|_2 \lesssim \epsilon_n\} \mid X^{(n)},\, \hat\gamma_n) \to 1$ in the same mode of convergence and the proof is complete. $\square$

5.5 Proofs for Aalen models

For any intensity λ, we still denote $M_\lambda = \int_\Omega \lambda(t)\mathrm{d}t$ and $\bar\lambda = M_\lambda^{-1} \times \lambda \in \mathcal{F}_1$. To prove Theorem 4 and Theorem 5, we need the following intermediate results that are based on classical tools subsequently defined. Recall that $\bar\epsilon_n = (\log n/n)^{1/3}$ and set $\epsilon_n = J_1\bar\epsilon_n$ for some $J_1 > 0$ large enough. The first result controls the Kullback-Leibler divergence and moments of $\ell_n(\lambda_0) - \ell_n(\lambda)$, where $\ell_n(\lambda)$ is the log-likelihood evaluated at λ, whose expression is given in (4.5).

Proposition 1 *Let v_n be a positive sequence converging to 0 such that $nv_n^2 \to \infty$. For $H > 0$, let*

$$B_{k,n}(H,\,\lambda_0,\,v_n) = \{\lambda : \ \bar{\lambda} \in \bar{B}_{k,n}(H,\,\bar{\lambda}_0,\,v_n),\ \ |M_\lambda - M_{\lambda_0}| \le v_n\}.$$

Then, for all $k \ge 2$ and $\lambda \in B_{k,n}(H,\,\lambda_0,\,v_n)$, under (4.2),

$$\mathrm{KL}(\lambda_0;\,\lambda) \le \kappa_0 n v_n^2, \quad V_k(\lambda_0;\,\lambda) \le \kappa (n v_n^2)^{\frac{k}{2}}$$

where κ, κ_0 depend only on k, C_{1k}, H, λ_0, m_1 and m_2; an expression of κ_0 is given in (6.5).

The second result establishes the existence of tests that control the numerator of posterior distributions.

Proposition 2 *Assume that conditions (i) and (ii) of Theorem 5 are satisfied. For any positive integer j, define*

$$S_{n,j}(v_n) = \{\lambda : \ \bar{\lambda} \in \mathcal{F}_n \ \text{and}\ jv_n \le \|\lambda - \lambda_0\|_1 \le (j+1)v_n\}.$$

Then, under (4.2)*, there exist J_0, ρ, $c > 0$ such that, for all $j \ge J_0$, there exists a test $\phi_{n,j} \in [0,1]$ such that*

$$\mathbb{E}_{\lambda_0}^{(n)}[1_{\Gamma_n}\phi_{n,j}] \le Ce^{-cnj^2v_n^2}, \qquad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_\lambda[1_{\Gamma_n}(1-\phi_{n,j})] \le Ce^{-cnj^2v_n^2} \quad \text{if } j \le \rho/v_n$$

$$\mathbb{E}_{\lambda_0}^{(n)}[1_{\Gamma_n}\phi_{n,j}] \le Ce^{-cnjv_n}, \qquad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_\lambda[1_{\Gamma_n}(1-\phi_{n,j})] \le Ce^{-cnjv_n} \quad \text{if } j \ge \rho/v_n$$

for C a constant.

5.5.1 Proof of Theorem 4

Proof. Without loss of generality, we 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 x,

$$\bar{\lambda}(x) = \sum_{j=1}^{\infty} p_j \frac{1\!\!1_{(0,\,\theta_j)}(x)}{\theta_j}, \quad \psi_{\gamma,\gamma'}(\bar{\lambda})(x) = \sum_{j=1}^{\infty} p_j \frac{1\!\!1_{(0,\,G_{\gamma'}^{-1}(G_\gamma(\theta_j)))}(x)}{G_{\gamma'}^{-1}(G_\gamma(\theta_j))},$$

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 Dirichlet process mixture of uniform distributions with base measure indexed by γ, then $\psi_{\gamma,\gamma'}(\bar{\lambda})$ is distributed according to a Dirichlet process mixture of uniform distributions with base measure indexed by γ'. We prove Theorem 4 for both types of base measure introduced in (4.7). Let G denote the cumulative distribution function of a $\Gamma(a,\,1)$ random variable and g its density. Then, for the first type of base measure we have

$$G_{\gamma'}^{-1}(G_\gamma(\theta)) = \frac{G^{-1}\left(G(\gamma\theta)G(\gamma' T)/G(\gamma T)\right)}{\gamma'} \quad \text{if } \theta \le T,$$

and $G_{\gamma'}^{-1}(G_\gamma(\theta)) = T$ if $\theta \ge T$. Therefore, for any $\theta \in [0,\,T]$, if $\gamma' \ge \gamma$ then

$$G_{\gamma'}^{-1}(G_\gamma(\theta)) \le \theta, \quad G_{\gamma'}^{-1}(G_\gamma(\theta)) \ge \frac{\gamma\theta}{\gamma'}. \tag{5.5}$$

The second inequality of (5.5) is straightforward. The first inequality of (5.5) is equivalent to $G(\gamma\theta)G(\gamma' T) \leq G(\gamma'\theta)G(\gamma T)$ 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

$$\Delta'(\theta) = \gamma g(\gamma\theta)G(\gamma' T) - \gamma' g(\gamma'\theta)G(\gamma T)$$

so 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 the brackets is increasing so that $\Delta'(\theta) \leq 0$ for $\theta \leq c$ and $\Delta'(\theta) \geq 0$ for $\theta \geq c$. 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 (5.5). For the second type of base measure, we have for $\theta \leq T$,

$$\forall\, \gamma,\, \gamma' > 0, \quad G_{\gamma'}^{-1}(G_\gamma(\theta)) = \frac{T\gamma}{\gamma'} \frac{\theta}{(T - \theta + \theta\gamma/\gamma')}$$

and (5.5) 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 $u_n = 1/(n\log n)$, so that $u_n = o(\bar\epsilon_n^2)$ and choose $k \geq 6$ so that $u_n^{-1} = o((n\bar\epsilon_n^2)^{k/2})$ and (2.2) holds. For κ_0 given in Proposition 1, we control $\mathbb{P}_{\lambda_0}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0+1)n\bar\epsilon_n^2\right)$. We follow most of the computations of Salomond (2013). Let $e_n = (n\bar\epsilon_n^2)^{-k/2}$ and set

$$\bar\lambda_{0n}(x) = \frac{\lambda_0(x)1\!\!1_{x\leq\theta_n}}{\int_0^{\theta_n}\lambda_0(x)\mathrm{d}x} \quad \text{with} \quad \theta_n = \inf\left\{\theta : \int_0^\theta \bar\lambda_0(x)\mathrm{d}x \geq 1 - \frac{e_n}{n}\right\}$$

and $\lambda_{0n} = M_{\lambda_0}\bar\lambda_{0n}$. Define the event $A_n = \{\forall\, X \in N,\, X \leq \theta_n\}$. We shall need the following result. Given $\tilde N$ a Poisson process with intensity $n(1+\alpha)m_2\lambda_0$. If $\tilde N[0, T] = k$, we denote $\tilde N = \{X_1, \ldots, X_k\}$ and conditionally on $\tilde N[0, T] = k$, $X_1, \ldots, X_k$ are i.i.d. with density $\bar\lambda_0$. So,

$$\begin{aligned}
\mathbb{P}_{\lambda_0}^{(n)}(A_n^c \mid \Gamma_n) &\leq \sum_{k=1}^\infty \mathbb{P}_{\lambda_0}^{(n)}(\exists\, X_i > \theta_n \mid \tilde N[0, T] = k)\mathbb{P}_{\lambda_0}^{(n)}(\tilde N[0, T] = k) \\
&\leq \sum_{k=1}^\infty \left(1 - \left(\int_0^{\theta_n}\bar\lambda_0(t)\mathrm{d}t\right)^k\right)\mathbb{P}_{\lambda_0}^{(n)}(\tilde N[0, T] = k) \\
&\leq \sum_{k=1}^\infty \left(1 - \left(1 - \frac{e_n}{n}\right)^k\right)\mathbb{P}_{\lambda_0}^{(n)}(\tilde N[0, T] = k) \\
&= O\left(\frac{e_n}{n}\mathbb{E}_{\lambda_0}^{(n)}[\tilde N[0, T]]\right) = O(e_n) = O((n\bar\epsilon_n^2)^{-k/2}).
\end{aligned}$$

Now, we have

$$\mathbb{P}_{\lambda_0}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0+2)n\bar\epsilon_n^2 \mid \Gamma_n\right) \leq \mathbb{P}_{\lambda_0}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0+2)n\bar\epsilon_n^2 \mid A_n, \Gamma_n\right) + \mathbb{P}_{\lambda_0}^{(n)}(A_n^c \mid \Gamma_n).$$

We now deal with the first term. On $\Gamma_n \cap A_n$,

$$
\begin{aligned}
\ell_n(\lambda_0) &= \ell_n(\lambda_{0n}) + \int_0^{\theta_n} \log\left(\frac{\lambda_0(t)}{\lambda_{0n}(t)}\right) \mathrm{d}N_t - \int_0^T (\lambda_0(t) - \lambda_{0n}(t)) Y_t \mathrm{d}t \\
&= \ell_n(\lambda_{0n}) + N[0,\, T] \log\left(\int_0^{\theta_n} \bar\lambda_0(t)\mathrm{d}t\right) - M_{\lambda_0} \int_0^T \bar\lambda_0(t) Y_t \mathrm{d}t + M_{\lambda_0} \frac{\int_0^{\theta_n} \bar\lambda_0(t) Y_t \mathrm{d}t}{\int_0^{\theta_n} \bar\lambda_0(t)\mathrm{d}t} \\
&\le \ell_n(\lambda_{0n}) + M_{\lambda_0} \frac{\int_{\theta_n}^T \bar\lambda_0(t)\mathrm{d}t \int_0^{\theta_n} \bar\lambda_0(t) Y_t \mathrm{d}t}{\int_0^{\theta_n} \bar\lambda_0(t)\mathrm{d}t} - M_{\lambda_0} \int_{\theta_n}^T \bar\lambda_0(t) Y_t \mathrm{d}t \\
&\le \ell_n(\lambda_{0n}) + M_{\lambda_0} \frac{e_n(1+\alpha)m_2}{1 - e_n/n}.
\end{aligned}
$$

So, for n large enough, for any λ,

$$
\begin{aligned}
\mathbb{P}_{\lambda_0}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_0) \le -(\kappa_0+2)n\bar\epsilon_n^2 \mid A_n,\, \Gamma_n\right) &\le \mathbb{P}_{\lambda_0}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_{0n}) \le -(\kappa_0+1)n\bar\epsilon_n^2 \mid A_n,\, \Gamma_n\right) \\
&= \mathbb{P}_{\lambda_{0n}}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_{0n}) \le -(\kappa_0+1)n\bar\epsilon_n^2 \mid \Gamma_n\right)
\end{aligned}
$$

because $\mathbb{P}_{\lambda_0}^{(n)}(\cdot \mid A_n) = \mathbb{P}_{\lambda_{0n}}^{(n)}(\cdot)$. For all $\lambda = M_\lambda \bar\lambda \in B_{k,n}(H,\, \lambda_{0n},\, \bar\epsilon_n)$, using Proposition 1, we obtain

$$
\mathbb{P}_{\lambda_{0n}}^{(n)}\left(\ell_n(\lambda) - \ell_n(\lambda_{0n}) \le -(\kappa_0+1)n\bar\epsilon_n^2 \mid \Gamma_n\right) = O((n\bar\epsilon_n^2)^{-k/2}). \tag{5.6}
$$

To prove (2.3), we need to control $\inf_{\gamma' \in [\gamma,\, \gamma+u_n]} \ell_n(M_\lambda \psi_{\gamma,\gamma'}(\bar\lambda))$. Using (5.5), we have for any $\gamma' \in [\gamma,\, \gamma+u_n]$, on Γ_n,

$$
\frac{\gamma}{\gamma+u_n}\psi_{\gamma,\gamma+u_n}(\bar\lambda)(t) \le \psi_{\gamma,\gamma'}(\bar\lambda)(t) \le \psi_{\gamma,\gamma+u_n}(\bar\lambda)(t) + \sum_{j=1}^\infty p_j \frac{1\!\!1_{\left(\frac{\gamma}{\gamma+u_n} G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j)),\, G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j))\right)}(t)}{G_{\gamma+u_n}^{-1}(G_\gamma(\theta_j))}
$$

so that for n large enough,

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

and, on the event $\{N[0,\, T] \le 2M_{\lambda_0} m_2 n\}$ which has probability going to 1,

$$
\inf_{\gamma' \in [\gamma,\, \gamma+u_n]} \ell_n(M_\lambda \psi_{\gamma,\gamma'}(\bar\lambda)) \ge \ell_n(M_\lambda \psi_{\gamma,\gamma+u_n}(\bar\lambda)) - n\gamma^{-1}(M_\lambda(1+\alpha)m_2 + 4m_2 M_{\lambda_0})u_n. \tag{5.7}
$$

Combining this lower bound with (5.6), we obtain that, for all $\lambda = M_\lambda \psi_{\gamma,\gamma+u_n}(\bar\lambda)$, with $\psi_{\gamma,\gamma+u_n}(\bar\lambda) \in \bar B_{k,n}(H,\, \bar\lambda_{0n},\, \bar\epsilon_n)$ and $|M_\lambda - M_{\lambda_0}| \le \bar\epsilon_n$,

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

and assumption [A1] is satisfied if $J_1^2 \ge \kappa_0 + 2$. We now verify assumption [A2]. First, mimicking the proof of Lemma 8 of Salomond (2013), we have that over any compact subset $\mathcal{K}'$ of $(0,\, \infty)$,

$$
\inf_{\gamma \in \mathcal{K}'} \pi_1\left(\bar B_{k,n}(H,\, \bar\lambda_{0n},\, \bar\epsilon_n) \mid \gamma\right) \ge e^{-C_k n\bar\epsilon_n^2} \tag{5.8}
$$

for some $C_k > 0$, when n is large enough. Let $B_n^\gamma = \{\lambda :\ \psi_{\gamma,\gamma+u_n}(\bar\lambda) \in \bar B_{k,n}(H,\, \bar\lambda_{0n},\, \bar\epsilon_n),\, M_\lambda \in [M_{\lambda_0} - \bar\epsilon_n,\, M_{\lambda_0} + \bar\epsilon_n]\}$. Then

$$\pi(B_n^\gamma \mid \gamma) = \pi_1(\bar B_{k,n}(H,\, \bar\lambda_{0n},\, \bar\epsilon_n) \mid \gamma + u_n)\pi_M([M_{\lambda_0} - \bar\epsilon_n,\, M_{\lambda_0} + \bar\epsilon_n])$$

which, together with (5.8), implies that

$$\inf_{\gamma\in\mathcal{K}} \pi(B_n^\gamma \mid \gamma) \geq e^{-2C_k n\bar\epsilon_n^2}$$

when n is large enough, so that (2.5) is satisfied as soon as J_1 is large enough. We now control the measure $Q_{\gamma,n}^\lambda$, with

$$\mathrm{d}Q_{\gamma,n}^\lambda = 1_{\Gamma_n} \times \sup_{\|\gamma'-\gamma\|\leq u_n} \exp(\ell_n(M_\lambda \psi_{\gamma,\gamma'}(\bar\lambda)))\mathrm{d}\mu$$

and μ is the measure such that under μ the process is an homogeneous Poisson process with intensity 1. Using (5.5) and similarly to (5.7), we obtain that, for all $\gamma' \in [\gamma,\, \gamma + u_n]$,

$$\bar\lambda(t) - \sum_{j=1}^\infty p_j \frac{\mathbb{1}_{\left(\frac{\gamma\theta_j}{\gamma'},\, \theta_j\right)}(t)}{\theta_j} \leq \psi_{\gamma,\gamma'}(\bar\lambda)(t) \leq \frac{\gamma+u_n}{\gamma}\bar\lambda(t)$$

and we have

$$\begin{aligned}
Q_{\gamma,n}^\lambda(\mathcal{X}^{(n)}) &= \mathbb{E}_\mu^{(n)}\left[1_{\Gamma_n} \sup_{\gamma'\in[\gamma,\,\gamma+u_n]} \exp\left(T - M_\lambda \int_0^T \psi_{\gamma,\gamma'}(\bar\lambda)(t) Y_t \mathrm{d}t + \int_0^T \log\left(M_\lambda \psi_{\gamma,\gamma'}(\bar\lambda)(t)\right)\mathrm{d}N_t\right)\right] \\
&\leq \mathbb{E}_\lambda^{(n)}\left[1_{\Gamma_n} \exp(nm_2(1+\alpha)M_\lambda u_n/(\gamma+u_n) + \log(1+u_n/\gamma)N[0,\,T])\right] \\
&\leq \mathbb{E}_\lambda^{(n)}\left[1_{\Gamma_n} \exp(nm_2(1+\alpha)M_\lambda \gamma^{-1}u_n + u_n\gamma^{-1}N[0,\,T])\right] \\
&\leq \exp\left(nm_2(1+\alpha)M_\lambda\gamma^{-1}u_n + (1+\alpha)nm_2M_\lambda(e^{u_n/\gamma}-1)\right) \\
&\leq \exp\left(3nm_2(1+\alpha)\gamma^{-1}M_\lambda u_n\right)
\end{aligned}$$

when n is large enough. Let $\phi_{n,j}$ be the tests defined in Proposition 2 over $S_{n,j}(\bar\epsilon_n)$. Using the previous computations, we have

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

As in Salomond (2013), we set $\mathcal{F}_n = \{\bar\lambda :\ \bar\lambda(0) \leq M_n\}$ with $M_n = \exp\left(c_1 n\epsilon_n^2\right)$ and c_1 is a positive constant. From Lemma 9 of Salomond (2013), there exists $a > 0$ such that $\sup_{\gamma\in\mathcal{K}'} \pi_1(\mathcal{F}_n^c|\gamma) \leq e^{-c_1(a+1)n\epsilon_n^2}$, so that when n is large enough,

$$\begin{aligned}
\sup_{\gamma\in\mathcal{K}'} \int_{\mathbb{R}^+}\int_{\mathcal{F}_n^c} Q_{\gamma,n}^\lambda(\mathcal{X}^{(n)})d\pi_1(\bar\lambda|\gamma)\pi_M(M_\lambda)dM_\lambda &\lesssim e^{-c_1(a+1)n\epsilon_n^2}\int_{\mathbb{R}^+}\exp\left(\delta M_\lambda\right)\pi_M(M_\lambda)dM_\lambda \\
&\lesssim e^{-c_1(a+1)n\epsilon_n^2},
\end{aligned}$$

with δ that can be chosen as small as needed since $nu_n = o(1)$. This proves (2.4) by choosing c_1 conveniently. Combining the above upper bound with Proposition 2, together with Remark 1, achieves the proof of Theorem 4. □

5.5.2 Proof of Theorem 5

The proof of Theorem 5 is similar to the proof of Theorem 1 of Ghosal and van der Vaart (2007a) once Proposition 1 and Proposition 2 are proved. Let $U_n = \{\lambda : \ \|\lambda - \lambda_0\|_1 \geq J_1 v_n\}$ and recall that

$$\pi(U_n \mid N) = \frac{\int_{U_n} e^{\ell_n(\lambda)-\ell_n(\lambda_0)} \mathrm{d}\pi(\lambda)}{\int_{\mathcal{F}} e^{\ell_n(\lambda)-\ell_n(\lambda_0)} \mathrm{d}\pi(\lambda)},$$

where $\ell_n(\lambda)$ is the log-likelihood evaluated at λ. We use notations of Proposition 1 and Proposition 2. Writing

$$D_n = \int_{\mathcal{F}} e^{\ell_n(\lambda)-\ell_n(\lambda_0)} \mathrm{d}\pi(\lambda),$$

we have

$$\begin{aligned}
&\mathbb{P}_{\lambda_0}^{(n)} \left(D_n \leq e^{-(\kappa_0+1)nv_n^2} \pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n)) \right) \\
&\leq \mathbb{P}_{\lambda_0}^{(n)} \left(\int_{B_{k,n}(H,\lambda_0,v_n)} \frac{\exp[\ell_n(\lambda) - \ell_n(\lambda_0)] d\pi(\lambda)}{\pi(B_{k,n}(H, \lambda_0, v_n))} \leq -(\kappa_0 + 1)nv_n^2 + \log \left(\frac{\pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n))}{\pi(B_{k,n}(H, \lambda_0, v_n))} \right) \right).
\end{aligned}$$

Note that since $v_n^2 \geq \log n/n$,

$$\pi(B_{k,n}(H, \lambda_0, v_n)) \gtrsim \pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n))v_n \gtrsim \pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n))e^{-nv_n^2/2}$$

so that using Proposition 1 and the Markov inequality, we obtain that

$$\mathbb{P}_{\lambda_0}^{(n)} \left(D_n \leq e^{-(\kappa_0+1)nv_n^2} \pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n)) \right) \lesssim (nv_n^2)^{-k/2}.$$

Moreover, (6.7) implies that $\pi(S_{n,j}(v_n)) \leq \pi_1(\bar{S}_{n,j})$ and using the tests $\phi_{n,j}$ of Proposition 2, we have for $J_1 \geq J_0$, mimicking the proof of Theorem 1 of Ghosal and van der Vaart (2007a),

$$\begin{aligned}
\mathbb{E}_{\lambda_0}^{(n)} \left[1_{\Gamma_n} \pi \left(\lambda : \ \|\lambda - \lambda_0\|_1 \geq J_1 v_n | N \right) \right] &\leq \sum_{j \geq J_1} \mathbb{E}_{\lambda_0}^{(n)} \left[1_{\Gamma_n} \phi_{n,j} \right] + \sum_{j=J_1}^{\lfloor \rho/v_n \rfloor} e^{(\kappa_0+1)nv_n^2} \frac{\pi_1(\bar{S}_{n,j}) e^{-cnj^2 v_n^2}}{\pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n))} \\
&+ \sum_{j > \rho/v_n} \frac{e^{(\kappa_0+1)nv_n^2} \pi_1(\bar{S}_{n,j}) e^{-cnjv_n}}{\pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n))} + \frac{e^{(\kappa_0+1)nv_n^2} \pi_1(\mathcal{F}_n^c)}{\pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n))} + \mathbb{P}_{\lambda_0}^{(n)} \left(D_n \leq e^{-(\kappa_0+1)nv_n^2} \pi_1(\bar{B}_{k,n}(H, \bar{\lambda}_0, v_n)) \right) \\
&\lesssim (nv_n^2)^{-k/2},
\end{aligned}$$

which proves the result since $\mathbb{P}_{\lambda_0}^{(n)}(\Gamma_n^c) = o(1)$.

6 Appendix

We use in the sequel that for any densities f and g, $\|f - g\|_1 \leq 2h(f, g)$.

6.1 Proof of Proposition 1

We recall that the log-likelihood evaluated at λ is given by

$$\ell_n(\lambda) = \int_0^T \log(\lambda(t)) \mathrm{d}N_t - \int_0^T \lambda(t) Y_t \mathrm{d}t,$$

see Andersen et al. (1993). Since on Ω^c, N is empty and $Y_t \equiv 0$ almost surely, in the sequel we assume, without loss of generality, that $\Omega = [0,\, T]$. We denote by

$$M_n(\lambda) = \int_0^T \lambda(x)\mu_n(x)\mathrm{d}x, \quad M_n(\lambda_0) = \int_0^T \lambda_0(x)\mu_n(x)\mathrm{d}x,$$

$$\bar\lambda_n(x) = \frac{\lambda(x)\mu_n(x)}{M_n(\lambda)} = \frac{\bar\lambda(x)\tilde\mu_n(x)}{\int_0^T \bar\lambda(t)\tilde\mu_n(t)\mathrm{d}t} \quad \text{and} \quad \bar\lambda_{0,n}(x) = \frac{\lambda_0(x)\mu_n(x)}{M_n(\lambda_0)} = \frac{\bar\lambda_0(x)\tilde\mu_n(x)}{\int_0^T \bar\lambda_0(t)\tilde\mu_n(t)\mathrm{d}t}.$$

Then, by using straightforward computations,

$$\begin{aligned}
\mathrm{KL}(\lambda_0;\, \lambda) = \mathbb{E}^{(n)}_{\lambda_0}[\ell_n(\lambda_0) - \ell_n(\lambda)] &= M_n(\lambda_0)\left(\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) + \frac{M_n(\lambda)}{M_n(\lambda_0)} - 1 - \log\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)\right)\\
&= M_n(\lambda_0)\left[\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) + \phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)\right]\\
&\le n m_2 M_{\lambda_0}\left[\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) + \phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)\right],
\end{aligned} \tag{6.1}$$

where $\phi(x) = x - 1 - \log x$ and

$$\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) = \int_0^T \log\left(\frac{\bar\lambda_{0,n}(t)}{\bar\lambda_n(t)}\right)\bar\lambda_{0,n}(t)\mathrm{d}t.$$

Now, we control $\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n)$ for $\lambda \in B_{k,n}(H, \lambda_0, v_n)$. By using Lemma 8.2 of Ghosal et al. (2000), we have

$$\begin{aligned}
\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) &\le 2h^2(\bar\lambda_{0,n},\, \bar\lambda_n)\left(1 + \log\left\|\frac{\bar\lambda_{0,n}}{\bar\lambda_n}\right\|_\infty\right)\\
&\le 2h^2(\bar\lambda_{0,n},\, \bar\lambda_n)\left(1 + \log\left(\frac{m_2}{m_1}\right) + \log\left\|\frac{\bar\lambda_0}{\bar\lambda}\right\|_\infty\right)\\
&\le 2\left(1 + \log\left(\frac{m_2}{m_1}\right)\right)h^2(\bar\lambda_{0,n},\, \bar\lambda_n)\left(1 + \log\left\|\frac{\bar\lambda_0}{\bar\lambda}\right\|_\infty\right)
\end{aligned} \tag{6.2}$$

since $1 + \log(m_2/m_1) \ge 1$. We now deal with $h^2(\bar\lambda_{0,n},\, \bar\lambda_n)$. We have

$$\begin{aligned}
h^2(\bar\lambda_{0,n},\, \bar\lambda_n) &= \int\left(\sqrt{\bar\lambda_{0,n}(x)} - \sqrt{\bar\lambda_n(x)}\right)^2\mathrm{d}x = \int\left(\sqrt{\frac{\bar\lambda_0(x)\tilde\mu_n(x)}{\int\bar\lambda_0(t)\tilde\mu_n(t)\mathrm{d}t}} - \sqrt{\frac{\bar\lambda(x)\tilde\mu_n(x)}{\int\bar\lambda(t)\tilde\mu_n(t)\mathrm{d}t}}\right)^2\mathrm{d}x\\
&\le 2m_2\int\left(\sqrt{\frac{\bar\lambda_0(x)}{\int\bar\lambda_0(t)\tilde\mu_n(t)\mathrm{d}t}} - \sqrt{\frac{\bar\lambda_0(x)}{\int\bar\lambda(t)\tilde\mu_n(t)\mathrm{d}t}}\right)^2 dx\\
&\quad +2m_2\int\left(\sqrt{\frac{\bar\lambda_0(x)}{\int\bar\lambda(t)\tilde\mu_n(t)\mathrm{d}t}} - \sqrt{\frac{\bar\lambda(x)}{\int\bar\lambda(t)\tilde\mu_n(t)\mathrm{d}t}}\right)^2\mathrm{d}x\\
&\le 2m_2 U_n + \frac{2m_2}{m_1}h^2(\bar\lambda_0,\, \bar\lambda),
\end{aligned}$$

with

$$U_n = \left(\sqrt{\frac{1}{\int\bar\lambda_0(x)\tilde\mu_n(x)\mathrm{d}x}} - \sqrt{\frac{1}{\int\bar\lambda(x)\tilde\mu_n(x)\mathrm{d}x}}\right)^2.$$

We denote by

$$\tilde{\epsilon}_n := \frac{1}{\int \bar{\lambda}_0(x)\tilde{\mu}_n(x)\mathrm{d}x} \int (\bar{\lambda}(x) - \bar{\lambda}_0(x))\tilde{\mu}_n(x)\mathrm{d}x,$$

so that

$$|\tilde{\epsilon}_n| \leq \frac{1}{m_1} \int |\bar{\lambda}(x) - \bar{\lambda}_0(x)|\tilde{\mu}_n(x)\mathrm{d}x \leq \frac{2m_2}{m_1} h(\bar{\lambda}_0,\, \bar{\lambda}).$$

Then,

$$U_n = \frac{1}{\int \bar{\lambda}_0(x)\tilde{\mu}_n(x)\mathrm{d}x} \left(1 - \frac{1}{\sqrt{1+\tilde{\epsilon}_n}}\right)^2 \leq \frac{\tilde{\epsilon}_n^2}{4m_1} \leq \frac{m_2^2}{m_1^3} h^2(\bar{\lambda}_0,\, \bar{\lambda}).$$

Finally,

$$h^2(\bar{\lambda}_{0,n},\, \bar{\lambda}_n) \leq \frac{2m_2}{m_1}\left(1 + \frac{m_2^2}{m_1^2}\right) h^2(\bar{\lambda}_0,\, \bar{\lambda}). \tag{6.3}$$

It remains to bound $\phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)$. We have

$$\begin{aligned}
|M_n(\lambda_0) - M_n(\lambda)| &\leq \int_0^T |\lambda(t) - \lambda_0(t)|\mu_n(t)\mathrm{d}t \leq n m_2 \int_0^T |\lambda(t) - \lambda_0(t)|\mathrm{d}t \\
&\leq \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0)\left(M_{\lambda_0}\|\bar{\lambda} - \bar{\lambda}_0\|_1 + |M_\lambda - M_{\lambda_0}|\right) \\
&\leq \frac{m_2 M_n(\lambda_0)}{m_1 M_{\lambda_0}}\left(2M_{\lambda_0} h(\bar{\lambda}, \bar{\lambda}_0) + |M_\lambda - M_{\lambda_0}|\right) \leq \frac{m_2(2M_{\lambda_0}+1)M_n(\lambda_0)v_n}{m_1 M_{\lambda_0}}.
\end{aligned}$$

Since $\phi(u+1) \leq u^2$ if $|u| \leq \frac{1}{2}$, we have for n large enough,

$$\phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right) \leq \frac{m_2^2(2M_{\lambda_0}+1)^2}{m_1^2 M_{\lambda_0}^2} v_n^2. \tag{6.4}$$

Combining (6.1), (6.2), (6.3) and (6.4), we have $\mathrm{KL}(\lambda_0;\, \lambda) \leq \kappa_0 n v_n^2$ for n large enough, with

$$\kappa_0 = m_2^2 M_{\lambda_0}\left(\frac{4}{m_1}\left(1 + \log\left(\frac{m_2}{m1}\right)\right)\left(1 + \frac{m_2^2}{m_1^2}\right) + \frac{m_2(2M_{\lambda_0}+1)^2}{m_1^2 M_{\lambda_0}^2}\right). \tag{6.5}$$

We now deal with

$$V_{2k}(\lambda_0;\, \lambda) = \mathbb{E}_{\lambda_0}^{(n)}[|\ell_n(\lambda_0) - \ell_n(\lambda) - \mathbb{E}_{\lambda_0}^{(n)}[\ell_n(\lambda_0) - \ell_n(\lambda)]|^{2k}],$$

with $k > 1$. In the sequel, we denote by C a constant that may change from line to line. Straightforward computations lead to

$$\begin{aligned}
V_{2k}(\lambda_0;\, \lambda) \;&=\; \mathbb{E}_{\lambda_0}^{(n)}\Bigg[\Bigg| -\int_0^T \left(\lambda_0(t) - \lambda(t) - \lambda_0(t)\log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\right)(Y_t - \mu_n(t))\mathrm{d}t \\
&\qquad\qquad + \int_0^T \log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t - Y_t\lambda_0(t)\mathrm{d}t)\Bigg|^{2k}\Bigg] \\
&\leq\; 2^{2k-1}(A_{2k} + B_{2k}),
\end{aligned}$$

with

$$B_{2k} := \mathbb{E}_{\lambda_0}^{(n)}\left[\left|\int_0^T \log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t - Y_t\lambda_0(t)\mathrm{d}t)\right|^{2k}\right]$$

and, by using (4.8),

$$\begin{aligned}
A_{2k} &:= \mathbb{E}^{(n)}_{\lambda_0}\left[\left|\int_0^T \left(\lambda_0(t)-\lambda(t)-\lambda_0(t)\log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\right)(Y_t-\mu_n(t))\mathrm{d}t\right|^{2k}\right]\\
&\leq \left(\int_0^T \left(\lambda_0(t)-\lambda(t)-\lambda_0(t)\log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\right)^2\mathrm{d}t\right)^k \times \mathbb{E}^{(n)}_{\lambda_0}\left[\left(\int_0^T (Y_t-\mu_n(t))^2\mathrm{d}t\right)^k\right]\\
&\leq 2^{2k-1}C_{1k}n^k\left(A_{2k,1}+A_{2k,2}\right),
\end{aligned}$$

with, for $\lambda\in B_{k,n}(H,\,\lambda_0,v_n)$,

$$\begin{aligned}
A_{2k,1} &:= \left(\int_0^T \lambda_0^2(t)\log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mathrm{d}t\right)^k \leq M_{\lambda_0}^{2k}\|\bar\lambda_0\|_\infty^k\left(\int_0^T \bar\lambda_0(t)\log^2\left(\frac{M_{\lambda_0}\bar\lambda_0(t)}{M_\lambda\bar\lambda(t)}\right)\mathrm{d}t\right)^k\\
&\leq 2^{2k-1}M_{\lambda_0}^{2k}\|\bar\lambda_0\|_\infty^k\left(E_2^k(\bar\lambda_0,\,\bar\lambda)+\left|\log\left(\frac{M_\lambda}{M_{\lambda_0}}\right)\right|^{2k}\right)\\
&\leq C\left(E_2^k(\bar\lambda_0,\,\bar\lambda)+|M_\lambda-M_{\lambda_0}|^{2k}\right)\leq Cv_n^{2k}
\end{aligned}$$

and

$$\begin{aligned}
A_{2k,2} &:= \left(\int_0^T (\lambda_0(t)-\lambda(t))^2\,\mathrm{d}t\right)^k\\
&= \left(\int_0^T \left((M_{\lambda_0}-M_\lambda)\bar\lambda_0(t)-M_\lambda(\bar\lambda(t)-\bar\lambda_0(t))\right)^2\mathrm{d}t\right)^k\\
&\leq 2^{2k-1}\left(\|\bar\lambda_0\|_\infty^k(M_{\lambda_0}-M_\lambda)^{2k}+M_\lambda^{2k}\left(\int_0^T\left(\sqrt{\bar\lambda_0(t)}-\sqrt{\bar\lambda(t)}\right)^2\left(\sqrt{\bar\lambda_0(t)}+\sqrt{\bar\lambda(t)}\right)^2\mathrm{d}t\right)^k\right)\\
&\leq 2^{2k-1}\left(\|\bar\lambda_0\|_\infty^k(M_{\lambda_0}-M_\lambda)^{2k}+2^kM_\lambda^{2k}(\|\bar\lambda_0\|_\infty+\|\bar\lambda\|_\infty)^kh^{2k}(\bar\lambda_0,\,\bar\lambda)\right)\leq Cv_n^{2k}.
\end{aligned}$$

Therefore,

$$A_{2k}\leq C(nv_n^2)^k.$$

To deal with B_{2k}, we set for any $T>0$,

$$M_T:=\int_0^T \log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t-Y_t\lambda_0(t)\mathrm{d}t),$$

so $(M_T)_T$ is a martingale. Using the Burkholder-Davis-Gundy Inequality (see Theorem B.15 in Karr (1991)), there exists a constant $C(k)$ only depending on k such that, since $2k>1$,

$$\mathbb{E}^{(n)}_{\lambda_0}[|M_T|^{2k}]\leq C(k)\mathbb{E}^{(n)}_{\lambda_0}\left[\left|\int_0^T\log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mathrm{d}N_t\right|^k\right].$$

Therefore, for $k>1$,

$$\begin{aligned}
B_{2k} &= \mathbb{E}^{(n)}_{\lambda_0}[|M_T|^{2k}]\\
&\leq 3^{k-1}C(k)\left(\mathbb{E}^{(n)}_{\lambda_0}\left[\left|\int_0^T\log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t-Y_t\lambda_0(t)\mathrm{d}t)\right|^k+\left|\int_0^T\log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(Y_t-\mu_n(t))\lambda_0(t)\mathrm{d}t\right|^k\right.\right.\\
&\qquad\left.\left.+\left|\int_0^T\log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mu_n(t)\lambda_0(t)\mathrm{d}t\right|^k\right]\right)\\
&\leq 3^{k-1}C(k)(B_{k,2}^{(0)}+B_{k,2}^{(1)}+B_{k,2}^{(2)}),
\end{aligned}$$

with

$$\begin{aligned}
B_{k,2}^{(0)} &= \mathbb{E}_{\lambda_0}^{(n)}\left[\left|\int_0^T \log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t - Y_t\lambda_0(t)\mathrm{d}t)\right|^k\right],\\
B_{k,2}^{(1)} &= \mathbb{E}_{\lambda_0}^{(n)}\left[\left|\int_0^T \log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(Y_t - \mu_n(t))\lambda_0(t)\mathrm{d}t\right|^k\right]\\
B_{k,2}^{(2)} &= \left|\int_0^T \log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mu_n(t)\lambda_0(t)\mathrm{d}t\right|^k.
\end{aligned}$$

This can be iterated: we set $J = \min\{j \in \mathbb{N}:\ 2^j \geq k\}$ so that $1 < k2^{1-J} \leq 2$. There exists a constant C_k, only depending on k, such that for

$$B_{k2^{1-j},2^j}^{(1)} = \mathbb{E}_{\lambda_0}^{(n)}\left[\left|\int_0^T \log^{2^j}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(Y_t - \mu_n(t))\lambda_0(t)\mathrm{d}t\right|^{k2^{1-j}}\right]$$

and

$$B_{k2^{1-j},2^j}^{(2)} = \left|\int_0^T \log^{2^j}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mu_n(t)\lambda_0(t)\mathrm{d}t\right|^{k2^{1-j}},$$

$$\begin{aligned}
B_{2k} &\leq C_k\left(\mathbb{E}_{\lambda_0}^{(n)}\left[\left|\int_0^T \log^{2^J}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t - Y_t\lambda_0(t)\mathrm{d}t)\right|^{k2^{1-J}}\right] + \sum_{j=1}^J (B_{k2^{1-j},2^j}^{(1)} + B_{k2^{1-j},2^j}^{(2)})\right)\\
&\leq C_k\left(\left(\mathbb{E}_{\lambda_0}^{(n)}\left[\left|\int_0^T \log^{2^J}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)(\mathrm{d}N_t - Y_t\lambda_0(t)\mathrm{d}t)\right|^2\right]\right)^{k2^{-J}} + \sum_{j=1}^J (B_{k2^{1-j},2^j}^{(1)} + B_{k2^{1-j},2^j}^{(2)})\right)\\
&= C_k\left(\left(\mathbb{E}_{\lambda_0}^{(n)}\left[\int_0^T \log^{2^{J+1}}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)Y_t\lambda_0(t)\mathrm{d}t\right]\right)^{k2^{-J}} + \sum_{j=1}^J (B_{k2^{1-j},2^j}^{(1)} + B_{k2^{1-j},2^j}^{(2)})\right)\\
&= C_k\left(\left(\int_0^T \log^{2^{J+1}}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mu_n(t)\lambda_0(t)\mathrm{d}t\right)^{k2^{-J}} + \sum_{j=1}^J (B_{k2^{1-j},2^j}^{(1)} + B_{k2^{1-j},2^j}^{(2)})\right)\\
&= C_k\left(B_{k2^{-J},2^{J+1}}^{(2)} + \sum_{j=1}^J (B_{k2^{1-j},2^j}^{(1)} + B_{k2^{1-j},2^j}^{(2)})\right).
\end{aligned}$$

Note that for any $1 \leq j \leq J$,

$$\begin{aligned}
B_{k2^{1-j},2^j}^{(1)} &\leq \left(\int_0^T \log^{2^{j+1}}\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\lambda_0^2(t)\mathrm{d}t\right)^{k2^{-j}} \times \mathbb{E}_{\lambda_0}^{(n)}\left[\left(\int_0^T (Y_t - \mu_n(t))^2\mathrm{d}t\right)^{k2^{-j}}\right]\\
&\leq C(M_{\lambda_0}^2\|\bar{\lambda}_0\|_\infty)^{k2^{-j}}\left(\int_0^T \log^{2^{j+1}}\left(\frac{M_{\lambda_0}\bar{\lambda}_0(t)}{M_\lambda\bar{\lambda}(t)}\right)\bar{\lambda}_0(t)\mathrm{d}t\right)^{k2^{-j}} \times n^{k2^{-j}}\\
&\leq C\left(\log^{2^{j+1}}\left(\frac{M_{\lambda_0}}{M_\lambda}\right) + E_{2^{j+1}}(\bar{\lambda}_0, \bar{\lambda})\right)^{k2^{-j}} \times n^{k2^{-j}}\\
&\leq C(nv_n^2)^{k2^{-j}} \leq C(nv_n^2)^k,
\end{aligned}$$

where we have used (4.8). Similarly, for any $j \geq 1$,

$$\begin{aligned} B^{(2)}_{k2^{1-j},2^j} &\leq (nm_2 M_{\lambda_0})^{k2^{1-j}} \left(\int_0^T \log^{2^j} \left(\frac{M_{\lambda_0} \bar{\lambda}_0(t)}{M_\lambda \bar{\lambda}(t)} \right) \bar{\lambda}_0(t) \mathrm{d}t \right)^{k2^{1-j}} \\ &\leq C \left(\log^{2^j} \left(\frac{M_{\lambda_0}}{M_\lambda} \right) + E_{2^j}(\bar{\lambda}_0, \bar{\lambda}) \right)^{k2^{1-j}} \times n^{k2^{1-j}} \\ &\leq C(nv_n^2)^{k2^{1-j}} \leq C(nv_n^2)^k. \end{aligned}$$

Therefore, for any $k > 1$,

$$V_{2k}(\lambda_0;\, \lambda) \leq \kappa (nv_n^2)^k,$$

where κ depends on C_{1k}, k, H, λ_0, m_1 and m_2. Using previous computations, the case $k = 1$ is straightforward. So, we obtain the result for $V_k(\lambda_0;\, \lambda)$ with $k \geq 2$.

6.2 Proof of Proposition 2

We shall use the following lemma whose proof is given in Section 6.3.

Lemma 1 *There exist constants ξ, $K > 0$, only depending on M_{λ_0}, α, m_1 and m_2, such that, for any λ_1, there exists a test ϕ_{λ_1} so that*

$$\mathbb{E}^{(n)}_{\lambda_0}[1_{\Gamma_n} \phi_{\lambda_1}] \leq 2 \exp\left(-Kn\|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1,\, m_1\}\right)$$

and

$$\sup_{\lambda:\ \|\lambda-\lambda_1\|_1 < \xi\|\lambda_1-\lambda_0\|_1} \mathbb{E}_\lambda[1_{\Gamma_n}(1 - \phi_{\lambda_1})] \leq 2 \exp\left(-Kn\|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1,\, m_1\}\right).$$

We consider the setting of Lemma 1 and a covering of $S_{n,j}(v_n)$ with $\mathbb{L}_1$-balls with radius $\xi j v_n$ and centers $(\lambda_{l,j})_{l=1,\dots,D_j}$, where D_j is the covering number of $S_{n,j}(v_n)$ by such balls. We then set $\phi_{n,j} = \max_{l=1,\dots,D_j} \phi_{\lambda_{l,j}}$, where the $\phi_{\lambda_{l,j}}$'s are defined in Lemma 1. So, there exists a constant $\rho > 0$ such that

$$\mathbb{E}^{(n)}_{\lambda_0}[1_{\Gamma_n}\phi_{n,j}] \leq 2D_j e^{-Knj^2v_n^2}, \qquad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}^{(n)}_\lambda[1_{\Gamma_n}(1-\phi_{n,j})] \leq 2e^{-Knj^2v_n^2}, \quad \text{if } j \leq \rho/v_n$$

and

$$\mathbb{E}^{(n)}_{\lambda_0}[1_{\Gamma_n}\phi_{n,j}] \leq 2D_j e^{-Knjv_n}, \qquad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}^{(n)}_\lambda[1_{\Gamma_n}(1-\phi_{n,j})] \leq 2e^{-Knjv_n}, \quad \text{if } j \geq \rho/v_n,$$

where K is a constant (see Lemma 1). We now bound D_j. First note that for any $\lambda = M_\lambda \bar{\lambda}$ and $\lambda' = M_{\lambda'} \bar{\lambda}'$,

$$\|\lambda - \lambda'\|_1 \leq M_\lambda \|\bar{\lambda} - \bar{\lambda}'\|_1 + |M_\lambda - M_{\lambda'}|. \tag{6.6}$$

Assume that $M_\lambda \geq M_{\lambda_0}$. Then, we have

$$\begin{aligned} \|\lambda - \lambda_0\|_1 &\geq \int_{\bar{\lambda} > \bar{\lambda}_0} (M_\lambda \bar{\lambda}(x) - M_{\lambda_0}\bar{\lambda}_0(x))\mathrm{d}x \\ &= M_\lambda \int_{\bar{\lambda} > \bar{\lambda}_0} (\bar{\lambda}(x) - \bar{\lambda}_0(x))\mathrm{d}x + (M_\lambda - M_{\lambda_0}) \int_{\bar{\lambda} > \bar{\lambda}_0} \bar{\lambda}_0(x)\mathrm{d}x \\ &\geq M_\lambda \int_{\bar{\lambda} > \bar{\lambda}_0} (\bar{\lambda}(x) - \bar{\lambda}_0(x))\mathrm{d}x = \frac{M_\lambda \|\bar{\lambda} - \bar{\lambda}_0\|_1}{2}. \end{aligned}$$

Conversely, if $M_\lambda < M_{\lambda_0}$,

$$\|\lambda - \lambda_0\|_1 \geq \int_{\bar{\lambda}_0 > \bar{\lambda}} (M_{\lambda_0}\bar{\lambda}_0(x) - M_\lambda \bar{\lambda}(x))\mathrm{d}x$$
$$\geq M_{\lambda_0} \int_{\bar{\lambda}_0 > \bar{\lambda}} (\bar{\lambda}_0(x) - \bar{\lambda}(x))\mathrm{d}x = \frac{M_{\lambda_0}\|\bar{\lambda} - \bar{\lambda}_0\|_1}{2}.$$

So, $2\|\lambda - \lambda_0\|_1 \geq (M_\lambda \vee M_{\lambda_0})\|\bar{\lambda} - \bar{\lambda}_0\|_1$ and we finally have

$$\|\lambda - \lambda_0\|_1 \geq \max\left\{\frac{(M_\lambda \vee M_{\lambda_0})\|\bar{\lambda} - \bar{\lambda}_0\|_1}{2}, |M_\lambda - M_{\lambda_0}|\right\} \tag{6.7}$$

So, for all $\lambda = M_\lambda \bar{\lambda} \in S_{n,j}(v_n)$,

$$\|\bar{\lambda} - \bar{\lambda}_0\|_1 \leq \frac{2(j+1)v_n}{M_{\lambda_0}}, \quad |M_\lambda - M_{\lambda_0}| \leq (j+1)v_n. \tag{6.8}$$

Therefore, $S_{n,j}(v_n) \subset (\bar{S}_{n,j} \cap \mathcal{F}_n) \times \{M : \ |M - M_{\lambda_0}| \leq (j+1)v_n\}$ and any covering of $(\bar{S}_{n,j} \cap \mathcal{F}_n) \times \{M : \ |M - M_{\lambda_0}| \leq (j+1)v_n\}$ will give a covering of $S_{n,j}(v_n)$. So, to bound D_j, we have to build a convenient covering of $(\bar{S}_{n,j} \cap \mathcal{F}_n) \times \{M : \ |M - M_{\lambda_0}| \leq (j+1)v_n\}$. We distinguish two cases.

- We assume that $(j+1)v_n \leq 2M_{\lambda_0}$. Then, (6.8) implies that $M_\lambda \leq 3M_{\lambda_0}$. Moreover, if

$$\|\bar{\lambda} - \bar{\lambda}'\|_1 \leq \frac{\xi j v_n}{3M_{\lambda_0} + 1} \quad \text{and} \quad |M_\lambda - M_{\lambda'}| \leq \frac{\xi j v_n}{3M_{\lambda_0} + 1},$$

  then, by using (6.6),

$$\|\lambda - \lambda'\|_1 \leq \frac{(M_\lambda + 1)\xi j v_n}{3M_{\lambda_0} + 1} \leq \xi j v_n.$$

  By Assumption (ii) of Theorem 5, this implies that, for any $\delta > 0$, there exists J_0 such that for $j \geq J_0$,

$$D_j \leq D((3M_{\lambda_0}+1)^{-1}\xi j v_n, \bar{S}_{n,j} \cap \mathcal{F}_n, \|\cdot\|_1) \times \left(2(j+1)v_n \times \frac{(3M_{\lambda_0}+1)}{\xi j v_n} + \frac{1}{2}\right) \lesssim \exp(\delta n (j+1)^2 v_n^2).$$

- We assume that $(j+1)v_n \geq 2M_{\lambda_0}$. If

$$\|\bar{\lambda} - \bar{\lambda}'\|_1 \leq \frac{\xi}{4} \quad \text{and} \quad |M_\lambda - M_{\lambda'}| \leq \frac{\xi(M_\lambda \vee M_{\lambda_0})}{4},$$

  then using again (6.6) and (6.8),

$$\|\lambda - \lambda'\|_1 \leq \frac{\xi M_\lambda}{4} + \frac{\xi(M_\lambda + M_{\lambda_0})}{4} \leq \frac{3\xi M_{\lambda_0}}{4} + \frac{\xi(j+1)v_n}{2} \leq \frac{7\xi(j+1)v_n}{8} \leq \xi j v_n,$$

  for n large enough. By Assumption (i) of Theorem 5, this implies that, for any $\delta > 0$,

$$D_j \lesssim D(\xi/4, \mathcal{F}_n, \|\cdot\|_1) \times \log((j+1)v_n) \lesssim \log(j v_n)\exp(\delta n).$$

It is enough to choose δ small enough to obtain the result of Proposition 2.

6.3 Proof of Lemma 1

For any λ, we denote $\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[\cdot] = \mathbb{E}^{(n)}_{\lambda}[1_{\Gamma_n} \times \cdot]$. We set for any λ, λ',

$$\|\lambda - \lambda'\|_{\tilde{\mu}_n} := \int_\Omega |\lambda(t) - \lambda'(t)|\tilde{\mu}_n(t)\mathrm{d}t.$$

On Γ_n we have

$$m_1\|\lambda - \lambda_0\|_1 \leq \|\lambda - \lambda_0\|_{\tilde{\mu}_n} \leq m_2\|\lambda - \lambda_0\|_1. \tag{6.9}$$

The main tool for building convenient tests is Theorem 3 of Hansen et al. (2013) (and its proof) applied in the univariate setting. By mimicking the proof of this theorem from Inequality (7.5) to Inequality (7.7), if H is a deterministic function bounded by b, we have that, for any $u \geq 0$,

$$\mathbb{P}^{(n)}_{\lambda}\left(\left|\int_0^T H_t(\mathrm{d}N_t - \mathrm{d}\Lambda_t)\right| \geq \sqrt{2vu} + \frac{bu}{3} \text{ and } \Gamma_n\right) \leq 2e^{-u}, \tag{6.10}$$

where we recall that $\Lambda_t = \int_0^t Y_s\lambda(s)\mathrm{d}s$ and v is a deterministic constant such that on Γ_n, almost surely,

$$\int_0^T H_t^2 Y_t\lambda(t)\mathrm{d}t \leq v.$$

For any non-negative function λ_1, we define the sets

$$A := \{t \in \Omega :\ \lambda_1(t) \geq \lambda_0(t)\}, \quad A^c := \{t \in \Omega :\ \lambda_1(t) < \lambda_0(t)\}$$

and the following pseudo-metrics

$$d_A(\lambda_1,\, \lambda_0) := \int_A [\lambda_1(t) - \lambda_0(t)]\tilde{\mu}_n(t)\mathrm{d}t \quad d_{A^c}(\lambda_1,\, \lambda_0) := \int_{A^c} [\lambda_0(t) - \lambda_1(t)]\tilde{\mu}_n(t)\mathrm{d}t.$$

Note that $\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} = d_A(\lambda_1,\, \lambda_0) + d_{A^c}(\lambda_1,\, \lambda_0)$. For $u > 0$, if $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$, define the test

$$\phi_{\lambda_1,A}(u) := 1\left\{N(A) - \int_A \lambda_0(t)Y_t\mathrm{d}t \geq \rho_n(u)\right\}, \text{ with } \rho_n(u) := \sqrt{2nv(\lambda_0)u} + \frac{u}{3},$$

where, for any non-negative function λ,

$$v(\lambda) := (1 + \alpha)\int_\Omega \lambda(t)\tilde{\mu}_n(t)\mathrm{d}t. \tag{6.11}$$

Similarly, if $d_A(\lambda_1,\, \lambda_0) < d_{A^c}(\lambda_1,\, \lambda_0)$, define

$$\phi_{\lambda_1,A^c}(u) := 1\left\{N(A^c) - \int_{A^c} \lambda_0(t)Y_t\mathrm{d}t \leq -\rho_n(u)\right\}.$$

Since for any non-negative function λ, on Γ_n, by using (4.4),

$$(1 - \alpha)\int_\Omega \lambda(t)\tilde{\mu}_n(t)\mathrm{d}t \leq \int_\Omega \lambda(t)\frac{Y_t}{n}\mathrm{d}t \leq (1 + \alpha)\int_\Omega \lambda(t)\tilde{\mu}_n(t)\mathrm{d}t, \tag{6.12}$$

then inequality (6.10) applied with $H = 1_A$ or $H = 1_{A^c}$, $b = 1$ and $v = nv(\lambda_0)$ implies that for any $u > 0$,

$$\mathbb{E}^{(n)}_{\lambda_0,\Gamma_n}[\phi_{\lambda_1,A}(u)] \leq 2e^{-u}, \quad \mathbb{E}^{(n)}_{\lambda_0,\Gamma_n}[\phi_{\lambda_1,A^c}(u)] \leq 2e^{-u}. \tag{6.13}$$

Now, we state the following lemma whose proof is given in Section 6.4.

Lemma 2 *Let λ be a non-negative function. Assume that*

$$\|\lambda - \lambda_1\|_{\tilde{\mu}_n} \leq \frac{1-\alpha}{4(1+\alpha)} \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}.$$

We set $\tilde{M}_n(\lambda_0) = \int_\Omega \lambda_0(t)\tilde{\mu}_n(t)\mathrm{d}t$ and we distinguish two cases.

1. *Assume that $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$. Then,*

$$\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1 - \phi_{\lambda_1,A}(u_A)] \leq 2\exp(-u_A),$$

where

$$u_A = \begin{cases} u_{0A} n d_A^2(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0), \\ u_{1A} n d_A(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} > 2\tilde{M}_n(\lambda_0), \end{cases}$$

and u_{0A} and u_{1A} are two constants only depending on α, M_{λ_0}, m_1 and m_2.

2. *Assume that $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$. Then,*

$$\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1 - \phi_{\lambda_1,A^c}(u_{A^c})] \leq 2\exp(-u_{A^c}),$$

where

$$u_{A^c} = \begin{cases} u_{0A^c} n d_{A^c}^2(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0), \\ u_{1A^c} n d_{A^c}(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} > 2\tilde{M}_n(\lambda_0), \end{cases}$$

and u_{0A^c} and u_{1A^c} are two constants only depending on α, M_{λ_0}, m_1 and m_2.

Note that, by using (6.9), if $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$,

$$\begin{aligned} u_A &\geq \min\{u_{0A} n d_A^2(\lambda_1, \lambda_0),\, u_{1A} n d_A(\lambda_1, \lambda_0)\} \\ &\geq n d_A(\lambda_1, \lambda_0) \times \min\{u_{0A} d_A(\lambda_1, \lambda_0),\, u_{1A}\} \\ &\geq \frac{n m_1 \|\lambda_1 - \lambda_0\|_1}{2} \times \min\left\{\frac{u_{0A} m_1 \|\lambda_1 - \lambda_0\|_1}{2},\, u_{1A}\right\} \\ &\geq K_A n \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1,\, m_1\}, \end{aligned}$$

for K_A a positive constant small enough only depending on α, M_{λ_0}, m_1 and m_2. Similarly, if $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$,

$$\begin{aligned} u_{A^c} &\geq \frac{n m_1 \|\lambda_1 - \lambda_0\|_1}{2} \times \min\left\{\frac{u_{0A^c} m_1 \|\lambda_1 - \lambda_0\|_1}{2},\, u_{1A^c}\right\} \\ &\geq K_{A^c} n \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1,\, m_1\}, \end{aligned}$$

for K_{A^c} a positive constant small enough only depending on α, M_{λ_0}, m_1 and m_2. Now, we set

$$\phi_{\lambda_1} = \phi_{\lambda_1,A}(u_A) 1_{\{d_A(\lambda_1,\lambda_0) \geq d_{A^c}(\lambda_1,\lambda_0)\}} + \phi_{\lambda_1,A^c}(u_{A^c}) 1_{\{d_A(\lambda_1,\lambda_0) < d_{A^c}(\lambda_1,\lambda_0)\}}$$

so that, with $K = \min\{K_A, K_{A^c}\}$, by using (6.13),

$$\begin{aligned} \mathbb{E}^{(n)}_{\lambda_0,\Gamma_n}[\phi_{\lambda_1}] &= \mathbb{E}^{(n)}_{\lambda_0,\Gamma_n}[\phi_{\lambda_1,A}(u_A)] 1_{\{d_A(\lambda_1,\lambda_0) \geq d_{A^c}(\lambda_1,\lambda_0)\}} + \mathbb{E}^{(n)}_{\lambda_0,\Gamma_n}[\phi_{\lambda_1,A^c}(u_{A^c})] 1_{\{d_A(\lambda_1,\lambda_0) < d_{A^c}(\lambda_1,\lambda_0)\}} \\ &\leq 2e^{-u_A} 1_{\{d_A(\lambda_1,\lambda_0) \geq d_{A^c}(\lambda_1,\lambda_0)\}} + 2e^{-u_{A^c}} 1_{\{d_A(\lambda_1,\lambda_0) < d_{A^c}(\lambda_1,\lambda_0)\}} \\ &\leq 2\exp\left(-Kn\|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1,\, m_1\}\right). \end{aligned}$$

If $\|\lambda - \lambda_1\|_1 < \xi \|\lambda_1 - \lambda_0\|_1$, $\xi = m_1(1-\alpha)/(4m_2(1+\alpha))$, then

$$\|\lambda - \lambda_1\|_{\tilde{\mu}_n} \leq \frac{1-\alpha}{4(1+\alpha)} \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}$$

and Lemma 2 shows that

$$\begin{aligned}
\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1 - \phi_{\lambda_1}] &\leq 2e^{-u_A} 1_{\{d_A(\lambda_1,\lambda_0) \geq d_{A^c}(\lambda_1,\lambda_0)\}} + 2e^{-u_{A^c}} 1_{\{d_A(\lambda_1,\lambda_0) < d_{A^c}(\lambda_1,\lambda_0)\}} \\
&\leq 2\exp\left(-Kn\|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\}\right),
\end{aligned}$$

which ends the proof of Lemma 1.

6.4 Proof of Lemma 2

We only consider the case $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$. The case $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$ can be dealt with by using similar arguments. So, we assume that $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$. On Γ_n we have

$$\begin{aligned}
\int_A (\lambda_1(t) - \lambda_0(t)) Y_t \mathrm{d}t &\geq n(1-\alpha) \int_A (\lambda_1(t) - \lambda_0(t)) \tilde{\mu}_n(t) \mathrm{d}t \\
&\geq \frac{n(1-\alpha)}{2} \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \geq 2n(1+\alpha)\|\lambda - \lambda_1\|_{\tilde{\mu}_n} \\
&\geq 2n(1+\alpha) \int_A |\lambda(t) - \lambda_1(t)| \tilde{\mu}_n(t) \mathrm{d}t \geq 2\int_A |\lambda(t) - \lambda_1(t)| Y_t \mathrm{d}t.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1 - \phi_{\lambda_1,A}(u_A)] &= \mathbb{P}^{(n)}_{\lambda,\Gamma_n}\left\{N(A) - \int_A \lambda(t) Y_t \mathrm{d}t < \rho_n(u_A) + \int_A (\lambda_0 - \lambda)(t) Y_t \mathrm{d}t\right\} \\
&= \mathbb{P}^{(n)}_{\lambda,\Gamma_n}\left\{N(A) - \int_A \lambda(t) Y_t \mathrm{d}t < \rho_n(u_A) - \int_A (\lambda_1 - \lambda_0)(t) Y_t \mathrm{d}t \right. \\
&\qquad\qquad \left. + \int_A (\lambda_1 - \lambda)(t) Y_t \mathrm{d}t\right\} \\
&\leq \mathbb{P}^{(n)}_{\lambda,\Gamma_n}\left\{N(A) - \int_A \lambda(t) Y_t \mathrm{d}t < \rho_n(u_A) - \frac{1}{2}\int_A (\lambda_1 - \lambda_0)(t) Y_t \mathrm{d}t\right\}.
\end{aligned}$$

<u>Assume $\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0)$</u>

This assumption implies that

$$d_A(\lambda_1, \lambda_0) \leq \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0) \leq 2m_2 M_{\lambda_0}.$$

Since $v(\lambda_0) = (1+\alpha)\tilde{M}_n(\lambda_0)$, with $u_A = u_{0A} n d_A^2(\lambda_1, \lambda_0)$, where $u_{0A} \leq 1$ is a constant depending on α, m_1 and m_2 chosen later, we have

$$\rho_n(u_A) \leq n d_A(\lambda_1, \lambda_0)\sqrt{2u_{0A}(1+\alpha)\tilde{M}_n(\lambda_0)} + \frac{u_{0A} n d_A^2(\lambda_1, \lambda_0)}{3} \leq \sqrt{u_{0A}}\, n d_A(\lambda_1, \lambda_0) K_1$$

as soon as $K_1 \geq \sqrt{2(1+\alpha)\tilde{M}_n(\lambda_0)} + \frac{2\tilde{M}_n(\lambda_0)\sqrt{u_{0A}}}{3}$. Observe that the definition of $v(\lambda)$ in (6.11) gives

$$\begin{aligned}
v(\lambda) &= (1+\alpha)\int_\Omega \lambda_0(t)\tilde{\mu}_n(t)\mathrm{d}t + (1+\alpha)\int_\Omega (\lambda(t) - \lambda_0(t))\tilde{\mu}_n(t)\mathrm{d}t \\
&\leq v(\lambda_0) + (1+\alpha)\|\lambda - \lambda_0\|_{\tilde{\mu}_n} \leq v(\lambda_0) + (1+\alpha)\left(\|\lambda - \lambda_1\|_{\tilde{\mu}_n} + \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}\right) \\
&\leq v(\lambda_0) + \frac{5+3\alpha}{4}\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq C_1,
\end{aligned}$$

where C_1 only depends on α, M_{λ_0}, m_1 and m_2. Combined with (6.12), this implies that, on Γ_n, if $K_1 \leq \frac{1-\alpha}{4\sqrt{u_{0A}}}$, which is true for u_{0A} small enough,

$$\begin{aligned}\frac{1}{2}\int_A (\lambda_1-\lambda_0)(t)Y_t\mathrm{d}t - \rho_n(u_A) &\geq \frac{(1-\alpha)n}{2} d_A(\lambda_1,\, \lambda_0)\left[1 - \frac{2K_1\sqrt{u_{0A}}}{1-\alpha}\right] \\ &\geq \frac{(1-\alpha)n}{4} d_A(\lambda_1,\, \lambda_0) \geq \sqrt{2nC_1 r} + \frac{r}{3} \geq \sqrt{2nv(\lambda)r} + \frac{r}{3},\end{aligned}$$

with

$$r = n\min\left\{\frac{(1-\alpha)^2 d_A^2(\lambda_1,\, \lambda_0)}{128C_1},\, \frac{3(1-\alpha)d_A(\lambda_1,\, \lambda_0)}{8}\right\}.$$

Inequality (6.10) then leads to

$$\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1-\phi_{\lambda_1,A}(u_A)] \leq 2e^{-r}. \tag{6.14}$$

For u_{0A} small enough only depending on M_{λ_0}, α, m_1 and m_2, we have

$$\frac{(1-\alpha)}{4\sqrt{u_{0A}}} \geq \sqrt{2(1+\alpha)\tilde{M}_n(\lambda_0)} + \frac{2\tilde{M}_n(\lambda_0)\sqrt{u_{0A}}}{3}$$

so (6.14) is true. Since $r \geq u_A$ for u_{0A} small enough, then

$$\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1-\phi_{\lambda_1,A}(u)] \leq 2e^{-u_A}.$$

<u>Assume $\|\lambda_1-\lambda_0\|_{\tilde{\mu}_n} \geq 2\tilde{M}_n(\lambda_0)$</u>

We take $u_A = u_{1A}nd_A(\lambda_1,\, \lambda_0)$, where $u_{1A} \leq 1$ is a constant depending on α chosen later. We still consider the same test $\phi_{\lambda_1,A}(u_A)$. Observe now that, since $d_A(\lambda_1,\, \lambda_0) \geq \frac{1}{2}\|\lambda_1-\lambda_0\|_{\tilde{\mu}_n} \geq \tilde{M}_n(\lambda_0)$,

$$\begin{aligned}\rho_n(u_A) &= \sqrt{2nu_A v(\lambda_0)} + \frac{u_A}{3} \\ &\leq n\sqrt{2(1+\alpha)u_{1A}\tilde{M}_n(\lambda_0)d_A(\lambda_1,\, \lambda_0)} + \frac{d_A(\lambda_1,\, \lambda_0)nu_{1A}}{3} \\ &\leq \left(\sqrt{2(1+\alpha)} + \frac{1}{3}\right)\sqrt{u_{1A}}nd_A(\lambda_1,\, \lambda_0)\end{aligned}$$

and, under the assumptions of the lemma,

$$v(\lambda) \leq (1+\alpha)\tilde{M}_n(\lambda_0) + (1+\alpha)\left(\|\lambda-\lambda_1\|_{\tilde{\mu}_n} + \|\lambda_1-\lambda_0\|_{\tilde{\mu}_n}\right) \leq C_2 d_A(\lambda_1,\, \lambda_0), \tag{6.15}$$

where C_2 only depends on α. Therefore,

$$\begin{aligned}\frac{1}{2}\int_A (\lambda_1-\lambda_0)(t)Y_t\mathrm{d}t - \rho_n(u_A) &\geq \frac{n(1-\alpha)}{2}\int_A (\lambda_1(t)-\lambda_0(t))\tilde{\mu}_n(t)\mathrm{d}t - \left(\sqrt{2(1+\alpha)}+\frac{1}{3}\right)\sqrt{u_{1A}}nd_A(\lambda_1,\, \lambda_0) \\ &\geq \left(\frac{1-\alpha}{2} - \left(\sqrt{2(1+\alpha)}+\frac{1}{3}\right)\sqrt{u_{1A}}\right)nd_A(\lambda_1,\, \lambda_0) \\ &\geq \frac{1-\alpha}{4}nd_A(\lambda_1,\, \lambda_0),\end{aligned}$$

where the last inequality is true for u_{1A} small enough depending only on α. Finally, using (6.15) and since $u_A = u_{1A}nd_A(\lambda_1,\, \lambda_0)$, we have

$$\begin{aligned}\frac{1-\alpha}{4}nd_A(\lambda_1,\, \lambda_0) &\geq \sqrt{2nC_2 d_A(\lambda_1,\, \lambda_0)u_{1A}nd_A(\lambda_1,\, \lambda_0)} + \frac{u_{1A}nd_A(\lambda_1,\, \lambda_0)}{3} \\ &\geq \sqrt{2nv(\lambda)u_A} + \frac{u_A}{3}\end{aligned}$$

for u_{1A} small enough depending only on α. We then obtain

$$\mathbb{E}^{(n)}_{\lambda,\Gamma_n}[1-\phi_{\lambda_1,A}(u_A)] \leq 2e^{-u_A},$$

which ends the proof of the lemma.

References

Adams, R. P., Murray, I., and MacKay, D. J. (2009). Tractable nonparametric bayesian inference in poisson processes with gaussian process intensities. *Proceedings of the 26th Annual International Conference on Machine Learning. ACM.*

Ahmed, S. and Reid, N. (2001). *Empirical Bayes and Likelihood Inference*. Lecture Notes in Statistics, Springer.

Andersen, P. K., Borgan, A., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.

Barron, A., Schervish, M., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.*, 27:536–561.

Belitser, E., Serra, P., and van Zanten, J. H. (2013). Rate optimal Bayesian intensity smoothing for inhomogeneous Poisson processes. *http://arxiv.org/pdf/1304.6017v2*.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, second edition.

Carroll, R. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. American Statist. Assoc.*, 83:1184–1186.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, 39:83–87.

Cheng, N. and Yuan, T. (2013). Nonparametric Bayesian lifetime data analysis using Dirichlet process lognormal mixture model. *Naval Res. Logist.*, 60(3):208–221.

Clyde, M. A. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *J. Royal Statist. Society Series B*, pages 681–698.

Cui, W. and George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *J. Statist. Plann. Inference*, 138:888–900.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition. Elementary theory and methods.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition. General theory and structure.

Dedecker, J. and Michel, B. (2013). Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension. *Journal of Multivariate Analysis*, 122:278–291.

Doss, H. and Sellke, T. (1982). The tails of probabilities chosen from a Dirichlet prior. *Ann. Statist.*, 10:1302–1305.

Efron, B. (2003). Robbins, empirical Bayes and microarrays. *Ann. Statist.*, 31:366–378.

Fall, M. D. and Barat, É. (2012). Gibbs sampling methods for Pitman-Yor mixture models. MAP5 2012-30.

Fan, J. (1991a). Global behavior of deconvolution kernel estimates. *Statistica Sinica*, 1:541–551.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19:1257–1272.

Fan, J. (1992). Deconvolution with supersmooth errors. *Canadian J. Statist.*, 20:155–169.

Ferguson, T. (1974). Prior distributions in spaces of probability measures. *Ann. Statist.*, 2:615–629.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28:500–531.

Ghosal, S. and van der Vaart, A. (2007a). Convergence rates of posterior distributions for non iid observations. *Ann. Statist.*, 35(1):192–223.

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

Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29(5):1233–1263.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. Springer-Verlag, New York.

Hansen, N., Reynaud-Bouret, P., and Rivoirard, V. (2013). Lasso and probabilistic inequalities for multivariate point processes. *Arxiv preprint arXiv:1208.0570*.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press, Cambridge, UK.

Karr, A. F. (1991). *Point processes and their statistical inference*, volume 7 of *Probability: Pure and Applied*. Marcel Dekker Inc., New York, second edition.

Knapik, B. T., Szabó, B. T., van der Vaart, A. W., and van Zanten, J. H. (2012). Bayes procedures for adaptive inference in nonparametric inverse problems. Technical report.

Kottas, A. and Sansó, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *J. Statist. Plann. Inference*, 137(10):3151–3163.

Kruijer, W., Rousseau, J., and van der Vaart, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Stat.*, 4:1225–1257.

Lo, A. Y. (1992). Bayesian inference for Poisson process models with censored data. *J. Nonparametr. Statist.*, 2(1):71–80.

McAuliffe, J. D., Blei, D. M., and Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16:5–14.

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

Petrone, S., Rousseau, J., and Scricciolo, C. (2014). Bayes and empirical Bayes: do they merge? *Biometrika*, 101:285–302.

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. Royal Statist. Society Series B*, 59:731–792.

Rousseau, J. (2010). Rates of convergence for the posterior distributions of mixtures of Betas and adaptive nonparametric estimation of the density. *Ann. Statist.*, 38:146–180.

Salomond, J. (2013). Concentration rate and consistency of the posterior under monotonicity constraints. Technical report, Univ. Paris Dauphine.

Sarkar, A., Pati, D., Mallick, B. K., and Carroll, R. J. (2013). Adaptive posterior convergence rates in Bayesian density deconvolution with supersmooth errors. Technical report, arXiv:1308.5427v2.

Schwartz, L. (1965). On Bayes procedures. *Z. Warsch. Verw. Gebiete*, 4:10–26.

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.

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

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

Szabó, B. T., van der Vaart, A. W., and van Zanten, J. H. (2013). Frequentist coverage of adaptive nonparametric Bayesian credible sets. Technical report.

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.*, 23:189–207.