Bayesian Analysis (2017) **12**, Number 1, pp. 53–87

Posterior Concentration Rates for Counting Processes with Aalen Multiplicative Intensities

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

Abstract. We provide sufficient conditions to derive posterior concentration rates for Aalen counting processes on a finite time horizon. The conditions are designed to resemble those proposed in the literature for the problem of density estimation, for instance, in Ghosal et al. (2000), so that existing results on density estimation can be adapted to the present setting. We apply the general theorem to some prior models including Dirichlet process mixtures of uniform densities to estimate monotone nondecreasing intensities and log-splines.

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

1 Introduction

Estimation of the intensity function of a point process is an important statistical problem with a long history. Most methods were initially employed for estimating intensities assumed to be of parametric or nonparametric form in Poisson point processes. However, in many fields such as genetics, seismology and neuroscience, the probability of observing a new jump of the studied temporal process may depend on covariates and, in this case, the intensity of the process is random so that such a feature is not captured by a classical Poisson model. Aalen models constitute a natural extension of Poisson models that allows for taking into account this aspect. Aalen (1978) revolutionized point process analysis developing a unified theory for frequentist nonparametric inference of multiplicative intensity models which, besides the Poisson model and other classical models such as right-censoring and Markov processes with finite state space, described in Section 2.2, encompass birth and death processes as well as branching processes. We refer the reader to Andersen et al. (1993) for a presentation of Aalen processes including various other illustrative examples. Classical probabilistic and statistical results about Aalen processes can be found in Karr (1991), Daley and Vere-Jones (2003, 2008). Recent nonparametric frequentist methodologies based on penalized least-squares contrasts have been proposed by Brunel and Comte (2005, 2008), Comte et al. (2011) and Reynaud-Bouret (2006). In the high-dimensional setting, more specific results have been established by Gaïffas and Guilloux (2012) and Hansen et al. (2015) who consider Lasso-type procedures.

Dykstra and Laud (1981) consider a Bayesian nonparametric approach to model hazard rates by extended gamma processes which have the advantage over Dirichlet

[*]MIA, INRA, UMR0518, AgroParisTech, sophie.donnet@agroparistech.fr
[†]CEREMADE, Université Paris Dauphine, rivoirard@ceremade.dauphine.fr
[‡]CEREMADE, Université Paris Dauphine, rousseau@ceremade.dauphine.fr
[§]Department of Decision Sciences, Bocconi University, catia.scricciolo@unibocconi.it

 DOI: 10.1214/15-BA986

processes that prior probability measures on the corresponding cumulative distribution functions select absolutely continuous rather than discrete distributions. Bayesian nonparametric inference for inhomogeneous Poisson point processes is considered by Lo (1982, 1992) who develops a prior-to-posterior analysis for weighted gamma process priors to model intensity functions. In the same spirit, Kuo and Ghosh (1997) employ several classes of nonparametric priors, including gamma, beta and extended gamma processes. Extension to multiplicative counting processes is treated in Lo and Weng (1989) who model intensities as kernel mixtures with mixing measure distributed according to a weighted gamma measure on the real line. Along the same lines, Ishwaran and James (2004) develop computational procedures for Bayesian non- and semi-parametric multiplicative intensity models using kernel mixtures of weighted gamma measures which can be viewed as a special case of kernel mixtures of dependent completely random measures proposed by Lijoi and Nipoti (2014).

Kim (1999) considers priors for the cumulative intensity function based on Lévy processes and, using conjugacy for the Aalen's multiplicative counting process model, derives formulas for the posterior process. Posterior inference is then exemplified in Poisson processes, right-censoring and Markov processes. Lévy processes are also considered for nonparametric inference with mixed Poisson processes by Gutiérrez-Peña and Nieto-Barajas (2003). Other articles mainly focus on exploring 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).

To the best of our knowledge, there are no results in the literature concerning aspects of the frequentist asymptotic behaviour of posterior distributions, like consistency and rates of convergence, for intensity estimation of general Aalen models. There are recent works on posterior contraction rates for inhomogeneous Poisson processes by Belitser et al. (2015), Gugushvili and Spreij (2013) and Kirichenko and van Zanten (2015) and a contribution on posterior consistency for hazard rate estimation with or without censoring by De Blasi et al. (2009). Both types of models are specific examples of Aalen processes. In this article, we generalize these results by studying rates of convergence for general Aalen multiplicative intensity models.

As in Belitser et al. (2015), Gugushvili and Spreij (2013) and Kirichenko and van Zanten (2015), we restrict attention to the estimation of the intensity function over a bounded interval $[0, T]$, with fixed $T > 0$. Although this is restrictive, this setup is realistic in a number of applications where the study takes place during a fixed period of time, but many subjects are meanwhile observed, see Section 2 for the mathematical formulation. Hence, T can be understood as a deterministic right-truncation. For estimating intensities of inhomogeneous Poisson processes, the extension to the case $T = +\infty$ can be performed in a similar way to what is done for density estimation, but this is not pursued here. Note that in the frequentist literature, minimax convergence rates for the intensity of a Poisson process on the real line have been derived by Reynaud-Bouret and Rivoirard (2010). In this case, the rates typically depend on tail conditions on the intensity or otherwise significantly deteriorate without such conditions.

Quoting Lo and Weng (1989), "the idea of our approach is that estimating a density and estimating a hazard rate are analogous affairs, and a successful attempt of one generally leads to a feasible approach for the other". Thus, in deriving general sufficient conditions for assessing posterior contraction rates in Theorem 1 of Section 3, we attempt at stating conditions that resemble those proposed by Ghosal et al. (2000) for density estimation with independent and identically distributed (i.i.d.) observations. This allows us to then derive in Section 4 posterior contraction rates for different families of prior distributions such as Dirichlet mixtures of uniform densities to estimate monotone nondecreasing intensities and log-splines by an adaptation of existing results on density estimation. Detailed proofs of the main results are reported in Section 6. Auxiliary results concerning the control of the Kullback–Leibler divergence for intensities in Aalen models and the existence of tests, which, to the best of our knowledge, are derived here for the first time and can also be of independent interest, are presented in Section 7 and Supplementary material (Donnet et al., 2015). Before exposing theoretical results, Section 2 introduces the setting of Aalen multiplicative intensity models and presents examples of such models like, for instance, right-censoring models in survival analysis whose treatment is a guideline of the article.

2 Aalen multiplicative intensity models

2.1 Set-up and notation

Let $(\mathcal{G}_t)_{t\geq 0}$ be a filtration on a probability space. Let $N = (N_t)_{t\geq 0}$ be a counting process on $\mathbb{R}_+$, namely, the sample paths of $(N_t)_{t\geq 0}$ are right-continuous step functions with value 0 at $t = 0$ and with positive jumps, each one of size 1. In the sequel, N_t denotes the number of jumps in $[0, t]$. We assume that, for any $t \geq 0$, $N_t < \infty$ almost surely. For any Borel set A, we denote by $N(A)$ the number of jumps of N in A. Let Λ be the compensator of N with respect to $(\mathcal{G}_t)_{t\geq 0}$, assumed to be finite, so that if $M_t = N_t - \Lambda_t$, then $(M_t)_t$ is a zero-mean $(\mathcal{G}_t)_t$-martingale. A non-negative predictable process $\tilde{\lambda}$ is called the stochastic intensity of N if Λ can be written as

$$\Lambda_t = \int_0^t \tilde{\lambda}(s)\mathrm{d}s, \quad t \geq 0,$$

see Section II.4.1. of Andersen et al. (1993) or Chapter 2 of Karr (1986) for more details. We say that N obeys the *Aalen multiplicative intensity model*, see Aalen (1978), if for any t,

$$\tilde{\lambda}(t) = \lambda(t) Y_t,$$

where $\lambda(\cdot)$ is a non-negative deterministic function called *intensity function* and $(Y_t)_t$ is a non-negative predictable process. We refer the reader to Kim (1999), Reynaud-Bouret (2006), Comte et al. (2011) or Hansen et al. (2015). For a detailed description of this model, see Chapter III of Andersen et al. (1993). Informally, using (2.41) of Karr (1986),

$$\begin{aligned}\mathbb{E}[N([t,\, t+\mathrm{d}t]) \mid \mathcal{G}_{t^-}] &= \mathbb{P}[N([t,\, t+\mathrm{d}t]) = 1 \mid \mathcal{G}_{t^-}] \\ &= \mathbb{P}[N([t,\, t+\mathrm{d}t]) > 0 \mid \mathcal{G}_{t^-}] = \lambda(t) Y_t \mathrm{d}t.\end{aligned} \tag{1}$$

Note that, almost surely, we have no jumps of N on sets where λ or Y vanishes.

In this article, we estimate λ on a compact set, say $[0, T]$, where $0 < T < \infty$, by using a Bayesian posterior distribution based on observations of $(N_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$. The posterior distribution is conditioned on the observables $(\mathcal{G}_t)_{t\in[0,T]}$. To simplify the presentation, the posterior distribution is denoted by $\pi(\cdot \mid \mathbf{D})$, where $\mathbf{D}$ represents the observed data up to time T. Omitting constants independent of λ, the log-likelihood at λ with respect to the filtration $(\mathcal{G}_t)_{t\geq 0}$ can be expressed as

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

see Proposition 7.2.III of Daley and Vere-Jones (2003) or Theorem 2.31 of Karr (1986).

We are herein interested in asymptotic results: both N and Y depend on an integer n, and we study estimation of λ (not depending on n) when T is kept fixed and $n \to \infty$. In Belitser et al. (2015) and the references given in Section 1, asymptotic results for Aalen multiplicative intensity models are also presented with fixed and finite T and n going to infinity. More precisely, in Belitser et al. (2015), an inhomogeneous Poisson process with a T-periodic intensity function is observed up to time nT and n goes to ∞.

2.2 Examples of Aalen multiplicative intensity models

The following examples justify the interest in the Aalen model.

Inhomogeneous Poisson processes

We refer the reader to Kingman (1993) for a good introduction to Poisson processes and some concrete illustrations. See also Reynaud-Bouret and Rivoirard (2010) or Belitser et al. (2015) who model the counts of phone calls arriving at a call centre by using inhomogeneous Poisson processes. Poisson processes correspond to the case where the process $(Y_t)_{t\in[0,T]}$ is equal to 1. Assume that we observe n independent Poisson processes with common intensity λ on $[0, T]$. This model is equivalent to the model where we observe a Poisson process with intensity $n \times \lambda$, so it corresponds to the case $Y_t = n$ for all $t \in [0, T]$. In this case, if $T_1, \dots, T_{N_T}$ are the jump times of N over $[0, T]$, we have

$$\ell_n(\lambda) = \sum_{i=1}^{N_T} \log(\lambda(T_i)) - n\int_0^T \lambda(t)\mathrm{d}t.$$

In this example, $\mathbf{D} = (N_t)_{t\leq T}$. Finally, note that when λ depends on covariates, inhomogeneous Poisson processes are referred to as Cox processes, see Comte et al. (2011) or Karr (1986). The setting where processes depend on covariates are not studied in this article.

Survival analysis

We refer the reader to Chapter I of Andersen et al. (1993) and Chapter 3 of Klein and Moeschberger (2003) for a wide class of concrete examples in survival analysis.
We first consider right-censoring models that are very popular in biomedical problems, see, for instance, Example I.3.9 of Andersen et al. (1993) concerning the survival analysis

with right-censoring of patients with malignant melanoma. We consider n patients and, for each patient i, we consider its lifetime T_i (a non-negative random variable) with density f that can be censored and we denote by C_i the censoring time assumed to be independent of T_i. We face with censoring when, for instance, the patient drops out of a hospital study: the time of death is not observed, but we know that the patient was still alive when he left the study. In right-censoring models, we observe $(Z_i,\, \delta_i)$ on $[0,\, T]$, with $Z_i = \min\{T_i,\, C_i\}$ and $\delta_i = \mathbf{1}_{T_i \le C_i}$. In this case, the processes to be considered are

$$N_t^i = \delta_i \times \mathbf{1}_{Z_i \le t} \quad \text{and} \quad Y_t^i = \mathbf{1}_{Z_i \ge t}.$$

We assume that the vectors $(T_i,\, C_i)_{1\le i\le n}$ are i.i.d. and we denote by λ the common hazard rate of the T_i's assumed to be finite at least on $[0,\, T]$:

$$\lambda(t) = \frac{f(t)}{\mathbb{P}(T_1 > t)}, \quad t \in [0,\, T]. \tag{2}$$

Note that we do not force the Z_i's to be supported on $[0,\, T]$. Finally, consider N (resp., Y) by aggregating the n independent processes N^i's (resp., the Y^i's), so

$$N_t = \sum_{i=1}^{n} N_t^i \quad \text{and} \quad Y_t = \sum_{i=1}^{n} Y_t^i,$$

and straightforward computations show that the compensator of N is

$$\Lambda_t = \int_0^t \lambda(s) Y_s \mathrm{d}s, \quad t \in [0,\, T],$$

thus right-censoring models obey the Aalen multiplicative model. Expressing the log-likelihood, we obtain

$$\begin{aligned} \ell_n(\lambda) &= \int_0^T \log(\lambda(t))\mathrm{d}N_t - \int_0^T \lambda(t) Y_t \mathrm{d}t \\ &= \sum_{i=1}^{n} \delta_i \log(\lambda(Z_i)) - \sum_{i=1}^{n} \int_0^{Z_i} \lambda(t)\mathrm{d}t. \end{aligned} \tag{3}$$

Then, using (2), the likelihood is proportional to

$$\prod_{i=1}^{n} [f(Z_i)]^{\delta_i} \times [S(Z_i)]^{1-\delta_i},$$

where $S(x) = \mathbb{P}(T_1 > x)$. This expression is expected and coherent with classical references, see, for instance, (3.5.6) of Klein and Moeschberger (2003), by interpreting the previous formula with fixed i: either $\delta_i = 1$ and we observe T_i whose density is f or $\delta_i = 0$ and we just know that $T_i > Z_i$ justifying the term $[S(Z_i)]^{1-\delta_i}$. In this example, $\mathbf{D} = (Z_i,\, \delta_i)_{i\le n}$. Note that left-censoring models, where the minimum between the lifetime and the censoring time is replaced with the maximum, do not obey the Aalen model since in this case $(Y_t)_t$ is not predictable. See Andersen et al. (1993).

Finite state Markov processes

Let $X = (X(t))_t$ be a Markov process with finite state space $\mathbb{S}$ and right-continuous sample paths, see Example I.3.10 in Andersen et al. (1993). 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 $X^1, \dots, X^n$. The filtration is given by $\mathcal{G}_t = \sigma((X(s)^1, \dots, X(s)^n), s \leq t)$. For any $i \in \{1, \dots, n\}$, 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 $Y_t^{ih} = \mathbf{1}_{\{X^i(t^-)=h\}}$. As before, we can consider $\mathfrak{N}$ (resp., Y^h) by aggregating the processes $\mathfrak{N}^i$ (resp., 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 $(\lambda_{hj}(t) Y_t^h)_t$ and the data $\mathbf{D} = ((X(s)^1, \dots, X(s)^n), s \leq T)$. Note that, for each (h, j), $(N_t^{hj})_t$ is a univariate Aalen process associated with the filtration $(\mathcal{G}_t)_t$. 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 refer the reader to Andersen et al. (1993), p. 126, Reynaud-Bouret (2006), Kim (1999) or Comte et al. (2011) for more details.

Censored processes

Previous models can be combined. For instance, as Kim (1999), following Lo (1992), we can consider censored Poisson processes. More precisely, let $M^1, \dots, M^n$ be n i.i.d. Poisson processes with common intensity λ and let $Y^1, \dots, Y^n$ be n i.i.d. non-negative predictable processes that are independent of the M^i's. For instance, we can consider $Z_1, \dots, Z_n$ n i.i.d. random variables and set, for any i, $Y_t^i = \mathbf{1}_{Z_i \geq t}$. Defining

$$N_t = \sum_{i=1}^n \int_0^t Y_s^i \mathrm{d}M_s^i,$$

we obtain a counting process obeying the Aalen multiplicative intensity model since its compensator can be written as

$$\Lambda_t = \int_0^t \lambda(s) Y_s \mathrm{d}s,$$

with $Y_t = \sum_{i=1}^n Y_t^i$ which is a non-negative predictable process. In this case, we have $\mathbf{D} = (Z_1, \dots, Z_n, M^1, \dots, M^n)$.

2.3 Assumptions

Let the true intensity λ_0 to be estimated be such that $\int_0^T \lambda_0(t)\mathrm{d}t < \infty$. We denote by $\mathbb{P}_{\lambda_0}^{(n)}$ and $\mathbb{E}_{\lambda_0}^{(n)}$ the probability measure and the expectation associated with λ_0, respectively. We now state some conditions concerning the asymptotic behaviour of Y_t. Define

$$\mu_n(t) := \mathbb{E}_{\lambda_0}^{(n)}[Y_t] \quad \text{and} \quad \tilde{\mu}_n(t) := \frac{1}{n}\mu_n(t). \tag{4}$$

We assume the existence of a non-random set $\Omega \subseteq [0, T]$ such that there are positive constants m_1 and m_2 satisfying for any n,

$$m_1 \leq \inf_{t\in\Omega} \tilde{\mu}_n(t) \leq \sup_{t\in\Omega} \tilde{\mu}_n(t) \leq m_2, \tag{5}$$

and there exists $\alpha \in (0, 1)$ such that, if

$$\Gamma_n := \left\{ \sup_{t\in\Omega} |n^{-1}Y_t - \tilde{\mu}_n(t)| \le \alpha m_1 \right\} \cap \left\{ \sup_{t\in[0,\,T]\setminus\Omega} Y_t = 0 \right\},$$

then

$$\lim_{n\to\infty} \mathbb{P}^{(n)}_{\lambda_0}(\Gamma_n) = 1. \tag{6}$$

These assumptions allow to control quite precisely the number of jumps of the process N on subsets of Ω. In particular, the number of jumps of N is bounded by the number of jumps of a Poisson process with intensity $n\lambda(\cdot)$. This trick allows us to use the classical machinery for density estimation developed by Ghosal and van der Vaart (2007) in the density estimation setting. Actually, assumption (6) is very mild as well as the right-hand side of (5). The left-hand side of (5) is most of the time unavoidable and variations of it are commonly used in the literature; see, for instance, Kim (1999), Reynaud-Bouret (2006), Comte et al. (2011) or Hansen et al. (2015). For inhomogeneous Poisson processes, conditions (5) and (6) are obviously satisfied with $m_1 = m_2 = 1$ and $\Omega = [0, T]$ since, for any $t \in [0, T]$, $Y_t = \mu_n(t) = n$. It may be the case for the other previously described examples, such as right-censoring, by using the following lemma.

Lemma 1. *Assume that Y_t can be written as*

$$Y_t = \sum_{i=1}^{n} Y_t^i,$$

where $Y_t^i = \mathbf{1}_{Z_i \ge t}$ and the Z_i's are i.i.d. with support denoted by $\mathcal{S}$. Then, (5) *and* (6) *are satisfied with*

- $\Omega = [0, T]$ *if* $M_{\mathcal{S}} > T$
- $\Omega = [0, M_{\mathcal{S}}]$ *if* $M_{\mathcal{S}} \le T$ *and* $\mathbb{P}(Z_1 = M_{\mathcal{S}}) > 0$,

where $M_{\mathcal{S}} = \max \mathcal{S}$.

Proof. For any $t \in [0, T]$, $\tilde{\mu}_n(t) = \mathbb{P}(Z_1 \ge t)$ and the right-hand side of (5) is true with $m_2 = 1$. For the left-hand side, we observe that we can take

$$m_1 = \mathbb{P}(Z_1 \ge \min\{T, M_{\mathcal{S}}\}).$$

If $M_{\mathcal{S}} > T$ then $m_1 > 0$ by definition of $\mathcal{S}$ and $[0, T] \setminus \Omega = \emptyset$. If $M_{\mathcal{S}} \le T$, then $m_1 = \mathbb{P}(Z_1 = M_{\mathcal{S}}) > 0$ and $[0, T] \setminus \Omega = (M_{\mathcal{S}}, T]$. By definition of $\mathcal{S}$, for any $t > M_{\mathcal{S}}$, $Y_t = 0$ almost surely. To prove (6), we write

$$\begin{aligned}\sup_{t\in\Omega} |n^{-1}Y_t - \tilde{\mu}_n(t)| &= \sup_{t\in\Omega} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{t\le Z_i} - \mathbb{P}(t \le Z_1) \right| \\ &\le \sup_{u\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{-Z_i\le u} - \mathbb{P}(-Z_1 \le u) \right|.\end{aligned}$$

So, for $\alpha \in (0, 1)$, the Dvoretzky–Kiefer–Wolfowitz inequality gives

$$\mathbb{P}\left(\sup_{t\in\Omega} |n^{-1}Y_t - \tilde{\mu}_n(t)| > \alpha m_1\right) \leq 2\exp(-2n(\alpha m_1)^2)$$

and $\gamma_n = \mathbb{P}_{\lambda_0}^{(n)}(\Gamma_n^c)$ goes to 0 at an exponential rate. □

The conditions of the previous lemma ensure that the stochastic intensity $\lambda(t)Y_t$ is bounded from below on Ω, which is classical in the literature (see the previously mentioned references). This implies that, on any non-empty open interval, the point process N has positive probability of jumping; see (1). In particular, if the distribution of the Z_i's is absolutely continuous and supported on $[0, T]$, then $M_{\mathcal{S}} = T$, but $\mathbb{P}(Z_1 = M_{\mathcal{S}}) = 0$ and the assumptions of Lemma 1 are not satisfied with $\Omega = [0, T]$. This case, which might be as difficult as the case of posterior asymptotics on the whole positive real line, is further discussed in Section 5.2.

In the following sections, performance of inferences is only measured over the set Ω assumed to be known, N has no jumps on $[0, T] \setminus \Omega$ almost surely.

3 Posterior contraction rates for Aalen counting processes

In this section, we present the main result providing sufficient conditions for assessing contraction rates of posterior distributions of intensities in general Aalen models.

Although Aalen processes do not lead to i.i.d. observations and estimating λ, the deterministic part of the stochastic intensity, is not as estimating a density, there are strong connections between the two problems and our aim is to provide sufficient conditions similar to those considered in the density estimation problem in Ghosal et al. (2000). This allows us to appeal to the large literature on posterior concentration rates for density estimation and apply the existing results that have been proved for various types of prior models for density estimation to the present framework; see Section 4 for an illustration of this through various examples.

Before stating the theorem, we need to introduce some more notation. We define the parameter space as

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

To emphasize the connections between Aalen intensities and density models, for any $\lambda \in \mathcal{F}$, we introduce the following parametrization

$$\lambda = M_\lambda \times \bar{\lambda},$$

where $M_\lambda = \int_\Omega \lambda(t)\mathrm{d}t$ and $\bar{\lambda} \in \mathcal{F}_1$, with $\mathcal{F}_1 = \{\lambda \in \mathcal{F} : \int_\Omega \lambda(t)\mathrm{d}t = 1\}$. We denote by $\|\cdot\|_1$ the $\mathbb{L}_1$-norm over $\mathcal{F}$: for all λ, $\lambda' \in \mathcal{F}$,

$$\|\lambda - \lambda'\|_1 = \int_\Omega |\lambda(t) - \lambda'(t)|\mathrm{d}t.$$

The Kullback–Leibler divergence of $\lambda \in \mathcal{F}$ from λ_0 is defined as

$$\mathrm{KL}(\lambda_0;\, \lambda) = \mathbb{E}_{\lambda_0}^{(n)}[\ell_n(\lambda_0) - \ell_n(\lambda)]. \tag{7}$$

For the sake of simplicity, we restrict attention to the case where M_λ and $\bar{\lambda}$ are a priori independent so that the prior probability measure π on $\mathcal{F}$ is the product measure $\pi_M \otimes \pi_1$, where π_M is a probability measure on $\mathbb{R}_+$ and π_1 is a probability measure on $\mathcal{F}_1$.

Let v_n be a positive sequence such that $v_n \to 0$ and $nv_n^2 \to \infty$. For every $j \in \mathbb{N}$, we define

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

where $M_{\lambda_0} = \int_\Omega \lambda_0(t)\mathrm{d}t$ and $\bar{\lambda}_0 = M_{\lambda_0}^{-1}\lambda_0$. For $H > 0$, we define

$$\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H) = \Big\{\bar{\lambda} \in \mathcal{F}_1 :\ h^2(\bar{\lambda}_0,\, \bar{\lambda}) \leq v_n^2/(1 + \log\|\bar{\lambda}_0/\bar{\lambda}\|_\infty),\ \ \|\bar{\lambda}_0/\bar{\lambda}\|_\infty \leq n^H,\ \ \|\bar{\lambda}\|_\infty \leq H\Big\},$$

where

$$h^2(\bar{\lambda}_0,\, \bar{\lambda}) = \int_\Omega \left(\sqrt{\bar{\lambda}_0(t)} - \sqrt{\bar{\lambda}(t)}\right)^2 \mathrm{d}t$$

is the squared Hellinger distance between $\bar{\lambda}_0$ and $\bar{\lambda}$ and $\|\cdot\|_\infty$ stands for the sup-norm.

In what follows, for any set Θ equipped with a semi-metric d and any real number $\epsilon > 0$, we denote by $D(\epsilon,\, \Theta,\, d)$ the ϵ-packing number of Θ, that is, the maximal number of points in Θ such that the d-distance between every pair is at least ϵ. Since $D(\epsilon,\, \Theta,\, d)$ is bounded above by the $(\epsilon/2)$-covering number, namely, the minimal number of balls of d-radius $\epsilon/2$ needed to cover Θ, with abuse of language, we will just speak of covering numbers.

Theorem 1. *Assume that conditions* (5) *and* (6) *are satisfied and that, for some $k \geq 1$, there exists a constant $C_{1k} > 0$ such that*

$$\mathbb{E}_{\lambda_0}^{(n)}\left[\left(\int_\Omega [Y_t - \mu_n(t)]^2 \mathrm{d}t\right)^k\right] \leq C_{1k} n^k. \tag{8}$$

Assume that the prior π_M on the mass M_λ is absolutely continuous with respect to Lebesgue measure and has positive and continuous density on $\mathbb{R}_+$, while the prior π_1 on $\bar{\lambda}$ satisfies the following conditions for some constant $H > 0$:

(*i*) *There exists $\mathcal{F}_n \subseteq \mathcal{F}_1$ such that, for a positive sequence $v_n = o(1)$ and $v_n^2 \geq (n/\log n)^{-1}$,*

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

with

$$\kappa_0 = m_2^2 M_{\lambda_0}\left\{\frac{4}{m_1}\left[1 + \log\left(\frac{m_2}{m_1}\right)\right]\left(1 + \frac{m_2^2}{m_1^2}\right) + \frac{m_2(2M_{\lambda_0}+1)^2}{m_1^2 M_{\lambda_0}^2}\right\}, \tag{9}$$

and, for any $\xi,\, \delta > 0$,

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

(*ii*) *For all* $\zeta,\, \delta > 0$, *there exists* $J_0 > 0$ *such that, for every* $j \ge J_0$,

$$\frac{\pi_1(\bar{S}_{n,j})}{\pi_1(\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H))} \le e^{\delta n \min\{(j+1)^2 v_n^2,\, 1\}}$$

and

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

Then, there exists a constant $J_1 > 0$ *such that*

$$\begin{aligned}&\mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 v_n \mid \mathbf{D})]\\ &\qquad = \gamma_n + O((\log n)^{3k/2}(n v_n^2)^{-3k/2} + (\log n)^k (n v_n^2)^{-k} + (n v_n^2)^{-2k+1}(\log n)^{2k-1}),\end{aligned}$$

with $\gamma_n = \mathbb{P}^{(n)}_{\lambda_0}(\Gamma_n^c)$.

If $\gamma_n = 0$, as for the Poisson case, or goes to 0 at an exponential rate, then it is negligible with respect to the other terms on the right-hand side of the previous equality. Furthermore, as soon as $n v_n^2 \gtrsim n^\delta$ for some $\delta > 0$, the above right-hand side satisfies

$$\mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 v_n \mid \mathbf{D})] = \gamma_n + O((\log n)^k (n v_n^2)^{-k})$$

for $k \ge 1$, so that (8) is verified. The exponent k in (8) can be any integer larger than or equal to 1 and does not influence the posterior contraction rate v_n. It however influences the quantity $\mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 v_n \mid \mathbf{D})]$ and the larger k, the better.

To the best of our knowledge, the only other papers dealing with posterior contraction rates in related models are those of Belitser et al. (2015), Gugushvili and Spreij (2013) and Kirichenko and van Zanten (2015), where inhomogeneous Poisson processes are considered. Theorem 1 differs in two aspects from their approach. First, we do not confine ourselves to inhomogeneous Poisson processes; an important consequence of this difference is that we cannot view the likelihood as that of i.i.d. observations, so that specific tests need to be constructed. Secondly, our conditions are different: we do not assume that λ_0 is bounded below away from zero and we lower bound the prior mass of Hellinger-type neighbourhoods of λ_0, as in Theorem 2.2 of Ghosal et al. (2000), instead of sup-norm neighbourhoods. This can lead to significant improvements on the rate in some cases; see, for instance, Section 4.1. In Theorem 1, our aim is to propose conditions to assess posterior concentration rates for intensity functions resembling those used in the density model obtained by parametrizing λ as $\lambda = M_\lambda \times \bar{\lambda}$, with $\bar{\lambda}$ a probability density on Ω.

The proof of Theorem 1 is reported in Section 6. It is an application of Theorems 1 and 3 of Ghosal and van der Vaart (2007) to the setup of counting processes by using some properties of martingale processes. Thus, we first prove that neighbourhoods $B_n(\bar{\lambda}_0;\, Cv_n,\, k)$ defined in Section 2 of Ghosal and van der Vaart (2007) contain $\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H)$ for some H and $C > 0$. For $\lambda \in \mathcal{F}$ such that $\bar{\lambda} \in \bar{B}_n(\bar{\lambda}_0;\, v_n,\, H)$, we first express $\mathrm{KL}(\lambda_0;\, \lambda)$ as a function of the Kullback–Leibler divergence between the renormalized intensities, which can be regarded as densities; see (20). The expression is then bounded by using the Hellinger distance $h(\bar{\lambda}_0,\, \bar{\lambda})$. Hence, the main difficulty in this step is to control $\mathbb{E}^{(n)}_{\lambda_0}[|\ell_n(\lambda_0) - \ell_n(\lambda) - \mathbb{E}^{(n)}_{\lambda_0}[\ell_n(\lambda_0) - \ell_n(\lambda)]|^{2k}]$ for $k \geq 1$. We proceed by using Rosenthal's inequalities for martingales associated with our counting processes. These results are presented in Proposition 1 of Section 6. Secondly, we construct tests based on the $\mathbb{L}_1$-distance between intensities. These tests are derived and controlled by using a specific concentration inequality for counting processes established by Hansen et al. (2015). See Lemma 2 for the construction of tests and Proposition 2 for the control of their type I and type II errors.

Remark 1. *Condition (8) is obviously satisfied for inhomogeneous Poisson processes and also when Y_t can be written as $Y_t = \sum_{i=1}^n Y_t^i$, where the $Y_t^i = \mathbf{1}_{Z_i \geq t}$ and the Z_i's are i.i.d. Indeed, if for every $i = 1, \ldots, n$, we set $V_i = \mathbf{1}_{Z_i \geq t} - \mathbb{P}(Z_1 \geq t)$, then, for $k \geq 2$,*

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

by Hölder and Rosenthal's inequalities; see, for instance, Theorem C.2 of Härdle et al. (1998). Under mild conditions, similar computations can be performed for finite state Markov processes.

As explained at the beginning of Section 3, our conditions intentionally resemble those considered in the density estimation problem. The entropy condition in (*ii*) of Theorem 1 is similar to the one of Ghosal et al. (2000). Apart from the mild constraints $\|\bar{\lambda}_0/\bar{\lambda}\|_\infty \leq n^H$ and $\|\bar{\lambda}\|_\infty \leq H$, the set $\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H)$ is the same as the one considered in Theorem 2.2 of Ghosal et al. (2000). One can sharpen the rate (to attain a rate close to the parametric case) by replacing $\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H)$ by

$$\tilde{B}_n = \left\{\bar{\lambda} \in \bar{B}_n(\bar{\lambda}_0;\, v_n,\, H) : \int_0^T \bar{\lambda}_0(t) \sum_{j=1}^k \log^{2j}(\bar{\lambda}_0(t)/\bar{\lambda}(t)) \mathrm{d}t \leq v_n^2\right\}.$$

In this case the result of Theorem 1 becomes

$$\mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 v_n \mid \mathbf{D})] = \gamma_n + O((nv_n^2)^{-k}).$$

This might be of interest to obtain the parametric rate $1/\sqrt{n}$ in some cases. We now apply Theorem 1 to various prior models.

4 Illustrations with different families of priors

As discussed in Section 3, the conditions of Theorem 1 to derive posterior contraction rates are very similar to those considered in the literature for density estimation so that existing results involving different families of prior distributions can be adapted to Aalen multiplicative intensity models. Some applications are presented below. We still denote $\gamma_n = \mathbb{P}^{(n)}_{\lambda_0}(\Gamma_n^c)$.

4.1 Monotone nondecreasing intensity functions

In this section, we deal with estimation of monotone nondecreasing intensity functions, which is equivalent to considering monotone nondecreasing density functions $\bar{\lambda}$ in the above described parametrization. To construct a prior on the set of monotone nondecreasing densities over $[0, T]$, we use their representation as mixtures of uniform densities as in Williamson (1956) and consider a Dirichlet process as a prior on the mixing distribution:

$$\bar{\lambda}(\cdot) = \int_0^\infty \frac{\mathbf{1}_{(T-\theta,\, T)}(\cdot)}{\theta} \mathrm{d}P(\theta), \qquad P \mid A,\, G \sim \mathrm{DP}(AG), \tag{10}$$

where G is a distribution on $[0, T]$ having density g with respect to Lebesgue measure. This prior has been studied by Salomond (2014). Here, we extend his results to the case of monotone nondecreasing intensity functions of Aalen processes. We consider the same assumption on G as in Salomond (2014): there exist $a_1,\, a_2 > 0$ such that

$$\theta^{a_1} \lesssim g(\theta) \lesssim \theta^{a_2} \quad \text{ for all } \theta \text{ in a neighbourhood of } 0. \tag{11}$$

The following result holds.

Corollary 1. *Assume that the counting process N verifies conditions (5) and (6) and that inequality (8) is satisfied for some $k \geq 1$. Consider a prior π_1 on $\bar{\lambda}$ satisfying conditions (10) and (11) and a prior π_M on M_λ that is absolutely continuous with respect to Lebesgue measure with positive and continuous density on $\mathbb{R}_+$. Suppose that λ_0 is monotone non-decreasing and bounded on $[0, T]$. Let $\bar{\epsilon}_n = (n/\log n)^{-1/3}$. Then, there exists a constant $J_1 > 0$ such that*

$$\mathbb{E}^{(n)}_{\lambda_0}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1 \bar{\epsilon}_n \mid \mathbf{D})] = \gamma_n + O((\log n)^k (n\bar{\epsilon}_n^2)^{-k}).$$

The proof is reported in Section 6.

4.2 Log-spline and log-linear priors on λ

For simplicity of presentation, we set $T = 1$. We consider a log-spline prior of order q as in Section 4 of Ghosal et al. (2000). In other words, $\bar{\lambda}$ is parametrized as

$$\log \bar{\lambda}_\theta(\cdot) = \theta^t \underline{B}_J(\cdot) - c(\theta), \qquad \text{with } \exp(c(\theta)) = \int_0^1 e^{\theta^t \underline{B}_J(x)} \mathrm{d}x,$$

where $\underline{B}_J = (B_1, \dots, B_J)$ is the qth order B-spline defined in de Boor (1978) associated with K fixed knots, so that $J = K + q - 1$, see Ghosal et al. (2000) for more details. Consider a prior on θ in the form $J = J_n = \lfloor n^{1/(2\alpha+1)} \rfloor$, $\alpha \in [1/2, q]$ and, conditionally on J, the prior is absolutely continuous with respect to Lebesgue measure on $[-M, M]^J$ with density bounded from below and above by c^J and C^J, respectively. Consider an absolutely continuous prior on M_λ having positive and continuous density on $\mathbb{R}_+$. We then have the following posterior concentration result.

Corollary 2. *For the above prior, if* $\|\log \lambda_0\|_\infty < \infty$ *and* λ_0 *is Hölder with regularity* $\alpha \in [1/2, q]$*, then, under condition* (8)*, there exists a constant* $J_1 > 0$ *so that*

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda : \ \|\lambda - \lambda_0\|_1 > J_1 n^{-\alpha/(2\alpha+1)} \mid \mathbf{D})] = \gamma_n + O(n^{-k/(2\alpha+1)}(\log n)^k).$$

Proof. Set $\epsilon_n = n^{-\alpha/(2\alpha+1)}$. Using Lemma 4.1 of Ghosal et al. (2000), there exists $\theta_0 \in \mathbb{R}^J$ such that $h(\bar{\lambda}_{\theta_0}, \bar{\lambda}_0) \lesssim \|\log \bar{\lambda}_{\theta_0} - \log \bar{\lambda}_0\|_\infty \lesssim J^{-\alpha}$, which, combined with Lemma 4.4 of Ghosal et al. (2000), leads to

$$\pi_1(\bar{B}_n(\bar{\lambda}_0; \epsilon_n, H)) \geq e^{-C_1 n \epsilon_n^2}.$$

Lemma 4.5, together with Theorem 4.5 of Ghosal et al. (2000), controls the entropy of $\bar{S}_{n,j}$ and its prior mass for j larger than some fixed constant J_0. □

With such families of priors, it is more interesting to work with non-normalized λ_θ. We can write

$$\lambda_{A,\theta}(\cdot) = A \exp\left(\theta^t \underline{B}_J(\cdot)\right), \quad A > 0,$$

so that a prior on λ is defined as a prior on A, say π_A, absolutely continuous with respect to Lebesgue measure, having positive and continuous density, and the same type of prior on θ as above is considered. Corollary 2 still holds although it is not a direct consequence of Theorem 1, since $M_{\lambda_{A,\theta}} = A \exp(c(\theta))$ is not a priori independent of $\bar{\lambda}_{A,\theta}$. However, introducing A allows adapting Theorem 1 to this case. The practical advantage of the latter representation is that it avoids computing the normalizing constant $c(\theta)$.

In a similar manner, we can replace spline basis with other orthonormal bases, as considered in Rivoirard and Rousseau (2012), leading to the same posterior concentration rates as in density estimation. More precisely, consider intensities parametrized as

$$\bar{\lambda}_\theta(\cdot) = e^{\sum_{j=1}^J \theta_j \phi_j(\cdot) - c(\theta)}, \quad e^{c(\theta)} = \int_{\mathbb{R}^J} e^{\sum_{j=1}^J \theta_j \phi_j(x)} \mathrm{d}x,$$

where $(\phi_j)_{j=1}^{\infty}$ is an orthonormal basis of $\mathbb{L}_2([0,\,1])$, with $\phi_1 = 1$. Write $\eta = (A,\,\theta)$, with $A > 0$, and

$$\lambda_\eta(\cdot) = Ae^{\sum_{j=1}^{J}\theta_j\phi_j(\cdot)} = Ae^{c(\theta)}\bar{\lambda}_\theta(\cdot).$$

Let $A \sim \pi_A$ and consider the same family of priors as in Rivoirard and Rousseau (2012):

$$J \sim \pi_J,$$
$$j^\beta\theta_j/\tau_0 \overset{\text{ind}}{\sim} g, \quad \forall\, j \le J, \quad \text{and} \quad \theta_j = 0, \quad \forall\, j > J,$$

where g is a positive and continuous density on $\mathbb{R}$ and there exist $s \ge 0$ and $p > 0$ such that

$$\log \pi_J(J) \asymp -J(\log J)^s, \qquad \log g(x) \asymp -|x|^p, \quad s = 0,\, 1,$$

when J and $|x|$ are large. Rivoirard and Rousseau (2012) prove that this prior leads to minimax adaptive posterior concentration rates over collections of positive and Sobolev (or more generally Besov) classes of densities. Their proof easily extends to prove assumptions (i) and (ii) of Theorem 1.

Corollary 3. *Consider the above described prior on an intensity function* λ *on* $[0,\,1]$*. Assume that* λ_0 *is positive and belongs to a Sobolev class with smoothness* $\alpha > 1/2$*. Under condition* (8)*, if* $\beta < 1/2 + \alpha$*, there exists a constant* $J_1 >$ *so that*

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda :\ \|\lambda - \lambda_0\|_1 > J_1(n/\log n)^{-\alpha/(2\alpha+1)}(\log n)^{(1-s)/2} \mid \mathbf{D})]$$
$$= \gamma_n + O((n/\log n)^{-k/(2\alpha+1)}(\log n)^{sk}).$$

Note that the constraint $\beta < 1/2 + \alpha$ is satisfied for all $\alpha > 1/2$ as soon as $\beta < 1$ and, as in Rivoirard and Rousseau (2012), the prior leads to adaptive minimax posterior contraction rates over collections of Sobolev balls.

5 Numerical illustration

We propose a numerical illustration for nonparametric Bayesian estimation of intensity functions in the right-censoring model. We first describe the prior model together with an ad-hoc MCMC algorithm designed for the right-censoring context and present numerical illustrations.

Recall that, for $i = 1,\, \ldots,\, n$, we observe $Z_i = \min\{T_i,\, C_i\}$, where $T_i \sim f(\cdot)$, T_i and C_i are independent, $C_i \in [0,\,1]$. In the following, we take $T = 1$, so that we observe the process on the interval $[0,\,1]$. Using the factorization $\lambda = M_\lambda \times \bar{\lambda}$, we set the following prior distribution on $(M_\lambda,\, \bar{\lambda})$:

$$\begin{aligned} M_\lambda &\sim \text{Gamma}(a_M,\, b_M), \\ \bar{\lambda}(t) &= \int_0^\infty \frac{1}{\theta}\mathbf{1}_{(1-\theta,\,1)}(t)\mathrm{d}P(\theta), \\ \text{with } P(\cdot) &\sim \text{DP}(AG), \\ \text{and } G(\cdot) &\sim \left(1 + \frac{1}{\text{Gamma}(\alpha,\,\beta)}\right)^{-1}. \end{aligned} \tag{12}$$

As a consequence, $\bar{\lambda}$ is a monotone nondecreasing density on [0, 1] and satisfies assumption (11) given in Section 4.1.

We propose to sample the posterior distribution of $\lambda = M_\lambda \times \bar{\lambda}$ using an adapted slice sampler MCMC algorithm based on the stick-breaking version of the Dirichlet process.

5.1 MCMC algorithm

In its stick-breaking version, $\bar{\lambda}$ is written as $\bar{\lambda}(t) = \sum_{k=1}^{\infty} w_k \frac{\mathbf{1}_{(1-\theta_k,\,1)}(t)}{\theta_k}$, $t \in [0,\,1]$, where

$$\begin{aligned} &w_1 = v_1, \quad w_k = v_k \prod_{j=1}^{k-1}(1-v_j), \quad \forall\, k \geq 2, \\ &v_k \overset{\text{i.i.d}}{\sim} \text{Beta}(1,\,A), \quad \forall\, k \geq 1, \\ &\theta_k \overset{\text{i.i.d}}{\sim} G(\cdot), \quad \forall\, k \geq 1. \end{aligned} \tag{13}$$

As a consequence,

$$\Lambda_t = M_\lambda \times \bar{\Lambda}(t) = M_\lambda \sum_{k=1}^{\infty} w_k F_{\mathcal{U}(1-\theta_k,\,1)}(t), \quad t \in [0,\,1],$$

where $F_{\mathcal{U}(1-\theta_k,\,1)}$ is the cumulative distribution function of a uniform distribution over $(1-\theta_k,\,1)$. We introduce $\mathcal{O} = \{i \in \{1,\,\ldots,\,n\} \mid \delta_i = 1\}$, $n^\star = \#\mathcal{O}$, $\boldsymbol{\theta} = (\theta_k)_{k\geq 1}$ and $\mathbf{v} = (v_k)_{k\geq 1}$. Combining with (3), the likelihood becomes

$$\begin{aligned} \mathcal{L}_n(\mathbf{Z};\,\mathbf{v},\,\boldsymbol{\theta},\,M_\lambda) &= M_\lambda^{n^\star}\left(\prod_{i\in\mathcal{O}} \bar{\lambda}(Z_i)\right)\exp\left(-M_\lambda \sum_{i=1}^{n} \bar{\Lambda}(Z_i)\right) \\ &= M_\lambda^{n^\star}\left(\prod_{i\in\mathcal{O}}\sum_{k=1}^{\infty} w_k \frac{\mathbf{1}_{(1-\theta_k,\,1)}(Z_i)}{\theta_k}\right)\exp\left(-M_\lambda \sum_{k=1}^{\infty} w_k H(\theta_k)\right), \end{aligned} \tag{14}$$

where

$$H(\theta_k) = \sum_{i=1}^{n} F_{\mathcal{U}(1-\theta_k,\,1)}(Z_i). \tag{15}$$

We use the slice sampling strategy proposed by Walker (2007) to deal with $\prod_{i\in\mathcal{O}} \bar{\lambda}(Z_i)$, based on the auxiliary variables $\mathbf{u} = (u_i)_{i\in\mathcal{O}}$, and we introduce a deterministic truncation K_t to approximate $\sum_{k=1}^{\infty} w_k H(\theta_k)$. The effect of the truncation is studied in the numerical illustration. This leads to the following approximation of (14):

$$\begin{aligned} \bar{\mathcal{L}}_{n,K_t}(\mathbf{u},\,\mathbf{Z};\,\mathbf{v},\,\boldsymbol{\theta},\,M_\lambda) = M_\lambda^{n^\star} &\left(\prod_{i\in\mathcal{O}}\sum_{k=1}^{\infty} w_k \frac{\mathbf{1}_{(1-\theta_k,1)}(Z_i)}{\theta_k}\frac{\mathbf{1}_{(0,\,w_k)}(u_i)}{w_k}\right) \\ &\times \exp\left(-M_\lambda \sum_{k=1}^{K_t} w_k H(\theta_k)\right). \end{aligned} \tag{16}$$

Because the sequence $(w_k)_{k\geq 1}$ is stochastically decreasing, the infinite sum in (16) only has (a.s.) a finite number of positive terms. We denote by $K_i^\star = \min\{k \in \mathbb{N}^* \mid \forall\, l \geq$

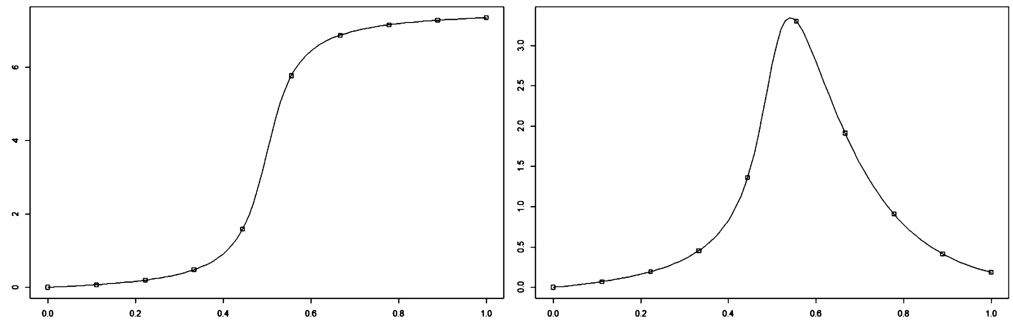


Figure 1: Simulation study. The hazard rate function λ (left panel) and its corresponding density function f (right panel).

$k, w_l \leq u_i\}$, $K^\star = \max\{K_t, (K_i^\star)_{i\in\mathcal{O}}\}$, $c_i \in \mathbb{N}^*$ the allocation variable of individual $i \in \mathcal{O}$ and $\mathbf{c} = (c_i)_{i\in\mathcal{O}}$. The augmented likelihood can then be written as

$$\widetilde{\mathcal{L}}_{n,K_t}(\mathbf{c},\, \mathbf{u},\, \mathbf{Z};\, \mathbf{v},\, \boldsymbol{\theta},\, M_\lambda) \;=\; M_\lambda^{n^\star}\left(\prod_{i\in\mathcal{O}} \frac{\mathbf{1}_{(1-\theta_{c_i},\,1)}(Z_i)}{\theta_{c_i}}\frac{\mathbf{1}_{(0,\,w_{c_i})}(u_i)}{w_{c_i}}\right) \times \exp\left(-M_\lambda \sum_{k=1}^{K_t} w_k H(\theta_k)\right) \times \prod_k w_k^{n_k}, \quad (17)$$

where $n_k = \#\{i \in \mathcal{O} \mid c_i = k\}$.

Following (17), the MCMC will sequentially sample M_λ, $\mathbf{u} = (u_i)_{i\in\mathcal{O}}$, $\mathbf{c} = (c_i)_{i\in\mathcal{O}}$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{K^\star})$ and $\mathbf{v} = (v_1, \ldots, v_{K^\star})$. We detail each step of the algorithm in Section A.2 in Supplementary material.

5.2 Numerical results

We conduct a simulation study to illustrate the performance of the MCMC algorithm based on the truncation. In a first paragraph, we present the parameters used to simulate the data and the prior distribution. In a second part, we study the influence of the truncation parameter K_t on the quality of the estimation. In this part, we also assess the convergence of the algorithm using Gelmand and Rubin diagnostic tools, implemented in the coda R-package.

Simulation parameters

We consider the following common hazard function:

$$\lambda(t) = 2.5\left[\arctan\left(20\,t - 10\right) - \arctan\left(-10\right)\right].$$

We plot λ, f, where $f(t) = \lambda(t)\exp\{-\int_0^t \lambda(u)\mathrm{d}u\}$, on Figure 1. The censoring times C_i are distributed as

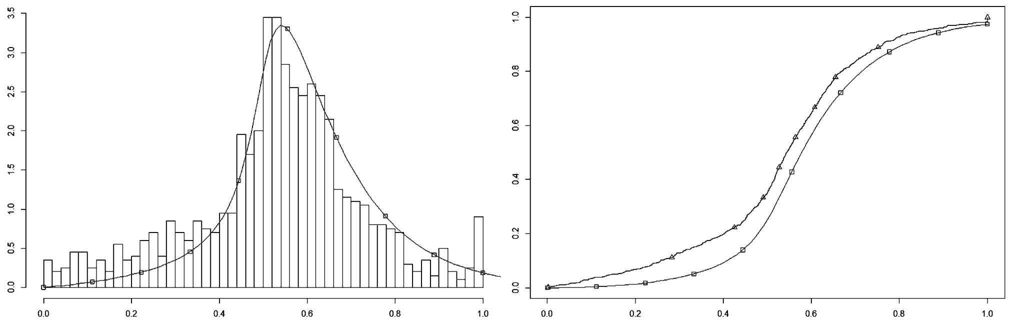


Figure 2: Simulated data. On the left, histogram of the $Z_i = \min\{X_i, C_i\}$, $i = 1, \ldots, n$, with $n = 2000$ and f (density of the X_i, line with squares). On the right, empirical cumulative distribution function (line with triangles) of the $(Z_i)_{i=1,\ldots,2000}$, and F (line with squares).

$$C_i \overset{\text{i.i.d}}{\sim} \frac{1}{3}\,\mathcal{U}_{(0,\,1)} + \frac{2}{3}\,\delta_{\{1\}}.$$

The chosen λ and censoring time distribution ensure a censoring rate equal to $\mathbb{P}(\delta_i = 0) = 1 - \mathbb{P}(T_i \leq C_i) = 1 - \frac{1}{3}\int_0^1 F(t)\mathrm{d}t - \frac{2}{3}F(1) \simeq 0.2146\%$.

We highlight that the assumptions of Lemma 1 are satisfied: $T = 1$, $M_{\mathcal{S}} = 1$ and $\mathbb{P}(Z_1 = M_{\mathcal{S}}) > 0$. Interestingly, we noticed in various simulations that when $\mathbb{P}(Z_1 = M_{\mathcal{S}}) = 0$, the estimates of λ are of very low quality.

With these parameters, we simulate 10 datasets (half of them with $n = 2000$, the others with $n = 1000$). An arbitrarily chosen dataset is plotted on Figure 2.

Hyperparameters

Going back to the prior distribution described in (12), we set the hyperparameters $(A,\, a_M,\, b_M,\, a)$ as follows:

$$A = 15, \quad (a_M,\, b_M) = (4, 1).$$

The choice of $(\alpha,\, \beta)$ can influence a lot the inference. To avoid this problem, we propose a hierarchical strategy on α, setting $\alpha \sim \text{Gamma}(1,\, 1)$ and $\beta = 3$. In Figure 3, we plot 100 realizations of λ under this prior distribution, illustrating the large support of the prior distribution on λ.

Remark 2. *The parameter A is fixed in this experiment but we could put a prior distribution on it; see Donnet et al. (2014), for instance.*

Effect of the truncation K_t

To study the effect of truncating with K_t, we have simulated one dataset with $n = 2000$, and run the MCMC algorithm with $K_t = 20, 80, 100, 500, 1000$. From the output in

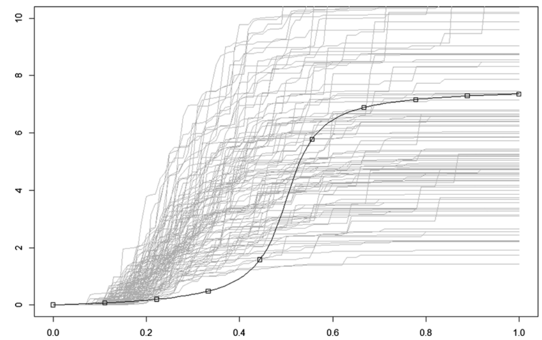
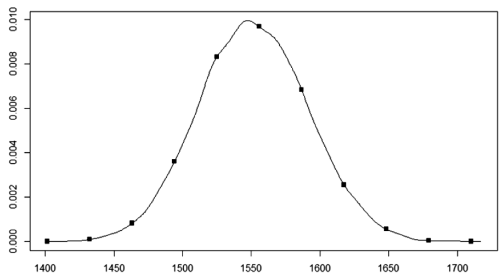
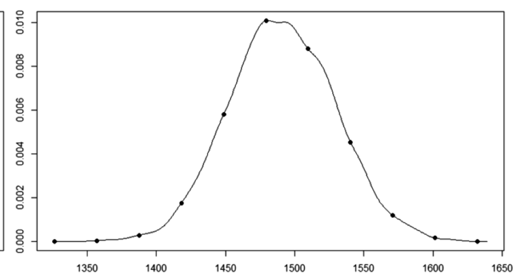


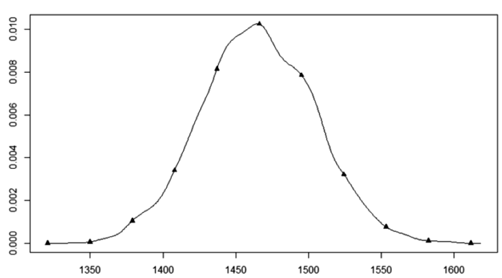
Figure 3: Prior distribution. 100 realizations of λ under the prior distribution (grey) and the true λ (line with squares).



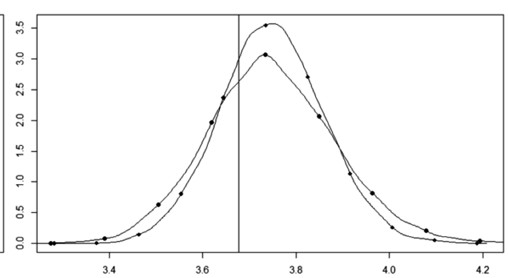
(a)



(b)



(c)



(d)

Figure 4: Posterior distributions of M_λ for different values of K_t: (a) $K_t = 20$; (b) $K_t = 80$; (c) $K_t = 100$; (d) $K_t = 500$ and 1000.

terms of the (approximation) of the posterior distribution of M_λ we observe that for $K_t = 500$ and $K_t = 1000$, the results are equivalent and the posterior distribution concentrates around the true value. Not surprisingly, for small values of K_t, the estimation degenerates and the posterior distributions concentrate around aberrant values. This is shown in Figure 4.

It appears that, for small values of K_t, K^* is much larger than K_t, which explains the bad behaviour of the approximated posterior distribution. This is illustrated in Figure 5

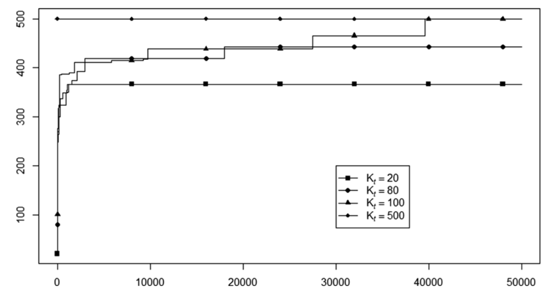


Figure 5: Evolution of $K^\star$ over the iterations of the MCMC algorithm for the runs with different values of K_t.

which represents the evolution of $K^\star$ throughout the iterations of the MCMC algorithm. From several experiments, we noted that when, over the iterations of the MCMC, $K^\star$ exceeds K_t, the estimation quality collapses. As a consequence, we propose – as a practical tool – to tune K_t along the iterations $K^\star \leq K_t$. As noted in our simulation experiments, when it happens, this phenomenon takes place early in the MCMC iterations. So this calibration is not exceedingly time consuming. This value will obviously depend on A (the mass parameter of the Dirichlet process), but our proposed calibration procedure has proved to be robust over many simulated datasets.

Convergence assessment of the MCMC

To assess the convergence of the MCMC algorithm, we run 5 MCMC chains starting from 5 different points, simulated with an inflated version of the prior distribution, as follows:

$$\begin{array}{rcl} M_\lambda^{(0)} & \sim & 2\ \mathrm{Gamma}(a_M,\, b_M), \\ K^{\star(0)} & = & K_t, \\ (v_k)^{(0)}_{k=1,\dots,K^{\star(0)}} & \overset{\text{i.i.d}}{\sim} & \mathrm{Beta}(1,\, A), \\ (\theta_k)^{(0)}_{k=1,\,\dots,\,K^{\star(0)}} & \overset{\text{i.i.d}}{\sim} & \left[1 + 1/\mathrm{Gamma}(\alpha,\, \beta)\right]^{-1}, \\ (c_i)^{(0)}_{i\in\mathcal{O}} & \overset{\text{i.i.d}}{\sim} & \mathcal{U}_{\{1,\,\dots,\,K^{\star(0)}\}}. \end{array}$$

The convergence diagnostic tests are performed using the `coda` `R`-package: these tools are designed for a parametric estimation; we propose to adapt them to the nonparametric paradigm. For one of the datasets, 5 chains are run during 50000 iterations and a burn-in period of 20000 iterations is removed. The algorithm is implemented in `R`.

In Figure 6, we plot the values of M_λ over the iterations of the MCMC and the autocorrelation function. In these graphs, we do not detect any convergence issue. We adapt the Potential Scale Reduction Factor (PRSF) (Gelman and Rubin, 1992) diagnostic.

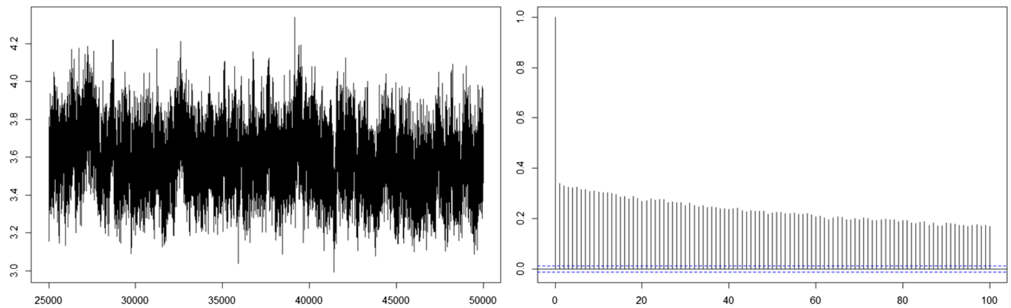


Figure 6: MCMC convergence assessment. On the left, trajectory of M_λ over 25000 iterations. On the right, autocorrelation function.

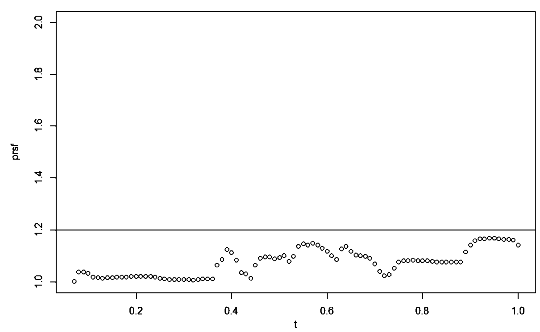


Figure 7: Convergence diagnostic. Potential Scale Reduction Factor of $\bar{\lambda}(t_j)$ for $t_j \in [0, 1]$.

For a fixed grid $(t_1, \ldots, t_J) \in [0, 1]$ (94 points regularly spaced between 0 and 1), we consider the 5 chains $\big((\bar{\lambda}(t_j))^{(\ell)}\big)_{\ell \geq 25000}$. The ratios between the within and between chain variances (Potential Scale Reduction Factor) are computed for each value of the grid t_j and plotted on Figure 7. The PRSF remains near 1.0, proving once again that no pathologic convergence can be found.

Results

With each simulated dataset, we concatenate the 5 chains to obtain a sample from the posterior distribution. For 4 of the datasets arbitrarily chosen, we plot 100 realizations of the posterior distribution of λ (Figure 8, left). Using the formula $S(t) = \exp(-\int_0^t \lambda(t)\mathrm{d}u)$, we also plot 100 posterior realizations of F and compare it with the true cumulative distribution function (Figure 8, right). The estimation of λ is of good quality over $[0, 0.7]$, the estimation is less accurate at the end of the interval, due to the increasing proportion of censored data. However, it corresponds to the tail of the distribution F and so this phenomenon is less noticeable on F.

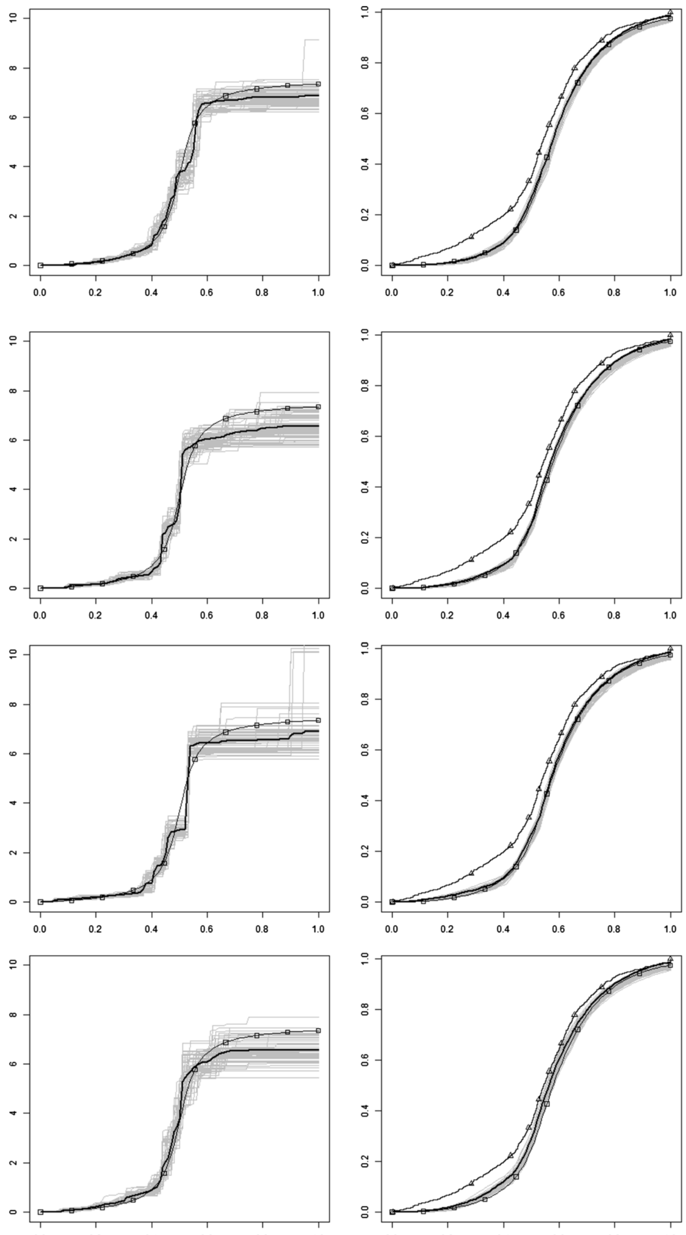


Figure 8: Posterior distributions. For 4 datasets, on the left 100 realizations (gray lines) of λ under the posterior distribution issued from the last iterations of the 5 MCMC chains: the posterior mean is plotted in plain line, the true λ is the line with squares. On the right, the corresponding curves for F: posterior simulation in gray, estimated in plain line, true F in line with squares; the empirical probability function of the Z_i is the line with triangles.

6 Proofs

In what follows, the symbols "$\lesssim$" and "$\gtrsim$" are used to denote inequalities valid up to constants that are universal or fixed throughout.

6.1 Proof of Theorem 1

Given Propositions 1 and 2, the proof is similar to that of Theorem 1 in Ghosal and van der Vaart (2007), which generalizes Theorem 2.4 of Ghosal et al. (2000). Write the posterior probability of the set $U_n = \{\lambda : \ \|\lambda - \lambda_0\|_1 > J_1 v_n\}$, given the observations, as

$$\pi(U_n \mid \mathbf{D}) = \frac{\int_{U_n} e^{\ell_n(\lambda)-\ell_n(\lambda_0)} \mathrm{d}\pi(\lambda)}{\int_{\mathcal{F}} e^{\ell_n(\lambda)-\ell_n(\lambda_0)} \mathrm{d}\pi(\lambda)} =: \frac{N_n}{D_n}.$$

We first show that, for the constant κ_0 introduced in Proposition 1, the probability of the event $A_n^c = (D_n \le e^{-(\kappa_0+1)nv_n^2}\pi_1(\bar{B}_n(\bar{\lambda}_0; v_n, H)))$ decays polynomially,

$$\mathbb{P}_{\lambda_0}^{(n)}(A_n^c) \lesssim (\log n)^{3k/2}(nv_n^2)^{-3k/2} + (\log n)^k (nv_n^2)^{-k} + (nv_n^2)^{-2k+1}(\log n)^{2k-1} =: p_n.$$

To the aim, we set

$$V_{2k}(\lambda_0; \lambda) = \mathbb{E}_{\lambda_0}^{(n)}[|\ell_n(\lambda_0) - \ell_n(\lambda) - \mathbb{E}_{\lambda_0}^{(n)}[\ell_n(\lambda_0) - \ell_n(\lambda)]|^{2k}], \quad k \ge 1.$$

Using Proposition 1, we have

$$B_n(\lambda_0; v_n, H) \subseteq \{\lambda : \mathrm{KL}(\lambda_0; \lambda) \le \kappa_0 n v_n^2 \text{ and } V_{2k}(\lambda_0; \lambda) \le \kappa p_n (nv_n^2)^{2k}\},$$

with

$$B_n(\lambda_0; v_n, H) = \{\lambda : \ \bar{\lambda} \in \bar{B}_n(\bar{\lambda}_0; v_n, H), \ \ |M_\lambda - M_{\lambda_0}| \le v_n\}.$$

By the assumption on the continuity and positivity of the Lebesgue density of the prior π_M and the requirement that $v_n^2 \ge (n/\log n)^{-1}$, we have

$$\pi(B_n(\lambda_0; v_n, H)) \gtrsim \pi_1(\bar{B}_n(\bar{\lambda}_0; v_n, H)) v_n \gtrsim \pi_1(\bar{B}_n(\bar{\lambda}_0; v_n, H)) e^{-nv_n^2/2}.$$

Thus, with $\mathrm{d}\bar{\pi}(\cdot) = \mathrm{d}\pi(\cdot) 1_{B_n(\lambda_0; v_n, H)}(\cdot)/\pi(B_n(\lambda_0; v_n, H))$, we get

$$\mathbb{P}_{\lambda_0}^{(n)}(A_n^c) \le \mathbb{P}_{\lambda_0}^{(n)}\left(\int_{B_n(\lambda_0; v_n, H)} e^{\ell_n(\lambda)-\ell_n(\lambda_0)} \mathrm{d}\bar{\pi}(\lambda) \lesssim e^{-(\kappa_0+1/2)nv_n^2}\right) \lesssim p_n, \tag{18}$$

by an application of Lemma 10 of Ghosal and van der Vaart (2007) to the probability on the right-hand side of (18).

Since $\mathbb{P}_{\lambda_0}^{(n)}(\Gamma_n^c) = o(1)$ and $\mathbb{P}_{\lambda_0}^{(n)}(A_n^c) \lesssim p_n$, to prove the assertion of the theorem, we can restrict attention to $\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n}\mathbf{1}_{A_n}\pi(\lambda : \ \|\lambda - \lambda_0\|_1 > J_1 v_n \mid \mathbf{D})]$, which can be decomposed into pieces mimicking the proof of Theorem 1 of Ghosal and van der Vaart

(2007). Thus, using tests $\phi_{n,j}$ of Proposition 2 and the fact that inequality (25) implies that $\pi(S_{n,j}(v_n)) \le \pi_1(\bar{S}_{n,j})$, we have, for $J_1 \ge J_0$,

$$
\begin{aligned}
&\mathbb{E}_{\lambda_0}^{(n)}\left[\mathbf{1}_{\Gamma_n}\mathbf{1}_{A_n}\pi\left(\lambda:\ \|\lambda-\lambda_0\|_1 > J_1 v_n \mid \mathbf{D}\right)\right] \\
&\le \sum_{j\ge J_1}\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n}\phi_{n,j}] + \sum_{j\ge J_1}\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n}\mathbf{1}_{A_n}(1-\phi_{n,j})\pi\left(S_{n,j}(v_n)\mid \mathbf{D}\right)] \\
&\qquad + \mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{A_n}\pi_1\left(\mathcal{F}_n^c \mid \mathbf{D}\right)] \\
&\lesssim \sum_{j\ge J_1}\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n}\phi_{n,j}] + \sum_{j=\lceil J_1\rceil}^{\lfloor \rho/v_n\rfloor} e^{(\kappa_0+1)nv_n^2}\frac{\pi_1(\bar{S}_{n,j})e^{-cnj^2v_n^2}}{\pi_1(\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H))} \\
&\qquad + \sum_{j>\rho/v_n}\frac{e^{(\kappa_0+1)nv_n^2}\pi_1(\bar{S}_{n,j})e^{-cnjv_n}}{\pi_1(\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H))} + \frac{e^{(\kappa_0+1)nv_n^2}\pi_1(\mathcal{F}_n^c)}{\pi_1(\bar{B}_n(\bar{\lambda}_0;\, v_n,\, H))}.
\end{aligned}
$$

The last expression converges to zero as $n \to \infty$ for fixed constants c, J_1, ρ. The conclusion follows. □

To prove Theorem 1, we have used the following intermediate results, whose proofs are postponed to Section 7. The first one controls the Kullback–Leibler divergence defined in (7) and the absolute moments of $\ell_n(\lambda_0) - \ell_n(\lambda)$.

Proposition 1. *Let v_n be a positive sequence such that $v_n \to 0$ and $nv_n^2 \to \infty$. For any $k \ge 1$ and $H > 0$, define the set*

$$B_n(\lambda_0;\, v_n,\, H) = \{\lambda:\ \bar{\lambda} \in \bar{B}_n(\bar{\lambda}_0;\, v_n,\, H),\ |M_\lambda - M_{\lambda_0}| \le v_n\}.$$

Under assumptions (5) and (8), for all $\lambda \in B_{k,n}(\lambda_0;\, v_n,\, H)$, we have

$$
\begin{aligned}
&\mathrm{KL}(\lambda_0;\, \lambda) \le \kappa_0 n v_n^2 \quad \textit{and} \\
&V_{2k}(\lambda_0;\, \lambda) \le \kappa[(nv_n^2\log n)^k + (nv_n^2(\log n)^3)^{k/2} + nv_n^2(\log n)^{2k-1}],
\end{aligned}
$$

where κ_0 and κ only depend on k, C_{1k}, H, λ_0, m_1 and m_2. An expression of κ_0 is given in (9).

The second result establishes the existence of tests that are used to control the numerator of posterior distributions. Recall that

$$\forall\, t \in \Omega, \quad (1-\alpha)\tilde{\mu}_n(t) \le \frac{Y_t}{n} \le (1+\alpha)\tilde{\mu}_n(t). \tag{19}$$

Proposition 2. *Assume that conditions (i) and (ii) of Theorem 1 are satisfied. For any $j \in \mathbb{N}$, define*

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

Then, under assumption (5), there are constants J_0, ρ, $c > 0$ such that, for every integer $j \ge J_0$, there exists a test $\phi_{n,j}$ so that, for a positive constant C,

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n}\phi_{n,j}] \le Ce^{-cnj^2v_n^2}, \quad \sup_{\lambda\in S_{n,j}(v_n)}\mathbb{E}_\lambda[\mathbf{1}_{\Gamma_n}(1-\phi_{n,j})] \le Ce^{-cnj^2v_n^2}, \qquad J_0 \le j \le \frac{\rho}{v_n},$$

and

$$\mathbb{E}^{(n)}_{\lambda_0}[\mathbf{1}_{\Gamma_n}\phi_{n,j}] \le Ce^{-cnjv_n}, \qquad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_\lambda[\mathbf{1}_{\Gamma_n}(1-\phi_{n,j})] \le Ce^{-cnjv_n}, \qquad j > \frac{\rho}{v_n}.$$

6.2 Proof of Corollary 1

Without loss of generality, we can assume that $\Omega = [0, T]$. At several places, using (1) and (19), we have that, under $\mathbb{P}^{(n)}_\lambda(\cdot \mid \Gamma_n)$, 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. Recall that $\bar\epsilon_n = (n/\log n)^{-1/3}$. For κ_0 as in (9), we control $\mathbb{P}^{(n)}_{\lambda_0}(\ell_n(\lambda) - \ell_n(\lambda_0) \le -(\kappa_0+2)n\bar\epsilon_n^2)$. We follow most of the computations of Salomond (2014). Let $e_n = (n\bar\epsilon_n^2)^{-k/2}$,

$$\bar\lambda_{0n}(t) = \frac{\lambda_0(t)\mathbf{1}_{t \ge \theta_n}}{\int_{\theta_n}^T \lambda_0(u)\mathrm{d}u}, \qquad \text{with} \;\; \theta_n = \sup\left\{\theta : \int_\theta^T \bar\lambda_0(t)\mathrm{d}t \ge 1 - \frac{e_n}{n}\right\},$$

and $\lambda_{0n} = M_{\lambda_0}\bar\lambda_{0n}$. Define the event $A_n = \{X \in N : X > \theta_n\}$. We make use of the following result. Let $\tilde N$ be a Poisson process with intensity $n(1+\alpha)m_2\lambda_0$. If $\tilde N_T = k$, denote by $\{T_1, \dots, T_k\}$ the jump times of $\tilde N$. Conditionally on $\tilde N_T = k$, the random variables $T_1, \dots, T_k$ are i.i.d. with density $\bar\lambda_0$. So,

$$\begin{aligned}
\mathbb{P}^{(n)}_{\lambda_0}(A_n^c \mid \Gamma_n) &\le \sum_{k=1}^\infty \mathbb{P}^{(n)}_{\lambda_0}(\exists\, T_i \le \theta_n \mid \tilde N_T = k)\,\mathbb{P}^{(n)}_{\lambda_0}(\tilde N_T = k)\\
&\le \sum_{k=1}^\infty \frac{ke_n}{n}\mathbb{P}^{(n)}_{\lambda_0}(\tilde N_T = k)\\
&= O\left(\frac{e_n}{n}\mathbb{E}^{(n)}_{\lambda_0}[\tilde N_T]\right) = O(e_n) = O((n\bar\epsilon_n^2)^{-k/2}).
\end{aligned}$$

Now,

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

We now deal with the first term on the right-hand side. On $\Gamma_n \cap A_n$,

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

So, for every λ and any n large enough,

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

because $\mathbb{P}_{\lambda_0}^{(n)}(\cdot \mid A_n)=\mathbb{P}_{\lambda_{0n}}^{(n)}(\cdot)$. Let $H>0$ be fixed. For all $\lambda\in B_n(\lambda_{0n};\,\bar\epsilon_n,\,H)$, using Proposition 1, we obtain

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

Mimicking the proof of Lemma 8 in Salomond (2014), we have that, for some constant $C_k>0$,

$$\pi_1\left(\bar B_n(\bar\lambda_{0n};\,\bar\epsilon_n,\,H)\right)\ge e^{-C_k n\bar\epsilon_n^2}\qquad\text{when } n \text{ is large enough,}$$

so that the first part of condition (ii) of Theorem 1 is verified. As in Salomond (2014), we set $\mathcal{F}_n=\{\bar\lambda:\ \bar\lambda(0)\le M_n\}$, with $M_n=\exp(c_1 n\bar\epsilon_n^2)$ and c_1 a positive constant. From Lemma 9 of Salomond (2014), there exists $a>0$ such that $\pi_1(\mathcal{F}_n^c)\le e^{-c_1(a+1)n\bar\epsilon_n^2}$ for n large enough, and the first part of condition (i) is satisfied. It is known from Groeneboom (1985) that the ϵ-entropy of $\mathcal{F}_n$ is of the order $(\log M_n)/\epsilon$, that is $o(n)$ for all $\epsilon>0$, and the second part of (i) holds. The second part of (ii) is a consequence of Salomond (2014). □

7 Proof of Propositions 1 and 2

This section reports the proofs of Propositions 1 and 2 that have been stated in Section 6. Proofs of intermediate results are deferred to Supplementary material.

We use the fact that for any pair of densities f and g, $\|f-g\|_1\le 2h(f,\,g)$.

7.1 Proof of Proposition 1

The proof of Proposition 1 relies on standard martingale properties of counting processes that can be found in Appendix B of Karr (1986). Recall that the log-likelihood evaluated at λ is given by $\ell_n(\lambda)=\int_0^T\log(\lambda(t))\mathrm{d}N_t-\int_0^T\lambda(t)Y_t\mathrm{d}t$. Since on $[0,\,T]\setminus\Omega$, N is empty and $Y_t\equiv 0$ almost surely, we can assume, without loss of generality, that $\Omega=[0,\,T]$. By using the definition of μ_n and $\tilde\mu_n$ given in (4), define

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

and the following density functions on $[0,\,T]$

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

Note that (5) gives

$$nm_1 M_{\lambda_0} \le M_n(\lambda_0) \le nm_2 M_{\lambda_0}, \quad nm_1 M_\lambda \le M_n(\lambda) \le nm_2 M_\lambda.$$

By using standard properties of counting processes, see Karr (1986), and straightforward computations,

$$\begin{aligned}
\mathrm{KL}(\lambda_0;\, \lambda) &= \mathbb{E}^{(n)}_{\lambda_0}[\ell_n(\lambda_0) - \ell_n(\lambda)] \\
&= \int_0^T \log\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\lambda_0(t)\mu_n(t)\mathrm{d}t - \int_0^T [\lambda_0(t) - \lambda(t)]\mu_n(t)\mathrm{d}t \\
&= M_n(\lambda_0)\left[\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) + \frac{M_n(\lambda)}{M_n(\lambda_0)} - 1 - \log\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)\right] \\
&= M_n(\lambda_0)\left[\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) + \phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)\right] \\
&\le nm_2 M_{\lambda_0}\left[\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) + \phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right)\right], \qquad (20)
\end{aligned}$$

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

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

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

$$\begin{aligned}
\mathrm{KL}(\bar\lambda_{0,n};\, \bar\lambda_n) &\le 2h^2(\bar\lambda_{0,n},\, \bar\lambda_n)\left(1 + \log\left\|\frac{\bar\lambda_{0,n}}{\bar\lambda_n}\right\|_\infty\right) \\
&\le 2h^2(\bar\lambda_{0,n},\, \bar\lambda_n)\left[1 + \log\left(\frac{m_2}{m_1}\right) + \log\left\|\frac{\bar\lambda_0}{\bar\lambda}\right\|_\infty\right] \\
&\le 2\left[1 + \log\left(\frac{m_2}{m_1}\right)\right] h^2(\bar\lambda_{0,n},\, \bar\lambda_n)\left(1 + \log\left\|\frac{\bar\lambda_0}{\bar\lambda}\right\|_\infty\right) \qquad (21)
\end{aligned}$$

since $1 + \log(m_2/m_1) \ge 1$. We now deal with $h^2(\bar\lambda_{0,n},\, \bar\lambda_n)$. By still using (5), we have

$$\begin{aligned}
h^2(\bar\lambda_{0,n},\, \bar\lambda_n) &= \int_0^T \left(\sqrt{\bar\lambda_{0,n}(t)} - \sqrt{\bar\lambda_n(t)}\right)^2 \mathrm{d}t \\
&= \int_0^T \left(\sqrt{\frac{\bar\lambda_0(t)\tilde\mu_n(t)}{\int_0^T \bar\lambda_0(u)\tilde\mu_n(u)\mathrm{d}u}} - \sqrt{\frac{\bar\lambda(t)\tilde\mu_n(t)}{\int_0^T \bar\lambda(u)\tilde\mu_n(u)\mathrm{d}u}}\right)^2 \mathrm{d}t \\
&\le 2m_2 \int_0^T \left(\sqrt{\frac{\bar\lambda_0(t)}{\int_0^T \bar\lambda_0(u)\tilde\mu_n(u)\mathrm{d}u}} - \sqrt{\frac{\bar\lambda_0(t)}{\int_0^T \bar\lambda(u)\tilde\mu_n(u)\mathrm{d}u}}\right)^2 \mathrm{d}t \\
&\qquad + 2m_2 \int_0^T \left(\sqrt{\frac{\bar\lambda_0(t)}{\int_0^T \bar\lambda(u)\tilde\mu_n(u)\mathrm{d}u}} - \sqrt{\frac{\bar\lambda(t)}{\int_0^T \bar\lambda(u)\tilde\mu_n(u)\mathrm{d}u}}\right)^2 \mathrm{d}t \\
&\le 2m_2 U_n + \frac{2m_2}{m_1} h^2(\bar\lambda_0,\, \bar\lambda),
\end{aligned}$$

with

$$U_n = \left(\sqrt{\frac{1}{\int_0^T \bar{\lambda}_0(t)\tilde{\mu}_n(t)\mathrm{d}t}} - \sqrt{\frac{1}{\int_0^T \bar{\lambda}(t)\tilde{\mu}_n(t)\mathrm{d}t}} \right)^2.$$

We denote by

$$\tilde{\epsilon}_n := \frac{1}{\int_0^T \bar{\lambda}_0(u)\tilde{\mu}_n(u)\mathrm{d}u} \int_0^T [\bar{\lambda}(t) - \bar{\lambda}_0(t)]\tilde{\mu}_n(t)\mathrm{d}t,$$

so that

$$|\tilde{\epsilon}_n| \le \frac{1}{m_1} \int_0^T |\bar{\lambda}(t) - \bar{\lambda}_0(t)|\tilde{\mu}_n(t)\mathrm{d}t \le \frac{2m_2}{m_1} h(\bar{\lambda}_0,\, \bar{\lambda}).$$

Then,

$$U_n = \frac{1}{\int_0^T \bar{\lambda}_0(t)\tilde{\mu}_n(t)\mathrm{d}t} \left(1 - \frac{1}{\sqrt{1+\tilde{\epsilon}_n}}\right)^2 \le \frac{\tilde{\epsilon}_n^2}{4m_1} \le \frac{m_2^2}{m_1^3} h^2(\bar{\lambda}_0,\, \bar{\lambda}).$$

Finally,

$$h^2(\bar{\lambda}_{0,n},\, \bar{\lambda}_n) \le \frac{2m_2}{m_1}\left(\frac{m_2^2}{m_1^2} + 1\right) h^2(\bar{\lambda}_0,\, \bar{\lambda}). \tag{22}$$

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

$$\begin{aligned}
|M_n(\lambda_0) - M_n(\lambda)| &\le \int_0^T |\lambda(t) - \lambda_0(t)|\mu_n(t)\mathrm{d}t \\
&\le nm_2 \int_0^T |\lambda(t) - \lambda_0(t)|\mathrm{d}t \\
&\le nm_2 \left[M_{\lambda_0}\|\bar{\lambda} - \bar{\lambda}_0\|_1 + |M_\lambda - M_{\lambda_0}|\right] \\
&\le \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0) \left[M_{\lambda_0}\|\bar{\lambda} - \bar{\lambda}_0\|_1 + |M_\lambda - M_{\lambda_0}|\right] \\
&\le \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0)[2M_{\lambda_0} h(\bar{\lambda},\, \bar{\lambda}_0) + |M_\lambda - M_{\lambda_0}|] \\
&\le \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0)(2M_{\lambda_0} + 1)v_n,
\end{aligned}$$

since $\lambda \in B_n(\lambda_0;\, v_n,\, H)$. Finally, since $\phi(u+1) \le u^2$ if $|u| \le 1/2$, the previous inequality gives

$$\phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right) \le \frac{m_2^2}{m_1^2 M_{\lambda_0}^2}(2M_{\lambda_0} + 1)^2 v_n^2 \quad \text{for } n \text{ large enough.} \tag{23}$$

Combining (20), (21), (22) and (23), we have $\mathrm{KL}(\lambda_0;\, \lambda) \le \kappa_0 n v_n^2$ for n large enough, with κ_0 as in (9). We now deal with $V_{2k}(\lambda_0;\, \lambda)$ for $k \ge 1$. In the sequel, we denote by C a constant that may change from line to line. For any j, let

$$E_j(\bar{\lambda}_0;\, \bar{\lambda}) = \int_0^T \bar{\lambda}_0(x)[\log \bar{\lambda}_0(x) - \log \bar{\lambda}(x)]^{2j}\mathrm{d}x.$$

Theorem 5 of Wong and Shen (1995) leads to $E_j(\bar{\lambda}_0;\,\bar{\lambda}) \le C(1+(\log n)^{2j-1})v_n^2$. Straightforward computations lead to

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

with

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

and

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

By (8),

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

where

$$A_{2k,1} = \left[\int_0^T \lambda_0^2(t)\log^2\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\mathrm{d}t\right]^k \quad \text{and} \quad A_{2k,2} = \left(\int_0^T [\lambda_0(t) - \lambda(t)]^2\mathrm{d}t\right)^k.$$

For $\lambda \in B_n(\lambda_0;\,v_n,\,H)$,

$$\begin{aligned} A_{2k,1} &\le M_{\lambda_0}^{2k}\|\bar{\lambda}_0\|_\infty^k\left[\int_0^T \bar{\lambda}_0(t)\log^2\left(\frac{M_{\lambda_0}\times\bar{\lambda}_0(t)}{M_\lambda\times\bar{\lambda}(t)}\right)\mathrm{d}t\right]^k \\ &\le 2^{2k-1}M_{\lambda_0}^{2k}\|\bar{\lambda}_0\|_\infty^k\left[E_1^k(\bar{\lambda}_0;\,\bar{\lambda}) + \left|\log\left(\frac{M_\lambda}{M_{\lambda_0}}\right)\right|^{2k}\right] \\ &\le C\left[E_1^k(\bar{\lambda}_0;\,\bar{\lambda}) + |M_\lambda - M_{\lambda_0}|^{2k}\right] \le Cv_n^{2k}(\log n)^k \end{aligned}$$

and

$$\begin{aligned}
A_{2k,2} &:= \left(\int_0^T [\lambda_0(t) - \lambda(t)]^2 \mathrm{d}t \right)^k \\
&= \left(\int_0^T \left\{ (M_{\lambda_0} - M_\lambda) \bar{\lambda}_0(t) - M_\lambda [\bar{\lambda}(t) - \bar{\lambda}_0(t)] \right\}^2 \mathrm{d}t \right)^k \\
&\leq 2^{2k-1} \|\bar{\lambda}_0\|_\infty^{2k} (M_{\lambda_0} - M_\lambda)^{2k} \\
&\qquad + 2^{2k-1} M_\lambda^{2k} \left[\int_0^T \left(\sqrt{\bar{\lambda}_0(t)} - \sqrt{\bar{\lambda}(t)} \right)^2 \left(\sqrt{\bar{\lambda}_0(t)} + \sqrt{\bar{\lambda}(t)} \right)^2 \mathrm{d}t \right]^k \\
&\leq 2^{2k-1} \|\bar{\lambda}_0\|_\infty^{2k} (M_{\lambda_0} - M_\lambda)^{2k} + 2^{3k-1} M_\lambda^{2k} (\|\bar{\lambda}_0\|_\infty + \|\bar{\lambda}\|_\infty)^k h^{2k}(\bar{\lambda}_0, \bar{\lambda}) \leq C v_n^{2k}.
\end{aligned}$$

Therefore,

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

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

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

so $(M_T)_T$ is a martingale and we use standard properties of continuous time martingales associated with counting processes, see Appendix B of Karr (1986). Assume that $k > 1$. Using Rosenthal's inequality for point process martingales, see Wood (1999), there exists a constant $C(k)$ only depending on k such that

$$\begin{aligned}
B_{2k} &= \mathbb{E}_{\lambda_0}^{(n)}[|M_T|^{2k}] \\
&\leq C(k) \left[\mathbb{E}_{\lambda_0}^{(n)} \left| \int_0^T \log^2\left(\frac{\lambda_0(t)}{\lambda(t)} \right) \lambda_0(t) Y_t \mathrm{d}t \right|^k + \int_0^T \log^{2k}\left(\frac{\lambda_0(t)}{\lambda(t)} \right) \lambda_0(t) \mu_n(t) \mathrm{d}t \right] \\
&\leq C \left(B_{k,2}^{(1)} + B_{k,2}^{(2)} + n m_2 M_{\lambda_0} \left([\log(M_{\lambda_0}/M_\lambda)]^{2k} + E_k(\bar{\lambda}_0; \bar{\lambda}) \right) \right),
\end{aligned}$$

with

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

Note that

$$B_{k,2}^{(1)} \leq \left[\int_0^T \log^4\left(\frac{\lambda_0(t)}{\lambda(t)}\right)\lambda_0^2(t)\mathrm{d}t\right]^{k/2} \times \mathbb{E}_{\lambda_0}^{(n)}\left[\left(\int_0^T [Y_t-\mu_n(t)]^2\mathrm{d}t\right)^{k/2}\right]$$
$$\leq (M_{\lambda_0}^2\|\bar\lambda_0\|_\infty)^{k/2}\left[\int_0^T \log^4\left(\frac{M_{\lambda_0}\times\bar\lambda_0(t)}{M_\lambda\times\bar\lambda(t)}\right)\bar\lambda_0(t)\mathrm{d}t\right]^{k/2}\times\sqrt{C_{1k}n^k}$$
$$\leq C\left[\log^4\left(\frac{M_{\lambda_0}}{M_\lambda}\right)+E_2(\bar\lambda_0;\,\bar\lambda)\right]^{k/2}\times n^{k/2},$$

where we have used (8) and the Jensen's inequality. Similarly,

$$B_{k,2}^{(2)} \leq (nm_2M_{\lambda_0})^k\left[\int_0^T \log^2\left(\frac{M_{\lambda_0}\times\bar\lambda_0(t)}{M_\lambda\times\bar\lambda(t)}\right)\bar\lambda_0(t)\mathrm{d}t\right]^k$$
$$\leq C\left[\log^2\left(\frac{M_{\lambda_0}}{M_\lambda}\right)+E_1(\bar\lambda_0;\bar\lambda)\right]^k\times n^k.$$

Therefore,

$$V_{2k}(\lambda_0;\,\lambda)\leq\kappa[(nv_n^2\log n)^k+(nv_n^2(\log n)^3)^{k/2}+nv_n^2(\log n)^{2k-1}],$$

where κ depends on C_{1k}, k, H, λ_0, m_1 and m_2. Using previous computations, the case $k=1$ is straightforward. □

7.2 Proof of Proposition 2

We consider the setting of Lemma 2, given in Section A.1 in Supplementary material and a covering of $S_{n,j}(v_n)$ with $\mathbb{L}_1$-balls of radius ξjv_n and centres $(\lambda_{l,j})_{l=1,\dots,D_j}$, where D_j is the covering number of $S_{n,j}(v_n)$ by such balls. We 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 2. So, there exists a constant $\rho>0$ such that

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

and

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n}\phi_{n,j}]\leq 2D_je^{-Knjv_n}\text{ and }\sup_{\lambda\in S_{n,j}(v_n)}\mathbb{E}_\lambda^{(n)}[\mathbf{1}_{\Gamma_n}(1-\phi_{n,j})]\leq 2e^{-Knjv_n},\quad\text{if } j>\frac{\rho}{v_n},$$

where K is a constant (see Lemma 2). We now bound D_j. First note that for any $\lambda=M_\lambda\times\bar\lambda$ and $\lambda'=M_{\lambda'}\times\bar\lambda'$,

$$\|\lambda-\lambda'\|_1\leq M_\lambda\|\bar\lambda-\bar\lambda'\|_1+|M_\lambda-M_{\lambda'}|. \tag{24}$$

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

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

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

$$\begin{aligned}\|\lambda - \lambda_0\|_1 &\geq \int_{\bar\lambda_0 > \bar\lambda} [M_{\lambda_0} \times \bar\lambda_0(t) - M_\lambda \times \bar\lambda(t)]\mathrm{d}t \\ &\geq M_{\lambda_0} \int_{\bar\lambda_0 > \bar\lambda} [\bar\lambda_0(t) - \bar\lambda(t)]\mathrm{d}t = \frac{M_{\lambda_0}}{2}\|\bar\lambda - \bar\lambda_0\|_1.\end{aligned}$$

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

$$\|\lambda - \lambda_0\|_1 \geq \max\left\{(M_\lambda \vee M_{\lambda_0})\|\bar\lambda - \bar\lambda_0\|_1/2,\ |M_\lambda - M_{\lambda_0}|\right\}. \tag{25}$$

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

$$\|\bar\lambda - \bar\lambda_0\|_1 \leq \frac{2(j+1)v_n}{M_{\lambda_0}} \quad \text{and} \quad |M_\lambda - M_{\lambda_0}| \leq (j+1)v_n. \tag{26}$$

Therefore, $S_{n,j}(v_n) \subseteq (\bar S_{n,j} \cap \mathcal{F}_n) \times \{M : \ |M - M_{\lambda_0}| \leq (j+1)v_n\}$ 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, (26) 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 (24),
$$\|\lambda - \lambda'\|_1 \leq \frac{(M_\lambda + 1)\xi j v_n}{3M_{\lambda_0} + 1} \leq \xi j v_n.$$
By assumption (ii) of Theorem 1, this implies that, for any $\delta > 0$, there exists J_0 such that for $j \geq J_0$,
$$\begin{aligned}D_j &\leq D((3M_{\lambda_0} + 1)^{-1}\xi j v_n,\ \bar S_{n,j} \cap \mathcal{F}_n,\ \|\cdot\|_1) \times \left[2(j+1)v_n \times \frac{(3M_{\lambda_0} + 1)}{\xi j v_n} + \frac{1}{2}\right] \\ &\lesssim \exp(\delta n[(j+1)^2 v_n^2 \wedge 1]).\end{aligned}$$
- We assume that $(j+1)v_n > 2M_{\lambda_0}$. If
$$\|\bar\lambda - \bar\lambda'\|_1 \leq \frac{\xi}{4} \quad \text{and} \quad |M_\lambda - M_{\lambda'}| \leq \frac{\xi(M_\lambda \vee M_{\lambda_0})}{4},$$

using again (24) and (26),

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

for n large enough. By assumption (i) of Theorem 1, this implies that, for any $\delta > 0$,

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

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

Supplementary Material

Supplementary material of "Posterior Concentration Rates for Counting Processes with Aalen Multiplicative Intensities" (DOI: 10.1214/15-BA986SUPP; .pdf).

References

Aalen, O., 1978. "Nonparametric inference for a family of counting processes." *The Annals of Statistics*, 6(4): 701–726. MR0491547. 53, 55

Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N., 1993. *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag, New York. MR1198884. doi: `http://dx.doi.org/10.1007/978-1-4612-4348-9`. 53, 55, 56, 57, 58

Belitser, E., Serra, P., van Zanten, H., 2015. "Rate-optimal Bayesian intensity smoothing for inhomogeneous Poisson processes." *Journal of Statistical Planning and Inference*, 166: 24–35. MR3390131. doi: `http://dx.doi.org/10.1016/j.jspi.2014.03.009`. 54, 56, 62

Brunel, E., Comte, F., 2005. "Penalized contrast estimation of density and hazard rate with censored data." *Sankhyā*, 67(3): 441–475. MR2235573. 53

Brunel, E., Comte, F., 2008. "Adaptive estimation of hazard rate with censored data." *Communications in Statistics – Theory and Methods*, 37(8): 1284–1305. MR2440441. doi: `http://dx.doi.org/10.1080/03610920701713302`. 53

Comte, F., Gaïffas, S., Guilloux, A., 2011. "Adaptive estimation of the conditional intensity of marker-dependent counting processes." *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 47(4): 1171–1196. MR2884230. doi: `http://dx.doi.org/10.1214/10-AIHP386`. 53, 55, 56, 58, 59

Daley, D. J., Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd Edition. Probability and Its Applications (New York). Springer-Verlag, New York. MR1950431. 53, 56

Daley, D. J., Vere-Jones, D., 2008. *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd Edition. Probability and Its Applications (New York). Springer-Verlag, New York. MR2371524. doi: http://dx.doi.org/10.1007/978-0-387-49835-5. 53

De Blasi, P., Peccati, G., Prünster, I., 2009. "Asymptotics for posterior hazards." *The Annals of Statistics*, 37(4): 1906–1945. MR2533475. doi: http://dx.doi.org/10.1214/08-AOS631. 54

de Boor, C., 1978. *A Practical Guide to Splines*. Springer, New York. MR0507062. 65

Donnet, S., Rivoirard, V., Rousseau, J., Scricciolo, C., 2014. "Posterior concentration rates for empirical Bayes procedures, with applications to Dirichlet process mixtures." Technical Report, arXiv:1406.4406. 69

Donnet, S., Rivoirard, V., Rousseau, J., Scricciolo, C., 2015. "Supplementary material of "Posterior Concentration Rates for Counting Processes with Aalen Multiplicative Intensities"." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/15-BA986SUPP. 55

Dykstra, R. L., Laud, P., 1981. "A Bayesian nonparametric approach to reliability." *The Annals of Statistics*, 9(2): 356–367. MR0606619. 53

Gaïffas, S., Guilloux, A., 2012. "High-dimensional additive hazards models and the Lasso." *Electronic Journal of Statistics*, 6: 522–546. MR2988418. doi: http://dx.doi.org/10.1214/12-EJS681. 53

Gelman, A., Rubin, D., 1992. "Inference from iterative simulation using multiple sequences (with discussion)." *Statistical Science*, 7: 457–511. MR1196276. 71

Ghosal, S., Ghosh, J. K., van der Vaart, A. W., 2000. "Convergence rates of posterior distributions." *The Annals of Statistics*, 28(2): 500–531. MR1790007. doi: http://dx.doi.org/10.1214/aos/1016218228. 53, 55, 60, 62, 63, 65, 74, 78

Ghosal, S., van der Vaart, A., 2007. "Convergence rates of posterior distributions for noniid observations." *The Annals of Statistics*, 35(1): 192–223. MR2332274. doi: http://dx.doi.org/10.1214/009053606000001172. 59, 62, 63, 74

Groeneboom, P., 1985. "Estimating a monotone density." In: *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Volume II*, Lucien M. Le Cam and Richard A. Olshen, eds., 539–555, Wadsworth and Brooks/Cole Statistics/Probability Series, Belmont, CA Wadsworth. MR0822052. 77

Gugushvili, S., Spreij, P., 2013. "A note on non-parametric Bayesian estimation for Poisson point processes." Technical Report, arXiv:1304.7353v2. 54, 62

Gutiérrez-Peña, E., Nieto-Barajas, L. E., 2003. Bayesian nonparametric inference for mixed Poisson processes (with discussion). In: *Bayesian Statistics, 7* (Tenerife, 2002), 163–179. Oxford: University Press, New York. MR2003172. 54

Hansen, N. R., Reynaud-Bouret, P., Rivoirard, V., 2015. "Lasso and probabilistic inequalities for multivariate point processes." *Bernoulli*, 21(1): 83–143. MR3322314. doi: http://dx.doi.org/10.3150/13-BEJ562. 53, 55, 59, 63

Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A., 1998. *Wavelets, Approximation, and Statistical Applications*. Vol. 129 of Lecture Notes in Statistics. Springer-Verlag, New York. MR1618204. doi: http://dx.doi.org/10.1007/978-1-4612-2222-4. 63

Ishwaran, H., James, L. F., 2004. "Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data." *Journal of the American Statistical Association*, 99: 175–190. MR2054297. doi: http://dx.doi.org/10.1198/016214504000000179. 54

Karr, A. F., 1986. *Point Processes and Their Statistical Inference*. Vol. 2 of Probability: Pure and Applied. Marcel Dekker, Inc., New York. MR0851982. 55, 56, 77, 78, 81

Karr, A. F., 1991. *Point Processes and Their Statistical Inference*, 2nd Edition. Vol. 7 of Probability: Pure and Applied. Marcel Dekker Inc., New York. MR1113698. 53

Kim, Y., 1999. "Nonparametric Bayesian estimators for counting processes." *The Annals of Statistics*, 27(2): 562–588. MR1714717. doi: http://dx.doi.org/10.1214/aos/1018031207. 54, 55, 58, 59

Kingman, J., 1993. *Poisson Processes*. Oxford Studies in Probability. MR1207584. 56

Kirichenko, A., van Zanten, H., 2015. "Optimality of Poisson processes intensity learning with Gaussian processes." Technical Report, arXiv:1409.5103. 54, 62

Klein, J. P., Moeschberger, M. L., 2003. *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, New York. 56, 57

Kottas, A., Sansó, B., 2007. "Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis." *Journal of Statistical Planning and Inference*, 137(10): 3151–3163. MR2365118. doi: http://dx.doi.org/10.1016/j.jspi.2006.05.022. 54

Kuo, L., Ghosh, S. K., 1997. "Bayesian nonparametric inference for nonhomogeneous Poisson processes." Technical Report, no. 97-18, University of Connecticut, Department of Statistics. 54

Lijoi, A., Nipoti, B., 2014. "A class of hazard rate mixtures for combining survival data from different experiments." *Journal of the American Statistical Association*, 109(506): 802–814. MR3223751. doi: http://dx.doi.org/10.1080/01621459.2013.869499. 54

Lo, A., Weng, C.-S., 1989. "On a class of Bayesian nonparametric estimates: II. Hazard rate estimates." *Annals of the Institute of Statistical Mathematics*, 41(2): 227–245. MR1006487. doi: http://dx.doi.org/10.1007/BF00049393. 54

Lo, A. Y., 1982. "Bayesian nonparametric statistical inference for Poisson point processes." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 59: 55–66. MR0643788. doi: http://dx.doi.org/10.1007/BF00575525. 54

Lo, A. Y., 1992. "Bayesian inference for Poisson process models with censored data." *Journal of Nonparametric Statistics*, 2(1): 71–80. MR1256374. doi: http://dx.doi.org/10.1080/10485259208832544. 54, 58

Reynaud-Bouret, P., 2006. "Penalized projection estimators of the Aalen multiplicative intensity." *Bernoulli*, 12(4): 633–661. MR2248231. doi: http://dx.doi.org/10.3150/bj/1155735930. 53, 55, 58, 59

Reynaud-Bouret, P., Rivoirard, V., 2010. "Near optimal thresholding estimation of a Poisson intensity on the real line." *Electronic Journal of Statistics*, 4: 172–238. MR2645482. doi: http://dx.doi.org/10.1214/08-EJS319. 54, 56

Rivoirard, V., Rousseau, J., 2012. "Posterior concentration rates for infinite dimensional exponential families." *Bayesian Analysis*, 7(2): 311–334. MR2934953. doi: http://dx.doi.org/10.1214/12-BA710. 65, 66

Salomond, J.-B., 2014. "Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints." *Electronic Journal of Statistics*, 8(1): 1380–1404. MR3263126. doi: http://dx.doi.org/10.1214/14-EJS929. 64, 76, 77

Walker, S. G., 2007. "Sampling the Dirichlet mixture model with slices." *Communications in Statistics – Simulation and Computation*, 36: 45–54. MR2370888. doi: http://dx.doi.org/10.1080/03610910601096262. 67

Williamson, R. E., 1956. "Multiply monotone functions and their Laplace transforms." *Duke Mathematical Journal*, 23(2): 189–207. MR0077581. 64

Wong, W. H., Shen, X., 1995. "Probability inequalities for likelihood ratios and convergence rates of sieves MLEs." *The Annals of Statistics*, 23(2): 339–362. MR1332570. doi: http://dx.doi.org/10.1214/aos/1176324524. 80

Wood, A. T. A., 1999. "Rosenthal's inequality for point process martingales." *Stochastic Processes and Applications*, 81(2): 231–246. MR1694561. doi: http://dx.doi.org/10.1016/S0304-4149(99)00005-8. 81

Acknowledgments

This research benefited from the support of the "Chaire Economie et Gestion des Nouvelles Données", under the auspices of the Institut Louis Bachelier, Havas-Media and Paris-Dauphine. The research of Sophie Donnet, Vincent Rivoirard and Judith Rousseau was partly supported by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 project Calibration). The research of Catia Scricciolo has been partially supported by Bocconi University. The authors wish to thank the Editor, the Associate Editor and anonymous Referees for providing suggestions that helped improving the final presentation of the article.