Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

SUPPLEMENTARY MATERIAL OF NONPARAMETRIC BAYESIAN ESTIMATION FOR MULTIVARIATE HAWKES PROCESSES

By Sophie Donnet*, Vincent Rivoirard† and Judith Rousseau†,‡

INRA, Université Paris Saclay, Université Paris-Dauphine† and University of Oxford‡*

In this supplement, we provide in Section 1 additional numerical results. Section 2 is devoted to the proofs of Theorems 1 and 2 (Sections 2.1 and 2.2). We also prove Corollary 1 in Section 2.3 and Lemma 1 in Section 2.4. In Section 2.5, we state and prove technical lemmas which are necessary for the main results of the paper. Finally, Section 2.6 is devoted to the proofs of results of Section 2.3 of [1].

1. Supplementary material for the simulation Section. In this section, we present two additional elements. First, in Table 1 we present a detailed version of the number of observations resulting from the chosen simulation parameters for the three scenarios.

Neuron	k	1	2	3	4	5	6	7	8
$K=2$	$T=5$	415.96	227.88						
	$T=10$	840.24	457.92						
	$T=20$	1667.16	903.44						
$K=8$	$T=10$	801.72	399.36	799.28	194.16	390.68	397.76	396.24	399.20
	$T=20$	1601.68	805.68	1599.20	392.36	788.20	796.64	794.40	802.32
K=2 with smooth $h_{k,\ell}$	$T=5$	428.24	206.60						
	$T=10$	841.64	412.80						
	$T=20$	1731.48	838.40						

Table 1

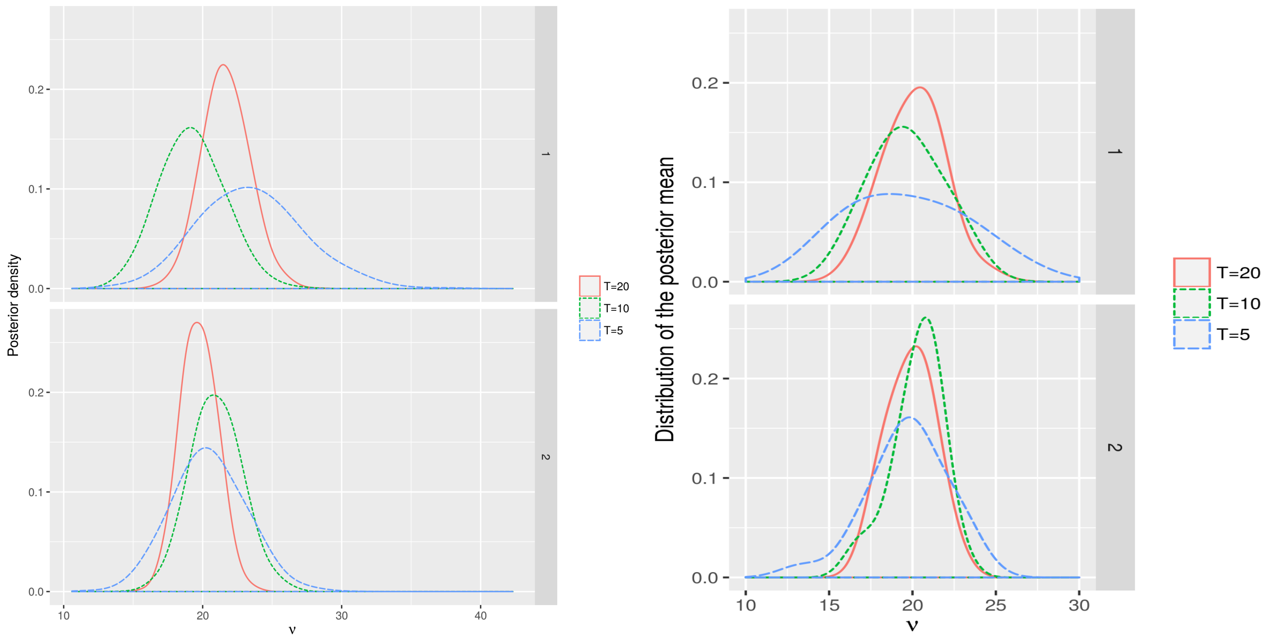
Mean numbers of events on each neuron for the three simulation scenarios (the average is done over all the simulated datasets)

In Figure 1, we then present the posterior density of ν_1 and ν_2 for one given dataset and the regularized distribution of the posterior mean of ν_1 and ν_2 over the 25 simulated datasets for the regular histogram prior (upper panel) and the random histogram prior (bottom panel). We thus illustrate the fact that both priors provide very similar results.

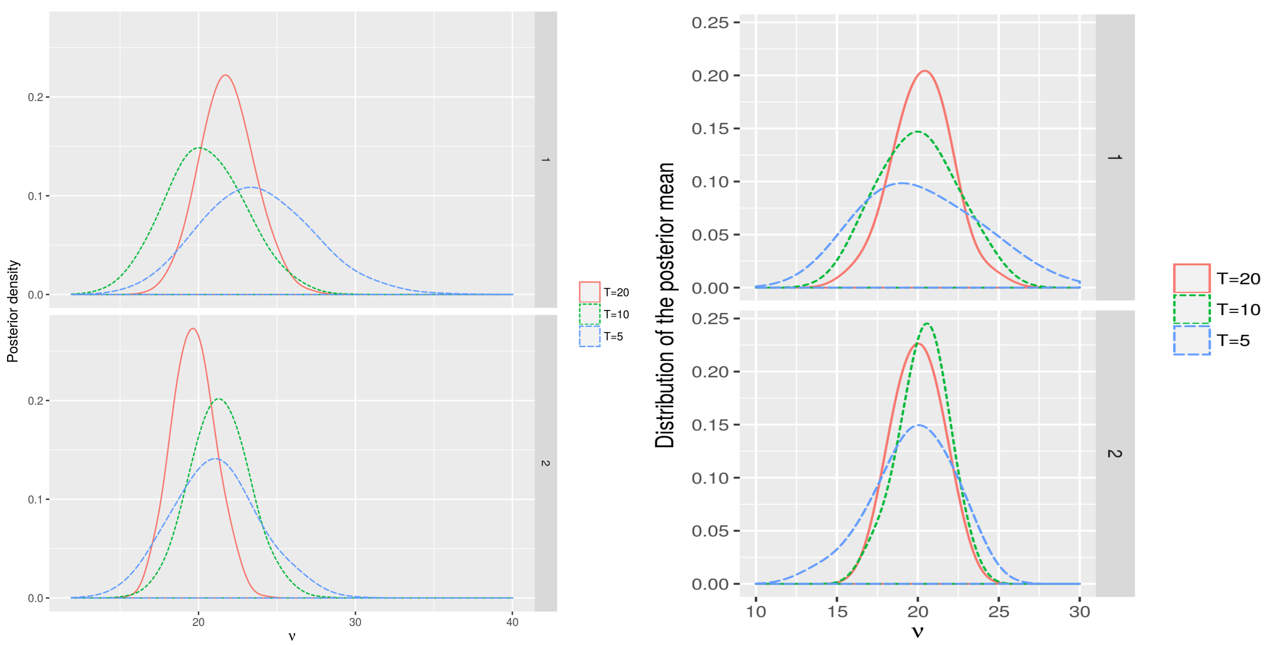
2. Supplementary material for the proof section.

2.1. *Proof of Theorem 1.* To prove Theorem 1, we apply the general methodology of Ghosal and van der Vaart [2], with modifications due to the fact that $\exp(L_T(f))$ is the likelihood of the distribution of $(N^k)_{k=1,\ldots,K}$ on $[0,T]$ conditional on $\mathcal{G}_0$ and that the metric $d_{1,T}$ depends on the observations. We set $M_T = M\sqrt{\log\log T}$, for M a positive constant. Let

$$A_\epsilon = \{f \in \mathcal{F};\ d_{1,T}(f_0, f) \le K\epsilon\}$$



(a) *Results for the regular histogram prior distribution* : *On the left*, posterior distribution of ν_1 (top) and ν_2 (bottom) with $T=5$, $T=10$ and $T=20$ for one dataset. *On the right*, regularized distribution of the posterior mean of (ν_1,ν_2) $\left(\widehat{\mathbb{E}}\left[\nu_\ell|(N_t^{sim})_{t\in[0,T]}\right]\right)_{sim=1\dots25}$ over the 25 simulated datasets.



(b) *Results for the random histogram prior distribution* : *On the left*, posterior distribution of ν_1 (top) and ν_2 (bottom) with $T=5$, $T=10$ and $T=20$ for one dataset. *On the right*, regularized distribution of the posterior mean of (ν_1,ν_2) $\left(\widehat{\mathbb{E}}\left[\nu_\ell|(N_t^{sim})_{t\in[0,T]}\right]\right)_{sim=1\dots25}$ over the 25 simulated datasets.

Fig 1: Results of *scenario 1* : influence of the prior distribution (random or regular histogram) on the inference of $(\nu_k)_{k=1,2}$

and for $j \geq 1$, we set

$$S_j = \{f \in \mathcal{F}_T;\ d_{1,T}(f, f_0) \in (Kj\epsilon_T, K(j+1)\epsilon_T]\}, \tag{2.1}$$

where $\mathcal{F}_T = \{f = ((\nu_k)_k, (h_{k,\ell})_{k,\ell}) \in \mathcal{F};\ (h_{k,\ell})_{k,\ell} \in \mathcal{H}_T\}$. So that, for any test function ϕ,

$$\begin{aligned}
\Pi\left(A^c_{M_T\epsilon_T}|N\right) &= \frac{\int_{A^c_{M_T\epsilon_T}} e^{L_T(f)-L_T(f_0)} d\Pi(f)}{\int_{\mathcal{F}} e^{L_T(f)-L_T(f_0)} d\Pi(f)} =: \frac{\bar{N}_T}{D_T} \\
&\leq \mathbb{1}_{\Omega^c_T} + \mathbb{1}_{\left\{D_T < \frac{\Pi(B(\epsilon_T,T))}{\exp(2(\kappa_T+1)T\epsilon_T^2)}\right\}} + \phi \mathbb{1}_{\Omega_T} + \frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T,T))} \int_{\mathcal{F}^c_T} e^{L_T(f)-L_T(f_0)} d\Pi(f) \\
&+ \mathbb{1}_{\Omega_T} \frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T,T))} \sum_{j=M_T}^{\infty} \int_{\mathcal{F}_T} \mathbb{1}_{f \in S_j} e^{L_T(f)-L_T(f_0)} (1-\phi) d\Pi(f)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_0\left[\Pi\left(A^c_{M_T\epsilon_T}|N\right)\right] &\leq \mathbb{P}_0(\Omega^c_T) + \mathbb{P}_0\left(D_T < e^{-2(\kappa_T+1)T\epsilon_T^2}\Pi(B(\epsilon_T,B))\right) + \mathbb{E}_0[\phi\mathbb{1}_{\Omega_T}] \\
&\quad + \frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T,B))} \left(\Pi(\mathcal{F}^c_T) + \sum_{j=M_T}^{\infty} \int_{\mathcal{F}_T} \mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi)|\mathcal{G}_0\right]\right] d\Pi(f)\right),
\end{aligned}$$

since

$$\mathbb{E}_0\left[\int_{\mathcal{F}^c_T} e^{L_T(f)-L_T(f_0)} d\Pi(f)\right] = \mathbb{E}_0\left[\mathbb{E}_0\left[\int_{\mathcal{F}^c_T} e^{L_T(f)-L_T(f_0)} d\Pi(f)|\mathcal{G}_0\right]\right] = \mathbb{E}_0\left[\mathbb{E}_f\left[\int_{\mathcal{F}^c_T} d\Pi(f)|\mathcal{G}_0\right]\right] = \Pi(\mathcal{F}^c_T),$$

which is controlled by using Assumption (ii). Since $e^{(\kappa_T+1)T\epsilon_T^2} e^{L_T(f)-L_T(f_0)} \geq \mathbb{1}_{\left\{L_T(f)-L_T(f_0)\geq -(\kappa_T+1)T\epsilon_T^2\right\}}$,

$$\begin{aligned}
\mathbb{P}_0\left(D_T \leq e^{-2(\kappa_T+1)T\epsilon_T^2}\Pi(B(\epsilon_T,B))\right) &\leq \mathbb{P}_0\left(\int_{B(\epsilon_T,B)} e^{L_T(f)-L_T(f_0)} \frac{d\Pi(f)}{\Pi(B(\epsilon_T,B))} \leq e^{-2(\kappa_T+1)T\epsilon_T^2}\right) \\
&\leq \mathbb{P}_0\left(\int_{B(\epsilon_T,B)} \mathbb{1}_{\left\{L_T(f)-L_T(f_0)\geq -(\kappa_T+1)T\epsilon_T^2\right\}} \frac{d\Pi(f)}{\Pi(B(\epsilon_T,B))} \leq e^{-(\kappa_T+1)T\epsilon_T^2}\right) \\
&\leq \frac{\mathbb{E}_0\left[\int_{B(\epsilon_T,B)} \mathbb{1}_{\left\{L_T(f)-L_T(f_0)< -(\kappa_T+1)T\epsilon_T^2\right\}} \frac{d\Pi(f)}{\Pi(B(\epsilon_T,B))}\right]}{\left(1-e^{-(\kappa_T+1)T\epsilon_T^2}\right)} \\
&\leq \frac{\int_{B(\epsilon_T,B)} \mathbb{P}_0\left(L_T(f_0)-L_T(f) > (\kappa_T+1)T\epsilon_T^2\right) d\Pi(f)}{\Pi(B(\epsilon_T,B))\left(1-e^{-(\kappa_T+1)T\epsilon_T^2}\right)} \\
&\lesssim \frac{\log\log(T)\log^3(T)}{T\epsilon_T^2},
\end{aligned}$$

by using Lemma 2 of Section 4.2 of [1]. Remember we have set $\rho^0_{k,\ell} := \|h^0_{k,\ell}\|_1$ and $\rho_{k,\ell} := \|h_{k,\ell}\|_1$. Since $h_{k,\ell}$ and $h^0_{k,\ell}$ are non-negative functions, $\int_{-s}^A h^0_{k,\ell}(u)du \le \rho^0_{k,\ell}$, $\int_0^{T-s} h^0_{k,\ell}(u)du \le \rho^0_{k,\ell}$, and note that

$$\begin{aligned} Td_{1,T}(f,f_0) &= \sum_{\ell=1}^K \int_0^T \left| \nu_\ell - \nu_\ell^0 + \sum_{k=1}^K \int_{t-A}^{t^-} (h_{k,\ell} - h^0_{k,\ell})(t-s) dN^k_s \right| dt \\ &\ge \sum_{\ell=1}^K \left| \int_0^T \left(\nu_\ell - \nu_\ell^0 + \sum_{k=1}^K \int_{t-A}^{t^-} (h_{k,\ell} - h^0_{k,\ell})(t-s) dN^k_s \right) dt \right| \\ &\ge \sum_{\ell=1}^K \left| T(\nu_\ell - \nu_\ell^0) + \int_0^T \left(\sum_{k=1}^K \int_{t-A}^{t^-} (h_{k,\ell} - h^0_{k,\ell})(t-s) dN^k_s \right) dt \right|, \end{aligned}$$

then for any $\ell = 1, \dots, K$,

$$\begin{aligned} d_{1,T}(f,f_0) &\ge \left| \nu_\ell - \nu_\ell^0 + \frac{1}{T} \sum_{k=1}^K \int_0^T \int_{t-A}^{t^-} (h_{k,\ell} - h^0_{k,\ell})(t-s) dN^k_s dt \right| \\ &= \Bigg| \nu_\ell - \nu_\ell^0 + \sum_{k=1}^K (\rho_{k,\ell} - \rho^0_{k,\ell}) \frac{N^k[0,T-A]}{T} \\ &\quad + \frac{1}{T} \int_{-A}^0 \int_{-s}^A (h_{k,\ell} - h^0_{k,\ell})(u) du dN^k_s + \frac{1}{T} \int_{T-A}^{T^-} \int_0^{T-s} (h_{k,\ell} - h^0_{k,\ell})(u) du dN^k_s \Bigg| \\ &= \Bigg| \nu_\ell + \sum_{k=1}^K \rho_{k,\ell} \frac{N^k[0,T-A]}{T} + \frac{1}{T} \int_{-A}^0 \int_{-s}^A h_{k,\ell}(u) du dN^k_s + \frac{1}{T} \int_{T-A}^{T^-} \int_0^{T-s} h_{k,\ell}(u) du dN^k_s \\ &\quad - \left(\nu_\ell^0 + \sum_{k=1}^K \rho^0_{k,\ell} \frac{N^k[0,T-A]}{T} + \frac{1}{T} \int_{-A}^0 \int_{-s}^A h^0_{k,\ell}(u) du dN^k_s + \frac{1}{T} \int_{T-A}^{T^-} \int_0^{T-s} h^0_{k,\ell}(u) du dN^k_s \right) \Bigg|. \end{aligned}$$

This implies for $f \in S_j$ that

$$\begin{aligned} \nu_\ell + \sum_{k=1}^K \rho_{k,\ell} \frac{N^k[0,T-A]}{T} &\le \nu_\ell^0 + \sum_{k=1}^K \rho^0_{k,\ell} \frac{N^k[-A,T]}{T} + K(j+1)\epsilon_T \\ \nu_\ell + \sum_{k=1}^K \rho_{k,\ell} \frac{N^k[-A,T]}{T} &\ge \nu_\ell^0 + \sum_{k=1}^K \rho^0_{k,\ell} \frac{N^k[0,T-A]}{T} - K(j+1)\epsilon_T. \end{aligned} \tag{2.2}$$

On Ω_T,

$$\sum_{k=1}^K \rho^0_{k,\ell} \frac{N^k[-A,T]}{T} \le \sum_{k=1}^K \rho^0_{k,\ell} (\mu^0_k + \delta_T),$$

so that, for T large enough, for all $j \geq 1$ $S_j \subset \mathcal{F}_j$ with

$$\mathcal{F}_j := \{f \in \mathcal{F}_T;\ \nu_\ell \leq \mu_\ell^0 + 1 + Kj\epsilon_T, \forall \ell \leq K\},$$

since

$$\mu_\ell^0 = \nu_\ell^0 + \sum_{k=1}^{K} \rho_{k,\ell}^0 \mu_k^0. \tag{2.3}$$

Let $(f_i)_{i=1,\dots,\mathcal{N}_j}$ be the centering points of a minimal $\mathbb{L}_1$-covering of $\mathcal{F}_j$ by balls of radius $\zeta j \epsilon_T$ with $\zeta = 1/(6N_0)$ (with N_0 defined in Section 2 of [1]) and define $\phi_{(j)} = \max_{i=1,\dots,\mathcal{N}_j} \phi_{f_i,j}$ where $\phi_{f_i,j}$ is the individual test defined in Lemma 1 associated to f_i and j (see Section 4.1 of [1]). Note also that there exists a constant C_0 such that

$$\mathcal{N}_j \leq (C_0(1 + j\epsilon_T)/j\epsilon_T)^K \mathcal{N}(\zeta j \epsilon_T/2, \mathcal{H}_T, \|.\|_1)$$

where $\mathcal{N}(\zeta j\epsilon_T/2, \mathcal{H}_T, \|.\|_1)$ is the covering number of $\mathcal{H}_T$ by $\mathbb{L}_1$-balls with radius $\zeta j \epsilon_T/2$. There exists C_K such that if $j\epsilon_T \leq 1$ then $\mathcal{N}_j \leq C_K e^{-K\log(j\epsilon_T)} \mathcal{N}(\zeta j\epsilon_T/2, \mathcal{H}_T, \|.\|_1)$ and if $j\epsilon_T > 1$ then $\mathcal{N}_j \leq C_K \mathcal{N}(\zeta j \epsilon_T/2, \mathcal{H}_T, \|.\|_1)$. Moreover $j \mapsto \mathcal{N}(\zeta j\epsilon_T/2, \mathcal{H}_T, \|.\|_1)$ is monotone non-increasing, choosing $j \geq 2\zeta_0/\zeta$, we obtain that

$$\mathcal{N}_j \leq C_K (\zeta/\zeta_0)^K e^{K\log T} e^{x_0 T \epsilon_T^2},$$

from hypothesis (iii) in Theorem 1. Combining this with Lemma 1, we have for all $j \geq 2\zeta_0/\zeta$,

$$\begin{aligned} \mathbb{E}_0[\mathbb{1}_{\Omega_T}\phi_{(j)}] &\lesssim \mathcal{N}_j e^{-Tx_2(j\epsilon_T \wedge j^2\epsilon_T^2)} \lesssim e^{K\log T} e^{x_0 T\epsilon_T^2} e^{-x_2 T(j\epsilon_T \wedge j^2\epsilon_T^2)} \\ \sup_{f\in\mathcal{F}_j} \mathbb{E}_0\left[\mathbb{E}_f[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi_{(j)})|\mathcal{G}_0]\right] &\lesssim e^{-x_2 T(j\epsilon_T \wedge j^2\epsilon_T^2)}, \end{aligned}$$

for x_2 a constant. Set $\phi = \max_{j\geq M_T} \phi_{(j)}$ with $M_T > 2\zeta_0/\zeta$, then

$$\mathbb{E}_0[\mathbb{1}_{\Omega_T}\phi] \lesssim e^{K\log T} e^{x_0 T\epsilon_T^2} \left[\sum_{j=M_T}^{\lfloor \epsilon_T^{-1} \rfloor} e^{-x_2 T\epsilon_T^2 j^2} + \sum_{j\geq \epsilon_T^{-1}} e^{-Tx_2\epsilon_T j} \right] \lesssim e^{-x_2 T\epsilon_T^2 M_T^2/2}$$

and

$$\sum_{j=M_T}^{\infty} \int_{\mathcal{F}_T} \mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi)|\mathcal{G}_0\right]\right] d\Pi(f) \lesssim e^{-x_2 T\epsilon_T^2 M_T^2/2}.$$

Therefore, using Assumption (i),

$$\frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T, B))} \sum_{j=M_T}^{\infty} \int_{\mathcal{F}_T} \mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi)|\mathcal{G}_0\right]\right] d\Pi(f) = o(1)$$

if M is a constant large enough, which terminates the proof of Theorem 1.

2.2. *Proof of Theorem 2.* The proof of Theorem 2 follows the same lines as for Theorem 1, except that the decomposition of $\mathcal{F}_T$ is based on the sets $\mathcal{F}_j$ and $\mathcal{H}_{T,i}$, $i \geq 1$ and $j \geq M_T$ for some $M_T > 0$. For each $i \geq 1$, $j \geq M_T$, consider $S'_{i,j}$ a maximal set of $\zeta j \epsilon_T$-separated points in $\mathcal{F}_j \cap \mathcal{H}_{T,i}$ (with a slight abuse of notations) and $\phi_{i,j} = \max_{f_1 \in S'_{i,j}} \phi_{f_1}$ with ϕ_{f_1} defined in Lemma 1. Then,

$$|S'_{i,j}| \leq C_K(\zeta/\zeta_0)^K e^{K\log(T)} \mathcal{N}(\zeta j \epsilon_T/2, \mathcal{H}_{T,i}, \|.\|_1).$$

Setting $\bar{N}_{T,ij} := \int_{\mathcal{F}_T \cap \mathcal{H}_{T,i}} \mathbb{1}_{f \in S_j} e^{L_T(f) - L_T(f_0)} d\Pi(f)$, using similar computations as for the proof of Theorem 1, we have

$$\begin{aligned}
\mathbb{E}_0\left[\Pi\left(A^c_{M_T\epsilon_T}|N\right)\right] &\leq \mathbb{P}_0(\Omega_T^c) + \mathbb{P}_0\left(D_T < e^{-2(\kappa_T+1)T\epsilon_T^2}\Pi(B(\epsilon_T, B))\right) + \frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T, B))}\Pi(\mathcal{F}_T^c) \\
&+ \mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\sum_{i=1}^{+\infty}\sum_{j=M_T}^{+\infty}\phi_{ij}\frac{\bar{N}_{T,ij}}{D_T}\right] + \frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T, B))}\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\sum_{i=1}^{+\infty}\sum_{j=M_T}^{+\infty}(1-\phi_{ij})\bar{N}_{T,ij}\right].
\end{aligned}$$

Assumptions of the theorem allow us to deal with the first three terms. So, we just have to bound the last two ones. Using the same arguments and the same notations as for Theorem 1,

$$\begin{aligned}
\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\sum_{i=1}^{+\infty}\sum_{j=M_T}^{+\infty}(1-\phi_{ij})\bar{N}_{T,ij}\right] &= \sum_{i=1}^{+\infty}\int_{\mathcal{F}_T\cap\mathcal{H}_{T,i}}\sum_{j=M_T}^{+\infty}\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi_{ij})e^{L_T(f)-L_T(f_0)}\right]d\Pi(f) \\
&= \sum_{i=1}^{+\infty}\int_{\mathcal{F}_T\cap\mathcal{H}_{T,i}}\sum_{j=M_T}^{+\infty}\mathbb{E}_0\left[\mathbb{E}_f[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi_{ij})|\mathcal{G}_0]\right]d\Pi(f) \\
&\lesssim \sum_{i=1}^{+\infty}\int_{\mathcal{F}_T\cap\mathcal{H}_{T,i}}d\Pi(f)\sum_{j=M_T}^{+\infty}e^{-x_2T(j\epsilon_T\wedge j^2\epsilon_T^2)} \lesssim e^{-x_2T\epsilon_T^2M_T^2/2}.
\end{aligned}$$

Now, for γ a fixed positive constant smaller than x_2, setting $\pi_{T,i} = \Pi(\mathcal{H}_{T,i})$, we have

$$\begin{aligned}
\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\sum_{i=1}^{+\infty}\sum_{j=M_T}^{+\infty}\phi_{ij}\frac{\bar{N}_{T,ij}}{D_T}\right] &\leq \mathbb{P}_0\left(D_T < e^{-2(\kappa_T+1)T\epsilon_T^2}\Pi(B(\epsilon_T,B))\right) + \mathbb{P}_0\left(\exists(i,j);\sqrt{\pi_{T,i}}\phi_{i,j} > e^{-\gamma T(j\epsilon_T\wedge j^2\epsilon_T^2)}\cap\Omega_T\right) \\
&+ \sum_{i=1}^{+\infty}\sum_{j=M_T}^{+\infty}e^{-\gamma T(j\epsilon_T\wedge j^2\epsilon_T^2)}\sqrt{\pi_{T,i}}\frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T,B))}\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\int_{\mathcal{F}_T}\mathbb{1}_{f\in S_j}e^{L_T(f)-L_T(f_0)}d\Pi(f|\mathcal{H}_{T,i})\right].
\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{P}_0\left(\exists(i,j); \sqrt{\pi_{T,i}}\phi_{i,j} > e^{-\gamma T(j\epsilon_T \wedge j^2\epsilon_T^2)} \cap \Omega_T\right) &\leq \sum_{i=1}^{+\infty}\sqrt{\pi_{T,i}}\sum_{j=M_T}^{+\infty} e^{\gamma T(j\epsilon_T\wedge j^2\epsilon_T^2)}\mathbb{E}_0[\mathbb{1}_{\Omega_T}\phi_{i,j}] \\ &\lesssim \sum_{i=1}^{+\infty}\sqrt{\pi_{T,i}}\sum_{j=M_T}^{+\infty} e^{(\gamma-x_2)T(j\epsilon_T\wedge j^2\epsilon_T^2)+K\log(T)}\mathcal{N}(\zeta j\epsilon_T/2, \mathcal{H}_{T,i}, \|.\|_1) \\ &\lesssim e^{(\gamma-x_2)T\epsilon_T^2 M_T^2/2}\sum_{i=1}^{+\infty}\sqrt{\pi_{T,i}}\mathcal{N}(\zeta_0\epsilon_T, \mathcal{H}_{T,i}, \|.\|_1) = o(1).\end{aligned}$$

But, we have

$$\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\int_{\mathcal{F}_T}\mathbb{1}_{f\in S_j}e^{L_T(f)-L_T(f_0)}d\Pi(f|\mathcal{H}_{T,i})\right] \quad \leq \quad 1$$

and

$$\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\sum_{i=1}^{+\infty}\sum_{j=M_T}^{+\infty}\phi_{ij}\frac{\bar{N}_{T,ij}}{D_T}\right] \quad \lesssim \quad \sum_{i=1}^{+\infty}\sqrt{\pi_{T,i}}e^{-\gamma T\epsilon_T^2 M_T^2}\frac{e^{2(\kappa_T+1)T\epsilon_T^2}}{\Pi(B(\epsilon_T, B))} + o(1) = o(1),$$

for M a constant large enough. This terminates the proof of Theorem 2.

2.3. *Proof of Corollary 1.* Let $w_T \to +\infty$. The proof of Corollary 1 follows from the usual convexity argument, so that

$$\|\hat{f} - f_0\|_1 \leq w_T\varepsilon_T + \mathbb{E}^\pi\left[\|f-f_0\|_1\mathbb{1}_{\|f-f_0\|_1 > w_T\varepsilon_T}|N\right],$$

together with a control of the second term of the right hand side similar to the proof of Theorem 3. We write

$$\mathbb{E}^\pi\left[\|f-f_0\|_1\mathbb{1}_{\|f-f_0\|_1>w_T\varepsilon_T}|N\right] \leq \mathbb{E}^\pi\left[\|f-f_0\|_1\mathbb{1}_{A_{L_1}(w_T\varepsilon_T)^c}\mathbb{1}_{A_{\varepsilon_T}}|N\right] + \mathbb{E}^\pi\left[\|f-f_0\|_1\mathbb{1}_{A^c_{\varepsilon_T}}|N\right]$$

and since $\int\|f-f_0\|_1 d\Pi(f) \leq \|f_0\|_1 + \int\|f\|_1 d\Pi(f) < \infty$,

$$\begin{aligned}\mathbb{P}_0\left(\mathbb{E}^\pi\left[\|f-f_0\|_1\mathbb{1}_{A_{L_1}(w_T\varepsilon_T)^c}\mathbb{1}_{A_{\varepsilon_T}}|N\right] > w_T\varepsilon_T\right) &\leq \mathbb{P}_0\left(\Omega_{1,T}^c\right) + \mathbb{P}_0\left(D_T < e^{-c_1T\epsilon_T^2}\right) \\ &+ \frac{e^{c_1T\epsilon_T^2}}{w_T\varepsilon_T}\int_{A_{L_1}(w_T\varepsilon_T)^c}\|f-f_0\|_1\mathbb{E}_0\left[\mathbb{P}_f\left(\Omega_{1,T}\cap\{d_{1,T}(f_0,f)\leq\varepsilon_T\}\right)|\mathcal{G}_0\right]d\Pi(f) \\ &\leq o(1) + o(1)\int\|f-f_0\|_1 d\Pi(f) = o(1),\end{aligned}$$

where the last inequality comes from the proof of Theorem 3. Similarly, using the proof of Theorem 1,

$$\begin{aligned}\mathbb{P}_0\left(\mathbb{E}^\pi\left[\|f-f_0\|_1\mathbb{1}_{A^c_{\varepsilon_T}}|N\right]>w_T\varepsilon_T\right)&\leq\mathbb{P}_0\left(\Omega_T^c\right)+\mathbb{P}_0\left(D_T<e^{-c_1T\epsilon_T^2}\right)+\mathbb{E}_0[\mathbb{1}_{\Omega_T}\phi]\\&+\frac{e^{c_1T\epsilon_T^2}}{w_T\varepsilon_T}\int_{A_{L_1}(w_T\varepsilon_T)^c}\|f-f_0\|_1\mathbb{E}_0\left[\mathbb{E}_f\left[(1-\phi)\mathbb{1}_{\Omega_T}\mathbb{1}_{\{d_{1,T}(f_0,f)>\varepsilon_T\}}\right]|\mathcal{G}_0\right]d\Pi(f)\\&\leq o(1)+o(1)\int\|f-f_0\|_1d\Pi(f),\end{aligned}$$

and $\mathbb{P}_0(\|\hat{f}-f_0\|_1>3w_T\varepsilon_T)=o(1)$. Since this is true for any $w_T\to+\infty$, this terminates the proof.

2.4. *Proof of Lemma 1.* For the sake of completeness, we recall the statement of Lemma 1.

LEMMA 1. *Let $j\geq 1$, $f_1\in\mathcal{F}_j$ and define the test*

$$\phi_{f_1,j}=\max_{\ell=1,\dots,K}\left(\mathbb{1}_{\{N^\ell(A_{1,\ell})-\Lambda^\ell(A_{1,\ell};f_0)\geq jT\epsilon_T/8\}}\vee\mathbb{1}_{\{N^\ell(A^c_{1,\ell})-\Lambda^\ell(A^c_{1,\ell};f_0)\geq jT\epsilon_T/8\}}\right),$$

with for all $\ell\leq K$, $A_{1,\ell}=\{t\in[0,T];\ \lambda^\ell_t(f_1)\geq\lambda^\ell_t(f_0)\}$, $\Lambda^\ell(A_{1,\ell};f_0)=\int_0^T\mathbb{1}_{A_{1,\ell}}(t)\lambda^\ell_t(f_0)dt$ and $\Lambda^\ell(A^c_{1,\ell};f_0)=\int_0^T\mathbb{1}_{A^c_{1,\ell}}(t)\lambda^\ell_t(f_0)dt$. Then

$$\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\phi_{f_1,j}\right]+\sup_{\|f-f_1\|_1\leq j\epsilon_T/(6N_0)}\mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\Omega_T}\mathbb{1}_{f\in S_j}(1-\phi_{f_1,j})|\mathcal{G}_0\right]\right]\leq(2K+1)\max_\ell e^{-x_{1,\ell}Tj\epsilon_T(\sqrt{\mu^0_\ell}\wedge j\epsilon_T)},$$

with N_0 is defined in Section 2 of [1] and

$$x_{1,\ell}=\min\left(36,1/(4096\mu^0_\ell),1/\left(1024K\sqrt{\mu^0_\ell}\right)\right).$$

PROOF OF LEMMA 1. Let $j\geq 1$ and $f_1=((\nu^1_k)_{k=1,\dots,K},(h^1_{\ell,k})_{k,\ell=1,\dots,K})\in\mathcal{F}_j$. Let $\ell\in\{1,\dots,K\}$ and let

$$\phi_{j,A_{1,\ell}}=\mathbb{1}_{\left\{N^\ell(A_{1,\ell})-\Lambda^\ell(A_{1,\ell};f_0)\geq jT\epsilon_T/8\right\}}.$$

By using (2.3), observe that on the event Ω_T,

$$\begin{aligned}\int_0^T\lambda^\ell_s(f_0)ds&=\nu^0_\ell T+\sum_{k=1}^K\int_0^T\int_{s-A}^{s^-}h^0_{k,\ell}(s-u)dN^k_u ds\\&\leq\nu^0_\ell T+\sum_{k=1}^K\int_{-A}^{T^-}\int_0^T\mathbb{1}_{u<s\leq A+u}h^0_{k,\ell}(s-u)dsdN^k_u\end{aligned}$$

and for T large enough,

$$\int_0^T\lambda^\ell_s(f_0)ds\leq\nu^0_\ell T+\sum_{k=1}^K\rho^0_{k,\ell}N^k[-A,T]\leq 2T\mu^0_\ell. \tag{2.4}$$

Let $j \leq \sqrt{\mu_\ell^0}\epsilon_T^{-1}$ and $x = x_1 j^2 T\epsilon_T^2$, for x_1 a constant. We use inequality (7.7) of [4], with $\tau = T$, $H_t = 1_{A_{1,\ell}}(t)$, $v = 2T\mu_\ell^0$ and $M_T = N^\ell(A_{1,\ell}) - \Lambda^\ell(A_{1,\ell}; f_0)$. So,

$$\mathbb{P}_0\left(\left\{N^\ell(A_{1,\ell}) - \Lambda^\ell(A_{1,\ell}; f_0) \geq \sqrt{2vx} + \frac{x}{3}\right\} \cap \Omega_T\right) \leq e^{-x_1 j^2 T\epsilon_T^2}.$$

If $x_1 \leq 1/(1024\mu_\ell^0)$ and $x_1 \leq 36$, we have that

$$\sqrt{2vx} + \frac{x}{3} = 2\sqrt{\mu_\ell^0 x_1} jT\epsilon_T + \frac{x_1 j^2 T\epsilon_T^2}{3} \leq 2\sqrt{\mu_\ell^0 x_1}\left(1 + \frac{\sqrt{x_1}}{6}\right) jT\epsilon_T \leq \frac{jT\epsilon_T}{8}. \tag{2.5}$$

Then

$$\mathbb{P}_0\left(\left\{N^\ell(A_{1,\ell}) - \Lambda^\ell(A_{1,\ell}; f_0) \geq \frac{jT\epsilon_T}{8}\right\} \cap \Omega_T\right) \leq e^{-x_1 j^2 T\epsilon_T^2}.$$

If $j \geq \sqrt{\mu_\ell^0}\epsilon_T^{-1}$, we apply the same inequality but with $x = x_0 jT\epsilon_T$ with $x_0 = \sqrt{\mu_\ell^0} \times x_1$. Then,

$$\sqrt{2vx} + \frac{x}{3} = 2\sqrt{\mu_\ell^0 x_1 \sqrt{\mu_\ell^0} j\epsilon_T T} + \frac{x_1\sqrt{\mu_\ell^0} jT\epsilon_T}{3} \leq 2\sqrt{\mu_\ell^0 x_1} jT\epsilon_T + \frac{x_1\sqrt{\mu_\ell^0} jT\epsilon_T}{3} \leq \frac{jT\epsilon_T}{8},$$

where we have used (2.5). It implies

$$\mathbb{P}_0\left(\left\{N^\ell(A_{1,\ell}) - \Lambda^\ell(A_{1,\ell}; f_0) \geq \frac{jT\epsilon_T}{8}\right\} \cap \Omega_T\right) \leq e^{-x_0 jT\epsilon_T}.$$

Finally $\mathbb{E}_0\left[\mathbb{1}_{\Omega_T}\phi_{j,A_{1,\ell}}\right] \leq e^{-x_1 Tj\epsilon_T(\sqrt{\mu_\ell^0}\wedge j\epsilon_T)}$. Now, assume that

$$\int_{A_{1,\ell}} (\lambda_t^\ell(f_1) - \lambda_t^\ell(f_0))dt \geq \int_{A_{1,\ell}^c} (\lambda_t^\ell(f_0) - \lambda_t^\ell(f_1))dt.$$

Then

$$\frac{\|\lambda^\ell(f_1) - \lambda^\ell(f_0)\|_1}{2} := \frac{\int_0^T |\lambda_t^\ell(f_1) - \lambda_t^\ell(f_0)|dt}{2} \leq \int_{A_{1,\ell}} (\lambda_t^\ell(f_1) - \lambda_t^\ell(f_0))dt. \tag{2.6}$$

Let $f = ((\nu_k)_{k=1,\dots,K}, (h_{\ell,k})_{k,\ell=1,\dots,K}) \in S_j$ satisfying $\|f - f_1\|_1 \leq \zeta j\epsilon_T$ for some $\zeta > 0$. Then,

$$\begin{aligned}\|\lambda^\ell(f) - \lambda^\ell(f_1)\|_1 &\leq T|\nu_\ell - \nu_\ell^1| + \int_0^T \left|\int_{t-A}^{t^-} \sum_k (h_{k,\ell} - h_{k,\ell}^1)(t-u)dN_u^k\right| dt \\ &\leq T|\nu_\ell - \nu_\ell^1| + \sum_k \int_0^T \int_{t-A}^{t^-} |(h_{k,\ell} - h_{k,\ell}^1)(t-u)|dN_u^k dt \\ &\leq T|\nu_\ell - \nu_\ell^1| + \max_k N^k[-A, T] \sum_k \|h_{k,\ell} - h_{k,\ell}^1\|_1 \leq TN_0\|f - f_1\|_1\end{aligned} \tag{2.7}$$

and $\|\lambda^\ell(f)-\lambda^\ell(f_1)\|_1 \le TN_0\zeta j\epsilon_T$. Since $f\in S_j$, there exists ℓ (depending on f) such that

$$\|\lambda^\ell(f)-\lambda^\ell(f_0)\|_1 \ge jT\epsilon_T.$$

This implies in particular that if $N_0\zeta<1$,

$$\|\lambda^\ell(f_1)-\lambda^\ell(f_0)\|_1 \ge \|\lambda^\ell(f)-\lambda^\ell(f_0)\|_1 - TN_0\zeta j\epsilon_T \ge (1-N_0\zeta)Tj\epsilon_T.$$

We then have

$$\begin{aligned}\Lambda^\ell(A_{1,\ell};f)-\Lambda^\ell(A_{1,\ell};f_0) &= \Lambda^\ell(A_{1,\ell};f)-\Lambda^\ell(A_{1,\ell};f_1)+\Lambda^\ell(A_{1,\ell};f_1)-\Lambda^\ell(A_{1,\ell};f_0)\\ &\ge -\|\lambda^\ell(f)-\lambda^\ell(f_1)\|_1+\int_{A_{1,\ell}}(\lambda^\ell_t(f_1)-\lambda^\ell_t(f_0))dt\\ &\ge -\|\lambda^\ell(f)-\lambda^\ell(f_1)\|_1+\frac{\|\lambda^\ell(f_1)-\lambda^\ell(f_0)\|_1}{2}\\ &\ge -TN_0\zeta j\epsilon_T+\frac{(1-N_0\zeta)Tj\epsilon_T}{2}=(1/2-3N_0\zeta/2)Tj\epsilon_T.\end{aligned}$$

Taking $\zeta=1/(6N_0)$ leads to

$$\begin{aligned}\mathbb{E}_f\left[\mathbb{1}_{f\in S_j}(1-\phi_{j,A_{1,\ell}})\mathbb{1}_{\Omega_T}|\mathcal{G}_0\right] &= \mathbb{E}_f\left[\mathbb{1}_{f\in S_j}\mathbb{1}_{\left\{N^\ell(A_{1,\ell})-\Lambda^\ell(A_{1,\ell};f_0)<jT\epsilon_T/8\right\}}\mathbb{1}_{\Omega_T}|\mathcal{G}_0\right]\\ &\le \mathbb{E}_f\left[\mathbb{1}_{f\in S_j}\mathbb{1}_{\left\{N^\ell(A_{1,\ell})-\Lambda^\ell(A_{1,\ell};f)\le -jT\epsilon_T/8\right\}}\mathbb{1}_{\Omega_T}|\mathcal{G}_0\right]\\ &\le \mathbb{E}_f\left[\mathbb{1}_{\left\{N^\ell(A_{1,\ell})-\Lambda^\ell(A_{1,\ell};f)\le -jT\epsilon_T/8\right\}}\mathbb{1}_{\Omega_T}|\mathcal{G}_0\right].\end{aligned}$$

Note that we can adapt inequality (7.7) of [4], with $H_t=\mathbb{1}_{A_{1,\ell}}(t)$ to the case of conditional probability given $\mathcal{G}_0$ since the process E_t defined in the proof of Theorem 3 of [4], being a supermartingale, satisfies $\mathbb{E}_f[E_t|\mathcal{G}_0]\le E_0=1$ and, given that from (2.2) and (2.4),

$$\int_0^T \lambda^\ell_s(f)ds \le \nu_\ell T+\sum_{k=1}^K \rho_{k,\ell}N^k[-A,T] \le 2T\mu^0_\ell+K(j+1)T\epsilon_T =: \tilde v$$

for T large enough, we obtain

$$\mathbb{E}_f\left[\mathbb{1}_{\left\{N^\ell(A_{1,\ell})-\Lambda^\ell(A_{1,\ell};f)\le -\sqrt{2\tilde v x}-\frac{x}{3}\right\}}\mathbb{1}_{\Omega_T}|\mathcal{G}_0\right]\le e^{-x}.$$

We use the same computations as before, observing that $\tilde v = v+K(j+1)T\epsilon_T$.

If $j \leq \sqrt{\mu_\ell^0}\epsilon_T^{-1}$ we set $x = x_1 j^2 T\epsilon_T^2$, for x_1 a constant. Then,

$$\begin{aligned}
\sqrt{2\tilde{v}x} + \frac{x}{3} &\leq \sqrt{2vx} + \frac{x}{3} + \sqrt{2K(j+1)T\epsilon_T x} \\
&\leq 2\sqrt{\mu_\ell^0 x_1} jT\epsilon_T + \frac{x_1 j^2 T\epsilon_T^2}{3} + \sqrt{2K(j+1)\epsilon_T x_1} jT\epsilon_T \\
&\leq 2\sqrt{\mu_\ell^0 x_1}\left(1 + \frac{\sqrt{x_1}}{6}\right) jT\epsilon_T + 2\sqrt{Kj\epsilon_T x_1} jT\epsilon_T \\
&\leq \left(2\sqrt{\mu_\ell^0 x_1}\left(1 + \frac{\sqrt{x_1}}{6}\right) + 2\sqrt{K\sqrt{\mu_\ell^0} x_1}\right) jT\epsilon_T.
\end{aligned}$$

Therefore, if $x_1 \leq \min\left(36, 1/(4096\mu_\ell^0), 1/\left(1024K\sqrt{\mu_\ell^0}\right)\right)$, then

$$\sqrt{2\tilde{v}x} + \frac{x}{3} \leq \frac{jT\epsilon_T}{8}.$$

If $j \geq \sqrt{\mu_\ell^0}\epsilon_T^{-1}$, we set $x = x_0 jT\epsilon_T$ with $x_0 = \sqrt{\mu_\ell^0} \times x_1$. Then,

$$\begin{aligned}
\sqrt{2\tilde{v}x} + \frac{x}{3} &\leq \sqrt{2vx} + \frac{x}{3} + \sqrt{2K(j+1)T\epsilon_T x} \\
&\leq 2\sqrt{\mu_\ell^0 x_1 \sqrt{\mu_\ell^0} j\epsilon_T T} + \frac{x_1\sqrt{\mu_\ell^0} jT\epsilon_T}{3} + \sqrt{2K(j+1)T\epsilon_T\sqrt{\mu_\ell^0} x_1 jT\epsilon_T} \\
&\leq 2\sqrt{\mu_\ell^0 x_1} jT\epsilon_T + \frac{x_1\sqrt{\mu_\ell^0} jT\epsilon_T}{3} + 2\sqrt{K\sqrt{\mu_\ell^0} x_1} jT\epsilon_T \leq \frac{jT\epsilon_T}{8}.
\end{aligned}$$

Therefore,

$$\mathbb{E}_f\left[\mathbb{1}_{\left\{N^\ell(A_{1,\ell}) - \Lambda^\ell(A_{1,\ell}; f) \leq -jT\epsilon_T/8\right\}} \mathbb{1}_{\Omega_T} | \mathcal{G}_0\right] \leq e^{-x_1 Tj\epsilon_T(\sqrt{\mu_\ell^0} \wedge j\epsilon_T)}.$$

Now, if

$$\int_{A_{1,\ell}} (\lambda_t^\ell(f_1) - \lambda_t^\ell(f_0))dt < \int_{A_{1,\ell}^c} (\lambda_t^\ell(f_0) - \lambda_t^\ell(f_1))dt,$$

then

$$\int_{A_{1,\ell}^c} (\lambda_t^\ell(f_1) - \lambda_t^\ell(f_0))dt \geq \frac{\|\lambda^\ell(f_1) - \lambda^\ell(f_0)\|_1}{2}$$

and the same computations are run with $A_{1,\ell}$ playing the role of $A_{1,\ell}^c$. This ends the proof of Lemma 1. □

2.5. *Technical lemmas.*

2.5.1. *Control of the number of occurrences of the process on a fixed interval.*

LEMMA 2. *For any $M \geq 1$, for any $\alpha > 0$, there exists a constant C_α only depending on f_0 such that for any $T > 0$, the set*

$$\tilde{\Omega}_T = \left\{ \max_{\ell \in \{1,\dots,K\}} \sup_{t \in [0,T]} N^\ell([t-A,t)) \leq C_\alpha \log T \right\}$$

satisfies

$$\mathbb{P}_0(\tilde{\Omega}_T^c) \leq T^{-\alpha}$$

and for any $1 \leq m \leq M$

$$\mathbb{E}_0 \left[\max_{\ell \in \{1,\dots,K\}} \sup_{t \in [0,T]} \Big(N^\ell([t-A,t)) \Big)^m \times 1_{\tilde{\Omega}_T^c} \right] \leq 2T^{-\alpha/2},$$

for T large enough.

PROOF. For the first part, we split the interval $[-A;T]$ into disjoint intervals of length A and we use Proposition 2 of [4]. For the second part, we set

$$X := \max_{\ell \in \{1,\dots,K\}} \sup_{t \in [0,T]} \Big(N^\ell([t-A,t)) \Big) \times 1_{\tilde{\Omega}_T^c} \geq 0$$

and the equality

$$\begin{aligned}
\mathbb{E}_0[X^m] &= \int_0^{+\infty} m x^{m-1} \mathbb{P}_0(X > x) dx \\
&= \int_0^{C_\alpha \log T} m x^{m-1} \mathbb{P}_0\left(X > x\right) dx + \int_{C_\alpha \log T}^{+\infty} m x^{m-1} \mathbb{P}_0\left(X > x\right) dx \\
&\leq m (C_\alpha \log T)^{m-1} \int_0^{C_\alpha \log T} \mathbb{P}_0(\tilde{\Omega}_T^c) dx + \int_{C_\alpha \log T}^{+\infty} m x^{m-1} \mathbb{P}_0\left(X > x\right) dx \\
&\leq m (C_\alpha \log T)^m T^{-\alpha} + \int_{C_\alpha \log T}^{+\infty} m x^{m-1} \mathbb{P}_0\left(X > x\right) dx.
\end{aligned}$$

Furthermore, for T large enough,

$$\begin{aligned}
\int_{C_\alpha \log T}^{+\infty} m x^{m-1} \mathbb{P}_0(X > x) dx &\leq \int_{C_\alpha \log T}^{+\infty} m x^{m-1} \mathbb{P}_0\left(\max_{\ell \in \{1,\dots,K\}} \sup_{t \in [0,T]} \Big(N^\ell([t-A,t)) \Big) > x \right) dx \\
&\leq \int_{C_\alpha \log T}^{+\infty} m x^{m-1} \mathbb{P}_0\left(\max_{\ell \in \{1,\dots,K\}} \sup_{t \in [0,e^{x/C_\alpha}]} \Big(N^\ell([t-A,t)) \Big) > x \right) dx \\
&\leq \int_{C_\alpha \log T}^{+\infty} m x^{m-1} \exp(-\alpha x / C_\alpha) dx \leq T^{-\alpha/2}.
\end{aligned}$$

□

2.5.2. *Control of* $N[0,T]$. Let $k \in \{1,\dots,K\}$. We have the following result.

LEMMA 3. *For any* $k \in \{1,\dots,K\}$*, for all* $\alpha > 0$ *there exists* $\delta_0 > 0$ *such that*

$$\mathbb{P}_0\left(\left|\frac{N^k[0,T]}{T} - \mu_k^0\right| \geq \delta_0\sqrt{\frac{(\log T)^3}{T}}\right) = O(T^{-\alpha}).$$

PROOF OF LEMMA 3. We use Proposition 3 of [4] and notations introduced for this result. We denote $N[-A,0)$ the total number of points of N in $[-A,0)$, all marks included. Let $\delta_T := \delta_0\sqrt{(\log T)^3/T}$, with δ_0 a constant. We have

$$\mathbb{P}_0\left(\left|\frac{N^k[0,T]}{T} - \mu_k^0\right| > \delta_T\right) \leq \mathbb{P}_0\left(\left|N^k[0,T] - \int_0^T \lambda_t^k(f_0)dt\right| > \frac{T\delta_T}{2}\right) + \mathbb{P}_0\left(\left|\int_0^T [\lambda_t^k(f_0) - \mu_k^0]dt\right| > \frac{T\delta_T}{2}\right) \tag{2.8}$$

and we observe that

$$\lambda_t^k(f_0) = \nu_k^0 + \int_{t-A}^{t^-}\sum_{\ell=1}^K h_{\ell,k}^0(t-s)dN_s^\ell = Z \circ \mathfrak{S}_t(N),$$

with $Z(N) = \lambda_0^k(f_0)$, where $\mathfrak{S}$ is the shift operator introduced in Proposition 3 of [4]. We then have

$$Z(N) \leq b(1 + N[-A,0))$$

with

$$b = \max_k \max\{\nu_k^0, \max_\ell \|h_{\ell,k}^0\|_\infty\}.$$

So, for any $\alpha > 0$, the second term of (2.8) is $O(T^{-\alpha})$ for δ_0 large enough depending on α and f_0. The first term is controlled by using Inequality (7.7) of [4] with $\tau = T$, $x = x_0 T\delta_T^2$, $H_t = 1$, $v = \mu_k^0 T + T\delta_T/2$ and

$$M_T = N^k[0,T] - \int_0^T \lambda_t^k(f_0)dt.$$

We take x_0 a positive constant such that $\sqrt{8\mu_k^0 x_0} < 1$, so that, for T large enough

$$\frac{T\delta_T}{2} \geq \sqrt{2vx} + x/3.$$

Therefore, we have

$$\begin{aligned}
\mathbb{P}_0\left(|M_T| > \frac{T\delta_T}{2}\right) &\leq \mathbb{P}_0\left(|M_T| \geq \sqrt{2vx} + x/3 \text{ and } \int_0^T \lambda_t^k(f_0)dt \leq v\right) + \mathbb{P}_0\left(\int_0^T \lambda_t^k(f_0)dt > v\right)\\
&\leq 2\exp(-x) + \mathbb{P}_0\left(\left|\int_0^T [\lambda_t^k(f_0) - \mu_k^0]dt\right| > \frac{T\delta_T}{2}\right)\\
&\leq 2\exp(-x_0\delta_0^2(\log T)^3) + O(T^{-\alpha}) = O(T^{-\alpha}),
\end{aligned}$$

which terminates the proof. □

2.5.3. *Lemma on* $\mathbb{E}_f[Z_{1,\ell}]$. We have the following result which is useful to prove Theorem 3 of [1].

LEMMA 4. *For for all* $f \in \mathcal{F}_T$ *such that* $d_{1,T}(f, f_0) \leq \varepsilon_T$, *there exists* ℓ *(depending on* f *and* f^0*) such that on* Ω_T,

$$\mathbb{E}_f[Z_{1,\ell}] \geq C\frac{T}{J_T}\|f - f^0\|_1,$$

where C *is a constant depending on* f^0.

PROOF. By using the first bound of (2.2), we observe that on Ω_T, for any ℓ, since $\inf_\ell \nu_\ell^0 > 0$, then $\inf_\ell \mu_\ell^0 > 0$ (by using (2.3)) and we obtain that $\sum_{k=1}^K \rho_{k,\ell}$ and $\sum_{k=1}^K \nu_k$ are bounded. Therefore $\|f\|_1$ is bounded. On Ω_T, since $\varepsilon_T \geq \delta_T$, still using (2.2), for any ℓ,

$$\nu_\ell + \sum_{k=1}^K \rho_{k,\ell}\mu_k^0 - M\varepsilon_T \leq \nu_\ell^0 + \sum_{k=1}^K \rho_{k,\ell}^0\mu_k^0 \leq \nu_\ell + \sum_{k=1}^K \rho_{k,\ell}\mu_k^0 + M\varepsilon_T$$

for M a constant large enough. By using the formula

$$\nu_\ell + \sum_{k=1}^K \rho_{k,\ell}\mu_k = \mu_\ell, \quad \nu_\ell^0 + \sum_{k=1}^K \rho_{k,\ell}^0\mu_k^0 = \mu_\ell^0,$$

we obtain

$$\left|(\mu_\ell - \mu_\ell^0) - \sum_k \rho_{k,\ell}(\mu_k - \mu_k^0)\right| \leq M\varepsilon_T,$$

which means that

$$\|(I_d - \rho^T)(\mu - \mu^0)\|_\infty \leq M\varepsilon_T.$$

Therefore, since $\|\rho\| = \|\rho^T\|$ ($\rho\rho^T$ and $\rho^T\rho$ have the same eigenvalues),

$$\begin{aligned}
\|\mu - \mu_0\|_2 &= \|(I_d - \rho^T)^{-1}(I_d - \rho^T)(\mu - \mu_0)\|_2 \\
&\leq (1 - \|\rho\|)^{-1}\sqrt{K}\|(I_d - \rho^T)(\mu - \mu^0)\|_\infty \\
&\leq (1 - \|\rho\|)^{-1}\sqrt{K}M\varepsilon_T.
\end{aligned}$$

Since $f \in \mathcal{F}_T$, $1 - \|\rho\| \geq u_T \gtrsim \varepsilon_T^{1/3}(\log T)^{1/6}$. Therefore, μ is bounded. As in [4], we denote $\mathbb{Q}_f$ a measure such that under $\mathbb{Q}_f$ the distribution of the full point process restricted to $(-\infty, 0]$ is identical to the distribution under $\mathbb{P}_f$ and such that on $(0, \infty)$ the process consists of independent components each being a homogeneous Poisson process with rate 1. Furthermore, the Poisson processes should be independent of the process on $(-\infty, 0]$. From Corollary 5.1.2 in [5] the likelihood process is given by

$$\mathcal{L}_t(f) = \exp\left(Kt - \sum_{k=1}^K \int_0^t \lambda_u^k(f)du + \sum_{k=1}^K \int_0^t \log(\lambda_u^k(f))dN_u^k\right).$$

Let $\tau > 0$ satisfying
$$0 < \frac{A\tau K^2}{1-\tau K} < \frac{1}{2} \quad \text{and} \quad \tau \leq \frac{\min_{\ell'} \nu^0_{\ell'}}{2C'_0},$$
with C'_0 an upper bound of $\|f - f_0\|_1$.

- Assume that for any ℓ', $\left|\nu_{\ell'} - \nu^0_{\ell'}\right| < \tau\|f - f_0\|_1$. Then, for any ℓ',
$$\left|\nu_{\ell'} - \nu^0_{\ell'}\right| < \tau\|f - f_0\|_1 = \tau\left(\sum_k \left|\nu_k - \nu^0_k\right| + \sum_{k,\ell} \|h_{k,\ell} - h^0_{k,\ell}\|_1\right)$$
and
$$\left|\nu_{\ell'} - \nu^0_{\ell'}\right| \leq \sum_\ell \left|\nu_\ell - \nu^0_\ell\right| < \frac{\tau K}{1-\tau K}\sum_{k,\ell}\|h_{k,\ell} - h^0_{k,\ell}\|_1.$$
Let ℓ such that
$$\sum_k \|h_{k,\ell} - h^0_{k,\ell}\|_1 = \max_{\ell'}\left\{\sum_k \|h_{k,\ell'} - h^0_{k,\ell'}\|_1\right\}.$$
Then, for any ℓ',
$$\left|\nu_{\ell'} - \nu^0_{\ell'}\right| < \frac{\tau K^2}{1-\tau K}\sum_k \|h_{k,\ell} - h^0_{k,\ell}\|_1, \tag{2.9}$$
and
$$\begin{aligned}\|f - f^0\|_1 &= \sum_{\ell'}\left|\nu_{\ell'} - \nu^0_{\ell'}\right| + \sum_{\ell'}\sum_k \|h_{k,\ell'} - h^0_{k,\ell'}\|_1 \\ &\leq \left(\frac{\tau K^2}{1-\tau K} + K\right)\sum_k \|h_{k,\ell} - h^0_{k,\ell}\|_1. \qquad (2.10)\end{aligned}$$
We denote
$$\Omega_k = \left\{\max_{k'\neq k} N^{k'}[0,A] = 0, \quad N^k[0,A] = 1, \quad N^{k'}[-A,0] \leq aA\mu_{k'}\ \forall k'\right\},$$

where a is a fixed constant chosen later. We then have

$$\begin{aligned}
\mathbb{E}_f[Z_{m,\ell}] &= \frac{T}{2J_T}\mathbb{E}_f\left[\left|\nu_\ell - \nu_\ell^0 + \sum_{k=1}^K \int_0^{A^-} (h_{k,\ell} - h_{k,\ell}^0)(A-s)dN_s^k\right|\right] \\
&\geq \frac{T}{2J_T}\sum_k \mathbb{E}_f\left[\mathbb{1}_{\max_{k'\neq k} N^{k'}[0,A]=0}\mathbb{1}_{N^k[0,A]=1}\left|\nu_\ell - \nu_\ell^0 + \int_0^{A^-} (h_{k,\ell} - h_{k,\ell}^0)(A-s)dN_s^k\right|\right] \\
&\geq \frac{T}{2J_T}\sum_k \mathbb{E}_{\mathbb{Q}_f}\left[\mathcal{L}_A(f)\mathbb{1}_{\max_{k'\neq k} N^{k'}[0,A]=0}\mathbb{1}_{N^k[0,A]=1}\left|\nu_\ell - \nu_\ell^0 + \int_0^{A^-} (h_{k,\ell} - h_{k,\ell}^0)(A-s)dN_s^k\right|\right] \\
&\geq \frac{T}{2J_T}\sum_k \mathbb{E}_{\mathbb{Q}_f}\left[\mathcal{L}_A(f)\mathbb{1}_{\Omega_k}\left|\nu_\ell - \nu_\ell^0 + \int_0^{A^-} (h_{k,\ell} - h_{k,\ell}^0)(A-s)dN_s^k\right|\right].
\end{aligned}$$

Note that on Ω_k,

$$\begin{aligned}
\mathcal{L}_A(f) &:= \exp\left(KA - \sum_{k'}\int_0^A \lambda_t^{k'}(f)dt + \sum_{k'}\int_0^A \log(\lambda_t^{k'}(f))dN_t^{k'}\right) \\
&\geq \nu_k \exp(KA)\exp\left(-\sum_{k'}\int_0^A \lambda_t^{k'}(f)dt\right) \\
&\geq \nu_k \exp(KA)\exp\left(-\sum_{k'}\int_0^A \left(\nu_{k'} + \int_{t-A}^{t-}\sum_{k''} h_{k''k'}(t-u)dN_u^{k''}\right)dt\right) \\
&\geq \nu_k \exp\left(KA - A\sum_{k'}\nu_{k'}\right)\exp\left(-\int_{-A}^{A^-}\sum_{k',k''}\rho_{k''k'}dN_u^{k''}\right) \\
&\geq \nu_k \exp\left(KA - A\sum_{k'}\nu_{k'}\right)\exp\left(-aA\sum_{k''}\mu_{k''}\sum_{k'}\rho_{k''k'} - \sum_{k'}\rho_{kk'}\right).
\end{aligned}$$

Since on $\mathcal{F}_T$,

$$\nu_k \exp\left(KA - A\sum_{k'}\nu_{k'}\right)\exp\left(-aA\sum_{k''}\mu_{k''}\sum_{k'}\rho_{k''k'} - \sum_{k'}\rho_{kk'}\right) \geq \nu_k e^{-KaAC_1} \geq \nu_k^0 e^{-KaAC_1}/2 \geq C(f_0),$$

where C_1 and $C(f_0)$ are some constants, we have, by definition of $\mathbb{Q}_f$,

$$\begin{aligned}
I_k &:= \mathbb{E}_{\mathbb{Q}_f}\left[\mathcal{L}_A(f)\mathbb{1}_{\Omega_k}\left|\nu_\ell - \nu_\ell^0 + \int_0^{A^-} (h_{k,\ell} - h_{k,\ell}^0)(A-s)dN_s^k\right|\right] \\
&\geq C(f_0)\mathbb{E}_{\mathbb{Q}_f}\left[\mathbb{1}_{N^k[0,A]=1}\left|\nu_\ell - \nu_\ell^0 + \int_0^{A^-} (h_{k,\ell} - h_{k,\ell}^0)(A-s)dN_s^k\right|\right] \\
&\quad \times \mathbb{Q}_f(N^{k'}[-A,0] \leq aA\mu_{k'}\ \forall k') \times \mathbb{Q}_f(\max_{k'\neq k} N^{k'}[0,A] = 0).
\end{aligned}$$

Under $\mathbb{Q}_f$, $N^k[0,A] \sim \text{Poisson}(A)$. If $U \sim Unif([0,A])$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}_f}\left[\mathbb{1}_{N^k[0,A]=1}\left|\int_0^{A^-}(h_{k,\ell}-h^0_{k,\ell})(A-s)dN^k_s\right|\right] &= \mathbb{E}\left[\left|(h_{k,\ell}-h^0_{k,\ell})(A-U)\right|\right]\mathbb{Q}_f(N^k[0,A]=1)\\
&= \frac{1}{A}\int_0^A \left|(h_{k,\ell}-h^0_{k,\ell})(A-s)\right| ds \times Ae^{-A}\\
&= e^{-A}\|h_{k,\ell}-h^0_{k,\ell}\|_1.
\end{aligned}$$

We also have, using (2.9),

$$\mathbb{E}_{\mathbb{Q}_f}\left[\mathbb{1}_{N^k[0,A]=1}\left|\nu_\ell-\nu^0_\ell\right|\right] = Ae^{-A}\left|\nu_\ell-\nu^0_\ell\right| \le Ae^{-A}\frac{\tau K^2}{1-\tau K}\sum_k \|h_{k,\ell}-h^0_{k,\ell}\|_1.$$

Furthermore,

$$\mathbb{Q}_f(\max_{k'\neq k} N^{k'}[0,A]=0) = \exp(-(K-1)A),$$

and

$$\begin{aligned}
\mathbb{Q}_f(N^{k'}[-A,0]\le aA\mu_{k'}\ \forall k') &\ge 1-\sum_{k'}\mathbb{Q}_f\left(N^{k'}[-A,0]>aA\mu_{k'}\right)\\
&\ge 1-\sum_{k'}\frac{\mu_{k'}A}{aA\mu_{k'}} = 1-\frac{K}{a}=\frac{1}{2},
\end{aligned}$$

with $a=2K$. Finally,

$$I_k \ge \frac{1}{2}C(f_0)\exp(-KA)\left(1-\frac{A\tau K^2}{1-\tau K}\right)\|h_{k,\ell}-h^0_{k,\ell}\|_1$$

and using (2.10),

$$\begin{aligned}
\mathbb{E}_f[Z_{m,\ell}] &\ge \frac{T}{2J_T}\sum_k I_k\\
&\ge \frac{T}{2J_T}\frac{1}{2}C(f_0)\exp(-KA)\left(1-\frac{A\tau K^2}{1-\tau K}\right)\sum_k\|h_{k,\ell}-h^0_{k,\ell}\|_1\\
&\ge C\frac{T}{J_T}\|f-f^0\|_1,
\end{aligned}$$

where C depends on f_0.

- We now assume that there exists ℓ such that

$$\left|\nu_\ell-\nu^0_\ell\right| \ge \tau\|f-f_0\|_1.$$

In this case, using similar arguments, still with $a = 2K$,

$$\begin{aligned}
\mathbb{E}_f[Z_{m,\ell}] &\geq \frac{T}{2J_T}\mathbb{P}_f[\{\max_k N^k[0,A]=0\}]\left|\nu_\ell - \nu_\ell^0\right| \\
&\geq \frac{\tau T}{2J_T}\|f-f_0\|_1 \mathbb{E}_{\mathbb{Q}_f}\left[\mathcal{L}_A(f)\mathbb{1}_{\{\max_k N^k[0,A]=0\}}\right] \\
&\geq \frac{\tau T}{2J_T}\|f-f_0\|_1 \mathbb{E}_{\mathbb{Q}_f}\left[\mathcal{L}_A(f)\mathbb{1}_{\{\max_k N^k[0,A]=0\}}\mathbb{1}_{\{N^k[-A,0]\leq aA\mu_k,\forall k\}}\right] \\
&\geq \frac{\tau T}{2J_T}\|f-f_0\|_1 \exp\left(KA - A\sum_{k'}\nu_{k'} - aA\sum_{k''}\mu_{k''}\sum_{k'}\rho_{k''k'}\right)\mathbb{E}_{\mathbb{Q}_f}\left[\mathbb{1}_{\{N[0,A]=0\}}\mathbb{1}_{\{\forall k\, N^k[-A,0]\leq aA\mu_k\}}\right] \\
&\geq \frac{\tau T}{4J_T}\|f-f_0\|_1 \exp\left(-A\sum_{k'}\nu_{k'} - aA\sum_{k'}(\mu_{k'}-\nu_{k'})\right) \geq C\frac{T}{J_T}\|f-f_0\|_1
\end{aligned}$$

for C depending on f_0. Lemma 4 is proved.

$\square$

2.5.4. *Upper bound for the Laplace transform of the number of points in a cluster.* In the next lemma, we refine the proof of Lemma 1 of [4]. Given an ancestor of type ℓ, we denote W^ℓ the number of points in its cluster. We have the following result.

LEMMA 5. *Assume $\|\rho\| < 1$ and consider t such that $0 \leq t \leq \frac{1-\|\rho\|}{2\sqrt{K}}\log\left(\frac{1+\|\rho\|}{2\|\rho\|}\right)$. Then, we have for any $\ell \in \{1,\ldots,K\}$,*

$$\mathbb{E}_f[\exp(tW^\ell)] \leq \frac{1+\|\rho\|}{2\|\rho\|}.$$

Moreover, if $\|\rho\| \leq 1/2$, then there exist two absolute constants c_0 and C_0 such that if $\sqrt{K}t \leq c_0$, then $\mathbb{E}_f[\exp(tW^\ell)] \leq C_0$. Finally,

$$\mathbb{E}_f[W^\ell] = \mathbb{1}^T(I-\rho^T)^{-1}\mathbf{e}_\ell.$$

PROOF OF LEMMA 5. We introduce $K^\ell(n) \in \mathbb{R}^K$ the vector of the number of descendants of the nth generation from a single ancestral point of type ℓ, with $K^\ell(0) = \mathbf{e}_\ell$, where $(\mathbf{e}_\ell)_k = \mathbb{1}_{k=\ell}$. More precisely, $(K^\ell(n))_k$ is the number of descendants of the nth generation and of the type k from a single ancestral point of type ℓ. Then,

$$W^\ell = \mathbb{1}^T \times \sum_{n=0}^{\infty} K^\ell(n).$$

We now set for any $\theta \in \mathbb{R}^K$,

$$\phi_\ell(\theta) = \log\left(\mathbb{E}_f[\exp(\theta^T K^\ell(1))]\right)$$

and

$$\phi(\theta) = (\phi_1(\theta),\ldots,\phi_K(\theta))^T.$$

Note that

$$K^\ell(1)_j \sim \mathcal{P}\left(\rho_{\ell,j}\right), \quad \forall j \leq K$$

and

$$\phi_\ell(\theta) = \sum_{j=1}^K \log\left(\mathbb{E}_f[\exp(\theta_j K^\ell(1)_j)]\right) = \sum_{j=1}^K \rho_{\ell,j}(\exp(\theta_j) - 1).$$

Therefore,

$$(D\phi(\theta))_{\ell,j} := \frac{\partial \phi_\ell(\theta)}{\partial \theta_j} = \rho_{\ell,j}\exp(\theta_j)$$

and for any $x \in \mathbb{R}^K$, since $\|\rho\| := \sup_{x,\|x\|_2=1}\|\rho x\|_2$,

$$\begin{aligned}
\|D\phi(\theta)x\|_2^2 &= \sum_{\ell=1}^K \left(\sum_{j=1}^K \rho_{\ell,j}\exp(\theta_j)x_j\right)^2 \\
&= \sum_j \sum_{j'} (\rho^T\rho)_{j,j'}\exp(\theta_j)x_j\exp(\theta_{j'})x_{j'} \\
&= v^T\rho^T\rho v \\
&\leq \|\rho\|^2\|v\|_2^2 = \|\rho\|^2\sum_{j=1}^K x_j^2\exp(2\theta_j)
\end{aligned}$$

with v the vector of $\mathbb{R}^K$ such that $v_j = \exp(\theta_j)x_j$. So,

$$|\|D\phi(\theta)\|| \quad \leq \quad \|\rho\|\max_j \exp(|\theta_j|) \leq \|\rho\|e^{\|\theta\|_2}.$$

So, by applying the mean value theorem,

$$\|\phi(\theta)\|_2 = \|\phi(\theta) - \phi(0)\|_2 \leq \|\rho\|e^{\|\theta\|_2}\|\theta\|_2.$$

We use a modification of the arguments in the proof of Lemma 1 of [4]. Writing $g_1(\theta) = \theta + \phi(\theta)$, we have for $n \geq 3$

$$\begin{aligned}
\mathbb{E}_f\left[e^{\theta^T(\sum_{k=0}^n K^\ell(k))}\right] &= \mathbb{E}_f\left[e^{\theta^T(\sum_{k=0}^{n-1} K^\ell(k))}\mathbb{E}_f\left[e^{\theta^T K^\ell(n)}|K^\ell(n-1),\dots,K^\ell(1)\right]\right] \\
&= \mathbb{E}_f\left[e^{\theta^T(\sum_{k=0}^{n-2} K^\ell(k))}e^{(\theta+\phi(\theta))^T K^\ell(n-1)}\right] = \mathbb{E}_f\left[e^{\theta^T(\sum_{k=0}^{n-2} K^\ell(k))}e^{g_1(\theta)^T K^\ell(n-1)}\right] \\
&= \mathbb{E}_f\left[e^{\theta^T(\sum_{k=0}^{n-3} K^\ell(k))}e^{(\theta+\phi(g_1(\theta)))^T K^\ell(n-2)}\right] = \mathbb{E}_f\left[e^{\theta^T(\sum_{k=0}^{n-3} K^\ell(k))}e^{g_2(\theta)^T K^\ell(n-2)}\right] \\
&= \mathbb{E}_f\left[e^{\theta^T K^\ell(0)}e^{g_{n-1}(\theta)^T K^\ell(1)}\right] = e^{(g_n(\theta)_\ell)},
\end{aligned}$$

with the induction formula: $g_n(\theta) = \theta + \phi(g_{n-1}(\theta))$ for $n \geq 2$. In particular,

$$\|g_1(\theta)\|_2 \leq \|\theta\|_2(1 + \|\rho\|e^{\|\theta\|_2}) \quad \text{and} \quad \|g_n(\theta)\|_2 \leq \|\theta\|_2 + \|\rho\|e^{\|g_{n-1}(\theta)\|_2}\|g_{n-1}(\theta)\|_2.$$

We now set $C := (1+\|\rho\|)/(1-\|\rho\|) > 1$. Then, if $\|g_{n-1}(\theta)\|_2 \leq \|\theta\|_2(1+C)$,

$$\|g_n(\theta)\|_2 \leq \|\theta\|_2(1+\|\rho\|(1+C)e^{\|\theta\|_2(1+C)}) \leq \|\theta\|_2(1+C)$$

as soon as

$$\|\theta\|_2 \leq (1+C)^{-1}\log(C/(\|\rho\|(1+C))) = \frac{1-\|\rho\|}{2}\log\left(\frac{1+\|\rho\|}{2\|\rho\|}\right). \tag{2.11}$$

Since $\|\rho\| < 1$, the previous upper bound is positive. Note that under (2.11), $\|\theta\|_2 \leq \log(C/\|\rho\|)$, and

$$\|g_1(\theta)\|_2 \leq \|\theta\|_2(1+\|\rho\|e^{\|\theta\|_2}) \leq \|\theta\|_2(1+\|\rho\|e^{\log(C/\|\rho\|)}) \leq \|\theta\|_2(1+C).$$

We finally obtain that under (2.11),

$$\|g_n(\theta)\|_2 \leq \|\theta\|_2(1+C), \quad \forall\, n \geq 1.$$

Since for any m, $n \mapsto \sum_{k=0}^n (K^\ell(k))_m$ is increasing and $W^\ell = \mathbb{1}^T \times \sum_{n=0}^\infty K^\ell(n)$, we have by monotone convergence that for $t > 0$,

$$\mathbb{E}_f[\exp(tW^\ell)] = \lim_{n\to\infty} \exp(g_n(t\mathbf{1})_\ell).$$

By the previous result, the right hand side is bounded if t is small enough. More precisely, for all $0 < t \leq (1+C)^{-1}\log(C/(\|\rho\|(1+C)))/\sqrt{K}$,

$$\mathbb{E}_f[\exp(tW^\ell)] \leq \exp(t\sqrt{K}(1+C)) \leq \frac{C}{\|\rho\|(1+C)} = \frac{1+\|\rho\|}{2\|\rho\|}.$$

The second point is obvious in view of previous computations. Moreover, since $\mathbb{E}_f[W^\ell] = \sum_{n=0}^\infty \mathbb{E}_f[\mathbb{1}^T K^\ell(n)]$ and since for any $v \in \mathbb{R}^K$

$$\mathbb{E}_f[v^T K^\ell(n) | K^\ell(0), \dots, K^\ell(n-1)] = \sum_{j=1}^K \sum_{k=1}^K K^\ell(n-1)_j v_k \rho_{j,k} = v^T \rho^T K^\ell(n-1).$$

We obtain by induction that $\mathbb{E}_f[\mathbb{1}^T K^\ell(n)] = \mathbb{1}^T (\rho^T)^n \mathbf{e}_\ell$ and taking the limit, since $\|\rho\| < 1$,

$$\mathbb{E}_f[W^\ell] = \mathbb{1}^T (I - \rho^T)^{-1} \mathbf{e}_\ell.$$

□

2.5.5. *Lemma on $\bar{N}^m$.*

LEMMA 6. *There exists $\tilde{c}_0$ such that for all $c_0 > 0$ such that for T large enough,*

$$\mathbb{P}_0\left(\sum_{m=1}^{J_T-1} \bar{N}^m(I_m) > c_0 T\right) \leq e^{-\tilde{c}_0 c_0 T}.$$

Furthermore, there exists a constant $\kappa_0 > 0$ (see the definition of J_T) such that for any $f \in \mathcal{F}_T$,

$$\sum_{m=1}^{J_T-1} \mathbb{E}_f[\bar{N}^m(I_m)] = o(T).$$

Proof of Lemma 6. We use computations of the proof of Proposition 2 of Hansen et al. [4]. To bound $\bar{N}^m(I_m)$, first observe that we only consider points of N whose ancestors are born before $(2m-1)T/(2J_T)$, i.e. the distance between the occurrence of an ancestor and I_m is at least $2mT/(2J_T) - A - (2m-1)T/(2J_T) = T/(2J_T) - A$ since

$$I_m = \left[\frac{2mT}{2J_T} - A, \frac{(2m+1)T}{2J_T}\right].$$

Using the cluster representations of the process, for any $p \in \mathbb{Z}$ and for any $\ell \in \{1, \dots, K\}$, we consider $B_{p,\ell}$ the number of ancestors of type ℓ born in the interval $[p, p+1]$. The $B_{p,\ell}$'s are iid Poisson random variables with parameter ν_ℓ. We have

$$\sum_{m=1}^{J_T-1} \bar{N}^m(I_m) \le \sum_{\ell=1}^{K} \sum_{p \in \mathcal{J}_T^+} \sum_{k=1}^{B_{p,\ell}} \left(W_{p,k}^\ell - \frac{1}{A}\left(\frac{T}{2J_T} - A\right)\right)_+ + \sum_{\ell=1}^{K} \sum_{p=-\infty}^{0} \sum_{k=1}^{B_{p,\ell}} \left(W_{p,k}^\ell - \frac{1}{A}\left(-p - 1 + \frac{T}{J_T} - A\right)\right)_+,$$

where $W_{p,k}^\ell$ is the number of points in the cluster generated by the ancestor k which is of type ℓ and

$$\mathcal{J}_T^+ = \{p : \ 1 \le p \le T - T/(2J_T)\}$$

since

$$\bigcup_{m=1}^{J_T-1} I_m \subset \left[\frac{T}{J_T} - A, T - \frac{T}{2J_T}\right].$$

For the first term of the previous right hand side, we have used same arguments as Hansen et al. [4] and the lower bound of the distance determined previously. For the second term of the right hand side, since $p \le 0$, this lower bound is at least $-p - 1 + \frac{T}{J_T} - A$. Conditioned on the $B_{p,\ell}$'s, the variables $(W_{p,k}^\ell)_k$ are iid with same distribution as W^ℓ introduced in Lemma 5. Furthermore, by Lemma 5 applied with $f = f_0$, since $\|\rho_0\| < 1$, we know that for $t_0 > 0$ small enough (only depending on $\|\rho_0\|$ and K),

$$\mathbb{E}_0[\exp(t_0 W^\ell)] \le C_0,$$

where C_0 is a constant. So, for any $c > 0$,

$$\begin{aligned}
\mathcal{P}_{T,1} &:= \mathbb{P}_0\left(\sum_{\ell=1}^{K}\sum_{p\in\mathcal{J}_T^+}\sum_{k=1}^{B_{p,\ell}}\left(W_{p,k}^{\ell}-\frac{1}{A}\left(\frac{T}{2J_T}-A\right)\right)_+\geq cT\right)\\
&\leq \exp(-t_0cT)\prod_{\ell=1}^{K}\prod_{p\in\mathcal{J}_T^+}\mathbb{E}_0\left[\prod_{k=1}^{B_{p,\ell}}\mathbb{E}_0\left[\exp\left(t_0\left(W_{p,k}^{\ell}-\frac{T}{2AJ_T}+1\right)_+\right)|B_{p,\ell}\right]\right]\\
&\leq \exp(-t_0cT)\prod_{\ell=1}^{K}\prod_{p\in\mathcal{J}_T^+}\mathbb{E}_0\left[(H_\ell(t_0))^{B_{p,\ell}}\right]=\exp\Big(-t_0cT+\sum_{\ell=1}^{K}\sum_{p\in\mathcal{J}_T^+}\nu_\ell^0(H_\ell(t_0)-1)\Big),
\end{aligned}$$

where

$$H_\ell(t_0) := \mathbb{E}_0\left[\exp\left(t_0\left(W^{\ell}-\frac{T}{2AJ_T}+1\right)_+\right)\right],$$

satisfying

$$\begin{aligned}
H_\ell(t_0) &\leq \mathbb{P}_0\left(W^\ell\leq\frac{T}{2AJ_T}-1\right)+\exp(t_0-Tt_0/(2AJ_T))\mathbb{E}_0\left[\exp\left(t_0W^\ell\right)\right]\\
&\leq 1+C_0\exp(t_0-Tt_0/(2AJ_T)).
\end{aligned}$$

Therefore,

$$\sum_{\ell=1}^{K}\sum_{p\in\mathcal{J}_T^+}\nu_\ell^0(H_\ell(t_0)-1)\lesssim(T-T/(2J_T)\exp(-Tt_0/(2AJ_T))\lesssim e^{-C'\kappa_0\log T}=o(t_0cT)$$

by choosing κ_0 large enough and then

$$\mathcal{P}_{T,1}\lesssim\exp(-t_0cT/2).$$

Similarly,

$$\begin{aligned}
\mathcal{P}_{T,2} &:= \mathbb{P}_0\left(\sum_{\ell=1}^{K}\sum_{p=-\infty}^{0}\sum_{k=1}^{B_{p,\ell}}\left(W_{p,k}^{\ell}-\frac{1}{A}\left(-p-1+\frac{T}{J_T}-A\right)\right)_+\geq cT\right)\\
&\leq \exp\Big(-t_0cT+\sum_{\ell=1}^{K}\sum_{p=-\infty}^{0}\nu_\ell^0(\tilde{H}_{\ell,p}(t_0)-1)\Big),
\end{aligned}$$

where

$$\tilde{H}_{\ell,p}(t_0) := \mathbb{E}_0\left[\exp\left(t_0\left(W^{\ell}-\frac{T}{AJ_T}+1+\frac{1}{A}+\frac{p}{A}\right)_+\right)\right],$$

satisfying

$$\tilde{H}_{\ell,p}(t_0) \leq 1 + C_0 \exp(t_0 + t_0/A - Tt_0/(AJ_T) + t_0 p/A).$$

Therefore,

$$\sum_{\ell=1}^{K} \sum_{p=-\infty}^{0} \nu_\ell^0 (\tilde{H}_{\ell,p}(t_0) - 1) \lesssim \exp(-Tt_0/(AJ_T)) = o(t_0 cT)$$

and then

$$\mathcal{P}_{T,2} \lesssim \exp(-t_0 cT/2).$$

Finally, there exists $\tilde{c}_0$ (only depending on t_0, so only depending on $\|\rho_0\|$ and K) such that for all $c_0 > 0$ such that for T large enough

$$\mathbb{P}_0 \left(\sum_{m=0}^{J_T - 1} \bar{N}^m(I_m) > c_0 T \right) \leq e^{-\tilde{c}_0 c_0 T}$$

and the first part of the lemma is proved.

For the second part, we only consider the case $1/2 \leq \|\rho\| < 1$. The case $\|\rho\| < 1/2$ can be derived easily using following computations. We have:

$$\sum_{m=1}^{J_T - 1} \mathbb{E}_f[\bar{N}^m(I_m)] = \mathcal{E}_{T,1} + \mathcal{E}_{T,2},$$

with

$$\mathcal{E}_{T,1} := \mathbb{E}_f \left[\sum_{\ell=1}^{K} \sum_{p \in \mathcal{J}_T^+} \sum_{k=1}^{B_{p,\ell}} \left(W_{p,k}^\ell - \frac{1}{A} \left(\frac{T}{2J_T} - A \right) \right)_+ \right]$$

and, with $t = \frac{1-\|\rho\|}{2\sqrt{K}} \log\left(\frac{1+\|\rho\|}{2\|\rho\|}\right) \gtrsim (1-\|\rho\|)^2 \gtrsim u_T^2$ on $\mathcal{F}_T$, since for $x > 0$, $x \le e^x$, by using Lemma 5,

$$\begin{aligned}
\mathcal{E}_{T,2} &:= \mathbb{E}_f\left[\sum_{\ell=1}^{K}\sum_{p=-\infty}^{0}\sum_{k=1}^{B_{p,\ell}}\left(W_{p,k}^\ell - \frac{1}{A}\left(-p-1+\frac{T}{J_T}-A\right)\right)_+\right] \\
&= \sum_{\ell=1}^{K}\nu_\ell\sum_{p=-\infty}^{0}\mathbb{E}_f\left[\left(W_{p,k}^\ell - \frac{1}{A}\left(-p-1+\frac{T}{J_T}-A\right)\right)_+\right] \\
&= t^{-1}\sum_{\ell=1}^{K}\nu_\ell\sum_{p=-\infty}^{0}\mathbb{E}_f\left[e^{t\left(W_{p,k}^\ell - \frac{1}{A}\left(-p-1+\frac{T}{J_T}-A\right)\right)}\right] \\
&\lesssim t^{-1}e^{-\frac{tT}{AJ_T}}(1-e^{-\frac{t}{A}})^{-1}\mathbb{E}_f\left[e^{tW^\ell}\right]\sum_{\ell=1}^{K}\nu_\ell \\
&\lesssim (1-\|\rho\|)^{-4}e^{-\frac{(1-\|\rho\|)^2T}{AJ_T}}\sum_{\ell=1}^{K}\nu_\ell \lesssim e^{-\kappa_0^{-1}C''\log T}(\log T)^{-2/3}\varepsilon_T^{-4/3},
\end{aligned}$$

for C'' depending on A and K. Similarly,

$$\mathcal{E}_{T,1} \lesssim T(1-\|\rho\|)^{-2}e^{-\frac{(1-\|\rho\|)^2T}{2AJ_T}}\sum_{\ell=1}^{K}\nu_\ell.$$

Choosing κ_0 small enough,

$$\sum_{m=1}^{J_T-1}\mathbb{E}_f[\bar{N}^m(I_m)] = o(T).$$

□

2.6. *Proofs of results of Section 2.3 of [1].* This section is devoted to the proofs of results of Section 2.3 of [1].

2.6.1. *Proof of Corollary 3.* The main difference with the case of the regular partition is the control of the $\mathbb{L}_1$-entropy. This is more complicated than the regular grid histogram prior and we apply instead Theorem 2 of [1]. Because of the equivalence between the parameterization in t or in u, we sometimes $\bar{h}_{w,t,J}$ as $\bar{h}_{w,u,J}$. Let J and $(\underline{w}, \underline{u})$ and $(\underline{w}', \underline{u}')$ belonging to $\mathcal{S}_J^2$. Then, for all $\zeta > 0$, if $\delta = \delta' = 1$, $|t_j' - t_j| \le \zeta\epsilon_T \min(|t_j - t_{j-1}|, |t_j - t_{j+1}|)$ for all j and $\sum_j |w_j - w_j'| \le \epsilon_T$ then

$$\begin{aligned}
\|\bar{h}_{w,t,J} - \bar{h}_{w',t',J}\|_1 &\le \|\bar{h}_{w,t,J} - \bar{h}_{w',t,J}\|_1 + \|\bar{h}_{w',t,J} - \bar{h}_{w',t'J}\|_1 \\
&\le \sum_{j=1}^{J}|w_j - w_j'| + 4\sum_{j=1}^{J}\zeta\epsilon_T w_j'.
\end{aligned}$$

Consider $e_T > 0$ and $\mathcal{U}_{J,T} = \{\underline{u} \in \mathcal{S}_J, \min_j u_j \geq e_T\}$, under the Dirichlet prior on $\underline{u}$

$$\Pi_u(\mathcal{U}_{J,T}^c|J) \leq \sum_{j\leq J} \Pi(u_j \leq e_T) = \sum_{j=1}^J Prob\,(\text{ Beta}(\alpha,(J-1)\alpha) \leq e_T) \lesssim Je_T^\alpha \leq e^{-cT\epsilon_T^2}$$

if $\log e_T \leq -(c/\alpha+1)T\epsilon_T^2$ if $J \leq J_1(T/\log T)^{1/(2\beta+1)} =: J_{1,T}$. We define $\overline{\mathcal{H}_T} = \{\bar{h}_{w,u,J}, J \leq J_{1,T}; \underline{u} \in \mathcal{U}_{J,T}\}$. To apply Theorem 2 of [1], we need to construct the slices $\mathcal{H}_{T,i}$ of $\overline{\mathcal{H}_T}$. Let $e_{T,\ell} = e_T^{1/\ell}$ for $1 \leq \ell \leq L = \log(e_T)/\log\tau$ and $0 < \tau < 1$ is fixed and $e_{T,L+1} = 1$. Without loss of generality we can assume that $\log(e_T)/\log\tau \in \mathbb{N}$. For $(u_1,\cdots,u_J)$ let k_i be defined by $u_i \in (e_{T,k_i}, e_{T,k_i+1})$ and $(N_1,\cdots,N_L)$ be given by $\text{card}\{j, u_j \in (e_{T,\ell}, e_{T,\ell+1})\} = N_\ell$ so that $\sum_\ell N_\ell = J$ and consider a configuration $\sigma = (k_1,\cdots,k_J)$; denote by $\mathcal{U}_{J,T}(\sigma)$ the set of $\underline{u} \in \mathcal{S}_J$ satisfying the configuration σ, we define $\mathcal{H}_{T,\sigma,J} = \{(\underline{w},\underline{u}) \in \mathcal{S}_J \times \mathcal{U}_{J,T}(\sigma)\}$ and $\mathcal{H}_{T,\sigma}$ the collection of $\mathcal{H}_{T,\sigma,J}$ with $J \leq J_{1,T}$. We have, by symmetry for all $\sigma = (k_1,\cdots,k_J)$ compatible with $(N_1,\cdots,N_L)$ writing $\bar{N}_\ell = N_1 + \cdots + N_\ell$

$$\begin{aligned}
\Pi_J\left(\mathcal{U}_{J,T}(\sigma)\right) &= \Pi_J\left(\cap_{\ell=1}^L\{(u_{\bar{N}_{\ell-1}+1},\cdots,u_{\bar{N}_\ell}) \in (e_{T,\ell-1},e_{T,\ell})^{N_\ell}\}\right)\\
&\leq \frac{\Gamma(\alpha J)}{\Gamma(\alpha)^J}\prod_{\ell=1}^L e_T^{(\alpha-1)/(\ell+1)}\text{Vol}\left(\cap_{\ell=1}^L\{(u_{\bar{N}_{\ell-1}+1},\cdots,u_{\bar{N}_\ell}) \in (e_{T,\ell-1},e_{T,\ell})^{N_\ell}\}\right)\\
&\leq \frac{\Gamma(\alpha J)}{\Gamma(\alpha)^J}\prod_{\ell=1}^{L-1} e_T^{(\alpha-1)N_\ell/(\ell+1)} e_T^{N_\ell/(\ell+1)}
\end{aligned}$$

We now construct a net $(\underline{u}^{(j)}, j \leq N_{\sigma,J})$ such that for all $\underline{u} \in \mathcal{U}_{J,T}(\sigma)$ there exists $\underline{u}^{(j)}$ satisfying $|t_i - t_i^{(j)}| \leq \epsilon_T u_i^{(j)} \wedge u_{i+1}^{(j)}$ for all i, with $t_i = \sum_{\ell=1}^i u_\ell$. If $|t_i - t_i^{(j)}| \leq \epsilon_T e_{T,k_i} \wedge e_{T,k_{i+1}}$ then $|t_i - t_i^{(j)}| \leq \epsilon_T u_i^{(j)} \wedge u_{i+1}^{(j)}$. Therefore, given a configuration $(k_1,\cdots,k_J)$ compatible with $(N_1,\cdots,N_L)$, we can cover $\mathcal{U}_{J,T}(\sigma)$ using

$$N_J(\sigma) \leq \prod_{i=1}^J e_T^{1/(k_i+1)-1/(k_i\wedge k_{i+1})} \leq \prod_{\ell=1}^L e_{T,\ell+1}^{N_\ell} e_{T,\ell}^{-2N_\ell}.$$

The covering number of $\mathcal{S}_J$ by balls of radius $\zeta\epsilon_T$ is bounded by $\left(\frac{1}{\zeta\epsilon_T}\right)^J$ and

$$
\begin{aligned}
I_T &:= \sum_J \sqrt{\Pi(J)}\mathcal{N}(\epsilon_T, \mathcal{H}_{T,\sigma,J}) \\
&\leq \sum_J \sqrt{\Pi(J)} \left(\frac{1}{\zeta\epsilon_T}\right)^J \sum_\sigma N_J(\sigma)\sqrt{\Pi_j(\mathcal{U}_{J,T}(\sigma))} \\
&\lesssim \sum_J \sqrt{\Pi(J)} \left(\frac{1}{\zeta\epsilon_T}\right)^J \sum_{(N_1,\cdots,N_L)} \frac{J!\Gamma(\alpha J)}{\Gamma(\alpha)^J N_1!\cdots N_L!} \exp\left[\log e_T \sum_{\ell=1}^{L-1} N_\ell \left(\frac{\alpha+2}{2(\ell+1)} - \frac{2}{\ell}\right)\right] e_T^{-2N_L/L} \\
&\lesssim \left(\frac{1}{\zeta\epsilon_T}\right)^{J_{1,T}} e^{2\alpha J_{1,T}\log J_{1,T}} \sum_{J=1}^{J_{1,T}} \sum_{(N_1,\cdots,N_L)} \frac{J!}{N_1!\cdots N_L!} \prod_{\ell=1}^{L} p_\ell^{N_\ell} \prod_{\ell=1}^{L-1} \frac{e_T^{N_\ell\left(\frac{\alpha+2}{2(\ell+1)}-\frac{2}{\ell}\right)}}{p_\ell} \frac{e_T^{-2N_L/L}}{p_L p_{L-1}}
\end{aligned}
$$

for any $p_1, \cdots, p_L \geq 0$ with $\sum_{\ell=1}^{L+1} p_\ell = 1$. Taking $p_L = 1/(L+1)$ and since $\alpha \geq 6$, $\frac{\alpha+2}{2(\ell+1)} - \frac{2}{\ell} \geq 0$ for all $\ell \geq 1$, leading to

$$
I_T \lesssim \tau^{-2J_{1,T}} \left(\frac{1}{\zeta\epsilon_T}\right)^{J_{1,T}} e^{2\alpha J_{1,T}\log J_{1,T} + (L+1)\log(L+1)} \sum_{J=1}^{J_{1,T}} \sum_{(N_1,\cdots,N_L)} \frac{J!}{N_1!\cdots N_L!} \prod_{\ell=1}^{L} p_\ell^{N_\ell} \lesssim e^{KJ_{1,T}\log T}
$$

for some $K > 0$ and condition (2.3) of [1] is verified.

2.6.2. *Proof of Corollary 4.* The proof is based on Rousseau [6], where mixtures of Beta densities are studied for density estimation, and using Theorem 2 of [1]. Note that for all h_1, h_2

$$
|(h_1(x))_+ - (h_2(x))_+| \leq |h_1(x) - h_2(x)|
$$

so that Corollary 4 of [1] is proved by studying

$$
\tilde{B}(\epsilon_T, B) = \left\{ (\nu_k, (g_{\ell,k})_\ell)_k; \quad \max_k |\nu_k - \nu_k^0| \leq \epsilon_T,\ \max_{\ell,k} \|g_{\ell,k} - g_{\ell,k}^0\|_2 \leq \epsilon_T,\ \max_{\ell,k} \|g_{\ell,k}\|_\infty \leq B \right\}
$$

in the place of $B(\epsilon_T, B)$ and by controlling the $\mathbb{L}_1$-entropy associated to

$$
\mathcal{G}_{1,T} = \left\{ g_{\alpha,P}; \quad P = \sum_{j=1}^{J} p_j \delta_{\epsilon_j},\ \epsilon_j \in [e_1, 1-e_1];\ \alpha \in [\alpha_{0T}, \alpha_{1T}];\ \sum_j |p_j| = 1,\ J \leq J_{1,T} \right\}
$$

where

$$
e_1 = e^{-a_0 T\epsilon_T^2}, \quad \alpha_{0T} = \exp\left(-Tc_0\epsilon_T^2\right); \quad \alpha_{1T} = \alpha_1 T^2\epsilon_T^4, \quad J_{1,T} = J_1 T^{1/(2\beta+1)}(\log T)^{(\beta-2)/(4\beta+2)},
$$

with $c_0, \alpha_1, a_0, J_1 > 0$ and $g_{\alpha,P} = \int_0^1 g_{\alpha,\epsilon} dP(\epsilon)$. From the proof of Theorem 2.1 in Rousseau [6], we have that for all $c_2 > 0$ we can choose $a_0, c_0, \alpha_1 > 0$ such that $\Pi\left(\mathcal{G}_{1,T}^c\right) \leq e^{-c_2 T \epsilon_T^2}$ and $\mathcal{G}_{1,T}$ can be cut into the following slices: we group the components into the intervals $[e_\ell, e_{\ell+1}]$ or $[1 - e_{\ell+1}, 1 - e_\ell]$ with $e_\ell = e_0^{1/\ell}$ and $e_{L_T} = T^{-t}$, for some $t > 0$, and the interval $[e_{L_T}, 1 - e_{L_T}]$. For each of these intervals we denote $N(\ell)$ the number of components which fall into the said interval, $N(\ell) = \sum_{i=1}^J \mathbb{1}_{\epsilon_i \in (e_\ell, e_{\ell+1}) \cup (1-e_{\ell+1}, 1-e_\ell)}$ if $\ell \leq J_T$, and $N(L_T+1) = \sum_{i=1}^k \mathbb{1}_{\epsilon_i \in (e_{L_T}, 1-e_{L_T})}$. Let $J \leq J_{1,T}$ $\mathcal{G}_{1,\sigma}(J) = \{g_{\alpha,P} \in \mathcal{G}_{1,T};\, N(\ell) = k_\ell,\, \sum_{\ell=1}^{L_T+1} k_\ell = J\}$ with σ denoting the configuration $(k_1, \cdots, k_{L_T+1})$. From Rousseau [6] Section 4.1, for all $\zeta > 0$, we have

$$\mathcal{N}(\zeta\epsilon_T, \mathcal{G}_{1,\sigma}(J), \|\cdot\|_1) \leq (\zeta\epsilon_T)^{-k_{L_T}} \prod_{\ell=1}^{L_T-1} \left(\frac{(\log e_{\ell+1} - \log e_\ell)}{\zeta\epsilon_T e_\ell} \right)^{k_\ell} \leq (\zeta\epsilon_T)^{-k_{L_T}} \prod_{\ell=1}^{L_T-1} \left(\frac{\log(1/e_1)}{\ell(\ell+1)\zeta\epsilon_T e_1^{1/\ell}} \right)^{k_\ell}$$

and

$$\sqrt{\Pi(\mathcal{G}_{1,\sigma})(J)} \leq \sqrt{\Pi_J(J)} \frac{\Gamma(J+1)^{1/2}}{\prod_{\ell=1}^{L_T+1} \Gamma(k_\ell)^{1/2}} \prod_{\ell=1}^{L_T} p_{T,\ell}^{k_\ell/2}, \quad p_{T,\ell} \leq c(e_{\ell+1}^{a+1} - e_\ell^{a+1}), \quad \ell \leq L_T - 1$$

and $p_{T,L_T} \leq 1$. Since $e_{\ell+1}^{a+1} - e_\ell^{a+1} \leq e_{\ell+1}^{a+1} \leq e_1^{(a+1)/(\ell+1)}$ and since $J! \geq \prod_{\ell=1}^{J_T+1} k_\ell!$, we obtain

$$\begin{aligned} \sum_{J \leq J_{1,T}} \sum_\sigma \mathcal{N}(\zeta\epsilon_T, \mathcal{G}_{1,\sigma}(J), \|\cdot\|_1) \sqrt{\Pi(\mathcal{G}_{1,\sigma}(J))} &\lesssim J_{1,T} e^{C J_{1,T} \log T} \sum_\sigma \frac{J!}{\prod_{\ell=1}^{L_T+1} k_\ell!} \prod_{\ell=1}^{L_T+1} \left(\frac{\bar{c}}{\ell(\ell+1)} \right)^{k_\ell} \\ &= J_{1,T} e^{C J_{1,T} \log T} \end{aligned}$$

as soon as $a \geq 3$, where $\bar{c}^{-1} = \sum_{\ell=1}^{L_T+1} 1/(\ell(\ell+1))$. Therefore condition (2.3) of [1] is verified. We now study the Kullback-Leibler condition (i). Again, we use Theorem 3.1 in Rousseau [6], so that for all $f_0 \in \mathcal{H}(\beta, L)$ and all $\beta > 0$ there exists f_1 such that $\|f_0 - g_{\alpha,f_1}\|_\infty \lesssim \alpha^{-\beta/2}$, when α is large enough and $g_{\alpha,f_1} = \int_0^1 g_{\alpha,\epsilon} f_1(\epsilon) d\epsilon$, and where f_1 is either equal to f_0 if $\beta \leq 2$ or $f_1 = f_0 \sum_{j=1}^{\lceil\beta\rceil-1} w_j/\alpha^{j/2}$, with w_j a polynomial function with coefficients depending on $f_0^{(l)}$ $l \leq j$. From that, we construct a finite mixture approximation of g_{α,f_1}. Note that even if f_0 is positive, f_1 is not necessarily so. Hence to use the convexity argument of Lemma A1 of Ghosal and van der Vaart [3] we write f_1 as $m_+ f_{1,+} - m_- f_{1,-}$ with $f_{1,+}, f_{1,-} \geq 0$ and probability densities. In the case where $m_- = 0$ then $f_{1,-} = 0$. We approximate $g_{\alpha,f_{1,+}}$ and $g_{\alpha,f_{1,-}}$ separately. Contrarywise to what happens in Rousseau [6], here we want to allow f_0 to be null in some sub-intervals of $[0, 1]$. Hence we adapt the proof of Theorem 3.2 of Rousseau [6] to this set up. Let f be a probability density on $[0, 1]$ we construct a discrete approximation of $g_{\alpha,f}$. Let $\epsilon_0 = \alpha^{-H_0}$ for some $H_0 > 0$ and define $\epsilon_j = \epsilon_0(1 + B\sqrt{\log\alpha/\alpha})^j$ for $j = 1, \cdots, J_\alpha$ with $J_\alpha = O(\sqrt{\alpha\log\alpha})$ and $B > 0$ a constant. We then have, from Lemma 7 below that there exists a signed measure P_0 with at most $N = O(\sqrt{\alpha}(\log\alpha)^{3/2})$ supporting points on $[\epsilon_1, 1 - \epsilon_1]$, such that

$$\|g_{\alpha,P_0} - f_0\|_2 \leq \|g_{\alpha,P_0} - g_{\alpha,f_1}\|_2 + \|g_{\alpha,f_1} - f_0\|_\infty \lesssim \alpha^{-\beta/2}; \quad \|g_{\alpha,P_0}\|_\infty \leq \|f_0\|_\infty + o(1), \quad P_0 = \sum_{i=1}^N p_i \delta_{\epsilon_i}.$$

As in Rousseau [6] Theorem 3.2, we can assume that $|p_i| \geq \alpha^{-A}$ for some fixed A large enough. Following from Section 4.1 of Rousseau [6], There exists $A' > 0$ such that if P satisfies $\max_i |P(U_i) - p_i| \leq \alpha^{-A'}|p_i|$, with $U_i = [\epsilon_i(1-\epsilon_i)(1-\alpha^{-A'}), \epsilon_i(1-\epsilon_i)(1+\alpha^{-A'})]$ then

$$\|g_{\alpha,P_0} - g_{\alpha,P}\|_2 \leq \alpha^{-\beta/2}, \quad \|g_{\alpha,P}\|_\infty \leq \|f_0\|_\infty + o(1).$$

As in Rousseau [6], if $\epsilon_T = \epsilon_0 T^{-\beta/(2\beta+1)}(\log T)^{5\beta/(4\beta+2)}$, then

$$\Pi\left(\tilde{B}(\epsilon_T, \|f_0\|_\infty + 1)\right) \geq e^{-c_1 T\epsilon_T^2}$$

for some $c_1 > 0$, which terminates the proof of Corollary 4.

Lemma 7. *Assume that f is a bounded probability density on $[0,1]$, then for all $B_0 > 0$ there exists $\tilde{N}_0 > 0$ and a signed measure P_0 with at most $N \leq \tilde{N}_0\sqrt{\alpha}(\log\alpha)^{3/2}$ on $[\epsilon_1, 1-\epsilon_1]$ such that*

$$\|g_{\alpha,f} - g_{\alpha,P_0}\|_2 \lesssim \alpha^{-B_0}, \quad \|g_{\alpha,P_0}\|_\infty \leq C,$$

where C is a constant depending on $\|f_0\|_\infty$.

Proof of Lemma 7. On each of the intervals $(\epsilon_{j-1}, \epsilon_j)$ we construct a probability P_j having support on $(\epsilon_{j-1}, \epsilon_j)$ with cardinality smaller than $N_j \leq N_0 \log\alpha$ and such that

$$\|g_{\alpha,f_j} - g_{\alpha,P_j}\|_2^2 \lesssim \alpha^{-B_0}, \quad f_j = \frac{f\mathbb{1}_{(\epsilon_{j-1},\epsilon_j)}}{\int_{\epsilon_{j-1}}^{\epsilon_j} f(\epsilon)d\epsilon} \tag{2.12}$$

where B_0 can be chosen arbitrarily large by choosing N_0 large enough. To prove (2.12) we use the same ideas as in the proof of Theorem 3.2 of [6]. For all $j = 2, \cdots, J-2$ on $(\epsilon_{j-1}, \epsilon_j)$, there exists P_j with at most $N_1 \log\alpha$ terms such that if $x \in [0,1]$,

$$\left|g_{\alpha,f_j} - g_{\alpha,P_j}\right|(x) \leq \frac{\alpha^{-H}}{x(1-x)}$$

where H can be chosen as large as need be, by choosing N_1 large enough. Moreover, let $x \leq \epsilon_0$ or $x > 1-\epsilon_0$, then for all $\epsilon \in (\epsilon_1, 1-\epsilon_1)$, if $x < \epsilon_0$ then $x/\epsilon \leq \delta_\alpha = (1 + B\sqrt{\log\alpha/\alpha})^{-1}$ and

$$g_{\alpha,\epsilon}(x) \lesssim \sqrt{\alpha}\exp\left(\alpha\left[\frac{\log(x/\epsilon)}{1-\epsilon} - (\log x)/\alpha + \frac{\log((1-x)/(1-\epsilon))}{\epsilon}\right]\right)$$

If $\epsilon_1 \leq \epsilon < 1/4$ then the function $\epsilon \to \frac{\log(\epsilon/x)}{1-\epsilon} - \log(\epsilon/x)/\alpha + \frac{\log((1-\epsilon)/(1-x))}{\epsilon}$ is increasing and

$$\begin{aligned} g_{\alpha,\epsilon}(x) &\lesssim \frac{\sqrt{\alpha}}{\epsilon}\exp\left(\alpha\left[\log(\delta_\alpha)\left(1 + x\delta_\alpha + \delta_\alpha^2 x^2) + O(x^3)\right) - 1 + \delta_\alpha^{-1}\right]\right) \\ &\lesssim \alpha^{-B^2/3+H_0} \lesssim \alpha^{-B^2/4}, \end{aligned}$$

by choosing $B^2 \geq 12H_0$. The same reasoning can be applied to $x > 1-\epsilon_0$, which terminates the proof. $\square$

References.

[1] Donnet, S., Rivoirard, V., and Rousseau, J. (2019). Nonparametric bayesian estimation for multivariate hawkes processes. Technical report.

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

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

[4] Hansen, N. R., Reynaud-Bouret, P., and Rivoirard, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143.

[5] Jacobsen, M. (2006). *Point process theory and applications*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA. Marked point and piecewise deterministic processes.

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

Address of Sophie Donnet
UMR MIA Paris AgroParisTech/INRA
Team MORSE
16 rue Claude Bernard
75 231 Paris Cedex 05
France
E-mail: sophie.donnet@agroparistech.fr

Address of Vincent Rivoirard
CEREMADE
CNRS, UMR 7534
Université Paris-Dauphine
PSL University
75016 Paris
France
E-mail: Vincent.Rivoirard@dauphine.fr

Address of Judith Rousseau
Department of Statistics
University of Oxford
24-29 St Giles'
Oxford, OX1 3LB
United Kingdom
E-mail: Judith.Rousseau@stats.ox.ac.uk