Submitted to Bernoulli

Supplementary material of Bayesian estimation of nonlinear Hawkes processes

DÉBORAH SULEM [1], VINCENT RIVOIRARD [2,†] and JUDITH ROUSSEAU [3,*]

[1]*University of Oxford, E-mail:* deborah.sulem@stats.ox.ac.uk; [*]judith.rousseau@stats.ox.ac.uk
[2]*Ceremade, CNRS, UMR 7534, Université Paris-Dauphine, PSL University, 75016 Paris, France.*
E-mail: [†]Vincent.Rivoirard@dauphine.fr
[3]*University of Oxford, E-mail:* deborah.sulem@stats.ox.ac.uk; [*]judith.rousseau@stats.ox.ac.uk
[4]*CEREMADE, Université Paris Dauphine, E-mail:* [†]Vincent.Rivoirard@dauphine.fr

This supplementary material contains additional results and proofs that could not be included in the main paper (Sulem, Rivoirard and Rousseau, 2022) due to space limitations. In Section S1, we prove the second case of Proposition 3.5, our concentration result for the shifted ReLU model. Then in Section S2, we report the proofs of Theorem 5.5 and Lemma A.2. Section S3 contains the proofs of two results in the graph estimation problem (second part of Theorem 3.11 and Proposition 3.10). In Section S4 we prove frequentist results of Corollary 3.8. Results regarding the construction of prior distributions can be found in Section S5. In Sections S6, S7 and S8 we report additional technical results and their proofs, notably on the tests used in the main theorems and on the Kullback-Leibler divergence defined for the Hawkes model. Lemmas A1 and A4 are proved in Section S9. Finally, we report multivariate extensions of existing results on the regenerative properties of the nonlinear Hawkes model in Section S10.

For the sake of simplicity, all sections, theorems, corollaries, lemmas and equations presented in the supplement are designed with a prefix S. Regarding the others, we refer to the material of the main text (Sulem, Rivoirard and Rousseau, 2022). This is not specified at each place.

S1. Proof of Case 2 of Proposition 3.5

We recall that in this case we consider a shifted ReLU model with unknown shift $\theta_0 = (\theta_1^0, \dots, \theta_K^0)$, corresponding to a particular case of partially known link functions $\phi_k(x;\theta_k) = \theta_k + (x)_+$, and for parameter $f \in \mathcal{F}$ and $\theta \in \Theta$, we denote $\lambda_t(f,\theta)$ the intensity process. We note that in this case, $r_0 = \theta_0 + \nu_0$ and similarly $r_f = \theta + \nu$, with r_f defined in (21). We then prove the posterior concentration rate on both f_0 and θ_0. First, we apply the same steps as in the proof of Theorem 3.2 in Section 5.2, replacing $\|f - f_0\|_1$ by $\left\|r_0 - r_f\right\|_1 + \|h - h_0\|_1 = \|\theta_0 + \nu_0 - \theta - \nu\|_1 + \|h - h_0\|_1$. In particular, we re-define the balls w.r.t. the L_1-distance as (for simplicity we keep the same notation)

We therefore obtain (see also Remark 3.7)

$$\mathbb{E}_0\left[\Pi(\|h - h_0\|_1 + \|\theta_0 + \nu_0 - \theta - \nu\|_1 > M\sqrt{\kappa_T}\epsilon_T \middle| N)\right] = o(1). \tag{S1.1}$$

Secondly, we design a test to separate θ_0 and ν_0. For this, we restrict again the set $\tilde{\Omega}_T$ to a high probability set Ω_A, where θ_0 can be correctly estimated. Let

$$A^k(T) = \{t \in [0,T];\ \tilde{\lambda}_t^k(\nu_0, h_0) < 0\}, \quad \Omega_A = \{|A^k(T)| > z_0 T,\ \forall k \in \mathcal{K}\}, \quad 1 \leqslant k \leqslant K,$$

with $z_0 > 0$ defined in the proof of Lemma A.1 (see Section S9.1), and define $\tilde{\Omega}'_T = \tilde{\Omega}_T \cap \Omega_A$. Moreover, we define a neighborhood around θ_0, $\bar{A}(R) := \{\theta \in \Theta;\ \|\theta - \theta_0\|_1 \leqslant R\}$ and $\tilde{M}_T = \tilde{M}\sqrt{\kappa_T}$ with $\tilde{M} > M$. Using again the decomposition (24), with $A = \bar{A}(\tilde{M}_T \epsilon_T)^c$, $B = A_{L_1}(M_T \epsilon_T)$, and the subset $\tilde{\Omega}'_T$, we thus only need to construct a test function $\phi \in [0,1]$ verifying:

$$\mathbb{E}_0\left[\phi \mathbb{1}_{\tilde{\Omega}'_T}\right] = o(1), \qquad \sup_{\theta \in \bar{A}(\tilde{M}_T \epsilon_T)^c, f \in A_{L_1}(M_T\epsilon_T) \cap \mathcal{F}_T} \mathbb{E}_0\left[\mathbb{E}_f\left[(1-\phi)\mathbb{1}_{\tilde{\Omega}'_T}\right] \middle| \mathcal{G}_0\right] = o(e^{-(\kappa_T + c_1)T\epsilon_T^2}). \tag{S1.2}$$

To construct this test, we first consider some arbitrary parameter $f_1 = ((\nu_k^1)_k, (h_{lk}^1)_{l,k}) \in A_{L_1}(M_T\epsilon_T)$ and $\theta_1 = (\theta_k^1)_k \in \bar{A}(\tilde{M}_T\epsilon_T)^c$, and for any $k \in [K]$, we define the following subset of the observation window

$$I_k^0(f_1, \theta_1) = \left\{t \in [0,T];\ \lambda_t^k(f_1,\theta_1) = \theta_k^1,\ \lambda_t^k(f_0,\theta_0) = \theta_k^0\right\}. \tag{S1.3}$$

By construction θ_k^0 and θ_k^1 can be identified on the set $I_k^0(f_1,\theta_1)$, hence we need $I_k^0(f_1,\theta_1)$ to be large enough in order to test between θ_k^0 and θ_k^1. We can ensure this by defining a controlled set of excursions $\mathcal{E}$. Let $l \in [K]$ such that $h_{lk}^{0-} \neq 0$, $\delta' = (x_2 - x_1)/3$ with x_1, x_2 defined in condition (S8.46), $c_\star = \min_{x \in [x_1, x_2]} h_{lk}^{0-}(x)$ and $n_1 = \lfloor 2\nu_k^1/(\kappa_1 c_\star)\rfloor + 1$ for some $0 < \kappa_1 < 1$. We consider the following subset of excursions:

$$\mathcal{E} := \{j \in [J_T];\ N[\tau_j, \tau_j + \delta') = N^l[\tau_j, \tau_j + \delta') = n_1, N[\tau_j + \delta',\ \tau_{j+1}) = 0\}, \tag{S1.4}$$

where the τ_j's are the regenerative times defined in Lemma 5.1. Using the intermediate result (S6.24) from the proof of Lemma A.5, if $|\mathcal{E}|$ is large enough, then we can find a lower bound on $|I_k^0(f_1,\theta_1)|$. We then define our generic test function:

$$\phi(f_1,\theta_1) := \max_{k \in [K]} \min\left(\mathbb{1}_{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k^0(I_k^0(f_1,\theta_1)) < -v_T} \vee \mathbb{1}_{|\mathcal{E}| < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}}, \mathbb{1}_{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k^0(I_k^0(f_1,\theta_1)) > v_T} \vee \mathbb{1}_{|\mathcal{E}| < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}}\right), \tag{S1.5}$$

where $p_0 = \mathbb{P}_0[j \in \mathcal{E}]$, $\Lambda_k^0(I_k^0(f_1,\theta_1)) = \int_0^T \mathbb{1}_{I_k^0(f_1,\theta_1)}\lambda_t^k(f_0,\theta_0)dt$, $v_T = w_T T\epsilon_T$, $w_T = 2\sqrt{\max_k \theta_k^0(\kappa_T + c_1)} + 2x_0$ and x_0 from assumption **(A2)**. From Lemma A.5, there exists $u_1 > 2x_0$ and $\zeta \in (0,1)$ such that

$$\mathbb{E}_0\left[\phi(f_1,\theta_1)\mathbb{1}_{\tilde{\Omega}'_T}\right] \leqslant e^{-u_1 T\epsilon_T^2}, \qquad \sup_{\|f - f_1\| + \|\theta - \theta_1\| \leqslant \zeta\epsilon_T} \mathbb{E}_0\left[\mathbb{E}_f\left[(1 - \phi(f_1,\theta_1))\mathbb{1}_{\tilde{\Omega}'_T}\right]\middle|\mathcal{G}_0\right] = o(e^{-(\kappa_T + c_1)T\epsilon_T^2}). \tag{S1.6}$$

To define our global test ϕ, we first cover the space $\bar{A}(\tilde{M}_T\epsilon_T)^c \times A_{L_1}(M_T\epsilon_T) \cap \mathcal{F}_T$ with L_1-balls $\{B_i\}_{1 \leqslant i \leqslant \mathcal{N}}$ of radius $\zeta\epsilon_T$, with $\zeta > 0$ and $\mathcal{N} \in \mathbb{N}$ the covering number. For each ball B_i centered at (f_i, θ_i), we define the elementary test $\phi(f_i,\theta_i)$ as in (S1.5). Then we define $\phi := \max_{i \in \mathcal{N}} \phi(f_i,\theta_i)$, and obtain that

$$\mathbb{E}_0\left[\phi\mathbb{1}_{\tilde{\Omega}'_T}\right] \leqslant \mathcal{N} e^{-u_1 T\epsilon_T^2}, \qquad \sup_{\theta \in \bar{A}(\tilde{M}_T\epsilon_T)^c, f \in A_{L_1}(M_T\epsilon_T)\cap\mathcal{F}_T} \mathbb{E}_0\left[\mathbb{E}_f\left[(1-\phi)\mathbb{1}_{\tilde{\Omega}'_T}\right]\middle|\mathcal{G}_0\right] = o(e^{-(\kappa_T + c_1)T\epsilon_T^2}).$$

Next, we find an upper bound of the covering number $\mathcal{N}$ using assumption **(A2)**. We note that if $f \in A_{L_1}(M_T\epsilon_T)$, then for any $(l,k) \in [K]^2$, $\theta_k \leqslant \theta_k + \nu_k = r_k^f \leqslant r_k^0 + \epsilon_T \leqslant 2(\theta_k^0 + \nu_k^0)$. Consequently, using similar computations as in the proof of Proposition 5.5 in Section S2), one can find $x'_0 > 0$ such that

$$\mathcal{N} \leqslant \left(\frac{2\max_k(\theta_k^0 + \nu_k^0)}{\zeta\epsilon_T}\right)^K \left(\frac{\max_k \nu_k^0 + \epsilon_T}{\zeta\epsilon_T}\right)^K \mathcal{N}(\zeta\epsilon_T, \mathcal{H}_T, \|.\|_1) \lesssim e^{-K\log\epsilon_T} e^{x'_0 T\epsilon_T^2} \lesssim e^{K\log T} e^{x'_0 T\epsilon_T^2} = o(e^{u_1 T\epsilon_T^2}),$$

since $\log T = o(T\epsilon_T^2)$ by assumption. Hence, reporting into (S1.6), this proves that (S1.2) holds and allows us to conclude that $\mathbb{E}_0\left[\Pi(\bar{A}(\tilde{M}_T\epsilon_T)^c|N)\right] = \mathbb{E}_0\left[\Pi(\|\theta-\theta_0\|_1 > \tilde{M}\sqrt{\kappa_T}\epsilon_T|N)\right] = o(1)$. Finally, since $\tilde{M} > M$, from (S1.1), we also have that $\mathbb{E}_0\left[\Pi(\|\nu+\theta-\nu_0-\theta_0\|_1 + \|h-h_0\|_1 > \tilde{M}\sqrt{\kappa_T}\epsilon_T|N)\right] = o(1)$,. Therefore it only remains to prove that $\mathbb{E}_0\left[\Pi(\|\nu-\nu_0\|_1 > \tilde{M}\sqrt{\kappa_T}\epsilon_T|N)\right] = o(1)$. By the triangle inequality, we have $\|\nu-\nu_0\|_1 \leqslant \|\nu+\theta-\nu_0-\theta_0\|_1 + \|\theta-\theta_0\|_1$, and, up to a modification of the constant $\tilde{M}$,

$$\mathbb{E}_0\left[\Pi(\left\|\nu-\nu_0\right\|_1 > \tilde{M}\sqrt{\kappa_T}\epsilon_T|N)\right] \leqslant \mathbb{E}_0\left[\Pi(\left\|\nu+\theta-\nu_0-\theta_0\right\|_1 > \tilde{M}\sqrt{\kappa_T}\epsilon_T|N)\right] + \mathbb{E}_0\left[\Pi(\left\|\theta-\theta_0\right\|_1 > \tilde{M}\sqrt{\kappa_T}\epsilon_T|N)\right] = o(1),$$

which terminates this proof.

S2. Proofs of Theorem 5.5 and of Lemma A.2

S2.1. Proof of Theorem 5.5

This section contains the proof of the posterior concentration rate w.r.t. the stochastic distance defined in (23) in Sulem, Rivoirard and Rousseau (2022). We use the well-known strategy of Ghosal and van der Vaart (2007) which has the following steps. First, the space of observations is restricted to a subset $\tilde{\Omega}_T$ defined in (25) which has high probability (see Lemma A.1). Secondly, we use a lower bound of the denominator D_T defined in (5) using Lemma A.2. Thirdly, we consider $A_{d_1}(M_T'\epsilon_T) \subset \mathcal{F}$, the ball centered at f_0 of radius $M_T'\epsilon_T$ w.r.t the auxiliary stochastic distance $\tilde{d}_{1T}$. To find an upper bound of the numerator $N_T(A_{d_1}(M_T'\epsilon_T)^c)$ as defined in (5), $A_{d_1}(M_T'\epsilon_T)^c$ is partitioned into slices S_i on which we can design tests that have exponentially decreasing type I and type II errors (see Lemma S6.1). We then define ϕ as the maximum of the tests on the individual slices S_i. Note that the following proof applies to all estimation scenarios, and for generality here, we consider θ_0 unknown.

We recall the notation $A_{d_1}(\epsilon) = \{f \in \mathcal{F};\ \tilde{d}_{1T}(f, f_0) \leqslant \epsilon\}$. and from (5), $D_T = \int_{\mathcal{F}} e^{L_T(f)-L_T(f_0)}d\Pi(f)$. For a sequence ϵ_T verifying the assumptions of Theorem 3.2 and for $i \geqslant 1$, we denote

$$S_i = \{f \in \mathcal{F}_T;\ Ki\epsilon_T \leqslant \tilde{d}_{1T}(f, f_0) \leqslant K(i+1)\epsilon_T\}, \tag{S2.7}$$

where $\mathcal{F}_T = \{f = (\nu, h) \in \mathcal{F};\ h = (h_{lk})_{l,k} \in \mathcal{H}_T,\ \nu \in \Upsilon_T\}$. Let $M_T' = M'\sqrt{\kappa_T}$ with $M' > 0$ and κ_T defined in (6). Using the decomposition (24) with $A = A_{d_1}(M_T'\epsilon_T)^c$ (and $B = \mathcal{F}$) , for any test function $\phi \in [0, 1]$, we have

$$\begin{aligned}\mathbb{E}_0[\Pi(A_{d_1}(M_T'\epsilon_T)^c|N)] &\leqslant \mathbb{P}_0(\tilde{\Omega}_T^c) + \mathbb{P}_0\left(\{D_T < e^{-\kappa_T T\epsilon_T^2}\Pi(B_\infty(\epsilon_T))\} \cap \tilde{\Omega}_T\right) + \mathbb{E}_0[\phi\mathbb{1}_{\tilde{\Omega}_T}] \\ &+ \frac{e^{\kappa_T T\epsilon_T^2}}{\Pi(B_\infty(\epsilon_T))}\Pi(\mathcal{F}_T^c) + \frac{e^{\kappa_T T\epsilon_T^2}}{\Pi(B_\infty(\epsilon_T))}\left(+\sum_{i=M_T'}^{+\infty}\int_{\mathcal{F}_T}\mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{f\in S_i}(1-\phi)\right]|\mathcal{G}_0\right]\right]d\Pi(f)\right).\end{aligned} \tag{S2.8}$$

For the first term on the RHS of (S2.8), we have $\mathbb{P}_0(\tilde{\Omega}_T^c) = o(1)$ by Lemma A.1. For the fourth term of the RHS of (S2.8), under **(A0)** and **(A1)**, we have that

$$\frac{e^{\kappa_T T\epsilon_T^2}}{\Pi(B_\infty(\epsilon_T))}\Pi(\mathcal{F}_T^c) \leqslant e^{(\kappa_T+c_1)T\epsilon_T^2}(\Pi(\mathcal{H}_T^c) + \Pi(\Upsilon_T^c)) = o(1).$$

The second term of (S2.8) is controlled by (26) and goes to 0.

We now deal with the third and fifth terms on the RHS of (S2.8), which require to define a suitable test function ϕ. Let $i \in \mathbb{N}, i \geqslant M'_T$ and $f \in S_i$. On $\tilde{\Omega}_T$, with $A_2(T)$ defined in (22), we have that

$$T\tilde{d}_{1T}(f,f_0) = \sum_{l=1}^{K} \int_{A_2(T)} \left|\lambda_l^k(f) - \lambda_l^k(f_0)\right| dt = \sum_{l=1}^{K} \sum_{j=1}^{J_T-1} \int_{\tau_j}^{\xi_j} \left|\lambda_l^k(f) - \lambda_l^k(f_0)\right| dt$$
$$\geqslant \sum_{l=1}^{K} \sum_{j=1}^{J_T-1} \int_{\tau_j}^{U_j^{(1)}} |r_l^f - r_l^0| dt \geqslant \sum_{j=1}^{J_T-1} (U_j^{(1)} - \tau_j) \sum_l |r_l^f - r_l^0| \geqslant \frac{T}{2\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1]} \sum_l |r_l^f - r_l^0|,$$

with $r_f = (\phi_1(\nu_1), \dots, \phi_K(\nu_K))$, $r_0 = (\phi_1(\nu_1^0), \dots, \phi_K(\nu_K^0))$ and τ_j, ξ_j, $U_j^{(1)}$, $1 \leqslant j \leqslant J_T - 1$ defined in Sections 5.1 and 5.2. Consequently, for any $l \in [K]$, since $\tilde{d}_{1T}(f,f_0) \leqslant K(i+1)\epsilon_T$, we obtain that

$$r_l^f \leqslant r_l^0 + 2K(i+1)\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1]\epsilon_T \leqslant r_l^0 + 1 + 2K\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1]\, i\epsilon_T, \tag{S2.9}$$

for T large enough. Moreover, using Assumption 3.1, ϕ_l^{-1} is L'-Lipschitz on $J_l = \phi_l(I_l)$ and $r_l^0 \in J_l$. With $\varepsilon > 0$ from Assumption 3.1, we now separate the set of indices i in two subsets.

Case 1: i is such that $2L'\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1] K(i+1)\epsilon_T < \varepsilon$. Then we have that $r_l^f \in J_l$ and $\nu_l \in I_l$ since $|r_l^f - r_l^0| = \leqslant 2\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1] K(i+1)\epsilon_T$. Consequently, $\frac{1}{L'}|\nu_l - \nu_l^0| \leqslant |r_l^f - r_l^0| \leqslant L|\nu_l - \nu_l^0|$ and in particular,

$$\nu_l \leqslant \nu_l^0 + 2KL'(i+1)\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1]\epsilon_T.$$

Defining

$$\mathcal{F}_i = \left\{ f \in \mathcal{F}_T;\ \nu_l^f \leqslant \nu_l^0 + 1 + 2KL'\|r_0\|_1 \mathbb{E}_0[\Delta\tau_1]\, i\epsilon_T, \forall l \in [K] \right\},$$

we therefore have that for any $f \in S_i$ and T large enough, $f \in \mathcal{F}_i$ Let $(f_{i,n})_{n=1}^{\mathcal{N}_i}$ be the centering points of a minimal L_1-covering of $\mathcal{F}_i$ by $\mathcal{N}_i$ balls of radius $\zeta i\epsilon_T$ with $\zeta = 1/(6N_0)$, and N_0 defined in the proof of Lemma S6.1 in Section S6.2. There exists $C_0 > 0$ such that we have

$$\mathcal{N}_i \leqslant \left(\frac{C_0(1+i\epsilon_T)}{\zeta i\epsilon_T/2}\right)^K \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1).$$

If $i\epsilon_T \leqslant 1$,

$$\mathcal{N}_i \leqslant \left(\frac{4C_0}{\zeta i\epsilon_T}\right)^K \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1) = \left(\frac{4C_0}{\zeta}\right)^K e^{-K\log(i\epsilon_T)} \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1).$$

Otherwise, if $i\epsilon_T \geqslant 1$,

$$\mathcal{N}_i \leqslant \left(\frac{4C_0}{\zeta}\right)^K \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1).$$

Moreover, since $i \mapsto \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1)$ is non-increasing, and if $i \geqslant 2\zeta_0/\zeta$, we have that $\mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T \|.\|_1) \leqslant \mathcal{N}(\zeta_0\epsilon_T, \mathcal{H}_T, \|.\|_1) \leqslant e^{x_0 T\epsilon_T^2}$ using **(A2)**. Consequently, since $\epsilon_T > \epsilon_T^2 > \frac{1}{T}$ when T is large enough, $e^{-\log(i\epsilon_T)} \leqslant e^{\log(\frac{\zeta}{2\zeta_0}T)}$ and we obtain

$$\mathcal{N}_i \leqslant \left(\frac{4C_0}{\zeta}\right)^K \left(\frac{\zeta}{2\zeta_0}\right)^K e^{K\log T} \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1) = \left(\frac{2C_0}{\zeta_0}\right)^K e^{K\log T} \mathcal{N}(\zeta i\epsilon_T/2, \mathcal{H}_T, \|.\|_1)$$

$$\leqslant C_K e^{K \log T} e^{x_0 T \epsilon_T^2},$$

denoting $C_K = \left(\frac{2C_0}{\zeta_0}\right)^K$.

Case 2: $2L' \|r_0\|_1 \mathbb{E}_0 [\Delta\tau_1] K(i+1)\epsilon_T > \varepsilon$. Then in this case we define $\mathcal{F}_i = \mathcal{F}_T$ and $\nu_T = e^{c_2 T \epsilon_T^2}$, and the L_1-covering number of $\mathcal{F}_i$ is now upper bounded by

$$\mathcal{N}_i \leqslant \left(\frac{\nu_T}{\zeta i \epsilon_T/2}\right)^K \mathcal{N}(\zeta i \epsilon_T/2, \mathcal{H}_T, \|.\|_1) \leqslant C_0' e^{(x_0 + c_2 K) T \epsilon_T^2},$$

with $C_0' > 0$ a constant.

In both cases, considering the tests $\phi_i = \max_{n \in [\mathcal{N}_i]} \phi_{f_{i,n}}$ with $\phi_{f_{i,n}}$, $\gamma_1 = \min_l x_{1l}$ defined in Lemma S6.1, and $C_K' = C_K \vee C_0'$, $x_0' = x_0 + c_K$, we have

$$\mathbb{E}_0[\mathbb{1}_{\tilde{\Omega}_T} \phi_i] \leqslant \mathcal{N}_i e^{-\gamma_1 T(i^2 \epsilon_T^2 \wedge i\epsilon_T)} \leqslant C_K'(2K+1) e^{K \log T} e^{x_0' T \epsilon_T^2} e^{-\gamma_1 T(i^2 \epsilon_T^2 \wedge i \epsilon_T)},$$

$$\mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T} \mathbb{1}_{f \in S_i}(1-\phi_i)|\mathcal{G}_0\right]\right] \leqslant (2K+1) e^{-\gamma_1 T(i^2 \epsilon_T^2 \wedge i\epsilon_T)}.$$

Choosing $\phi = \max_{M_T' \leqslant i \leqslant \mathcal{N}_i} \phi_i$ and since $M_T' \geqslant 2\zeta_0/\zeta$ for T large enough, we obtain

$$\begin{aligned}
\mathbb{E}_0[\mathbb{1}_{\tilde{\Omega}_T} \phi] &\leqslant C_K'(2K+1) e^{K \log T} e^{x_0' T \epsilon_T^2} \left[\sum_{i=M_T'}^{\epsilon_T^{-1}} e^{-\gamma_1 i^2 T \epsilon_T^2} + \sum_{i > \epsilon_T^{-1}} e^{-\gamma_1 i T \epsilon_T}\right] \\
&\leqslant C_K'(2K+1) e^{K \log T} e^{x_0' T \epsilon_T^2} \left[\sum_{i=M_T'}^{\epsilon_T^{-1}} e^{-\gamma_1 i M_T' T \epsilon_T^2} + \sum_{i > \epsilon_T^{-1}} e^{-\gamma_1 T i \epsilon_T}\right] \\
&\leqslant C_K'(2K+1) e^{K \log T} e^{x_0' T \epsilon_T^2} \left[2 e^{-\gamma_1 M_T'^2 T \epsilon_T^2} + 2 e^{-\gamma_1 T}\right] \\
&\leqslant 4 C_K'(2K+1)[e^{-\gamma_1 M_T'^2 T \epsilon_T^2} + e^{-\gamma_1 T}],
\end{aligned} \tag{S2.10}$$

since $\log^3 T = O(T\epsilon_T^2)$ by assumption. Therefore, we arrive at $\mathbb{E}_0[\mathbb{1}_{\tilde{\Omega}_T} \phi] = o(1)$. Similarly, we can obtain

$$\begin{aligned}
\mathbb{E}_0\left[\sum_{i \geqslant M_T'} \int_{\mathcal{F}_T} \mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T} \mathbb{1}_{f \in S_i}(1-\phi)|\mathcal{G}_0\right] d\Pi(f)\right] &\leqslant (2K+1)\left[\sum_{i=M_T'}^{\epsilon_T^{-1}} e^{-\gamma_1 i^2 T \epsilon_T^2} + \sum_{i > \epsilon_T^{-1}} e^{-\gamma_1 T i \epsilon_T}\right] \\
&\leqslant 4(2K+1)[e^{-\gamma_1 M_T'^2 T \epsilon_T^2} + e^{-\gamma_1 T}].
\end{aligned}$$

Therefore, using **(A0)**, we have for the second term in (S2.8),

$$\begin{aligned}
&\frac{e^{\kappa_T T \epsilon_T^2}}{\Pi(B_\infty(\epsilon_T))}\left(\sum_{i=M_T'}^{+\infty} \int_{\mathcal{F}_T} \mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T} \mathbb{1}_{f \in S_i}(1-\phi)|\mathcal{G}_0\right]\right] d\Pi(f)\right) \leqslant \frac{e^{\kappa_T T \epsilon_T^2}}{e^{-c_1 T \epsilon_T^2}} 4(2K+1)[e^{-\gamma_1 M_T'^2 T \epsilon_T^2} + e^{-\gamma_1 T}] \\
&\leqslant 4(2K+1) e^{-\gamma_1 M_T'^2 T \epsilon_T^2/2} = o(1),
\end{aligned} \tag{S2.11}$$

for $M'_T > \sqrt{c_1 + \kappa_T}$, which holds true if $M'_T = M' \sqrt{\kappa_T}$ with M' large enough. Aggregating the upper bounds previously obtained, we can finally conclude that

$$\mathbb{E}_0[\Pi(A_{d_1}(M'_T \epsilon_T)^c | N)] \leqslant \mathbb{P}_0(\tilde{\Omega}_T^c) + o(1) = o(1),$$

which terminates the proof of Theorem 5.5.

S2.2. Proof of Lemma A.2

In this section, we prove a control on the log-likelihood ratio of the form $\mathbb{P}_0\left[L_T(f_0) - L_T(f) \geqslant 5z_T\right] = o(1)$, where $z_T = T\epsilon_T^2(\log T)^r$ where $r = 0, 1, 2$ is defined in Lemma S7.3 and depends on the assumptions on the link function. We have

$$\begin{aligned} L_T(f_0) - L_T(f) &= \sum_k \int_0^T \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_0^T (\lambda_t^k(f_0) - \lambda_t^k(f)) dt \\ &= W_0 + \sum_{j=1}^{J_T - 1} T_j + W_T, \end{aligned}$$

with

$$\begin{aligned} W_0 &:= \sum_k \int_0^{\tau_1} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_0^{\tau_1} (\lambda_t^k(f_0) - \lambda_t^k(f)) dt, \\ W_T &:= \sum_k \int_{\tau_{J_T}}^T \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_{\tau_{J_T}}^T (\lambda_t^k(f_0) - \lambda_t^k(f)) dt. \end{aligned}$$

Let $\mathcal{L}_T = L_T(f_0) - L_T(f) - \mathbb{E}_0\left[L_T(f_0) - L_T(f)\right] = L_T(f_0) - L_T(f) - KL(f_0, f)$, with $KL(f_0, f)$ the Kullback-Leibler divergence defined in (S7.29). Then

$$\begin{aligned} &\mathbb{P}_0\left[\mathcal{L}_T \geqslant 4z_T\right] = \mathbb{P}_0\left[\sum_{j=1}^{J_T-1} T_j + W_0 + W_T - KL(f_0, f) \geqslant 4z_T\right] \\ &= \mathbb{P}_0\left[\sum_{j=1}^{J_T-1}(T_j - \mathbb{E}_0\left[T_j\right]) + \sum_{j=1}^{J_T-1} \mathbb{E}_0\left[T_j\right] - \mathbb{E}_0\left[\sum_{j=1}^{J_T-1} T_j\right] + W_T - \mathbb{E}_0\left[W_T\right] + W_0 - \mathbb{E}_0\left[W_0\right] \geqslant 4z_T\right] \\ &\leqslant \mathbb{P}_0\left[\sum_{j=1}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right] \geqslant z_T\right] + \mathbb{P}_0\left[(J_T - \mathbb{E}_0\left[J_T\right])\mathbb{E}_0\left[T_1\right] - \mathbb{E}_0\left[\sum_{j=0}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right]\right] \geqslant z_T\right] + \mathbb{P}_0\left[W_T - \mathbb{E}_0\left[W_T\right] \geqslant z_T\right] \\ &+ \mathbb{P}_0\left[W_0 - \mathbb{E}_0\left[W_0\right] \geqslant z_T\right], \end{aligned} \tag{S2.12}$$

using equation (S7.31) and that

$$KL(f_0, f) = \underbrace{\sum_k \mathbb{E}_0\left[\int_{\tau_0}^{\tau_1} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_0^{\tau_1} (\lambda_t^k(f_0) - \lambda_t^k(f)) dt\right]}_{\mathbb{E}_0[W_0]}$$

$$+ \underbrace{\sum_k \mathbb{E}_0\left[\int_0^{\tau_{J_T}} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_0^{\tau_{J_T}} (\lambda_t^k(f_0) - \lambda_t^k(f))dt\right]}_{=\mathbb{E}_0\left[\sum_{j=1}^{J_T-1} T_j\right]}$$

$$+ \underbrace{\sum_k \mathbb{E}_0\left[\int_{\tau_{J_T}}^{T} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_{\tau_{J_T}}^{T} (\lambda_t^k(f_0) - \lambda_t^k(f))dt\right]}_{\mathbb{E}_0[W_T]}.$$

From Lemma S7.3, we have that $\mathbb{P}_0\left[\sum_{j=1}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right] \geqslant z_T\right] = o(1)$. We now deal with the second term on the RHS of (S2.12). Using Lemma S7.3, we have

$$\begin{aligned}
\mathbb{E}_0\left[\sum_{j=1}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right]\right] &= \mathbb{E}_0\left[\sum_{j=\lfloor T/\mathbb{E}_0[\Delta\tau_1]\rfloor}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right]\right] \\
&\leqslant \mathbb{E}_0\left[\sum_{J\in\mathcal{J}_T} \mathbb{1}_{J_T=J}\left(\sum_{j=\lfloor T/\mathbb{E}_0[\Delta\tau_1]\rfloor}^{J-1} |T_j - \mathbb{E}_0\left[T_j\right]|\right)\right] + \sqrt{\mathbb{P}_0\left[J_T \notin \mathcal{J}_T\right]}\sqrt{T^2\mathbb{E}_0\left[T_1^2\right]} \\
&\leqslant \mathbb{E}_0\left[\sum_{j=\lfloor \frac{T}{\mathbb{E}_0[\Delta\tau_1]}(1-c_\beta\sqrt{\frac{\log T}{T}})\rfloor}^{\lfloor \frac{T}{\mathbb{E}_0[\Delta\tau_1]}(1+c_\beta\sqrt{\frac{\log T}{T}})\rfloor} |T_j - \mathbb{E}_0\left[T_j\right]|\right] + T^{1-\beta/2}\sqrt{\mathbb{E}_0\left[T_1^2\right]} \\
&\leqslant \frac{2c_\beta}{\mathbb{E}_0\left[\Delta\tau_1\right]}\mathbb{E}_0\left[|T_1 - \mathbb{E}_0\left[T_j\right]|\right]\sqrt{T\log T} + T^{1-\beta/2}\sqrt{\mathbb{E}_0\left[T_1^2\right]} \\
&\lesssim \sqrt{\mathbb{E}_0\left[T_1^2\right]}\sqrt{T\log T} \lesssim \sqrt{T}(\log T)^{3/2}\epsilon_T = o(z_T),
\end{aligned}$$

since $\log^3 T = O(z_T)$ by assumption. Consequently,

$$\begin{aligned}
\mathbb{P}_0\left[(J_T - \mathbb{E}_0\left[J_T\right])\mathbb{E}_0\left[T_1\right] - \mathbb{E}_0\left[\sum_{j=0}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right]\right] \geqslant z_T\right] &\leqslant \mathbb{P}_0\left[J_T - \mathbb{E}_0\left[J_T\right] \geqslant \frac{z_T}{2\mathbb{E}_0\left[T_1\right]}\right] \\
&\leqslant \mathbb{P}_0\left[J_T - \frac{T}{\mathbb{E}_0\left[\Delta\tau_1\right]} \geqslant \frac{z_T}{4\mathbb{E}_0\left[T_1\right]}\right],
\end{aligned}$$

using that $J_T - \mathbb{E}_0\left[J_T\right] = J_T - \frac{T}{\mathbb{E}_0[\Delta\tau_1]} + \frac{T}{\mathbb{E}_0[\Delta\tau_1]} - \mathbb{E}_0\left[J_T\right]$ and $\frac{T}{\mathbb{E}_0[\Delta\tau_1]} - \mathbb{E}_0\left[J_T\right] \leqslant \frac{z_T}{4\mathbb{E}_0[T_1]}$ for T large enough. Consequently, since $\mathbb{E}_0\left[T_1\right] \leqslant \sqrt{\frac{z_T}{T}}$, we have with $\eta_T = \sqrt{\frac{z_T}{4\mathbb{E}_0[T_1]}}$ and $B_j = \tau_j - \tau_{j-1} - \mathbb{E}_0\left[\Delta\tau_1\right]$, and using the computations as for the proof of Lemma A.1,

$$\begin{aligned}
\mathbb{P}_0\left[J_T - \frac{T}{\mathbb{E}_0\left[\Delta\tau_1\right]} \geqslant \eta_T\right] &\leqslant \mathbb{P}_0\left[\tau_{\lfloor T/\mathbb{E}_0[\Delta\tau_1]+\eta_T\rfloor} \leqslant T\right] \\
&= \mathbb{P}_0\left[\sum_{j=1}^{\lfloor T/\mathbb{E}_0[\Delta\tau_1]+\eta_T\rfloor} B_j \leqslant T - \lfloor T/\mathbb{E}_0\left[\Delta\tau_1\right] + \eta_T\rfloor\mathbb{E}_0\left[\Delta\tau_1\right]\right]
\end{aligned}$$

$$\leqslant \mathbb{P}_0\left[\sum_{j=1}^{\lfloor T/\mathbb{E}_0[\Delta\tau_1]+\eta_T\rfloor} B_j \leqslant -\mathbb{E}_0\left[\Delta\tau_1\right]\eta_T + \mathbb{E}_0\left[\Delta\tau_1\right]\right]$$

$$\leqslant \frac{4\lfloor T/\mathbb{E}_0\left[\Delta\tau_1\right]+\eta_T\rfloor \mathbb{E}_0\left[\Delta\tau_1^2\right]}{\mathbb{E}_0\left[\Delta\tau_1\right]^2\eta_T^2} \lesssim \frac{T}{\eta_T^2} + \frac{1}{\eta_T} \lesssim \frac{1}{z_T} = o(1).$$

For the third term on the RHS of (S2.12), applying Bienayme-Chebyshev's inequality, we have

$$\mathbb{P}_0\left[W_T - \mathbb{E}_0\left[W_T\right] \geqslant z_T\right] \leqslant \frac{\mathbb{E}_0\left[W_T^2\right]}{z_T^2}. \tag{S2.13}$$

Using similarly computations as in Lemma S7.3, we obtain

$$\mathbb{E}_0\left[W_T^2\right] = \mathbb{E}_0\left[\left(\sum_k \int_{\tau_{J_T}}^T \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)dN_t^k - \int_{\tau_{J_T}}^T (\lambda_t^k(f_0) - \lambda_t^k(f))dt\right)^2\right]$$

$$\lesssim \mathbb{E}_0\left[(T-\tau_{J_T})\int_{\tau_{J_T}}^T\left[\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0) - (\lambda_t^k(f_0)-\lambda_t^k(f))\right]^2 dt\right] + \mathbb{E}_0\left[\int_{\tau_{J_T}}^T \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right].$$

Then since

$$\mathbb{E}_0\left[(T-\tau_{J_T})\int_{\tau_{J_T}}^T\left[\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0) - (\lambda_t^k(f_0)-\lambda_t^k(f))\right]^2 dt\right] \leqslant \mathbb{E}_0\left[\Delta\tau_1\int_{\tau_1}^{\tau_2}\chi\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)^2\lambda_t^k(f_0)^2 dt\right],$$

$$\mathbb{E}_0\left[\int_{\tau_{J_T}}^T \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right] \leqslant \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right],$$

we can use the bounds derived for $\mathbb{E}_0\left[T_j^2\right]$ in Lemma S7.3.

We finally obtain

$$\mathbb{P}_0\left[W_T - \mathbb{E}_0\left[W_T\right] \geqslant z_T\right] \lesssim \frac{(\log^2 T)\epsilon_T^2}{z_T^2} \lesssim \frac{\log^2 T}{T^2\epsilon_T^2} = o(1).$$

With similar computations, we also obtain that $\mathbb{P}_0\left[W_0 - \mathbb{E}_0\left[W_0\right] \geqslant z_T\right] = o(1)$. Consequently, reporting into (S2.12) and using Lemma S7.1, we finally obtain that

$$\mathbb{P}_0\left[L_T(f_0) - L_T(f) > 5z_T\right] \leqslant \mathbb{P}_0\left[\mathcal{L}_T > 5z_T - u_T\right] \leqslant \mathbb{P}_0\left[\mathcal{L}_T > 4z_T\right] = o(1),$$

since $KL(f_0, f) \leqslant u_T \leqslant z_T$ using Lemmas S7.1 and S7.3.

S3. Proofs of Theorem 3.9, Theorem 3.11 and Proposition 3.10

S3.1. Proof of Theorem 3.9

In this section, we show that in all the models satisfying the assumptions of Theorem 3.2 or Proposition 3.5, the posterior distribution is consistent on the connectivity graph parameter δ_0. For ease of exposition, we here report the proof for the models considered in Theorem 3.2. We first recall the notation

$M_T = M\sqrt{\kappa_T}$, $A_{L_1}(M_T\epsilon_T) = \{f \in \mathcal{F};\ \|r_f - r_0\|_1 + \|h - h_0\|_1 \leqslant M_T\epsilon_T\}$, and $I(\delta_0) = \{(l,k) \in [K]^2,\ \delta^0_{lk} = 1\}$. We first note that

$$\Pi(\delta \neq \delta_0|N) = \Pi\left(\exists (l,k) \in [K]^2, \delta^0_{lk} \neq \delta_{lk}\Big|N\right) \leqslant \Pi\left(\exists (l,k) \in I(\delta_0), \delta_{lk} = 0\Big|N\right) + \sum_{(l,k)\notin I(\delta_0)} \Pi\left(\delta_{lk} = 1\Big|N\right). \quad \text{(S3.14)}$$

For the first term on the RHS of (S3.14), using Theorem 3.2, we have that

$$\Pi\left(\exists (l,k) \in I(\delta_0), \delta_{lk} = 0\Big|N\right) \leqslant \sum_{(l,k)\in I(\delta_0)} \Pi\left(\{\delta_{lk} = 0\} \cap A_{L_1}(M_T\epsilon_T)\Big|N\right) + o_{\mathbb{P}_0}(1).$$

For large enough T, if $\|h^0_{lk}\|_1 > M_0 M_T\epsilon_T$ with $M_0 > 1$, then

$$\{f \in \mathcal{F}; \delta_{lk} = 0\} \subset \{f \in \mathcal{F}; \left\|h^0_{lk} - h_{lk}\right\|_1 = \left\|h^0_{lk}\right\|_1\} \subset \left\{f \in \mathcal{F}; \left\|h^0_{lk} - h_{lk}\right\|_1 > \frac{\|h^0_{lk}\|_1}{2}\right\} \subset A_{L_1}(M_T\epsilon_T)^c,$$

therefore $\Pi\left(\{\delta_{lk} = 0\} \cap A_{L_1}(M_T\epsilon_T)\Big|N\right) = 0$. For the second term on the RHS of (S3.14), since $(l,k) \notin I(\delta_0)$ implies that $\left\|h^0_{lk}\right\|_1 = 0$ and $\{\delta_{lk} = 1\} \cap A_{L_1}(M_T\epsilon_T) \subset \{f \in \mathcal{F};\ 0 < \|h_{lk}\|_1 \leqslant M_T\epsilon_T\}$, defining $N_T = \int_{\{\delta_{lk}=1\}\cap A_{L_1}(M_T\epsilon_T)} e^{L_T(f)-L_T(f_0)} d\Pi(f)$, and using the decomposition (24) with $A = A_{L_1}(M_T\epsilon_T), B = \{\delta_{lk} = 1\}$ and $\phi = 1$, we obtain that

$$\begin{aligned}\mathbb{E}_0\left[\Pi(\{\delta_{lk} = 1\} \cap A_{L_1}(M_T\epsilon_T)|N)\right] &\leqslant \mathbb{P}_0(D_T < e^{-(\kappa_T+c_1)T\epsilon_T^2} \cap \tilde{\Omega}_T) + \mathbb{P}_0(\tilde{\Omega}_T^c) + e^{(\kappa_T+c_1)T\epsilon_T^2}\Pi(\{\delta_{lk} = 1\} \cap A_{L_1}(M_T\epsilon_T)) \\ &\leqslant o(1) + e^{(\kappa_T+c_1)T\epsilon_T^2} \sum_{\delta\in\{0,1\}^{K^2}} \mathbb{1}_{\delta_{lk}=1} \Pi_{h|\delta}(\|h_{lk}\|_1 \leqslant M_T\epsilon_T|\delta) = o(1),\end{aligned}$$

where in the last inequality we have used assumptions **(A0)-(A1)**, (11), and the construction of the prior from Section 4. Consequently, from (S3.14), we finally arrive at $\mathbb{E}_0[\Pi(\delta \neq \delta_0|N)] = o(1)$.

S3.2. Proof of Theorem 3.11

We here prove the consistency of the penalised estimator defined in (13). We consider the models satisfying the assumptions of Theorem 3.2, although our proof is also valid for the ReLU-type models of Proposition 3.5. Besides, for $f \in \mathcal{F}$, we use the shortened notation $d_{1T} := \tilde{d}_{1T}(f, f_0)$ and $\hat{\delta}^{\Pi,L} := \hat{\delta}^{\Pi,L}(N)$. We recall that for $(l,k) \in [K]^2$, $S_{lk} = \|h_{lk}\|_1$ and the notation from previous proofs, $M_T = M\sqrt{\kappa_T}$, $M'_T = M'\sqrt{\kappa_T}$ with $M > M' > 0$. We first note that $\mathbb{P}_0\left[\hat{\delta}^{\Pi,L} \neq \delta_0\right] \leqslant \sum_{l,k} \mathbb{P}_0\left[\hat{\delta}^{\Pi,L}_{lk} \neq \delta^0_{lk}\right]$ and consider two cases for each (l,k).

- **Case 1:** $(l,k) \notin I(\delta_0)$, i.e, $\delta^0_{lk} = 0$. Using (13) and (14), there exists $a > 0$ such that with $c'_1 := a + c_1 + \kappa_T$, for any $\gamma > 0$, we have

$$\begin{aligned}\mathbb{P}_0\left[\hat{\delta}^{\Pi,L}_{lk} \neq \delta^0_{lk}\right] &= \mathbb{P}_0\left[\hat{\delta}^{\Pi,L}_{lk} = 1\right] \\ &\leqslant \mathbb{P}_0\left[e^{-c'_1T\epsilon_T^2}\Pi(\delta_{lk} = 1,\ S_{lk} \leqslant M_T\epsilon_T|N) \geqslant \Pi(\delta_{lk} = 0|N) - \Pi(S_{lk} > M_T\epsilon_T|N)\right] \\ &\leqslant \mathbb{P}_0\left[e^{-c'_1T\epsilon_T^2}\Pi(\delta_{lk} = 1,\ S_{lk} \leqslant M_T\epsilon_T|N) \geqslant \Pi(\delta_{lk} = 0|N)/2\right] \\ &+ \mathbb{P}_0\left[\Pi(S_{lk} > M_T\epsilon_T|N) > \Pi(\delta_{lk} = 0|N)/2\right]. \quad \text{(S3.15)}\end{aligned}$$

To show that the second term in the previous equation is $o(1)$, it is enough to show that

$$\mathbb{P}_0\left[\Pi(d_{1T} > M'_T\epsilon_T|N) > \Pi(\delta_{lk} = 0|N)/4\right] = o(1), \tag{S3.16}$$

$$\mathbb{P}_0\left[\Pi(d_{1T} \leqslant M'_T\epsilon_T,\, S_{lk} > M_T\epsilon_T|N) > \Pi(\delta_{lk} = 0|N)/4\right] = o(1). \tag{S3.17}$$

Let $m_T(\delta_{lk} = 0) := \int_{\mathcal{F}_T} e^{L_T(f)-L_T(f_0)} d\Pi(f|\delta_{lk} = 0)$. Similarly to the computations of the lower bound of D_T in Section S2, we have under **(A0')** that $\mathbb{P}_0\left[m_T(\delta_{lk} = 0) \leqslant e^{-\kappa'_T T\epsilon_T^2}\right] = o(1)$ with $\kappa'_T := \kappa_T + c_1$. Using the test function from the proof of Theorem 3.5 in Section S2, $\phi = \max_i \phi(f_i)$ (with $\phi(f_i)$ defined in Lemma S6.1), we have

$$\begin{aligned}
\mathbb{P}_0\left[\Pi(d_{1T} > M'_T\epsilon_T|N) > \Pi(\delta_{lk} = 0|N)/4\right] &\leqslant \mathbb{E}_0\left[\phi \mathbb{1}_{\tilde{\Omega}_T}\right] + \mathbb{P}_0\left[\tilde{\Omega}_T^c\right] + \Pi(\mathcal{F}_T^c) \\
&\quad + \mathbb{E}_0\left[(1-\phi)\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{\int_{\mathcal{F}_T}\mathbb{1}_{d_{1T}>M'_T\epsilon_T} e^{L_T(f)-L_T(f_0)}d\Pi(f) > \Pi(\delta_{lk}=0)m_T(\delta_{lk}=0)/4}\right] \\
&\leqslant o(1) + \mathbb{E}_0\left[(1-\phi)\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{\int_{\mathcal{F}_T}\mathbb{1}_{d_{1T}>M'_T\epsilon_T} e^{L_T(f)-L_T(f_0)}d\Pi(f) > e^{-\kappa'_T T\epsilon_T^2}/4}\right] \\
&\leqslant o(1) + 4e^{\kappa'_T T\epsilon_T^2}\int_{\mathcal{F}_T}\mathbb{E}_0\left[\mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{d_{1T}>M'_T\epsilon_T}(1-\phi)|\mathcal{G}_0\right]d\Pi(f|\delta_{lk}=0)\right].
\end{aligned}$$

In the second inequality, we have notably used (S2.10) $\mathbb{E}_0\left[\phi\mathbb{1}_{\tilde{\Omega}_T}\right] = o(1)$ from Section S2. Moreover, from (S2.11), there exists $\gamma_1 > 0$ such that

$$\sum_{i \geqslant M'_T}\int_{\mathcal{F}_T}\mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{f\in S_i}(1-\phi)|\mathcal{G}_0\right]d\Pi(f|\delta_{lk}=0) \leqslant 4(2K+1)e^{-\gamma_1 M_T'^2 T\epsilon_T^2},$$

where the S_i's are the slices defined in (S2.7). Therefore, we obtain (S3.16) using that

$$\mathbb{P}_0\left[\Pi(d_{1T} > M'_T\epsilon_T|N) > \Pi(\delta_{lk} = 0|N)/4\right] \leqslant o(1) + 4e^{\kappa'_T T\epsilon_T^2}4(2K+1)e^{-(M'_T)^2 T\epsilon_T^2} = o(1).$$

To prove (S3.17), using Markov's inequality and Fubini's theorem, we have, for M' large enough, that

$$\begin{aligned}
&\mathbb{P}_0\left[\Pi(d_{1T} \leqslant M'_T\epsilon_T,\, S_{lk} > M_T\epsilon_T|N) > \Pi(\delta_{lk} = 0|N)/4\right] \\
&\quad\leqslant \mathbb{P}_0\left[\{m_T(\delta_{lk}=0) < e^{-\kappa'_T T\epsilon_T^2}\}\cap\tilde{\Omega}_T\right] + \mathbb{P}_0\left[\tilde{\Omega}_T^c\right] \\
&\quad + 4e^{\kappa'_T T\epsilon_T^2}\mathbb{E}_0\left[\int_{\mathcal{F}_T\cap\{S_{lk}>M_T\epsilon_T\}}\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{d_{1T}\leqslant M'_T\epsilon_T}e^{L_T(f)-L_T(f_0)}d\Pi(f|\delta_{lk}=0)\right] \\
&\quad = o(1) + 4e^{\kappa'_T T\epsilon_T^2}\int_{S_{lk}>M_T\epsilon_T}\mathbb{E}_0\left[\mathbb{P}_f\left[\tilde{\Omega}_T\cap\{d_{1T}\leqslant M'_T\epsilon_T\}|\mathcal{G}_0\right]\right]d\Pi(f).
\end{aligned}$$

Moreover, from (27), we have that $\sup_{f\in A_{L_1}(M_T\epsilon_T)^c\cap\mathcal{F}_T}\mathbb{P}_f\left[\tilde{\Omega}_T\cap\{d_{1T}\leqslant M'_T\epsilon_T\}|\mathcal{G}_0\right] = o(e^{-\kappa'_T T\epsilon_T^2})$. Finally, since $\delta_{lk}^0 = 0$, $S_{lk} > M_T\epsilon_T$ implies that $f \in A_{L_1}^c(M_T\epsilon_T)$, which thus leads to (S3.17).

Reporting into (S3.15), we now have

$$
\begin{aligned}
\mathbb{P}_0\left[\hat{\delta}_{lk}^{\Pi,L}=1\right] &\leqslant \mathbb{P}_0\left[e^{-c_1' T\epsilon_T^2}\Pi(\delta_{lk}=1,\, S_{lk}\leqslant M_T\epsilon_T|N)\geqslant \Pi(\delta_{lk}=0|N)/2\right]+o(1)\\
&\leqslant \mathbb{P}_0\left[e^{-c_1' T\epsilon_T^2}\Pi(\delta_{lk}=1|N)\geqslant \Pi(\delta_{lk}=0|N)/2\right]+o(1)\\
&= \mathbb{P}_0\left[e^{-c_1' T\epsilon_T^2}m_T(\delta_{lk}=1)\geqslant \frac{\Pi(\delta_{lk}=0)}{2\Pi(\delta_{lk}=1)}m_T(\delta_{lk}=0)\right]+o(1)\\
&\leqslant \mathbb{P}_0\left[\left\{e^{-c_1' T\epsilon_T^2}m_T(\delta_{lk}=1)\geqslant \frac{\Pi(\delta_{lk}=0)}{2\Pi(\delta_{lk}=1)}e^{-\kappa_T' T\epsilon_T^2}\right\}\cap\tilde{\Omega}_T\right]+\mathbb{P}_0\left[m_T(\delta_{lk}=0)<e^{-\kappa_T' T\epsilon_T^2}\right]+o(1)\\
&\leqslant \mathbb{P}_0\left[\left\{m_T(\delta_{lk}=1)\geqslant \frac{\Pi(\delta_{lk}=0)}{2\Pi(\delta_{lk}=1)}e^{(c_1'-\kappa_T')T\epsilon_T^2}\right\}\cap\tilde{\Omega}_T\right]+o(1)\\
&\leqslant \mathbb{E}_0\left[m_T(\delta_{lk}=1)\right]\frac{2\Pi(\delta=1)}{\Pi(\delta_{lk}=0)}e^{-(c_1'-\kappa_T')T\epsilon_T^2}+o(1)\leqslant \frac{2\Pi(\delta_{lk}=1)^2}{\Pi(\delta_{lk}=0)}e^{-(c_1'-\kappa_T')T\epsilon_T^2}+o(1)=o(1),
\end{aligned}
$$

since $c_1'>\kappa_T+c_1=\kappa_T'$ and that $\mathbb{E}_0\left[m_T(\delta=1)\right]=\Pi(\delta_{lk}=1)$ with Fubini's theorem.

- **Case 2:** $(l,k)\in I(\delta_0)$, i.e, $\delta_{lk}^0=1$. In the case, the computations are slightly simpler since $\{\delta_{lk}=0\}\implies f\in A_{L_1}(M\sqrt{\kappa_T}\epsilon_T)^c$ and for T large enough, $S_{lk}^0-M_T\epsilon_T>0$. Thus we can use the fact that $\Pi(\delta_{lk}=0|N)\leqslant\Pi(A_{L_1}(M\sqrt{\kappa_T}\epsilon_T)^c|N)$.

  We first note that if $S_{lk}^0>M_1\sqrt{\kappa_T}\epsilon_T$ with $M_1>M$ and $1-F(S_{lk}^0/2)\geqslant 2e^{-\gamma T\epsilon_T^2}$ for some $\gamma>\kappa_T+c_1=:\kappa_T'$, then if $\delta_{lk}=0$, $f\in A_{L_1}(M_1\sqrt{\kappa_T}\epsilon_T)^c$ and

$$
\Pi(\delta_{lk}=0|N)\leqslant\Pi(A_{L_1}(M_T\epsilon_T)^c|N),\quad\text{ and }\quad S_{lk}^0-M_T\epsilon_T\geqslant S_{lk}^0/2.
$$

Therefore, since F is non-increasing, $F(S_{lk}^0-M_T\epsilon_T)\leqslant F(S_{lk}^0/2)$ and

$$
\begin{aligned}
&\mathbb{P}_0\left[\hat{\delta}_{lk}^{\Pi,L}=0\right]\leqslant\mathbb{P}_0\left[\Pi((1-F(S_{lk}))\mathbb{1}_{\delta=1}(\mathbb{1}_{S_{lk}\geqslant S_{lk}^0-M_T\epsilon_T}+\mathbb{1}_{S_{lk}<S_{lk}^0-M_T\epsilon_T})|N)\leqslant\Pi(A_{L_1}(2M_T\epsilon_T)^c|N)\right]\\
&\leqslant\mathbb{P}_0\left[(1-F(S_{lk}^0/2))\Pi(S_{lk}>S_{lk}^0-M_T\epsilon_T|N)+\Pi((1-F(S_{lk}))\mathbb{1}_{S_{lk}<S_{lk}^0-M_T\epsilon_T})|N)\leqslant\Pi(A_{L_1}(M_1\sqrt{\kappa_T}\epsilon_T)^c|N)\right]\\
&\leqslant\mathbb{P}_0\left[2e^{-\gamma T\epsilon_T^2}\Pi(S_{lk}>S_{lk}^0-M_T\epsilon_T|N)\leqslant\Pi(A_{L_1}(M_1\sqrt{\kappa_T}\epsilon_T)^c|N)\right]\\
&\leqslant\mathbb{P}_0\left[\Pi(S_{lk}>S_{lk}^0-M_T\epsilon_T|N)\leqslant 1/2\right]+\mathbb{P}_0\left[\tilde{\Omega}_T\cap\left\{e^{-\gamma T\epsilon_T^2}\leqslant\Pi(A_{L_1}(M_1\sqrt{\kappa_T}\epsilon_T)^c|N)\right\}\right]+\mathbb{P}_0\left[\tilde{\Omega}_T^c\right].
\end{aligned}
$$

Similar to the first case where $\delta_{lk}^0=0$, we have that $\mathbb{P}_0\left[\tilde{\Omega}_T\cap\left\{e^{-\kappa_T' T\epsilon_T^2}\leqslant\Pi(A_{L_1}(M_T\epsilon_T)^c|N)\right\}\right]=o(1)$, and since $\gamma\geqslant\kappa_T'$,

$$
\begin{aligned}
\mathbb{P}_0\left[\hat{\delta}_{lk}^{\Pi,L}=0\right]&\leqslant\mathbb{P}_0\left[\Pi(S_{lk}>S_{lk}^0-M_T\epsilon_T|N)\leqslant 1/2\right]+o(1)=\mathbb{P}_0\left[\Pi(S_{lk}<S_{lk}^0-M_T\epsilon_T|N)>1/2\right]+o(1)\\
&\leqslant\mathbb{P}_0\left[\tilde{\Omega}_T\cap\{\Pi(A_{L_1}(M_1\sqrt{\kappa_T}\epsilon_T)^c|N)>1/2\}\right]+\mathbb{P}_0\left[\tilde{\Omega}_T^c\right]=o(1),
\end{aligned}
$$

which terminates this proof.

S3.3. Proof of Proposition 3.10

In this section, we prove our posterior consistency result on the posterior distribution in the restricted models, the **All equal model** and **Receiver dependent model**, defined in Section 3.2 of Sulem, Rivoirard and Rousseau (2022).

In the **All equal model**, if $I(\delta_0) \neq \emptyset$ then $\exists (l_1, k_1) \in [K]^2, \delta^0_{l_1 k_1} = 1$, and $h_0 \neq 0$. Consequently, for T large enough,

$$\{f \in \mathcal{F};\ \delta_{l_1k_1} \neq \delta^0_{l_1k_1}\} = \left\{f \in \mathcal{F};\ \delta_{l_1k_1} = 0\right\} \subset \left\{f \in \mathcal{F};\ \|h^0_{l_1k_1} - h_{l_1k_1}\|_1 = \|h_0\|_1\right\} \subset A_{L_1}(M_T\epsilon_T)^c,$$

leading to $\mathbb{E}_0\left[\Pi(\delta_{l_1k_1} \neq \delta^0_{l_1k_1}|N)\right] = o(1)$ using Theorem 3.2. This would hold for the same reasons for any $(l,k) \in I(\delta_0)$. For $(l,k) \notin I(\delta_0)$, we have instead that for T large enough,

$$\begin{aligned}\{f \in \mathcal{F};\ \delta_{lk} \neq \delta^0_{lk}\} = \{f \in \mathcal{F};\ \delta_{lk} = 1\} &\subset \{f \in \mathcal{F};\ \left\|h^0_{lk} - h_{lk}\right\|_1 = \|h\|_1\} \\ &\subset \{f \in \mathcal{F};\ \|h\|_1 + \left\|h^0_{l_1k_1} - h_{l_1k_1}\right\|_1 \geqslant \|h_0\|_1\} \subset A_{L_1}(M_T\epsilon_T)^c,\end{aligned}$$

as soon as $\|h_0\|_1 \geqslant 3M_T\epsilon_T$, since $\|h\|_1 + \left\|h^0_{l_1k_1} - h_{l_1k_1}\right\|_1 \geqslant \|h\|_1 + \|h_0\|_1 \wedge \left\|h - h^0\right\|_1 \geqslant (\|h\|_1 + \|h_0\|_1) \wedge (\|h\|_1 + \|h - h_0\|_1) \geqslant \|h_0\|$. Similarly to the proof of Theorem 3.9 in Section S3.1, we then obtain $\mathbb{E}_0\left[\Pi(\delta \neq \delta_0|N)\right] = o(1)$.

If $I(\delta_0) = \emptyset$, then $\forall (l,k) \in [K]^2, \delta^0_{lk} = 0$, and $h_0 = 0$, and in this case we first show that there exists $C > 0$ such that

$$\mathbb{P}_0\left[\left\{D_T < CT^{-K/2}\right\} \cap \tilde{\Omega}_T\right] = o(1). \tag{S3.18}$$

Since $h_0 = 0$, the log-likelihood function is the one of a K independent homogeneous Poisson PP with parameter r_0, i.e.,

$$L_T(f_0) = L_T(r_0) = \sum_k \log(r^0_k)N^k[0,T) - r^0_kT,$$

with $r^0_k = \phi_k(\nu^0_k)$. Let $\bar{A} = \{f \in \mathcal{F}_T;\ h = 0\}$. For any $f \in \bar{A}$, we also have $L_T(f) = L_T(r_f) = \sum_k \log(r^f_k)N^k[0,T) - r^f_kT$ and the model is also a Poisson PP, which is a regular model, and which parameter is $\phi(\nu)$. Therefore, we have

$$\begin{aligned}L_T(r) - L_T(r_0) &= \sum_k \log(\frac{r_k}{r^0_k})N^k[0,T) - (r^f_k - r^0_k)T \\ &= \sum_k \left[\frac{r^f_k - r^0_k}{r^0_k} - \frac{1}{2}\left(\frac{r^f_k - r^0_k}{r^0_k}\right)^2 + O_{\mathbb{P}_0}(r^f_k - r^0_k)^3\right]N^k[0,T) - (r^f_k - r^0_k)T \\ &= \sum_k \left(\frac{N^k[0,T)}{r^0_k} - T\right)(r^f_k - r^0_k) - \frac{N^k[0,T)}{2}\left(\frac{r^f_k - r^0_k}{r^0_k}\right)^2 + O_{\mathbb{P}_0}(T(r^f_k - r^0_k)^3).\end{aligned}$$

Also, let $\tilde{\pi}_r$ be the prior density of $r^f_k = \phi_k(\nu_k)$ given by $\tilde{\pi}_r(x) = \phi(\nu)\pi_\nu(\nu)$. Note that in the case of partially known link functions of the form $\phi_k(x) = \theta_k + \psi(x)$, the parameter of the Poisson PP is now

(ν, θ) and we can consider a marginal prior density of $r_k^f = \theta_k + \psi(\nu_k)$ given by

$$\tilde{\pi}_r(x) = \int_0^{\psi^{-1}(x)} \pi_\theta(x - \psi(\nu))\pi_\nu(\nu)d\nu.$$

The regularity assumptions on π_ν (and π_θ) and ϕ^{-1} imply that $\tilde{\pi}_r$ is continuous and positive at r_k^0 for all k.

Defining $\bar{A}_T = \bar{A} \cap \{\|r_f - r_0\|_1 \leqslant \epsilon\}$ for $\epsilon > 0$ small enough, we thus have

$$\begin{aligned}
D_T &= \int_{\mathcal{F}_T} e^{L_T(f) - L_T(f_0)} d\Pi(f) \geqslant \int_{\bar{A}_T} e^{L_T(r) - L_T(r_0)} d\Pi(f) \\
&\geqslant \int_{\bar{A}_T} \prod_{k=1}^K \exp\left\{ \left(\frac{N^k[0,T]}{r_k^0} - T \right)(r_k^f - r_k^0) - \frac{N^k[0,T]}{2} \left(\frac{r_k^f - r_k^0}{r_k^0} \right)^2 (1+\epsilon) \right\} \tilde{\pi}(r_k) dr_k \\
&= \prod_{k=1}^K \tilde{\pi}_r(r_k^0)(1 + o_{\mathbb{P}_0}(1)) e^{\frac{r_k^0}{2(1+\epsilon)N^k[0,T)} \left(\frac{N^k[0,T)}{r_k^0} - T \right)^2} \times \\
&\int_{|r_k^f - r_k^0| \leqslant \epsilon/K} \exp\left\{ -\frac{N^k[0,T)}{2(r_k^0)^2}(1-\epsilon) \left(r_k^f - r_k^0 - \frac{(r_k^0)^2}{(1+\epsilon)N^k[0,T)} \left(\frac{N^k[0,T)}{r_k^0} - T \right) \right)^2 \right\} dr_k^f \\
&\geqslant \prod_{k=1}^K \tilde{\pi}_r(r_k^0) r_k^0 \frac{\sqrt{2\pi}}{[N^k[0,T)(1+\epsilon)]^{1/2}} (1 + o_{\mathbb{P}_0}(1)) \geqslant \prod_{k=1}^K \frac{\sqrt{2\pi}\tilde{\pi}_r(r_k^0) r_k^0}{[T(1+\epsilon)]^{1/2}} (1 + o_{\mathbb{P}_0}(1)),
\end{aligned}$$

since $N^k[0,T)$ is a Poisson random variable with parameter $r_k^0 T$ so that $|N^k[0,T)/T - r_k^0| \leqslant M_T/\sqrt{T}$ with probability going to 1 and $\{|r_k^f - r_k^0| \leqslant \epsilon/K\}$ contains the set

$$\left| r_k^f - r_k^0 - \frac{(r_k^0)^2}{(1-\epsilon)N^k[0,T)} \left(\frac{N^k[0,T)}{r_k^0} - T \right) \right| \leqslant \frac{\epsilon}{2K},$$

for T large enough. Therefore we obtain (S3.18) and deduce that $\epsilon_T \lesssim \sqrt{\log T/T}$ using the same arguments as in the proofs of Theorem 3.2. As in Theorem 3.2, it is thus sufficient that

$$\begin{aligned}
&\Pi(\{0 < \|h\|_1 \leqslant M\sqrt{\log T/T}\} \cap \{\max_k |r_k^f - r_k^0| \leqslant M\sqrt{\log T/T}\}) \\
&\leqslant \Pi(\{0 < \|h\|_1 \leqslant M\sqrt{\log T/T}\} \cap \{\max_k |\nu_k - \nu_k^0| \leqslant \frac{M}{L}\sqrt{\log T/T}\}) = o(T^{-K/2}),
\end{aligned}$$

for M large enough which boils down to assuming that

$$\Pi(\{0 < \|h\|_1 \leqslant M\sqrt{\log T/T}\}) = o((\log T)^{-K/2}),$$

to conclude that $\mathbb{E}_0[\Pi(\delta \neq \delta_0|N)] = o(1)$.

In the **Receiver node dependent model**, i.e., $\forall (l,k) \in [K]^2, h_{lk} = \delta_{lk} h_k$, we obtain the result similarly to the **All equal model** since the likelihood is also a product of likelihoods per node:

$$L_T(f) = \sum_{k=1}^K L_T(\nu_k, h_k, \delta(k), \theta_k), \qquad \text{with } \delta(k) := (\delta_{lk}, 1 \leqslant l \leqslant K),$$

$$L_T(\nu_k, h_k, \delta(k), \theta_k) := \sum_{T_i^k} \log \lambda^k_{T_i^k}(f_k) - \int_0^T \lambda^k_t(f_k)dt, \quad f_k = (\nu_k, h_k, \delta(k), \theta_k), \quad k \in [K].$$

If the priors on $(\theta_k, \nu_k, h_k, \delta(k))$ are independent, the posteriors are also independent and we can directly apply the previous result.

S4. Proof of Corollary 3.8

In this section we prove our result on the convergence rate of the posterior mean estimator. In all the considered models with known link functions, the convergence of the posterior mean $\hat{f} = (\hat{\nu}, \hat{h})$ results from the same arguments as in Corollary 1 of Donnet, Rivoirard and Rousseau (2020) (proof in Section 2.3 in the supplementary material). In the case of the shifted ReLU model with unknown shift, we can also use similar computations for $\hat{f} = (\hat{\nu}, \hat{h})$ and $\hat{\theta}$. We first recall some notation from the proofs of Theorem 3.2 and Proposition 3.5: $\bar{A}(\tilde{M}_T\epsilon_T) = \{\theta \in \Theta, \|\theta - \theta_0\|_1 < \tilde{M}_T\epsilon_T\}$, $A_{L_1}(M_T\epsilon_T) = \{(f, \theta) \in \mathcal{F} \times \Theta, \|\theta + \nu - \theta_0 - \nu_0\|_1 + \|h - h_0\|_1 < M_T\epsilon_T\}$ and $\tilde{M}_T = \tilde{M}\sqrt{\kappa_T}$, $M_T = M\sqrt{\kappa_T}$, $\tilde{M} > M > 0$. We note that

$$\left\|\hat{\theta} - \theta_0\right\|_1 \leqslant \tilde{M}_T\epsilon_T + \mathbb{E}^{\Pi}[\|\theta - \theta_0\|_1 \mathbb{1}_{\bar{A}(\tilde{M}_T\epsilon_T)^c}|N].$$

Then, splitting $\bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}_T$ into $\bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}_T \cap A_{L_1}(M_T\epsilon_T)$ and $\bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}_T \cap A_{L_1}(M_T\epsilon_T)^c$, we control $\mathbb{E}^{\Pi}[\|\theta - \theta_0\|_1 \mathbb{1}_{B_T}|N]$ using the following arguments with B_T representing either $\bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}_T \cap A_{L_1}(M_T\epsilon_T)$ or $A_{L_1}(M_T\epsilon_T)^c$. Using the decomposition (24), with $\kappa'_T = \kappa_T + c_1$, we have

$$\begin{aligned}
\mathbb{P}_0\left[\mathbb{E}^{\Pi}[\|\theta - \theta_0\|_1 \mathbb{1}_{B_T}|N] > \epsilon_T\right] &\leqslant \mathbb{E}_0\left[\phi\mathbb{1}_{\tilde{\Omega}_T}\right] + \mathbb{P}_0\left[\{D_T < e^{-\kappa'_T T\epsilon_T^2}\} \cap \tilde{\Omega}_T\right] + \mathbb{P}_0\left[\tilde{\Omega}_T^c\right] + \frac{e^{\kappa'_T T\epsilon_T^2}}{\epsilon_T}\Pi(\mathcal{F}_T^c) \\
&\quad + \frac{e^{\kappa'_T T\epsilon_T^2}}{\epsilon_T}\int_{\mathcal{F}_T \cap B_T} \|\theta - \theta_0\|_1 \mathbb{E}_0\left[\mathbb{E}_f\left[(1-\phi)\mathbb{1}_{\tilde{\Omega}_T}\right]\middle|\mathcal{G}_0\right] d\Pi(f, \theta) \\
&\leqslant o(1) + o\left(\int_{\mathcal{F}_T \cap B_T} \|\theta - \theta_0\|_1 d\Pi(f, \theta)\right) = o(1),
\end{aligned}$$

using the tests defined in Lemma A.4 if $B_T = \bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}_T \cap A_{L_1}(M_T\epsilon_T)$ or the tests defined in Lemma S6.1 if $B_T = A_{L_1}(M_T\epsilon_T)^c$, and also that $\log T = o(T\epsilon_T^2)$ to obtain that $\frac{e^{\kappa'_T T\epsilon_T^2}}{\epsilon_T}\Pi(\mathcal{F}_T^c) \leqslant \Pi(\mathcal{H}_T^c)e^{\kappa'_T T\epsilon_T^2 - \log\epsilon_T} = o(1)$, whichs terminates this proof.

S5. Proofs of some results on prior distributions

In this section, we present an alternative construction of the prior distribution using mixtures of Beta distributions and the proof of Lemma 4.3, which gives one example of model where the condition (8) can be verified.

S5.1. Mixtures of Betas priors

This family of prior distributions can be also considered alongside the ones presented in Section 4 of Sulem, Rivoirard and Rousseau (2022). The following construction is similar to Section 2.3.2 of

Donnet, Rivoirard and Rousseau (2020), which is based on Rousseau (2010). Using the hierarchical structure (15) from Section 4 , we define $\pi_{\tilde{h}}$ as follows. For simplicity, we here consider that $A = 1$. Let

$$\tilde{h}_{\alpha,M}(x) = \int_u g_{\alpha,u}(x) dM(u), \quad g_{\alpha,u}(x) = \frac{\Gamma(\alpha/u(1-u))}{\Gamma(\alpha/u)\Gamma(\alpha/(1-u))} x^{-\alpha/(1-u)-1}(1-x)^{-\alpha/u-1},$$

and $\pi_{\tilde{h}}$ be the push forward distribution of $\Pi_\alpha \times \Pi_M$ by the transformation $(\alpha, M) \to h_{\alpha,M}$, where Π_α and Π_M are respectively the probability distribution on α and M. Therefore $\pi_{\tilde{h}}$ is a bounded signed measure on $[0, 1]$. As in Donnet, Rivoirard and Rousseau (2020), we choose $\sqrt{\alpha}$ to follow a Gamma distribution and define Π_M by

$$M(u) = \sum_{j=1}^{J} r_j p_j \delta_{u_j}(u), \quad u_j \overset{i.i.d.}{\sim} G_0, \quad J \sim \mathcal{P}(\lambda),$$

where G_0 is a base measure and the r_j's are independent Rademacher random variables and $(p_1, \cdots, p_J) \sim \mathcal{D}(a_1, \cdots, a_J)$ with $\sum_{j=1}^{J} a_j \leqslant C$ for some fixed $C > 0$. Note that since $\|\bar{h}_{\alpha,M}\|_1 \leqslant 1$, we can define

$$h_{lk} = \tilde{S}_{lk} \tilde{h}_{lk}, \quad \|\tilde{S}^+\| \leqslant 1, \quad \tilde{h}_{lk} \overset{i.i.d.}{\sim} \pi_{\tilde{h}},$$

so that the prior distribution on h is the push forward distribution of $\pi_{\tilde{h}}^{\otimes|I(\delta)|} \times \pi_{S|\delta}$ by the above transformation, with π_S defined in **(S2)** in Section 4. Since $\tilde{S}$ is a (component-wise) upper bound on the matrix S, $\|\tilde{S}^+\| \leqslant 1$ implies $\|S^+\| \leqslant 1$. We then arrive at the following result.

Corollary S5.1. *Let N be a Hawkes process with link functions $\phi = (\phi_k)_k$ and parameter $f_0 = (\nu_0, h_0)$ such that (ϕ, f_0) verify the conditions of Lemma 2.1, and Assumption 3.1. Under the above spline prior, if the prior on S satisfies the conditions defined in **(S1)** (see Section 4), and also if $\forall (l,k) \in [K]^2$, $h^0_{lk} \in \mathcal{H}(\beta, L)$ with $\beta > 0$ and $\|S_0^+\| < 1$ then for M large enough,*

$$\mathbb{E}_0 \left[\Pi(\|f - f_0\|_1 > M T^{-\beta/(2\beta+1)} \sqrt{\log\log T} (\log T)^q \big| N) \right] = o(1),$$

where $q = 5\beta/(4\beta+2)$ if ϕ verifies Assumption 3.1(i) , and $q = 1/2 + 5\beta/(4\beta+2)$ if ϕ verifies Assumption 3.1(ii) .

S5.2. Proof of Lemma 4.3

Lemma S5.2 (Lemma 4.3)**.** *Let N be a Hawkes process with ReLU link functions $\phi_k(x) = (x)_+, \forall k \in [K]$, and parameter $f_0 = (\nu_0, h_0)$ such that (ϕ, f_0) verify condition **(C1bis)** and for all l, there exists $J_0 \in \mathbb{N}^*$ such that*

$$h^0_{lk}(t) = \sum_{j=1}^{J_0} \omega^{lk}_{j0} \mathbb{1}_{I_j}(t), \quad \omega^{lk}_{j0} \in \mathbb{Q}, \quad \forall j \in [J_0],$$

with $\{I_j\}_{j=1}^{J_0}$ a partition of $[0, A]$. Then, condition (8) of Proposition 3.5 holds, i.e.,

$$\limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_0 \left(\int_0^T \frac{\mathbb{1}_{\lambda^k_t(f_0)>0}}{\lambda^k_t(f_0)} dt \right) < +\infty, \quad k \in [K].$$

Proof. Let f_0 verifying the conditions of the lemma. We first show that there exists $c_0 > 0$ that depends only on the parameters $\{\nu_k^0, \{\omega_{j0}^{kl}\}_{j=1}^J\}_{k,l=1}^K$ such that $\forall k \in [K], \forall t \geqslant 0$, $\lambda_t^k(f_0) > 0 \implies \lambda_t^k(f_0) \geqslant c_0$. We prove here the result for the unidimensional Hawkes model with $K = 1$, but our proof can be easily generalized to $K > 1$. We therefore use the notation ν_0 and w_{j0} for ν_1^0 and w_{j0}^{11}.

Since $w_{j0} \in \mathbb{Q}$, let $p_j, q_j \in \mathbb{Z}$ such that $w_{j0} = p_j/q_j$ and let $q \in \mathbb{Z}$ be the least common multiple of (p_j, q_j). Thus there exists $a_j \in \mathbb{Z}$ such that $\omega_{j0} = a_j/q$. and for any $t \geqslant 0$, let $n_j(t) = \int_{t-A}^t \mathbb{1}_{I_j}(t-s)dN_s$ be the number of events that "activate" the bin j at t. With this notation, we can then write

$$\lambda_t(f_0) = \left(\nu_0 + \sum_{j=1}^{J_0} n_j(t)\frac{a_j}{q}\right)_+ = \left(\nu_0 + \sum_{j=1}^{J_0} n_j(t)\frac{a_j}{q}\right)_+$$
$$= \left(\frac{1}{q}\left[\nu_0 q + \sum_{j=1}^{J_0} n_j(t)a_j\right]\right)_+ .$$

Let $\varepsilon > 0$ such that $\varepsilon = \min_{u\in\mathbb{Z}, \nu_0 q+u>0} \nu_0 q + u$. Then $\varepsilon > 0$ and for any $t \geqslant 0$ such that $\tilde{\lambda}_t(f_0) > 0$, since $\sum_{j=1}^{J_0} n_j(t)a_j \in \mathbb{Z}$, then $\nu_0 q + \sum_{j=1}^{J_0} n_j(t)a_j \geqslant \varepsilon > 0$ and $\lambda_t(f_0) \geqslant \varepsilon/q =: c_0 > 0$, which proves that (i) holds. Therefore, in this model, we have

$$\frac{1}{T}\mathbb{E}_0\left(\int_0^T \frac{\mathbb{1}_{\lambda_t(f_0)>0}}{\lambda_t(f_0)}dt\right) \leqslant \frac{1}{T}\mathbb{E}_0\left(\int_0^T \frac{\mathbb{1}_{\lambda_t(f_0)>0}}{c_0}dt\right) \leqslant \frac{1}{T}\mathbb{E}_0\left(\int_0^T \frac{1}{c_0}dt\right) = \frac{1}{c_0} < +\infty,$$

which proves that (8) is satisfied. □

***Remark* S5.3.** We could similarly show that if also $\forall l \in [K], \forall j \in [J]$, $\nu_k^0 \in \mathbb{R}\backslash\mathbb{Q}$, then there exists $d_0 < 0$ depending on $\{\nu_k^0, \{\omega_{j0}^{kl}\}_{j=1}^J\}_{k,l=1}^K$ such that $\forall k \in [K], \forall t \geqslant 0$, $\lambda_t^k(f_0) = 0 \implies \tilde{\lambda}_t^k(\nu_0, h_0) \leqslant d_0$.

S6. Lemmas on tests

In this section we prove two technical lemmas on the test functions used in the proofs of Theorem 5.5 and Proposition 3.5. In Section S6.1, we state and prove our first lemma, Lemma S6.1, which relates to the elementary test functions used in the proof of Theorem 5.5 (Section S2) and in Section S6.2, we prove Lemma A.5, which provides the bound on the error of the tests used in the proof of Case 2 of Proposition 3.5.

S6.1. Test used in the proof of Theorem 5.5

Lemma S6.1. *For $i \geqslant 1$, let $\mathcal{F}_i = \{f \in \mathcal{F}_T; \nu_l \leqslant \nu_l^0 + 2K\|r_0\|_1 \mathbb{E}_0(\Delta\tau_1) i\epsilon_T,\ \forall l \in [K]\}$ and $f_1 \in \mathcal{F}_i$. We define the test*

$$\phi_{f_1,i} = \max_{l\in[K]} \mathbb{1}_{\{N^l(A_{1l})-\Lambda^l(A_{1l},f_0)\geqslant iT\epsilon_T/8\}} \wedge \mathbb{1}_{\{N^l(A_{1l}^c)-\Lambda^l(A_{1l}^c,f_0)\geqslant iT\epsilon_T/8\}},$$

where for all $l \in [K]$, $A_{1l} = \{t \in [0,T]; \lambda_t^l(f_1) \geqslant \lambda_t^l(f_0)\}$, $\Lambda^l(A_{1l}, f_0) = \int_0^T \mathbb{1}_{A_{1l}}(t)\lambda_t^l(f_0)dt$ and $\Lambda^l(A_{1l}^c, f_0) = \int_0^T \mathbb{1}_{A_{1l}^c}(t)\lambda_t^l(f_0)dt$. Then

$$\mathbb{E}_0[\mathbb{1}_{\tilde{\Omega}_T}\phi_{f_1,i}] + \sup_{\|f-f_1\|_1\leqslant i\epsilon_T/(12N_0)} \mathbb{E}_0\left[\mathbb{E}_f[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{f\in S_i}(1-\phi_{f_1,i})|\mathcal{G}_0]\right] \leqslant (2K+1)\max_{l\in[K]} e^{-x_{1l}Ti\epsilon_T(\sqrt{\mu_l^0}\wedge i\epsilon_T)},$$

where for $l \in [K]$*,* $x_{1l} > 0$ *is an absolute constant,* $\mu_l^0 = \mathbb{E}_0\left[\lambda_t^l(f_0)\right]$*,* $N_0 = 1 + \sum_{l=1}^K \mu_l^0$ *and* S_i *is defined in* (S2.7)*.*

Proof. For $l \in [K]$, let

$$\phi_{il} = \phi_{il}(f_1) = \mathbb{1}_{\{N^l(A_{1l}) - \Lambda^l(A_{1l}, f_0) \geqslant iT\epsilon_T/8\}}.$$

Mimicking the proof of Lemma 1 of Donnet, Rivoirard and Rousseau (2020), we obtain that

$$\mathbb{E}_0\left[\phi_{il}\mathbb{1}_{\tilde{\Omega}_T}\right] \leqslant e^{-x_1 iT\epsilon_T \min(\sqrt{\mu_l^0}, i\epsilon_T)}. \tag{S6.19}$$

We first consider the event $\{\Lambda^l(A_{1l}, f_1) - \Lambda^l(A_{1l}, f_0) \geqslant \Lambda^l(A_{1l}^c, f_1) - \Lambda^l(A_{1l}^c, f_0)\}$. Let $f \in \mathcal{F}_i$ such that $\|f - f_1\|_1 \leqslant \zeta i\epsilon_T$ with $\zeta = 1/(6N_0)$ and $N_0 = 1 + \sum_l \mu_l^0$. On $\tilde{\Omega}_T$, using that ϕ_l is L-Lipschitz for any l, we have

$$\begin{aligned}
T\tilde{d}_{1T}(f, f_1) &= \sum_{l=1}^K \int_0^T \mathbb{1}_{A_2(T)}(t)|\lambda_t^l(f) - \lambda_t^l(f_1)|dt \leqslant \sum_{l=1}^K \int_0^T |\lambda_t^l(f) - \lambda_t^l(f_1)|dt \\
&\leqslant L\sum_{l=1}^K \int_0^T |\tilde{\lambda}_t^l(\nu, h) - \tilde{\lambda}_t^l(\nu_1, h_1)|dt \\
&\leqslant TL\sum_{l=1}^K |\nu_l - \nu_l^1| + L\sum_{l=1}^K\sum_{k=1}^K \int_0^T \int_{t-A}^t |(h_{kl} - h_{kl}^1)(t-s)|N^k(ds) \\
&\leqslant TL\|\nu - \nu_1\|_1 + \max_l N^l[-A, T]L\sum_{l=1}^K\sum_{k=1}^K \|h_{kl} - h_{kl}^1\|_1 \\
&\leqslant LN_0T\|f - f_1\|_1 \leqslant LN_0T\zeta i\epsilon_T.
\end{aligned}$$

Moreover, since $f \in S_i$, on $\tilde{\Omega}_T$, we also have that

$$\int_0^T \mathbb{1}_{A_2(T)}\lambda_t^l(f)dt \leqslant \int_0^T \mathbb{1}_{A_2(T)}\lambda_t^l(f_0)dt + KT(i+1)\epsilon_T \leqslant 2T\mu_l^0 + KT(i+1)\epsilon_T =: \tilde{v}.$$

Applying again inequality (7.7) of Hansen, Reynaud-Bouret and Rivoirard (2015) with $v = \tilde{v}$ and using the computations of Donnet, Rivoirard and Rousseau (2020), we arrive at

$$\mathbb{E}_f\left[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{f\in S_i}(1 - \phi_{il})\Big|\mathcal{G}_0\right] \leqslant 2Ke^{-x_{1l}iT\epsilon_T \min(\sqrt{\mu_l^0}, i\epsilon_T)},$$

for some $x_{1l} > 0$. We can obtain similar results for

$$\phi'_{il} = \mathbb{1}_{\{N^l(A_{1l}^c) - \Lambda^l(\bar{A}_{1l}^c, f_0) \geqslant iT\epsilon_T/8\}}.$$

Finally, with $\phi_{f_1,i} = \max_{l\in[K]} \phi_{il} \wedge \phi'_{il}$, we arrive at the final results of this lemma:

$$\mathbb{E}_0\left[\phi_{f_1,i}\mathbb{1}_{\tilde{\Omega}_T}\right] \leqslant \max_l e^{-x_{1l}iT\epsilon_T \min(\sqrt{\mu_l^0}, i\epsilon_T)} \leqslant e^{-(\min_l x_{1l})iT\epsilon_T \min(\sqrt{\mu_l^0}, i\epsilon_T)}$$

$$\mathbb{E}_f[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{f\in S_i}(1 - \phi_{f_1,i})|\mathcal{G}_0] \leqslant \min_l \mathbb{E}_f[\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{f\in S_i}(1 - \phi_{il})|\mathcal{G}_0] \leqslant 2Ke^{-(\min_l x_{1l})iT\epsilon_T \min(\sqrt{\mu_l^0}, i\epsilon_T)}.$$

□

S6.2. Proof of Lemma A.5

In Lemma A.5, we establish the bound on the type I and type II errors of the tests to estimate the parameter θ in the shifted ReLU link function considered in Case 2 of Proposition 3.5.

We recall that $\Theta = \mathbb{R}_+\backslash\{0\}^K$ and $\bar{A}(R) = \{\theta \in \Theta;\ \|\theta - \theta_0\|_1 \leqslant R\}$. Let $\zeta > 0$ and

$$(f_1, \theta_1) = (\nu_1, h_1, \theta_1) = ((\nu_k^1)_k, (h_{lk}^1)_{l,k}, (\theta_k^1)_k) \in (\bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}) \cap A_{L_1}(M_T\epsilon_T),$$

with $\tilde{M}_T = \tilde{M}\sqrt{\kappa_T}$, $M_T = M\sqrt{\kappa_T}$ and $\tilde{M} \geqslant M$. Let $(f, \theta) \in (\bar{A}(\tilde{M}_T\epsilon_T)^c \times \mathcal{F}) \cap A_{L_1}(M_T\epsilon_T)$ such that $\|f - f_1\|_1 \leqslant \zeta\epsilon_T$, i.e,

$$\sum_k |\nu_k - \nu_k^1| + |\theta_k - \theta_k^1| + \sum_{l,k} \left\|h_{lk} - h_{lk}^1\right\|_1 \leqslant \zeta\epsilon_T.$$

Since $\theta \in \bar{A}(\tilde{M}_T\epsilon_T)^c$, there exists $k \in [K]$ such that $|\theta_k^0 - \theta_k| \geqslant \tilde{M}_T\epsilon_T/K$. For this k, from assumption (S8.46), there exists $l \in [K]$ and $x_1, x_2, c_\star > 0$ such that $\forall x \in [x_1, x_2]$, $h_{lk}^0(x) \leqslant -c_\star < 0$.

We first consider the case $\theta_k < \theta_k^0 - \tilde{M}_T\epsilon_T/K$ and recall the notation of Section S1: $\delta' = (x_2 - x_1)/3$, $n_1 = \lfloor 2\nu_k^1/(\kappa_1 c_\star)\rfloor + 1$ for some $\kappa_1 \in (0, 1)$ and the subset of excursions

$$\mathcal{E} = \{j \in [J_T];\ N[\tau_j, \tau_j + \delta') = N^l[\tau_j, \tau_j + \delta') = n_1,\ N[\tau_j + \delta',\ \tau_{j+1}) = 0\}.$$

We recall that

$$I_k^0(f_1, \theta_1) = \left\{t \in [0, T];\ \lambda_t^k(f_1, \theta_1) = \theta_k^1,\ \lambda_t^k(f_0, \theta_0) = \theta_k^0\right\},$$

and we first state a preliminary lemma on $I_k^0(f_1, \theta_1)$, which is proved at the end of this proof.

Lemma S6.2. *In the Hawkes model with shifted ReLU link function , for any $f_0 \in \mathcal{F}$ such that (S8.46) is satisfied and any $(f_1, \theta_1) \in (\bar{A}(\tilde{M}_T\epsilon_T)^c \times \Theta) \cap A_{L_1}(M_T\epsilon_T)$, on $\tilde{\Omega}_T'$, it holds that*

$$|I_k^0(f_1, \theta_1)| \geqslant \frac{x_2 - x_1}{2} \sum_{j \in [J_T]} \mathbb{1}_{j \in \mathcal{E}},$$

with $\mathcal{E}$ defined in (S1.4).

Let

$$\phi_k(f_1, \theta_1) := \mathbb{1}_{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f_0) < -v_T} \vee \mathbb{1}_{|\mathcal{E}| < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}},$$

with $\Lambda_k(I_k^0(f_1, \theta_1), f_0) = \int_0^T \mathbb{1}_{I_k^0(f_1,\theta_1)}(t)\lambda_t^k(f_0)dt$, $p_0 = \mathbb{P}_0[j \in \mathcal{E}]$, $v_T = w_T T\epsilon_T > 0$ with $w_T > 0$ chosen later. We have by definition

$$\mathbb{E}_0\left[\phi_k(f_1, \theta_1)\mathbb{1}_{\tilde{\Omega}_T'}\right] \leqslant \mathbb{P}_0\left[\left\{|\mathcal{E}| < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}\right\} \cap \tilde{\Omega}_T'\right] + \mathbb{P}_0\left[\left\{N^k(I_k^0(f_1, \theta_1)) - \Lambda_k(I_k^0(f_1, \theta_1), f_0) < -v_T\right\} \cap \tilde{\Omega}_T'\right]. \tag{S6.20}$$

For the first term on the RHS of (S6.20), we apply Hoeffding's inequality with $X_j = \mathbb{1}_{j \in \mathcal{E}} \overset{i.i.d.}{\sim} \mathcal{B}(p_0)$:

$$\mathbb{P}_0\left[\left\{|\mathcal{E}| < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}\right\} \cap \tilde{\Omega}'_T\right] \leqslant \mathbb{P}_0\left[\left\{\sum_{j=1}^{J_T} X_j < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}\right\} \cap \tilde{\Omega}'_T\right]$$

$$\leqslant \mathbb{P}_0\left[\sum_{j=1}^{T/(2\mathbb{E}_0[\Delta\tau_1])} X_j < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}\right] \lesssim e^{-\frac{T p_0^2}{8\mathbb{E}_0[\Delta\tau_1]}} = o(e^{-u_0 T \epsilon_T^2}),$$

for $u_0 < p_0^2/(8\mathbb{E}_0[\Delta\tau_1])$ and using that on $\tilde{\Omega}'_T$, $J_T > T/(2\mathbb{E}_0[\Delta\tau_1])$. For the second term of the RHS of (S6.20), we apply inequality (7.7) in Hansen, Reynaud-Bouret and Rivoirard (2015), with $H_t = \mathbb{1}_{I_k^0(f_1,\theta_1)}(t)$, $H_t^2 \circ \Lambda_t^k(f_0) = \int_0^T \mathbb{1}_{I_k^0(f_1,\theta_1)}(t)\theta_k^0 dt = \theta_k^0 |I_k^0(f_1,\theta_1)| \leqslant \theta_k^0 T$, $x = x_3 T\epsilon_T^2$, $x_3 > 0$. If $\sqrt{2\theta_k^0 T x} + x/3 \leqslant w_T T\epsilon_T$ and $x_3 > u_0$, then by (7.7) of Hansen, Reynaud-Bouret and Rivoirard (2015),

$$\mathbb{P}_0\left[\left\{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f_0) < -v_T\right\} \cap \tilde{\Omega}'_T\right] \leqslant e^{-x_3 T\epsilon_T^2} = o(e^{-u_0 T\epsilon_T^2}).$$

Reporting into (S6.20), we obtain that $\mathbb{E}_0\left[\phi_k(f_1)\mathbb{1}_{\tilde{\Omega}'_T}\right] = o(e^{-u_0 T\epsilon_T^2})$, which proves the first part of Lemma A.5. To prove the second part of Lemma A.5, we first note that

$$\mathbb{E}_f\left[(1-\phi_k(f_1,\theta_1))\mathbb{1}_{\tilde{\Omega}'_T}\right] = \mathbb{P}_f\left[\left\{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f_0) \geqslant -v_T\right\} \cap \left\{|\mathcal{E}| \geqslant \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}\right\} \cap \tilde{\Omega}'_T\right]. \tag{S6.21}$$

We also have

$$\Lambda_k(I_k^0(f_1,\theta_1), f_0) - \Lambda_k(I_k^0(f_1,\theta_1), f) = \Lambda_k(I_k^0(f_1,\theta_1), f_0) - \Lambda_k(I_k^0(f_1,\theta_1), f_1) \tag{S6.22}$$

$$+ \Lambda_k(I_k^0(f_1,\theta_1), f_1) - \Lambda_k(I_k^0(f_1,\theta_1), f). \tag{S6.23}$$

Firstly, if $|\mathcal{E}| > \frac{p_0}{2\mathbb{E}_0[\Delta\tau_1]}T$, then from Lemma S6.2,

$$|I_k^0(f_1,\theta_1)| \geqslant \frac{(x_2 - x_1)p_0}{4\mathbb{E}_0[\Delta\tau_1]}T \tag{S6.24}$$

and

$$\Lambda_k(I_k^0(f_1,\theta_1), f_0) - \Lambda_k(I_k^0(f_1,\theta_1), f_1) = (\theta_k^0 - \theta_k^1)|I_k^0(f_1,\theta_1)| \geqslant \frac{(x_2 - x_1)p_0}{8K\mathbb{E}_0[\Delta\tau_1]}\tilde{M}_T T\epsilon_T, \tag{S6.25}$$

since $\|\theta - \theta_1\|_1 \leqslant \zeta\epsilon_T$ therefore $\theta_k^0 - \theta_k^1 \geqslant |\theta_k^0 - \theta_k| - |\theta_k - \theta_k^1| \geqslant \tilde{M}_T\epsilon_T/K - \zeta\epsilon_T \geqslant \frac{\tilde{M}_T}{2K}\epsilon_T$ for T large enough.

Secondly, since $\forall t \in I_k^0(f_1,\theta_1)$, $\tilde{\lambda}_t^k(\nu_1,h_1) \leqslant 0$ and $\tilde{\lambda}_t^k(\nu,h) \leqslant 0$, we have

$$\Lambda_k(I_k^0(f_1,\theta_1), f_1) - \Lambda_k(I_k^0(f_1,\theta_1), f) = (\theta_k^1 - \theta_k)|I_k^0(f_1,\theta_1)| - \int_{I_k^0(f_1,\theta_1)} \left((\tilde{\lambda}_t^k(\nu,h))_+ - (\tilde{\lambda}_t^k(\nu_1,h_1))_+\right)dt$$

$$\geqslant (\theta_k^1 - \theta_k)|I_k^0(f_1,\theta_1)| - \int_{I_k^0(f_1,\theta_1)} |\tilde{\lambda}_t^k(\nu,h) - \tilde{\lambda}_t^k(\nu_1,h_1)|dt$$

$$\geqslant -\zeta T\epsilon_T - \int_0^T |\tilde{\lambda}_t^k(\nu,h) - \tilde{\lambda}_t^k(\nu_1,h_1))|dt, \tag{S6.26}$$

where we have used the fact that by definition $|I_k^0(f_1,\theta_1)| \leqslant T$. Using Fubini's theorem, for any $l \in [K]$, we have

$$\begin{aligned}
\int_0^T |\tilde{\lambda}_t^k(\nu,h) - \tilde{\lambda}_t^k(\nu_1,h_1)|dt &= \int_0^T \left| \nu_k - \nu_k^1 + \sum_l \int_{t-A}^{t^-} (h_{lk} - h_{lk}^1)(t-s)dN_s^l \right| dt \\
&\leqslant T|\nu_k - \nu_k^1| + \sum_l \int_{T-A}^{T} \int_s^{s+A} |h_{lk} - h_{lk}^1|(t)|dtdN_s^l = T|\nu_k - \nu_k^1| + \sum_l \left\| h_{lk} - h_{lk}^1 \right\|_1 N^l[-A,T] \\
&\leqslant T\left\| f - f_1 \right\| \left(1 + \sum_l (\mu_l^0 + \delta_T) \right) \leqslant \zeta T \epsilon_T \left(1 + 2\sum_l \mu_l^0 \right), \qquad \text{(S6.27)}
\end{aligned}$$

using the definition of $\tilde{\Omega}_T'$ in Section 5.2. Consequently, reporting the previous upper bound into (S6.26), we obtain

$$\Lambda_k(I_k^0(f_1,\theta_1), f_1) - \Lambda_k(I_k^0(f_1,\theta_1), f) \geqslant -\zeta T \epsilon_T (2 + 2\sum_l \mu_l^0).$$

Therefore, using now (S6.25) and (S6.26) in (S6.22), we arrive at

$$\Lambda_k(I_k^0(f_1,\theta_1), f_0) - \Lambda_k(I_k^0(f_1,\theta_1), f) \geqslant \frac{\tilde{M}_T(x_2-x_1)p_0}{8K\mathbb{E}_0[\Delta\tau_1]} T\epsilon_T - \zeta T\epsilon_T(2 + 2\sum_l \mu_l^0) \geqslant \frac{\tilde{M}_T(x_2-x_1)p_0}{16K\mathbb{E}_0[\Delta\tau_1]} T\epsilon_T,$$

since for T large enough, $\tilde{M}_T > \frac{16K\zeta\mathbb{E}_0[\Delta\tau_1](2+2\sum_l \mu_l^0)}{(x_2-x_1)p_0}$. Reporting into (S6.21), we obtain

$$\begin{aligned}
&\mathbb{P}_f\left[\left\{ N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f_0) \geqslant -v_T \right\} \cap \left\{ |\mathcal{E}| \geqslant \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]} \right\} \cap \tilde{\Omega}_T' \right] \\
&\leqslant \mathbb{P}_f\left[\{ N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f) \geqslant -v_T + \frac{\tilde{M}_T(x_2-x_1)p_0}{16\mathbb{E}_0[\Delta\tau_1]} T\epsilon_T \} \cap \tilde{\Omega}_T' \right] \\
&\leqslant \mathbb{P}_f\left[\{ N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f) \geqslant v_T \} \cap \tilde{\Omega}_T' \right],
\end{aligned}$$

if $\tilde{M}_T > \frac{16 w_T \mathbb{E}_0[\Delta\tau_1]}{(x_2-x_1)p_0}$, which is true for $\tilde{M}$ large enough (recall that $\tilde{M}_T = \tilde{M}\sqrt{\kappa_T}$) if $w_T \leqslant C\sqrt{\kappa_T}$ with $C > 0$ a constant.

Similarly to the proof of Lemma 1 in Donnet, Rivoirard and Rousseau (2020), we can adapt inequality (7.7) from Hansen, Reynaud-Bouret and Rivoirard (2015) with $H_t = \mathbb{1}_{I_k^0(f_1,\theta_1)}(t)$ to the conditional probability $\mathbb{E}_f[.|\mathcal{G}_0]$ and the supermartingale $\int_0^T \mathbb{1}_{I_k^0(f_1,\theta_1)}(t)(dN_t - \lambda_t^k(f,\theta)dt)$. With $\tau = T$, $x_T = x_1 T\epsilon_T^2$, we obtain

$$\mathbb{P}_f\left[\{ N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1), f) > v_T \} \cap \tilde{\Omega}_T' \right] \leqslant e^{-x_T T\epsilon_T^2} = o(e^{-(\kappa_T + c_1)T\epsilon_T^2}), \quad \text{if } x_T > \kappa_T + c_1. \qquad \text{(S6.28)}$$

For this to be true, we also need $v_T > \sqrt{2\tilde{v}(\kappa_T + c_1)T\epsilon_T^2} + (\kappa_T + c_1)T\epsilon_T^2/3$ where $\tilde{v}$ is an upper bound of $H_t^2 \circ \Lambda_t^k(f)$. Using the fact that $\forall t \in I_k^0(f_1,\theta_1)$, $\tilde{\lambda}_t^k(\nu_1,h_1) \leqslant 0$, we have

$$H_t^2 \circ \Lambda_t^k(f) = \int_{I_k^0(f_1,\theta_1)} \lambda_t^k(f,\theta)dt = \theta_k |I_k^0(f_1,\theta_1)| + \int_{I_k^0(f_1,\theta_1) \cap \{\tilde{\lambda}_t^k(\nu,h) > 0\}} \tilde{\lambda}_t^k(\nu,h)dt$$

$$\leqslant \theta_k |I_k^0(f_1,\theta_1)| + \int_{I_k^0(f_1,\theta_1)\cap\{\tilde{\lambda}_t^k(\nu,h)>0\}} |\tilde{\lambda}_t^k(\nu,h) - \tilde{\lambda}_t^k(\nu_1,h_1)| dt$$

$$\leqslant \theta_k |I_k^0(f_1,\theta_1)| + \zeta T\epsilon_T \left(1 + 2\sum_l \mu_l^0\right) \leqslant T(\theta_k + \tilde{M}_T\epsilon_T/K) \leqslant \theta_k^0 T =: \tilde{v},$$

using (S6.27) and since for T large enough, $\zeta K(1+2\sum_l \mu_l^0) < M_T \leqslant \tilde{M}_T$. Consequently, if $w_T > \sqrt{2\theta_k^0(\kappa_T + c_1)} + (\kappa_T + c_1)\epsilon_T/3$ and $w_T \leqslant C\sqrt{\kappa_T}$ (which is possible since $\epsilon_T = o(1/\sqrt{\kappa_T})$ by assumption), then (S6.28) holds and we can finally conclude that $\mathbb{E}_f\left[(1-\phi_k(f_1,\theta_1))\mathbb{1}_{\tilde{\Omega}'_T}\right] = o(e^{-(\kappa_T+c_1)T\epsilon_T^2})$ is verified, which leads to the second part of Lemma A.5.

In the alternative case where $\theta_k > \theta_k^0 + \tilde{M}_T\epsilon_T/K$, similar arguments can be applied with $I_k^0(f_1,\theta_1)$ defined as in (S1.3) and $\mathcal{E}$ defined as in (S1.4) except that $n_1 = \lfloor 2\nu_k^0/(\kappa_1 c_\star)\rfloor + 1$. We then use the following test, with $v_T = w_T T\epsilon_T$

$$\phi_k(f_1,\theta_1) := \mathbb{1}_{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1),f_0) > v_T} \vee \mathbb{1}_{|\mathcal{E}| < \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}}.$$

Then Hoeffding's inequality and inequality (7.7) from Hansen, Reynaud-Bouret and Rivoirard (2015) lead to $\mathbb{E}_0\left[\phi_k(f_1,\theta_1)\mathbb{1}_{\tilde{\Omega}_T}\right] = o(e^{-u_0 T\epsilon_T^2})$. For the second part of Lemma A.5, we first note that in this case, since $\forall t \in I_k^0(f_1,\theta_1)$, $\lambda_t^k(f,\theta) \geqslant \theta_k$ (and also $\lambda_t^k(f_0,\theta_0) = \theta_k^0$, $\lambda_t^k(f_1,\theta_1) = \theta_k^1$), then on the event $|\mathcal{E}| \geqslant \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}$,

$$\Lambda_k(I_k^0(f_1,\theta_1),f_0) - \Lambda_k(I_k^0(f_1,\theta_1),f) \leqslant (\theta_k^0 - \theta_k^1)|I_k^0(f_1,\theta_1)| + (\theta_k^1 - \theta_k)|I_k^0(f_1,\theta_1)|$$

$$\leqslant (-\tilde{M}_T\epsilon_T/K + \zeta\epsilon_T)|I_k^0(f_1,\theta_1)| \leqslant -\frac{\tilde{M}_T\epsilon_T|I_k^0(f_1,\theta_1)|}{2K} \leqslant -\frac{(x_2-x_1)p_0}{8K\mathbb{E}_0[\Delta\tau_1]}\tilde{M}_T T\epsilon_T,$$

for T large enough and using (S6.24). Consequently,

$$\mathbb{P}_f\left[\{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1),f_0) \leqslant v_T\} \cap \left\{|\mathcal{E}| \geqslant \frac{p_0 T}{2\mathbb{E}_0[\Delta\tau_1]}\right\} \cap \tilde{\Omega}'_T\right]$$

$$\leqslant \mathbb{P}_f\left[\{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1),f) \leqslant v_T - \frac{(x_2-x_1)p_0}{8\mathbb{E}_0[\Delta\tau_1]}\tilde{M}_T T\epsilon_T\} \cap \tilde{\Omega}'_T\right]$$

$$\leqslant \mathbb{P}_f\left[\{N^k(I_k^0(f_1,\theta_1)) - \Lambda_k(I_k^0(f_1,\theta_1),f) \leqslant -v_T\} \cap \tilde{\Omega}'_T\right],$$

if $\tilde{M}_T > \frac{16K\mathbb{E}_0[\Delta\tau_1]}{(x_2-x_1)p_0} w_T$. Applying inequality (7.7) from Hansen, Reynaud-Bouret and Rivoirard (2015), we can finally obtain

$$\mathbb{E}_f\left[(1-\phi_k(f_1,\theta_1))\mathbb{1}_{\tilde{\Omega}'_T}\right] = o(e^{-(\kappa_T+c_1)T\epsilon_T^2}),$$

which ends the proof of Lemma A.5.

Proof of Lemma S6.2

Let $(f_0,\theta_0) \in \mathcal{F} \times \Theta$, $(f_1,\theta_1) \in (\mathcal{F} \times \bar{A}(\tilde{M}_T\epsilon_T)^c) \cap A_{L_1}(M_T\epsilon_T)$ and $k \in [K]$ such that $|\theta_k^1 - \theta_k^0| > \tilde{M}_T\epsilon_T/K$. For this k, from assumption (S8.46), there exists $l \in [K]$ and $x_1, x_2, c_\star > 0$ such that $\forall x \in [x_1,x_2]$, $h_{lk}^0(x) \leqslant -c_\star < 0$. We first consider the case $\theta_k^1 < \theta_k^0 - \tilde{M}_T\epsilon_T/K$. Since $(f_1,\theta_1) \in A_{L_1}(M_T\epsilon_T)$, we also have that $|\theta_k^1 + \nu_k^1 - \theta_k^0 - \nu_k^0| \leqslant M_T \epsilon_T$, which implies that $\nu_k^1 > \nu_k^0 - (M_T - \tilde{M}_T/K)\epsilon_T > \nu_k^0/2$. For $0 < \kappa_1 < 1$, we define

$$B_1 = \{x \in [0, A];\ h_{lk}^{1-}(x) > \kappa_1 c_\star\}, \quad n_1 = \left\lfloor \frac{2\nu_k^1}{\kappa_1 c_\star} \right\rfloor + 1.$$

Moreover, since $\left\|h_{lk}^0 - h_{lk}^1\right\|_1 \leqslant M_T \epsilon_T$ and $h_{lk}^{0-}(x) \geqslant c_\star$ for $x \in [x_1, x_2]$,

$$|[x_1, x_2] \cap B_1^c| c_\star (1 - \kappa_1) \leqslant \int_{[x_1,x_2]\cap B_1^c} (h_{lk}^1 - h_{lk}^0)(x) dx \leqslant M_T \epsilon_T$$
$$\implies |[x_1, x_2] \cap B_1| \geqslant (x_2 - x_1) - \frac{M_T \epsilon_T}{c_\star (1 - \kappa_1)} \geqslant 3(x_2 - x_1)/4,$$

for T large enough.

Now let $\delta' = (x_2 - x_1)/4$. For $j \in \mathcal{E}$, we denote $T_1, \dots, T_{n_1}$ the n_1 events occurring on $[\tau_j, \tau_j + \delta']$. For $t \in [\tau_j + x_1 + \delta', \tau_j + x_2]$, we have $t - T_i \in [x_1, x_2]$ for any $i \in [n_1]$ and

$$\tilde{\lambda}_t^k(\nu_0, h_0) = \nu_k^0 + \sum_{i\in[n_1]} h_{lk}^0(t - T_i) < \nu_k^0 - n_1 c_\star < 2\nu_k^1 - n_1 \kappa_1 c_\star < 0,$$

by definition of n_1. Similarly, for $t \in B_1 + [\tau_j, \tau_j + \delta']$, we have $t - T_i \in B_1$ and therefore

$$\tilde{\lambda}_t(\nu_1, h_1) = \nu_k^1 + \sum_{i\in[n_1]} h_{lk}^1(t - T_i) < 2\nu_k^1 - n_1 \kappa_1 c_\star < 0.$$

Consequently, for $t \in ([x_1, x_2] \cap B_1) + [\tau_j, \tau_j + \delta']$, $\lambda_t^k(f_0, \theta_0) = \theta_k^0$ and $\lambda_t^k(f_1, \theta_1) = \theta_k^1$, and thus $([x_1, x_2] \cap B_1) + [\tau_j, \tau_j + \delta'] \subset I_k^0(f_1, \theta_1)$. Moreover, we have

$$\left|([x_1, x_2] \cap B_1) + [\tau_j, \tau_j + \delta']\right| \geqslant 3(x_2 - x_1)/4 - (x_2 - x_1)/4 \geqslant (x_2 - x_1)/2.$$

Consequently,

$$|I_k^0(f_1, \theta_1)| = \sum_{j=0}^{J_T} [\tau_j, \tau_{j+1}] \cap \{t \geqslant 0;\ \lambda_t^k(f_0, \theta_0) = \theta_0,\ \lambda_t^k(f_1, \theta_1) = \theta_1\} \geqslant \sum_{j\in[J_T]} \frac{x_2 - x_1}{2} \mathbb{1}_{j\in\mathcal{E}}.$$

In the alternative case $\theta_k^1 > \theta_k^0 + \tilde{M}_T \epsilon_T / K$, similar computations can be derived by defining n_1 as $n_1 = \min\{n \in \mathbb{N};\ n\kappa_1 c_\star > \nu_k^0\}$.

S7. Lemmas on $L_T(f_0) - L_T(f)$

For $f_0, f \in \mathcal{F}$, we define the Kullback-Leibler (KL) divergence in the Hawkes model as

$$KL(f_0, f) = \mathbb{E}_0[L_T(f_0) - L_T(f)]. \tag{S7.29}$$

With a slight abuse of notation, we still use the same notations $L_T(f_0), L_T(f), KL(f_0, f)$ in the nonlinear model with shifted ReLU link function with the additional shift parameter θ. We also note that with the standard ReLU link function, the KL divergence can be infinite for some $f \in \mathcal{F}$, e.g., if there exists $t \in [0, T]$ such that $dN_t^k = 1$ and $\lambda_t^k(f) = 0$. However, in this model, for any $f \in B_\infty(\epsilon_T)$, $\lambda_t^k(\nu, h) \geqslant \lambda_t^k(\nu_0, h_0)$, which implies that $KL(f_0, f) < +\infty$. The next lemma provides some upper bound on the KL divergence on $B_\infty(\epsilon_T)$ with all the link functions considered in Theorem 3.2 and Proposition 3.5.

S7.1. Lemma to bound the Kullback - Leibler divergence

Lemma S7.1. *Under the assumptions of Theorem 3.2 and of Case 2 of Proposition 3.5, for any $f \in B_\infty(\epsilon_T)$ and T large enough,*

$$0 \leqslant KL(f_0, f) \leqslant \kappa_1 T \epsilon_T^2,$$

and, under the assumptions of Case 1 of Proposition 3.5, we similarly have

$$0 \leqslant KL(f_0, f) \leqslant \kappa_2 (\log T)^2 T,$$

with $\kappa_1, \kappa_2 > 0$ constants that only depends on $(\phi_k)_k$ and f_0.

***Remark* S7.2.** For the models considered in Theorem 3.2 and with the shifted ReLU link function (Case 2 of Proposition 3.5), for $f \in B_2(\epsilon_T, B)$, we instead obtain

$$0 \leqslant KL(f_0, f) \lesssim (\log\log T) T \epsilon_T^2.$$

Moreover, with the standard ReLU link function (Case 1 of Proposition 3.5), without assuming that the additional condition (8) holds, we can also obtain the sub-obtimal bound

$$0 \leqslant KL(f_0, f) \lesssim T\epsilon_T,$$

which would also lead to the sub-optimal posterior concentration rate $\sqrt{\epsilon_T}$.

Proof. For simplicity of exposition, throughout this proof, we use the notation $\lambda_t^k(f), \lambda_t^k(f_0)$ for the intensity in all models, therefore including the case $\lambda_t^k(f,\theta), \lambda_t^k(f_0,\theta_0)$ (Case 2 of Proposition 3.5).

Firstly, similarly to the proof of Lemma 2 of Donnet, Rivoirard and Rousseau (2020), we can easily prove that $KL(f_0, f) \geqslant 0$. Secondly, since intensities are predictable, we have

$$\mathbb{E}_0\left[\int_0^T \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)(dN_t^k - \lambda_t^k(f_0)dt)\right] = 0. \tag{S7.30}$$

Since

$$KL(f_0, f) = \sum_k \mathbb{E}_0\left[\int_0^T \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)dN_t^k + \int_0^T (\lambda_t^k(f) - \lambda_t^k(f_0))dt\right], \tag{S7.31}$$

then, with

$$R_T = \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_0^T \lambda_t^k(f_0)\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)dt\right] + \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_0^T (\lambda_t^k(f) - \lambda_t^k(f_0))dt\right], \tag{S7.32}$$

$$KL(f_0, f) = \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\left(\int_0^T \lambda_t^k(f_0)\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)dt + \int_0^T (\lambda_t^k(f) - \lambda_t^k(f_0))dt\right)\right] + R_T. \tag{S7.33}$$

We first show that $R_T = o(T\epsilon_T^2)$. For the first term on the RHS of (S7.32), if $f \in B_\infty(\epsilon_T)$, we use that $\log x \leqslant x - 1$ for $x \geqslant 1$ and we have

$$\sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_0^T \log\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\lambda_t^k(f_0)dt\right] \leqslant \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_0^T \mathbb{1}_{\lambda_t^k(f) > \lambda_t^k(f_0)}\log\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\lambda_t^k(f_0)dt\right]$$

$$
\begin{aligned}
&\leqslant \sum_k \mathbb{E}_0\left[\int_0^T \mathbb{1}_{\tilde{\Omega}_T^c}\mathbb{1}_{\lambda_t^k(f_0)>0}\left(\lambda_t^k(f)-\lambda_t^k(f_0)\right)dt\right] \\
&\leqslant \sum_k TL\left(|\nu_k^0-\nu_k|+\sum_l \left\|h_{lk}-h_{lk}^0\right\|_\infty \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\sup_{t\in[0,T]} N^l[t-A,t)\right]\right) \\
&\leqslant TL\sum_k\left(|\nu_k^0-\nu_k|+\sum_l \left\|h_{lk}-h_{lk}^0\right\|_\infty\right)\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\max_l \sup_{t\in[0,T]} N^l[t-A,t)\right]\leqslant LT^{1-\beta}\epsilon_T \qquad \text{(S7.34)}
\end{aligned}
$$

for T large enough, using Lemma A.1 for $\beta>0$. If the model verifies Assumption 3.1(i), and $f\in B_2(\epsilon_T,B)$, we have

$$
\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\vee\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\leqslant 2\frac{2\theta_k^0+2L\nu_k^0+L(B+\max_l\left\|h_{lk}^0\right\|_\infty)\sup_t N[t-A,t)}{\inf_x\phi_k(x)},
$$

therefore

$$
\begin{aligned}
\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_0^T\left|\log\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\right|\lambda_t^k(f_0)dt\right] &\lesssim \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\max_l\sup_{t\in[0,T]}N^l[t-A,t)\int_0^T\lambda_t^k(f_0)dt\right] \\
&\lesssim T\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\left(\sup_{t\in[0,T]}N[t-A,t)\right)\left(\nu_k^0+\max_l\left\|h_{lk}^0\right\|_\infty\sup_{t\in[0,T]}N^l[t-A,t)\right)\right] \\
&\lesssim T\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\max_l\left(\sup_{t\in[0,T]}N^l[t-A,t)\right)^2\right]\lesssim T^{1-\beta}.
\end{aligned}
$$

If instead the model verifies Assumption 3.1(ii), using that $\log\phi_k$ is L_1-Lipschitz for any k, we can alternatively use that

$$
\begin{aligned}
\sum_k\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_0^T\left|\log\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\right|\lambda_t^k(f_0)dt\right] &\leqslant L_1\sum_k\mathbb{E}_0\left[\int_0^T\mathbb{1}_{\tilde{\Omega}_T^c}\lambda_t^k(f_0)|\tilde{\lambda}_t^k(f)-\tilde{\lambda}_t^k(f_0)|dt\right] \\
&\lesssim\sum_k T\left(|\nu_k^0-\nu_k|+\sum_l\left\|h_{lk}-h_{lk}^0\right\|_\infty\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\max_l\left(\sup_{t\in[0,T]}N^l[t-A,t)\right)^2\right]\right)\lesssim T^{1-\beta}.
\end{aligned}
$$

We can additionally bound the second term of (S7.32) in a similar fashion and conclude that, in all cases, $R_T=O(T^{1-\beta})=o(T\epsilon_T^2)$ for β large enough.

To bound the first term of the RHS of (S7.33), we consider separately the models satisfying Assumption 3.1(i) and (ii) and Case 1 and Case 2 of Proposition 3.5.

Scenario 1: under Assumption 3.1(i) or Case 2 of Proposition 3.5

Under Assumption 3.1(i), for any $f\in B_\infty(\epsilon_T)$ or $f\in B_2(\epsilon_T,B)$ and $t\geqslant 0$, $\lambda_t^k(f)\geqslant\inf_x\phi_k(x)\geqslant\min_k\inf_x\phi_k(x)$ and $\lambda_t^k(f_0)\leqslant L\nu_k^0+L\sup_{t\in[0,T]}N[t-A,t)\sum_l\|h_{lk}^0\|_\infty$. In Case 2 of Proposition 3.2, for T large enough, $t\in[0,T]$ and $\theta\in B_\infty^\Theta(\epsilon_T)$, $\lambda_t^k(f,\theta)\geqslant\theta_k\geqslant\theta_k^0/2$ and $\lambda_t^k(f_0,\theta_0)\leqslant\theta_k^0+L\nu_k^0+L\sup_{t\in[0,T]}N[t-A,t)\sum_l\|h_{lk}^0\|_\infty$. Therefore, in this scenario, on $\tilde{\Omega}_T$, $\lambda_t^k(f_0)/\lambda_t^k(f)\leqslant\ell_0\log T$ for some $\ell_0>0$. Thus, with $\chi(x)=-\log x+x-1$, we have

$$
KL(f_0,f)-R_T=\sum_k\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\left(\int_0^T\lambda_t^k(f_0)\left(\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)+\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}-1\right)dt\right)\right]
$$

$$= \sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \left(\int_0^T \lambda_t^k(f_0) \chi \left(\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)} \right) dt \right) \right]$$
$$\leqslant \frac{4 \log(\ell_0 \log T)}{\min_k \inf_x \phi_k(x)} \sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \int_0^T (\lambda_t^k(f_0) - \lambda_t^k(f))^2 dt \right],$$

since for any $r_T \in (0, 1/2]$ and $x \geqslant r_T$, we have $\chi(x) \leqslant 4 \log r_T^{-1} (x-1)^2$ (see the proof of Lemma 2 of Donnet, Rivoirard and Rousseau (2020)). Note that if $f \in B_\infty(\epsilon_T)$, $\forall t \in [0, T]$, $\lambda_t^k(f) \geqslant \lambda_t^k(f_0)$ and we obtain instead

$$KL(f_0, f) - R_T \leqslant \frac{1}{\min_k \inf_x \phi_k(x)} \sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \int_0^T (\lambda_t^k(f_0) - \lambda_t^k(f))^2 dt \right].$$

Moreover, since ϕ_k is L-Lipschitz, under Assumption 3.1,

$$|\lambda_t^k(f_0) - \lambda_t^k(f)| = |\phi_k(\tilde{\lambda}_t^k(\nu_0, h_0)) - \phi_k(\tilde{\lambda}_t^k(\nu, h))| \leqslant L |\tilde{\lambda}_t^k(\nu_0, h_0) - \tilde{\lambda}_t^k(\nu, h)|$$
$$\leqslant L |\nu_k - \nu_k^0| + L \sum_l \int_{t-A}^{t^-} |h_{lk} - h_{lk}^0|(t-s) dN_s^l,$$

and in Case 2 of Proposition 3.5, we have

$$|\lambda_t^k(f_0, \theta_0) - \lambda_t^k(f, \theta)| = |\theta_k^0 + \phi_k(\tilde{\lambda}_t^k(\nu_0, h_0)) - \theta_k - \phi_k \tilde{\lambda}_t^k(\nu, h))| \leqslant |\theta_k^0 - \theta_k| + L |\tilde{\lambda}_t^k(\nu_0, h_0) - \tilde{\lambda}_t^k(\nu, h)|$$
$$\leqslant |\theta_k - \theta_k^0| + L |\nu_k - \nu_k^0| + L \sum_l \int_{t-A}^{t^-} |h_{lk} - h_{lk}^0|(t-s) dN_s^l.$$

Using the same computations as in the proof of Lemma 2 of Donnet, Rivoirard and Rousseau (2020), we obtain

$$\sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \left(\int_0^T (\lambda_t^k(f_0) - \lambda_t^k(f))^2 \right) dt \right] \leqslant \gamma_0 T \left(|\nu_k - \nu_k^0|^2 + \sum_l \|h_{lk} - h_{lk}^0\|_2^2 \right) \leqslant \gamma_0 T \epsilon_T^2,$$

or

$$\sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \left(\int_0^T (\lambda_t^k(f_0, \theta_0) - \lambda_t^k(f, \theta))^2 \right) dt \right] \leqslant \gamma_0 T \left(\sum_k |\theta_k - \theta_k^0|^2 + |\nu_k - \nu_k^0|^2 + \sum_l \|h_{lk} - h_{lk}^0\|_2^2 \right) \leqslant \gamma_0 T \epsilon_T^2,$$

with $\gamma_0 := \max(1, L) \left[3 + 6K \sum_k \left(A \mathbb{E}_0 \left[\lambda_0^k(f_0)^2 \right] + \mathbb{E}_0 \left[\lambda_0^k(f_0) \right] \right) \right]$. Consequently,

$$KL(f_0, f) - R_T \leqslant \begin{cases} \frac{4 \log(\ell_0 \log T)}{\min_k \inf_x \phi_k(x)} \gamma_0 T \epsilon_T^2 & \text{if} \quad f \in B_2(\epsilon_T, B) \\ \frac{\gamma_0}{\min_k \inf_x \phi_k(x)} T \epsilon_T^2 & \text{if} \quad f \in B_\infty(\epsilon_T). \end{cases} \tag{S7.35}$$

Therefore, $KL(f_0, f) \leqslant \kappa_1' (\log \log T) T \epsilon_T^2$, with $\kappa_1' = \frac{8\gamma_0}{\min_k \inf_x \phi_k(x)}$ if $f \in B_2(\epsilon_T, B)$ - or $KL(f_0, f) \leqslant \kappa_1 T \epsilon_T^2$ with $\kappa_1 = \frac{2}{\min_k \inf_x \phi_k(x)}$ if $f \in B_\infty(\epsilon_T)$.

Scenario 2: Under Assumption 3.1(ii), i.e., $\phi_k > 0$, and $\log \phi_k$ and $\sqrt{\phi_k}$ are L_1-Lipschitz, $L_1 > 0$.

For $k \in [K]$, let $\Lambda^k(f) := \int_0^T \lambda_t^k(f)dt$. Then for $t \in [0,T]$, we define

$$\alpha_t^k(f) = \frac{\lambda_t^k(f)}{\Lambda^k(f)}.$$

From (S7.33), we have

$$\begin{aligned} KL(f_0, f) - R_T &= \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\left(\int_0^T \lambda_t^k(f_0)\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)dt + \int_0^T (\lambda_t^k(f) - \lambda_t^k(f_0))dt\right)\right] \\ &= \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\left(\Lambda_A^k(f_0)\int_{A^k(T)} \alpha_t^k(f_0)\log\left(\frac{\alpha_t^k(f_0)}{\alpha_t^k(f)}\right)dt + \Lambda^k(f_0)\log\left(\frac{\Lambda^k(f_0)}{\Lambda^k(f)}\right) + (\Lambda^k(f) - \Lambda^k(f_0))\right)\right] \\ &\leqslant \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\left(\Lambda^k(f_0)\int_0^T \alpha_t^k(f_0)\log\left(\frac{\alpha_t^k(f_0)}{\alpha_t^k(f)}\right)dt + \frac{(\Lambda^k(f_0) - \Lambda^k(f))^2}{\Lambda^k(f_0)}\right)\right], \end{aligned}$$

where in the last inequality we have used that $\chi(x) \leqslant (x-1)^2$ for $x \geqslant 1/2$, with $x = \frac{\Lambda^k(f)}{\Lambda^k(f_0)}$. In fact, we have

$$|\Lambda^k(f) - \Lambda^k(f_0)| \leqslant TL|\nu_k - \nu_k^0| + L\sum_l \left\|h_{lk} - h_{lk}^0\right\|_1 N^l[-A,T] \leqslant TL\epsilon_T(1 + 2\max_l \mu_l^0),$$

using that on $\tilde{\Omega}_T$, $N^l[-A,T] \leqslant T\mu_l^0 + T\delta_T \leqslant 2T\mu_l^0$. Moreover, on $\tilde{\Omega}_T$, using the notations of Section 5.2, we have

$$\Lambda^k(f_0) \geqslant \phi_k(\nu_k^0)\sum_{j=1}^{J_T-1}(U_j^{(1)} - \tau_j) \geqslant \phi_k(\nu_k^0)\frac{T}{2\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1} =: y_0 T,$$

for some $y_0 > 0$. Similarly, for $f \in B_2(\epsilon_T, B)$ or $f \in B_\infty(\epsilon_T)$, we have

$$\Lambda^k(f) \geqslant \phi_k(\nu_k)\sum_{j=1}^{J_T-1}(U_j^{(1)} - \tau_j) \geqslant \phi_k(\nu_k^0/2)\frac{T}{2\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1}.$$

Consequently,

$$\frac{1}{2} \leqslant 1 - \frac{|\Lambda^k(f) - \Lambda^k(f_0)|}{\Lambda_A(f_0)} \leqslant \frac{\Lambda^k(f)}{\Lambda^k(f_0)} \leqslant 1 + \frac{|\Lambda^k(f) - \Lambda^k(f_0)|}{\Lambda_A(f_0)} \leqslant 1 + \frac{1 + 2A\max_l \mu_l^0}{y_0}\epsilon_T = 1 + O(\epsilon_T),$$

for T large enough, and

$$\frac{(\Lambda^k(f) - \Lambda^k(f_0))^2}{\Lambda^k(f_0)} \leqslant \frac{L^2T^2\epsilon_T^2(1 + 2\max_l \mu_l^0)^2}{\Lambda^k(f_0)} \leqslant \frac{L^2T\epsilon_T^2(1 + 2A\max_l \mu_l^0)^2}{y_0}.$$

Additionally, on $\tilde{\Omega}_T$, on the one hand, for $f \in B_2(\epsilon_T, B)$, we also have that for any $t \in [0,T]$, since $\lambda_t^k(f_0) \leqslant \lambda_t^k(f) + \epsilon_T + BC_\beta \log T \implies \frac{\lambda_t^k(f_0)}{\lambda_t^k(f)} \leqslant M_0 \log T$ for some $M_0 > 0$, then

$$\frac{\alpha_t^k(f_0)}{\alpha_t^k(f)} = \frac{\lambda_t^k(f_0)\Lambda^k(f)}{\lambda_t^k(f)\Lambda^k(f_0)} \leqslant M_0 \log T\frac{\Lambda^k(f)}{\Lambda^k(f_0)} \leqslant M\log T + O(M_0 \log T\epsilon_T).$$

Applying Lemma 8.7 from Ghosal, Ghosh and van der Vaart (2000), we have, for any $M \geqslant M_0$,

$$\int_0^T \alpha_t^k(f_0) \log\left(\frac{\alpha_t^k(f_0)}{\alpha_t^k(f)}\right) dt \leqslant \log(M \log T) \int_0^T \left(\sqrt{\alpha_t^k(f_0)} - \sqrt{\alpha_t^k(f)}\right)^2 dt.$$

Moreover,

$$\begin{aligned}\int_0^T \left(\sqrt{\alpha_t^k(f_0)} - \sqrt{\alpha_t^k(f)}\right)^2 dt &\leqslant \int_0^T \frac{1}{\Lambda^k(f_0)}\left(\sqrt{\lambda_t^k(f_0)} - \sqrt{\frac{\Lambda^k(f_0)}{\Lambda^k(f)}\lambda_t^k(f)}\right)^2 dt \\ &\leqslant \frac{2}{\Lambda^k(f_0)} \int_0^T \left(\sqrt{\lambda_t^k(f_0)} - \sqrt{\lambda_t^k(f)}\right)^2 dt + \frac{1}{\Lambda^k(f_0)} \int_0^T \lambda_t^k(f)\left(1 - \sqrt{\frac{\Lambda^k(f_0)}{\Lambda^k(f)}}\right)^2 dt \\ &\lesssim \frac{1}{\Lambda^k(f_0)} \int_0^T \left(\sqrt{\lambda_t^k(f_0)} - \sqrt{\lambda_t^k(f)}\right)^2 dt + \frac{(\Lambda^k(f) - \Lambda^k(f_0))^2}{\Lambda^k(f_0)^2}.\end{aligned}$$

On the other hand, if $f \in B_\infty(\epsilon_T)$, then $\lambda_t^k(f_0) \leqslant \lambda_t^k(f)$ and we have

$$\int_0^T \alpha_t^k(f_0) \log\left(\frac{\alpha_t^k(f_0)}{\alpha_t^k(f)}\right) dt \leqslant \frac{2}{\Lambda^k(f_0)} \int_0^T \left(\sqrt{\lambda_t^k(f_0)} - \sqrt{\lambda_t^k(f)}\right)^2 dt + \frac{4(\Lambda^k(f) - \Lambda^k(f_0))^2}{\Lambda^k(f_0)^2}.$$

Moreover, in this case,

$$\begin{aligned}\int_0^T \left(\sqrt{\lambda_t^k(f_0)} - \sqrt{\lambda_t^k(f)}\right)^2 dt &= \int_0^T \left(\sqrt{\phi_k(\tilde\lambda_t^k(\nu_0, h_0))} - \sqrt{\phi_k(\tilde\lambda_t^k(\nu, h))}\right)^2 dt \\ &\leqslant L_1^2 \int_{A^k(T)} \left(\tilde\lambda_t^k(\nu_0, h_0) - \tilde\lambda_t^k(\nu, h)\right)^2 dt \lesssim T\epsilon_T^2.\end{aligned}$$

Finally, we obtain that

$$KL(f_0, f) \lesssim \begin{cases}(\log\log T) T\epsilon_T^2 & \text{if} \quad f \in B_2(\epsilon_T, B) \\ T\epsilon_T^2 & \text{if} \quad f \in B_\infty(\epsilon_T)\end{cases}.$$

Scenario 3: Case 1 of Proposition 3.5, i.e., $\phi_k(x) = (x)_+, \forall k \in [K]$.

In a Hawkes model with the standard ReLU link function, we can obtain two types of rates, under and without condition (8). We consider $f \in B_\infty(\epsilon_T)$ so that $\forall t \in [0,T], \tilde\lambda_t^k(\nu, h) \geqslant \tilde\lambda_t^k(\nu_0, h_0)$. Since for any $t \in [0,T]$, $\log(\lambda_t^k(f_0)/\lambda_t^k(f)) \leqslant 0$, we can use that

$$KL(f_0, f) \leqslant \sum_k \mathbb{E}_0\left[\int_0^T (\lambda_t^k(f) - \lambda_t^k(f_0)) dt\right] = \sum_k \mathbb{E}_0\left[\Lambda^k(f) - \Lambda^k(f_0)\right],$$

with for any $1 \leqslant k \leqslant K$, $\Lambda^k(f) := \int_0^T \lambda_t^k(f) dt$, and $\Lambda^k(f_0) := \int_0^T \lambda_t^k(f_0) dt$. Since for any t, $\tilde\lambda_t^k(\nu, h) \geqslant \tilde\lambda_t^k(\nu_0, h_0)$, we have

$$0 \leqslant \Lambda^k(f) - \Lambda^k(f_0) = \int_0^T ((\tilde\lambda_t^k(\nu, h))_+ - (\tilde\lambda_t^k(\nu_0, h_0))_+ dt \leqslant \int_0^T |\tilde\lambda_t^k(\nu, h) - \tilde\lambda_t^k(\nu_0, h_0)| dt$$

$$\leqslant T|\nu_k - \nu_k^0| + \sum_l \int_0^T \int_{t-A}^{t^-} |h_{lk} - h_{lk}^0|(t-s) dN_s^l dt \leqslant T(\nu_k - \nu_k^0) + \sum_l \left\| h_{lk} - h_{lk}^0 \right\|_1 N^l[-A, T]. \tag{S7.36}$$

Consequently, we arrive at

$$\begin{aligned} KL(f_0, f) &\leqslant KT\epsilon_T (1 + \max_l \mathbb{E}_0 \left[N^l[-A, T] \right]) + R_T \\ &\leqslant T\epsilon_T K(1 + 2 \max_l \mu_l^0) + o(T\epsilon_T^2) \lesssim T\epsilon_T. \end{aligned}$$

To refine this bound, we will assume that (8) holds. For $k \in [K]$ and $t \in [0, T]$, we define $p_t^k(f) = \lambda_t^k(f)/\Lambda^k(f)$ and similarly for $p_t^k(f_0)$. Using (S7.33), we then have

$$\begin{aligned} KL(f_0, f) - R_T &= \sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \left(\Lambda^k(f_0) \int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0} p_t^k(f_0) \log \left(\frac{p_t^k(f_0)}{p_t^k(f)} \right) dt + \Lambda^k(f_0) \log \left(\frac{\Lambda^k(f_0)}{\Lambda^k(f)} \right) + (\Lambda^k(f) - \Lambda^k(f_0)) \right) \right] \\ &\leqslant \sum_k \mathbb{E}_0 \left[\mathbb{1}_{\tilde{\Omega}_T} \left(\Lambda^k(f_0) \int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0} p_t^k(f_0) \log \left(\frac{p_t^k(f_0)}{p_t^k(f)} \right) dt + \frac{(\Lambda^k(f_0) - \Lambda^k(f))^2}{\Lambda^k(f_0)} \right) \right], \end{aligned} \tag{S7.37}$$

where in the last inequality, we have used the fact that $-\log x + x - 1 \leqslant (x-1)^2$ for $x \geqslant 1/2$, with $x = \frac{\Lambda^k(f)}{\Lambda^k(f_0)} \geqslant 1$. Moreover, from (S7.36), we have on $\tilde{\Omega}_T$,

$$\Lambda^k(f) - \Lambda^k(f_0) \leqslant T\epsilon_T(1 + 2 \max_l \mu_l^0).$$

Besides, on $\tilde{\Omega}_T$, using $A_2(T)$ defined in (22) and noting that in this case, $r_k^0 = \nu_k^0, \forall k$,

$$\begin{aligned} \Lambda^k(f_0) &\geqslant \int_{A_2(T)} \lambda_t^k(f_0) dt \geqslant \sum_{j=1}^{J_T - 1} \int_{\tau_j}^{U_j^{(1)}} \lambda_t^k(f_0) dt = \nu_k^0 \sum_{j=1}^{J_T-1} (U_j^{(1)} - \tau_j) \\ &\geqslant \frac{\nu_k^0 T}{\mathbb{E}_0(\Delta\tau_1) \|\nu_0\|_1} \left(1 - 2c_\beta \sqrt{\frac{\log T}{T}} \right) \geqslant \frac{\nu_k^0 T}{2\mathbb{E}_0(\Delta\tau_1)\|\nu_0\|_1}. \end{aligned}$$

Therefore,

$$\begin{aligned} \Lambda^k(f_0) \leqslant \Lambda^k(f) &\leqslant \Lambda^k(f_0) + T\epsilon_T(1 + 2 \max_l \mu_l^0) \\ &\leqslant \Lambda^k(f_0) + \frac{2\Lambda^k(f_0)(1 + 2A \max_l \mu_l^0)\mathbb{E}_0(\Delta\tau_1)\|\nu_0\|_1}{\nu_k^0} \epsilon_T \\ &\leqslant \Lambda^k(f_0) \left(1 + \frac{2(1 + 2A \max_l \mu_l^0)\mathbb{E}_0(\Delta\tau_1)\|\nu_0\|_1}{\nu_k^0} \epsilon_T \right) \leqslant 2\Lambda^k(f_0), \end{aligned} \tag{S7.38}$$

for T large enough. Besides, this implies that $p_t^k(f) = \frac{\lambda_t^k(f)}{\Lambda^k(f)} \geqslant \frac{\lambda_t^k(f_0)}{2\Lambda^k(f_0)} \geqslant p_t^k(f_0)/2$. Using again the inequality $-\log x + x - 1 \leqslant (x-1)^2$ with $x = \frac{p_t^k(f)}{p_t^k(f_0)} \geqslant \frac{1}{2}$ and the fact that $\int_0^T p_t^k(f)dt = \int_0^T p_t^k(f_0)dt = 1$, we have

$$\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0} p_t^k(f_0)\log\left(\frac{p_t^k(f_0)}{p_t^k(f)}\right)dt = \int_0^T p_t^k(f_0)\log\left(\frac{p_t^k(f_0)}{p_t^k(f)}\right)dt + \int_0^T (p_t^k(f) - p_t^k(f_0))dt$$

$$= \int_0^T p_t^k(f_0)\left(\log\left(\frac{p_t^k(f_0)}{p_t^k(f)}\right) + \frac{p_t^k(f)}{p_t^k(f_0)} - 1\right)dt \leqslant \int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{(p_t^k(f_0) - p_t^k(f))^2}{p_t^k(f_0)}dt$$

$$\leqslant \frac{1}{\Lambda^k(f_0)}\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{2\left(\lambda_t^k(f_0) - \lambda_t^k(f)\right)^2 + 2\lambda_t^k(f)^2\left(1 - \frac{\Lambda^k(f_0)}{\Lambda^k(f)}\right)^2}{\lambda_t^k(f_0)}dt$$

$$\leqslant \frac{2}{\Lambda^k(f_0)}\left[\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{3\left(\lambda_t^k(f_0) - \lambda_t^k(f)\right)^2}{\lambda_t^k(f_0)} + 2\Lambda^k(f_0)\times\frac{(\Lambda^k(f) - \Lambda^k(f_0))^2}{\Lambda^k(f)^2}\right]$$

$$\leqslant \frac{6}{\Lambda^k(f_0)}\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{2\left(\lambda_t^k(f_0) - \lambda_t^k(f)\right)^2}{\lambda_t^k(f_0)}dt + 4\frac{(\Lambda^k(f) - \Lambda^k(f_0))^2}{\Lambda^k(f_0)^2}.$$

In the previous inequalities, we have used $\Lambda^k(f_0) \leqslant \Lambda^k(f)$, and for T large enough, we have the following intermediate result:

$$KL(f_0, f) - R_T \leqslant \sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\left(6\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{(\lambda_t^k(f_0) - \lambda_t^k(f))^2}{\lambda_t^k(f_0)}dt + 4\frac{(\Lambda^k(f_0) - \Lambda^k(f))^2}{\Lambda^k(f_0)}\right)\right]. \quad \text{(S7.39)}$$

Moreover, on $\tilde{\Omega}_T$, using (S7.38)

$$\Lambda^k(f_0) = \int_0^T \left(\nu_k^0 + \sum_l \int_{t-A}^{t^-} h_{lk}^0(t-s)dN_s^l\right)_+ dt \leqslant T\nu_k^0 + \sum_l \|h_{lk}^{0+}\|_1 N^l[-A, T)$$

$$\leqslant T\nu_k^0 + \frac{3}{2}T\sum_l \|h_{lk}^{0+}\|_1(\mu_l^0 + \delta_T) \leqslant 2T\left(\nu_k^0 + \sum_l \|h_{lk}^{0+}\|_1\mu_l^0\right),$$

for T large enough, since $\delta_T = \delta_0\sqrt{\frac{\log T}{T}}$. Thus,

$$\frac{(\Lambda^k(f_0) - \Lambda^k(f))^2}{\Lambda^k(f_0)} \leqslant \Lambda^k(f_0)\left(\frac{2(1 + 2A\max_l \mu_l^0)\mathbb{E}_0(\Delta\tau_1)\|\nu_0\|_1}{\nu_k^0}\right)^2 \epsilon_T^2 \leqslant= c_2^0 T\epsilon_T^2,$$

with

$$c_2^0 = 8\left(\nu_k^0 + \sum_l \|h_{lk}^{0+}\|_1\mu_l^0\right)\left(\frac{(1 + 2A\max_l \mu_l^0)\mathbb{E}_0(\Delta\tau_1)\|\nu_0\|_1}{\nu_k^0}\right)^2.$$

Therefore, reporting into (S7.39) we have

$$KL(f_0, f) - R_T \leqslant 6\sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{(\lambda_t^k(f_0) - \lambda_t^k(f))^2}{\lambda_t^k(f_0)}dt\right] + 4Kc_2^0 T\epsilon_T^2.$$

We now bound the first term on the RHS of the previous equation.

$$\sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{(\lambda_t^k(f_0)-\lambda_t^k(f))^2}{\lambda_t^k(f_0)}dt\right] \leqslant \sum_k \mathbb{E}_0\left[\mathbb{1}_{\Omega_{\tilde{\Omega}_T}}\sup_{t\in[0,T]}\mathbb{1}_{\lambda_t^k(f_0)>0}(\lambda_t^k(f)-\lambda_t^k(f_0))^2\int_0^T\frac{\mathbb{1}_{\lambda_t^k(f_0)>0}}{\lambda_t^k(f_0)}dt\right].$$

Moreover, for any $k\in[K]$ and $t\in[0,T]$, we have on $B_\infty(\epsilon_T)$

$$\begin{aligned}\mathbb{1}_{\tilde{\Omega}_T}\mathbb{1}_{\lambda_t^k(f_0)>0}(\lambda_t^k(f)-\lambda_t^k(f_0))^2 dt &\leqslant 2(\nu_k-\nu_k^0)^2+2K\max_l\|h_{lk}-h_{lk}^0\|_\infty^2\sup_{t\in[0,T]}N^l[t-A,t)^2\\ &\leqslant 2\epsilon_T^2+2KC_\beta^2\log^2 T\epsilon_T^2\leqslant 4KC_\beta^2\log^2 T\epsilon_T^2.\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_k \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\int_0^T \mathbb{1}_{\lambda_t^k(f_0)>0}\frac{(\lambda_t^k(f_0)-\lambda_t^k(f))^2}{\lambda_t^k(f_0)}dt\right] &\leqslant 4C_\beta^2K(\log T)^2T\epsilon_T^2\sum_k\mathbb{E}_0\left[\frac{1}{T}\int_0^T\frac{\mathbb{1}_{\lambda_t^k(f_0)>0}}{\lambda_t^k(f_0)}dt\right]\\ &= 4C_\beta^2c_1^0K(\log T)^2T\epsilon_T^2,\end{aligned}$$

using (8), with

$$c_1^0 := \lim\sup_{T\to\infty}\mathbb{E}_0\left[\frac{1}{T}\int_0^T\frac{\mathbb{1}_{\lambda_t^k(f_0)>0}}{\lambda_t^k(f_0)}dt\right]<+\infty.$$

Consequently, reporting into (S7.39), we finally obtain

$$\begin{aligned}KL(f_0,f)&\leqslant 4C_\beta^2c_1^0KL(\log T)^2T\epsilon_T^2+4Kc_2^0T\epsilon_T^2+o(T\epsilon_T^2)\\ &\leqslant 8KC_\beta^2c_1^0(\log T)^2T\epsilon_T^2=\kappa_2(\log T)^2T\epsilon_T^2,\end{aligned}$$

with $\kappa_2 := 8KC_\beta^2c_1^0$, which terminates the proof of this lemma.

□

S7.2. Deviations on the log likelihood ratio: Lemma S7.3

The next lemma is a control under $\mathbb{P}_0$ over the centered sum of i.i.d. variables that are used to decompose the log-likelihood ratio in Lemma A.2.

Lemma S7.3. *Under the assumptions of Lemma S7.1, for $f\in B_\infty(\epsilon_T)$ and $j\geqslant 1$, let*

$$T_j := \sum_k\int_{\tau_j}^{\tau_{j+1}}\log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)dN_t^k-\int_{\tau_j}^{\tau_{j+1}}(\lambda_t^k(f_0)-\lambda_t^k(f))dt. \tag{S7.40}$$

Then it holds that $\mathbb{E}_0\left[T_j^2\right]\lesssim z_T/T$, with

$$z_T=\begin{cases}T\epsilon_T^2 & \text{(under Assumption 3.1(i))}\\ (\log T)T\epsilon_T^2 & \text{(under Assumption 3.1(ii))}\\ (\log T)^2T\epsilon_T^2 & \text{(ReLU link)}\end{cases}$$

Moreover, if $\log^3 T = O(z_T)$,

$$\mathbb{P}_0\left[\sum_{j=0}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right] \geqslant z_T\right] = o(1).$$

***Remark* S7.4.** Under Assumption 3.5, for $f \in B_2(\epsilon_T, B)$, we also obtain similar results with $z_T = (\log\log T)^2 T\epsilon_T^2$.

Proof. Firstly, using the fact that τ_1, τ_2 are stopping times, we have

$$\begin{aligned}
\mathbb{E}_0\left[T_1^2\right] &= \mathbb{E}_0\left[\left(\sum_k \int_{\tau_1}^{\tau_2} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right) dN_t^k - \int_{\tau_1}^{\tau_2} (\lambda_t^k(f_0) - \lambda_t^k(f)) dt\right)^2\right] \\
&\lesssim \sum_k \mathbb{E}_0\left[\left(\int_{\tau_1}^{\tau_2} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt + \int_{\tau_1}^{\tau_2} \log\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)(dN_t^k - \lambda_t^k(f_0)dt) - \int_{\tau_1}^{\tau_2}(\lambda_t^k(f_0) - \lambda_t^k(f))dt\right)^2\right] \\
&\lesssim \mathbb{E}_0\left[\Delta\tau_1 \int_{\tau_1}^{\tau_2} \chi\left(\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\right)^2 \lambda_t^k(f_0)^2 dt\right] + \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right],
\end{aligned} \tag{S7.41}$$

with $\chi(x) = -\log x + x - 1$. For any $x > 0$, we have $\chi^2(x) \leqslant 2\log^2 x + 2(x-1)^2$. Now, if $f \in B_\infty(\epsilon_T)$, using that $\log^2 x \leqslant (x-1)^2$ for $x = \lambda_t^k(f)/\lambda_t^k(f_0) \geqslant 1$, we have $\chi\left(\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\right)^2 \lambda_t^k(f_0)^2 \lesssim (\lambda_t^k(f_0) - \lambda_t^k(f))^2$ and $\log^2\left(\frac{\lambda_t^k(f)}{\lambda_t^k(f_0)}\right)\lambda_t^k(f_0) \lesssim \frac{(\lambda_t^k(f_0)-\lambda_t^k(f))^2}{\lambda_t^k(f_0)}$. Therefore, (S7.41) becomes

$$\begin{aligned}
\mathbb{E}_0\left[T_1^2\right] \lesssim \mathbb{E}_0\left[\Delta\tau_1 \int_{\tau_1}^{\tau_2} (\lambda_t^k(f_0) - \lambda_t^k(f))^2 dt\right] + \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c} \int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right] \\
+ \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T} \int_{\tau_1}^{\tau_2} \mathbb{1}_{\lambda_t^k(f_0)>0} \frac{(\lambda_t^k(f_0) - \lambda_t^k(f))^2}{\lambda_t^k(f_0)} dt\right].
\end{aligned} \tag{S7.42}$$

With the ReLU link function, we can easily bound the third term on the RHS of (S7.42) using (8):

$$\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T} \int_{\tau_1}^{\tau_2} \mathbb{1}_{\lambda_t^k(f_0)>0} \frac{(\lambda_t^k(f_0) - \lambda_t^k(f))^2}{\lambda_t^k(f_0)} dt\right] \lesssim \log^2 T\epsilon_T^2 \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} \frac{\mathbb{1}_{\lambda_t^k(f_0)>0}}{\lambda_t^k(f_0)} dt\right] \lesssim \log^2 T\epsilon_T^2.$$

For the second term on the RHS of (S7.42), using that $\log^2(\lambda_t^k(f))\lambda_t^k(f) \lesssim (\sup_t N[t-A,t))^3$ and similarly for $\lambda_t^k(f_0)$, we have

$$\begin{aligned}
\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c} \int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right] &\lesssim \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c} \int_{\tau_1}^{\tau_2} \log^2(\lambda_t^k(f_0))\lambda_t^k(f_0)dt\right] + \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c} \int_{\tau_1}^{\tau_2} \log^2(\lambda_t^k(f))\lambda_t^k(f)dt\right] \\
&\lesssim \sqrt{\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}(\sup_t N[t-A,t))^6\right]}\sqrt{\mathbb{E}_0\left[\Delta\tau_1^2\right]} \lesssim T^{-\beta/2} = o(\epsilon_T^2),
\end{aligned}$$

using Lemma A.1. For the first term on the RHS of (S7.42), we have

$$\mathbb{E}_0\left[\Delta\tau_1 \int_{\tau_1}^{\tau_2} (\lambda_t^k(f_0) - \lambda_t^k(f))^2 dt\right] \lesssim \mathbb{E}_0\left[\Delta\tau_1 \int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(f_0) - \tilde{\lambda}_t^k(f))^2 dt\right]$$

$$\leqslant \mathbb{E}_0\left[\Delta\tau_1 \int_{\tau_1}^{\tau_2} (2|\nu_k - \nu_k^0|^2 + 2K\sum_{l=1}^K \left(\int_{t-A}^t (h_{lk} - h_{lk}^0)(t-s)dN_s^l\right)^2 dt\right]$$

$$\leqslant 2|\nu_k - \nu_k^0|^2 \mathbb{E}_0\left[\Delta\tau_1^2\right] + 2K\sum_{l=1}^K \mathbb{E}_0\left[\Delta\tau_1 \int_{\tau_1}^{\tau_2} N^l(t-A,t)\int_{t-A}^t (h_{lk} - h_{lk}^0)^2(t-s)dN_s^l dt\right]$$

$$= 2|\nu_k - \nu_k^0|^2 \mathbb{E}_0\left[\Delta\tau_1^2\right] + 2K\sum_{l=1}^K \left\|h_{lk} - h_{lk}^0\right\|_2^2 \mathbb{E}_0\left[\Delta\tau_1 N^l[\tau_1,\tau_2)^2\right]$$

$$\leqslant 2|\nu_k - \nu_k^0|^2 \mathbb{E}_0\left[\Delta\tau_1\right] + 2K\sum_{l=1}^K \left\|h_{lk} - h_{lk}^0\right\|_2^2 \sqrt{\mathbb{E}_0\left[N^l[\tau_1,\tau_2)^4\right]}\sqrt{\mathbb{E}_0\left[\Delta\tau_1^2\right]} \lesssim \epsilon_T^2.$$

Thus, reporting into (S7.42), we can conclude that if (8) holds, $\mathbb{E}_0\left[T_1^2\right] \lesssim \log^2 T\epsilon_T^2$.

Under Assumption 3.1(i), if $f \in B_\infty(\epsilon_T)$, we can use the same computations. If $f \in B_2(\epsilon_T, B)$, for the first term on the RHS of (S7.42) and for the second term, we use instead that $\log^2 x \leqslant 4\log^2(r_T^{-1})(x-1)^2$ for $x \geqslant r_T$ with $x = \frac{\lambda_t^k(f_0)}{\lambda_t^k(f)} \gtrsim r_T := (\log T)^{-1}$ and we obtain,

$$\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right] \lesssim (\log\log T)^2 \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} (\lambda_t^k(f_0) - \lambda_t^k(f))^2 dt\right]$$
$$\lesssim (\log\log T)^2 \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(\nu_0,h_0) - \tilde{\lambda}_t^k(\nu,h))^2 dt\right] \lesssim (\log\log T)^2\epsilon_T^2,$$

or, in the shifted ReLU model with unknown link (Case 2 of Proposition 3.5),

$$\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T}\int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0,\theta_0)}{\lambda_t^k(f,\theta)}\right)\lambda_t^k(f_0,\theta_0)dt\right]$$
$$\lesssim (\log\log T)^2\left[\mathbb{E}_0\left[\Delta\tau_1\right](\theta_k - \theta_k^0)^2 + \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(\nu_0,h_0) - \tilde{\lambda}_t^k(\nu,h))^2 dt\right]\right] \lesssim (\log\log T)^2\epsilon_T^2,$$

using similar computations to the control of the first term of (S7.42). The remaining term, i.e.,

$$\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right],$$

is bounded as the second term of (S7.42).

Finally, under Assumption 3.1(ii), using the fact that $\log\phi_k$ L_1-Lipschitz for any k, we have

$$\mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} \log^2\left(\frac{\lambda_t^k(f_0)}{\lambda_t^k(f)}\right)\lambda_t^k(f_0)dt\right] \lesssim \mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(\nu_0,h_0) - \tilde{\lambda}_t^k(\nu,h))^2\lambda_t^k(f_0)dt\right]$$
$$\lesssim \log T\,\mathbb{E}_0\left[\int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(\nu_0,h_0) - \tilde{\lambda}_t^k(\nu,h))^2 dt\right] + \mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c}\int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(f_0) - \tilde{\lambda}_t^k(f))^2\lambda_t^k(f_0)dt\right]$$
$$\lesssim (\log T)\epsilon_T^2,$$

and the first term of (S7.41) can be bounded similarly.

We now prove the second part of the lemma. We first note that

$$\mathbb{P}_0\left[\sum_{j=0}^{J_T-1} T_j - \mathbb{E}_0\left[T_j\right] \geqslant z_T\right] \leqslant \sum_{J\in\mathcal{J}_T} \mathbb{P}_0\left(\sum_{j=0}^{J-1} T_j - \mathbb{E}_0\left[T_j\right] \geqslant z_T\right) + \mathbb{P}_0\left(\tilde{\Omega}_T^c\right)$$
$$\leqslant T\mathbb{P}_0\left(\sum_{j=0}^{J-1} T_j - \mathbb{E}_0\left[T_j\right] \geqslant z_T\right) + o(1). \quad \text{(S7.43)}$$

Let $J \in \mathcal{J}_T$. Since the $\{T_j\}_{1\leqslant j\leqslant J}$ are i.i.d.. random variables, we apply Fuk-Nagaev inequality (see Proposition S10.3) to the sum of centered variables $T_j - \mathbb{E}[T_j]$ with $\lambda := z_T$ and $x := x_T$ with $x_T \to \infty$ a sequence determined later. We denote $v := J\mathbb{E}_0\left[T_1^2\right] \leqslant T\mathbb{E}_0\left[T_1^2\right] \lesssim z_T$. Hence, we have $x\lambda/v = x_T z_T/v \gtrsim x_T$. Since $x_T \to \infty$,

$$\left(1 + \frac{x\lambda}{v}\right)\log\left(1 + \frac{x\lambda}{v}\right) - \frac{x\lambda}{v} \geqslant \frac{x_T\lambda}{v}.$$

From Fuk-Nagaev inequality, we have

$$\mathbb{P}_0\left(\sum_{j=1}^{J}(T_j - \mathbb{E}[T_j]) \geqslant z_T\right) \leqslant J\mathbb{P}_0\left[T_1 - \mathbb{E}[T_1] \geqslant x_T\right] + \exp\left\{-\frac{z_T}{x_T}\right\}. \quad \text{(S7.44)}$$

We note that in the second term on the RHS of (S7.44), if $\frac{z_T}{x_T} \geqslant x_0 \log T$ with $x_0 > 0$ large enough, then $\exp\left\{-\frac{z_T}{x_T}\right\} = o(\frac{1}{T})$. Since by assumption, $\log T = o(T\epsilon_T^2)$, then we can choose $x_T = x_0' \frac{z_T}{\log T} \to \infty$ with $x_0' > 0$ a constant small enough. For the first term on the RHS of (S7.44), let us consider $j \in [J]$. From (S7.40), we have

$$T_1 \leqslant \sum_k \left\{\int_{\tau_1}^{\tau_2} |\lambda_t^k(f) - \lambda_t^k(f_0)|dt + \int_{[\tau_1,\tau_2)} |\log\lambda_t^k(f) - \log\lambda_t^k(f_0)|dN_t^k\right\}.$$

Using the first part of the lemma and Cauchy-Schwarz inequality, we have that $\mathbb{E}_0\left[T_1\right] \leqslant \sqrt{\frac{z_T}{T}} \leqslant x_T$ since $x_T \gtrsim z_T/\log T$ and $\log^3 T = O(z_T)$. Therefore,

$$\mathbb{P}_0\left[T_1 - \mathbb{E}_0\left[T_1\right] \geqslant x_T\right] \leqslant \mathbb{P}_0\left[\tilde{\Omega}_T \cap \left\{\int_{\tau_1}^{\tau_2} |\lambda_t^k(f) - \lambda_t^k(f_0)|dt + \int_{[\tau_1,\tau_2)} |\log\lambda_t^k(f) - \log\lambda_t^k(f_0)| \geqslant x_T\right\}\right] + \mathbb{P}_0\left[\tilde{\Omega}_T^c\right].$$

On the one hand, on $\tilde{\Omega}_T$, under Assumption 3.1(i), using that $|\log x - \log y| \leqslant \frac{|x-y|}{y}$ for $x \geqslant y$,

$$\begin{aligned}\int_{[\tau_1,\tau_2)} |\log\lambda_t^k(f) - \log\lambda_t^k(f_0)|dN_t^k &\leqslant \frac{2}{\min_k \inf_x \phi_k(x)} \int_{[\tau_1,\tau_2)} |\log\lambda_t^k(f) - \log\lambda_t^k(f_0)|dN_t^k \\ &\leqslant \frac{2LN[\tau_1,\tau_2)}{\min_k \inf_x \phi_k(x)}|\nu_k - \nu_k^0| + \frac{2L}{\min_k \inf_x \phi_k(x)} \int_{[\tau_1,\tau_2)^2} |h_{lk} - h_{lk}^0|(t-s)dN_t^k dN_s^k \\ &\leqslant \frac{4L}{\min_k \inf_x \phi_k(x)}(\epsilon_T N[\tau_1,\tau_2) + N[\tau_1,\tau_2)^2 \left\|h_{lk} - h_{lk}^0\right\|_\infty) \leqslant 3LBN[\tau_1,\tau_2)^2,\end{aligned}$$

for T large enough. In Case 2 of Proposition 3.5, we similarly have

$$\int_{[\tau_1,\tau_2)} |\log \lambda_t^k(f) - \log \lambda_t^k(f_0)| dN_t^k \leqslant \frac{2}{\theta_k^0} \int_{[\tau_1,\tau_2)} |\log \lambda_t^k(f) - \log \lambda_t^k(f_0)| dN_t^k$$

$$\leqslant \frac{2N[\tau_1,\tau_2)}{\theta_k^0}(|\theta_k - \theta_k^0| + |\nu_k - \nu_k^0|) + \frac{2}{\theta_k^0}\int_{[\tau_1,\tau_2)}\int_{[\tau_1,\tau_2)} |h_{lk} - h_{lk}^0|(t-s) dN_t^k dN_s^k$$

$$\leqslant \frac{4}{\theta_k^0}\epsilon_T N[\tau_1,\tau_2) + 2N[\tau_1,\tau_2)^2 \left\|h_{lk} - h_{lk}^0\right\|_\infty \leqslant 3BN[\tau_1,\tau_2)^2,$$

Under Assumption 3.1(ii), $\log \phi_k$ is L_1-Lipschitz, therefore,

$$\sum_{t_i \in [\tau_1,\tau_2)} |\log \lambda_{t_i}^k(f) - \log \lambda_{t_i}^k(f_0)| \leqslant L_1 \sum_{t_i \in [\tau_1,\tau_2)} |\tilde{\lambda}_{t_i}^k(\nu,h) - \tilde{\lambda}_{t_i}^k(\nu_0,h_0)| \leqslant L_1 BN[\tau_1,\tau_2)^2.$$

With the ReLU link function, we directly have that $T_1 \leqslant \sum_k \int_{\tau_1}^{\tau_2} (\lambda_t^k(f) - \lambda_t^k(f_0))dt$.
In Case 2 of Proposition 3.5,

$$\int_{\tau_1}^{\tau_2} |\lambda_t^k(f,\theta) - \lambda_t^k(f_0,\theta_0)|dt \leqslant |\theta_k^0 - \theta_k|\Delta\tau_1 + \int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(\nu,h) - \tilde{\lambda}_t^k(\nu_0,h_0))dt$$

$$\leqslant (|\theta_k^0 - \theta_k| + |\nu_k - \nu_k^0|)\Delta\tau_1 + \sum_l \left\|h_{lk} - h_{lk}^0\right\|_1 N^l[\tau_1,\tau_2) \leqslant [2\Delta\tau_1 + N[\tau_1,\tau_2)]\epsilon_T.$$

and in all other cases,

$$\int_{\tau_1}^{\tau_2} |\lambda_t^k(f) - \lambda_t^k(f_0)|dt \leqslant L \int_{\tau_1}^{\tau_2} (\tilde{\lambda}_t^k(\nu,h) - \tilde{\lambda}_t^k(\nu_0,h_0))dt$$

$$\leqslant L|\nu_k - \nu_k^0|)\Delta\tau_1 + L\sum_l \left\|h_{lk} - h_{lk}^0\right\|_1 N^l[\tau_1,\tau_2) \leqslant L[2\Delta\tau_1 + N[\tau_1,\tau_2)]\epsilon_T.$$

Consequently,

$$T_1 \leqslant KC[2\Delta\tau_1 + N[\tau_1,\tau_2)]\epsilon_T + 3KCBN[\tau_1,\tau_2)^2 \leqslant 4KCBN[\tau_1,\tau_2)^2,$$

with $C = \max(1, L, L_1)$ or $C = \max(1, L)$ depending on the assumptions on the link functions, and

$$\mathbb{P}_0 \left[T_1 - \mathbb{E}_0[T_1] \geqslant 2x_T\right] \leqslant \mathbb{P}_0 \left[N[\tau_1,\tau_2)^2 > \frac{x_T}{2KCB}\right].$$

Using Lemma 5.1, we have for some $s > 0$

$$\mathbb{P}_0 \left[N[\tau_1,\tau_2)^2 > \frac{x_T}{2KCB}\right] \leqslant \mathbb{E}_0 \left[e^{sN[\tau_1,\tau_2)}\right] e^{-s\sqrt{x_T/(2KCB)}} = o(T^{-2}),$$

if $x_T \geqslant x_0'' \log^2 T$ for some $x_0'' > 0$ large enough, implying that $z_T \geqslant z_0 \log^3 T$ for some $z_0 > 0$. Finally, reporting into (S7.43), we can conclude that

$$\mathbb{P}_0\left(\sum_{j=1}^{J_T}(T_j - \mathbb{E}[T_j]) \geqslant z_T\right) \leqslant T^2 \mathbb{P}_0\left[T_1 - \mathbb{E}[T_1] \geqslant x_T\right] + T\mathbb{P}_0\left[\tilde{\Omega}_T^c\right] + T\exp\left\{-\frac{z_T}{x_T}\right\} + o(1) = o(1).$$

□

S8. Proofs of identifiability results and regenerative properties of nonlinear Hawkes models

S8.1. Proofs of Proposition 2.3, Proposition 2.5 and Lemma 2.6

In this section, we prove our two propositions on the model identifiability, i.e., Propositions 2.3 and 2.5, as well as Lemma 2.6 in the mutually-exciting Hawkes model. We recall the results in each case.

Proposition S8.1 (Proposition 2.3). *Let N be a nonlinear Hawkes process as defined in (1) with link functions $(\phi_k)_k$ and parameter $f = (\nu, h)$ satisfying the conditions of Lemma 2.1 and Assumption 2.2. If N' is a Hawkes processes with the same link functions $(\phi_k)_k$ and parameter $f' = (\nu', h')$, then*

$$N \overset{\mathcal{L}}{=} N' \implies \nu = \nu' \quad \text{and} \quad h = h'.$$

Proof. Let $f' = (\nu', h')$ and $N' \sim \mathbb{P}_{f'}$. We recall that $N \sim \mathbb{P}_f$ and $N \overset{\mathcal{L}}{=} N'$ is equivalent to $\lambda_t^l(f) = \lambda_t^l(f')$ for all $t > 0$ and $l \in [K]$. Let τ_1 be the first renewal time of the process N, as defined in Section 5.1. From the proof of Lemma 5.1, with $U_1^{(1)}$ the time of the first event after τ_1 and $V^{(1)} \in [K]$ the index of the component associated with this event, we have that $U_1^{(1)} \sim Exp(\|r_f\|_1) \perp\!\!\!\perp V_1^{(1)}$ with $r_f = (r_1^f, \dots, r_K^f)$ and $r_k^f = \phi_k(\nu_k), \forall k,$ and

$$V_1^{(1)} \sim Multi\left(1; \frac{r_1^f}{\|r_f\|_1}, \dots, \frac{r_K^f}{\|r_f\|_1}\right).$$

Therefore we can conclude that

$$N \overset{\mathcal{L}}{=} N' \implies r_f = r_{f'} \iff \phi_k(\nu_k) = \phi_k(\nu_k'), \ \forall k \in [K]. \tag{S8.45}$$

Since for all k, $\nu_k \in I_k$ defined by Assumption 2.2 (ii), then $\nu_k' = \phi_k^{-1}(\phi_k(\nu_k))$ and since ϕ_k is monotone non-decreasing, we obtain $\nu_k = \nu_k', \forall k$.

Moreover, for each $k \in [K]$, we define the event Ω_k as

$$\Omega_k = \left\{ \max_{k' \neq k} N^{k'}[\tau_1, \tau_2) = 0, N^k[\tau_1, \tau_1 + A] = 1, N^k[\tau_1 + A, \tau_2) = 0 \right\}.$$

On Ω_k, for $t \in [\tau_1, \tau_2) \cap [U_1^{(1)}, U_1^{(1)} + A]$ and $l \in [K]$, $\lambda_t^l(f) = \phi_l(\nu_l + h_{kl}(t - U_1^{(1)}))$ and similarly for $\lambda_t^l(f')$. Then, for any $s = t - U_1^{(1)} \in [0, A]$, $\lambda^l_{U_1^{(1)}+s}(f) = \phi_l(\nu_l + h_{kl}(s)) = \phi_l(\nu_l + h'_{kl}(s))$. Consequently, using that ϕ_l is injective on I_l, $h_{kl} = h'_{kl}$ for all $1 \leqslant k, l \leqslant K$ which concludes the proof of this proposition. □

Proposition S8.2 (Proposition 2.5). *Let N be a Hawkes process with parameter $f = (\nu, h)$ and link function $\phi_k(x; \theta_k) = \theta_k + \psi_k(x)$ with $\theta_k \geqslant 0$ for any $k \in [K]$ satisfying the conditions of Lemma 2.1 and Assumption 2.2. We also assume that for all $k \in [K]$, $\lim_{x \to -\infty} \psi_k(x) = 0$ and*

$$\exists l \in [K],\ x_1 < x_2, \quad \text{such that } h_{lk}^-(x) > 0, \quad \forall x \in [x_1, x_2]. \tag{S8.46}$$

Then if N' is a Hawkes processes with link functions $\phi_k(x;\theta'_k) = \theta'_k + \psi_k(x)$, $\theta'_k \geqslant 0$ and parameter $f' = (\nu', h')$,

$$N \overset{\mathcal{L}}{=} N' \implies \nu = \nu', \quad h = h', \quad \text{and} \quad \theta = \theta', \quad \theta = (\theta_k)_{k=1}^K, \quad \theta' = (\theta'_k)_{k=1}^K.$$

Besides, in this case we have $\mathbb{P}_0\left[\inf_{t\geqslant 0} \lambda_t^k(f,\theta) = \theta_k\right] = 1$.

Proof. Using the proof of Proposition 2.3, we first obtain that $\phi_k(\nu_k) = \phi_k(\nu'_k)$, therefore

$$\theta_k + \psi_k(\nu_k) = \theta'_k + \psi_k(\nu'_k),\ \forall k \in [K].$$

Secondly, we also have that $\theta_l + \psi_k(\nu_l + h_{kl}(s)) = \theta'_l + \psi_k(\nu'_l + h'_{kl}(s))$ for any $s \in [0, A]$ and all $1 \leqslant k, l \leqslant K$.

We first prove that $\theta = \theta'$ and from the latter we can deduce that $\nu = \nu'$ and finally that $h = h'$ by the injectivity of ψ_k on I_k, for any k. The proof of the identification of θ relies on the construction of a specific excursion for each $k \in [K]$ in which there exists $t > 0$ such that $\lambda_t^k(f) \in [\theta_k, \theta_k + \epsilon]$ for any $\epsilon > 0$. From that, we will deduce that $N \overset{\mathcal{L}}{=} N' \implies \theta = \theta'$.

Let $k \in [K]$ and consider $l \in [K]$ such that h_{lk} satisfies Assumption S8.46. We first note that

$$\lambda_t^k(f) = \theta_k + \psi_k(\tilde{\lambda}_t^k(\nu, h)) \geqslant \theta_k.$$

Thus, we directly have that $\theta_k \leqslant \inf_{t>0} \lambda_t^k(f)$, a.s. Let $\epsilon > 0$. Using Assumption S8.46 (i), $\exists M > 0, \forall x \leqslant M$, $\psi_k(x) \leqslant \epsilon$. Using now Assumption S8.46 (ii), let $l \in [K]$ and $x_1 < x_2$ such that $[x_1, x_2] \subset B_0 := \{x \in [0, A], h_{lk}(x) \leqslant -c_*\}$. Define $n_1 = \min\{n \in \mathbb{N};\ nc_* > \nu_k^0 - M\}$, $\delta' = (x_2 - x_1)/3$, and we consider an excursion, which we write $[0, \tau]$, and which satisfies

$$\mathcal{E} = \{N[0, \delta'] = N^l[0, \delta'] = n_1,\ N[\delta', \delta' + A] = 0\}.$$

In other words the events only occur on the l-th component of the Hawkes process and only on $[0, \delta']$. Since ψ_k is Lipschitz and injective on $I_k = (\nu_k - \max_l \left\|h_{lk}^-\right\|_\infty - \varepsilon, \nu_k + \max_l \left\|h_{lk}^+\right\|_\infty + \varepsilon)$, it holds that $\mathbb{P}_f[\mathcal{E}] > 0$. For $t \in [x_1 + \delta', x_2]$, $\forall i \in [n_1]$, we have $x_1 \leqslant t - t_i \leqslant x_2$, and therefore,

$$\tilde{\lambda}_t^k(\nu, h) = \nu_k + \sum_{i \in [n_1]} h_{lk}(t - t_i) \leqslant \nu_k - n_1 c_* \leqslant M.$$

Consequently, for $t \in [x_1 + \delta', x_2]$, $\lambda_t^k(f_0) = \theta_k + \psi_k(\tilde{\lambda}_t^k(\nu, h)) \leqslant \theta_k + \epsilon$. We can then conclude that

$$\mathbb{P}_0\left[\exists t \geqslant 0,\ \lambda_t^k(f) \in [\theta_k, \theta_k + \epsilon]\right] > 0,$$

for any $\epsilon > 0$. This is equivalent to

$$\theta_k = \inf_{\omega \in \Omega} \inf_{t \in [0,\tau]} \lambda_t^k(f)(\omega),$$

where $\lambda_t^k(f_0)(\omega)$ denotes the value of the random process $(\lambda_t(f_0))_t$ at time t.

Now, if N' is a Hawkes process with parameter $f' \in \mathcal{F}$ and link functions $\phi_k = \theta'_k + \psi_k$, $k \in [K]$ such that $N \overset{\mathcal{L}}{=} N'$, then for any $t \geqslant 0$ and k such $\lambda_t^k(f) \leqslant \theta_k + \epsilon$, we have $\theta'_k \leqslant \lambda_t^k(f') \leqslant \theta_k + \epsilon$ and thus, $\theta_k \geqslant \theta'_k$. Inversely, if $\lambda_t^k(f') \leqslant \theta'_k + \epsilon$ then $\theta_k \leqslant \theta'_k$ and finally we can conclude that $\theta = \theta'$.

□

Lemma S8.3 (Lemma 2.6). *Let N be a Hawkes process with parameter $f = (\nu, h)$ and link functions $\phi_k(x;\theta_k) = \theta_k + (x)_+$, $\theta_k \geqslant 0$, $k \in [K]$ satisfying Assumption 2.2, and let $k \in [K]$. If $\forall l \in [K], h_{lk} \geqslant 0$, then for any $\theta'_k \geqslant 0$ such that $\theta_k + \nu_k - \theta'_k > 0$, let N' be the Hawkes process driven by the same underlying Poisson process Q as N (see Lemma S10.2) with parameter $f' = (\nu', h')$ and link functions $\phi_k(x;\theta'_k) = \theta'_k + (x)_+, k \in [K]$ with $\nu' = (\nu_1, \dots, \nu_k + \theta_k - \theta'_k, \dots, \nu_K) \neq \nu$, $h' = h$, and $\theta' = (\theta_1, \dots, \theta'_k, \dots, \theta_K) \neq \theta$. Then for any $t \geqslant 0$, $\lambda^k_t(f,\theta) = \lambda^k_t(f',\theta')$, and therefore $N \overset{\mathcal{L}}{=} N'$.*

Proof. We consider $k \in [K]$ such that $\forall l \in [K]$, $h_{lk} \geqslant 0$. For any $t \geqslant 0$, we have

$$\tilde{\lambda}^k_t(\nu, h) = \nu_k + \sum_l \int_{t-A}^{t^-} h_{lk}(t-s)dN^l_s \geqslant \nu_k > 0,$$

and thus $\lambda^k_t(f) = \theta_k + (\tilde{\lambda}^k_t(\nu,h))_+ = \theta_k + \tilde{\lambda}^k_t(\nu,h)$. Moreover, for any $t \geqslant 0$, we have

$$\tilde{\lambda}^k_t(\nu', h') = \nu_k + \theta_k - \theta'_k + \sum_l \int_{t-A}^{t^-} h_{lk}(t-s)dN^l_s \geqslant \nu_k + \theta_k - \theta'_k > 0,$$

and

$$\begin{aligned}\lambda^k_t(f') &= \theta'_k + (\tilde{\lambda}^k_t(\nu',h'))_+ = \theta'_k + \tilde{\lambda}^k_t(\nu',h') \\ &= \theta'_k + \nu_k + \theta_k - \theta'_k + \sum_l \int_{t-A}^{t^-} h_{lk}(t-s)dN^l_s = \theta_k + \tilde{\lambda}^k_t(\nu,h) = \lambda^k_t(f).\end{aligned}$$

Therefore, we obtain that $N =^{\mathcal{L}} N'$.

□

S8.2. Proofs of Lemmas 5.2 and 5.4

In this section, we prove our lemmas related to the renewal properties of the nonlinear Hawles processes, in particular the existence of exponential moments for the generic renewal time $\Delta\tau_1$, and a concentration inequality on J_T. the nunber of excursions in the interval of observation $[0,T]$.

Lemma S8.4 (Lemma 5.2). *Under the assumptions of Lemma 5.1, the random variables $\Delta\tau_1$ and $N[\tau_1,\tau_2)$ admit exponential moments. More precisely, under condition **(C1bis)**, with $m = \|S^+\| < 1$, we have*

$$\forall s < \min(\|r_f\|_1, \gamma/A), \quad \mathbb{E}_f\left[e^{s\Delta\tau_1}\right] < +\infty, \quad \text{and} \quad \mathbb{E}_f\left[e^{sN[\tau_1,\tau_2)}\right] < +\infty, \quad \gamma = \frac{1-m}{2\sqrt{K}} \log\left(\frac{1+m}{2m}\right).$$

*Under condition **(C2)**, we have $\forall s < \min_k \Lambda_k$, $\mathbb{E}_f\left[e^{s\Delta\tau_1}\right] \leqslant \frac{\|\Lambda\|_1^2}{(\min_k \Lambda_k - s)^2}$ and $\mathbb{E}_f\left[e^{sN[\tau_1,\tau_2)}\right] < +\infty$. In particular, this implies that $\mathbb{E}_f\left[N[\tau_1,\tau_2) + N[\tau_1,\tau_2)^2\right] < +\infty$.*

Proof. Under condition **(C1bis)**, similarly to Costa et al. (2020), we use the fact that the multivariate Hawkes model is stochastically dominated by a mutually-exciting process N^+ with parameter $f^+ = (\nu, (h^+_{lk})_{l,k})$, and driven by the same Poisson process as N (see Lemma S10.2). For N^+, the stopping time $\Delta\tau_1^+$ corresponds to the length of the busy period of a $M^K/G^K/\infty$ queue (see Lemma S10.1, which is a multi-type extension of existing results).

More precisely, since N^+ is mutually-exciting, the cluster representation is available Reynaud-Bouret and Roy (2007), with the ancestor arrival process being a Poisson Point Process equal to the baseline rate r_f, defined in (21). For this process, the duration of the clusters then corresponds to the generic service time H of a queue with an infinite number of servers. In the multidimensional case, this duration may depend on the type of the ancestor (or "customer" in the queuing framework) but the generic service time can be written in a compact form, and is independent of the arrival process

$$H = \sum_{k=1}^{K} \delta_k H^k,$$

where $\delta_k = 1$ if and only if the ancestor is of type $k \in [K]$. To apply Lemma S10.1, we only need to check that the cluster length H^k, $k \in [K]$ has exponential moments. This can be proved using results from Donnet, Rivoirard and Rousseau (2020).

For the process N^+, let W^k be the number of events in a cluster with an ancestor of type k. By definition of a cluster of events, $H^k \leqslant AW^k$. Moreover, from Lemma 5 in the Supplementary Materials of Donnet, Rivoirard and Rousseau (2020), for a mutually-exciting Hawkes process and for any $t \leqslant \frac{1-\|S^+\|_1}{2\sqrt{K}} \log\left(\frac{1+\|S^+\|}{2\|S^+\|}\right)$ and $k \in [K]$,

$$\mathbb{E}_f\left[e^{tW^k}\right] \leqslant \frac{1+\|S^+\|}{2\|S^+\|}.$$

Therefore, we define $\gamma = (1-\|S^+\|)\left[\log\left(1+\|S^+\|\right) - \log(2\|S^+\|)\right]/(2\sqrt{K})$ and $s_0 = \frac{1+\|S^+\|}{2\|S^+\|}$. For all $0 < t \leqslant \gamma$, we thus have $\mathbb{E}_f\left[e^{tH^k/A}\right] \leqslant s_0$. Consequently, we deduce that the service time H^k has exponential tails, i.e., $\mathbb{P}_f\left[H^k \geqslant t\right] \leqslant s_0 e^{-t\gamma/A}$. We can now use the fact that a.s. $\mathcal{T}_1 = \Delta\tau_1^+$ (cf Lemma S10.2), so that for any $s < \|r_f\|_1 \wedge \gamma/A$, we have $\mathbb{E}_f\left[e^{s\Delta\tau_1^+}\right] < \infty$. Finally using the second part of Lemma S10.2, we have that $\mathbb{P}_f\left[\Delta\tau_1 \leqslant \Delta\tau_1^+\right] = 1$ and, using Lemma S10.1, we arrive at $\forall s < \|r_f\|_1 \wedge \gamma/A$, $\mathbb{E}_f\left[e^{s\Delta\tau_1}\right] < \infty$.

Under condition **(C2)**, we use the fact that the process N is dominated by a K-dimensional homogeneous Poisson point process $N_P = (N_P^1, \dots, N_P^K)$ with rate $\Lambda = (\Lambda_1, \dots, \Lambda_K)$. For the latter process, the generic service time of an ancestor of type k, H_k, is exponentially distributed with mean Λ_k, i.e.,

$$\mathbb{P}_f\left[H_k > t\right] = e^{-\Lambda_k t}, \quad t \geqslant 0.$$

Therefore, denoting $\Delta\tau_1^P$, the corresponding generic stopping time of N^P - with the same definition as in Lemma 5.1 for the Hawkes process (note that the Poisson point process is a renewal process), we have

$$\mathbb{P}_f\left[\Delta\tau_1^P > t\right] \leqslant \mathbb{E}_f\left[N^P[0,t]\right] e^{-\min_k \Lambda_k (t-A)} = \|\Lambda\|_1\, t e^{-\min_k \Lambda_k (t-A)}.$$

Therefore, for any $s < \min_k \Lambda_k$,

$$\mathbb{E}_f\left[e^{s\Delta\tau_1}\right] \leqslant \mathbb{E}_f\left[e^{s\Delta\tau_1^P}\right] = \int_0^{+\infty} s e^{st} \mathbb{P}_f\left[\Delta\tau_1^P \geqslant t\right] dt \leqslant \|\Lambda\|_1^2\, e^{\min_k \Lambda_k A} \int_0^{+\infty} t e^{t(s-\min_k \Lambda_k)} dt$$

$$\leqslant \|\Lambda\|_1^2 \int_0^{+\infty} \frac{e^{t(s-\min_k \Lambda_k)}}{\min_k \Lambda_k - s} dt = \frac{\|\Lambda\|_1^2}{(\min_k \Lambda_k - s)^2}.$$

We now consider the number of events in a excursion $N[\tau_1, \tau_2)$. Under condition **(C1bis)**, From Lemma S10.2, we can also deduce that $\mathbb{E}_f[N[\tau_1, \tau_2)] \leqslant \mathbb{E}_f\left[N^+[\tau_1^+, \tau_2^+)\right]$. We once again use the cluster representation available for N^+. For the latter, let n^τ be the number of ancestors arriving in $[\tau_1^+, \tau_2^+)$ and W_i be the number of points in the cluster with ancestor i for $1 \leqslant i \leqslant n_\tau$. We denote $(NP_t)_t$ the homogeneous Poisson process of intensity $\left\|r_f\right\|_1$ corresponding to the arrival times of the ancestors. By definition of τ_1^+, τ_2^+, we have

$$N^+[\tau_1^+, \tau_2^+) = \sum_{i=1}^{n_\tau} W_i. \tag{S8.47}$$

Let $\gamma > s > 0$ and $u < \left\|r_f\right\|_1 \wedge \gamma/A$. With $t = \mathbb{E}_f\left[e^{sW_1}\right] \leqslant s_0$, since the W_i's are independent conditionally on n_τ,

$$\begin{aligned}
\mathbb{E}_f\left[e^{sN[\tau_1,\tau_2)}\right] &\leqslant \mathbb{E}_f\left[e^{s\sum_{i=1}^{n_\tau} W_i}\right] = \mathbb{E}_f\left[\mathbb{E}_f\left[e^{s\sum_{i=1}^{n_\tau} W_i}|n_\tau\right]\right] = \mathbb{E}_f\left[\mathbb{E}_f\left[e^{sW_1}\right]^{n_\tau}\right] = \mathbb{E}_f\left[\sum_{l=A}^{+\infty} e^{sn_\tau}\mathbb{1}_{\Delta\tau_1\in[l,l+1)}\right] \\
&\leqslant \sum_{l=A}^{+\infty} \mathbb{E}_f\left[e^{sNP[\tau_1,\tau_1+l+1)}\mathbb{1}_{\Delta\tau_1\geqslant l}\right] \leqslant \sum_{l=A}^{+\infty} \sqrt{\mathbb{E}_f\left[e^{2sNP[\tau_1,\tau_1+l+1)}\right]}\sqrt{\mathbb{P}_f\left[\Delta\tau_1 > l\right]} \\
&\leqslant \sqrt{\mathbb{E}_f\left[e^{u\Delta\tau_1}\right]}\sum_{l=A}^{+\infty} \sqrt{\mathbb{E}_f\left[e^{2sNP[\tau_1,\tau_1+l+1)}\right]}e^{-ul/2} = \sqrt{\mathbb{E}_f\left[e^{u\Delta\tau_1}\right]}\sum_{l=A}^{+\infty} e^{\left\|r_f\right\|_1(l+1)(e^{2s}-1)/2}e^{-ul/2},
\end{aligned}$$

since NP is a homogoneous Poisson process with rate $\left\|r_f\right\|_1$. Moreover, since for any $\alpha \in (0,1)$, $\mathbb{E}_f\left[e^{\alpha sW_1}\right] = (\mathbb{E}_f\left[e^{\alpha sW_1}\right]^{1/\alpha})^\alpha \leqslant \mathbb{E}_f\left[e^{sW_1}\right]^\alpha \leqslant s_0^\alpha$, with $t' = \mathbb{E}_f\left[e^{\alpha sW_1}\right]$, we have that $\left\|r_f\right\|_1(e^{2t'} - 1) < u/2$ for α small enough. Consequently,

$$\mathbb{E}_f\left[e^{sN[\tau_1,\tau_2)}\right] \leqslant \sqrt{\mathbb{E}_f\left[e^{u\Delta\tau_1}\right]}\sum_{l=A}^{+\infty} e^{-ul/4} = \frac{\sqrt{\mathbb{E}_0\left[e^{u\Delta\tau_1}\right]}}{1 - e^{-u/4}} < \infty.$$

In particular, this implies that $\mathbb{E}_f[N[\tau_1, \tau_2)] + \mathbb{E}_f\left[N[\tau_1, \tau_2)^2\right] < \infty$. Under condition **(C2)**, the dominating process N^+ is a homogeneous Poisson process with intensity $\Lambda = (\Lambda_1, \dots, \Lambda_K)$ and the previous computations remain valid by replacing r_f by Λ and with $W_i = 1$ for any $i \in [n_\tau]$ (since in this case each cluster only contains the "ancestor" event).

□

Lemma S8.5 (Lemma 5.4)**.** *Under the assumptions of Lemma 5.1, for any $\beta > 0$, there exists a constant $c_\beta > 0$ such that $\mathbb{P}_f\left[J_T \notin [J_{T,\beta,1}, J_{T,\beta,2}]\right] \leqslant T^{-\beta}$, with J_T defined in (19) and*

$$J_{T,\beta,1} = \left\lfloor \frac{T}{\mathbb{E}_f[\Delta\tau_1]}\left(1 - c_\beta\sqrt{\frac{\log T}{T}}\right)\right\rfloor, \quad J_{T,\beta,2} = \left\lfloor \frac{T}{\mathbb{E}_f[\Delta\tau_1]}\left(1 + c_\beta\sqrt{\frac{\log T}{T}}\right)\right\rfloor.$$

Proof. Let $c_\beta > 0$ and for $2 \leqslant j \leqslant J_T$, $B_j = \tau_j - \tau_{j-1} - \mathbb{E}_f[\Delta\tau_1]$. Using Lemma 5.1, the random variables $\{B_j\}_{2\leqslant j\leqslant J_T}$ are i.i.d.. By definition of $J_{T,\beta,2}$, we have

$$\frac{T}{\mathbb{E}_f[\Delta\tau_1]}\left(1 + c_\beta\sqrt{\frac{\log T}{T}}\right) - 1 < J_{T,\beta,2} \leqslant \frac{T}{\mathbb{E}_f[\Delta\tau_1]}\left(1 + c_\beta\sqrt{\frac{\log T}{T}}\right).$$

Therefore,

$$\mathbb{P}_f\left[J_T \geqslant J_{T,\beta,2}\right] = \mathbb{P}_0\left[\tau_{J_{T,\beta,2}} \leqslant T\right] = \mathbb{P}_f\left[\tau_0 + \sum_{j=1}^{J_{T,\beta,2}} B_j \leqslant T - J_{T,\beta,2}\mathbb{E}_f\left[\Delta\tau_1\right]\right]$$

$$= \mathbb{P}_f\left[\sum_{j=1}^{J_{T,\beta,2}} B_j \leqslant T - J_{T,\beta,2}\mathbb{E}_f\left[\Delta\tau_1\right]\right] \leqslant \mathbb{P}_f\left[\sum_{j=1}^{J_{T,\beta,2}} B_j \leqslant T - T\left(1 + c_\beta\sqrt{\frac{\log T}{T}}\right) + \mathbb{E}_f\left[\Delta\tau_1\right]\right]$$

$$= \mathbb{P}_f\left[\sum_{j=0}^{J_{T,\beta,2}} B_j \leqslant -c_\beta\sqrt{T\log T} + \mathbb{E}_f\left[\Delta\tau_1\right]\right] \leqslant \mathbb{P}_f\left[\sum_{j=1}^{J_{T,\beta,2}} B_j \leqslant -\frac{c_\beta\sqrt{T\log T}}{2}\right].$$

We can now apply the Bernstein's inequality. Using Lemma 5.2, there exists $\alpha > 0$, such that $\mathbb{E}_f\left[e^{\alpha\Delta\tau_1}\right] < +\infty$. Since

$$\mathbb{E}_f\left[e^{\alpha\Delta\tau_1}\right] = \sum_{k=1}^{+\infty}\frac{\alpha^k\mathbb{E}_f\left[(\Delta\tau_1)^k\right]}{k!},$$

we therefore have that

$$\mathbb{E}_f\left[(\Delta\tau_1)^k\right] \leqslant \frac{k!}{\alpha^k}\mathbb{E}_f\left[e^{\alpha\Delta\tau_1}\right] = \frac{1}{2}k!\alpha^{-k+2}\times 2\frac{\mathbb{E}_f\left[e^{\alpha\Delta\tau_1}\right]}{\alpha^2}.$$

In particular, $\mathbb{E}_f\left[(\Delta\tau_1)^2\right] \leqslant 2\frac{\mathbb{E}_0\left[e^{\alpha\Delta\tau_1}\right]}{\alpha^2} =: v$. Consequently, with $b := 1/\alpha$, we obtain $\mathbb{E}_f\left[(\Delta\tau_1)^k\right] \leqslant \frac{1}{2}k!b^{k-2}v$, and therefore,

$$\mathbb{P}_f\left[J_T \geqslant J_{T,\beta,2}\right] \leqslant \exp\left\{\frac{-c_\beta^2 T\log T}{8(\sigma^2 + \frac{c_\beta}{2}\sqrt{T\log T}b)}\right\},$$

with

$$\sigma^2 = \sum_{j=1}^{J_{T,\beta,2}}\mathbb{V}_f(B_j) = J_{T,\beta,2}\mathbb{V}_f(\Delta\tau_1) \leqslant T\left(1 + c_\beta\sqrt{\frac{\log T}{T}}\right)\frac{\mathbb{E}_f\left[\Delta\tau_1^2\right]}{\mathbb{E}_f\left[\Delta\tau_1\right]} \leqslant 2T\frac{\mathbb{E}_f\left[\Delta\tau_1^2\right]}{\mathbb{E}_f\left[\Delta\tau_1\right]},$$

for T large enough. Therefore, $\sigma^2 + \frac{c_\beta}{2}\sqrt{T\log T}b \leqslant 4T\frac{\mathbb{E}_f\left[\Delta\tau_1^2\right]}{\mathbb{E}_f\left[\Delta\tau_1\right]}$ and

$$\mathbb{P}_f\left[J_T \geqslant J_{T,\beta,2}\right] \leqslant \exp\left\{\frac{-c_\beta^2\log T\mathbb{E}_f\left[\Delta\tau_1\right]}{32\mathbb{E}_f\left[\Delta\tau_1^2\right]}\right\} = o(T^{-\beta}),$$

for any $\beta > 0$, if $c_\beta > 0$ is chosen large enough. Consequently, with probability greater than $1 - \frac{1}{2}T^{-\beta}$, we have that $J_T \leqslant \frac{T}{\mathbb{E}_f[\Delta\tau_1]}\left(1 + c_\beta\sqrt{\frac{\log T}{T}}\right)$. Similarly, we obtain that

$$\mathbb{P}_f\left[J_T \leqslant J_{T,\beta,1}\right] \leqslant \mathbb{P}_f\left[\sum_{j=1}^{J_{T,\beta,1}} B_j \geqslant c_\beta\sqrt{T\log T}\right] \leqslant \exp\left\{\frac{-c_\beta^2 T\log T}{2(\sigma^2 + c_\beta\sqrt{T\log T}b)}\right\}$$

$$\leqslant \exp\left\{\frac{-c_\beta^2 \log T \mathbb{E}_f[\Delta\tau_1]}{4\mathbb{E}_f\left[\Delta\tau_1^2\right]}\right\} = o(T^{-\beta}).$$

Finally, we conclude that with probability greater than $1 - T^{-\beta}$, $J_{T,\beta,1} \leqslant J_T \leqslant J_{T,\beta,2}$.

□

S9. Proof of Lemmas A1 and A.4

S9.1. Proof of Lemma A.1

Lemma S9.1 (Lemma A.1). *Let $Q > 0$. We consider $\tilde{\Omega}_T$ defined in (25) in Section 5.2. For any $\beta > 0$, we can choose C_β and c_β in the definition of $\tilde{\Omega}_T$ such that*

$$\mathbb{P}_0[\tilde{\Omega}_T^c] \leqslant T^{-\beta}.$$

Moreover, for any $1 \leqslant q \leqslant Q$, $\mathbb{E}_0\left[\mathbb{1}_{\tilde{\Omega}_T^c} \max_l \sup_{t\in[0,T]} \left(N^l[t-A,t)\right)^q\right] \leqslant 2T^{-\beta/2}$. Finally, the previous results hold when replacing $\tilde{\Omega}_T$ by $\tilde{\Omega}'_T = \tilde{\Omega}_T \cap \Omega_A$ with Ω_A defined in Section S1 for the model with shifted ReLU link and unknown shift.

Proof. Let $\beta > 0$. From the definition of $\tilde{\Omega}_T$, we have that

$$\mathbb{P}_0[\tilde{\Omega}_T^c] \leqslant \mathbb{P}_0[\Omega_N^c] + 3\mathbb{P}_0[\Omega_J^c] + \mathbb{P}_0[\Omega_J \cap \Omega_U^c]. \tag{S9.48}$$

For the second term on the RHS of (S9.48), we can directly use Lemma 5.4, and we obtain $\mathbb{P}_0[\Omega_J^c] \leqslant \frac{1}{12}T^{-\beta}$ for c_β large enough. For the first term on the RHS of (S9.48), we use the same strategy as in Donnet, Rivoirard and Rousseau (2020). Firstly we have

$$\mathbb{P}_0[\Omega_N^c] \leqslant \mathbb{P}_0\left[\max_{k\in[K]} \sup_{t\in[0,T]} N^k[t-A,t) > C_\beta \log T\right] + \sum_{k=1}^K \mathbb{P}_0\left[\left|\frac{N^k[0,T]}{T} - \mu_k^0\right| \geqslant \delta_T\right]. \tag{S9.49}$$

For the first term on the RHS of (S9.49), we use the coupling with the process N^+, i.e., the Hawkes process with parameter $f_0^+ = (\nu_0, h_0^+)$ driven by the same Poisson process. Then for any $l \in [K]$, $\sup_{t\in[0,T]} N^l[t-A,t) \leqslant \sup_{t\in[0,T]} (N^+)^l[t-A,t)$ and consequently,

$$\mathbb{P}_0\left[\max_{k\in[K]} \sup_{t\in[0,T]} N^k[t-A,t) > C_\beta \log T\right] \leqslant \mathbb{P}_0\left[\max_{k\in[K]} \sup_{t\in[0,T]} (N^+)^k[t-A,t) > C_\beta \log T\right].$$

Using Lemma 2 from Donnet, Rivoirard and Rousseau (2020), we obtain that for any $\beta > 0$, there exists $C_\beta > 0$ such that

$$\mathbb{P}_0\left[\max_{k\in[K]} \sup_{t\in[0,T]} (N^+)^k[t-A,t) > C_\beta \log T\right] \leqslant \frac{1}{4}T^{-\beta}.$$

For the second term on the RHS of (S9.49), we use the same arguments as in the proof of Lemma 3 in Donnet, Rivoirard and Rousseau (2020). For $k \in [K]$, we have

$$\mathbb{P}_0\left[\left|\frac{N^k[0,T]}{T} - \mu_k^0\right| \geqslant \delta_T\right] \leqslant \mathbb{P}_0\left[\left|N^k[0,T] - \int_0^T \lambda_t^k(f_0)\right| \geqslant T\delta_T/2\right] + \mathbb{P}_0\left[\left|\int_0^T \lambda_t^k(f_0) - \mu_k^0 T\right| \geqslant T\delta_T/2\right]. \quad \text{(S9.50)}$$

For the second term on the RHS of (S9.50), we can use Corollary 1.1 from Costa et al. (2020). We have that $\lambda_t^k(f_0) = Z(S_t N)$, with

$$Z(N) = \lambda_0^k(f_0) = \phi_k\left(\nu_k^0 + \sum_l \int_{-A}^{0^-} h_{lk}(t-s)dN_s^l\right) \leqslant Lb(1 + N[-A,0)),$$

with $b = \max(\nu_k^0, \max_l \left\|h_{lk}^{0+}\right\|_\infty)$ and for $t \in \mathbb{R}$, $S_t : \mathcal{N}(\mathbb{R}) \to S_t N = N(.+t)$ the shift operator by t units of time. Applying Corollary 1.1 of Costa et al. (2020) with $f = Z$, $\pi_A f = \mathbb{E}_0\left[\lambda_0^k(f_0)\right] = \mu_k^0$, $\varepsilon = \delta_T/2$ and $\eta = \frac{1}{4}T^{-\beta}$, we obtain that for δ_0 large enough,

$$\mathbb{P}_0\left[\left|\int_0^T \lambda_t^k(f_0) - \mu_k^0 T\right| \geqslant T\delta_T/2\right] \leqslant \frac{1}{4}T^{-\beta}.$$

For the first term on the RHS of (S9.50), we use the computations of the proof Lemma 3 in the Supplementary Materials of Donnet, Rivoirard and Rousseau (2020) and obtain

$$\mathbb{P}_0\left[\left|N^k[0,T] - \int_0^T \lambda_t^k(f_0)\right| \geqslant T\delta_T/2\right] \leqslant \frac{1}{4}T^{-\beta},$$

for δ_0 large enough.

For the third term on the RHS of (S9.48), we denote $X_j = U_j^{(1)} - \tau_j$ for $1 \leqslant j \geqslant J_T - 1$. We recall that the X_j's are i.i.d. and follow an exponential law with rate $\|r_0\|_1$ under $\mathbb{P}_0$ and $\mathbb{E}_0\left[X_j\right] = \frac{1}{\|r_0\|_1}$. We thus have

$$\begin{aligned}
\mathbb{P}_0[\Omega_J \cap \Omega_U^c] &\leqslant \mathbb{P}_0\left[\Omega_J \cap \left\{\sum_{j=1}^{J_T-1} X_j \leqslant \frac{T}{\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1}\left(1 - 2c_\beta\sqrt{\frac{\log T}{T}}\right)\right\}\right] \\
&\leqslant \mathbb{P}_0\left[\Omega_J \cap \left\{\sum_{j=1}^{J_T-1} X_j - \frac{J_T-1}{\|r_0\|_1} \leqslant \frac{T}{\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1}\left(1 - 2c_\beta\sqrt{\frac{\log T}{T}} - 1 + c_\beta\sqrt{\frac{\log T}{T}}\right)\right\}\right] \\
&= \mathbb{P}_0\left[\Omega_J \cap \left\{\sum_{j=1}^{J_T-1} X_j - \frac{J_T-1}{\|r_0\|_1} \leqslant -\frac{c_\beta\sqrt{T\log T}}{\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1}\right\}\right] \leqslant \sum_{J \in \mathcal{J}_T} \mathbb{P}_0\left[\sum_{j=1}^{J-1} X_j - \frac{J-1}{\|r_0\|_1} \leqslant -\frac{c_\beta\sqrt{T\log T}}{\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1}\right],
\end{aligned}$$

where in the first inequality we have used the fact that on Ω_J,

$$J_T - 1 \geqslant \frac{T}{\mathbb{E}_0[\Delta\tau_1]}\left(1 - c_\beta\sqrt{\frac{\log T}{T}}\right).$$

We apply the Bernstein's inequality using that for any $k \geqslant 1$, $\mathbb{E}_0\left[X_1^k\right] \leqslant k!(\|r_0\|_1)^{-k+2}\mathbb{E}_0\left[X_1^2\right]/2$. Therefore, since $\mathbb{E}_0\left[X_1^2\right] = \|r_0\|_1^{-2}$, we obtain

$$\mathbb{P}_0\left[\sum_{j=1}^{J-1} X_j - \frac{J-1}{\|r_0\|_1} \leqslant -\frac{c_\beta\sqrt{T\log T}}{\mathbb{E}_0[\Delta\tau_1]\|r_0\|_1}\right] \leqslant \exp -\left\{\frac{c_\beta^2 \log T}{\mathbb{E}_0\left[\Delta\tau_1\right]^2\left(1 + \frac{c_\beta\sqrt{\log T}}{\mathbb{E}_0[\Delta\tau_1]\sqrt{T}}\right)}\right\}$$

$$\leqslant \exp -\left\{\frac{c_\beta^2\log T}{2\mathbb{E}_0\left[\Delta\tau_1\right]}\right\} \leqslant \frac{1}{4}T^{-\beta},$$

for $c_\beta > 0$ large enough. Finally, reporting into (S9.48) we can conclude that for $C_\beta, c_\beta, \delta_0$ large enough,

$$\mathbb{P}_0\left[\tilde{\Omega}_T^c\right] \leqslant T^{-\beta}.$$

For the second part of the lemma, we can use the exact same arguments as in the proof of Lemma 2 in Donnet, Rivoirard and Rousseau (2020) to obtain the result.

For the case of shifted ReLU link function with unknown shift, we similarly have that

$$\mathbb{P}_0[\tilde{\Omega}_T'^c] \leqslant \mathbb{P}_0[\Omega_N^c] + 3\mathbb{P}_0[\Omega_J^c] + \mathbb{P}_0[\Omega_J \cap \Omega_U^c] + \mathbb{P}_0\left[\Omega_J \cap \Omega_A^c\right], \tag{S9.51}$$

and therefore it only remains to bound the last term on the RHS of the previous inequality. Using Assumption S8.46 (ii), let $0 < x_1 < x_2$ and $c_\star$ such that $[x_1, x_2] \subset B_0 = \{x \in [0, A], h_{lk}^0(x) \leqslant -c_*\}$, $n_1 = \min\{n \in \mathbb{N}; nc_* > \nu_k^0\}$, $\delta' = (x_2 - x_1)/3$. We denote $\mathcal{E}_0$ the set of indices satisfying

$$\mathcal{E}_0 = \{j \in [J_T];\ N[\tau_j, \tau_j + \delta'] = N^l[\tau_j, \tau_j + \delta'] = n_1,\ N[\tau_j + \delta', \tau_{j+1}) = 0\}.$$

Since $\forall t \in [\tau_j + x_1 + \delta', \tau_j + x_2]$, $\tilde{\lambda}_t^k(f) < 0$, then $|A^k(f_0)| \geqslant \frac{2(x_2-x_1)}{3}|\mathcal{E}_0|$ and, with $p_0 = \mathbb{P}_0\left[j \in \mathcal{E}_0\right]$,

$$\mathbb{P}_0\left[|A^k(f_0)| < z_0 T\right] \leqslant \mathbb{P}_0\left[|\mathcal{E}_0| < \frac{3z_0}{2(x_2 - x_1)}T\right] \leqslant \mathbb{P}_0\left[|\mathcal{E}_0| < p_0T/2\right],$$

if $z_0 < 2p_0(x_2 - x_1)/3$. Consequently, applying Hoeffding's inequality with $Y_j = \mathbb{1}_{j\in\mathcal{E}_0} \overset{i.i.d.}{\sim} \mathcal{B}(p_0)$ for $j \in [J_T]$ with $J_T \geqslant 2T/3\mathbb{E}_0\left[\Delta\tau_1\right]$, we obtain

$$\mathbb{P}_0\left[|\mathcal{E}_0| < \frac{p_0T}{2}\right] \leqslant \mathbb{P}_0\left[\sum_{j=1}^{2T/3\mathbb{E}_0[\Delta\tau_1]} Y_j < \frac{p_0T}{2}\right] \lesssim e^{-\frac{Tp_0^2}{6\mathbb{E}_0[\Delta\tau_1]}} \leqslant \frac{1}{4}T^{-\beta}.$$

Consequently, $\mathbb{P}_0\left[\Omega_J \cap \Omega_A^c\right] = o(T^{-\beta})$, which terminates the proof of this lemma.

□

S9.2. Proof of Lemma A.4

Lemma S9.2 (Lemma A.4)**.** *For $f \in \mathcal{F}_T$ and $l \in [K]$, let*

$$Z_{1l} = \int_{\tau_1}^{\xi_1} |\lambda_t^l(f) - \lambda_t^l(f_0)|dt,$$

where ξ_1 is defined in (22) in Section 5.2. Under the assumptions of Theorem 3.2 and Case 1 of Proposition 3.5, for $M_T \to \infty$ such that $M_T > M\sqrt{\kappa_T}$ with $M > 0$ and for any $f \in \mathcal{F}_T$ such that $\|\nu - \nu_0\|_1 \leqslant \max(\|\nu_0\|_1, \tilde{C})$ with $\tilde{C} > 0$, there exists $l \in [K]$ such that on $\tilde{\Omega}_T$,

$$\mathbb{E}_f[Z_{1l}] \geqslant C(f_0)\|f - f_0\|_1,$$

with $C(f_0) > 0$ a constant that depends only on f_0 and $\phi = (\phi_k)_k$.

Similarly, under the assumptions of Case 2 of Proposition 3.5, for $f \in \mathcal{F}_T$ and $\theta \in \Theta$, let $r_0 = (r_k^0)_k$, $r_f = (r_k^f)_k$ with $r_k^0 = \phi_k(\nu_k^0) = \theta_k^0 + \nu_k^0$, $r_k^f = \phi_k(\nu_k) = \theta_k + \nu_k$, $\forall k$. If $\left\|r_f - r_0\right\|_1 \leqslant \max(\|r_0\|, \tilde{C}')$ with $\tilde{C}' > 0$, then there exists $l \in [K]$ such that on $\tilde{\Omega}_T$,

$$\mathbb{E}_f[Z_{1l}] \geqslant C'(f_0)(\|r_f - r_0\|_1 + \|h - h_0\|_1), \quad C'(f_0) > 0. \tag{S9.52}$$

Proof. In this proof, we will show that (S9.52) holds for all the models satisfying the assumptions of Theorem 3.2 and Proposition 3.5, with $r_k^0 = \phi_k(\nu_k^0)$ and $r_k^f = \phi_k(\nu_k)$ for all k. Then, excluding Case 2, we use the fact that for any k, ϕ_k^{-1} is fully known and L'-Lipshitz on $J_k = \phi_k(I_k)$ with I_k defined in Assumption 3.1 (which also holds for the ReLU link function by Assumption 2.2), to show that

$$\begin{aligned}\|r_f - r_0\|_1 + \|h - h_0\|_1 &\geqslant 1/L'\|\nu - \nu_0\|_1 + \|h - h_0\|_1 \\ &\geqslant \min(1, 1/L')(\|\nu - \nu_0\|_1 + \|h - h_0\|_1) = \min(1, 1/L')\|f - f_0\|_1.\end{aligned}$$

The proof of (S9.52) is inspired by the proof of Lemma 4 in the supplementary material of Donnet, Rivoirard and Rousseau (2020). The following computations are valid in all our estimation scenarios. We recall that for any k, $r_k^f = \nu_k$ for the ReLU link (Case 1 of Proposition 3.5) and $r_k^f = \theta_k + \nu_k$ for the shifted ReLU link (Case 2 of Proposition 3.5).

Let $A > x > 0$ and $\eta > 0$ such that

$$0 < \frac{(A + x)^2 \eta K^2}{1 - \eta K} < \frac{1}{2} \qquad \text{and} \qquad \eta \leqslant \frac{\min_l r_l^0}{2C_0'}, \tag{S9.53}$$

with C_0' such that $\|r_f - r_0\|_1 + \|h - h_0\|_1 \leqslant C_0'$. Assume that for any $1 \leqslant l' \leqslant K$, $|r_{l'}^f - r_{l'}^0| \leqslant \eta(\|r_f - r_0\|_1 + \|h - h_0\|_1)$ and let $l \in [K]$ such that $\sum_k \|h_{kl} - h_{kl}^0\|_1 = \max_{l'} \sum_k \|h_{kl'} - h_{kl'}^0\|_1$.

Then we have

$$\|r_f - r_0\|_1 + \|h - h_0\|_1 \leqslant \left(\frac{\eta K^2}{1 - \eta K} + K\right) \sum_k \|h_{kl} - h_{kl}^0\|_1. \tag{S9.54}$$

For each $k \in [K]$, we define the event Ω_k as

$$\Omega_k = \left\{\max_{k' \neq k} N^{k'}[\tau_1, \tau_2) = 0,\ N^k[\tau_1, \tau_1 + x] = 0,\ N^k[\tau_1 + x, \tau_1 + x + A] = 1,\ N^k[\tau_1 + x + A, \tau_2) = 0\right\}.$$

On Ω_k, we have $\xi_1 = U_1^{(1)} + A$ and thus,

$$\mathbb{E}_f[Z_{1l}] \geqslant \sum_k \mathbb{E}_f\left[\mathbb{1}_{\Omega_k} \int_{\tau_1}^{A + U_1^{(1)}} |\lambda_t^l(f) - \lambda_t^l(f_0)| dt\right].$$

Let $\mathbb{Q}$ be the point process measure of a homogeneous Poisson process with unit intensity on $\mathbb{R}^+$ and equal to the null measure on $[-A, 0)$. Then

$$\mathbb{E}_f [Z_{1l}] \geqslant \sum_k \mathbb{E}_{\mathbb{Q}} \left[\int_{\tau_1}^{U_1^{(1)}+A} \mathcal{L}_t(f) \mathbb{1}_{\Omega_k} |\lambda_t^l(f) - \lambda_t^l(f_0)| \right] dt,$$

with $\mathcal{L}_t(f)$ the likelihood process given by

$$\mathcal{L}_t(f) = \exp\left(Kt - \sum_k \int_{\tau_1}^t \lambda_u^k(f) du + \sum_k \int_{\tau_1}^t \log(\lambda_u^k(f)) dN_u^k \right).$$

For $t \in [\tau_1, U_1^{(1)} + A)$, since on Ω_k, $\tau_1 + x \leqslant U_1^{(1)} \leqslant \tau_1 + A + x$, we have

$$\mathcal{L}_t(f) \geqslant e^{Kt} \lambda^k_{U_1^{(1)}}(f) \exp\left\{ - \sum_{k'} \int_{\tau_1}^t \phi_{k'}(\tilde{\lambda}_u^{k'}(f)) du \right\}.$$

Under condition **(C2)**, since $\phi_{k'} \leqslant \Lambda_{k'}, \forall k'$, with $\Lambda = (\Lambda_1, \dots, \Lambda_K)$, we directly have that

$$\mathcal{L}_t(f) \geqslant e^{Kt} \lambda^k_{U_1^{(1)}}(f) e^{-(A+x)\|\Lambda\|_1} \geqslant r_k^f e^{-\|\Lambda\|_1},$$

since at $\lambda^k_{U_1^{(1)}} = r_k^f = \phi_k(\nu_k)$.

Under condition **(C1bis)**, using that ϕ_k is L-Lipschitz, we have

$$\begin{aligned}
\mathcal{L}_t(f) &\geqslant e^{-\sum_{k'} \phi_{k'}(0)(A+U_1^{(1)}-\tau_1)} \lambda^k_{U_1^{(1)}}(f) \exp\left\{ - \sum_{k'} \int_{\tau_1}^{A+U_1^{(1)}} (\phi_{k'}(\tilde{\lambda}_u^{k'}(f)) - \phi_{k'}(0)) du \right\} \\
&\geqslant e^{-\sum_{k'} \phi_{k'}(0)(A+U_1^{(1)}-\tau_1)} \lambda^k_{U_1^{(1)}}(f) \exp\left\{ -L \sum_{k'} \left((A + U_1^{(1)} - \tau_1)\nu_{k'} + \int_{U_1^{(1)}}^{A+U_1^{(1)}} h_{kk'}(u - U_1^{(1)}) du \right) \right\} \\
&\geqslant e^{-\sum_{k'} \phi_{k'}(0)(2A+x)} \lambda^k_{U_1^{(1)}}(f) \exp\left\{ -L \sum_{k'} \left((2A + x)\nu_{k'} + \int_{U_1^{(1)}}^{A+U_1^{(1)}} h^+_{kk'}(u - U_1^{(1)}) du \right) \right\} \\
&\geqslant e^{-\sum_{k'} \phi_{k'}(0)(2A+x)} r_k^f \exp\left\{ -L \sum_{k'} \left((2A + x)\nu_{k'} + \|h^+_{kk'}\|_1 \right) \right\}.
\end{aligned}$$

Moreover, since $\|S^+\|_1 < 1$, then $\forall (k, k') \in [K]^2$, $\|h^+_{kk'}\|_1 < 1$. Thus, we obtain

$$\begin{aligned}
\mathcal{L}_t(f) &\geqslant e^{-\sum_{k'} \phi_{k'}(0)(2A+x)} r_k^f e^{-LK - L(2A+x)\sum_{k'} \nu_{k'}} \\
&\geqslant \frac{e^{-\sum_{k'} \phi_{k'}(0)(2A+x)} r_k^0}{2} e^{-LK - 6AL \max(\tilde{C}, \|\nu_0\|_1)} =: C.
\end{aligned}$$

In the last inequality, we have used our assumption $\|\nu - \nu_0\|_1 \leqslant \max(\|\nu_0\|_1, \tilde{C})$ which implies that

$$\sum_{k'} \nu_{k'} \leqslant 2 \max(\|\nu_0\|_1, \tilde{C}).$$

Moreover, we have that

$$\mathbb{E}_f\left[Z_{1l}\right] \geqslant C \sum_k \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} \left|\phi_l(\tilde{\lambda}_t^l(f)) - \phi_l(\tilde{\lambda}_t^l(f_0))\right| dt\right]$$
$$\geqslant \frac{C}{L'} \sum_k \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} \left|(\nu_l - \nu_l^0) + (h_{kl} - h_{kl}^0)(t - U_1^{(1)})\right| dt\right],$$

in all models except Case 2. In fact, in the latter case, we obtain

$$\mathbb{E}_f\left[Z_{1l}\right] \geqslant C \sum_k \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} \left|(\theta_l + \nu_l - \theta_l^0 - \nu_l^0) + (h_{kl} - h_{kl}^0)(t - U_1^{(1)})\right| dt\right]$$
$$= C \sum_k \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} \left|(r_l^f - r_l^0) + (h_{kl} - h_{kl}^0)(t - U_1^{(1)})\right| dt\right].$$

On the one hand,

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} |\nu_l - \nu_l^0| dt\right] = A|\nu_l - \nu_l^0| \mathbb{Q}(\Omega_k) \leqslant AL' |\phi_l(\nu_l) - \phi_l(\nu_l^0)| \mathbb{Q}(\Omega_k) = AL' |r_l^f - r_l^0| \mathbb{Q}(\Omega_k)$$
$$\leqslant AL' \frac{\eta K^2}{1 - \eta K} \sum_{k'} \|h_{k'l} - h_{k'l}^0\|_1,$$

and in Case 2 we have

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} |r_l^f - r_l^0| dt\right] = A|r_l - r_l^0| \mathbb{Q}(\Omega_k) \leqslant A \frac{\eta K^2}{1 - \eta K} \sum_{k'} \|h_{k'l} - h_{k'l}^0\|_1.$$

On the other hand, by definition of $\mathbb{Q}$, $N^k[\tau_1, \tau_1 + x + A] \sim \text{Poisson}(x + A)$. Consequently, with U a random variable with uniform distribution on $[\tau_1 + x, \tau_1 + x + A]$, we obtain

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\Omega_k} \int_{U_1^{(1)}}^{U_1^{(1)}+A} \left|(h_{kl} - h_{kl}^0)(t - U_1^{(1)})\right| dt\right] = \mathbb{Q}(\Omega_k) \mathbb{E}\left[\int_U^{U+A} |(h_{kl} - h_{kl}^0)(t - U)| dt\right]$$
$$= \frac{\mathbb{Q}(\Omega_k)}{A} \int_{\tau_1+x}^{\tau_1+A+x} \left[\int_s^{A+s} |h_{kl} - h_{kl}^0|(t - s) dt\right] ds \geqslant \mathbb{Q}(\Omega_k) \|h_{kl} - h_{kl}^0\|_1.$$

Moreover, we have

$$\mathbb{Q}(\Omega_k) \geqslant \mathbb{Q}(\max_{k' \neq k} N^{k'}[\tau_1, \tau_1 + x + 2A] = 0, N^k[\tau_1, \tau_1 + x] = 0, N^k[\tau_1 + x, \tau_1 + x + A] = 1)$$
$$= \mathbb{Q}(\max_{k' \neq k} N^{k'}[\tau_1, \tau_1 + x + 2A] = 0) \mathbb{Q}(N^k[\tau_1, \tau_1 + x] = 0) \mathbb{Q}(N^k[\tau_1 + x, \tau_1 + x + A] = 1)$$
$$= e^{-(K-1)(x+2A)} \times e^{-x} \times A e^{-A} := C'.$$

Using (S9.53) together with (S9.54), we obtain

$$\mathbb{E}_f[Z_{1l}] \geqslant \frac{C}{L'} \sum_k \frac{\mathbb{Q}(\Omega_k)}{A} \left(\|h_{kl} - h_{kl}^0\|_1 - A^2 L' \frac{\eta K^2}{1-\eta K} \|h_{kl} - h_{kl}^0\|_1 \right) \geqslant \frac{C}{L'} \frac{C'}{2} \sum_k \|h_{kl} - h_{kl}^0\|_1$$
$$\geqslant C(f_0)(\|r - r_0\|_1 + \|h - h_0\|_1), \quad C(f_0) = \frac{C}{L'} \frac{C'}{2(K + \eta K^2/(1-\eta K))}.$$

If there exists $l \in [K]$ such that $|r_l^f - r_l^0| \geqslant \eta(\|r - f - r_0\|_1 + \|h - h_0\|_1)$, we can use similar arguments as in the proof of Lemma 4 of Donnet, Rivoirard and Rousseau (2020):

$$\mathbb{E}_f[Z_{1l}] \geqslant \mathbb{P}_f\left[\max_k N^k[\tau_1, \tau_1 + A] = 0\right] \times A|r_l^f - r_l^0|,$$

and

$$\mathbb{P}_f\left[\max_k N^k[\tau_1, \tau_1 + A] = 0\right] = \mathbb{E}_{\mathbb{Q}}\left[\int_{\tau_1}^{\tau_1 + A} \mathcal{L}_t(f) \mathbb{1}_{\max_k N^k[\tau_1, \tau_1 + A] = 0} dt\right] = \mathbb{E}_{\mathbb{Q}}\left[\int_{\tau_1}^{\tau_1 + A} e^{A\|r\|_1} \mathbb{1}_{\max_k N^k[\tau_1, \tau_1 + A] = 0} dt\right]$$
$$\geqslant A e^{A\|r_f\|_1} e^{-KA},$$

so that

$$\mathbb{E}_f[Z_{1l}] \geqslant C'(f_0)(\|r_f - r_0\|_1 + \|h - h_0\|_1), \quad C'(f_0) = A^2 \eta e^{A\|r_0\|_1/2} e^{-KA}.$$

We can conclude that in all cases,

$$\mathbb{E}_f[Z_{1l}] \geqslant \min(C(f_0), C'(f_0))(\|r_f - r_0\|_1 + \|h - h_0\|_1),$$

and except in Case 2 of Proposition 3.5,

$$\mathbb{E}_f[Z_{1l}] \geqslant \min(C(f_0), C'(f_0), \frac{1}{L'}, 1) \|f - f_0\|_1.$$

□

S10. Additional results

In this section we recall some useful results on the regenerative properties of the nonlinear Hawkes model, which are mainly straightforward extensions of Costa et al. (2020) to our multivariate and general nonlinear setup. Besides, we recall the well-known Fuk-Nagaev's inequality.

The first lemma is an extension of Theorem A.1 Costa et al. (2020) for a $M^K/G^K/\infty$ queue when the arrival process is the superposition of K Poisson Point processes, corresponding to K types of customers.

Lemma S10.1. *Consider a $M^K/G^K/\infty$ queue with K types of customers that arrive according to a Poisson process with rate $r = (r_1, \dots, r_K)$. Assume that for each $k \in [K]$, the generic service time H^k for a customer of type k satisfies for some $\gamma > 0$ and for any $t \geqslant 0$:*

$$\mathbb{P}\left[H^k \geqslant t\right] = o(e^{-\gamma t}).$$

Let $\mathcal{T}_1$ the first time of return of the queue to zero.

1. *If $\|r\|_1 < \gamma$, then*

$$\mathbb{P}[\mathcal{T}_1 \geqslant t] \leqslant \left[1 + \frac{\mathbb{E}\left[e^{\gamma B}\right]}{\gamma - \|r\|_1}\right] e^{-\|r\|_1 t},$$

where B is the length of a busy period of the queue, i.e. $B = \mathcal{T}_1 - V_1$ with V_1 the arrival time of the first customer.

2. *If $\gamma \leqslant \|r\|_1$, then for any $0 < \alpha < \gamma$, $\mathbb{P}[\mathcal{T}_1 \geqslant t] \leqslant c_1(\alpha) e^{-\alpha t}$, with*

$$c_1(\alpha) = \left[1 + \frac{\mathbb{E}\left[e^{\alpha B}\right]}{\|r\|_1 - \alpha}\right].$$

3. $\forall \alpha < \|r\|_1 \wedge \gamma$, $\mathbb{E}\left[e^{\alpha \mathcal{T}_1}\right] \leqslant \frac{\|r\|_1}{\|r\|_1 - \alpha} \mathbb{E}\left[e^{\alpha B}\right] < +\infty$.

Proof. In this situation, the arrival process of customers, *regardless of their type*, is a superposition of K Poisson processes with individual rate r_k, $k \in [K]$. Consequently, it is equivalent to a Poisson process with rate $\|r\|_1 = \sum_k r_k$. Moreover, the generic service time H of a customer can be written as $H = \sum_k \delta_k H^k$, with $\delta = (\delta_k)_{k \in [K]}$ a one-hot vector indicating the type of customer. We can easily see that

$$\delta \sim \mathrm{Mult}\left(1, \frac{r_1}{\|r\|_1}, \dots, \frac{r_K}{\|r\|_1}\right), \quad H|\delta \sim \delta \mathcal{P},$$

with $\mathcal{P}$ the vector of service time distributions of the K types of customers. We note that the service time H is independent of the arrival process. Consequently, for $t \geqslant 0$,

$$\mathbb{P}[H \geqslant t] = \sum_k \mathbb{P}\left[H^k \geqslant t,\, \delta_k = 1\right] \leqslant \sum_k \mathbb{P}\left[H^k \geqslant t\right] = o(e^{-\gamma t}).$$

We can therefore conclude that this queue is equivalent to a $M/G/\infty$ queue with rate $\|r\|_1$ and generic service time satisfying $\mathbb{P}[H \geqslant t] = o(e^{-\gamma t})$. We can then apply Theorem A.1 in Costa et al. (2020) to obtain the results. □

The next lemma is a direct multivariate extension of the results in Propositions 2.1 and 3.1 and Lemma 3.2 of Costa et al. (2020). It introduces the mutually-exciting process dominating (in the sense of measure) a nonlinear Hawkes process.

Lemma S10.2. *Let Q be a K-dimensional Poisson point process on $(0, +\infty) \times (0, +\infty)^K$ with unit intensity. Let N be the Hawkes process with immigration rate $\nu = (\nu_1, \dots, \nu_K)$, $\nu_k > 0$, $k \in [K]$, interaction functions $h_{lk} : \mathbb{R}_+ \to \mathbb{R}$, $(l,k) \in [K]^2$ and initial measure N_0 on $[-A, 0]$ driven by $(Q_t)_{t \geqslant 0}$ and satisfying one condition of Lemma 2.1. N is the pathwise unique strong solution of the following system of stochastic equations*

$$\begin{cases} N^k = N_0^k + \int_{(0,+\infty)\times(0,+\infty)} \delta(u) \mathbb{1}_{\theta \leqslant \lambda^k(u)} Q^k(du, d\theta), \\ \lambda^k(u) = \phi_k\left(\nu_k + \sum_{l=1}^K \int_{u-A}^u h_{lk}(u-s) dN_s^l\right),\ u > 0, \quad k \in [K] \end{cases}.$$

with $\delta(.)$ the Dirac delta function. Consider the similar equation for a point process N^+ in which h_{lk} is replaced by h_{lk}^+ for any $l, k \in [K]^2$. Then

1. *there exists a pathwise unique strong solution N;*

2. *the same holds for N^+ and $N \leqslant N^+$ a.s. in the sense of measures.*

This also implies that, with $\Delta\tau_1^+$ defined similarly to $\Delta\tau_1$ in (20) for the process N^+,

$$\mathbb{P}\left[\Delta\tau_1 \leqslant \Delta\tau_1^+\right] = 1.$$

Moreover, with $\mathcal{T}_1$ defined as in Lemma S10.1, we also have $\mathbb{P}\left[\Delta\tau_1^+ = \mathcal{T}_1\right] = 1$.

Finally, the last proposition is the Fuk-Nagaev's inequality.

Proposition S10.3. *Let $(X_i)_{i\geqslant 1}$ a sequence of independent and centered random variables with finite variance and $S_n = \sum_{i=1}^n X_i$. With $v = \sum_{i=1}^n \mathbb{V}(X_i)$, for any $x \geqslant 0$ and $\lambda \geqslant 0$, it holds that*

$$\mathbb{P}\left[S_n \geqslant \lambda\right] \leqslant \sum_{i=1}^n \mathbb{P}\left[X_i > x\right] + \exp\left\{-\frac{v}{x^2} h\left(\frac{x\lambda}{v}\right)\right\},$$

where $h(u) = (1+u)\log(1+u) - u,\ u \geqslant 0$.

References

Costa, M., Graham, C., Marsalle, L. and Tran, V. C. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability* **52** 879–915.

Donnet, S., Rivoirard, V. and Rousseau, J. (2020). Nonparametric Bayesian estimation for multivariate Hawkes processes. *The Annals of Statistics* **48** 2698 – 2727.

Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *The Annals of Statistics* **28** 500 – 531.

Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for non iid observations. *The Annals of Statistics* **35** 192-223.

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

Reynaud-Bouret, P. and Roy, E. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bulletin of the Belgian Mathematical Society-Simon Stevin* **13** 883–896.

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

Sulem, D., Rivoirard, V. and Rousseau, J. (2022). Bayesian estimation of nonlinear Hawkes processes.