Supplementary Material to "Estimation for the Prediction of Point Processes with Many Covariates" by Alessio Sancetta

A.1 Proofs of Results

The notation is collected in the next subsection so that the reader can refer to it when needed.

A.1.1 Preliminary Lemmas and Notation

Write $\mathcal{L}_0 := \mathcal{L}(B_0, \Theta, \mathcal{W})$, $\bar{\mathcal{L}} := \mathcal{L}\left(\bar{B}, \Theta, \mathcal{W}\right)$ and $\mathcal{L} := \mathcal{L}(B, \Theta, \mathcal{W})$ for arbitrary, but fixed B. By Condition 2, the envelope function of $\bar{\mathcal{L}}$, is

$$\sup_{g \in \bar{\mathcal{L}}} \sup_{z \in \mathbb{R}} |g(z)| \leq \bar{B}\bar{\theta}/\underline{w} =: \bar{g}. \tag{A.1}$$

From the main text, recall that $\bar{B}_w := \bar{B}/\underline{w}$. Throughout, to keep notation simpler, suppose that $K > 1$.

To ease notation, write $\Lambda(t)$ for $\int_0^t d\Lambda(s) = \int_0^t \lambda(X(s))\, ds$, $\int_0^t e^g d\mu$ for $\int_0^t e^{g(X(s))} ds$ and similarly for $\int_0^t g dN$, $\int_0^t g d\Lambda$ $\int_0^t g d\mu$, etc., where μ is the Lebesgue measure. Hence, arguments $X(t)$ and t are dropped, but this should cause no confusion: all integrals here are w.r.t. $dN(t)$, $d\mu(t)$ etc., and the argument of all the functions is $X(t)$. Also, $\lambda(X(s)) = e^{g_0(X(s))}$, where $\bar{g}_0 := |g_0|_\infty$. With no loss of generality, to keep notation simple also suppose that $|g_{B_0}|_\infty \leq \bar{g}_0$. If this were not the case, we can just redefine $\bar{g}_0$ to be an upper bound for the uniform norms of g_0 and g_{B_0} (recall the definition of B_0 in (6)). It then follows from (6) that $\sup_{B>0} |g_B|_\infty \leq \bar{g}_0$ because g_B is the best uniform approximation for g_0 in $\mathcal{L}(B)$, and for $B \geq B_0$, (6) implies $g_B = g_{B_0}$. These facts will be used freely in the proofs without further mention. Define the following random Hellinger metric $d_T(g, g_0) = \sqrt{\frac{1}{2}\int_0^T \left(e^{g/2} - e^{g_0/2}\right)^2 d\mu}$. Sometimes, it will be useful to consider the identity $d_T^2(g, 0) = \frac{1}{2}\int_0^T \left|e^{g/2} - 1\right|^2 d\mu$.

Lemma 3 *Suppose that f, f' are functions on $\mathbb{R}^K$. Then,*

$$\frac{1}{8}\int_0^T \left(f - f'\right)^2 e^{f'} d\mu \leq d_T^2\left(f, f'\right). \tag{A.2}$$

Proof. Multiplying and dividing by $e^{f'}$,

$$d_T^2\left(f, f'\right) = \frac{1}{2}\int_0^T e^{f'}\left(e^{(f-f')/2} - 1\right)^2 d\mu. \tag{A.3}$$

Expand the square in the above display

$$\left(e^{(f-f')/2}-1\right)^2 = e^{(f-f')} - 2e^{(f-f')/2} + 1.$$

By Taylor expansion of the two exponentials, the above is equal to

$$\sum_{j=0}^{\infty}\frac{(f-f')^j}{j!} - 2\sum_{j=0}^{\infty}\frac{(f-f')^j}{j!}\left(\frac{1}{2}\right)^j + 1 = \sum_{j=2}^{\infty}\frac{(f-f')^j}{j!}\left(1-\frac{1}{2^{j-1}}\right) \geq \frac{(f-f')^2}{4}.$$

Inserting in (A.3) deduce (A.2). ∎

Lemma 4 *Suppose that* $|g_{B_0}|_\infty \leq \bar{g}_0$*. Then,*

$$0 \leq \int_0^T \left[(g_0 - g_{B_0})\, d\Lambda - \left(e^{g_0} - e^{g_{B_0}}\right) d\mu\right] \leq \frac{1}{2}e^{2\bar{g}_0}\int_0^T (g_0 - g)^2\, d\Lambda.$$

Proof. By definition of $d\Lambda = e^{g_0}d\mu$,

$$\begin{aligned}\int_0^T \left[(g_0 - g)\, d\Lambda - \left(e^{g_0} - e^{g}\right) d\mu\right] &= \int_0^T \left[(g_0 - g)\, e^{g_0} - \left(e^{g_0} - e^{g}\right)\right] d\mu \\ &= \int_0^T \left[(g_0 - g) + e^{-(g_0-g)} - 1\right] e^{g_0} d\mu. \qquad \text{(A.4)}\end{aligned}$$

For any fixed real x, by Taylor series with remainder, for some x_* in the convex hull of $\{0, x\}$,

$$e^{-x} - 1 + x = \frac{x^2}{2}e^{-x_*}. \qquad \text{(A.5)}$$

Apply this equality to $x = g_0 - g$ and insert it in the square brackets on the r.h.s. of (A.4) to deduce the upper bound in the lemma because $|g_0 - g_{B_0}|_\infty \leq 2\bar{g}_0$. For any $x > 0$, the following inequality holds:

$$0 \leq (x - \ln x - 1) \qquad \text{(A.6)}$$

with equality only if $x = 1$. Apply this inequality to $x = \exp\{-(g_0 - g_{B_0})\}$ and insert it in the square brackets on the r.h.s. of (A.4) to deduce the lower bound in the lemma. ∎

A.1.2 Solution of the Population Likelihood

For simplicity, as in Condition 1 suppose that $T_0 = 0$. Then, by Lemma 2 in Ogata (1978),

$$L(g) = \lim_T \frac{L_T(g)}{T} = \lim_T \frac{1}{T}\int_0^T \left(g dN - e^g d\mu\right) = P\left(ge^{g_0} - e^g\right)$$

almost surely, where L_T is the log-likelihood at time T (e.g., Ogata, 1978, eq.1.3). Taking first derivatives, the first order condition is $P\left(he^{g_0} - he^g\right) = 0$ for any $h \in \bar{\mathcal{L}}$. Hence, if $g = g_0$, the condition is satisfied. To check uniqueness, verify that the second order condition for concavity, i.e., $-Ph^2e^g < 0$, holds for any $h \neq 0$. Using the lower bound $e^{-\bar{g}} \leq e^g$, deduce that $-Ph^2e^g \leq -e^{-\bar{g}}Ph^2 < 0$ holds for any $h \neq 0$ P-almost everywhere. Given that $-L(g)$ is convex and $\bar{\mathcal{L}}$ is convex and closed, the maximizer of $L(g)$ is unique.

A.1.3 Proof of Theorem 1

The result is derived for the Hellinger distance d_T rather than the norm $|\cdot|_{\lambda,T}$.

Define $C_T^2 := C^2 \times T \max\left\{r_T^{-2}, 2e^{3\bar{g}_0}|g_0 - g_{\bar{B}}|_\infty^2\right\}$ and the martingale $M = N - \Lambda$ (Λ in (1) is the compensator of N). Here, r_T is a nondecreasing sequence which will be defined in due course. With the present notation, the last display in the proof of lemma 4.1 in van de Geer (1995) states that

$$\frac{1}{2}\int_0^T (g - g_0)\, dM \geq d_T^2(g, g_0) + \frac{1}{2}L_T(g, g_0), \tag{A.7}$$

where $L_T(g, g_0) := L_T(g) - L_T(g_0)$ for any g, so also for $g = g_T$. (The above display is only valid when g_0 is the true function, but it is not required that $g_0 \in \mathcal{L}(B)$ for some B.) By Condition 3, and the inequality $L_T(g_T, g_{\bar{B}}) \geq L_T(g_T) - \sup_{g\in\bar{\mathcal{L}}} L_T(g)$, deduce that

$$L_T(g_T, g_0) = L_T(g_T, g_{\bar{B}}) + L_T(g_{\bar{B}}, g_0) \geq -\left(C_T^2/2\right) + L_T(g_{\bar{B}}, g_0) \tag{A.8}$$

choosing C large enough, in the definition of C_T. Hence, inserting (A.8) in (A.7), deduce that

$$\begin{aligned} & \Pr\left(d_T(g_T, g_0) > C_T\right) \\ \leq\ & \Pr\Bigg(\frac{1}{2}\left[\int_0^T (g - g_0)\, dM - L_T(g_{\bar{B}}, g_0)\right] \geq d_T^2(g, g_0) - \frac{C_T^2}{4} \\ & \text{and } d_T^2(g, g_0) > C_T^2 \text{ for some } g \in \bar{\mathcal{L}}\Bigg) \end{aligned} \tag{A.9}$$

To bound the term in the square bracket, add and subtract $\int_0^T g_{\bar{B}} dM$ and note that $L_T(g_{\bar{B}}, g_0)$ can be written as $\int_0^T [(g_{\bar{B}} - g_0)\, dM + (g_{\bar{B}} - g_0)\, d\Lambda - (e^{g_{\bar{B}}} - e^{g_0})\, d\mu]$. This implies that

$$\begin{aligned} \int_0^T (g - g_0)\, dM - L_T(g_{\bar{B}}, g_0) &= \int_0^T [(g - g_{\bar{B}}) + (g_{\bar{B}} - g_0)]\, dM \\ &\quad - \int_0^T [(g_{\bar{B}} - g_0)\, dM + (g_{\bar{B}} - g_0)\, d\Lambda - (e^{g_{\bar{B}}} - e^{g_0})\, d\mu] \\ &= \int_0^T (g - g_{\bar{B}})\, dM + \int_0^T [(g_0 - g_{\bar{B}})\, d\Lambda - (e^{g_0} - e^{g_{\bar{B}}})\, d\mu] \\ &\leq \int_0^T (g - g_{\bar{B}})\, dM + \frac{1}{2}e^{2\bar{g}_0}\int_0^T (g_0 - g_{\bar{B}})^2\, d\Lambda \end{aligned}$$

using Lemma 4 in the inequality. From the above calculations, and the fact that $\int_0^T (g_0 - g_{\bar{B}})^2 d\Lambda \leq Te^{\bar{g}_0} |g_0 - g_{\bar{B}}|_\infty^2$, deduce that (A.9) is less than

$$\begin{aligned} & \Pr\left(\frac{1}{2}\int_0^T (g - g_{\bar{B}})\, dM \geq d_T^2(g, g_0) - \frac{C_T^2}{4} - \frac{1}{2}Te^{3\bar{g}_0} |g_{\bar{B}} - g_0|_\infty^2 \right. \\ & \quad \left. \text{and } d_T^2(g, g_0) > C_T^2 \text{ for some } g \in \bar{\mathcal{L}}\right) \\ \leq \; & \Pr\left(\frac{1}{2}\int_0^T (g - g_{\bar{B}})\, dM \geq d_T^2(g, g_0) - \frac{C_T^2}{2} \text{ and } d_T^2(g, g_0) > C_T^2 \text{ for some } g \in \bar{\mathcal{L}}\right), \end{aligned}$$

using the definition of C_T. The above is bounded by $\Pr\left(\sup_{g \in \bar{\mathcal{L}}} \int_0^T (g - g_{\bar{B}})\, dM \geq C_T^2\right)$, which is further bounded by

$$\frac{1}{C_T^2}\mathbb{E}\left|\sup_{g \in \bar{\mathcal{L}}} \int_0^T (g - g_{\bar{B}})\, dM\right| \leq \frac{2}{C_T^2}\mathbb{E}\left|\sup_{g \in \bar{\mathcal{L}}} \int_0^T g dM\right|$$

using Markov inequality and then the triangle inequality because $g_{\bar{B}} \in \bar{\mathcal{L}}$. Write $g = \sum_\theta b_\theta \theta$. Note that

$$\sup_{g \in \bar{\mathcal{L}}} \left|\int_0^T g dM\right| = \sup_{b_\theta, \theta \in \Theta} \left|\int_0^T \left(\sum_\theta b_\theta \theta\right) dM\right| \leq \bar{B}_w \sup_{\theta \in \Theta} \left|\int_0^T \theta dM\right|$$

where the supremum runs over all the b_θ's such that $\sum_\theta |b_\theta| \leq \bar{B}_w$. According to these calculations, to bound (A.9) it is sufficient to bound

$$\frac{2\bar{B}_w}{C_T^2}\mathbb{E}\sup_{\theta \in \Theta}\left|\int_0^T \theta dM\right|. \tag{A.10}$$

Let $\{\Pi_l(\epsilon) : v = 1, 2, .., N_\Pi\}$ be a partition of Θ into $N_\Pi(\epsilon)$ elements such that $\sup_{\theta, \theta' \in \Pi_l(\epsilon)} |\theta - \theta'| \leq \epsilon$. By Condition 2, one can construct such partition with $N_\Pi(\epsilon) \lesssim N(\epsilon, \Theta)$ and such that

$$\sup_{\theta, \theta' \in \Pi_l(\epsilon)} |\theta - \theta'|_\infty \leq |\theta_{U,l} - \theta_{L,l}|_\infty \tag{A.11}$$

where $[\theta_{L,l}, \theta_{U,l}]$ is an ϵ-bracket for the functions in Π_l, under the uniform norm. It follows that $N_\Pi(2\bar{\theta}) = 1$ because the diameter of Θ under the uniform norm is bounded by $2\bar{\theta}$. To bound (A.10), use the following maximal inequality from Nishiyama (1998, Theorem 2.2.3), which is specialized to the present framework.

Lemma 5 *Under Conditions 1 and 2,*

$$\mathbb{E}\max_{t \in [0,T]}\max_{\theta \in \Theta}\left|\int_0^t \theta dM\right| \lesssim C_{1,T}\int_0^{2\bar{\theta}} \sqrt{\ln(1 + N_\Pi(\epsilon))} d\epsilon + \frac{C_{2,T}}{\bar{\theta}C_{1,T}} \tag{A.12}$$

for any $C_{2,T} \geq \int_0^T \bar{\theta}^2 d\Lambda$, *and* $C_{1,T} \geq |\Theta|_{\Pi,T}$, *where*

$$|\Theta|_{\Pi,T} := \sup_{\epsilon\in(0,\bar{\theta}]} \max_{l\leq N_\Pi(\epsilon)} \frac{\sqrt{\int_0^T \left(\sup_{\theta,\theta'\in\Pi_l(\epsilon)} |\theta-\theta'|\right)^2 d\Lambda}}{\epsilon}.$$

From the discussion around (A.11) replace $N_\Pi(\epsilon)$ with $N(\epsilon,\Theta)$. The application of Lemma 5 essentially requires to find a bound for $C_{1,T}$ and $C_{2,T}$. Given that $\lambda = d\Lambda/d\mu$ is bounded by $e^{\bar{g}_0}$, from the discussion around (A.11), $|\Theta|_{\Pi,T} \leq \sqrt{e^{\bar{g}_0}T}$ and we set $C_{1T} = C_1\sqrt{e^{\bar{g}_0}T}$ for some C_1 to be chosen later. Also, deduce that we can choose $C_{2,T} = \bar{\theta}e^{\bar{g}_0}T$. This implies that $C_{2,T}/\bar{\theta}C_{1,T} = \sqrt{e^{\bar{g}_0}T/C_1}$. Hence, the first term on the r.h.s. of (A.12) is of no smaller order of magnitude than the second (i.e., not smaller than a constant multiple of $T^{1/2}$). Thus, in what follows, we can incorporate $C_{2,T}/\bar{\theta}C_{1,T}$ into it without further mention. Hence, an application of Lemma 5 bounds (A.10) by

$$\frac{2\bar{B}_w}{C_T^2}\mathbb{E}\sup_{\theta\in\Theta}\left|\int_0^T \theta dM\right| \lesssim \frac{2\bar{B}_w\sqrt{e^{\bar{g}_0}T}}{C_T^2}\int_0^{2\bar{\theta}}\sqrt{\ln\left(1+N(\epsilon,\Theta)\right)}d\epsilon. \tag{A.13}$$

Using the definition of C_T, and choosing $r_T^2 \lesssim \left[e^{3\bar{g}_0}|g_0-g|_\infty^2\right]^{-1}$, the above is a constant multiple of

$$r_T^2\frac{\bar{B}_w e^{\bar{g}_0/2}}{T^{1/2}}\int_0^{2\bar{\theta}}\sqrt{\ln\left(1+N(\epsilon,\Theta)\right)}d\epsilon$$

which is required to be $O(1)$, as it is an upper bound for (A.9) . This implies

$$r_T^2 \lesssim \frac{T^{1/2}}{\bar{B}_w e^{\bar{g}_0/2}\int_0^{2\bar{\theta}}\sqrt{\ln\left(1+N(\epsilon,\Theta)\right)}d\epsilon}.$$

But, r_T is also required not to go to zero, and in fact it is supposed to diverge to infinity unless the approximation error is nonvanishing. Therefore, the r.h.s. of the above display needs to be bounded away from zero.

To bound the entropy integral, recall that $\Theta = \bigcup_{k=1}^K \Theta_k$. The bracketing number of a union of sets is bounded above by the sum of the bracketing numbers of the individual sets. Hence, $N(\epsilon,\Theta) \leq \sum_{k=1}^K N(\epsilon,\Theta_k)$. Using the inequality $\ln(1+xy) \leq \ln x + \ln(1+y)$ for real $x, y \geq 1$, this implies that

$$\begin{aligned}\int_0^{2\bar{\theta}}\sqrt{\ln\left(1+N(\epsilon,\Theta_k)\right)}d\epsilon &\leq \int_0^{2\bar{\theta}}\max_{k\leq K}\sqrt{\ln K+\ln\left(1+N(\epsilon,\Theta_k)\right)}d\epsilon \\ &\leq 2\bar{\theta}\sqrt{\ln K}+\max_{k\leq K}\int_0^{2\bar{\theta}}\sqrt{\ln\left(1+N(\epsilon,\Theta_k)\right)}d\epsilon.\end{aligned}$$

Also, given that $\bar{\theta}$ is bounded and the entropy above is decreasing in ϵ, the above display can be bounded by a multiple of

$$\sqrt{\ln K} + \max_{k \le K} \int_0^1 \sqrt{\ln \left(1 + N\left(\epsilon, \Theta_k\right)\right)} d\epsilon. \tag{A.14}$$

Also, we can discard the terms that are bounded, i.e., $\bar{g}_0$ and $\bar{\theta}$, but kept so far just to highlight what their contribution might be. Similarly, $\bar{B}_w$ can be replaced by $\bar{B}$ because it enters the bound as a multiplicative constant. These calculations imply that there is a sequence r_T as in the statement in the theorem such that for C large enough,

$$\Pr\left(\frac{r_T^2}{T} d_T^2\left(g_T, g_0\right) > C\right) \le \frac{1}{C^2}.$$

By the relation between $d_T^2\left(g_T, g_0\right)/T$ and $\left|g_T - g_0\right|_{\lambda,T}^2$ (see (A.2)), the theorem follows.

A.1.4 Proof of Theorem 2

To ease notation, $T = T_n$. We adapt the calculations in the proof of Theorem 2 in Tsybakov (2003). This requires an upper bound for the Kullback-Leibler distance between two intensity densities, and the construction of a suitable subset of $\mathcal{L}(1)$ (using the notation of our theorem). The result in Tsybakov (2003) will then provide the necessary lower bound as stated in our Theorem 2.

To this end, let $N^{(1)}$ and $N^{(2)}$ be point processes with intensities e^{g_1} and e^{g_2} such that $\left|g_k\right|_\infty \le \bar{g}$, $k = 1, 2$. Let the sigma algebra generated by the process $X = \left(X\left(t\right)\right)_{t \ge 0}$ be denoted by $\mathcal{F}^X$. The Kullback-Leibler distance between two intensity densities e^{g_1} and e^{g_2}, restricted to $[0, T]$, and conditioning on $\mathcal{F}^X$ is

$$K\left(g_1, g_2 | \mathcal{F}^X\right) = \mathbb{E}_X \int_0^T \left(g_1 - g_2\right) dN^{(1)} - \int_0^T \left(e^{g_1} - e^{g_2}\right) d\mu$$

where $\mathbb{E}_X$ is the expectation conditional on $\mathcal{F}^X$. The above follows noting that conditioning on $\mathcal{F}^X$, durations are exponentially distributed with intensity density $\exp\left\{g_1\left(X\left(t\right)\right)\right\}$. Then,

$$K\left(g_1, g_1 | \mathcal{F}^X\right) = \int_0^T \left(g_1 - g_2\right) e^{g_1} d\mu - \int_0^T \left(e^{g_1} - e^{g_2}\right) d\mu \le \frac{e^{3\bar{g}}}{2} \int_0^T \left|g_1 - g_2\right|^2 d\mu$$

using (A.5) and the fact that $\left|g_k\right|_\infty \le \bar{g}$, $k = 1, 2$. This provides the necessary upper bound for the Kullback-Leibler distance, to be used in the proof of Theorem 2 in Tsybakov (2003).

Now, follow Bunea et al. (2007, p. 1693) with minor adjustments. For each k, we shall construct a function, say f_k, in Θ_k. Let $A_j = \sum_{i=1}^{j} 1\left\{T_i - T_{i-1} \ge a\right\}$, i.e. the number of durations greater than a amongst the first j durations. Throughout, $1\{\cdot\}$ is the indicator function. Clearly, $A_n \leq n$ with equality only if $a = 0$. Define

$$f_k(x) = \gamma \sum_{j=1}^{n} \phi_k\left(\frac{A_j}{A_n}\right) \frac{1\{x_k = X_k(T_{j-1})\} 1\{T_j - T_{j-1} \geq a\}}{\sqrt{T_j - T_{j-1}}}$$

where $\gamma > 0$ is a constant to be chosen in due course, and $\{\phi_k(s) : k = 1, 2, ..., K\}$ are bounded functions w.r.t. $s \in [0, 1]$, and such that $\frac{1}{A_n} \sum_{j=1}^{A_n} \phi_k\left(\frac{j}{A_n}\right) \phi_l\left(\frac{j}{A_n}\right) = \delta_{kl}$, where $\delta_{kl} = 1$ if $k = l$, zero otherwise (e.g., mutatis mutandis, as in Bunea et al., 2007, p. 1693). The functions f_k's are uniformly bounded in absolute value by a constant multiple of $\gamma/\sqrt{a}$. Hence $f_k \in \Theta_k$, for each k, choosing γ small enough. It follows that

$$\begin{aligned}
\int_0^T f_k(X(t)) f_l(X(t)) dt &= \sum_{j=1}^{n} f_k(X(T_{j-1})) f_l(X(T_{j-1}))(T_j - T_{j-1}) \\
&= \gamma^2 \sum_{j=1}^{A_n} \phi_k\left(\frac{j}{A_n}\right) \phi_l\left(\frac{j}{A_n}\right) = \gamma^2 A_n \delta_{kl}.
\end{aligned}$$

The first step follows because $X(t)$ is predictable and only changes after a jump. The second step follows by the definition of the f_k's because by continuity of the distribution of $X(0)$ and stationarity, $\Pr(X(T_i) = X(T_j)) = 0$ for $i \neq j$. Also, note that unless $\{T_j - T_{j-1} \geq a\}$ is true, the j^{th} term in the definition of f_k will be zero.

Let $\mathcal{C}$ be the subset of $\mathcal{L}(1)$ which consists of arbitrary convex combinations of $m \leq K/6$ of the f_k's with weight $1/m$ so that the weights sum to one. In consequence, for any $g_1, g_2 \in \mathcal{C}$,

$$\int_0^T (g_1 - g_2)^2 d\mu \asymp A_n \gamma^2 / m.$$

Let $p_a := \Pr(T_j - T_{j-1} \geq a)$. We claim that $\Pr(A_n < np_a/2) \to 0$ exponentially fast. Hence, the r.h.s. of the above display is proportional to $n\gamma^2/m$ with probability going to one. This claim will be verified at the end of the proof.

Now, by suitable choice of small γ, it is possible to follow line by line the argument after eq. (10) in Tsybakov (2003, proof of Theorem 2). This would give us a result for $\int_0^T (g_T - g_0)^2 d\mu$ rather than $\int_0^T (g_T - g_0)^2 \lambda d\mu$ and in terms of n rather than $T = T_n$. To replace n with T_n as in the statement of the theorem, note that T_n/n converges almost surely to $(P\lambda)^{-1}$, which is bounded. Finally, $\int_0^T (g_T - g_0)^2 \lambda d\mu \gtrsim \int_0^T (g_T - g_0)^2 d\mu$ by the conditions of the theorem.

It remains to show that the claim on A_n holds true. For any positive decreasing function h on the reals, the sets $\{A_n < cn\}$ and $\{h(A_n) > h(cn)\}$ are the same; here $c \in (0, 1)$ is a constant to be chosen in due course. Hence, by Markov inequality $\Pr(A_n < cn) \leq$

$\mathbb{E}h\left(n^{-1/2}A_n\right)/h\left(cn^{1/2}\right)$, which implies the following lower bound

$$\Pr\left(A_n \geq cn\right) \geq 1 - \frac{\mathbb{E}h\left(A_n/\sqrt{n}\right)}{h\left(c\sqrt{n}\right)}.$$

It remains to show that the second term on the r.h.s. goes to zero. To this end, let $h\left(s\right) = e^{-ts}$, for some fixed $t > 0$. For p_a as previously defined in the proof, write

$$\frac{A_n}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1\left\{T_i - T_{i-1} \geq a\right\} - p_a\right) + \sqrt{n}p_a.$$

The first term on the r.h.s. is a root-n standardized sum of i.i.d. centered Bernoulli random variables. Hence, it has a moment generating function which is bounded (use the proof of the central limit theorem for Bernoulli random variables). By this remark,

$$\frac{\mathbb{E}h\left(A_n/\sqrt{n}\right)}{h\left(c\sqrt{n}\right)} = \frac{\mathbb{E}\exp\left\{-tA_n/\sqrt{n}\right\}}{\exp\left\{-tc\sqrt{n}\right\}} \lesssim e^{-t(p_a-c)\sqrt{n}}.$$

Choose $c = p_a/2$ to see that the r.h.s. goes to zero exponentially fast for any $t > 0$, as previously claimed.

A.1.5 Proof of Lemma 1 and Corollaries

Proof. [Lemma 1] The proof is a minor re-adaptation of Lemma 4 in Sancetta (2015). Note that if $B \geq B_0$, the lemma is clearly true because in this case, $\mathcal{L}_0 \subseteq \mathcal{L} := \mathcal{L}\left(B, \Theta, \mathcal{W}\right)$. Hence, assume $B < B_0$ and w.n.l.g. $B = \rho B_0$ for $\rho \in (0, 1)$. Write

$$g_0 = \sum_{\theta\in\Theta} b_\theta \theta = \sum_{\theta\in\Theta} \lambda_\theta \bar{b}\theta$$

where the λ_θ's are nonnegative and add to one, and $\bar{b} = \sum_{\theta\in\Theta}\left|b_\theta\right|$. Note that the constraint $\sum_{\theta\in\Theta} w_\theta\left|b_\theta\right| \leq B_0$ for functions in $\mathcal{L}_0$ implies $\bar{b} \leq B_0/\underline{w}$. Define $g'\left(x\right) = \rho g_0\left(x\right)$ for ρ such that $B = \rho B_0$ so that $g' \in \mathcal{L}$. Using this choice of g', by standard inequalities,

$$\left|g_0 - g'\right|_r \leq \left|\sum_{\theta\in\Theta}\lambda_\theta\bar{b}\theta - \sum_{\theta\in\Theta}\lambda_\theta\rho\bar{b}\theta\right|_r \leq \left|\bar{b}\left(1-\rho\right)\right|\sum_{\theta\in\Theta}\lambda_\theta\left|\theta\right|_r \leq \bar{b}\left(1-\rho\right)\max_{\theta\in\Theta}\left|\theta\right|_r \leq \frac{\bar{\theta}_r}{\underline{w}}\left(B_0 - B\right)$$

using the definition of ρ. This proves the result, because for g' above, $\inf_{g\in\mathcal{L}}\left|g_0 - g\right|_r \leq \left|g_0 - g'\right|_r$. ■

Proof. [Corollary 2] We need to show that $L_T\left(\tilde{g}_T, g_B\right) \geq -\left(C_T^2/2\right)$ with C_T as in the proof of Theorem 1 and r_T as in (9), e.g., $C_T^2 \gtrsim \bar{B}\sqrt{T\ln K}$. To this end, recall that $\tilde{L}_T\left(g\right) = \int_0^T g\left(\tilde{X}\left(t\right)\right)dN\left(t\right) - \int_0^T \exp\left\{g\left(\tilde{X}\left(t\right)\right)\right\}dt$, which is the log-likelihood when we use $\tilde{X}$ instead of X. Note that the counting process N is still the same whether we use X or $\tilde{X}$, as jumps are observable. By definition, $\tilde{g}$ is the approximate maximizer of $\tilde{L}_T(g)$, but not necessarily the maximizer of $L_T(g)$. It would be enough to show that $L_T(\tilde{g}_T, g_B) \gtrsim -C_T^2$ in probability, as by a re-definition of the constant in C_T, the proof in Theorem 1 would go through. Given these remarks, write

$$L_T(\tilde{g}_T, g_B) \geq \tilde{L}_T(\tilde{g}_T, g_B) - \left| L_T(\tilde{g}_T, g_B) - \tilde{L}_T(\tilde{g}_T, g_B) \right|.$$

Using (11) we have that $\tilde{L}_T(\tilde{g}_T, g_B) \gtrsim -C_T^2$ as in (A.8). To bound the second term on the r.h.s. of the above display, it is sufficient to bound a constant multiple of

$$\begin{aligned}
& \sup_{g\in\bar{\mathcal{L}}} \left| L_T(g) - \tilde{L}_T(g) \right| \\
= & \sup_{g\in\bar{\mathcal{L}}} \left| \int_0^T \left[g(X(t)) - g\left(\tilde{X}(t)\right) \right] dN(t) - \int_0^T \left[\exp\{g(X(t))\} - \exp\left\{ g\left(\tilde{X}(t)\right) \right\} \right] dt \right| \\
\leq & \sup_{g\in\bar{\mathcal{L}}} \left| \int_0^T \left[g(X(t)) - g\left(\tilde{X}(t)\right) \right] dN(t) \right| + \sup_{g\in\bar{\mathcal{L}}} \left| \int_0^T \left[\exp\{g(X(t))\} - \exp\left\{ g\left(\tilde{X}(t)\right) \right\} \right] dt \right| \\
=: & \; I + II.
\end{aligned}$$

First, find a bound for II. By the mean value theorem in Banach spaces,

$$II \leq \sup_{g\in\bar{\mathcal{L}}} e^{\bar{g}} \int_0^T \left| g(X(t)) - g\left(\tilde{X}(t)\right) \right| dt. \tag{A.15}$$

Now,

$$\begin{aligned}
\sup_{g\in\bar{\mathcal{L}}} \int_0^T \left| g(X(t)) - g\left(\tilde{X}(t)\right) \right| dt &\leq \sup_{\left\{b_\theta : \sum_{\theta\in\Theta} |b_\theta| \leq \bar{B}_w\right\}} \int_0^T \sum_{\theta\in\Theta} |b_\theta| \left| \theta\left(\tilde{X}(t)\right) - \theta(X(t)) \right| dt \\
&\leq \bar{B}_w \max_{\theta\in\Theta} \int_0^T \left| \theta\left(\tilde{X}(t)\right) - \theta(X(t)) \right| dt
\end{aligned}$$

because the supremum over the simplex is achieved at one of its edges. By the conditions of the corollary, the above display is $O_p\left(\bar{B}e^{-\bar{g}}\sqrt{T \ln K}\right)$. Hence, deduce that (A.15) is $O_p\left(\bar{B}\sqrt{T \ln K}\right) = O_p(C_T)$ (recall the notation in (A.1)).

It remains to bound I. Adding and subtracting $\int_0^T \left[g(X(t)) - g\left(\tilde{X}(t)\right) \right] d\Lambda(t)$, and using the triangle inequality,

$$I \leq \sup_{g\in\bar{\mathcal{L}}} \left| \int_0^T \left[g(X(t)) - g\left(\tilde{X}(t)\right) \right] dM(t) \right| + \sup_{g\in\bar{\mathcal{L}}} \left| \int_0^T \left[g(X(t)) - g\left(\tilde{X}(t)\right) \right] d\Lambda(t) \right|.$$

The first term in the above display can be incorporated in the l.h.s. of (A.7) and bounded as in the proof of Theorem 1. To bound the second term on the above display by definition of $d\Lambda$,

$$\sup_{g\in\bar{\mathcal{L}}}\left|\int_0^T\left[g\left(X\left(t\right)\right)-g\left(\tilde{X}\left(t\right)\right)\right]\exp\left\{g_0\left(X\left(t\right)\right)\right\}dt\right|\leq\sup_{g\in\bar{\mathcal{L}}}e^{\bar{g}_0}\int_0^T\left|g\left(X\left(t\right)\right)-g\left(\tilde{X}\left(t\right)\right)\right|dt.$$

From the derived bound for II deduce that the r.h.s. is $O_p\left(C_T^2\right)$. This completes the proof of the first statement in the corollary, as all the conditions of Theorem 1 are satisfied. To show the last statement of the corollary, use the inequality $\left|g\left(X\left(t\right)\right)-g\left(\tilde{X}\left(t\right)\right)\right|^2\leq 2\bar{g}\left|g\left(X\left(t\right)\right)-g\left(\tilde{X}\left(t\right)\right)\right|$ together with a trivial modification of the previous display. ■

Proof. [Corollary 4] The approximation error is zero by assumption. Given that Θ_k has one single element, the entropy integral is trivially finite. Hence, (9) simplifies as in the statement of the corollary. ■

Proof. [Corollary 5] Define the set

$$\mathcal{B}:=\left\{\sup_{t>0}\left|\int_0^t\left(t-s\right)e^{-\underline{a}\left(t-s\right)}dN\left(s\right)\right|\leq\beta\right\}$$

for some $\beta<\infty$. In the proof of Theorem 1 write

$$\Pr\left(d_T\left(g_T,g_0\right)>C_T,\right)\leq\Pr\left(d_T\left(g_T,g_0\right)>C_T,\ \text{and }\mathcal{B}\right)+\Pr\left(\mathcal{B}^c\right)$$

where $\mathcal{B}^c$ is the complement of $\mathcal{B}$. We shall apply Corollary 2 to the first term on the r.h.s., and then show that the last term in the above display is negligible.

At first, show that the process with intensity density $\lambda\left(t\right)=\exp\left\{f_{a_0}\left(t\right)+g_0\left(X\left(t\right)\right)\right\}$ is stationary. To this end, we apply Theorem 2 in Brémaud and Massoulié (1996). Using their notation, their nonlinear function $\phi\left(\cdot\right)$ in their eq.(1) is here defined as $\exp\left\{f\left(\cdot\right)\right\}\exp\left\{g_0\left(X\left(t\right)\right)\right\}$, which is random, unlike their case. However, in the proof of their Theorem 2 they only use the fact that $\left|\phi\left(y\right)-\phi\left(y'\right)\right|\leq\alpha\left|y-y'\right|$ for some finite constant α (see their eq.(23) and first display on p.1580). This is the case here as well. To see this, recall the definition of f (see Section 3.6.2) which is bounded and Lipschitz. Then,

$$\left|\exp\left\{f\left(y\right)\right\}\exp\left\{g_0\left(X\left(t\right)\right)\right\}-\exp\left\{f\left(y'\right)\right\}\exp\left\{g_0\left(X\left(t\right)\right)\right\}\right|\leq\exp\left\{\bar{g}_0\right\}\left|f\left(y\right)-f\left(y'\right)\right|$$

(recall $\bar{g}_0$ is the uniform norm of g_0). We also need to note that $\exp\left\{g_0\left(X\left(t\right)\right)\right\}$ is stationary, bounded and predictable. This ensures that the intensity $\lambda\left(t\right)$ is bounded and predictable, which is required in the lemmas used in Brémaud and Massoulié (1996). Hence Condition 1 is satisfied.

To verify Condition 2, we verify that the entropy integral of the process $\tilde{f}_a$ is finite in a sense to be made clear below. We shall postpone this to the end of the proof.

Hence, mutatis mutandis, we now verify (10) in Corollary 2. To this end, we bound $c_T := \mathbb{E}\max_{a\in[\underline{a},\bar{a}]}\int_0^T\left|f_a(t)-\tilde{f}_a(t)\right|dt$. Corollary 2 requires c_T to be $O\left(e^{-\bar{B}_w\bar{\theta}}\sqrt{T\ln K}\right)$. By the Lipschitz condition and $a\in[\underline{a},\bar{a}]$,

$$\int_0^T\left|f_a(t)-\tilde{f}_a(t)\right|dt \lesssim \int_0^T e^{-\underline{a}t}\left(\int_{(-\infty,0)}e^{\underline{a}s}dN(s)\right)dt.$$

Using the fact that Λ is the compensator of N, and that Λ has bounded density $\exp\{f_{a_0}(t)+g_0(X(t))\}$, deduce that

$$\begin{aligned}\mathbb{E}\max_{a\in[\underline{a},\bar{a}]}\int_0^T\left|f_a(t)-\tilde{f}_a(t)\right|dt &\leq \mathbb{E}\left[\left(\int_{(-\infty,0)}e^{\underline{a}s}dN(s)\right)\left(\int_0^T e^{-\underline{a}t}dt\right)\right]\\ &\lesssim \frac{1}{\underline{a}}\mathbb{E}\int_{(-\infty,0)}e^{\underline{a}s}d\Lambda(s)\lesssim\frac{1}{\underline{a}^2}<\infty.\end{aligned}$$

This verifies (10) in Corollary 2.

To verify Condition 2 for $\tilde{f}_a$, we need an estimate of the entropy integral for the family of stochastic processes $\mathcal{A}:=\left\{\left(\tilde{f}_a(t)\right)_{t\geq0}:a\in[\underline{a},\bar{a}]\right\}$. This means that we need to bound

$$\begin{aligned}\sup_{t>0}\left|\tilde{f}_a(t)-\tilde{f}_{a'}(t)\right| &\lesssim \sup_{t>0}\left|\int_0^t\left(e^{-a(t-s)}-e^{-a'(t-s)}\right)dN(s)\right|\\ &\leq \sup_{t>0}\left|\int_0^t(t-s)\,e^{-\underline{a}(t-s)}dN(s)\right|dt\,|a-a'|\end{aligned}$$

using a first order Taylor expansion, and the lower bound on a,a'. On $\mathcal{B}$, the above is $\beta\,|a-a'|$. It is then easy to see that the entropy integral is a constant multiple of $\beta^{1/2}$ because the uniform ϵ-bracketing number of $[\underline{a}\beta,\bar{a}\beta]$ has size $\beta(\bar{a}-\underline{a})/\epsilon$.

In consequence, we can apply Corollary 2. Let $\beta=O(\ln T)$). There is no approximation error, so that r_T^{-2} (r_T as in (9)) becomes as in (14). The term $\sqrt{\ln T}$, in the numerator of (14), is proportional to the entropy integral of $\mathcal{A}$.

To conclude, we show that $\mathcal{B}^c$, the complement of $\mathcal{B}$, is such that $\Pr(\mathcal{B}^c)\to0$ as $\beta\to\infty$. By Markov inequality,

$$\Pr(\mathcal{B}^c)\leq\frac{\mathbb{E}\sup_{t>0}\left|\int_0^t(t-s)\,e^{-\underline{a}(t-s)}dN(s)\right|}{\beta}.$$

Recalling that $M=N-\Lambda$, by the triangle inequality, the numerator on the r.h.s. can be bounded by

$$\mathbb{E}\sup_{t>0}\left|\int_0^t(t-s)\,e^{-\underline{a}(t-s)}dM(s)\right|+\mathbb{E}\sup_{t>0}\left|\int_0^t(t-s)\,e^{-\underline{a}(t-s)}d\Lambda(s)\right|=:I+II.$$

The first integral inside the square is a bounded predictable function w.r.t. a martingale, and is a martingale. By the Burkholder-Davis-Gundy inequality,

$$I^2 \lesssim \sup_{t>0} \mathbb{E} \int_0^t \left| (t-s)\, e^{-\underline{a}(t-s)} \right|^2 d\Lambda(s) \leq e^{\bar{g}_0} \sup_{t>0} \int_0^t \left| (t-s)\, e^{-\underline{a}(t-s)} \right|^2 ds = O(1).$$

By a similar argument $II = O(1)$. These bounds imply that $\Pr(\mathcal{B}^c) \to 0$. The last statement in the corollary is deduced from the proof of Corollary 4. ■

Proof. [Corollary 6] By Lemma 1, the approximation error will be zero as soon as $\bar{B} \geq B_0$, which will be eventually the case as $\bar{B} \to \infty$ and B_0 is finite. By the remarks in Section 3.6.3 the entropy integral is finite. Hence, the bound follows from (9). ■

Proof. [Corollary 7] By Lemma 2 and (13) the approximation error is a constant multiple of $V^{-2\alpha} + \max\left\{c_\alpha - \bar{B}, 0\right\}^2$. The univariate square uniform approximation rate $V^{-2\alpha}$ follows by the remarks in Section 3.6.4. Given that there are V elements in each Θ_k the entropy integral is a constant multiple of $\sqrt{\ln(1+V)}$. Inserting in (9), the bound is deduced as long as $V > 1$. In particular for $V \gtrsim (T/\ln T)^{1/(4\alpha)}$ the bound simplifies further. ■

Proof. [Corollary 8] The proof is the same as for Corollary 7. ■

Proof. [Corollary 9] As stated in Section 3.6.6, the approximation rate of Bernstein polynomials under the squared uniform loss is a constant multiple of $\alpha^2 V^{-1}$. Hence, by Lemma 2 and (13), the approximation error is a constant multiple of $\alpha^2 V^{-1} + \max\left\{B_0 - \bar{B}, 0\right\}^2$. In consequence, as $\bar{B} \to \infty$, the approximation error is eventually $O\left(\sqrt{\alpha/T}\right)$ when $V \gtrsim T^{1/2}\alpha^{3/2}$. By the remarks in Section 3.6.6, the entropy integral is $\alpha^{1/2}$. Inserting in (9) the bound follows. ■

A.1.6 Proof of Theorem 3

Define $h := b\theta$, and let $t \in [0,1]$. Let

$$h_m := \arg\sup_{h \in \bar{\mathcal{L}}} D_T(F_{m-1}, h - F_{m-1}).$$

By linearity, the maximum is obtained by a function $h = b\theta$ with $\theta \in \Theta_k$ for some k and $|b| \leq \bar{B}$. Hence, it is sufficient to maximize the absolute value of D_T w.r.t. θ as the coefficient b is not constrained in sign. Define,

$$G(F_{m-1}) := D_T(F_{m-1}, h_m - F_{m-1}),$$

so that for any $g \in \bar{\mathcal{L}}$,

$$L_T(g) - L_T(F_{m-1}) \leq G(F_{m-1}) \tag{A.16}$$

by concavity. For $m \geq 0$, define $\bar{\rho}_m = 2/(m+2)$. By concavity, again,

$$L_T(F_m) = \max_{\rho \in [0,1]} L_T(F_{m-1} + \rho(h - F_{m-1})) \geq L_T(F_{m-1}) + D_T(F_{m-1}, h - F_{m-1})\bar{\rho}_m + \frac{\bar{C}}{2}\bar{\rho}_m^2$$

where

$$\bar{C} := \min_{h,g \in \bar{\mathcal{L}}, t \in [0,1]} \frac{2}{t^2}\left[L_T(g + t(h-g)) - L_T(g) - D_T(g, t(h-g))\right] < 0.$$

The above two displays together with (A.16), imply

$$\begin{aligned} L_T(F_m) - L_T(g) &\geq L_T(F_{m-1}) - L_T(g) + \bar{\rho}_m G(F_{m-1}) + \frac{\bar{C}}{2}\bar{\rho}_m^2 \\ &\geq L_T(F_{m-1}) - L_T(g) + \bar{\rho}_m (L_T(g) - L_T(F_{m-1})) + \frac{\bar{C}}{2}\bar{\rho}_m^2 \\ &= (1 - \bar{\rho}_m)(L_T(F_{m-1}) - L_T(g)) + \frac{\bar{C}}{2}\bar{\rho}_m^2 \\ &\geq \frac{2\bar{C}}{m+2} \end{aligned} \tag{A.17}$$

for the given choice of $\bar{\rho}_m$ (mutatis mutandis, as in the proof of Theorem 1 in Jaggi (2013)). It remains to bound $\bar{C}$. By Taylor's expansion in Banach spaces,

$$L_T(g + t(h-g)) = L_T(g) + D_T(g, t(h-g)) + \frac{1}{2}H_T\left(g_*, t^2(h-g)^2\right),$$

for $g_* = t_* g + (1 - t_*)h$, and some $t_* \in [0,1]$, where

$$H_T\left(g, t^2(h-g)\right) = -\int_0^T t^2(h-g)^2 e^g ds.$$

This means that

$$\bar{C} \geq \min_{h,g \in \bar{\mathcal{L}}, t \in [0,1]} \frac{2}{t^2}\left[-\frac{1}{2}\int_0^T t^2 (h(X(s)) - g(X(s)))^2 e^{\bar{g}} ds\right] \geq -4Te^{\bar{g}}\bar{g}^2 \geq -4Te^{\bar{B}\bar{\theta}/\underline{w}}\left(\bar{B}\bar{\theta}/\underline{w}\right)^2$$

using (A.1). Substituting in (A.17) gives the result.

A.1.7 Proof of Proposition 1

Let $M := N - \Lambda$ and $h_t := g_t - g'_t$. To ease notation, suppose for the moment that S is an integer. Then, under the conditions of the proposition (the null hypothesis),

$$L_S(g, g') = \sum_{s=1}^{S} \int_{s-1}^{s} h_t(X(t))\, dM(t) = \sum_{s=1}^{S} Y_s.$$

Then, $\{Y_s : s = 1, 2, ...\}$ is a sequence of martingale differences. This follows from the law of iterated expectations and the fact that h_t is a predictable process. Denote the expectation conditioning on $\{Y_i : i \leq s\}$ by $\mathbb{E}_s$. The result will follow by an application of Theorem 2.3 in McLeish (1974). To this end, it is sufficient to show that (i.) $\mathbb{E}\left|\frac{1}{S}\sum_{s=1}^{S} Y_s^2\right| \to \sigma^2$, (ii.) $\lim_{S\to\infty} \mathbb{E}\max_{s\leq S} Y_s^2/S < \infty$ and (iii.) $\max_{s\leq S}\left|Y_s/\sqrt{S}\right| \to 0$ in probability. Note that

$$\mathbb{E}\left|\frac{1}{S}\sum_{s=1}^{S} Y_s^2\right| = \mathbb{E}\left|\frac{1}{S}\sum_{s=1}^{S} \mathbb{E}_{s-1}Y_s^2\right| \tag{A.18}$$

using iterated expectations and the fact that the elements in the sum are positive. Note that

$$\begin{aligned}\mathbb{E}_{s-1}Y_s^2 &= \mathbb{E}_{s-1}\left[\int_{s-1}^{s} h_t\left(X\left(t\right)\right)dM\left(t\right)\right]^2 \\ &= \mathbb{E}_{s-1}\left[\int_{s-1}^{s} h_t^2\left(X\left(t\right)\right)d\Lambda\left(t\right)\right]\end{aligned}$$

(e.g., Ogata, 1978, e.q. 2.1). Hence,

$$\frac{1}{S}\sum_{s=1}^{S}\mathbb{E}_{s-1}Y_s^2 = \left[\frac{1}{S}\sum_{s=1}^{S}\mathbb{E}_{s-1}\int_{s-1}^{s} h_t^2\left(X\left(t\right)\right)d\Lambda\left(t\right)\right].$$

By these remarks, (A.18) is equal to

$$\begin{aligned}\mathbb{E}\left|\frac{1}{S}\sum_{s=1}^{S}\mathbb{E}_{s-1}\int_{s-1}^{s} h_t^2\left(X\left(t\right)\right)d\Lambda\left(t\right)\right| &= \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}\int_{s-1}^{s} h_t^2\left(X\left(t\right)\right)d\Lambda\left(t\right) \\ &= \mathbb{E}\frac{1}{S}\int_{0}^{S} h_t^2\left(X\left(t\right)\right)d\Lambda\left(t\right),\end{aligned}$$

using the fact that the terms in the sum are positive. By the conditions of the proposition

$$\sigma_S^2 := \frac{1}{S}\int_0^S h_t^2\left(X\left(t\right)\right)d\Lambda\left(t\right) \to \sigma^2 > 0$$

in probability. The sequence $\left(\sigma_S^2\right)_{S\geq 1}$ is uniformly bounded. In consequence, convergence in probability implies convergence in L_1, i.e. $\mathbb{E}\sigma_S^2 \to \sigma^2$. This verifies the first condition (i.). Now,

$$\mathbb{E}\max_{s\leq S}\frac{Y_s^2}{S} \leq \frac{1}{S}\mathbb{E}\sum_{s=1}^{S} Y_s^2$$

bounding the maximum by the sum. By the previous calculations deduce that the above is bounded, which then verifies the second condition (ii.). Finally,

$$\begin{aligned}\max_{s\leq S}|Y_s|/\sqrt{S} &= \frac{1}{\sqrt{S}}\max_{s\leq S}\left|\int_{s-1}^{s} h_t\left(X\left(t\right)\right)dM\left(t\right)\right| \\ &\lesssim \frac{1}{\sqrt{S}}\max_{s\leq S}\left|\int_{s-1}^{s} dN\left(t\right)\right| + \frac{1}{\sqrt{S}}\max_{s\leq S}\left|\int_{s-1}^{s} d\Lambda\left(t\right)\right| \\ &= \frac{1}{\sqrt{S}}\max_{s\leq S}\left[N\left(s\right)-N\left(s-1\right)\right] + \frac{1}{\sqrt{S}}\max_{s\leq S}\Lambda\left(\left[s-1,s\right]\right)\end{aligned}$$

where the inequality uses the fact that h_t is bounded. The last term on the r.h.s. is $O_p\left(S^{-1/2}\right)$. A counting process N is increasing with the intensity. Since $\lambda\left(X\left(s\right)\right)\leq e^{\bar{g}_0}$ uniformly in s, there is a counting process N' with intensity density $e^{\bar{g}_0}$ such $\Pr\left(N\left(s\right)>n\right)\leq\Pr\left(N'\left(s\right)>n\right)$. In consequence, for any s, $\mathbb{E}\left[N\left(s\right)-N\left(s-1\right)\right]^4\leq\mathbb{E}\left[N'\left(s\right)-N'\left(s-1\right)\right]^4\leq C$ for some absolute constant C that depends on $\bar{g}_0$ only. The last inequality follows because N' is Poisson with intensity $e^{\bar{g}_0}$. By these remarks,

$$\begin{aligned}\mathbb{E}\frac{1}{\sqrt{S}}\max_{s\leq S}\left[N\left(s\right)-N\left(s-1\right)\right] &\leq \frac{1}{\sqrt{S}}\left(\mathbb{E}\max_{s\leq S}\left|N\left(s\right)-N\left(s-1\right)\right|^4\right)^{1/4} \\ &\leq \frac{1}{\sqrt{S}}\left(\sum_{s=1}^{S}\mathbb{E}\left|N\left(s\right)-N\left(s-1\right)\right|^4\right)^{1/4}\end{aligned}$$

bounding the maximum by the sum. Deduce that the above is $\left(C/S\right)^{1/4}=o\left(1\right)$. This verifies the third condition (iii.) required for the application of Theorem 2.3 in McLeish (1974).

If S is not an integer, write $\lfloor S\rfloor$ for its integer part. Then,

$$\frac{1}{\sqrt{S}}L_S\left(g,g'\right)=\left(\frac{\lfloor S\rfloor}{S}\right)^{1/2}\frac{1}{\sqrt{\lfloor S\rfloor}}\sum_{s=1}^{\lfloor S\rfloor}Y_s+\frac{1}{\sqrt{S}}\int_{\lfloor S\rfloor}^{S}h_t\left(X\left(t\right)\right)dM\left(t\right).$$

Clearly, $\lfloor S\rfloor/S\to1$. Moreover, by arguments similar to the ones used to verify the third condition (iii.) above, we deduce that the last term on the r.h.s. is $o_p\left(1\right)$. This shows the result using σ_S as scaling sequence rather than $\hat{\sigma}_S$. However, $\left|\hat{\sigma}_S^2-\sigma_S^2\right|=\left|\frac{1}{S}\int_0^S h_t^2\left(X\left(t\right)\right)dM\left(t\right)\right|\to$ 0 a.s., and we can use $\hat{\sigma}_S^2$ to define the t-statistic. This completes the proof.

A.2 Details Regarding Section 5.3.1

Define $Y_i := \exp\{g_0(X(T_i))\}$ and $Z_i := \sum_{T_j \le T_i} e^{-a_0(T_i - T_j)}$, and recall $R(T_{i+1}) = T_{i+1} - T_i$. Note that for $t \in (T_i, T_{i+1}]$, $\lambda(t) = \left(c_0 + Z_i e^{-a_0(t-T_i)}\right) Y_i$. In consequence,

$$\Lambda((T_i, T_{i+1}]) = \int_{T_i}^{T_{i+1}} \lambda(t)\, dt = \left[c_0 R(T_{i+1}) + \frac{Z_i}{a_0}\left(1 - e^{-a_0 R(T_{i+1})}\right)\right] Y_i$$

is exponentially distributed with mean one, conditioning on $\mathcal{F}_i := (T_i, Z_i, Y_i)$. Moreover, $Z_i = Z_{i-1} e^{-a_0(T_i - T_{i-1})} + 1$ with $Z_0 = 1$. Hence, define $c_1 = c_0 Y_i$, $c_2 = Y_i Z_i$, and simulate i.i.d. $[0,1]$ uniform random variables U_i's. We simulate $R(T_i)$ setting it equal to the s that solves $c_1 s + \frac{c_2}{a_0}\left(1 - e^{-a_0 s}\right) = -\ln U_i$. Given an initial guess $(2, 1.5)$ of of the true $(c_0, a_0) = (2, 1.3)$ we estimate $\exp\{g_T(X(t))\}$. Given $\exp\{g_T(X(t))\}$ we estimate c and a in $\left(c + \sum_{T_i < t} e^{-a(t-T_i)}\right)\exp\{g_T(X(t))\}$. We perform a second iteration.

Estimation of g is done using the algorithm in Section 3.5. In this case, the relevant part of the likelihood is

$$\sum_{i=1}^{n} g(T_{i-1}) - \sum_{i=1}^{n} \exp\{g(T_{i-1})\}\,\Delta_i$$

where

$$\Delta_i = cR(T_i) + \frac{Z_{i-1}}{a}\left(1 - e^{-aR(T_i)}\right)$$

and c and a are set to their guess/estimated values. Estimation of c and a is via maximum likelihood given $\exp\{g_T(X(t))\}$.