Multiclass Classification for Hawkes Processes (Supplementary material)

Christophe Denis[1] **Charlotte Dion-Blanc**[2] **Laure Sansonnet**[3]

[1]LAMA, Université Gustave Eiffel, France
[2]LPSM, Sorbonne Université, France
[3]AgroParisTech, MIA Paris-Saclay, Université Paris-Saclay, France

In this supplementary material, we give first a technical result in Section 1. Then, Section 2 proposes the proofs of main results.

For the sake of simplicity we denote $\mathcal{T}$ for $\mathcal{T}_T$. We use in the sequel the notation C which represents a positive constant that does not depend on n. Each time C is written in some equation, one should understand that there exists a positive constant such that the equation holds. Therefore, the values of C may change from line to line and even change in the same equation. When an index K appears, C_K represents a constant depending on K (and not on n).

1 A TECHNICAL RESULT

Let us remind the reader that $\mathcal{E}\,(g) = \mathcal{R}(g) - \mathcal{R}(g^*)$ for any classifier $g \in \mathcal{G}$.

Proposition 1.1. *For any classifier $g \in \mathcal{G}$, we have*

$$\mathcal{E}\,(g) = \mathbb{E}\left[\sum_{i,\, k \neq i}^{K} |\pi_i^*(\mathcal{T}) - \pi_k^*(\mathcal{T})| \mathbb{1}_{\{g^*(\mathcal{T})=i, g(\mathcal{T})=k\}}\right].$$

Proof. Let $g \in \mathcal{G}$, we have:

$$\begin{aligned}
\mathcal{E}\,(g) &= \mathbb{E}\left[\mathbb{1}_{\{g(\mathcal{T})\neq Y\}} - \mathbb{1}_{\{g^*(\mathcal{T})\neq Y\}}\right]\\
&= \mathbb{E}\Bigg[\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{k=1}^{K} \pi_i^*(\mathcal{T})\left(\mathbb{1}_{\{g(\mathcal{T})\neq i\}}\right.\\
&\qquad \left.-\mathbb{1}_{\{g^*(\mathcal{T})\neq i\}}\right)\mathbb{1}_{\{g^*(\mathcal{T})=j\}}\mathbb{1}_{\{g(\mathcal{T})=k\}}\Bigg]\\
&= \mathbb{E}\Bigg[\sum_{i=1}^{K}\sum_{k\neq i} \pi_i^*(\mathcal{T})\mathbb{1}_{\{g(\mathcal{T})=k\}}\mathbb{1}_{\{g^*(\mathcal{T})=i\}}\\
&\qquad -\sum_{k=1}^{K}\sum_{i\neq k} \pi_k^*(\mathcal{T})\mathbb{1}_{\{g(\mathcal{T})=k\}}\mathbb{1}_{\{g^*(\mathcal{T})=i\}}\Bigg]\\
&= \mathbb{E}\left[\sum_{i,\, k\neq i}^{K} (\pi_i^*(\mathcal{T}) - \pi_k^*(\mathcal{T}))\mathbb{1}_{\{g(\mathcal{T})=k\}}\mathbb{1}_{\{g*(\mathcal{T})=i\}}\right].
\end{aligned}$$

We deduce the result of Proposition 1.1 from the following observation on the event $\{g^*(\mathcal{T}) = i\}$

$$\pi_i^*(\mathcal{T}) - \pi_k^*(\mathcal{T}) = |\pi_i^*(\mathcal{T}) - \pi_k^*(\mathcal{T})|.$$

□

2 PROOFS OF MAIN RESULTS

Proof of Proposition 2.1. We first denote for all $k \in \mathcal{Y}$

$$\Phi_t^k := \frac{d\mathbb{P}_k|_{\mathcal{F}_t^N}}{d\mathbb{P}_0|_{\mathcal{F}_t^N}},$$

with $\mathcal{F}_T^N := \sigma\,(\mathcal{T}_T) = \sigma\,(N_t, 0 \leq t \leq T)$. We classically obtain:

$$\log(\Phi_t^k) = -\int_0^t (\lambda_k^*(s) - 1)\,\mathrm{d}s + \int_0^t \log(\lambda_k^*(s))\,\mathrm{d}N_s,$$

by writing *w.r.t.* a Poisson process measure of intensity 1 (see Chapter 13 of [Daley and Vere-Jones, 2003]). Thus,

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

for $t \geq 0$, we have the following equation for the mixture measure

$$d\mathbb{P}|_{\mathcal{F}_t^N} = \sum_{k=1}^K p_k d\mathbb{P}_k|_{\mathcal{F}_t^N} = \sum_{k=1}^K p_k \Phi_t^k d\mathbb{P}_0|_{\mathcal{F}_t^N}$$

and then

$$\frac{d\mathbb{P}_k|_{\mathcal{F}_t^N}}{d\mathbb{P}|_{\mathcal{F}_t^N}} = \frac{p_k \Phi_t^k d\mathbb{P}_0|_{\mathcal{F}_t^N}}{\sum_{j=1}^K p_j \Phi_t^j d\mathbb{P}_0|_{\mathcal{F}_t^N}} = \frac{\Phi_t^k}{\sum_{j=1}^K p_j \Phi_t^j}.$$

Finally, by using the definition of F_k^*, it comes

$$\pi_k^*(\mathcal{T}_T) = \frac{p_k^* \mathrm{e}^{F_k^*}}{\sum_{j=1}^K p_j^* \mathrm{e}^{F_j^*}},$$

that concludes the proof. □

Proof of Proposition 3.4. Let $(\mathbf{p}, \mu, \mathbf{h})$ and $(\mathbf{p}^{'}, \mu^{'}, \mathbf{h}^{'})$ two tuples. We denote π and $\pi^{'}$ the associated elements in Π (see Equation (5)). We have that

$$\begin{aligned}\left\|\pi(\mathcal{T}) - \pi^{'}(\mathcal{T})\right\|_1 \quad \leq \quad & \left\|\pi(\mathcal{T}) - \boldsymbol{\pi}_{\mathbf{p},\mu^{'},\mathbf{h}^{'}}(\mathcal{T})\right\|_1 \\ & + \left\|\boldsymbol{\pi}_{\mathbf{p},\mu^{'},\mathbf{h}^{'}}(\mathcal{T}) - \pi^{'}(\mathcal{T})\right\|_1 \quad (1)\end{aligned}$$

Since for any k, j and $(x_1, \ldots, x_K)$,

$$\left|\frac{\partial \phi_k^{\mathbf{p}}(x_1, \ldots, x_K)}{\partial p_j}\right| \leq \frac{1}{p_0},$$

we deduce by mean value inequality

$$\left\|\boldsymbol{\pi}_{\mathbf{p},\mu^{'},\mathbf{h}^{'}}(\mathcal{T}) - \pi^{'}(\mathcal{T})\right\|_1 \leq \frac{K}{p_0}\left\|\mathbf{p} - \mathbf{p}^{'}\right\|_1.$$

Besides for any k, j and $\mathbf{p}$,

$$\left|\frac{\partial \phi_k^{\mathbf{p}}(x_1, \ldots, x_K)}{\partial x_j}\right| \leq 1,$$

we also deduce

$$\begin{aligned}&\left\|\pi(\mathcal{T}) - \boldsymbol{\pi}_{\mathbf{p},\mu^{'},\mathbf{h}^{'}}(\mathcal{T})\right\|_1 \\ &\leq K \sum_{k=1}^K \left|F^{(\mu,h_k)}(\mathcal{T}) - F^{(\mu^{'},h_k^{'})}(\mathcal{T})\right|.\end{aligned}$$

Therefore, from Equation (1), we obtain

$$\begin{aligned}\mathbb{E}\left[\left\|\pi(\mathcal{T}) - \pi^{'}(\mathcal{T})\right\|_1\right] \leq & \frac{K}{p_0}\left\|\mathbf{p} - \mathbf{p}^{'}\right\|_1 \\ & + K \sum_{k=1}^K \mathbb{E}\left[\left|F^{(\mu,h_k)}(\mathcal{T}) - F^{(\mu^{'},h_k^{'})}(\mathcal{T})\right|\right].\end{aligned}$$

Hence, it remains to bound the second term in the *r.h.s.* of the above inequality. Using Cauchy-Schwarz inequality, for each k, we have that

$$\begin{aligned}&\mathbb{E}\left[\left|F^{(\mu,h_k)}(\mathcal{T}) - F^{(\mu^{'},h_k^{'})}(\mathcal{T})\right|\right] \\ &= \mathbb{E}\left[\left|\int_0^T \log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu^{'},h_k^{'})}(t)}\right) \mathrm{d}N_t - \int_0^T \left(\lambda^{(\mu,h_k)}(t) - \lambda^{(\mu^{'},h_k^{'})}(t)\right) \mathrm{d}t\right|\right] \\ &\leq \mathbb{E}\left[\left(\int_0^T \left|\log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu^{'},h_k^{'})}(t)}\right)\right| \mathrm{d}N_t\right)^2\right]^{1/2} \\ &\quad + \mathbb{E}\left[\int_0^T \left|\lambda^{(\mu,h_k)}(t) - \lambda^{(\mu^{'},h_k^{'})}(t)\right| \mathrm{d}t\right]. \quad (2)\end{aligned}$$

Now, we observe that

$$\left|\lambda^{(\mu,h_k)}(t) - \lambda^{(\mu^{'},h_k^{'})}(t)\right| \leq |\mu^{'} - \mu| + \|\mathbf{h} - \mathbf{h}^{'}\|_{\infty,T} N_T,$$

where $N_T = N_{[0,T]}$ denotes the number of jump times of the observed process lying on $[0, T]$. Therefore we deduce

$$\begin{aligned}&\mathbb{E}\left[\int_0^T \left|\lambda^{(\mu,h_k)}(t) - \lambda^{(\mu^{'},h_k^{'})}(t)\right| \mathrm{d}t\right] \\ &\leq T\left(|\mu^{'} - \mu| + \|\mathbf{h} - \mathbf{h}^{'}\|_{\infty,T}\mathbb{E}\left[N_T\right]\right). \quad (3)\end{aligned}$$

Now, we bound the first term in the *r.h.s.* of Equation (2). Using that $x \mapsto \log(1 + x)$ is Lipschitz we obtain:

$$\begin{aligned}&\left|\log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu^{'},h_k^{'})}(t)}\right)\right| \leq \left|\log\left(\frac{\mu}{\mu^{'}}\right)\right| \\ &\quad + \left|\frac{\lambda^{(\mu,h_k)}(t)}{\mu^{'}} - \frac{\lambda^{(\mu^{'},h_k^{'})}(t)}{\mu}\right| \\ &\leq \frac{1}{\mu_0}\left|\mu - \mu^{'}\right| + \frac{1}{\mu_0^2}\left|\mu\lambda^{(\mu,h_k)}(t) - \mu^{'}\lambda^{(\mu^{'},h_k^{'})}(t)\right| \\ &\leq \frac{1}{\mu_0}\left|\mu - \mu^{'}\right| + \frac{1}{\mu_0^2}\left(|\mu - \mu^{'}|\lambda^{(\mu^{'},h_k^{'})}(t) + \mu_1\left|\lambda^{(\mu,h_k)}(t) - \lambda^{(\mu^{'},h_k^{'})}(t)\right|\right) \\ &\leq \frac{1}{\mu_0}\left|\mu - \mu^{'}\right| + \frac{1}{\mu_0^2}\left(|\mu - \mu^{'}|\lambda^{(\mu^{'},h_k^{'})}(t) + \mu_1\left(|\mu^{'} - \mu| + \|\mathbf{h} - \mathbf{h}^{'}\|_{\infty,T} N_T\right)\right). \quad (4)\end{aligned}$$

Using Doob's decomposition, we get

$$\mathbb{E}\left[\left(\int_0^T \left|\log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right| \mathrm{d}N_t\right)^2\right] = \mathbb{E}\left[\int_0^T \log^2\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\lambda_Y^*(t)\,\mathrm{d}t\right] + \mathbb{E}\left[\left(\int_0^T \left|\log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right| \lambda_Y^*(t)\,\mathrm{d}t\right)^2\right]. \quad (5)$$

Using that $\mathbb{E}\left[(\lambda_Y^*(t))^2\right] < \infty$, the first term in the *r.h.s.* in Equation (5) can be bounded as follows

$$\begin{aligned}\mathbb{E}&\left[\int_0^T \log^2\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\lambda_Y^*(t)\,\mathrm{d}t\right] \\ &\leq \int_0^T \mathbb{E}\left[\log^4\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right]^{1/2} \mathbb{E}\left[(\lambda_Y^*(t))^2\right]^{1/2} \mathrm{d}t \\ &\leq CT \sup_{t\in[0,T]} \mathbb{E}\left[\log^4\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right]^{1/2}.\end{aligned}$$

Similarly, we obtain:

$$\begin{aligned}\mathbb{E}&\left[\left(\int_0^T \left|\log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right| \lambda_Y^*(t)\,\mathrm{d}t\right)^2\right] \\ &\leq T\mathbb{E}\left[\int_0^T \log^2\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)(\lambda_Y^*(t))^2\,\mathrm{d}t\right] \\ &\leq CT^2 \sup_{t\in[0,T]} \mathbb{E}\left[\log^4\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right]^{1/2}.\end{aligned}$$

Then, by Assumption 3.1, we get

$$\begin{aligned}\mathbb{E}&\left[\left(\int_0^T \left|\log\left(\frac{\lambda^{(\mu,h_k)}(t)}{\lambda^{(\mu',h_k')}(t)}\right)\right| \mathrm{d}N_t\right)^2\right] \\ &\leq C\left(|\mu-\mu'|^2 + \|\mathbf{h}-\mathbf{h}'\|_{\infty,T}^2\right) \\ &\leq C\left(2\mu_1|\mu-\mu'| + \|\mathbf{h}-\mathbf{h}'\|_{\infty,T}^2\right),\end{aligned}$$

where C is constant which depends on $\mu_0, \mu_1, \mathbf{h}^*, A_1$, and T. Finally, combining the above equation, Equations (3) and (2) yields the desired result. □

Proof of Corollary 3.5. Let $\pi \in \Pi$. We recall that

$$g_\pi(\mathcal{T}) = \operatorname*{argmax}_{k\in\mathcal{Y}} \pi^k(\mathcal{T})$$

for $h \in \mathcal{H}$. By Proposition 1.1 we then get

$$\begin{aligned}0 &\leq \mathcal{E}(g_\pi) \\ &= \mathbb{E}\left[\sum_{i,\,k\neq i}^K |\pi_i^*(\mathcal{T}) - \pi_k^*(\mathcal{T})|\mathbb{1}_{\{g_\pi(\mathcal{T})=k\}}\mathbb{1}_{\{g^*(\mathcal{T})=i\}}\right] \\ &\leq 2\mathbb{E}\left[\max_{k\in\mathcal{Y}}|\pi^k(\mathcal{T}) - \pi_k^*(\mathcal{T})|\mathbb{1}_{\{g_\pi(\mathcal{T})\neq g^*(\mathcal{T})\}}\right] \\ &\leq 2\sum_{k=1}^K \mathbb{E}\left[|\pi^k(\mathcal{T}) - \pi_k^*(\mathcal{T})|\right].\end{aligned}$$

Finally, applying Proposition 3.4, we obtain the desired result. □

Proof of Theorem 4.2. Let us remind the reader that $\widehat{\mathbf{p}} = (\widehat{p}_k)_{k=1,\ldots,K}$ with $\widehat{p}_k = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Y_i=k}$. We consider the following set $\mathcal{A} = \left\{\widehat{\mathbf{p}} : \min(\widehat{\mathbf{p}}) \geq \frac{p_0}{2}\right\}$, where p_0 is defined in Assumption 3.3.

On the one hand, note that on $\mathcal{A}^c$ we have

$$|\min(\mathbf{p}^*) - \min(\widehat{\mathbf{p}})| \geq \frac{p_0}{2},$$

which implies that there exists $k \in \mathcal{Y}$ s.t. $|p_k^* - \widehat{p}_k| \geq \frac{p_0}{2}$. Thus, by using Hoeffding's inequality we get

$$\begin{aligned}\mathbb{P}(\mathcal{A}^c) &\leq \sum_{k=1}^K \mathbb{P}\left(|p_k^* - \widehat{p}_k| \geq \frac{p_0}{2}\right) \\ &\leq 2K\mathrm{e}^{-np_0^2/2}. \quad (6)\end{aligned}$$

On the other hand, we focus on what happens on the event $\mathcal{A}$. First, we define

$$\tilde{\mathbf{f}} = \mathbf{f}_{(\widehat{\mathbf{p}},\tilde{\mu},\tilde{\mathbf{h}})} = \operatorname*{argmin}_{\mathbf{f}\in\widehat{\mathcal{F}}} \mathcal{R}(\mathbf{f}), \quad (7)$$

and then consider the following decomposition

$$\begin{aligned}\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*) &= (\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\tilde{\mathbf{f}})) + (\mathcal{R}(\tilde{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)) \\ &=: T_1 + T_2.\end{aligned}$$

By Equation (7), we have that

$$\begin{aligned}T_2 &= \mathcal{R}(\tilde{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*) \\ &= \mathcal{R}(\mathbf{f}_{(\widehat{\mathbf{p}},\tilde{\mu},\tilde{\mathbf{h}})}) - \mathcal{R}(\mathbf{f}_{(\widehat{\mathbf{p}},\mu^*,\mathbf{h}^*)}) \\ &\quad + \mathcal{R}(\mathbf{f}_{(\widehat{\mathbf{p}},\mu^*,\mathbf{h}^*)}) - \mathcal{R}(\mathbf{f}_{(\mathbf{p}^*,\mu^*,\mathbf{h}^*)}) \\ &\leq \mathcal{R}(\mathbf{f}_{(\widehat{\mathbf{p}},\mu^*,\mathbf{h}^*)}) - \mathcal{R}(\mathbf{f}_{(\mathbf{p}^*,\mu^*,\mathbf{h}^*)}).\end{aligned}$$

Therefore, on $\mathcal{A}$, we deduce from the mean value inequality that

$$T_2 \leq C_K \sum_{k=1}^K |\widehat{p}_k - p_k^*|^2, \quad (8)$$

where C_K is a constant depending on K. For establishing an upper bound for T_1, we first recall the definition (8) of the empirical risk minimizer over $\widehat{\mathcal{F}}$:

$$\widehat{\mathbf{f}} \in \underset{\mathbf{f}\in\widehat{\mathcal{F}}}{\operatorname{argmin}}\ \widehat{\mathcal{R}}(\mathbf{f}),$$

with

$$\widehat{\mathcal{R}}(\mathbf{f}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(Z_k^i - \mathbf{f}^k(\mathcal{T}^i)\right)^2.$$

Besides, let us introduce the set of parameters

$$\mathcal{S} = \{(\mathbf{p},\mu,\mathbf{h}):\ \mathbf{p}\in\mathcal{P}_{p_0/2},\ \mu\in[\mu_0,\mu_1],\ \mathbf{h}\in\mathcal{H}_A^K\}.$$

Then, on $\mathcal{A}$, we have by definition (7) of $\tilde{f}$,

$$\begin{aligned} T_1 &= \mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\tilde{\mathbf{f}}) \\ &= \mathcal{R}(\widehat{\mathbf{f}}) - \widehat{\mathcal{R}}(\widehat{\mathbf{f}}) + \widehat{\mathcal{R}}(\widehat{\mathbf{f}}) - \mathcal{R}(\tilde{\mathbf{f}}) \\ &\leq \mathcal{R}(\widehat{\mathbf{f}}) - \widehat{\mathcal{R}}(\widehat{\mathbf{f}}) + \widehat{\mathcal{R}}(\tilde{\mathbf{f}}) - \mathcal{R}(\tilde{\mathbf{f}}) \\ &\leq 2\sup_{(\mathbf{p},\mu,\mathbf{h})\in\mathcal{S}} |\mathcal{R}(\mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})}) - \widehat{\mathcal{R}}(\mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})})|. \quad (9)\end{aligned}$$

By combining (8) and (9), we obtain

$$\begin{aligned}&\mathbb{E}[\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)] \\ &\leq 2\mathbb{E}\left[\sup_{(\mathbf{p},\mu,\mathbf{h})\in\mathcal{S}} |\mathcal{R}(\mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})}) - \widehat{\mathcal{R}}(\mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})})|\mathbb{1}_{\mathcal{A}}\right] \\ &\quad + \mathbb{E}\left[C_K\sum_{k=1}^{K}|\widehat{p}_k - p_k^*|^2\mathbb{1}_{\mathcal{A}}\right] \\ &\quad + \mathbb{E}\left[\left(\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)\right)\mathbb{1}_{\mathcal{A}^c}\right].\end{aligned}$$

Since for $k\in\mathcal{Y}$, $\mathbb{E}[|\widehat{p}_k - p_k^*|^2] \leq C/n$ with C an absolute constant and $\widehat{\mathbf{f}}$ and $\mathbf{f}^*$ are bounded, by using Equation (6), we obtain:

$$\begin{aligned}&\mathbb{E}[\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)] \\ &\leq 2\mathbb{E}\left[\sup_{(\mathbf{p},\mu,\mathbf{h})\in\mathcal{S}} |\mathcal{R}(\mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})}) - \widehat{\mathcal{R}}(\mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})})|\right] \\ &\quad + C_K\left(\frac{1}{n} + \exp\left(-\frac{np_0^2}{2}\right)\right). \quad (10)\end{aligned}$$

It remains to control the first term in the right hand side of the above inequality. By Assumption 4.1 with $\varepsilon = 1/n$ and since $\mathbf{p}\in\mathcal{P}_{p_0/2}$, and $\mu\in[\mu_0,\mu_1]$, there exists a finite set $\mathcal{S}_n\subset\mathcal{S}$ such that for each $(\mathbf{p},\mu,\mathbf{h})\in\mathcal{S}$, there exists $(\mathbf{p}_n,\mu_n,\mathbf{h}_n)\in\mathcal{S}_n$ satisfying

$$\|\mathbf{p}_n - \mathbf{p}\|_1 \leq \frac{C_K}{n},\ \ |\mu_n - \mu| \leq \frac{1}{n},\ \ \|\mathbf{h}_n - \mathbf{h}\|_{\infty,T} \leq \frac{1}{n}.$$

Moreover, we have $\log(\operatorname{card}(\mathcal{S}_n)) \leq C_K\log(n^d)$. For $(\mathbf{p},\mu,\mathbf{h}) \in \mathcal{S}$, let us denote $\mathbf{f} = \mathbf{f}_{(\mathbf{p},\mu,\mathbf{h})}$ and $\mathbf{f}_n = \mathbf{f}_{(\mathbf{p}_n,\mu_n,\mathbf{h}_n)}$ the corresponding element of $\mathcal{S}_n$. Then, we have

$$\begin{aligned}|\mathcal{R}(\mathbf{f}) - \widehat{\mathcal{R}}(\mathbf{f})| &\leq |\mathcal{R}(\mathbf{f}) - \mathcal{R}(\mathbf{f}_n)| \\ &\quad + |\mathcal{R}(\mathbf{f}_n) - \widehat{\mathcal{R}}(\mathbf{f}_n)| + \left|\widehat{\mathcal{R}}(\mathbf{f}_n) - \widehat{\mathcal{R}}(\mathbf{f})\right|.\end{aligned}$$

Moreover, since $\mathbf{f}$ and $\mathbf{f_n}$ are bounded, we deduce that by denoting $\pi_n := \boldsymbol{\pi}_{\mathbf{p}_n,\mu_n,\mathbf{h}_n}$

$$\mathbb{E}\left[|\mathcal{R}(\mathbf{f}) - \mathcal{R}(\mathbf{f}_n)|\right] \leq \mathbb{E}\left[\|\pi(\mathcal{T}) - \pi_n(\mathcal{T})\|_1\right] \leq \frac{C}{n},$$

where the last inequality is obtained with the same arguments as in the proof of Proposition 3.4. In the same way, we also get

$$\mathbb{E}\left[\left|\widehat{\mathcal{R}}(\mathbf{f}) - \widehat{\mathcal{R}}(\mathbf{f}_n)\right|\right] \leq \frac{C}{n}.$$

Finally, from the above inequalities, we obtain that

$$\begin{aligned}&\mathbb{E}\left[\sup_{\mathcal{S}}\left|\mathcal{R}(\mathbf{f}) - \widehat{\mathcal{R}}(\mathbf{f})\right|\right] \\ &\leq \frac{2C}{n} + \mathbb{E}\left[\max_{\mathcal{S}_n}\left|\mathcal{R}(\mathbf{f}) - \widehat{\mathcal{R}}(\mathbf{f})\right|\right].\end{aligned}$$

Moreover, by Hoeffding's inequality, it comes for $t\geq 0$,

$$\begin{aligned}&\mathbb{P}\left(\max_{\mathcal{S}_n}|\widehat{\mathcal{R}}(\mathbf{f}) - \mathcal{R}(\mathbf{f})| \geq t\right) \\ &\leq \min(1, 2\operatorname{card}(\mathcal{S}_n)\exp(-2nt^2)).\end{aligned}$$

Integrating the previous equation leads to

$$\begin{aligned}&\mathbb{E}\left[\max_{\mathcal{S}_n}|\widehat{\mathcal{R}}(\mathbf{f}) - \mathcal{R}(\mathbf{f})|\right] \\ &\leq \int_0^\infty \min(1, \exp(\log(2\operatorname{card}(\mathcal{S}_n)) - 2nt^2))\ \mathrm{d}t \\ &\leq \int_0^\infty \exp\left(-(2nt^2 - \log(2\operatorname{card}(\mathcal{S}_n)))_+\right)\ \mathrm{d}t \\ &\leq \sqrt{\frac{\log(2\operatorname{card}(\mathcal{S}_n))}{2n}} + \frac{\sqrt{\pi}}{2\sqrt{2n}}.\end{aligned}$$

Finally, since there are at least two elements in $\mathcal{S}_n$, combining the above inequality and Equation (10) yields

$$\mathbb{E}[\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)] \leq \sqrt{\frac{\log(2\operatorname{card}(\mathcal{S}_n))}{2n}} + \frac{C}{n},$$

which concludes the proof. □

Proof of Theorem 4.3. Let us denote

$$\Delta_n := \sum_{k=1}^{K}(\widehat{p}_k - p_k^*)^2,$$

where based on $\mathcal{D}_{n_1} := \mathcal{D}_n^1$, $\widehat{p}_k = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbb{1}_{Y_i=k}$. Note that Δ_n is independent from $\mathcal{D}_{n_2} := \mathcal{D}_n^2$. Recall that n is assumed to be even and $n_1 = n_2 = n/2$.

Let us work again on the set $\mathcal{A} = \left\{\widehat{\mathbf{p}} :\ \min(\widehat{\mathbf{p}}) \geq \frac{p_0}{2}\right\}$. As in proof of Theorem 4.2, we can write

$$\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*) \leq \mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\tilde{\mathbf{f}}) + \mathcal{R}(\tilde{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*),$$

and from Equation (8), the second term in the right hand side of the above inequality is bounded by $C_K\Delta_n$.

Let us denote

$$D_{\mathbf{f}} := \mathcal{R}(\mathbf{f}) - \mathcal{R}(\tilde{\mathbf{f}})$$

and

$$\widehat{D}_{\mathbf{f}} := \widehat{\mathcal{R}}(\mathbf{f}) - \widehat{\mathcal{R}}(\tilde{\mathbf{f}}).$$

Furthermore, let us introduce

$$\tilde{\mathcal{S}} = \{(\mu, \mathbf{h}) :\ \mu \in [\mu_0, \mu_1],\ \mathbf{h} \in \mathcal{H}_A^K\}.$$

By Assumption 4.1, there exists a subset $\tilde{\mathcal{S}_n} \subset \tilde{\mathcal{S}}$ with $\log(\mathrm{card}(\tilde{\mathcal{S}}_n)) \leq C\log(n^d)$, such that for each $(\mu, \mathbf{h}) \in \tilde{\mathcal{S}}$, there exists $(\mu_n, \mathbf{h}_n) \in \tilde{\mathcal{S}}_n$ satisfying

$$|\mu_n - \mu| \leq \frac{1}{n} \quad \text{and} \quad \|\mathbf{h}_n - \mathbf{h}\|_{\infty,T} \leq \frac{1}{n}.$$

For $(\mu, \mathbf{h}) \in \tilde{\mathcal{S}}$, let us denote $\mathbf{f} = \mathbf{f}_{(\widehat{\mathbf{p}},\mu,\mathbf{h})}$ and $\mathbf{f}_n = \mathbf{f}_{(\widehat{\mathbf{p}},\mu_n,\mathbf{h}_n)}$ the associated element of $\tilde{\mathcal{S}_n}$. Then, the following decomposition holds

$$\begin{aligned} D_{\widehat{\mathbf{f}}} &\leq D_{\widehat{\mathbf{f}}} - 2\widehat{D}_{\widehat{\mathbf{f}}} \\ &= (D_{\widehat{\mathbf{f}}} - D_{\mathbf{f}_n}) + (2\widehat{D}_{\mathbf{f}_n} - 2\widehat{D}_{\widehat{\mathbf{f}}}) \\ &\quad + (D_{\mathbf{f}_n} - 2\widehat{D}_{\mathbf{f}_n}) \\ &=: T_1 + T_2 + T_3. \end{aligned}$$

As in proof of Theorem 4.2 and using same arguments as in proof of Proposition 3.4, we have

$$\mathbb{E}\left[T_i\right] \leq \frac{C}{n}, \quad \text{for } i = 1, 2.$$

Besides,

$$T_3 \leq \max_{\tilde{\mathcal{S}_n}}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}}).$$

Therefore, gathering the previous inequalities, we deduce that

$$\begin{aligned} &\mathbb{E}[\mathcal{R}(\widehat{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)] \\ &\qquad \leq \mathbb{E}\left[\max_{\tilde{\mathcal{S}_n}}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}})\mathbb{1}_{\mathcal{A}}\right] \\ &\qquad\quad + C_K\left(\frac{1}{n} + \exp\left(-\frac{np_0^2}{4}\right)\right). \qquad (11) \end{aligned}$$

Therefore to finish the proof it remains to control the first term in the right hand side of Inequality (11). For $u \geq 0$, on $\mathcal{A}$ and conditionally on $\mathcal{D}_{n_1}$, it holds that,

$$\begin{aligned} &\mathbb{E}\left[\max_{\tilde{\mathcal{S}}_n}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}})\right] \\ &\qquad \leq u + \int_u^{\infty} \mathbb{P}\left(\max_{\tilde{\mathcal{S}}_n}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}}) \geq t\right) dt. \qquad (12) \end{aligned}$$

Let us introduce the least squares function

$$l_{\mathbf{f}}(Z, \mathcal{T}) := \sum_{k=1}^{K}(Z_k - \mathbf{f}^k(\mathcal{T}))^2.$$

Since for each $(\mu, \mathbf{h}) \in \tilde{\mathcal{S}}$, $\mathbf{f}_{(\widehat{\mathbf{p}},\mu,\mathbf{h})}$ are uniformly bounded by 1, we get from Bernstein's inequality, conditionally on $\mathcal{D}_{n_1}$, for $t \geq 0$

$$\begin{aligned} \mathbb{P}\left(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}} \geq t\right) &\leq \mathbb{P}\left(2(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}}) \geq t + D_{\mathbf{f}}\right) \\ &\leq \exp\left(\frac{-n(t + D_{\mathbf{f}})^2/8}{B_{\mathbf{f}} + (t + D_{\mathbf{f}})4K/3}\right), \qquad (13) \end{aligned}$$

with

$$B_{\mathbf{f}} := \mathbb{E}\left[\left(l_{\mathbf{f}}(Z, \mathcal{T}) - l_{\tilde{\mathbf{f}}}(Z, \mathcal{T})\right)^2\right].$$

Besides, conditionally on $\mathcal{D}_{n_1}$, we have

$$l_{\mathbf{f}}(Z, \mathcal{T}) - l_{\mathbf{f}^*}(Z, \mathcal{T}) \leq C\sum_{k=1}^{K}\left(\mathbf{f}^k(\mathcal{T}) - \mathbf{f}^{*k}(\mathcal{T})\right).$$

Therefore, conditionally on $\mathcal{D}_{n_1}$, we deduce from Cauchy-Schwartz Inequality

$$\begin{aligned} &\mathbb{E}\left[\left(l_{\mathbf{f}}(Z, \mathcal{T}) - l_{\mathbf{f}^*}(Z, \mathcal{T})\right)^2\right] \\ &\qquad \leq C_K\sum_{k=1}^{K}\mathbb{E}\left[(\mathbf{f}^k(\mathcal{T}) - \mathbf{f}^{*k}(\mathcal{T}))^2\right] \\ &\qquad = C_K\left(\mathcal{R}(\mathbf{f}) - \mathcal{R}(\mathbf{f}^*)\right). \end{aligned}$$

Thus, writing

$$\begin{aligned} B_{\mathbf{f}} &\leq 2\mathbb{E}\left[\left(l_{\mathbf{f}}(Z, \mathcal{T}) - l_{\mathbf{f}^*}(Z, \mathcal{T})\right)^2\right] \\ &\quad + 2\mathbb{E}\left[\left(l_{\tilde{\mathbf{f}}}(Z, \mathcal{T}) - l_{\mathbf{f}^*}(Z, \mathcal{T})\right)^2\right], \end{aligned}$$

we deduce

$$B_{\mathbf{f}} \leq C_K\left(\mathcal{R}(\mathbf{f}) - \mathcal{R}(\mathbf{f}^*) + \mathcal{R}(\tilde{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)\right).$$

Then, as $\mathcal{R}(\mathbf{f}) - \mathcal{R}(\mathbf{f}^*) = \mathcal{R}(\mathbf{f}) - \mathcal{R}(\tilde{\mathbf{f}}) + \mathcal{R}(\tilde{\mathbf{f}}) - \mathcal{R}(\mathbf{f}^*)$, conditionally on $\mathcal{D}_{n_1}$ and on the event $\mathcal{A}$, we deduce from the above inequality and Equation (8) that

$$B_{\mathbf{f}} \leq C_K\left(D_{\mathbf{f}} + \Delta_n\right).$$

Hence, from Inequality (13), we get for $t \geq \Delta_n$,

$$\mathbb{P}\left(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}} \geq t\right) \leq \exp\left(-C_K n t\right),$$

which leads to

$$\mathbb{P}\left(\max_{\tilde{\mathcal{S}}_n}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}}) \geq t\right) \leq \text{card}(\tilde{\mathcal{S}}_n) \exp\left(-C_K n t\right).$$

In view of Equation (12), we then obtain that, conditionally on $\mathcal{D}_{n_1}$,

$$\mathbb{E}\left[\max_{\tilde{\mathcal{S}}_n}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}})\mathbb{1}_{\mathcal{A}}\right] \leq \max\left(\Delta_n, \frac{C_K \log(\tilde{\mathcal{S}}_n)}{n}\right) + \int_{C_K \log(\tilde{\mathcal{S}}_n)/n}^{+\infty} \exp(-C_K n t) dt.$$

Finally, integrating the above inequality ,*w.r.t.* $\mathcal{D}_{n_1}$, yields

$$\mathbb{E}\left[\max_{\tilde{\mathcal{S}}_n}(D_{\mathbf{f}} - 2\widehat{D}_{\mathbf{f}})\mathbb{1}_{\mathcal{A}}\right] \leq \frac{C_K \log(\tilde{\mathcal{S}}_n)}{n}.$$

Hence, this inequality combined with Equation (11) give the desired result.

□

References

DJ Daley and D Vere-Jones. Basic properties of the poisson process. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, pages 19–40, 2003.