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 $T$ for $T_T$. 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

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

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

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

## 2 PROOFS OF MAIN RESULTS

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

$$\Phi_k^i := \frac{d\mathbb{P}_k|_{\mathcal{F}_T}}{d\mathbb{P}_i|_{\mathcal{F}_T}},$$

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

$$\log(\Phi_k^i) = - \int_0^t (\lambda_k^*(s) - 1) \, ds + \int_0^t \log(\lambda_k^*(s)) \, dN_s,$$

by writing w.r.t. a Poisson process measure of intensity 1 (see Chapter 13 of [Daley and Vere-Jones 2003]). Thus,
for \( t \geq 0 \), we have the following equation for the mixture measure
\[
dP |_{\mathcal{F}_t^N} = \sum_{k=1}^K p_k dP_k |_{\mathcal{F}_t^N} = \sum_{k=1}^K p_k \Phi_k^t dP_0 |_{\mathcal{F}_t^N}
\]
and then
\[
dP_k |_{\mathcal{F}_t^N} = \frac{p_k \Phi_k^t dP_0 |_{\mathcal{F}_t^N}}{\sum_{j=1}^K p_j \Phi_j^t dP_0 |_{\mathcal{F}_t^N}} = \frac{\Phi_k^t}{\sum_{j=1}^K p_j \Phi_j^t}.
\]
Finally, by using the definition of \( F_k^* \), it comes
\[
\pi_k^* (T_T) = \frac{p_k^* e^{F_k^*}}{\sum_{j=1}^K p_j^* e^{F_j^*}},
\]
that concludes the proof. □

**Proof of Proposition 3.2.** Let \((p, \mu, h)\) and \((p', \mu', h')\) two tuples. We denote \( \pi \) and \( \pi' \) the associated elements in \( \Pi \) (see Equation \( 3 \)). We have that
\[
\left\| \pi(T) - \pi'(T) \right\|_1 \leq \left\| \pi(T) - \pi_{p, \mu', h'}(T) \right\|_1 + \left\| \pi_{p, \mu', h'}(T) - \pi'(T) \right\|_1
\]
(1)
Since for any \( k, j \) and \((x_1, \ldots, x_K)\),
\[
\left| \frac{\partial \Phi_k^t(x_1, \ldots, x_K)}{\partial p_j} \right| \leq \frac{1}{p_0},
\]
we deduce by mean value inequality
\[
\left\| \pi_{p, \mu', h'}(T) - \pi'(T) \right\|_1 \leq \frac{K}{p_0} \left\| p - p' \right\|_1.
\]
Besides for any \( k, j \) and \( p \),
\[
\left| \frac{\partial \Phi_k^t(x_1, \ldots, x_K)}{\partial x_j} \right| \leq 1,
\]
we also deduce
\[
\left\| \pi(T) - \pi_{p, \mu', h'}(T) \right\|_1 \leq K \sum_{k=1}^K \left| F_{p, (\mu, h)}(T) - F_{p, (\mu', h')}(T) \right|.
\]
Therefore, from Equation \( 3 \), we obtain
\[
\mathbb{E} \left[ \left\| \pi(T) - \pi'(T) \right\|_1 \right] \leq \frac{K}{p_0} \left\| p - p' \right\|_1
\]
\[
+ K \sum_{k=1}^K \mathbb{E} \left[ \left\| F_{p, (\mu, h)}(T) - F_{p, (\mu', h')}(T) \right\|_1 \right].
\]
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
\[
\mathbb{E} \left[ \left\| F_{p, (\mu, h)}(T) - F_{p, (\mu', h')}(T) \right\|_1 \right]
\]
\[
= \mathbb{E} \left[ \left\| \int_0^T \log \left( \frac{\lambda_{p, (\mu, h)}(t)}{\lambda_{p, (\mu', h')}(t)} \right) dN_t \right\|_1 \right]
\]
\[
- \int_0^T \left( \lambda_{p, (\mu, h)}(t) - \lambda_{p, (\mu', h')}(t) \right) dt
\]
\[
\leq \mathbb{E} \left[ \left( \int_0^T \left| \frac{\lambda_{p, (\mu, h)}(t)}{\lambda_{p, (\mu', h')}(t)} \right| dN_t \right)^{1/2} \right] + \mathbb{E} \left[ \int_0^T \left| \lambda_{p, (\mu, h)}(t) - \lambda_{p, (\mu', h')}(t) \right| dt \right].
\]
(2)
Now, we observe that
\[
\left| \lambda_{p, (\mu, h)}(t) - \lambda_{p, (\mu', h')}(t) \right| \leq |\mu' - \mu| + \|h - h'\|_\infty, T N_T,
\]
where \( N_T = N_{[0, T]} \) denotes the number of jump times of the observed process lying on \([0, T]\). Therefore we deduce
\[
\mathbb{E} \left[ \int_0^T \left| \lambda_{p, (\mu, h)}(t) - \lambda_{p, (\mu', h')}(t) \right| dt \right]
\]
\[
\leq T \left( |\mu' - \mu| + \|h - h'\|_\infty, T \mathbb{E} [N_T] \right).
\]
(3)
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:
\[
\log \left( \frac{\lambda_{p, (\mu, h)}(t)}{\lambda_{p, (\mu', h')}(t)} \right) \leq \log \left( \frac{\mu'}{\mu} \right)
\]
\[
+ \left| \frac{\lambda_{p, (\mu, h)}(t)}{\mu'} - \frac{\lambda_{p, (\mu', h')}(t)}{\mu} \right|
\]
\[
\leq \frac{1}{\mu_0} |\mu' - \mu| + \frac{1}{\mu_0^2} \left| \mu \lambda_{p, (\mu, h)}(t) - \mu' \lambda_{p, (\mu', h')}(t) \right|
\]
\[
\leq \frac{1}{\mu_0} |\mu' - \mu| + \frac{1}{\mu_0^3} \left| \mu' \lambda_{p, (\mu, h)}(t) - \mu \lambda_{p, (\mu', h')}(t) \right|
\]
\[
+ \frac{\mu_3}{\mu_0^2} \left| \lambda_{p, (\mu, h)}(t) - \lambda_{p, (\mu', h')}(t) \right|
\]
\[
\leq \frac{1}{\mu_0} |\mu' - \mu| + \frac{1}{\mu_0^3} \left| \mu - \mu' \lambda_{p, (\mu', h')}(t) \right|
\]
\[
+ \mu_1 \left| \mu' - \mu \right| + \|h - h'\|_\infty, T N_T \right).
\]
(4)
Using Doob’s decomposition, we get
\[
\mathbb{E} \left[ \left( \int_0^T \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \mathcal{L}_Y(t) \, dt \right)^2 \right] =
\mathbb{E} \left[ \int_0^T \log \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \lambda_Y(t) \, dt \right]
+ \mathbb{E} \left[ \left( \int_0^T \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \lambda_Y(t) \, dt \right)^2 \right].
\] (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
\[
\mathbb{E} \left[ \int_0^T \log \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \lambda_Y(t) \, dt \right] \leq C T \sup_{t \in [0,T]} \mathbb{E} \left[ \log^4 \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \right]^{1/2} \mathbb{E} \left[ (\lambda_Y(t))^2 \right]^{1/2} dt
\leq C T \sup_{t \in [0,T]} \mathbb{E} \left[ \log^4 \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \right]^{1/2}.\]

Similarly, we obtain:
\[
\mathbb{E} \left[ \left( \int_0^T \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \mathcal{L}_Y(t) \, dt \right)^2 \right] \leq CZ^2 \sup_{t \in [0,T]} \mathbb{E} \left[ \log^4 \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \right]^{1/2}.
\]

Then, by Assumption 3.1, we get
\[
\mathbb{E} \left[ \left( \int_0^T \left( \frac{\lambda(\mu, h_k(t))}{\lambda(\mu', h_k(t))} \right) \, dN_t \right)^2 \right] \leq C \left( |\mu - \mu'|^2 + \|h - h\|_{\infty,T}^2 \right)
\leq C \left( 2\mu_1 |\mu - \mu'| + \|h - h\|_{\infty,T}^2 \right),
\]
where \( C \) is constant which depends on \( \mu_0, \mu_1, h^*, A_1 \), and \( T \). Finally, combining the above equation, Equations (3) and (2) yields the desired result.

\[\text{Proof of Corollary 3.5}\] Let \( \pi \in \Pi \). We recall that
\[g(\pi^k) = \arg \max_{k \in \mathcal{Y}} \pi^k(T)\]
for \( h \in \mathcal{H} \). By Proposition 1.1 we then get
\[
0 \leq \mathcal{E}(g_{\pi})
= \mathbb{E} \left[ \sum_{i, k \neq i} |\pi_i^k(T) - \pi_{\hat{k}}^*(T)| \mathbb{I}_{\{g_{\pi}(T) = k\}} \mathbb{I}_{\{g_{\hat{k}}^*(T) = i\}} \right]
\leq 2 \mathbb{E} \left[ \max_{k \in \mathcal{Y}} |\pi^k(T) - \pi_{\hat{k}}^*(T)| \mathbb{I}_{\{g_{\pi}(T) \neq g_{\hat{k}}^*(T)\}} \right]
\leq 2 \sum_{k = 1}^K \mathbb{E} \left[ |\pi^k(T) - \pi_{\hat{k}}^*(T)| \right].
\]

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

\[\text{Proof of Theorem 2.2}\] Let us remind the reader that \( \hat{\mathbf{p}} = (\hat{p}_k)_{k=1}^\infty \) with \( \hat{p}_k = \frac{1}{T} \sum_{t=1}^n \mathbb{I}_{Y_t = k} \). We consider the following set \( \mathcal{A} = \{ \hat{\mathbf{p}} : \min(\hat{\mathbf{p}}) \geq \frac{p_0}{2} \} \), where \( p_0 \) is defined in Assumption 3.3.
On the one hand, note that on \( \mathcal{A}^c \) we have
\[
|\min(\mathbf{p}^*) - \min(\hat{\mathbf{p}})| \geq \frac{p_0}{2},
\]
which implies that there exists \( k \in \mathcal{Y} \) s.t. \( |p_k^* - \hat{p}_k| \geq \frac{p_0}{2} \).
Thus, by using Hoeffding’s inequality we get
\[
\mathbb{P}(\mathcal{A}^c) \leq \sum_{k=1}^K \mathbb{P} \left( |p_k^* - \hat{p}_k| \geq \frac{p_0}{2} \right)
\leq 2K e^{-np_0^2/2}.
\] (6)

On the other hand, we focus on what happens on the event \( \mathcal{A} \). First, we define
\[
\hat{f} = f_{(\hat{\mathbf{p}}, \hat{\mathbf{h}}, \hat{\mathbf{h}})} = \arg \min_{f \in \mathcal{F}} \mathcal{R}(f),
\] (7)
and then consider the following decomposition
\[
\mathcal{R}(\hat{f}) - \mathcal{R}(f^*) = (\mathcal{R}(\hat{f}) - \mathcal{R}(\hat{f})) + (\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)) =: T_1 + T_2.
\]
By Equation (7), we have that
\[
T_2 = \mathcal{R}(\hat{f}) - \mathcal{R}(f^*)
= \mathcal{R}(f_{(\hat{\mathbf{p}}, \hat{\mathbf{h}}, \hat{\mathbf{h}})}) - \mathcal{R}(f_{(\hat{\mathbf{p}}, \mu, h^*)})
+ \mathcal{R}(f_{(\hat{\mathbf{p}}, \mu, h^*)}) - \mathcal{R}(f_{(\mathbf{p}^*, \mu, h^*)})
\leq \mathcal{R}(f_{(\hat{\mathbf{p}}, \mu, h^*)}) - \mathcal{R}(f_{(\mathbf{p}^*, \mu, h^*)}).
\]
Therefore, on \( \mathcal{A} \), we deduce from the mean value inequality that
\[
T_2 \leq C_{\mathcal{H}} \sum_{k=1}^K |\hat{p}_k - p_k^*|^2,
\] (8)
where $C_K$ is a constant depending on $K$. For establishing an upper bound for $T_1$, we first recall the definition of the empirical risk minimizer over $\hat{\mathcal{F}}$:

$$\hat{f} \in \text{argmin}_{f \in \hat{\mathcal{F}}} \mathcal{R}(f),$$

with

$$\mathcal{R}(f) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} (Z_k - f^k(T_i))^2.$$

Besides, let us introduce the set of parameters

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

Then, on $A$, we have by definition of $\hat{f}$,

$$T_1 = \mathcal{R}(\hat{f}) - \mathcal{R}(f^\ast)$$

$$= \mathcal{R}(\hat{f}) - \mathcal{R}(\hat{f}) + \hat{\mathcal{R}}(\hat{f}) - \mathcal{R}(\hat{f})$$

$$\leq \mathcal{R}(\hat{f}) - \mathcal{R}(\hat{f}) + \hat{\mathcal{R}}(\hat{f}) - \mathcal{R}(\hat{f})$$

$$\leq 2 \sup_{(p, \mu, h) \in \mathcal{S}} \left| \mathcal{R}(f_{(p, \mu, h)}) - \hat{\mathcal{R}}(f_{(p, \mu, h)}) \right|. \quad (9)$$

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

$$\mathbb{E}[\mathcal{R}(\hat{f}) - \mathcal{R}(f^\ast)]$$

$$\leq 2 \mathbb{E} \left[ \sup_{(p, \mu, h) \in \mathcal{S}} \left| \mathcal{R}(f_{(p, \mu, h)}) - \hat{\mathcal{R}}(f_{(p, \mu, h)}) \right| \right]$$

$$\leq C_K \sum_{k=1}^{K} \left| \hat{p}_k - p_k^\ast \right|^2 \mathbb{I}_{A_4} + \mathbb{E} \left[ \left( \mathcal{R}(\hat{f}) - \mathcal{R}(f^\ast) \right) \mathbb{I}_{A_4} \right].$$

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

$$\mathbb{E}[\mathcal{R}(\hat{f}) - \mathcal{R}(f^\ast)]$$

$$\leq 2 \mathbb{E} \left[ \sup_{(p, \mu, h) \in \mathcal{S}} \left| \mathcal{R}(f_{(p, \mu, h)}) - \hat{\mathcal{R}}(f_{(p, \mu, h)}) \right| \right]$$

$$\leq C_K \left( \frac{1}{n} + \exp \left( - \frac{n\beta_0^2}{2} \right) \right). \quad (10)$$

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 $p \in \mathcal{P}_{p_0/2}$, and $\mu \in [\mu_0, \mu_1]$, there exists a finite set $\mathcal{S}_n \subset \mathcal{S}$ such that for each $(p, \mu, h) \in \mathcal{S}$, there exists $(p_n, \mu_n, h_n) \in \mathcal{S}_n$ satisfying

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

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

$$\mathbb{E}[\mathcal{R}(f) - \hat{\mathcal{R}}(f)] \leq \mathbb{E}[\mathcal{R}(f) - \mathcal{R}(f_n)]$$

$$\leq |\mathcal{R}(f) - \hat{\mathcal{R}}(f)| + |\mathcal{R}(f_n) - \hat{\mathcal{R}}(f_n)| + |\hat{\mathcal{R}}(f_n) - \hat{\mathcal{R}}(f)|.$$

Moreover, since $f$ and $f_n$ are bounded, we deduce that by denoting $\pi_n := \pi_{n, p_n, \mu_n, h_n}$,

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

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

$$\mathbb{E}[|\hat{\mathcal{R}}(f) - \hat{\mathcal{R}}(f_n)|] \leq \frac{C}{n}.$$

Finally, from the above inequalities, we obtain that

$$\mathbb{E} \left[ \sup_{\mathcal{S}_n} \left| \mathcal{R}(f) - \hat{\mathcal{R}}(f) \right| \right]$$

$$\leq \frac{2C}{n} + \mathbb{E} \left[ \max_{\mathcal{S}_n} \left| \mathcal{R}(f) - \hat{\mathcal{R}}(f) \right| \right].$$

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

$$\mathbb{P} \left( \max_{\mathcal{S}_n} \left| \mathcal{R}(f) - \hat{\mathcal{R}}(f) \right| \geq t \right)$$

$$\leq \min(1, 2 \text{card}(\mathcal{S}_n) \exp(-2nt^2)).$$

Integrating the previous equation leads to

$$\mathbb{E} \left[ \max_{\mathcal{S}_n} \left| \mathcal{R}(f) - \hat{\mathcal{R}}(f) \right| \right]$$

$$\leq \int_0^{\infty} \min(1, \exp(-2 \text{card}(\mathcal{S}_n) - 2nt^2)) \, dt$$

$$\leq \int_0^{\infty} \exp \left( -(2nt^2 - \log(2 \text{card}(\mathcal{S}_n))) \right) \, dt$$

$$\leq \sqrt{\frac{\log(2 \text{card}(\mathcal{S}_n))}{2n}} + \frac{\sqrt{n}}{2 \sqrt{2n}}.$$

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

$$\mathbb{E}[\mathcal{R}(\hat{f}) - \mathcal{R}(f^\ast)] \leq \sqrt{\frac{\log(2 \text{card}(\mathcal{S}_n))}{2n}} + \frac{C}{n},$$

which concludes the proof.

\textbf{Proof of Theorem 4.3} Let us denote

$$\Delta_n := \sum_{k=1}^{K} (\hat{p}_k - p_k^\ast)^2,$$
where based on $D_{n_1} := D_1$, $\hat{p}_k = \frac{1}{n_1} \sum_{i=1}^{n_1} I_{Y_i = k}$. Note that $\Delta_n$ is independent from $D_{n_2} := D_2^2$. Recall that $n$ is assumed to be even and $n_1 = n_2 = n/2$.

Let us work again on the set $\mathcal{A} = \{\hat{p} : \min(\hat{p}) \geq \frac{p_0}{2}\}$. As in proof of Theorem 4.2 and using same arguments as in where based on $D$, we can write

$$\mathcal{R}(\hat{f}) - \mathcal{R}(f^*) \leq \mathcal{R}(\hat{f}) - \mathcal{R}(\hat{f}) + \mathcal{R}(\hat{f}) - \mathcal{R}(f^*),$$

and from Equation (3), the second term in the right hand side of the above inequality is bounded by $C_K \Delta_n$.

Let us denote

$$D_f := \mathcal{R}(f) - \mathcal{R}(\hat{f})$$

and

$$\hat{D}_f := \hat{\mathcal{R}}(f) - \mathcal{R}(\hat{f}).$$

Furthermore, let us introduce

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

By Assumption 4.4, there exists a subset $\tilde{\mathcal{S}}_n \subset \tilde{\mathcal{S}}$ with $\log(\text{card}(\tilde{\mathcal{S}}_n)) \leq C \log(n^d)$, such that for each $(\mu, h) \in \tilde{\mathcal{S}}_n$, there exists $(\mu_n, h_n) \in \mathcal{S}_n$ satisfying

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

For $(\mu, h) \in \tilde{\mathcal{S}}$, let us denote $f = f_{(\hat{\mu}, \mu, h)}$ and $f_n = f_{(\hat{\mu}, \mu_n, h_n)}$ the associated element of $\tilde{\mathcal{S}}_n$, then, the following decomposition holds

$$D_{\tilde{f}} \leq D_{\tilde{f}} - 2\hat{D}_{\tilde{f}}$$

$$= (D_{\tilde{f}} - D_{\tilde{f}}) + (2\hat{D}_{\tilde{f}} - 2\hat{D}_{\tilde{f}})$$

$$= (D_{\tilde{f}} - D_{\tilde{f}})$$

$$=: T_1 + T_2 + T_3.$$

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

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

Besides,

$$T_3 \leq \max_{S_n}(D_{\tilde{f}} - 2\hat{D}_{\tilde{f}}).$$

Therefore, gathering the previous inequalities, we deduce that

$$\mathbb{E}[\mathcal{R}(\tilde{f}) - \mathcal{R}(f^*)]$$

$$\leq \mathbb{E}[\max_{S_n}(D_{\tilde{f}} - 2\hat{D}_{\tilde{f}})1_{\mathcal{A}}] + C_K \left(\frac{1}{n} + \exp\left(-\frac{n p_0^2}{4}\right)\right).$$

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 $D_{n_1}$, it holds that,

$$\mathbb{E}\left[\max_{\mathcal{S}_n}(D_{\tilde{f}} - 2\hat{D}_{\tilde{f}})\right]$$

$$\leq u + \int_{u}^{\infty} \mathbb{P}\left(\max_{\mathcal{S}_n}(D_{\tilde{f}} - 2\hat{D}_{\tilde{f}}) \geq t\right) \, dt. \quad (12)$$

Let us introduce the least squares function

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

Since for each $(\mu, h) \in \tilde{\mathcal{S}}, f_{(\mu, h)}$ are uniformly bounded by 1, we get from Bernstein’s inequality, conditionally on $D_{n_1}$, for $t \geq 0$

$$\mathbb{P}\left(D_f - 2\hat{D}_f \geq t\right) \leq \mathbb{P}\left(2(D_f - 2\hat{D}_f) \geq t + D_f\right)$$

$$\leq \exp\left(-\frac{n(t + D_f)^2}{8B_f + (t + D_f)AK/3}\right), \quad (13)$$

with

$$B_f := \mathbb{E}\left[(l_f(Z, T) - l_{f^*}(Z, T))^2\right].$$

Besides, conditionally on $D_{n_1}$, we have

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

Therefore, conditionally on $D_{n_1}$, we deduce from Cauchy-Schwarz Inequality

$$\mathbb{E}\left[(l_f(Z, T) - l_{f^*}(Z, T))^2\right]$$

$$\leq C_K \sum_{k=1}^{K} \mathbb{E}\left[(f^k(T) - f^{*k}(T))^2\right]$$

$$= C_K (\mathcal{R}(f) - \mathcal{R}(f^*)).$$

Thus, writing

$$B_f \leq 2\mathbb{E}\left[(l_f(Z, T) - l_{f^*}(Z, T))^2\right]$$

$$+ 2\mathbb{E}\left[(l_{f^*}(Z, T) - l_{f^*}(Z, T))^2\right],$$

we deduce

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

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

$$B_f \leq C_K (D_f + \Delta_n).$$
Hence, from Inequality (13), we get for \( t \geq \Delta_n \),
\[
\mathbb{P}\left( D_f - 2 \hat{D}_f \geq t \right) \leq \exp\left( -C_K nt \right),
\]
which leads to
\[
\mathbb{P}\left( \max_{\mathcal{S}_n}(D_f - 2 \hat{D}_f) \geq t \right) \leq \text{card}(\mathcal{S}_n) \exp\left( -C_K nt \right).
\]
In view of Equation (12), we then obtain that, conditionally on \( D_n \),
\[
\mathbb{E}\left[ \max_{\mathcal{S}_n}(D_f - 2 \hat{D}_f) \mathbb{I}_A \right] \leq \max\left( \Delta_n, \frac{C_K \log(\mathcal{S}_n)}{n} \right)
\]
\[
+ \int_{C_K \log(\mathcal{S}_n)/n}^{+\infty} \exp( -C_K nt ) dt.
\]
Finally, integrating the above inequality, w.r.t. \( D_n \), yields
\[
\mathbb{E}\left[ \max_{\mathcal{S}_n}(D_f - 2 \hat{D}_f) \mathbb{I}_A \right] \leq \frac{C_K \log(\mathcal{S}_n)}{n}.
\]
Hence, this inequality combined with Equation (11) give the desired result.

References