

# Supplementary material for the paper: Finite sample improvement of Akaike's Information Criterion

Adrien Saumard<sup>1</sup>      Fabien Navarro<sup>2</sup>

<sup>1</sup>CREST, ENSAI, Université Rennes

<sup>2</sup>SAMM, Université Paris 1 Panthéon-Sorbonne

This supplement contains in Sections 1 and 2 the proofs of the results described in the main part of this article as well as some theoretical extensions that complement the analysis of the over-penalization procedure.

## 1 First proofs

### 1.1 Proofs Related to Section IV-A

**Proof of Proposition IV.1.** We fix  $x, \theta > 0$  and we set  $z > 0$  to be chosen later. Let us set for any  $I \in m$ ,  $\varphi_I = (P(I))^{-1/2} \mathbb{1}_I$ . The family of functions  $(\varphi_I)_{I \in m}$  forms an orthonormal basis of  $(m, \|\cdot\|_2)$ . By Cauchy-Schwarz inequality, we have

$$\chi_n(m) = \sup_{(a_I)_{I \in m} \in B_2} \left| (P_n - P) \left( \sum_{I \in m} a_I \varphi_I \right) \right| ,$$

where  $B_2 := \{(a_I)_{I \in m} ; \sum_{I \in m} a_I^2 \leq 1\}$ . Furthermore, the case of equality in Cauchy-Schwarz inequality gives us

$$\chi_n(m) = (P_n - P) \left( \sum_{I \in m} a_I^\infty \varphi_I \right) \text{ with } a_I^\infty = \frac{(P_n - P)(\varphi_I)}{\chi_n(m)} .$$

Hence, by setting  $\mathcal{A}(s) := B_2 \cap \{(a_I)_{I \in m} ; \sup_{I \in m} |a_I (P(I))^{-1/2}| \leq s\}$  for any  $s \geq 0$ , we get

$$\chi_n(m) \mathbf{1}_{\Omega_m(\theta)} \mathbf{1}_{\{\chi_n(m) \geq z\}} \leq \sup_{(a_I)_{I \in m} \in \mathcal{A}(\theta/z)} \left| (P_n - P) \left( \sum_{I \in m} a_I \varphi_I \right) \right| . \quad (1)$$

Apply now Bousquet's inequality ([3]) to the supremum in the right-hand side of (1). This gives for any  $\delta > 0$ ,

$$\begin{aligned} & \mathbb{P} \left( \chi_n(m) \mathbf{1}_{\Omega_m(\theta)} \mathbf{1}_{\{\chi_n(m) \geq z\}} \geq (1 + \delta) E_m + \sqrt{\frac{2\sigma_m^2 x}{n}} + \left( \frac{1}{\delta} + \frac{1}{3} \right) \frac{b_m x}{n} \right) \\ & \leq \mathbb{P} \left( \sup_{(a_I)_{I \in m} \in \mathcal{A}(\theta/z)} \left| (P_n - P) \left( \sum_{I \in m} a_I \varphi_I \right) \right| \geq (1 + \delta) E_m + \sqrt{\frac{2\sigma_m^2 x}{n}} + \left( \frac{1}{\delta} + \frac{1}{3} \right) \frac{b_m x}{n} \right) \\ & \leq \exp(-x) , \end{aligned} \quad (2)$$

with

$$E_m \leq \mathbb{E}[\chi_n(m)] \leq \mathbb{E}^{1/2}[\chi_n^2(m)] = \sqrt{\frac{D_m}{n}} ; \quad \sigma_m^2 \leq \sup_{(a_I)_{I \in m} \in B_2} \mathbb{V} \left( \sum_{I \in m} a_I \varphi_I(\xi_1) \right) \leq 1$$

and

$$b_m = \sup_{(a_I)_{I \in m} \in \mathcal{A}(\theta/z)} \left\| \sum_{I \in m} a_I \varphi_I \right\|_{\infty} \leq \frac{\theta}{z} .$$

Injecting the latter bounds in (2) and taking  $z = \sqrt{(D_m)/n} + \sqrt{2x/n}$ , we obtain that with probability at least  $1 - \exp(-x)$ ,

$$\chi_n(m) \mathbf{1}_{\Omega_m(\theta)} < (1 + \delta) \sqrt{\frac{D_m}{n}} + \sqrt{\frac{2x}{n}} + \left( \frac{1}{\delta} + \frac{1}{3} \right) \frac{\theta x}{(\sqrt{D_m} + \sqrt{2x}) \sqrt{n}} . \quad (3)$$

By taking  $\delta = \sqrt{\theta x} (D_m + \sqrt{2x D_m})^{-1/2}$ , the right-hand side of (3) becomes

$$\begin{aligned} & \sqrt{\frac{D_m}{n}} + \sqrt{\frac{2x}{n}} + 2\sqrt{\frac{\theta x}{n}} \frac{\sqrt{D_m}}{\sqrt{D_m} + \sqrt{2x D_m}} + \frac{\theta x}{3(\sqrt{D_m} + \sqrt{2x}) \sqrt{n}} \\ & \leq \sqrt{\frac{D_m}{n}} + \sqrt{\frac{2x}{n}} + 2\sqrt{\frac{\theta}{n}} \left( \sqrt{x} \wedge \left( \frac{x D_m}{2} \right)^{1/4} \right) + \frac{\theta}{3} \sqrt{\frac{x}{n}} \left( \sqrt{\frac{x}{D_m}} \wedge \frac{1}{\sqrt{2}} \right) , \end{aligned}$$

which gives (20). Inequality (19) is a direct consequence of (20). ■

**Proof of Proposition IV.2.** Let us set for any  $I \in m$ ,  $\varphi_I = (P(I))^{-1/2} \mathbb{1}_I$ . The family of functions  $(\varphi_I)_{I \in m}$  forms an orthonormal basis of  $(m, \|\cdot\|_2)$ . By Cauchy-Schwarz inequality, we have

$$\chi_n(m) = \sup_{(a_I)_{I \in m} \in B_2} \left| (P_n - P) \left( \sum_{I \in m} a_I \varphi_I \right) \right| ,$$

where  $B_2 = \{(a_I)_{I \in m} ; \sum_{I \in m} a_I^2 \leq 1\}$ . As

$$\sigma_m^2 := \sup_{(a_I)_{I \in m} \in B_2} \mathbb{V} \left( \sum_{I \in m} a_I \varphi_I(\xi_1) \right) \leq 1$$

and

$$b_m = \sup_{(a_I)_{I \in m} \in B_2} \left\| \sum_{I \in m} a_I \varphi_I \right\|_{\infty} \leq \sup_{I \in m} \|\varphi_I\|_{\infty} = \sup_{I \in m} \frac{1}{\sqrt{P(I)}} ,$$

we get by Klein-Rio's inequality (see [5]), for every  $x, \delta > 0$ ,

$$\mathbb{P} \left( \chi_n(m) \leq (1 - \delta) \mathbb{E}[\chi_n(m)] - \sqrt{\frac{2x}{n}} - \left( \frac{1}{\delta} + 1 \right) \frac{b_m x}{n} \right) \leq \exp(-x) . \quad (4)$$

Note that we have  $b_m \leq \sqrt{D_m A_{\Lambda}^{-1}}$ . Now, we bound  $\mathbb{E}[\chi_n(m)]$  by below. By Theorem 11.10 in [2] applied to  $\chi_n(m)$ , we get, for any  $\zeta > 0$ ,

$$\begin{aligned} \mathbb{E}[\chi_n^2(m)] - \mathbb{E}^2[\chi_n(m)] & \leq \frac{\sigma_m^2}{n} + 4 \frac{b_m}{n} \mathbb{E}[\chi_n(m)] \\ & \leq \frac{1}{n} + 4 \frac{\sqrt{D_m A_{\Lambda}^{-1}}}{n} \mathbb{E}[\chi_n(m)] \\ & \leq \frac{1}{n} + 4\zeta^{-1} \frac{D_m A_{\Lambda}^{-1}}{n^2} + \zeta \mathbb{E}^2[\chi_n(m)] . \end{aligned} \quad (5)$$

The latter inequality results from  $2ab \leq \zeta a^2 + \zeta^{-1} b^2$  applied with  $a = \mathbb{E}[\chi_n(m)]$  and  $b = 2\sqrt{D_m A_{\Lambda}^{-1}}/n$ . As  $\mathbb{E}[\chi_n^2(m)] = D_m/n$ , (5) applied with  $\zeta = n^{-1/2}$  gives

$$\begin{aligned} \mathbb{E}[\chi_n(m)] & \geq \sqrt{\frac{1}{1 + n^{-1/2}} \left( \frac{D_m}{n} - 4A_{\Lambda}^{-1} \frac{D_m}{n^{3/2}} - \frac{1}{n} \right)}_+ \\ & \geq \sqrt{\frac{D_m}{n}} \left( 1 - L_{A_{\Lambda}} D_m^{-1/2} \vee n^{-1/4} \right) . \end{aligned} \quad (6)$$

Hence, by using (6) and taking  $x = \alpha \ln(n+1)$  and  $\delta = n^{-1/4} \sqrt{\ln(n+1)}$  in (4), we obtain with probability at least  $1 - (n+1)^{-\alpha}$ ,

$$\begin{aligned}
& \chi_n(m) \\
& \geq \left(1 - \frac{\sqrt{\ln(n+1)}}{n^{1/4}}\right) \sqrt{\frac{D_m}{n}} \left(1 - L_{A_\Lambda} \frac{1}{\sqrt{D_m}} \vee \frac{1}{n^{1/4}}\right) \\
& \quad - \sqrt{\frac{2\alpha \ln(n+1)}{n}} - \left(\frac{n^{1/4}}{\sqrt{\ln(n+1)}} + 1\right) \frac{\alpha \sqrt{D_m} \ln(n+1)}{\sqrt{A_\Lambda} n} \\
& \geq \sqrt{\frac{D_m}{n}} \left(1 - L_{A_\Lambda} \frac{\sqrt{\ln(n+1)}}{n^{1/4}} \vee \frac{1}{\sqrt{D_m}} - \sqrt{\frac{2\alpha \ln(n+1)}{D_m}} - L_{A_\Lambda, \alpha} \frac{\sqrt{\ln(n+1)}}{n^{1/4}}\right) \\
& \geq \sqrt{\frac{D_m}{n}} \left(1 - L_{A_\Lambda, \alpha} \frac{\sqrt{\ln(n+1)}}{n^{1/4}} \vee \sqrt{\frac{\ln(n+1)}{D_m}}\right),
\end{aligned}$$

which concludes the proof. ■

## 1.2 Proofs related to Section IV-B

Concentration inequalities for the centered empirical bias are provided in Section IV-B of the main part of the article and we give below the proofs of these results.

The results in this section are based on the Cramèr-Chernoff method (see [2] for instance). Let us recall that if we set  $S := \sum_{i=1}^n X_i - \mathbb{E}[X_i]$ , where  $(X_i)_{i=1}^n$  are  $n$  i.i.d. real random variables, and for any  $\lambda \geq 0$  and  $y \in \mathbb{R}_+$ ,

$$\varphi_S(\lambda) := \ln(\mathbb{E}[\exp(\lambda S)]) = n(\ln(\mathbb{E}[\exp(\lambda X_1)]) - \lambda \mathbb{E}[X_1])$$

and

$$\varphi_S^*(y) := \sup_{\lambda \in \mathbb{R}_+} \{\lambda y - \varphi_S(\lambda)\},$$

then

$$\mathbb{P}(S \geq y) \leq \exp(-\varphi_S^*(y)). \quad (7)$$

**Proof of Proposition IV.3.** We first prove concentration inequality (22). We set  $X_i := \ln(f/f_*)(\xi_i)$  and use Inequality (7). For  $\lambda \in [0, 1]$ , as  $\mathbb{E}[X_1] = -\mathcal{K}(f_*, f)$ , we have

$$\begin{aligned}
\varphi_S(\lambda)/n &= \ln\left(P\left[\left(\frac{f}{f_*}\right)^\lambda\right]\right) + \lambda \mathcal{K}(f_*, f) \\
&\leq \lambda \ln(P[f/f_*]) + \lambda \mathcal{K}(f_*, f) = \lambda \mathcal{K}(f_*, f),
\end{aligned}$$

where the inequality derives from the concavity of the function  $x \mapsto x^\lambda$ . By setting  $\mathcal{K} = \mathcal{K}(f_*, f)$ , we thus get

$$\varphi_S^*(y) \geq \sup_{\lambda \in [0, 1]} \{\lambda y - \varphi_S(\lambda)\} \geq (y - n\mathcal{K})_+,$$

which implies,

$$\mathbb{P}((P_n - P)(\ln(f/f_*)) \geq x) \leq \exp(-n(x - \mathcal{K})_+). \quad (8)$$

Inequality (22) is a direct consequence of (8). Moreover, we notice that for any  $u \in \mathbb{R}$ ,  $\exp(u) \leq 1 + u + \frac{u^2}{2} \exp(u_+)$  and  $\ln(1 + u) \leq u$ , where  $u_+ = u \vee 0$ . By consequence, for  $\lambda \in [0, 1]$ , it holds

$$\begin{aligned}
\varphi_S(\lambda)/n &= \ln(\mathbb{E}[\exp(\lambda X_1)]) - \lambda \mathbb{E}[X_1] \\
&\leq \ln\left(1 + \lambda \mathbb{E}[X_1] + \frac{\lambda^2}{2} \mathbb{E}[Y_1^2 \exp(\lambda (X_1)_+)]\right) - \lambda \mathbb{E}[Y_1] \\
&\leq \frac{\lambda^2}{2} \mathbb{E}[X_1^2 \exp(\lambda (X_1)_+)] \leq \frac{\lambda^2}{2} \mathbb{E}[X_1^2 \exp((X_1)_+)] \leq \frac{\lambda^2}{2} v.
\end{aligned}$$

Now, we get, for any  $y \geq 0$ ,

$$\begin{aligned} \varphi_S^*(y) &\geq \sup_{\lambda \in [0,1]} \{\lambda y - \varphi_S(\lambda)\} \\ &\geq \sup_{\lambda \in [0,1]} \left\{ \lambda y - n \frac{\lambda^2 v}{2} \right\} = \left( \frac{y^2}{2nv} \mathbb{1}_{y \leq nv} + \left( y - \frac{nv}{2} \right) \mathbb{1}_{y > nv} \right). \end{aligned} \quad (9)$$

So, by using (7) with (9) taken with  $x = y/n$ , it holds

$$\mathbb{P}((P_n - P)(\ln(f/f_*)) \geq x) \leq \exp\left(-n \left( \frac{x^2}{2v} \mathbb{1}_{x \leq v} + \left( x - \frac{v}{2} \right) \mathbb{1}_{x > v} \right)\right). \quad (10)$$

To obtain (23), we notice that Inequality (10) implies by simple calculations, for any  $z \geq 0$ ,

$$\mathbb{P}\left((P_n - P)(\ln(f/f_*)) \geq \sqrt{\frac{2vz}{n}} \mathbb{1}_{z \leq nv/2} + \left( \frac{z}{n} + \frac{v}{2} \right) \mathbb{1}_{z > nv/2}\right) \leq \exp(-z).$$

To conclude the proof, it suffices to remark that

$$\sqrt{2vz/n} \mathbb{1}_{z \leq nv/2} + (z/n + v/2) \mathbb{1}_{z > nv/2} \leq \sqrt{2vz/n} + 2z/n.$$

■

**Proof of Proposition IV.4.** Let us first prove the inequality of concentration (24). We set  $Y_i := \ln(f_*/f)(\xi_i)$  and use (7) with  $S = \sum_{i=1}^n Y_i - \mathbb{E}[Y_i]$ . For  $\lambda \in [0, r]$ , we have by Hölder's inequality,  $P[(f_*/f)^\lambda] \leq P[(f_*/f)^r]^{\lambda/r}$ . Hence,

$$\begin{aligned} \varphi_S(\lambda)/n &= \ln\left(P[(f_*/f)^\lambda]\right) - \lambda \mathcal{K}(f_*, f) \\ &\leq \lambda \left( \frac{1}{r} \ln(P[(f_*/f)^r]) - \mathcal{K}(f_*, f) \right). \end{aligned}$$

Let us notice that by concavity of  $\ln$ , we have  $\frac{1}{r} \ln(P[(f_*/f)^r]) - \mathcal{K}(f_*, f) \geq 0$ . Now we get, for any  $y \geq 0$ ,

$$\begin{aligned} \varphi_S^*(y) &\geq \sup_{\lambda \in [0,r]} \{\lambda y - \varphi_S(\lambda)\} \\ &\geq \sup_{\lambda \in [0,r]} \left\{ \lambda \left( y - n \left( \frac{1}{r} \ln(P[(f_*/f)^r]) - \mathcal{K}(f_*, f) \right) \right) \right\} \\ &= (ry - n(\ln(P[(f_*/f)^r]) - r\mathcal{K}(f_*, f)))_+. \end{aligned}$$

Using (7), we obtain, for any  $x \geq 0$ ,

$$\mathbb{P}((P_n - P)(\ln(f/f_*)) \leq -x) \leq \exp(-n(rx - \ln[(f_*/f)^r] + r\mathcal{K}(f_*, f))_+). \quad (11)$$

Inequality (24) is a straightforward consequence of (11). As in the proof of Proposition IV.3, we notice that for any  $u \in \mathbb{R}$ ,  $\exp(u) \leq 1 + u + \frac{u^2}{2} \exp(u_+)$  and  $\ln(1 + u) \leq u$ , where  $u_+ = u \vee 0$ . By consequence, for  $\lambda \in [0, r]$ , it holds

$$\begin{aligned} \varphi_S(\lambda)/n &= -\lambda \mathbb{E}[Y_1] + \ln(\mathbb{E}[\exp(\lambda Y_1)]) \\ &\leq \ln\left(1 + \lambda \mathbb{E}[Y_1] + \frac{\lambda^2}{2} \mathbb{E}[Y_1^2 \exp(\lambda (Y_1)_+)]\right) - \lambda \mathbb{E}[Y_1] \\ &\leq \frac{\lambda^2}{2} \mathbb{E}[Y_1^2 \exp(\lambda (Y_1)_+)] \leq \frac{\lambda^2}{2} \mathbb{E}[Y_1^2 \exp(r (Y_1)_+)] \leq \frac{\lambda^2}{2} w_r. \end{aligned}$$

Now we get, for any  $y \geq 0$ ,

$$\begin{aligned} \varphi_S^*(y) &\geq \sup_{\lambda \in [0, r]} \{ \lambda y - \varphi_S(\lambda) \} \\ &\geq \sup_{\lambda \in [0, r]} \left\{ \lambda y - \frac{n\lambda^2 w_r}{2} \right\} = \frac{y^2}{2nw_r} \mathbb{1}_{y \leq rnw_r} + r \left( y - \frac{rnw_r}{2} \right) \mathbb{1}_{y > rnw_r}, \end{aligned}$$

which gives

$$\mathbb{P} \left( (P_n - P)(\ln(f/f_*)) \leq -x \right) \leq \exp \left( -n \left( \frac{x^2}{2nw_r} \mathbb{1}_{x \leq rnw_r} + r \left( x - \frac{rnw_r}{2} \right) \mathbb{1}_{x > rnw_r} \right) \right) \quad (12)$$

Inequality (25) is again a consequence of (12), by the same kind of arguments as those involved in the proof of (23) in Lemma IV.3. ■

### 1.3 Proofs related to Section IV-C

We provide here the proofs of the margin-like relations that stated in Section IV-C of the main part of the article.

**Proof of Proposition IV.5.** Let us take  $q > 1$  such that  $1/p + 1/q = 1$ . It holds

$$\begin{aligned} P \left[ \left( \frac{f}{f_*} \vee 1 \right) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \right] &= \int (f \vee f_*) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu \\ &= \int \frac{f_* \vee f}{f_* \wedge f} \left( (f_* \wedge f) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \right)^{\frac{1}{p} + \frac{1}{q}} d\mu \\ &= \int \left( \frac{f_* \vee f}{(f_* \wedge f)^{\frac{1}{q}}} \left| \ln \left( \frac{f}{f_*} \right) \right|^{\frac{2}{p}} \right) \left( (f_* \wedge f) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \right)^{\frac{1}{q}} d\mu \\ &\leq \left( \int (f_* \wedge f) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu \right)^{\frac{1}{q}} \underbrace{\left( \int \frac{(f_* \vee f)^p}{(f_* \wedge f)^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu \right)^{\frac{1}{p}}}_{:=I} \end{aligned}$$

where in the last step we used Hölder's inequality. Now, by [6, Lemma 7.24], it also holds

$$\frac{1}{2} \int (f_* \wedge f) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu \leq \mathcal{K}(f_*, f).$$

In order to prove (27), it thus remains to bound  $I$  in terms of  $p, c_+$  and  $c_-$  only. First, we decompose  $I$  into two parts,

$$\begin{aligned} \int \frac{(f_* \vee f)^p}{(f_* \wedge f)^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu &= \int \frac{f_*^p}{f^{p-1}} \left( \ln \left( \frac{f_*}{f} \right) \right)^2 \mathbb{1}_{f_* \geq f} d\mu \\ &\quad + \int \frac{f^p}{f_*^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \mathbb{1}_{f \geq f_*} d\mu. \end{aligned} \quad (13)$$

For the first term in the right-hand side of (13), we get

$$\begin{aligned} \int \frac{f_*^p}{f^{p-1}} \left( \ln \left( \frac{f_*}{f} \right) \right)^2 \mathbb{1}_{f_* \geq f} d\mu &\leq c_-^{1-p} \int f_*^p \left( \ln \left( \frac{f_*}{c_-} \right) \right)^2 d\mu \\ &\leq 4c_-^{1-p} \left( (\ln c_-)^2 \vee 1 \right) \int f_*^p \left( (\ln f_*)^2 \vee 1 \right) d\mu < +\infty, \end{aligned} \quad (14)$$

where in the second inequality we used the following fact:  $(a + b)^2 \leq 4(a^2 \vee 1)(b^2 \vee 1)$ , for any real numbers  $a$  and  $b$ . The finiteness of the upper bound is guaranteed by Assumption (26). For the second term in the right-hand side of (13), it holds by same kind of arguments that lead to (14),

$$\begin{aligned} \int \frac{f^p}{f_*^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \mathbb{1}_{f \geq f_*} d\mu &\leq c_+^p \int \frac{1}{f_*^{p-1}} \left( \ln \left( \frac{c_+}{f_*} \right) \right)^2 d\mu \\ &\leq 4c_+^p \left( (\ln c_+)^2 \vee 1 \right) \int f_*^{1-p} \left( (\ln f_*)^2 \vee 1 \right) d\mu < +\infty, \end{aligned} \quad (15)$$

where in the last inequality we used the following fact:  $(a - b)^2 \leq 4(a^2 \vee 1)(b^2 \vee 1)$ . Again, the finiteness of the upper bound is guaranteed by Assumption (26). Inequality (27) then follows from combining (13), (14) and (15).

Inequality (28) follows from the same kind of computations. Indeed, we have by the use of Hölder's inequality,

$$\begin{aligned} P \left[ \left( \frac{f_*}{f} \vee 1 \right)^r \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \right] &= \int \left( \frac{f_*^{r+1}}{f^r} \vee f_* \right) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu \\ &= \int \left( \frac{f_*^{r+1} \vee f_* f^r}{f^r (f_* \wedge f)^{1-\frac{r+1}{p}}} \left| \ln \left( \frac{f}{f_*} \right) \right|^{\frac{2(r+1)}{p}} \right) \left( (f_* \wedge f) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \right)^{1-\frac{r+1}{p}} d\mu \\ &\leq \left( \int (f_* \wedge f) \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu \right)^{1-\frac{r+1}{p}} \left( \underbrace{\int \frac{f_*^p \vee f_*^{\frac{p}{r+1}} f^{\frac{rp}{r+1}}}{f^{\frac{rp}{r+1}} f_*^{\frac{p}{r+1}-1} \wedge f^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu}_{:=I_r} \right)^{\frac{r+1}{p}}. \end{aligned}$$

In order to prove (28), it thus remains to bound  $I_r$  in terms of  $p, c_+$  and  $c_-$  only. Again, we split  $I_r$  into two parts,

$$\begin{aligned} \int \frac{f_*^p \vee f_*^{\frac{p}{r+1}} f^{\frac{rp}{r+1}}}{f^{\frac{rp}{r+1}} f_*^{\frac{p}{r+1}-1} \wedge f^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 d\mu &= \int \frac{f_*^p}{f^{p-1}} \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \mathbb{1}_{f_* \geq f} d\mu \\ &\quad + \int f_* \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \mathbb{1}_{f \geq f_*} d\mu. \end{aligned} \quad (16)$$

The first term in the right-hand side of (16) is given by (14) above. For the second term in the right-hand side of (16), it holds

$$\begin{aligned} \int f_* \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \mathbb{1}_{f \geq f_*} d\mu &\leq \int f_* \left( \ln \left( \frac{c_+}{f_*} \right) \right)^2 d\mu \\ &\leq 2(\ln(c_+))^2 + 2P \left( (\ln f_*)^2 \right). \end{aligned} \quad (17)$$

Furthermore we have  $f_*^{p-1} + f_*^{-p} \geq 1$ , so  $P \left( (\ln f_*)^2 \right) \leq P \left( f_*^{p-1} (\ln f_*)^2 \right) + P \left( f_*^{-p} (\ln f_*)^2 \right) \leq J + Q$  and by (17),

$$\int f_* \left( \ln \left( \frac{f}{f_*} \right) \right)^2 \mathbb{1}_{f \geq f_*} d\mu \leq 2 \left( (\ln(c_+))^2 + J + Q \right) < +\infty,$$

where the finiteness of the upper bound comes from Assumption (26). Inequality (28) then easily follows. ■

**Proof of Proposition IV.6.** Let us first prove Inequality (29). Considering the proof of Inequality (27) of proposition IV.5 of the main part of the article, we see that it is sufficient to bound the second term in the right-hand side of (13), applied with  $f = f_m$ , in terms of  $A_{\min}, J$  and  $p$  only. It holds

$$\int \frac{f_m}{f_*^{p-1}} \left( \ln \left( \frac{f_m}{f_*} \right) \right)^2 \mathbb{1}_{f_m \geq f_*} d\mu \leq A_{\min} \int \left( \frac{f_m}{A_{\min}} \right)^p \left( \ln \left( \frac{f_m}{A_{\min}} \right) \right)^2 d\mu. \quad (18)$$

Now, we set  $h$  an auxiliary function, defined by  $h(x) = x^p (\ln x)^2$  for any  $x \geq 1$ . It is easily seen that  $h$  is convex. As the function  $f_m$  is piecewise constant relatively to the partition  $\Lambda_m$ , the notation  $f_m(I) = f_m(x)$  for any  $x \in I$  and  $I \in \Lambda_m$  is consistent. Using this notation it holds, for any  $I \in \Lambda_m$ ,

$$h\left(\frac{f_m(I)}{A_{\min}}\right) = h\left(\int_I \frac{f_*}{A_{\min}} \frac{d\mu}{\mu(I)}\right) \leq \int_I h\left(\frac{f_*}{A_{\min}}\right) \frac{d\mu}{\mu(I)}.$$

From the latter inequality and from (18), we deduce

$$\begin{aligned} \int \frac{f_m}{f_*^{p-1}} \left(\ln\left(\frac{f_m}{f_*}\right)\right)^2 \mathbb{1}_{f_m \geq f_*} d\mu &\leq A_{\min} \int h\left(\frac{f_m}{A_{\min}}\right) d\mu \\ &\leq A_{\min} \int h\left(\frac{f_*}{A_{\min}}\right) d\mu \\ &\leq 4A_{\min}^{1-p} \left((\ln A_{\min})^2 \vee 1\right) J. \end{aligned}$$

Inequality (29) is thus proved.

In the same manner, to establish Inequality (30) it suffices to adapt the proof of Inequality (28) given above by controlling the second term in the right-hand side of (16), applied with  $f = f_m$ , in terms of  $A_{\min}, p$  and  $P(\ln f_*)^2$  only. Let us notice that the function  $f$  defined on  $[1, +\infty)$  by  $f(x) = x(\ln x)^2$  is convex. We have

$$\int f_* \left(\ln\left(\frac{f_m}{f_*}\right)\right)^2 \mathbb{1}_{f_m \geq f_*} d\mu \leq \int f_m \left(\ln\left(\frac{f_m}{A_{\min}}\right)\right)^2 d\mu = A_{\min} \int f\left(\frac{f_m}{A_{\min}}\right) d\mu.$$

Now, for any  $I \in \Lambda_m$ , it holds  $f\left(\frac{f_m(I)}{A_{\min}}\right) = f\left(\int_I \frac{f_*}{A_{\min}} \frac{d\mu}{\mu(I)}\right) \leq \int_I f\left(\frac{f_*}{A_{\min}}\right) \frac{d\mu}{\mu(I)}$ . Hence,

$$\int f_* \left(\ln\left(\frac{f_m}{f_*}\right)\right)^2 \mathbb{1}_{f_m \geq f_*} d\mu \leq A_{\min} \int f\left(\frac{f_*}{A_{\min}}\right) d\mu \leq 2P(\ln f_*)^2 + 2(\ln A_{\min})^2,$$

which gives the desired upper-bound and proves (30). In the event that  $f_* \in L_\infty(\mu)$ , we have to prove (31).

We have  $\inf_{z \in \mathcal{Z}} f_*(z) \leq \inf_{z \in \mathcal{Z}} f_m(z) \leq \|f_m\|_\infty \leq \|f_*\|_\infty$ , so it holds

$$P\left[\left(\frac{f_m}{f_*} \vee 1\right) \left(\ln\left(\frac{f_m}{f_*}\right)\right)^2\right] \vee P\left[\left(\frac{f_*}{f_m} \vee 1\right)^r \left(\ln\left(\frac{f_m}{f_*}\right)\right)^2\right] \leq \left(\frac{\|f_*\|_\infty}{A_{\min}}\right)^{r \vee 1} P\left[\left(\ln\left(\frac{f_m}{f_*}\right)\right)^2\right].$$

Now, Inequality (31) is a direct consequence of Lemma 1 of Barron and Sheu [1], which contains the following inequality,

$$P\left[\left(\ln\left(\frac{f_m}{f_*}\right)\right)^2\right] \leq 2 \exp\left(\left\|\ln\left(\frac{f_m}{f_*}\right)\right\|_\infty\right) \mathcal{K}(f_*, f_m).$$

This finishes the proof of Proposition Proposition IV.6. ■

## 2 Further Proofs and Theoretical Results

### 2.1 Proofs Related to Section III-A

Most of the arguments given in the proofs of this section are borrowed from Castellan [4]. We essentially rearrange these arguments in a more efficient way, thus obtaining better concentration bounds than in [4] (or [6]).

We also set, for any  $\varepsilon > 0$ , the event  $\Omega_m(\varepsilon)$  where some control of  $\hat{f}_m$  in sup-norm is achieved,

$$\Omega_m(\varepsilon) = \left\{ \left\| \frac{\hat{f}_m - f_m}{f_m} \right\|_\infty \leq \varepsilon \right\}.$$

As we have the following formulas for the estimators and the projections of the target,

$$\hat{f}_m = \sum_{I \in \Lambda_m} \frac{P_n(I)}{\mu(I)} \mathbb{1}_I \quad , \quad f_m = \sum_{I \in \Lambda_m} \frac{P(I)}{\mu(I)} \mathbb{1}_I \quad ,$$

we deduce that,

$$\left\| \frac{\hat{f}_m - f_m}{f_m} \right\|_{\infty} = \sup_{I \in \Lambda_m} \frac{|(P_n - P)(I)|}{P(I)} . \quad (19)$$

Hence, it holds  $\Omega_m(\varepsilon) = \bigcap_{I \in m} \{|P_n(I) - P(I)| \leq \varepsilon P(I)\}$ .

Before giving the proof of Theorem III.1, the following lemma will be useful. It describes the consistency in sup-norm of the histogram estimators, suitably normalized by the projections of the target on each model.

**Lemma 2.1** *Let  $\alpha, A_+$  and  $A_{\Lambda}$  be positive constants. Consider a finite partition  $m$  of  $\mathcal{Z}$ , with cardinality  $D_m$ . Assume*

$$0 < A_{\Lambda} \leq D_m \inf_{I \in m} \{P(I)\} \quad \text{and} \quad 0 < D_m \leq A_+ \frac{n}{\ln(n+1)} \leq n . \quad (20)$$

Then by setting

$$R_n^{\infty}(m) = \sqrt{\frac{2(\alpha+1)D_m \ln(n+1)}{A_{\Lambda} n}} + \frac{(\alpha+1)D_m \ln(n+1)}{3A_{\Lambda} n} , \quad (21)$$

we get

$$\mathbb{P} \left( \left\| \frac{\hat{f}_m - f_m}{f_m} \right\|_{\infty} \leq R_n^{\infty}(m) \right) \geq 1 - 2(n+1)^{-\alpha} . \quad (22)$$

In other words,  $\mathbb{P}(\Omega_m(R_n^{\infty}(m))) \geq 1 - 2(n+1)^{-\alpha}$ . In addition, there exists a positive constant  $A_c$ , only depending on  $\alpha, A_+$  and  $A_{\Lambda}$ , such that  $R_n^{\infty}(m) \leq A_c \sqrt{\frac{D_m \ln(n+1)}{n}}$ . Furthermore, if

$$\frac{(\alpha+1)A_+}{A_{\Lambda}} \leq \tau = \sqrt{\sqrt{6} - 3/\sqrt{2}} < 0.57 ,$$

or if  $n \geq n_0(\alpha, A_+, A_{\Lambda})$ , then  $R_n^{\infty}(m) \leq 1/2$ .

**Proof of Lemma 2.1.** Let  $\beta > 0$  to be fixed later. Recall that, by (19) we have

$$\left\| \frac{\hat{f}_m - f_m}{f_m} \right\|_{\infty} = \sup_{I \in \Lambda_m} \frac{|(P_n - P)(I)|}{P(I)} . \quad (23)$$

By Bernstein's inequality (see Proposition 2.9 in [6]) applied to the random variables  $\mathbb{1}_{\xi_i \in I}$  we get, for all  $x > 0$ ,

$$\mathbb{P} \left[ |(P_n - P)(I)| \geq \sqrt{\frac{2P(I)x}{n}} + \frac{x}{3n} \right] \leq 2 \exp(-x) .$$

Taking  $x = \beta \ln(n+1)$  and normalizing by the quantity  $P(I) > 0$  we get

$$\mathbb{P} \left[ \frac{|(P_n - P)(I)|}{P(I)} \geq \sqrt{\frac{2\beta \ln(n+1)}{P(I)n}} + \frac{\beta \ln(n+1)}{P(I)3n} \right] \leq 2(n+1)^{-\beta} . \quad (24)$$

Now, by the first inequality in (20), we have  $0 < P(I)^{-1} \leq A_{\Lambda}^{-1} D_m$ . Hence, using (24) we get

$$\mathbb{P} \left[ \frac{|(P_n - P)(I)|}{P(I)} \geq \sqrt{\frac{2\beta D_m \ln(n+1)}{A_{\Lambda} n}} + \frac{\beta D_m \ln(n+1)}{3A_{\Lambda} n} \right] \leq 2(n+1)^{-\beta} , \quad (25)$$



We then deduce from (23) and (25) that

$$\mathbb{P} \left[ \left\| \frac{\hat{f}_m - f_m}{f_m} \right\|_{\infty} \geq R_n^{\infty}(m) \right] \leq \frac{2D_m}{(n+1)^{\beta}}$$

and, since  $D_m \leq n$ , taking  $\beta = \alpha + 1$  yields Inequality (22). The other facts of Lemma 2.1 then follow from simple computations. In particular, the value of  $\tau$  is fixed to be the largest value satisfying the following inequality,  $\sqrt{2\tau} + \tau/3 \leq 1/2$ . ■

**Proof of Theorem III.1.** Recall that  $\alpha$  is fixed. By Inequality (19) in Proposition IV.1 applied with  $\theta = R_n^{\infty}(m)$ —where  $R_n^{\infty}(m)$  is defined in (21) with our fixed value of  $\alpha$ —and  $x = \alpha \ln(n+1)$ , it holds with probability at least  $1 - (n+1)^{-\alpha}$ ,

$$\chi_n(m) \mathbf{1}_{\Omega_m(R_n^{\infty}(m))} \leq \sqrt{\frac{D_m}{n}} + \left( 1 + \sqrt{2R_n^{\infty}(m)} + \frac{R_n^{\infty}(m)}{6} \right) \sqrt{\frac{2\alpha \ln(n+1)}{n}}. \quad (26)$$

As  $R_n^{\infty}(m) \leq L_{\alpha, A_+, A_{\Lambda}} \sqrt{D_m \ln(n+1)/n} \leq L_{\alpha, A_+, A_{\Lambda}}$ , (26) gives

$$\begin{aligned} \chi_n(m) \mathbf{1}_{\Omega_m(R_n^{\infty}(m))} &\leq \sqrt{\frac{D_m}{n}} + L_{\alpha, A_+, A_{\Lambda}} \sqrt{\frac{\ln(n+1)}{n}} \\ &= \sqrt{\frac{D_m}{n}} \left( 1 + L_{\alpha, A_+, A_{\Lambda}} \sqrt{\frac{\ln(n+1)}{D_m}} \right). \end{aligned} \quad (27)$$

We set the event  $\Omega_0$  on which we have

$$\begin{aligned} \left\| \frac{\hat{f}_m - f_m}{f_m} \right\|_{\infty} &\leq R_n^{\infty}(m), \\ \chi_n(m) &\leq \sqrt{\frac{D_m}{n}} \left( 1 + L_{\alpha, A_+, A_{\Lambda}} \sqrt{\frac{\ln(n+1)}{D_m}} \right), \end{aligned} \quad (28)$$

and

$$\chi_n(m) \geq \left( 1 - A_g \left( \sqrt{\frac{\ln(n+1)}{D_m}} \vee \frac{\sqrt{\ln(n+1)}}{n^{1/4}} \right) \right) \sqrt{\frac{D_m}{n}}.$$

In particular,  $\Omega_0 \subset \Omega_m(R_n^{\infty}(m))$ . By (27), Lemma 2.1 and Proposition IV.2, it holds  $\mathbb{P}(\Omega_0) \geq 1 - 4(n+1)^{-\alpha}$ . It suffices to prove the inequalities of Theorem III.1 on  $\Omega_0$ . The following inequalities, between the excess risk on  $m$  and the chi-square statistics, are shown in [4] (Inequalities (2.13)). For any  $\varepsilon \in (0, 1)$ , on  $\Omega_m(\varepsilon)$ ,

$$\frac{1 - \varepsilon}{2(1 + \varepsilon)^2} \chi_n^2(m) \leq \mathcal{K}(f_m, \hat{f}_m) \leq \frac{1 + \varepsilon}{2(1 - \varepsilon)^2} \chi_n^2(m). \quad (29)$$

Under the condition  $(\alpha + 1) A_+ A_{\Lambda}^{-1} < \tau$ , Lemma 2.1 gives  $R_n^{\infty}(m) \leq 1/2$ . Hence, by applying the right-hand side of (29) with  $\varepsilon = R_n^{\infty}(m) \leq 1/2$ , using (28) and the fact that  $(1 - \varepsilon)^{-1} \leq 1 + 2\varepsilon$  for  $\varepsilon \leq 1/2$ , we get on  $\Omega_0$ ,

$$\begin{aligned} \mathcal{K}(f_m, \hat{f}_m) &\leq \frac{1 + R_n^{\infty}(m)}{2(1 - R_n^{\infty}(m))^2} \chi_n^2(m) \\ &\leq \left( \frac{1}{2} + L_{\alpha, A_+, A_{\Lambda}} \sqrt{\frac{D_m \ln(n+1)}{n}} \right) \frac{D_m}{n} \left( 1 + L_{\alpha, A_+, A_{\Lambda}} \sqrt{\frac{\ln(n+1)}{D_m}} \right)^2. \end{aligned}$$

Then simple computations allow to get the right-hand side inequality in (12)).

By applying the left-hand side of (29) with  $\varepsilon = R_n^\infty(m) \leq 1/2$  and using the fact that  $(1+\varepsilon)^{-1} \geq 1 - \varepsilon$ , we also get on  $\Omega_0$ ,

$$\begin{aligned} \mathcal{K}(f_m, \hat{f}_m) &\geq \frac{1 - R_n^\infty(m)}{2(1 + R_n^\infty(m))^2} \chi_n^2(m) \\ &\geq (1 - R_n^\infty(m))^3 \left( 1 - A_g \left( \sqrt{\frac{\ln(n+1)}{D_m}} \vee \frac{\sqrt{\ln(n+1)}}{n^{1/4}} \right) \right)^2 \frac{D_m}{2n} \\ &\geq (1 - (3R_n^\infty(m) \wedge 1)) \left( 1 - 2A_g \left( \sqrt{\frac{\ln(n+1)}{D_m}} \vee \frac{\sqrt{\ln(n+1)}}{n^{1/4}} \right) \right)^2 \frac{D_m}{2n} \end{aligned}$$

The left-hand side inequality in (12) then follows by simple computations, noticing in particular that  $n^{-1/4} \sqrt{\ln(n+1)} \leq \sqrt{\ln(n+1)/D_m} \vee \sqrt{D_m \ln(n+1)/n}$ .

Inequalities in (13) follow from the same kind of arguments as those involved in the proofs of inequalities in (12). Indeed, from [6, Lemma 7.24]—or [4, Lemma 2.3]—, it holds

$$\frac{1}{2} \int (\hat{f}_m \wedge f_m) \left( \ln \frac{\hat{f}_m}{f_m} \right)^2 d\mu \leq \mathcal{K}(\hat{f}_m, f_m) \leq \frac{1}{2} \int (\hat{f}_m \vee f_m) \left( \ln \frac{\hat{f}_m}{f_m} \right)^2 d\mu .$$

We deduce that for  $\varepsilon \in (0, 1)$ , we have on  $\Omega_m(\varepsilon)$ ,

$$\frac{1 - \varepsilon}{2} \int f_m \left( \ln \frac{\hat{f}_m}{f_m} \right)^2 d\mu \leq \mathcal{K}(\hat{f}_m, f_m) \leq \frac{1 + \varepsilon}{2} \int f_m \left( \ln \frac{\hat{f}_m}{f_m} \right)^2 d\mu . \quad (30)$$

As for every  $x > 0$ , we have  $(1 \vee x)^{-1} \leq (x - 1)^{-1} \ln x \leq (1 \wedge x)^{-1}$ , (30) leads by simple computations to the following inequalities

$$\frac{1 - \varepsilon}{2(1 + \varepsilon)^2} \chi_n^2(m) \leq \mathcal{K}(\hat{f}_m, f_m) \leq \frac{1 + \varepsilon}{2(1 - \varepsilon)^2} \chi_n^2(m) .$$

We thus have the same upper and lower bounds, in terms of the chi-square statistic  $\chi_n^2(m)$ , for the empirical excess risk as for the true excess risk. ■

## 2.2 Oracle inequalities and dimension guarantees

Using the notations of Section III-B of the main part of the article, we define the set of assumptions  $(\mathbf{SA}_0)$  to be the conjunction of assumptions  $(\mathbf{P1})$ ,  $(\mathbf{P2})$ ,  $(\mathbf{P3})$ ,  $(\mathbf{A}sm)$  and  $(\mathbf{A}lr)$ . The set of assumptions  $(\mathbf{SA})$  of Section III-B of the main part thus consists on assuming  $(\mathbf{SA}_0)$  together with  $(\mathbf{A}p)$ .

For some of the following results, we will also need the following assumptions.

$(\mathbf{A}p_u)$  The bias decreases as a power of  $D_m$ : there exist  $\beta_+ > 0$  and  $C_+ > 0$  such that

$$\mathcal{K}(f_*, f_m) \leq C_+ D_m^{-\beta_+} .$$

$(\mathbf{A}p)$  The bias decreases like a power of  $D_m$ : there exist  $\beta_- \geq \beta_+ > 0$  and  $C_+, C_- > 0$  such that

$$C_- D_m^{-\beta_-} \leq \mathcal{K}(f_*, f_m) \leq C_+ D_m^{-\beta_+} .$$

Recall that an oracle model  $m_*$  is defined as follows:

$$m_* \in \arg \min_{m \in \mathcal{M}_n} \left\{ \mathcal{K}(f_*, \hat{f}_m) \right\} .$$

We obtain in the following theorem some oracle inequalities and dimension bounds for the oracle and selected models, under a slightly less restrictive set of assumptions than for Theorem III.2 of the main part of the article. Indeed, when putting assumption on the decay of the bias of the models, we use  $(\mathbf{A}p_u)$  instead of  $(\mathbf{A}p)$ .

**Theorem 2.2** Take  $n \geq 1$  and  $r \in (0, p-1)$ . Assume that the set of assumptions  $(\mathbf{SA}_0)$  holds and that for some  $\theta \in (1/2, +\infty)$  and  $\Delta > 0$ ,

$$\text{pen}(m) = (\theta + \Delta \varepsilon_n^+(m)) \frac{D_m}{n},$$

for every model  $m \in \mathcal{M}_n$ . Then there exists an event  $\Omega_n$  of probability  $1 - (n+1)^{-2}$  and some positive constants  $A_1, A_2$  depending only on the constants defined in  $(\mathbf{SA}_0)$  such that, if  $\Delta \geq (\theta-1)_- A_1 + A_2 > 0$  then we have on  $\Omega_n$ ,

$$\begin{aligned} \mathcal{K}(f_*, \hat{f}_{\hat{m}}) &\leq \frac{1 + 2(\theta-1)_+ + L_{(\mathbf{SA}), \theta, r} (\ln(n+1))^{-1/2}}{1 - 2(\theta-1)_-} \inf_{m \in \mathcal{M}_n} \left\{ \mathcal{K}(f_*, \hat{f}_m) \right\} \\ &\quad + L_{(\mathbf{SA}), \theta, r} \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}}. \end{aligned} \quad (31)$$

Assume furthermore that Assumption  $(\mathbf{Ap}_u)$  holds. Then, for any oracle model  $m_*$ , that is

$$m_* \in \arg \min_{m \in \mathcal{M}_n} \mathcal{K}(f_*, \hat{f}_m),$$

it holds on  $\Omega_n$ ,

$$D_{\hat{m}} \leq L_{(\mathbf{SA}), \Delta, \theta, r} n^{\frac{1}{2+\beta_+ \left(1 - \frac{r+1}{p}\right)}} \sqrt{\ln(n+1)}, \quad D_{m_*} \leq L_{(\mathbf{SA})} n^{\frac{1}{1+\beta_+}}.$$

In particular, if we are in the case where  $p < \beta_+$  then Inequality (31) reduces to

$$\mathcal{K}(f_*, \hat{f}_{\hat{m}}) \leq L_{(\mathbf{SA}), \theta, r} \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}}. \quad (32)$$

Note that in Inequality (32) above, the rate is sub-optimal in a minimax sense, since a regular histogram  $m$  of dimension  $D_m \sim n^{1/(1+\beta_+)}$  achieves on  $\Omega_n$  a better rate of order  $n^{-\beta_+/(1+\beta_+)}$ . We do not know if the fact that the selected model cannot recover the rate of an oracle model in this case where  $p < \beta_+$  - which cannot happen in the bounded setting where  $p = +\infty$  - is an artefact of our proof or a real limitation of model selection by penalization.

Theorem III.2 of the main part of the article is a direct corollary of the following theorem, where we use indeed the same set of assumptions  $(\mathbf{SA})$  as in Theorem III.2 of the main part, consisting in the conjunction of  $(\mathbf{SA}_0)$  with  $(\mathbf{Ap})$ .

**Theorem 2.3** Take  $n \geq 1$  and  $r \in (0, p-1)$ . Assume that the set of assumptions  $(\mathbf{SA})$  - that is  $(\mathbf{SA}_0)$  and  $(\mathbf{Ap})$  - holds and that for some  $\theta \in (1/2, +\infty)$  and  $\Delta > 0$ ,

$$\text{pen}(m) = (\theta + \Delta \varepsilon_n^+(m)) \frac{D_m}{n},$$

for every model  $m \in \mathcal{M}_n$ . Then there exists an event  $\Omega_n$  of probability  $1 - (n+1)^{-2}$  and some positive constants  $A_1, A_2$  depending only on the constants defined in  $(\mathbf{SA}_0)$  such that, if  $\Delta \geq (\theta-1)_- A_1 + A_2 > 0$  then it holds on  $\Omega_n$ ,

$$L_{\Delta, \theta, (\mathbf{SA})}^{(1)} \frac{n^{\frac{\beta_+}{\beta_- (1+\beta_+)}}}{(\ln(n+1))^{\frac{1}{\beta_-}}} \leq D_{\hat{m}} \leq L_{\Delta, \theta, (\mathbf{SA})}^{(2)} n^{\frac{1}{2+\beta_+ \left(1 - \frac{r+1}{p}\right)}} \sqrt{\ln(n+1)},$$

$$L_{(\mathbf{SA})}^{(1)} n^{\frac{\beta_+}{(1+\beta_+) \beta_-}} \leq D_{m_*} \leq L_{(\mathbf{SA})}^{(2)} n^{\frac{1}{1+\beta_+}}$$

and

$$\begin{aligned} \mathcal{K} \left( f_*, \hat{f}_{\hat{m}} \right) &\leq \frac{1 + 2(\theta - 1)_+ + L_{(\mathbf{SA}), \theta, r} n^{-\frac{\beta_+}{(1+\beta_+)\beta_-}} \sqrt{\ln(n+1)}}{1 - 2(\theta - 1)_-} \inf_{m \in \mathcal{M}_n} \left\{ \mathcal{K} \left( f_*, \hat{f}_m \right) \right\} \\ &\quad + L_{(\mathbf{SA}), \theta, r} \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}} . \end{aligned} \quad (33)$$

Furthermore, if  $\beta_- < p(1 + \beta_+) / (1 + p + r)$  or  $p / (1 + r) > \beta_- + \beta_- / \beta_+ - 1$ , then we have on  $\Omega_n$ ,

$$\mathcal{K} \left( f_*, \hat{f}_{\hat{m}} \right) \leq \frac{1 + 2(\theta - 1)_+ + L_{(\mathbf{SA}), \theta, r} (\ln(n+1))^{-1/2}}{1 - 2(\theta - 1)_-} \inf_{m \in \mathcal{M}_n} \left\{ \mathcal{K} \left( f_*, \hat{f}_m \right) \right\} . \quad (34)$$

In order to avoid cumbersome notations in the proofs of Theorem 2.2 and Theorem 2.3, when a generic constant  $L$  or an integer  $n_0$  depends on constants defined in the sets of assumptions  $(\mathbf{SA}_0)$  or  $(\mathbf{SA})$ , we will note  $L_{(\mathbf{SA})}$  and  $n_0((\mathbf{SA}))$ . The values of these constants may change from line to line, or even within one line.

### Proof of Theorem 2.2

- Proof of oracle Inequality (31):

From the definition of the selected model  $\hat{m}$  given in ((3)) of the main part,  $\hat{m}$  minimizes

$$\text{crit}(m) := P_n \left( -\ln \hat{f}_m \right) + \text{pen}(m) ,$$

over the models  $m \in \mathcal{M}_n$ . Hence,  $\hat{m}$  also minimizes

$$\text{crit}'(m) := \text{crit}(m) - P_n(-\ln f_*) \quad (35)$$

over the collection  $\mathcal{M}_n$ . Let us write

$$\begin{aligned} \mathcal{K} \left( f_*, \hat{f}_m \right) &= P_n \left( -\ln \hat{f}_m \right) + \mathcal{K} \left( \hat{f}_m, f_m \right) + (P_n - P) \left( \gamma(f_*) - \gamma(f_m) \right) \\ &\quad + \mathcal{K} \left( f_m, \hat{f}_m \right) - P_n \left( -\ln f_* \right) . \end{aligned}$$

By setting

$$p_1(m) = \mathcal{K} \left( f_m, \hat{f}_m \right) ,$$

$$p_2(m) = \mathcal{K} \left( \hat{f}_m, f_m \right) ,$$

$$\bar{\delta}(m) = (P_n - P) \left( \ln(f_m / f_*) \right)$$

and

$$\text{pen}'_{\text{id}}(m) = p_1(m) + p_2(m) + \bar{\delta}(m) ,$$

we have

$$\mathcal{K} \left( f_*, \hat{f}_m \right) = P_n \left( \gamma(\hat{f}_m) \right) + p_1(m) + p_2(m) + \bar{\delta}(m) - P_n \left( \gamma(f_*) \right)$$

and by (35),

$$\text{crit}'(m) = \mathcal{K} \left( f_*, \hat{f}_m \right) + \left( \text{pen}(m) - \text{pen}'_{\text{id}}(m) \right) . \quad (36)$$

As  $\hat{m}$  minimizes  $\text{crit}'$  over  $\mathcal{M}_n$ , it is therefore sufficient by (36) to control  $\text{pen}(m) - \text{pen}'_{\text{id}}(m)$  in terms of the excess risk  $\mathcal{K} \left( f_*, \hat{f}_m \right)$ , for every  $m \in \mathcal{M}_n$ , in order to derive oracle inequalities. We further set

$$\mathcal{K}_m = \mathcal{K} \left( f_*, f_m \right) \quad , \quad v_m = P \left[ \left( \frac{f_m}{f_*} \vee 1 \right) \left( \ln \left( \frac{f_m}{f_*} \right) \right)^2 \right]$$

and

$$w_m = P \left[ \left( \frac{f_*}{f_m} \vee 1 \right)^r \left( \ln \left( \frac{f_m}{f_*} \right) \right)^2 \right].$$

Set  $z_n = (2 + \alpha_{\mathcal{M}}) \ln(n+1) + 2 \ln 2$ . Let  $\Omega_n$  be the event on which we have, for all models  $m \in \mathcal{M}_n$ ,

$$\bar{\delta}(m) \leq \sqrt{\frac{2v_m z_n}{n}} + \frac{2z_n}{n} \quad (37)$$

$$-\bar{\delta}(m) \leq \sqrt{\frac{2w_m z_n}{n}} + \frac{2z_n}{nr} \quad (38)$$

$$-L_{(\mathbf{SA}), \alpha \varepsilon_n^-}(m) \frac{D_m}{2n} \leq p_1(m) - \frac{D_m}{2n} \leq L_{(\mathbf{SA}), \alpha \varepsilon_n^+}(m) \frac{D_m}{2n} \quad (39)$$

$$-L_{(\mathbf{SA}), \alpha \varepsilon_n^-}(m) \frac{D_m}{2n} \leq p_2(m) - \frac{D_m}{2n} \leq L_{(\mathbf{SA}), \alpha \varepsilon_n^+}(m) \frac{D_m}{2n} \quad (40)$$

By Theorem III.1 applied with  $\alpha = 5 + \alpha_{\mathcal{M}}$  and Propositions IV.3 and IV.4 applied with  $z = z_n$ , we get

$$\begin{aligned} \mathbb{P}(\Omega_n) &\geq 1 - \sum_{m \in \mathcal{M}_n} \left[ 4(n+1)^{-5-\alpha_{\mathcal{M}}} + \frac{(n+1)^{-2-\alpha_{\mathcal{M}}}}{2} \right] \\ &= 1 - \sum_{m \in \mathcal{M}_n} (n+1)^{-2-\alpha_{\mathcal{M}}} \geq 1 - (n+1)^{-2}. \end{aligned}$$

The following simple remark will be used along the proof: for any  $m \in \mathcal{M}_n$ ,  $z_n/n \leq L_{(\mathbf{SA})} \varepsilon_n^+(m) \frac{D_m}{n}$ . Notice also that  $\varepsilon_n^-(\hat{m}) \leq \varepsilon_n^+(\hat{m})$  (see Theorem III.1).

By using (29) of the main part, (36), (37), (39) and (40), we get that on  $\Omega_n$ , for  $\Delta$  of the form  $(\theta - 1)_- L_{(\mathbf{SA})}^{(1)} + L_{(\mathbf{SA})}^{(2)}$  with  $L_{(\mathbf{SA})}^{(1)}$  and  $L_{(\mathbf{SA})}^{(2)}$  sufficiently large,

$$\begin{aligned} &\text{crit}'(\hat{m}) \\ &\geq \mathcal{K}(f_*, \hat{f}_{\hat{m}}) + \text{pen}(\hat{m}) - p_1(\hat{m}) - p_2(\hat{m}) - \sqrt{\frac{2v_{\hat{m}} z_n}{n}} - \frac{2z_n}{n} \\ &\geq \mathcal{K}(f_*, \hat{f}_{\hat{m}}) + (\theta - 1) \frac{D_{\hat{m}} - 1}{n} + (\Delta - L_{(\mathbf{SA})}) \varepsilon_n^+(\hat{m}) \frac{D_{\hat{m}}}{n} - \sqrt{\frac{2A_{MR, -} \mathcal{K}_{\hat{m}}^{1-1/p} z_n}{n}} \\ &\geq \mathcal{K}(f_*, \hat{f}_{\hat{m}}) - 2(\theta - 1)_- \left( \frac{D_{\hat{m}} - 1}{2n} - L_{(\mathbf{SA})} \varepsilon_n^-(\hat{m}) \frac{D_{\hat{m}}}{n} \right) \\ &\quad + \left( \Delta - \left( L_{(\mathbf{SA})}^{(1)} (\theta - 1)_- + L_{(\mathbf{SA})}^{(2)} \right) \right) \varepsilon_n^+(\hat{m}) \frac{D_{\hat{m}}}{n} - \sqrt{\frac{2A_{MR, -} \mathcal{K}_{\hat{m}}^{1-1/p} z_n}{n}} \\ &\geq (1 - 2(\theta - 1)_-) \mathcal{K}(f_*, \hat{f}_{\hat{m}}) - \sqrt{\frac{2A_{MR, -} \mathcal{K}_{\hat{m}}^{1-1/p} z_n}{n}}. \end{aligned} \quad (41)$$

Note that  $1 - 2(\theta - 1)_- > 0$ . Let us take  $\eta \in (0, 1/2 - (\theta - 1)_-)$ , so that

$$1 - 2(\theta - 1)_- - \eta > 1/2 - (\theta - 1)_- > 0.$$

By Lemma 2.6 given at the end of this supplementary material, applied with  $a = (\eta \mathcal{K}_{\hat{m}})^{\frac{p-1}{2p}}$ ,  $b = \eta^{-\frac{p-1}{2p}} \sqrt{2A_{MR, -} z_n/n}$ ,  $u = \frac{2p}{p-1}$  and  $v = \frac{2p}{p+1}$ , we have

$$\sqrt{\frac{2A_{MR, -} \mathcal{K}_{\hat{m}}^{1-1/p} z_n}{n}} \leq \eta \mathcal{K}_{\hat{m}} + L_{(\mathbf{SA}), \alpha} \left( \frac{\ln(n+1)}{n} \right)^{\frac{p}{p+1}} \left( \frac{1}{\eta} \right)^{\frac{p-1}{p+1}}.$$

By using the latter inequality in (41) we obtain,

$$\text{crit}'(\hat{m}) \geq (1 - 2(\theta - 1)_- - \eta) \mathcal{K}(f_*, \hat{f}_{\hat{m}}) - L_{(\mathbf{SA}),\alpha} \left( \frac{\ln(n+1)}{n} \right)^{\frac{p}{p+1}} \left( \frac{1}{\eta} \right)^{\frac{p-1}{p+1}}. \quad (42)$$

We now compute an upper bound on  $\text{crit}'$  for each model  $m$ . By Lemma 2.6 applied with  $a = (\eta \mathcal{K}_m)^{\frac{p-r-1}{2p}}$ ,  $b = \eta^{-\frac{p-1-r}{2p}} \sqrt{2A_{MR,-} z_n/n}$ ,  $u = \frac{2p}{p-1-r}$  and  $v = \frac{2p}{p+1+r}$ , we have

$$\sqrt{\frac{2A_{MR,-} \mathcal{K}_m^{1-\frac{r+1}{p}} z_n}{n}} \leq \eta \mathcal{K}_m + L_{(\mathbf{SA}),r} \left( \frac{\ln(n+1)}{n} \right)^{\frac{p}{p+1+r}} \left( \frac{1}{\eta} \right)^{\frac{p-1-r}{p+1+r}}.$$

By (30) of the main part, (36), (38), (39), (40) and by using Lemma 2.6 we have on  $\Omega_n$ ,

$$\begin{aligned} \text{crit}'(m) &= \mathcal{K}(f_*, \hat{f}_m) + \text{pen}(m) - p_1(m) - p_2(m) - \bar{\delta}(m) \\ &\leq \mathcal{K}(f_*, \hat{f}_m) + 2(\theta - 1)_+ \left( \frac{D_m}{2n} - L_{(\mathbf{SA})} \varepsilon_n^-(m) \frac{D_m}{n} \right) \\ &\quad + \left( \Delta + (\theta - 1)_+ L_{(\mathbf{SA})}^{(1)} + L_{(\mathbf{SA})}^{(2)} \right) \varepsilon_n^+(m) \frac{D_m}{n} + \sqrt{\frac{2A_{MR,-} \mathcal{K}_m^{1-\frac{r+1}{p}} z_n}{n}} + \frac{2z_n}{nr} \\ &\leq (1 + 2(\theta - 1)_+ + \eta) \mathcal{K}(f_*, \hat{f}_m) + \left( \Delta + (\theta - 1)_+ L_{(\mathbf{SA})}^{(1)} + L_{(\mathbf{SA}),r}^{(2)} \right) \varepsilon_n^+(m) \frac{D_m}{n} \\ &\quad + L_{(\mathbf{SA}),r} \left( \frac{\ln(n+1)}{n} \right)^{\frac{p}{p+1+r}} \left( \frac{1}{\eta} \right)^{\frac{p-1-r}{p+1+r}}. \end{aligned}$$

Recall that we took  $\Delta = L_{(\mathbf{SA})}^{(1)} (\theta - 1)_- + L_{(\mathbf{SA})}^{(2)}$  for some positive constants sufficiently large, so

$$\Delta + (\theta - 1)_+ L_{(\mathbf{SA})}^{(1)} + L_{(\mathbf{SA}),r}^{(2)} \leq |\theta - 1| L_{(\mathbf{SA})}^{(1)} + L_{(\mathbf{SA}),r}^{(2)}$$

and we finally get,

$$\begin{aligned} \text{crit}'(m) &\leq (1 + 2(\theta - 1)_+ + \eta) \mathcal{K}(f_*, \hat{f}_m) + \left( |\theta - 1| L_{(\mathbf{SA})}^{(1)} + L_{(\mathbf{SA}),r}^{(2)} \right) \varepsilon_n^+(m) \frac{D_m}{n} \\ &\quad + L_{(\mathbf{SA}),r} \left( \frac{\ln(n+1)}{n} \right)^{\frac{p}{p+1+r}} \left( \frac{1}{\eta} \right)^{\frac{p-1-r}{p+1+r}}. \end{aligned} \quad (43)$$

Now, as  $\hat{m}$  minimizes  $\text{crit}'$  we get from (42) and (43), on  $\Omega_n$ ,

$$\begin{aligned} \mathcal{K}(f_*, \hat{f}_{\hat{m}}) &\leq \frac{1 + 2(\theta - 1)_+ + \eta}{1 - 2(\theta - 1)_- - \eta} \left( \mathcal{K}(f_*, \hat{f}_{m_*}) + L_{(\mathbf{SA}),\theta,r} \varepsilon_n^+(m_*) \frac{D_{m_*}}{n} \right) \\ &\quad + L_{(\mathbf{SA}),\theta,r} \left( \frac{\ln(n+1)}{n} \right)^{\frac{p}{p+1+r}} \left( \frac{1}{\eta} \right)^{\frac{p-1-r}{p+1+r}}. \end{aligned} \quad (44)$$

We distinguish two cases. If  $D_{m_*} \geq L_{(\mathbf{SA})} (\ln(n+1))^2$  with a constant  $L_{(\mathbf{SA})}$  chosen such that

$$A_0 \varepsilon_n^+(m_*) \leq \frac{1}{2\sqrt{\ln(n+1)}} \leq \frac{1}{2\sqrt{\ln 2}} < 1,$$

where  $A_0$  and  $\varepsilon_n^+(m)$  are defined in Theorem III.1, then by Theorem III.1 it holds on  $\Omega_n$ ,

$$\mathcal{K}(f_*, \hat{f}_{m_*}) \geq \left( 1 - \frac{1}{2\sqrt{\ln(n+1)}} \right) \frac{D_{m_*}}{2n}.$$

On the other hand, if  $D_{m_*} \leq L_{(\mathbf{SA})} (\ln(n+1))^2$  then by Theorem III.1 we have on  $\Omega_n$ ,

$$\mathcal{K}(f_*, \hat{f}_{m_*}) + L_{(\mathbf{SA}),\theta,r} \varepsilon_n^+(m_*) \frac{D_{m_*}}{n} \leq L_{(\mathbf{SA}),\theta,r} \frac{(\ln(n+1))^3}{n}.$$

Hence, in any case we always have on  $\Omega_n$ ,

$$\begin{aligned} & \mathcal{K}(f_*, \hat{f}_{m_*}) + L_{(\mathbf{SA}),\theta,r} \varepsilon_n^+(m_*) \frac{D_{m_*}}{n} \\ & \leq L_{(\mathbf{SA}),\theta,r} \frac{(\ln(n+1))^3}{n} + \left(1 + L_{(\mathbf{SA}),\theta,r} (\ln(n+1))^{-1/2}\right) \mathcal{K}(f_*, \hat{f}_{m_*}). \end{aligned}$$

By taking  $\eta = (\ln(n+1))^{-1/2} (1/2 - (\theta-1)_-)$  and using the fact that in this case,

$$\frac{1 + 2(\theta-1)_+ + \eta}{1 - 2(\theta-1)_- - \eta} \leq \frac{1 + 2(\theta-1)_+ + L_{\theta}\eta}{1 - 2(\theta-1)_-},$$

we deduce that Inequality (44) gives,

$$\begin{aligned} \mathcal{K}(f_*, \hat{f}_{\hat{m}}) & \leq \frac{1 + 2(\theta-1)_+ + L_{(\mathbf{SA}),\theta,r} (\ln(n+1))^{-1/2}}{1 - 2(\theta-1)_-} \inf_{m \in \mathcal{M}_n} \left\{ \mathcal{K}(f_*, \hat{f}_m) \right\} \\ & \quad + L_{(\mathbf{SA}),\theta,r} \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}} \end{aligned}$$

which is Inequality (31).

- Proof of Inequality (32):

By Lemma 2.5 below, we know that  $D_{m_*} \leq L_{(\mathbf{SA})} n^{\frac{1}{1+\beta_+}}$  on  $\Omega_n$ . Furthermore, we have on  $\Omega_n$ , by simple computations,

$$\begin{aligned} \mathcal{K}(f_*, \hat{f}_m) & = \mathcal{K}(f_*, f_m) + \mathcal{K}(f_m, \hat{f}_m) \\ & \leq C_+ D_m^{-\beta_+} + (1 + L_{(\mathbf{SA})} \varepsilon_n^+(m)) \frac{D_m}{2n} \\ & \leq L_{(\mathbf{SA})} \left( D_m^{-\beta_+} + \frac{D_m}{n} + \frac{\ln(n+1)}{n} \right). \end{aligned}$$

This yields

$$\begin{aligned} \inf_{m \in \mathcal{M}_n} \left\{ \mathcal{K}(f_*, \hat{f}_m) \right\} & \leq L_{(\mathbf{SA})} \inf \left\{ D_m^{-\beta_+} + \frac{D_m}{n} + \frac{\ln(n+1)}{n}; m \in \mathcal{M}_n, D_m \leq L_{(\mathbf{SA})} n^{\frac{1}{1+\beta_+}} \right\} \\ & \leq L_{(\mathbf{SA})} n^{-\frac{\beta_+}{1+\beta_+}}. \end{aligned}$$

To conclude, it suffices to notice that if  $\beta_+ > p$  then

$$n^{-\frac{\beta_+}{1+\beta_+}} \leq n^{-\frac{p}{p+1}} \leq n^{-\frac{p}{p+1+r}},$$

which finally gives

$$\frac{1 + 2(\theta-1)_+ + L_{(\mathbf{SA}),\theta,r} (\ln(n+1))^{-1/2}}{1 - 2(\theta-1)_-} \inf_{m \in \mathcal{M}_n} \left\{ \mathcal{K}(f_*, \hat{f}_m) \right\} \leq L_{(\mathbf{SA}),\theta,r} \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}}.$$

### Proof of Theorem 2.3

Note first that Theorem 2.2 is valid under the assumptions of Theorem 2.3, so we may use any result established in the proof of Theorem 2.2.

- Proof of Inequality (33):

From **(A<sub>P</sub>)**, we know by Lemma 2.5 below that there exist  $L_{(\mathbf{SA})}^{(1)}, L_{(\mathbf{SA})}^{(2)} > 0$  such that

$$L_{(\mathbf{SA})}^{(1)} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}} \leq D_{m_*} \leq L_{(\mathbf{SA})}^{(2)} n^{\frac{1}{1+\beta_+}}$$

and so

$$\begin{aligned} \varepsilon_n^+(m_*) &\leq \max \left\{ \sqrt{\frac{D_{m_*} \ln(n+1)}{n}}; \sqrt{\frac{\ln(n+1)}{D_{m_*}}}; \frac{\ln(n+1)}{D_{m_*}} \right\} \\ &\leq L_{(\mathbf{SA})} \max \left\{ n^{-\frac{\beta_+}{2(1+\beta_+)}}; n^{-\frac{\beta_+}{2(1+\beta_+)\beta_-}} \right\} \sqrt{\ln(n+1)} \\ &= L_{(\mathbf{SA})} n^{-\frac{\beta_+}{(1+\beta_+)\beta_-}} \sqrt{\ln(n+1)}. \end{aligned} \quad (45)$$

Assume for now that we also have

$$A_0 \varepsilon_n^+(m_*) \leq (\ln(n+1))^{-1/2}, \quad (46)$$

where  $A_0$  and  $\varepsilon_n^+(m)$  are defined in Theorem III.1. Then by Theorem III.1 it holds on  $\Omega_n$ ,

$$\mathcal{K}(f_*, \hat{f}_{m_*}) \geq \left(1 - (\ln(n+1))^{-1/2}\right) \frac{D_{m_*}}{2n}.$$

In this case, we deduce that

$$\begin{aligned} &\mathcal{K}(f_*, \hat{f}_{m_*}) + L_{(\mathbf{SA}),\theta,r} \varepsilon_n^+(m_*) \frac{D_{m_*}}{n} \\ &\leq \left(1 + L_{(\mathbf{SA}),\theta,r} n^{-\frac{\beta_+}{(1+\beta_+)\beta_-}} \sqrt{\ln(n+1)}\right) \mathcal{K}(f_*, \hat{f}_{m_*}), \end{aligned} \quad (47)$$

and Inequality (33) simply follows from using Inequality (44).

If Inequality (46) is not satisfied, that is

$$A_0 \varepsilon_n^+(m_*) > (\ln(n+1))^{-1/2}, \quad (48)$$

then by (45), this means that there exists a positive constant  $L_{(\mathbf{SA})}$  such that

$$L_{(\mathbf{SA})} n^{-\frac{\beta_+}{(1+\beta_+)\beta_-}} \sqrt{\ln(n+1)} > (\ln(n+1))^{-1/2}.$$

Consequently, this ensures that in the case where (48) is true, we also have  $n \leq n_0((\mathbf{SA}))$ . Hence, as

$$\mathcal{K}(f_*, \hat{f}_{m_*}) \geq C_- D_{m_*}^{-\beta_-} \geq L_{(\mathbf{SA})} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}} > 0,$$

this yields Inequality (47) with a positive constant  $L_{(\mathbf{SA}),\theta,r}$  in the right-hand term sufficiently large and then the result easily follows from using Inequality (44).

- Proof of Inequality (34):

If  $D_{m_*} \geq L_{(\mathbf{SA})} (\ln(n+1))^2$  with a constant  $L_{(\mathbf{SA})}$  chosen such that

$$A_0 \varepsilon_n^+(m_*) \leq 1/2,$$

where  $A_0$  and  $\varepsilon_n^+(m)$  are defined in Theorem III.1, then by Theorem III.1 it holds on  $\Omega_n$ ,

$$\mathcal{K}(f_*, \hat{f}_{m_*}) \geq C_- D_{m_*}^{-\beta_-} + \frac{D_{m_*}}{4n}.$$



By Lemma 2.5 below we know that  $L_{(\mathbf{SA})}^{(1)} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}} \leq D_{m_*} \leq L_{(\mathbf{SA})}^{(2)} n^{\frac{1}{1+\beta_+}}$  on  $\Omega_n$ . This gives

$$\mathcal{K}(f_*, \hat{f}_{m_*}) \geq L_{(\mathbf{SA})} n^{-\frac{\beta_-}{1+\beta_+}} + L_{(\mathbf{SA})} n^{-1 + \frac{\beta_+}{(1+\beta_+)\beta_-}}$$

and we deduce by simple algebra that if  $\beta_- < p(1+\beta_+)/(1+p+r)$  or  $p/(1+r) > \beta_+ / (\beta_- (1+\beta_+)) - 1$ , then

$$L_{(\mathbf{SA})} (\ln(n+1))^{-1/2} \mathcal{K}(f_*, \hat{f}_{m_*}) \geq \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}}.$$

On the other hand, if  $D_{m_*} \leq L_{(\mathbf{SA})} (\ln(n+1))^2$ , this implies in particular

$$L_{(\mathbf{SA})}^{(1)} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}} \leq L_{(\mathbf{SA})} (\ln(n+1))^2.$$

Hence, there exists an integer  $n_0((\mathbf{SA}))$  such that  $n \leq n_0((\mathbf{SA}))$ . In this case, we can find a constant  $L_{(\mathbf{SA})}$  such that

$$L_{(\mathbf{SA})} (\ln(n+1))^{-1/2} \mathcal{K}(f_*, \hat{f}_{m_*}) \geq \frac{(\ln(n+1))^{\frac{3p-1-r}{2(p+1+r)}}}{n^{\frac{p}{p+1+r}}}$$

and through the use of inequality (33), this conclude the proof of Inequality (34).

**Lemma 2.4 (Control on the dimension of the selected model)** *Assume that  $(\mathbf{SA}_0)$  holds together with  $(\mathbf{Ap}_u)$ . If  $\beta_+ \leq \frac{p}{r+1}$  then, on the event  $\Omega_n$  defined in the proof of Theorem 2.2, we have*

$$D_{\hat{m}} \leq L_{\Delta, \theta, r, (\mathbf{SA})} n^{\frac{1}{2+\beta_+(1-\frac{r+1}{p})}} \sqrt{\ln(n+1)}. \quad (49)$$

If moreover  $(\mathbf{Ap})$  holds, then we get on the event  $\Omega_n$ ,

$$L_{\Delta, \theta, (\mathbf{SA})}^{(1)} \frac{n^{\frac{\beta_+}{\beta_-(1+\beta_+)}}}{(\ln(n+1))^{\frac{1}{\beta_-}}} \leq D_{\hat{m}} \leq L_{\Delta, \theta, (\mathbf{SA})}^{(2)} n^{\frac{1}{2+\beta_+(1-\frac{r+1}{p})}} \sqrt{\ln(n+1)}. \quad (50)$$

**Lemma 2.5 (Control over the dimension of oracle models)** *Assume that  $(\mathbf{SA}_0)$  holds together with  $(\mathbf{Ap}_u)$ . We have on the event  $\Omega_n$  defined in the proof of Theorem 2.2,*

$$D_{m_*} \leq L_{(\mathbf{SA})} n^{\frac{1}{1+\beta_+}}.$$

If moreover  $(\mathbf{Ap})$  holds, then we get on the event  $\Omega_n$ ,

$$L_{(\mathbf{SA})}^{(1)} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}} \leq D_{m_*} \leq L_{(\mathbf{SA})}^{(2)} n^{\frac{1}{1+\beta_+}}. \quad (51)$$

In the proofs of Lemmas 2.4 and 2.5 below, we assume for ease of presentation that for any  $\lambda \in (0, 1)$ , there exists a model  $m \in \mathcal{M}_n$  such that  $D_m = \lceil n^\lambda \rceil$ . Note that our reasoning would still work by invoking assumption **(P3)** of  $(\mathbf{SA}_0)$ .

**Proof of Lemma 2.4.** Recall that  $\hat{m}$  minimizes

$$\text{crit}'(m) = \text{crit}(m) - P_n(-\ln f_*) = \mathcal{K}_m - p_2(m) - \bar{\delta}(m) + \text{pen}(m)$$

over the models  $m \in \mathcal{M}_n$ . Moreover,  $\text{pen}(m) = (\theta + \Delta \varepsilon_n^+(m)) D_m/n$ . The analysis is restricted on  $\Omega_n$ .

1. Upper bound on  $\text{crit}'(m)$ :

$$\begin{aligned} p_2(m) &\geq \left( \frac{1}{2} - L_{(\mathbf{SA})} \varepsilon_n^+(m) \right) \frac{D_m}{n} \\ -\bar{\delta}(m) &\leq \sqrt{\frac{2w_m z_n}{n}} + \frac{2z_n}{nr}. \end{aligned}$$

Moreover, by Proposition IV.5, we have  $w_m \leq A_{MR,-} \mathcal{K}_m^{1-\frac{r+1}{p}}$  and so,

$$\begin{aligned} \text{crit}'(m) &\leq \mathcal{K}_m + \left( \theta - \frac{1}{2} + L_{\Delta,(\mathbf{SA})} \varepsilon_n^+(m) \right) \frac{D_m}{n} + \sqrt{\frac{2A_{MR,-} \mathcal{K}_m^{1-\frac{r+1}{p}} z_n}{n}} \\ &\leq L_{\Delta,\theta,(\mathbf{SA})} \left( D_m^{-\beta_+} + \frac{D_m}{n} + \frac{\ln(n+1)}{n} + \sqrt{\frac{D_m^{-\beta_+ \left(1-\frac{r+1}{p}\right)} \ln(n+1)}{n}} \right). \end{aligned}$$

Now, if  $\beta_+ \leq \frac{p}{r+1}$ , then for  $m_0$  such that  $D_{m_0} = \left\lceil n^{\frac{1}{1+\beta_+}} \right\rceil$  we have

$$\frac{D_{m_0}}{n} \leq 2n^{-\frac{\beta_+}{1+\beta_+}}; \quad D_{m_0}^{-\beta_+} \leq n^{-\frac{\beta_+}{1+\beta_+}}; \quad \sqrt{\frac{D_{m_0}^{-\beta_+ \left(1-\frac{r+1}{p}\right)}}{n}} \leq n^{-\frac{\beta_+ \left(2-\frac{r+1}{p}\right)+1}{2(1+\beta_+)}} \leq n^{-\frac{\beta_+}{1+\beta_+}},$$

so we get

$$\text{crit}'(m_0) \leq L_{\Delta,\theta,(\mathbf{SA})} n^{-\frac{\beta_+}{1+\beta_+}} \sqrt{\ln(n+1)}. \quad (52)$$

Otherwise, if  $\beta_+ > \frac{p}{r+1}$ , then for  $m_1$  such that  $D_{m_1} = \left\lceil n^{\frac{1}{2+\beta_+ \left(1-\frac{r+1}{p}\right)}} \right\rceil$ , we have

$$\begin{aligned} \frac{D_{m_1}}{n} &\leq 2n^{-\frac{1+\beta_+ \left(1-\frac{r+1}{p}\right)}{2+\beta_+ \left(1-\frac{r+1}{p}\right)}}; \quad \sqrt{\frac{D_{m_1}^{-\beta_+ \left(1-\frac{r+1}{p}\right)}}{n}} \leq n^{-\frac{1+\beta_+ \left(1-\frac{r+1}{p}\right)}{2+\beta_+ \left(1-\frac{r+1}{p}\right)}}; \\ D_{m_1}^{-\beta_+} &\leq n^{-\frac{\beta_+}{2+\beta_+ \left(1-\frac{r+1}{p}\right)}} \leq n^{-\frac{1+\beta_+ \left(1-\frac{r+1}{p}\right)}{2+\beta_+ \left(1-\frac{r+1}{p}\right)}} \end{aligned}$$

and

$$\text{crit}'(m_1) \leq L_{\Delta,\theta,(\mathbf{SA})} n^{-\frac{1+\beta_+ \left(1-\frac{r+1}{p}\right)}{2+\beta_+ \left(1-\frac{r+1}{p}\right)}} \sqrt{\ln(n+1)}. \quad (53)$$

**2. Lower bound on  $\text{crit}'(m)$ :** we have

$$\begin{aligned} p_2(m) &\leq \left( \frac{1}{2} + L_{(\mathbf{SA})} \varepsilon_n^+(m) \right) \frac{D_m}{n} \\ -\bar{\delta}(m) &\geq -\sqrt{\frac{2v_m z_n}{n}} - \frac{2z_n}{n}. \end{aligned}$$

Moreover, by Proposition IV.5 we have for some constant  $A_{MR,-} > 0$ ,  $v_m \leq A_{MR,-} \mathcal{K}_m^{1-\frac{1}{p}}$ . For  $\Delta$  large enough we thus get,

$$\text{crit}'(m) \geq \mathcal{K}_m + \left( \theta - \frac{1}{2} \right) \frac{D_m}{n} - \sqrt{\frac{2A_{MR,-} \mathcal{K}_m^{1-\frac{1}{p}} z_n}{n}} \quad (54)$$

Assume that  $\beta_+ \leq \frac{p}{r+1}$ . We take  $D_m \leq L(n/\ln(n+1))^{\frac{p}{(p+1)\beta_-}}$  for some constant  $L > 0$ . If  $L$  is small enough, we have by **(A<sub>p</sub>)**  $\mathcal{K}_m \geq (8A_{MR,-} z_n/n)^{p/(p+1)}$  and by (54),  $\text{crit}'(m) \geq \mathcal{K}_m/2$ . Now if

$$D_m \leq L \left( \left( \frac{n}{\ln(n+1)} \right)^{\frac{p}{(1+p)\beta_-}} \wedge \frac{n^{\frac{\beta_+}{\beta_-(1+\beta_+)}}}{(\ln(n+1))^{\frac{1}{2\beta_-}}} \right) \quad (55)$$

with  $L$  sufficiently small, only depending on  $\Delta, \theta$  and constants in **(SA)**, then by (52) we obtain  $\text{crit}'(m) > \text{crit}'(m_0)$ . As  $\beta_+ \leq \frac{p}{r+1} < p$ , the upper bound in (55) reduces to  $D_m \leq Ln^{\frac{\beta_+}{\beta_-(1+\beta_+)}} / (\ln(n+1))^{\frac{1}{\beta_-}}$ . This proves the left-hand side of (50). Assume now that  $\beta_+ > \frac{p}{r+1}$ . If

$$D_m \leq L \left( \left( \frac{n}{\ln(n+1)} \right)^{\frac{p}{(1+p)\beta_-}} \wedge \frac{n^{\frac{1+\beta_+(1-\frac{r+1}{p})}{\beta_-(2+\beta_+(1-\frac{r+1}{p}))}}}{(\ln(n+1))^{\frac{1}{2\beta_-}}} \right)$$

with  $L$  sufficiently small, only depending on  $\Delta, \theta$  and constants in **(SA)**, then by (53) we obtain  $\text{crit}'(m) > \text{crit}'(m_1)$ . As  $\beta_+ \leq \frac{p}{r+1} < p$ , the upper bound in (55) reduces to  $D_m \leq Ln^{\frac{\beta_+}{\beta_-(1+\beta_+)}} / (\ln(n+1))^{\frac{1}{\beta_-}}$ . This proves the left-hand side of (50).

Assume that  $\beta_+ \leq \frac{p}{r+1}$ . We take  $D_m \leq L(n \ln(n+1))^{\frac{1}{2+\beta_+(1-\frac{1}{p})}}$  for some constant  $L > 0$ . If  $L$  is large enough, then we get by (54) and simple calculations,  $\text{crit}'(m) \geq (\theta/2 - 1/4) D_m/n$ . Furthermore, if

$$D_m \geq L \left( (n \ln(n+1))^{\frac{1}{2+\beta_+(1-\frac{1}{p})}} \wedge n^{\frac{1}{1+\beta_+}} \sqrt{\ln(n+1)} \right) \quad (56)$$

with  $L$  sufficiently small, only depending on  $\Delta, \theta$  and constants in **(SA)**, then by (52) we obtain  $\text{crit}'(m) > \text{crit}'(m_0)$ . As  $\beta_+ \leq \frac{p}{r+1} < p$ , the lower bound in (56) reduces to  $D_m \leq L(n \ln(n+1))^{\frac{1}{2+\beta_+(1-\frac{1}{p})}}$ . This proves the left-hand side of (50).

Assume now that  $\beta_+ > \frac{p}{r+1}$ . If

$$D_m \geq L \left( (n \ln(n+1))^{\frac{1}{2+\beta_+(1-\frac{1}{p})}} \vee n^{\frac{1}{2+\beta_+(1-\frac{r+1}{p})}} \sqrt{\ln(n+1)} \right) \quad (57)$$

with  $L$  sufficiently large, only depending on  $\Delta, \theta$  and constants in **(SA)**, then by (52) we obtain  $\text{crit}'(m) > \text{crit}'(m_1)$ . As  $\beta_+ \leq \frac{p}{r+1} < p$ , the upper bound in (55) reduces to  $D_m \leq Ln^{\frac{\beta_+}{\beta_-(1+\beta_+)}} / (\ln(n+1))^{\frac{1}{\beta_-}}$ . This proves the left-hand side of (50). As  $r > 0$ , (57) reduces to  $D_m \geq Ln^{\frac{1}{2+\beta_+(1-\frac{r+1}{p})}} \sqrt{\ln(n+1)}$ , which proves (49) and the right-hand side of (50).

**Proof of Lemma 2.5.** By definition,  $m_*$  minimizes

$$\mathcal{K}(f_*, \hat{f}_m) = \mathcal{K}_m + p_1(m)$$

over the models  $m \in \mathcal{M}_n$ . The analysis is restricted on  $\Omega_n$ .

1. Upper bound on  $\mathcal{K}(f_*, \hat{f}_m)$ : we have

$$\begin{aligned} \mathcal{K}(f_*, \hat{f}_m) &\leq C_+ D_m^{-\beta_+} + \left( \frac{1}{2} + L_{(\mathbf{SA})} \varepsilon_n^+(m) \right) \frac{D_m}{n} \\ &\leq L_{(\mathbf{SA})} \left( D_m^{-\beta_+} + \frac{D_m}{n} + \frac{\ln(n+1)}{n} \right). \end{aligned}$$

Hence, if  $m_0$  is such that  $D_{m_0} = n^{\frac{1}{1+\beta_+}}$ , then

$$\mathcal{K}(f_*, \hat{f}_{m_0}) \leq L_{(\mathbf{SA})} n^{-\frac{\beta_+}{1+\beta_+}}. \quad (58)$$

2. Lower bound on  $\mathcal{K}(f_*, \hat{f}_m)$ : there exists a constant  $A_0$ , only depending on constants in **(SA)**, such that

$$\begin{aligned} \mathcal{K}(f_*, \hat{f}_m) &\geq \mathcal{K}_m + \left(\frac{1}{2} - A_0 \varepsilon_n^-(m)\right) \frac{D_m}{n} \\ &\geq \mathcal{K}_m + \frac{D_m}{2n} - A_0 \max \left\{ \left(\frac{D_m}{n}\right)^{3/2} \sqrt{\ln(n+1)}; \frac{\sqrt{D_m \ln(n+1)}}{n} \right\}. \end{aligned} \quad (59)$$

If **(Ap)** holds, then for

$$D_m \leq L_{(\mathbf{SA})} \min \left\{ \frac{n^{\frac{3}{3+2\beta_-}}}{(\ln(n+1))^{\frac{1}{3+2\beta_-}}}; \frac{n^{\frac{2}{1+2\beta_-}}}{(\ln(n+1))^{\frac{1}{1+2\beta_-}}} \right\}$$

with  $L_{(\mathbf{SA})}$  sufficiently small, we have

$$\frac{\mathcal{K}_m}{2} \geq C_- \frac{D_m^{-\beta_-}}{2} \geq A_0 \max \left\{ \left(\frac{D_m}{n}\right)^{3/2} \sqrt{\ln(n+1)}; \frac{\sqrt{D_m \ln(n+1)}}{n} \right\}.$$

In this case, we have by (59),  $\mathcal{K}(f_*, \hat{f}_m) \geq \mathcal{K}_m/2 + (D_m)/2n \geq C_- D_m^{-\beta_-}/2$ . Moreover, if  $m$  is such that  $D_m \leq L_{(\mathbf{SA})} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}}$  with  $L_{(\mathbf{SA})}$  sufficiently small, we also have

$$\begin{aligned} D_m &\leq L_{(\mathbf{SA})} n^{\frac{\beta_+}{(1+\beta_+)\beta_-}} \\ &\leq L_{(\mathbf{SA})} \min \left\{ n^{\frac{\beta_+}{(1+\beta_+)\beta_-}}; \frac{n^{\frac{3}{3+2\beta_-}}}{(\ln(n+1))^{\frac{1}{3+2\beta_-}}}; \frac{n^{\frac{2}{1+2\beta_-}}}{(\ln(n+1))^{\frac{1}{1+2\beta_-}}} \right\} \end{aligned}$$

and by (58) we get  $\mathcal{K}(f_*, \hat{f}_{m_0}) < \mathcal{K}(f_*, \hat{f}_m)$ , which gives the left-hand side of (51).

We turn now to the proof of the right-hand side of (51). Let  $m \in \mathcal{M}_n$  be such that  $D_m \geq L_1 n^{\frac{1}{1+\beta_+}}$ . By (58) we deduce that if  $L_1$  is large enough, depending only on constants in **(SA)**, then we have

$$\frac{D_m}{4n} > \mathcal{K}(f_*, \hat{f}_{m_0}).$$

In addition, if  $D_m \geq L_2 (\ln(n+1))^2$  and  $D_m \leq L_2^{-1} n / \ln(n+1)$  for some constant  $L_2$  sufficiently large, then

$$\frac{D_m}{4n} \geq A_0 \max \left\{ \left(\frac{D_m}{n}\right)^{3/2} \sqrt{\ln(n+1)}; \frac{\sqrt{D_m \ln(n+1)}}{n} \right\}$$

and by (59), we deduce that  $\mathcal{K}(f_*, \hat{f}_m) > \mathcal{K}(f_*, \hat{f}_{m_0})$ . The latter inequality implies that  $D_{m_*} \leq L_1 n^{\frac{1}{1+\beta_+}}$ . Our reasoning is valid if  $n$  is such that  $L_2 (\ln(n+1))^2 \leq L_1 n^{\frac{1}{1+\beta_+}} \leq L_2^{-1} n / \ln(n+1)$ . At the price of enlarging  $L_1$ , we can always achieve  $L_2 (\ln(n+1))^2 \leq L_1 n^{\frac{1}{1+\beta_+}}$ , with  $L_1$  not depending on  $n$ . Then if  $L_2^{-1} n / \ln(n+1) < L_1 n^{\frac{1}{1+\beta_+}}$ , we still have

$$D_{m_*} \leq \max_{m \in \mathcal{M}_n} D_m \leq A_{\mathcal{M},+} \frac{n}{(\ln(n+1))^2} \leq A_{\mathcal{M},+} L_2 L_1 n^{\frac{1}{1+\beta_+}}.$$

In every case, there exists  $L > 0$  only depending on constants in **(SA)** such that  $D_{m_*} \leq L n^{\frac{1}{1+\beta_+}}$ .

**Lemma 2.6** *Let  $(a, b) \in \mathbb{R}_+^2$  and  $(u, v) \in [1, \infty]$  such that  $1/u + 1/v = 1$ . Then*

$$ab \leq \max \{a^u; b^v\} \leq a^u + b^v.$$

**Proof.** By symmetry, we can assume  $a^u \geq b^v$ . Then  $b = (b^v)^{1/v} \leq a^{u/v} = a^{u-1}$  and so,  $ab \leq a a^{u-1} = a^u \leq a^u + b^v$ . ■

## References

- [1] A. Barron and C. Sheu. Approximation of density functions by sequences of exponential families. *Ann. Statist.*, 19(3):1347–1369, 1991.
- [2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013.
- [3] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6):495–500, 2002.
- [4] G. Castellán. Modified Akaike’s criterion for histogram density estimation. *Technical report #99.61, Université Paris-Sud*, 1999.
- [5] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.
- [6] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.