

Technical appendix to “Model selection by resampling penalization”

Sylvain Arlot

Sylvain Arlot
Univ Paris-Sud, UMR 8628,
Laboratoire de Mathématiques,
Orsay, F-91405 ; CNRS, Orsay, F-91405 ;
INRIA-Futurs, Projet Select
e-mail: sylvain.arlot@math.u-psud.fr

Abstract: This is a technical appendix to “Model selection by resampling penalization”. We present additional sufficient conditions for the assumptions of our main results, results from a wider simulation experiment and several proofs that have been skipped in the main paper.

AMS 2000 subject classifications: Primary 62G09 ; secondary 62G08, 62M20.

Keywords and phrases: non-parametric statistics, resampling, non-asymptotic, exchangeable weighted bootstrap, model selection, penalization, non-parametric regression, adaptivity, heteroscedastic data, histogram.

Throughout this appendix, we use the notations of the main paper (2). In order to distinguish references within the appendix from references to the main paper, we denote the former ones by (1) or 1, and the latter ones by **(1)** or **1**.

Following the ordering of (2), we first complete Sect. **4.3** with other alternative assumptions to Thm. **1** (Sect. 1). Then, we state some approximation theory results for histogram models that give sufficient condition for assumption **(Ap)** on the bias of the models (Sect. 2). In Sect. 3 the additional simulation studies mentioned in Sect. **5**. An extensive simulation experiment is provided in Sect. 3. Finally, some concentration inequalities stated or used in Sect. **9.9** are proven in Sect. 4, and the technical lemmas of Sect. **9.10** are proven in Sect. 4.

1. Other assumption sets for oracle inequalities for RP

In this section, we consider alternative assumption sets for the results of Sect. **4** about Resampling Penalization. Since Thm. **1** relies on a general result (Lemma **7**), giving alternative assumptions for Thm. **1** remains to give sufficient conditions for **(Bg)** or **(Ug)**.

1.1. Bounded case

We already suggested one way of removing **(An)** in Sect. **4**. It is actually possible to replace it in the assumptions of Thm. **1** by

1. ($\mathbf{A}_{\text{gauss}}$) the noise is sub-gaussian
and ($\mathbf{Ar}_{\mathbf{u}}^X$) the partition is “upper-regular” for $\mathcal{L}(X)$, i.e. $D_m \max_{\lambda \in \Lambda_m} p_\lambda \leq c_{r,u}^X$.
2. $X \subset \mathbb{R}^k$, ($\mathbf{A}_{\text{gauss}}$) the noise is sub-gaussian,
($\mathbf{Ar}_{\mathbf{u}}$) the partition is “upper-regular” for Leb : $D_m \max_{\lambda \in \Lambda_m} \text{Leb}(I_\lambda) \leq c_{r,u} \text{Leb}(\mathcal{X})$
and the density of X w.r.t. Leb is bounded from above ($\mathbf{Ad}_{\mathbf{u}}$)

$$\forall I \subset \mathcal{X}, \quad \mathbb{P}(X \in I) \leq c_X^{\max} \frac{\text{Leb}(I)}{\text{Leb}(\mathcal{X})} .$$

Proof. Following the proof given in App. 9.4, we only have to give a lower bound on $Q_m^{(p)} := D_m^{-1} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$.

In the first case, we have

$$\begin{aligned} Q_m^{(p)} &\geq \frac{1}{D_m} \sum_{\lambda \in \Lambda_m} (\sigma_\lambda^r)^2 = \sum_{\lambda \in \Lambda_m} \frac{p_\lambda (\sigma_\lambda^r)^2}{D_m p_\lambda} \\ &\geq \sum_{\lambda \in \Lambda_m} \frac{p_\lambda (\sigma_\lambda^r)^2}{\max_{\lambda \in \Lambda_m} \{D_m p_\lambda\}} \geq \frac{\|\sigma(X)\|_2^2}{c_{r,u}^X} . \end{aligned}$$

The second case is a consequence of the first one since

$$\max_{\lambda \in \Lambda_m} \{p_\lambda\} \leq c_X^{\max} \max_{\lambda \in \Lambda_m} \left\{ \frac{\text{Leb}(I_\lambda)}{\text{Leb}(\mathcal{X})} \right\} \leq c_X^{\max} c_{r,u} D_m^{-1} .$$

□

Moreover, in all the assumption sets above and in those of Sect. 4, the sub-gaussian assumption on the noise ($\mathbf{A}_{\text{gauss}}$) can be replaced by a general moment inequality:

($\mathbf{A}\epsilon$) Pointwise moment inequality for the noise: there exists P^{pt} growing as some power of q such that

$$\forall q \geq 2, \forall x \in \mathcal{X}, \quad \mathbb{E}[|\epsilon|^q | X = x]^{1/q} \leq P^{pt}(q) \sigma(x) .$$

For instance, when $P^{pt}(q) \leq cq$ for every $q \geq 2$ for some constant c , this means that ϵ is *sub-poissonian*.

1.2. Unbounded case

In Sect. 4, we also give a set of assumptions for Thm. 1 in the unbounded case. One can actually remove both (\mathbf{Ab}) and the lower bound on the noise (\mathbf{An}) from the assumptions of Thm. 1, at the price of adding

(**A**_{gauss}) The noise is sub-gaussian: there exists $c_{\text{gauss}} > 0$ such that

$$\forall q \geq 2, \forall x \in \mathcal{X}, \quad \mathbb{E}[|\epsilon|^q | X = x]^{1/q} \leq c_{\text{gauss}} \sqrt{q} \sigma(x) .$$

(**A** δ) Global moment assumption for the bias: there is a constant $c_{\Delta, m}^g > 0$ such that, for every $m \in \mathcal{M}_n$ of dimension $D_m \geq D_0$,

$$\|s - s_m\|_{\infty} \leq c_{\Delta, m}^g \|s(X) - s_m(X)\|_2$$

(**A** σ_{max}) Noise-level bounded from above: $\sigma^2(X) \leq \sigma_{\text{max}}^2 < +\infty$ a.s.

(**A** s_{max}) Bound on the target function: $\|s\|_{\infty} \leq A$.

and one among the following

1. (**A** $r_{\mathbf{u}}^X$) the partition is ‘‘upper-regular’’ for $\mathcal{L}(X)$, i.e. $D_m \max_{\lambda \in \Lambda_m} p_{\lambda} \leq c_{r, \mathbf{u}}^X$.
2. $X \subset \mathbb{R}^k$,
 (**A** $r_{\mathbf{u}}$) the partition is ‘‘upper-regular’’ for Leb: $D_m \max_{\lambda \in \Lambda_m} \text{Leb}(I_{\lambda}) \leq c_{r, \mathbf{u}} \text{Leb}(\mathcal{X})$
 and the density of X w.r.t. Leb is bounded from above (**A** $d_{\mathbf{u}}$):

$$\forall I \subset \mathcal{X}, \quad \mathbb{P}(X \in I) \leq c_X^{\max} \frac{\text{Leb}(I)}{\text{Leb}(\mathcal{X})} .$$

3. $X \subset \mathbb{R}^k$ is bounded, equipped with $\|\cdot\|_{\infty}$,
 (**A** $r_{\mathbf{u}}^d$) the partition is ‘‘upper-regular’’: $\max_{\lambda \in \Lambda_m} \{\text{diam}(I_{\lambda})\} \leq c_{r, \mathbf{u}}^d D_m^{-\alpha_d} \text{diam}(X)$
 (**A** $r_{\mathbf{u}}$) the partition is ‘‘upper-regular’’ for Leb: $\max_{\lambda \in \Lambda_m} \{\text{Leb}(I_{\lambda})\} \leq c_{r, \mathbf{u}} D_m^{-1} \text{Leb}(X)$
 and (**A** σ) σ is piecewise K_{σ} -Lipschitz with at most J_{σ} jumps.

Proof. In Sect. 9.4, we made this proof with (**A** n) as last additional assumption. It is actually only used to give a lower bound on $Q_m^{(p)}$. The two first cases thus follows from the proof given in Sect. 1.1 above. The last one follows from Lemma 16. \square

As in the bounded case, the sub-gaussian assumption on the noise (**A**_{gauss}) can be replaced everywhere by the more general moment assumption (**A** ϵ).

Sufficient conditions for (**A** δ) can be derived either from Lemma 17 or from Lemma 1 below.

1.3. An alternative to Lemma 17

In Sect. 9.10, we give a sufficient condition for (**A** δ) that relies on the regularity of s , a lower bound on the density of X w.r.t. Leb and the regularity of the partition (Lemma 17). We state below a lemma which gives a more accurate estimation of the constant $c_{\Delta, m}^g$ when $\mathcal{X} \subset \mathbb{R}$ is bounded.

Lemma 1. *Let s be a C_{Lip} -Lipschitz function on $\mathcal{X} \subset \mathbb{R}$ and μ a probability measure on \mathcal{X} . We assume that μ and Leb are mutually absolutely continuous. Let $\left((I_\lambda)_{\lambda \in \Lambda_{m_k}} \right)_{k \in \mathbb{N}}$ be a sequence of partitions of \mathcal{X} . We assume that their sizes D_{m_k} are going to infinity and (\mathbf{Ar}_u^d) :*

$$\max_{\lambda \in \Lambda_m} \{ \text{diam}(I_\lambda) \} \leq c_{r,u}^d D_m^{-1} \text{diam}(\mathcal{X}) .$$

Then, there exists a constant $c_{\Delta,m}^g$ (depending on s , μ and $c_{r,u}^d$) such that for every $k \in \mathbb{N}$,

$$\|s - s_m\|_\infty \leq c_{\Delta,m}^g \|s - s_m\|_{L^2(\mu)} .$$

Proof. If s is constant, the result is obvious. Otherwise, both $\|s - s_m\|_\infty$ and $\|s - s_m\|_2$ are positive.

Since s is Lipschitz, with constant C_{Lip} , we have

$$\|s - s_m\|_\infty \leq C_{Lip} \max_{\lambda \in \Lambda_m} \{ \text{diam}(I_\lambda) \} \leq D_m^{-1} \text{diam}(\mathcal{X}) c_{r,u}^d C_{Lip} .$$

On the other hand, $\|s - s_{m_k}\|_{L^2(\mu)}^2$ is equivalent to $\frac{\|s'\|_{L^2(\mu)}^2}{12D_{m_k}^2}$ as long as the Riemann sums of s' are converging. The result follows. \square

Remark 1. If one assumes further regularity conditions on s , then the difference between $\|s - s_{m_k}\|_{L^2(\mu)}^2$ and its limit when $k \rightarrow \infty$ can be controlled. Then, the constant $c_{\Delta,m}^g$ only depends on s through $\|s'\|_{L^2(\mu)}^2$ and these regularity conditions, at least for $k \geq k_0$ for some k_0 depending on the same conditions.

2. Approximation properties of histograms

In Thm. 1, we use the following assumption:

(Ap) Polynomial decreasing of the bias: there exists $\beta_1 \geq \beta_2 > 0$ and $C_b^+, C_b^- > 0$ such that

$$C_b^- D_m^{-\beta_1} \leq \|s - s_m\|_{L^2(\mu)} \leq C_b^+ D_m^{-\beta_2} .$$

where $\mu = \mathcal{L}(X)$ and s_m is the $L^2(\mu)$ projection onto some histogram model $(S_m)_{m \in \mathcal{M}}$. It is somehow unintuitive, since it assumes that s is not too well approximated by histograms. For instance, it excludes the case of constant functions, which are both α -hölderian (for any α) and histogram functions. Lemma 2 below shows that it is the only excluded function among the hölderian ones. On approximation theory, we refer to the book of DeVore and Lorentz (5). All the proofs of the following results can be found in Chap. 8.10 of (1), where they were initially stated.

Let (\mathcal{X}, d) be a metric space. For every $\alpha \in (0; 1]$, $\delta, \epsilon, R > 0$, we define $\mathcal{H}_{\delta,\epsilon}(\alpha, R)$ the set of α -hölderian functions f on \mathcal{X} , *i.e.*:

$$\forall x, y \in [0; 1], \quad |s(x) - s(y)| \leq R d(x, y)^\alpha$$

such that there exists $x_1, x_2 \in \mathcal{X}$ such that

$$d(x_1, x_2) \leq \delta \quad \text{and} \quad |s(x_1) - s(x_2)| \geq \epsilon .$$

When \mathcal{X} is bounded, we also define

$$\mathcal{H}_\epsilon(\alpha, R) := \mathcal{H}_{\text{diam}(\mathcal{X}), \epsilon}(\alpha, R) .$$

2.1. Regular histograms in $[0; 1]$

We first investigate the simplest case, where $(\mathcal{X}, d) = ([0; 1], \|\cdot\|_\infty)$ and regular histograms.

Lemma 2. *Let $\alpha \in (0; 1]$, $\delta, \epsilon, R > 0$ and $s \in \mathcal{H}_{\delta, \epsilon}(\alpha, R)$.*

For every $D \in \mathbb{N}$, denote by s_D the $L^2(\text{Leb})$ projection of s on the space of regular histograms with D pieces. Then, there exists a constant

$$C_1 = L(\alpha) R^{-\alpha-1} \epsilon^{2+\alpha-1} |x_1 - x_2|^{-1-\alpha-1} > 0$$

and $C_2 = R^2 \quad \beta_1 = 1 + \frac{1}{\alpha} \quad \beta_2 = 2\alpha$

such that for all $D > 0$,

$$\frac{C_1}{D^{\beta_1}} \leq \|s - s_D\|_{L^2(\text{Leb})}^2 \leq \frac{C_2}{D^{\beta_2}} . \quad (1)$$

Remark 2. The upper bound holds with any probability measure μ on \mathcal{X} instead of Leb, since

$$\ell(s, s_m) \leq \|s - s_m\|_\infty^2 \leq R^2 D_m^{-2\alpha} .$$

If **(Ad_ℓ)** holds, then

$$\ell(s, s_m) \geq c_{\min}^X \text{Leb}(\mathcal{X})^{-1} \|s - s_m\|_{L^2(\text{Leb})}^2 \geq c \|s - s_{D_m}\|_{L^2(\text{Leb})}^2$$

and thus the lower bound is still valid.

Remark 3. The lower bound in (1) cannot be improved, as shown in Sect. 2.3: for every $\alpha, R, \delta, \epsilon > 0$, there exists C'_1 such that for every D ,

$$\inf_{s \in \mathcal{H}_{\delta, \epsilon}(\alpha, R)} \left\{ \|s - s_D\|_{L^2(\text{Leb})}^2 \right\} \leq \frac{C'_1}{D^{1+\alpha-1}} .$$

2.2. Regular histograms in dimension k

We now generalize the previous result to subsets of \mathbb{R}^k . For the sake of simplicity, we assume that \mathcal{X} is a ball of $(\mathbb{R}^k, \|\cdot\|_\infty)$. Otherwise, if $\overset{\circ}{\mathcal{X}}$ is connex and non-empty, any non-constant continuous function s on \mathcal{X} is non-constant on some ball $B(s) \subset \mathcal{X}$. Then, we can apply Lemma 3 on $\mathcal{B}(s)$ in order to derive **(Ap)**. The constants $\delta, \epsilon > 0$ have to take into account the restriction $x_1, x_2 \in B(s) \subset \mathcal{X}$ in the definition of $\mathcal{H}_{\delta, \epsilon}(\alpha, R)$. When \mathcal{X} is a ball, this condition is automatically satisfied.

Lemma 3. Let \mathcal{X} be a non-empty closed ball of $(\mathbb{R}^k, \|\cdot\|_\infty)$ and $s \in \mathcal{H}_{\delta,\epsilon}(\alpha, R)$. Let $D > 0$ and consider a “regular” partition¹ (I_λ) of \mathcal{X} of pace D^{-1} . Let s_D be the piecewise constant function, defined on each piece I_λ of this partition by

$$s_D \equiv \frac{1}{\text{Leb}(I_\lambda)} \int_{I_\lambda} s(t) dt .$$

Then,

$$\int_{\mathcal{X}} (s(t) - s_D)^2 dt \geq L_{k,\alpha} \epsilon^{2+k\alpha^{-1}} \delta^{-1-k\alpha^{-1}} R^{1-k(1+\alpha^{-1})} \quad (2)$$

$$\times (D \vee \delta^{-1})^{-1-k\alpha^{-1}+(k-1)\alpha} .$$

Remark 4. 1. The number of pieces in the partition is not D but (approximately, depending on the shape of \mathcal{X}) $\text{Leb}(\mathcal{X})D^k$. Then, if \mathcal{X} has a lower bounded density w.r.t. Leb on \mathcal{X} , under the assumptions of Lemma 3, **(Ap)** is satisfied with $\beta_1 = k^{-1} + \alpha^{-1} - (k-1)k^{-1}\alpha$.
2. The following upper bound on the bias is straightforward:

$$\frac{1}{\text{Leb}(\mathcal{X})} \int_{\mathcal{X}} (s(t) - s_D)^2 dt \leq \|s - s_D\|_\infty^2 \leq R^2 D^{-2\alpha} .$$

3. When \mathcal{X} is not a ball of \mathbb{R}^k , we can use a general argument assuming only that for every $x_1, x_2 \in \overset{\circ}{\mathcal{X}}$, there is a path from x_1 to x_2 that has an η -enlargement in \mathcal{X} for some $\eta > 0$. We then obtain a weaker lower bound than (2), but which still implies **(Ap)**.

2.3. Optimality of the lower bound in $[0;1]$

When \mathcal{X} is a non-empty compact interval of \mathbb{R} , the exponent $1+\alpha^{-1}$ in Lemma 2 is unimprovable in the following sense. Without any loss of generality, we assume that $\mathcal{X} = [0; 1]$.

Lemma 4. Let $\mathcal{X} = [0; 1]$, $R > 0$, $\alpha \in (0; 1]$, $1 \geq \delta \geq (1 + \eta)D^{-1}$ (for some $\eta > 0$) and $L(\alpha)R[D\delta]D^{-1} \geq \epsilon > 0$.

$$\inf_{s \in \mathcal{H}_{\delta,\epsilon}(\alpha, R)} \left\{ \int_{\mathcal{X}} (s(t) - s_D(t))^2 dt \right\} \leq L(\alpha, \eta) R^{-\alpha^{-1}} \epsilon^{2+\alpha^{-1}} \delta^{-1-\alpha^{-1}} D^{-1-\alpha^{-1}} .$$

Remark 5. If $D \geq 2\delta^{-1}$, one can replace η by 1 and this upper bound is (up to some factor $L(\alpha)$) the same as the lower bound in Lemma 2.

¹When $\mathcal{X} = [0, 1]^k$ (which is equivalent to any closed ball, up to a translation and an homothety of \mathbb{R}^k), it is the partition $\left(\prod_{i=1}^k \left[\frac{j_i}{D}; \frac{j_i+1}{D} \right) \right)_{0 \leq j_1, \dots, j_k \leq D-1}$. For more general \mathcal{X} , it can be defined as the collection of non-empty intersections between \mathcal{X} and the family $\left(\prod_{i=1}^k \left[\frac{j_i}{D}; \frac{j_i+1}{D} \right) \right)_{j_1, \dots, j_k \in \mathbb{Z}}$.

Thus, the exponent $\beta_1 = 1 + \alpha^{-1}$ cannot be improved as long as we look for a uniform bound on $\mathcal{H}_{\delta,\epsilon}(\alpha, R)$. However, this does not mean that there exists a function $s \in \mathcal{H}(\alpha, R)$ approximated by regular histograms at the rate $D^{-1-\alpha^{-1}}$. To our knowledge, this question remains unsolved. Some references about this problem (and the equivalent one when the knots of the partition are no longer fixed) may be found in Burchard and Hale (4). See also the book of DeVore and Lorentz (5), in particular Chap. 12.

3. Simulation study

We consider in this section eight experiments (called S1000, S $\sqrt{0.1}$, S0.1, Svar2, Sqrt, His6, DopReg and Dop2bin) in which we have compared the same procedures as in Sect. 4, with the same benchmarks, but with only $N = 250$ samples for each experiment.

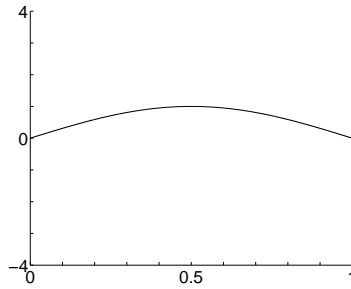
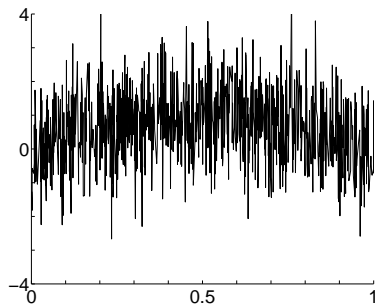
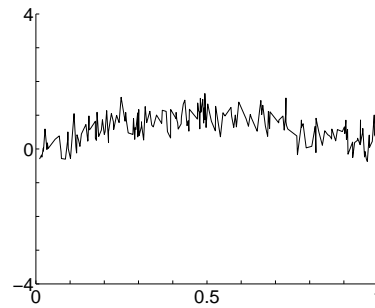
Data are generated according to

$$Y_i = s(X_i) + \sigma(X_i)\epsilon_i$$

with X_i i.i.d. uniform on $\mathcal{X} = [0; 1]$ and $\epsilon_i \sim \mathcal{N}(0, 1)$ independent from X_i . The experiments differ from

- the regression function s :
 - S1000, S $\sqrt{0.1}$, S0.1 and Svar2 have the same smooth function as S1 and S2, see Fig. 1.
 - Sqrt has $s(x) = \sqrt{x}$, which is smooth except around 0, see Fig. 6.
 - His6 has a regular histogram with 5 jumps (hence it belongs to the regular histogram model of dimension 6), see Fig. 8.
 - DopReg and Dop2bin have the Doppler function, as defined by Donoho and Johnstone (6), see Fig. 10.
- the noise level σ :
 - $\sigma(x) = 1$ for S1000, Sqrt, His6, DopReg and Dop2bin.
 - $\sigma(x) = \sqrt{0.1}$ for S $\sqrt{0.1}$.
 - $\sigma(x) = 0.1$ for S0.1.
 - $\sigma(x) = \mathbb{1}_{x \geq 1/2}$ for Svar2.
- the sample size n :
 - $n = 200$ for S $\sqrt{0.1}$, S0.1, Svar2, Sqrt and His6.
 - $n = 1000$ for S1000.
 - $n = 2048$ for DopReg and Dop2bin.
- the family of models: with the notations introduced in Sect. 4,
 - for S1000, S $\sqrt{0.1}$, S0.1, Sqrt and His6, we use the “regular” collection, as for S1:

$$\mathcal{M}_n = \left\{ 1, \dots, \left\lfloor \frac{n}{\ln(n)} \right\rfloor \right\} .$$

FIG 1. $s(x) = \sin(\pi x)$ FIG 2. Data sample for $S1000$ FIG 3. Data sample for $S\sqrt{0.1}$

- for Svar2, we use the “regular with two bin sizes” collection, as for S2:

$$\mathcal{M}_n = \{1\} \cup \left\{ 1, \dots, \left\lfloor \frac{n}{2 \ln(n)} \right\rfloor \right\}^2 .$$

- for DopReg, we use the “regular dyadic” collection, as for HSd1:

$$\mathcal{M}_n = \{2^k \text{ s.t. } 0 \leq k \leq \ln_2(n) - 1\} .$$

- for Dop2bin, we use the “regular dyadic with two bin sizes” collection, as for HSd2:

$$\mathcal{M}_n = \{1\} \cup \{2^k \text{ s.t. } 0 \leq k \leq \ln_2(n) - 2\}^2 .$$

Notice that contrary to HSd2, Dop2bin is an homoscedastic problem. The interest of considering two bin sizes for it is that the smoothness of the Doppler function is quite different for small x and for $x \geq 1/2$.

Instances of data sets for each experiment are given in Fig. 2–5, 7, 9 and 11.

Compared to S1, S2, HSd1 and HSd2, these eight experiments consider larger signal-to-noise ratio data (S1000, $S\sqrt{0.1}$, S0.1), another kind of heteroscedasticity (Svar2) and other regression functions, with different kinds of unsmoothness (Sqrt, His6, DopReg and Dop2bin).

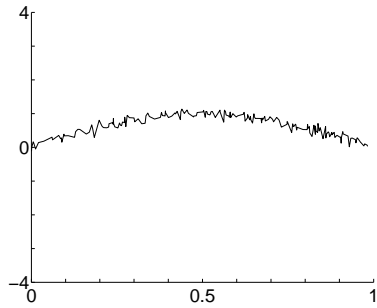


FIG 4. Data sample for $S0.1$

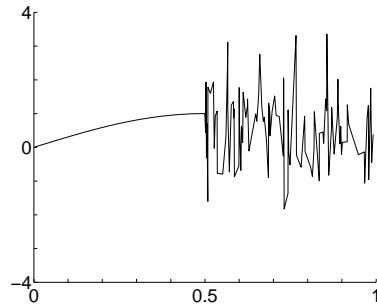


FIG 5. Data sample for $Svar2$

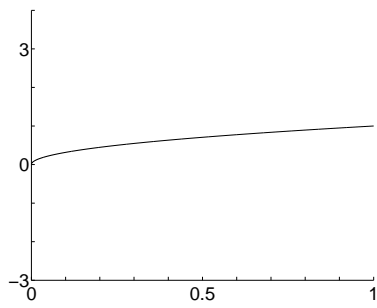


FIG 6. $s(x) = \sqrt{x}$

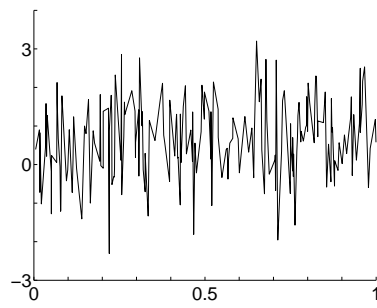


FIG 7. Data sample for $Sqrt$

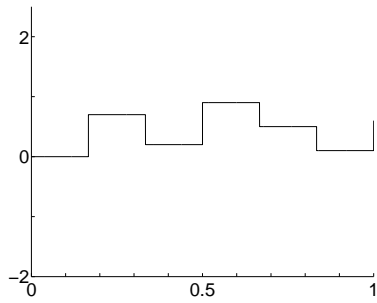


FIG 8. $s(x) = His_6(x)$

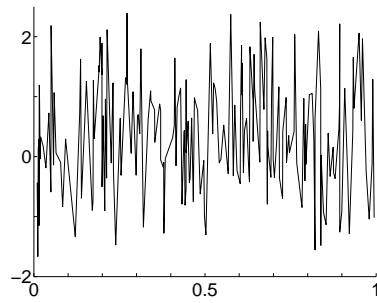
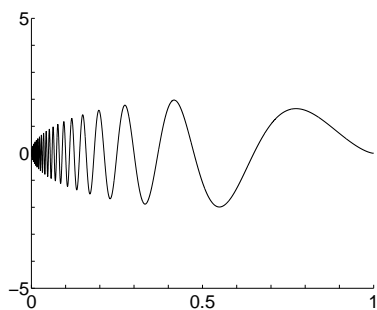
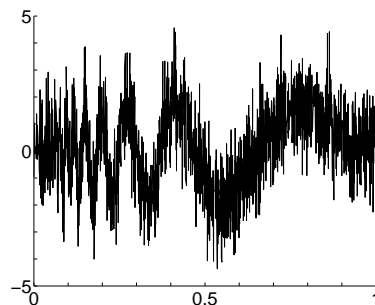


FIG 9. Data sample for $His6$

FIG 10. $s(x) = \text{Doppler}(x)$ (see (6))FIG 11. Data sample for *DopReg* and *Dop2bin*

We consider for each of these experiments the same algorithms as in Sect. 4, and we added to them:

1. Mal^{*}, which is Mallows' C_p penalty with the true value of the variance: $\text{pen}(m) = 2\mathbb{E}[\sigma^2(X)] D_m n^{-1}$. Although it can not be used on real data sets, it is an interesting point of comparison, which does not have possible weaknesses coming from the variance estimator $\hat{\sigma}^2$.
2. penPoi, which is the Poisson (1) penalty, with $C = C_W = 1$.
3. for each of the exchangeable penalties (except penLoo), we also considered Monte-Carlo approximations of their exact values, drawing 20 resampling weights at random. These procedures are denoted by a (20) following the shortened names.
4. penV-F, for $V \in \{2, 5, 10, 20\}$, which are the V -fold subsampling version of Algorithm. 2. The corresponding results come from (3), where these penalties are properly defined and studied.

In the case of Mal^{*}, penPoi and penVF, we also tested overpenalization by multiplying these penalties by 5/4. The corresponding algorithms are denoted by a + sign, as in the main paper.

Our estimates of C_{or} (and uncertainties for these estimates) for the procedures we consider are reported in Tab. 1 to 3 (we report here again the results for S1, S2, HSd1 and HSd2 to make comparisons easier). On the last line of these Tables, we also report

$$\frac{\mathbb{E}[\inf_{m \in \mathcal{M}_n} \ell(s, \hat{s}_m)]}{\inf_{m \in \mathcal{M}_n} \{\mathbb{E}[\ell(s, \hat{s}_m)]\}} = \frac{C'_{\text{or}}}{C_{\text{or}}} \quad \text{where} \quad C'_{\text{or}} := \frac{\mathbb{E}[\ell(s, \hat{s}_m)]}{\inf_{m \in \mathcal{M}_n} \{\mathbb{E}[\ell(s, \hat{s}_m)]\}}$$

is the leading constant which appear in most of the classical oracle inequalities. Notice that C'_{or} is always smaller than C_{or} .

The comparison between Mallows', VFCV and Resampling Penalization is quite the same: in “easy” homoscedastic frameworks (S1000, $S\sqrt{0.1}$, S0.1), their performances are similar. A harder problem such as Svar2 (which is het-

eroscedastic but different from S2) make Mallows' fail whereas the two others only get a bit worse.

As expected, taking n larger (S1000) or σ smaller ($S\sqrt{0.1}$ and S0.1) make the constant C_{or} closer to 1. Notice also that overpenalization often improves the quality of the algorithm (but not always: see DopReg and S0.1). We have for instance $C_{\text{or}}(\text{penLoo}) < C_{\text{or}}(\text{penRho}) < C_{\text{or}}(\text{penRho+})$ in S0.1 (with only small differences), although penLoo may slightly underpenalize. This also shows that the factor 5/4 is not optimal in general (actually, it is certainly not optimal for all these 12 experiments, it was fixed arbitrarily and *a priori*).

We can also compare the exact exchangeable resampling penalties with their Monte-Carlo approximations, or with non-exchangeable ones (with V -fold penalties). Exact formulas with exchangeable schemes (Rad, Rho) are better than the other ones in most of the experiments (and sometimes significantly, for instance in S2 or HSd1). But the difference is not so large, making *e.g.* penLoo worse than penRad(20) for HSd1. The extreme case is Sqrt where penRad(20) performs (unsignificantly) better than penRad! This shows that such an approximation could be used in practice without loosing too much (actually, choosing the right overpenalization factor is much more important than making exact computations). On the other hand, it is quite unclear whether one should prefer Monte-Carlo approximations or V -fold subsampling schemes (taking $V = 20$ to make complexities comparable). According to the experiments, pen20-F can be better (S2), worse (HSd1) or not significantly² different (S1,HSd2) from penRad(20).

To conclude, these eight experiments confirm the strenghts of RP already pointed out in Sect. 5, and show that the assumptions of Thm. 1 (or even the ones of Lemma 7 are not necessary for the resampling penalties to be efficient.

For the sake of completeness, we also reported the results for the twelve experiments in terms of the other benchmark

$$C_{\text{path-or}} := \mathbb{E} \left[\frac{\ell(s, \widehat{s}_m)}{\inf_{m \in \mathcal{M}_n} \ell(s, \widehat{s}_m)} \right]$$

in Tab. 4 to Tab. 6. They are indeed quite similar to the previous ones.

4. Complete proofs of concentration inequalities

We here given complete proofs and key results of our concentration inequalities, most of them coming from (1).

complete proof of Prop. 12. According to the explicit expressions (46) and (47), $\widehat{p}_1(m)$ and $p_2(m)$ are both U-statistics of order 2 conditionally to $(\mathbb{1}_{X_i \in I_\lambda})_{(i,\lambda)}$.

²Given the number of experiments made here, we should also mention that we do not take into account multiplicity when we say that we have found "significant" differences. Then, the situation observed between pen20-F and penRad(20) is typically the one of two procedures which have exactly the same performance.

TABLE 1
Accuracy indexes C_{or} for experiments S1, S2, HSd1 and HSd2 ($N = 1000$). Uncertainties reported are empirical standard deviations divided by \sqrt{N} . In each column, the more accurate data-dependent algorithms (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\text{pen}_{\text{id}}]$ and $\mathbb{E}[\text{pen}_{\text{id}}] +$ when they have better or comparable performances.

Experiment	S1	S2	HSd1	HSd2
s	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$	HeaviSine	HeaviSine
$\sigma(x)$	1	x	1	x
n (sample size)	200	200	2048	2048
\mathcal{M}_n	regular	2 bin sizes	dyadic, regular	dyadic, 2 bin sizes
Mal	1.928 ± 0.04	3.687 ± 0.07	1.015 ± 0.003	1.373 ± 0.010
Mal+	1.800 ± 0.03	3.173 ± 0.07	1.002 ± 0.003	1.411 ± 0.008
Mal*	2.028 ± 0.04	2.657 ± 0.06	1.044 ± 0.004	1.513 ± 0.005
Mal*+	1.827 ± 0.03	2.437 ± 0.05	1.004 ± 0.003	1.548 ± 0.003
$\mathbb{E}[\text{pen}_{\text{id}}]$	1.919 ± 0.03	2.296 ± 0.05	1.028 ± 0.004	1.102 ± 0.004
$\mathbb{E}[\text{pen}_{\text{id}}] +$	1.792 ± 0.03	2.028 ± 0.04	1.003 ± 0.003	1.089 ± 0.004
2-FCV	2.078 ± 0.04	2.542 ± 0.05	1.002 ± 0.003	1.184 ± 0.004
5-FCV	2.137 ± 0.04	2.582 ± 0.06	1.014 ± 0.003	1.115 ± 0.005
10-FCV	2.097 ± 0.04	2.603 ± 0.06	1.021 ± 0.003	1.109 ± 0.004
20-FCV	2.088 ± 0.04	2.578 ± 0.06	1.029 ± 0.004	1.105 ± 0.004
LOO	2.077 ± 0.04	2.593 ± 0.06	1.034 ± 0.004	1.105 ± 0.004
penRad	1.973 ± 0.04	2.485 ± 0.06	1.018 ± 0.003	1.102 ± 0.004
penRho	1.982 ± 0.04	2.502 ± 0.06	1.018 ± 0.003	1.103 ± 0.004
penLoo	2.080 ± 0.04	2.593 ± 0.06	1.034 ± 0.004	1.105 ± 0.004
penEfr	2.597 ± 0.07	3.152 ± 0.07	1.067 ± 0.005	1.114 ± 0.005
penPoi	2.650 ± 0.07	3.191 ± 0.07	1.067 ± 0.005	1.115 ± 0.005
penRad(20)	2.027 ± 0.04	2.762 ± 0.06	1.021 ± 0.003	1.107 ± 0.004
penRho(20)	2.039 ± 0.04	2.706 ± 0.06	1.020 ± 0.004	1.107 ± 0.004
penEfr(20)	2.688 ± 0.07	3.339 ± 0.07	1.065 ± 0.005	1.119 ± 0.006
penPoi(20)	2.627 ± 0.07	3.405 ± 0.07	1.067 ± 0.005	1.120 ± 0.006
pen2-F	2.578 ± 0.06	3.061 ± 0.07	1.038 ± 0.004	1.103 ± 0.004
pen5-F	2.219 ± 0.05	2.750 ± 0.06	1.037 ± 0.004	1.104 ± 0.004
pen10-F	2.121 ± 0.04	2.653 ± 0.06	1.034 ± 0.004	1.104 ± 0.004
pen20-F	2.085 ± 0.04	2.639 ± 0.06	1.034 ± 0.004	1.105 ± 0.004
penRad+	1.799 ± 0.03	2.137 ± 0.05	1.002 ± 0.003	1.095 ± 0.004
penRho+	1.798 ± 0.03	2.142 ± 0.05	1.002 ± 0.003	1.095 ± 0.004
penLoo+	1.844 ± 0.03	2.215 ± 0.05	1.004 ± 0.003	1.096 ± 0.004
penEfr+	2.016 ± 0.05	2.605 ± 0.06	1.011 ± 0.003	1.097 ± 0.004
penPoi+	2.039 ± 0.05	2.620 ± 0.06	1.011 ± 0.003	1.097 ± 0.004
pen2-F+	2.175 ± 0.05	2.748 ± 0.06	1.011 ± 0.003	1.106 ± 0.004
pen5-F+	1.913 ± 0.03	2.378 ± 0.05	1.006 ± 0.003	1.102 ± 0.004
pen10-F+	1.872 ± 0.03	2.285 ± 0.05	1.005 ± 0.003	1.098 ± 0.004
pen20-F+	1.898 ± 0.03	2.254 ± 0.05	1.004 ± 0.003	1.098 ± 0.004
$C'_{\text{or}}/C_{\text{or}}$	0.768	0.753	0.999	0.854

TABLE 2

Accuracy indexes C_{or} for experiments S1000, $S\sqrt{0.1}$, S0.1 and Svar2 ($N = 250$).
 Uncertainties reported are empirical standard deviations divided by \sqrt{N} . In each column, the more accurate data-dependent algorithms (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\text{pen}_{\text{id}}]$ and $\mathbb{E}[\text{pen}_{\text{id}}] +$ when they have better or comparable performances.

Experiment	S1000	$S\sqrt{0.1}$	S0.1	Svar2
s	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$
$\sigma(x)$	1	$\sqrt{0.1}$	0.1	$\mathbb{1}_{x \geq 1/2}$
n (sample size)	1000	200	200	200
\mathcal{M}_n	regular	regular	regular	2 bin sizes
Mal	1.667 ± 0.04	1.611 ± 0.03	1.400 ± 0.02	5.643 ± 0.22
Mal+	1.619 ± 0.03	1.593 ± 0.03	1.426 ± 0.02	4.647 ± 0.22
Mal*	1.745 ± 0.05	1.925 ± 0.03	3.204 ± 0.05	4.481 ± 0.21
Mal*+	1.617 ± 0.03	2.073 ± 0.04	3.641 ± 0.07	3.544 ± 0.17
$\mathbb{E}[\text{pen}_{\text{id}}]$	1.745 ± 0.05	1.571 ± 0.03	1.373 ± 0.02	2.409 ± 0.13
$\mathbb{E}[\text{pen}_{\text{id}}] +$	1.617 ± 0.03	1.554 ± 0.03	1.392 ± 0.02	2.005 ± 0.10
2-FCV	1.668 ± 0.04	1.663 ± 0.04	1.394 ± 0.02	2.960 ± 0.15
5-FCV	1.756 ± 0.07	1.693 ± 0.04	1.393 ± 0.02	2.950 ± 0.16
10-FCV	1.746 ± 0.04	1.684 ± 0.04	1.385 ± 0.02	2.681 ± 0.14
20-FCV	1.774 ± 0.05	1.645 ± 0.03	1.382 ± 0.02	2.742 ± 0.16
LOO	1.768 ± 0.05	1.639 ± 0.04	1.379 ± 0.02	2.641 ± 0.15
penRad	1.748 ± 0.05	1.609 ± 0.03	1.405 ± 0.02	2.510 ± 0.15
penRho	1.748 ± 0.05	1.619 ± 0.03	1.404 ± 0.02	2.518 ± 0.15
penLoo	1.776 ± 0.05	1.641 ± 0.04	1.379 ± 0.02	2.656 ± 0.15
penEfr	1.813 ± 0.05	1.888 ± 0.05	1.417 ± 0.02	3.451 ± 0.20
penPoi	1.813 ± 0.05	1.922 ± 0.05	1.419 ± 0.02	3.548 ± 0.21
penRad(20)	1.794 ± 0.05	1.636 ± 0.04	1.415 ± 0.02	2.966 ± 0.17
penRho(20)	1.725 ± 0.05	1.641 ± 0.04	1.397 ± 0.02	2.961 ± 0.17
penEfr(20)	1.808 ± 0.05	1.875 ± 0.05	1.410 ± 0.02	3.974 ± 0.22
penPoi(20)	1.868 ± 0.06	1.908 ± 0.05	1.414 ± 0.02	3.866 ± 0.21
pen2-F	2.066 ± 0.08	1.809 ± 0.05	1.390 ± 0.02	3.209 ± 0.18
pen5-F	1.816 ± 0.05	1.638 ± 0.04	1.400 ± 0.02	2.749 ± 0.15
pen10-F	1.783 ± 0.05	1.706 ± 0.04	1.374 ± 0.02	2.598 ± 0.15
pen20-F	1.801 ± 0.05	1.657 ± 0.03	1.385 ± 0.02	2.684 ± 0.15
penRad+	1.619 ± 0.03	1.574 ± 0.03	1.417 ± 0.02	2.232 ± 0.12
penRho+	1.619 ± 0.03	1.578 ± 0.03	1.417 ± 0.02	2.243 ± 0.12
penLoo+	1.626 ± 0.03	1.587 ± 0.03	1.401 ± 0.02	2.349 ± 0.13
penEfr+	1.636 ± 0.03	1.670 ± 0.04	1.407 ± 0.02	2.614 ± 0.16
penPoi+	1.636 ± 0.03	1.669 ± 0.04	1.420 ± 0.02	2.668 ± 0.17
pen2-F+	1.809 ± 0.05	1.714 ± 0.04	1.416 ± 0.02	2.808 ± 0.16
pen5-F+	1.683 ± 0.04	1.616 ± 0.03	1.399 ± 0.02	2.460 ± 0.14
pen10-F+	1.627 ± 0.04	1.613 ± 0.03	1.385 ± 0.02	2.398 ± 0.14
pen20-F+	1.644 ± 0.04	1.583 ± 0.03	1.390 ± 0.02	2.316 ± 0.13
$C'_{\text{or}}/C_{\text{or}}$	0.8	0.801	0.816	0.779

TABLE 3

Accuracy indexes C_{or} for experiments Sqrt, His6, DopReg and Dop2bin ($N = 250$).
 Uncertainties reported are empirical standard deviations divided by \sqrt{N} . In each column, the more accurate data-dependent algorithms (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\text{pen}_{\text{id}}]$ and $\mathbb{E}[\text{pen}_{\text{id}}] +$ when they have better or comparable performances.

Experiment	Sqrt	His6	DopReg	Dop2bin
s	$\sqrt{\cdot}$	His6	Doppler	Doppler
$\sigma(x)$	1	1	1	1
n (sample size)	200	200	2048	2048
\mathcal{M}_n	regular	regular	dyadic, regular	dyadic, 2 bin sizes
Mal	2.295 ± 0.11	1.969 ± 0.11	1.039 ± 0.01	1.052 ± 0.01
Mal+	1.989 ± 0.08	1.799 ± 0.09	1.090 ± 0.00	1.047 ± 0.01
Mal*	2.483 ± 0.12	2.021 ± 0.11	1.013 ± 0.01	1.061 ± 0.01
Mal*+	2.075 ± 0.09	1.836 ± 0.10	1.070 ± 0.00	1.041 ± 0.01
$\mathbb{E}[\text{pen}_{\text{id}}]$	2.365 ± 0.11	1.805 ± 0.10	1.025 ± 0.01	1.056 ± 0.01
$\mathbb{E}[\text{pen}_{\text{id}}] +$	2.012 ± 0.09	1.632 ± 0.08	1.083 ± 0.00	1.040 ± 0.01
2-FCV	2.489 ± 0.12	2.788 ± 0.13	1.097 ± 0.00	1.165 ± 0.01
5-FCV	2.777 ± 0.16	2.316 ± 0.12	1.064 ± 0.01	1.049 ± 0.01
10-FCV	2.571 ± 0.13	2.074 ± 0.11	1.043 ± 0.01	1.051 ± 0.01
20-FCV	2.561 ± 0.12	2.071 ± 0.11	1.034 ± 0.01	1.053 ± 0.01
LOO	2.695 ± 0.14	2.059 ± 0.11	1.026 ± 0.01	1.058 ± 0.01
penRad	2.396 ± 0.11	1.884 ± 0.10	1.043 ± 0.01	1.055 ± 0.01
penRho	2.448 ± 0.12	1.907 ± 0.11	1.043 ± 0.01	1.055 ± 0.01
penLoo	2.695 ± 0.14	2.063 ± 0.12	1.026 ± 0.01	1.058 ± 0.01
penEfr	3.468 ± 0.22	2.721 ± 0.16	1.030 ± 0.01	1.064 ± 0.01
penPoi	3.525 ± 0.22	2.878 ± 0.18	1.040 ± 0.01	1.064 ± 0.01
penRad(20)	2.361 ± 0.10	2.083 ± 0.11	1.044 ± 0.01	1.058 ± 0.01
penRho(20)	2.499 ± 0.12	2.039 ± 0.11	1.046 ± 0.01	1.057 ± 0.01
penEfr(20)	3.558 ± 0.22	2.928 ± 0.16	1.036 ± 0.01	1.058 ± 0.01
penPoi(20)	3.588 ± 0.21	2.899 ± 0.15	1.033 ± 0.01	1.066 ± 0.01
pen2-F	4.088 ± 0.23	3.210 ± 0.14	1.048 ± 0.01	1.062 ± 0.01
pen5-F	3.024 ± 0.18	2.485 ± 0.13	1.033 ± 0.01	1.055 ± 0.01
pen10-F	3.009 ± 0.18	2.192 ± 0.12	1.029 ± 0.01	1.056 ± 0.01
pen20-F	2.723 ± 0.14	2.150 ± 0.12	1.031 ± 0.01	1.056 ± 0.01
penRad+	2.036 ± 0.09	1.746 ± 0.09	1.092 ± 0.00	1.058 ± 0.01
penRho+	2.053 ± 0.09	1.747 ± 0.09	1.091 ± 0.00	1.059 ± 0.01
penLoo+	2.152 ± 0.10	1.858 ± 0.10	1.082 ± 0.00	1.048 ± 0.01
penEfr+	2.205 ± 0.11	1.924 ± 0.11	1.056 ± 0.01	1.057 ± 0.01
penPoi+	2.249 ± 0.11	2.017 ± 0.11	1.056 ± 0.01	1.058 ± 0.01
pen2-F+	3.015 ± 0.17	2.728 ± 0.12	1.084 ± 0.00	1.084 ± 0.01
pen5-F+	2.409 ± 0.13	2.080 ± 0.09	1.080 ± 0.00	1.063 ± 0.01
pen10-F+	2.305 ± 0.11	1.869 ± 0.09	1.082 ± 0.00	1.050 ± 0.01
pen20-F+	2.180 ± 0.10	1.832 ± 0.09	1.079 ± 0.00	1.052 ± 0.01
$C'_{\text{or}}/C_{\text{or}}$	0.795	0.996	0.998	0.977

TABLE 4

Accuracy indexes $C_{\text{path-or}}$ for experiments S1, S2, HSd1 and HSd2 ($N = 1000$).
 Uncertainties reported are empirical standard deviations divided by \sqrt{N} . In each column, the more accurate data-dependent algorithms (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\text{pen}_{\text{id}}]$ and $\mathbb{E}[\text{pen}_{\text{id}}] +$ when they have better or comparable performances.

Experiment	S1	S2	HSd1	HSd2
s	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$	HeaviSine	HeaviSine
$\sigma(x)$	1	x	1	x
n (sample size)	200	200	2048	2048
\mathcal{M}_n	regular	2 bin sizes	dyadic, regular	dyadic, 2 bin sizes
Mal	2.064 ± 0.04	4.129 ± 0.10	1.015 ± 0.002	1.316 ± 0.010
Mal+	1.921 ± 0.03	3.500 ± 0.09	1.002 ± 0.001	1.354 ± 0.008
Mal*	2.168 ± 0.04	2.907 ± 0.07	1.045 ± 0.003	1.453 ± 0.006
Mal*+	1.941 ± 0.03	2.645 ± 0.06	1.004 ± 0.001	1.487 ± 0.005
$\mathbb{E}[\text{pen}_{\text{id}}]$	2.053 ± 0.04	2.458 ± 0.06	1.029 ± 0.003	1.050 ± 0.002
$\mathbb{E}[\text{pen}_{\text{id}}] +$	1.903 ± 0.03	2.142 ± 0.04	1.003 ± 0.001	1.038 ± 0.002
2-FCV	2.230 ± 0.05	2.755 ± 0.06	1.002 ± 0.001	1.134 ± 0.004
5-FCV	2.290 ± 0.05	2.827 ± 0.08	1.014 ± 0.002	1.064 ± 0.003
10-FCV	2.237 ± 0.05	2.832 ± 0.08	1.021 ± 0.002	1.057 ± 0.002
20-FCV	2.225 ± 0.05	2.794 ± 0.07	1.029 ± 0.003	1.054 ± 0.002
LOO	2.212 ± 0.05	2.832 ± 0.08	1.034 ± 0.003	1.053 ± 0.002
penRad	2.102 ± 0.04	2.705 ± 0.07	1.018 ± 0.002	1.051 ± 0.002
penRho	2.111 ± 0.04	2.726 ± 0.07	1.018 ± 0.002	1.051 ± 0.002
penLoo	2.215 ± 0.05	2.832 ± 0.08	1.034 ± 0.003	1.053 ± 0.002
penEfr	2.818 ± 0.08	3.468 ± 0.09	1.067 ± 0.004	1.062 ± 0.003
penPoi	2.874 ± 0.09	3.522 ± 0.09	1.067 ± 0.004	1.062 ± 0.003
penRad(20)	2.148 ± 0.04	3.000 ± 0.08	1.022 ± 0.002	1.056 ± 0.002
penRho(20)	2.159 ± 0.04	2.941 ± 0.08	1.020 ± 0.002	1.055 ± 0.002
penEfr(20)	2.899 ± 0.08	3.695 ± 0.09	1.065 ± 0.004	1.066 ± 0.004
penPoi(20)	2.842 ± 0.08	3.807 ± 0.10	1.068 ± 0.004	1.067 ± 0.004
pen2-F	2.770 ± 0.07	3.340 ± 0.08	1.039 ± 0.003	1.052 ± 0.003
pen5-F	2.383 ± 0.06	2.982 ± 0.08	1.038 ± 0.003	1.053 ± 0.002
pen10-F	2.256 ± 0.05	2.867 ± 0.07	1.035 ± 0.003	1.053 ± 0.002
pen20-F	2.219 ± 0.05	2.869 ± 0.08	1.035 ± 0.003	1.053 ± 0.002
penRad+	1.917 ± 0.03	2.304 ± 0.06	1.002 ± 0.001	1.045 ± 0.002
penRho+	1.915 ± 0.03	2.308 ± 0.06	1.002 ± 0.001	1.045 ± 0.002
penLoo+	1.959 ± 0.03	2.397 ± 0.06	1.004 ± 0.001	1.045 ± 0.002
penEfr+	2.155 ± 0.05	2.841 ± 0.08	1.011 ± 0.002	1.046 ± 0.002
penPoi+	2.179 ± 0.05	2.855 ± 0.08	1.011 ± 0.002	1.045 ± 0.002
pen2-F+	2.328 ± 0.05	2.979 ± 0.07	1.011 ± 0.002	1.056 ± 0.003
pen5-F+	2.050 ± 0.04	2.540 ± 0.06	1.006 ± 0.001	1.052 ± 0.002
pen10-F+	1.997 ± 0.03	2.436 ± 0.05	1.005 ± 0.001	1.048 ± 0.002
pen20-F+	2.018 ± 0.04	2.416 ± 0.06	1.004 ± 0.001	1.047 ± 0.002

TABLE 5

Accuracy indexes $C_{\text{path-or}}$ for experiments S1000, $S\sqrt{0.1}$, S0.1 and Svar2 ($N = 250$).
 Uncertainties reported are empirical standard deviations divided by \sqrt{N} . In each column, the more accurate data-dependent algorithms (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\text{pen}_{\text{id}}]$ and $\mathbb{E}[\text{pen}_{\text{id}}] +$ when they have better or comparable performances.

Experiment	S1000	$S\sqrt{0.1}$	S0.1	Svar2
s	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$	$\sin(\pi \cdot)$
$\sigma(x)$	1	$\sqrt{0.1}$	0.1	$\mathbb{1}_{x \geq 1/2}$
n (sample size)	1000	200	200	200
\mathcal{M}_n	regular	regular	regular	2 bin sizes
Mal	1.704 ± 0.04	1.654 ± 0.03	1.407 ± 0.02	7.212 ± 0.40
Mal+	1.670 ± 0.03	1.636 ± 0.03	1.436 ± 0.02	5.740 ± 0.34
Mal*	1.793 ± 0.04	2.018 ± 0.04	3.273 ± 0.06	5.597 ± 0.33
Mal*+	1.664 ± 0.03	2.175 ± 0.05	3.719 ± 0.08	4.284 ± 0.25
$\mathbb{E}[\text{pen}_{\text{id}}]$	1.793 ± 0.04	1.611 ± 0.03	1.378 ± 0.01	2.785 ± 0.19
$\mathbb{E}[\text{pen}_{\text{id}}] +$	1.194 ± 0.02	1.177 ± 0.02	1.128 ± 0.01	1.337 ± 0.07
2-FCV	1.721 ± 0.04	1.723 ± 0.04	1.400 ± 0.02	3.507 ± 0.19
5-FCV	1.801 ± 0.06	1.740 ± 0.04	1.399 ± 0.02	3.486 ± 0.24
10-FCV	1.802 ± 0.05	1.735 ± 0.04	1.388 ± 0.02	3.149 ± 0.20
20-FCV	1.832 ± 0.05	1.687 ± 0.03	1.388 ± 0.02	3.257 ± 0.23
LOO	1.815 ± 0.05	1.685 ± 0.04	1.385 ± 0.01	3.127 ± 0.24
penRad	1.796 ± 0.05	1.655 ± 0.04	1.411 ± 0.02	2.932 ± 0.22
penRho	1.796 ± 0.05	1.666 ± 0.04	1.409 ± 0.02	2.951 ± 0.23
penLoo	1.825 ± 0.05	1.687 ± 0.04	1.385 ± 0.01	3.152 ± 0.24
penEfr	1.865 ± 0.05	1.941 ± 0.06	1.423 ± 0.02	4.181 ± 0.31
penPoi	1.865 ± 0.05	1.972 ± 0.06	1.425 ± 0.02	4.310 ± 0.32
penRad(20)	1.836 ± 0.05	1.675 ± 0.04	1.423 ± 0.02	3.650 ± 0.29
penRho(20)	1.768 ± 0.05	1.689 ± 0.04	1.403 ± 0.02	3.567 ± 0.27
penEfr(20)	1.836 ± 0.05	1.924 ± 0.05	1.416 ± 0.02	4.854 ± 0.34
penPoi(20)	1.924 ± 0.07	1.974 ± 0.06	1.418 ± 0.02	4.678 ± 0.33
pen2-F	2.108 ± 0.07	1.864 ± 0.05	1.394 ± 0.02	3.839 ± 0.27
pen5-F	1.852 ± 0.05	1.675 ± 0.04	1.404 ± 0.02	3.237 ± 0.23
pen10-F	1.812 ± 0.05	1.767 ± 0.04	1.381 ± 0.01	3.093 ± 0.23
pen20-F	1.839 ± 0.05	1.706 ± 0.03	1.391 ± 0.01	3.123 ± 0.23
penRad+	1.665 ± 0.03	1.615 ± 0.03	1.427 ± 0.02	2.502 ± 0.15
penRho+	1.665 ± 0.03	1.619 ± 0.03	1.428 ± 0.02	2.511 ± 0.15
penLoo+	1.673 ± 0.03	1.624 ± 0.03	1.409 ± 0.02	2.659 ± 0.18
penEfr+	1.683 ± 0.03	1.730 ± 0.04	1.413 ± 0.02	3.098 ± 0.25
penPoi+	1.683 ± 0.03	1.729 ± 0.04	1.427 ± 0.02	3.133 ± 0.25
pen2-F+	1.852 ± 0.05	1.765 ± 0.05	1.420 ± 0.02	3.336 ± 0.23
pen5-F+	1.732 ± 0.04	1.664 ± 0.03	1.408 ± 0.02	2.890 ± 0.22
pen10-F+	1.663 ± 0.04	1.657 ± 0.03	1.394 ± 0.02	2.810 ± 0.21
pen20-F+	1.680 ± 0.04	1.623 ± 0.03	1.397 ± 0.01	2.657 ± 0.19

TABLE 6

Accuracy indexes $C_{\text{path-or}}$ for experiments Sqrt, His6, DopReg and Dop2bin ($N = 250$).
 Uncertainties reported are empirical standard deviations divided by \sqrt{N} . In each column, the
 more accurate data-dependent algorithms (taking the uncertainty into account) are bolded,
 as well as $\mathbb{E}[\text{pen}_{\text{id}}]$ and $\mathbb{E}[\text{pen}_{\text{id}}] +$ when they have better or comparable performances.

Experiment	Sqrt	His6	DopReg	Dop2bin
s	$\sqrt{\cdot}$	His6	Doppler	Doppler
$\sigma(x)$	1	1	1	1
n (sample size)	200	200	2048	2048
\mathcal{M}_n	regular	regular	dyadic, regular	dyadic, 2 bin sizes
Mal	2.557 ± 0.12	2.356 ± 0.18	1.040 ± 0.00	1.049 ± 0.00
Mal+	2.232 ± 0.10	2.041 ± 0.12	1.094 ± 0.00	1.045 ± 0.01
Mal*	2.838 ± 0.15	2.533 ± 0.21	1.013 ± 0.00	1.057 ± 0.00
Mal*+	2.349 ± 0.11	2.168 ± 0.16	1.073 ± 0.00	1.038 ± 0.00
$\mathbb{E}[\text{pen}_{\text{id}}]$	2.678 ± 0.14	2.182 ± 0.17	1.026 ± 0.00	1.053 ± 0.00
$\mathbb{E}[\text{pen}_{\text{id}}] +$	1.348 ± 0.07	1.230 ± 0.06	1.050 ± 0.00	1.038 ± 0.00
2-FCV	2.974 ± 0.17	3.713 ± 0.25	1.100 ± 0.00	1.164 ± 0.01
5-FCV	3.209 ± 0.21	2.977 ± 0.24	1.066 ± 0.00	1.046 ± 0.00
10-FCV	2.912 ± 0.16	2.639 ± 0.21	1.045 ± 0.00	1.047 ± 0.00
20-FCV	2.889 ± 0.15	2.584 ± 0.20	1.035 ± 0.00	1.050 ± 0.00
LOO	3.061 ± 0.17	2.568 ± 0.21	1.027 ± 0.00	1.055 ± 0.00
penRad	2.708 ± 0.13	2.272 ± 0.19	1.044 ± 0.00	1.052 ± 0.00
penRho	2.755 ± 0.14	2.291 ± 0.19	1.044 ± 0.00	1.052 ± 0.00
penLoo	3.063 ± 0.17	2.571 ± 0.21	1.027 ± 0.00	1.055 ± 0.00
penEfr	4.091 ± 0.32	3.560 ± 0.29	1.031 ± 0.01	1.061 ± 0.00
penPoi	4.126 ± 0.32	3.790 ± 0.32	1.042 ± 0.01	1.061 ± 0.00
penRad(20)	2.690 ± 0.13	2.685 ± 0.24	1.046 ± 0.00	1.055 ± 0.00
penRho(20)	2.822 ± 0.14	2.560 ± 0.20	1.048 ± 0.00	1.054 ± 0.00
penEfr(20)	4.103 ± 0.29	3.931 ± 0.32	1.038 ± 0.01	1.055 ± 0.00
penPoi(20)	4.107 ± 0.26	3.753 ± 0.28	1.034 ± 0.01	1.062 ± 0.01
pen2-F	5.062 ± 0.37	4.462 ± 0.30	1.050 ± 0.00	1.059 ± 0.01
pen5-F	3.595 ± 0.25	3.458 ± 0.28	1.034 ± 0.00	1.052 ± 0.00
pen10-F	3.445 ± 0.22	2.744 ± 0.21	1.031 ± 0.00	1.053 ± 0.00
pen20-F	3.120 ± 0.17	2.670 ± 0.21	1.032 ± 0.00	1.053 ± 0.00
penRad+	2.291 ± 0.11	2.018 ± 0.14	1.095 ± 0.00	1.056 ± 0.01
penRho+	2.317 ± 0.11	2.019 ± 0.14	1.095 ± 0.00	1.057 ± 0.01
penLoo+	2.437 ± 0.12	2.218 ± 0.18	1.085 ± 0.00	1.045 ± 0.00
penEfr+	2.495 ± 0.13	2.348 ± 0.19	1.058 ± 0.00	1.054 ± 0.00
penPoi+	2.531 ± 0.13	2.446 ± 0.20	1.058 ± 0.00	1.054 ± 0.00
pen2-F+	3.723 ± 0.29	3.777 ± 0.26	1.087 ± 0.00	1.082 ± 0.01
pen5-F+	2.790 ± 0.18	2.698 ± 0.19	1.083 ± 0.00	1.061 ± 0.01
pen10-F+	2.653 ± 0.14	2.364 ± 0.20	1.085 ± 0.00	1.047 ± 0.01
pen20-F+	2.497 ± 0.13	2.318 ± 0.20	1.082 ± 0.00	1.049 ± 0.01

Then, we use Lemma 5, with $\xi_{i,\lambda} = Y_i - \beta_\lambda$, $a_\lambda = 0$, $b_\lambda = p_\lambda(n\hat{p}_\lambda)^{-2}$ for \tilde{p}_1 and $b_\lambda = (n^2\hat{p}_\lambda)^{-1}$ for p_2 . This proves, for all $q \geq 2$,

$$\|\tilde{p}_1(m) - \mathbb{E}^{\Lambda_m}[\tilde{p}_1(m)]\|_q^{(\Lambda_m)} \leq \max_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{\hat{p}_\lambda} \mathbb{1}_{\hat{p}_\lambda > 0} \right\} L_{A,\sigma_{\min}} D_m^{-1/2} q \mathbb{E}[p_2(m)] \quad (3)$$

$$\|p_2(m) - \mathbb{E}[p_2(m)]\|_q^{(\Lambda_m)} \leq L_{A,\sigma_{\min}} D_m^{-1/2} q \mathbb{E}[p_2(m)] \quad . \quad (4)$$

We deduce conditional concentration inequalities from those moment inequalities (for instance by Lemma 8.9 of (1)), with a deterministic probability bound $1 - Le^{-x} = 1 - n^{-\gamma}$. Hence, we deduce unconditional concentration inequalities, and the result follows for p_2 . To control the remainder term for \tilde{p}_1 , we use (9) in Lemma 9.

We now have to control the distance between $\mathbb{E}^{\Lambda_m}[\tilde{p}_1]$ and $\mathbb{E}[\tilde{p}_1]$. First, if $B_n \geq 1$, we can use Lemma 6: taking $X_\lambda = n\hat{p}_\lambda$ and $a_\lambda = p_\lambda(\sigma_\lambda)^2$, according to (46), we have $\tilde{p}_1(m) = Z_{m,1}$ and the concentration inequality for \tilde{p}_1 follows. On the other hand, if we only know that $B_n > 0$, instead of using Lemma 6, we remark that

$$\mathbb{E}^{\Lambda_m}[\tilde{p}_1(m)] \geq \min_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{\hat{p}_\lambda} \right\} \mathbb{E}^{\Lambda_m}[p_2(m)] \quad ,$$

and the result follows thanks to (10) in Lemma 9. \square

Lemma 5 (Prop. 5.5 of (1)). *Let $(a_\lambda)_{\lambda \in \Lambda_m}$ and $(b_\lambda)_{\lambda \in \Lambda_m}$ be two families of real numbers, $(r_\lambda)_{\lambda \in \Lambda_m}$ a family of integers. For all $\lambda \in \Lambda_m$, let $(\xi_{\lambda,i})_{1 \leq i \leq r_\lambda}$ be independent centered random variables admitting $2q$ -th moments $m_{2q,\lambda,i}$ for some $q \geq 2$. We define $S_{\lambda,1}$, $S_{\lambda,2}$ and Z as follows:*

$$Z = \sum_{\lambda \in \Lambda_m} (a_\lambda S_{\lambda,2} + b_\lambda S_{\lambda,1}^2) \quad \text{with} \quad S_{\lambda,1} = \sum_{i=1}^{r_\lambda} \xi_{\lambda,i} \quad \text{and} \quad S_{\lambda,2} = \sum_{i=1}^{r_\lambda} \xi_{\lambda,i}^2 \quad . \quad (5)$$

Then, there is a numerical constant $\kappa \leq 1.271$ such that, for every $q \geq 2$,

$$\|Z - \mathbb{E}[Z]\|_q \leq 4\sqrt{\kappa}\sqrt{q} \sqrt{\sum_{\lambda \in \Lambda_m} \left((a_\lambda + b_\lambda)^2 \sum_{i=1}^{r_\lambda} m_{2q,\lambda,i}^4 \right)} + 8\sqrt{2}\kappa q \sqrt{\sum_{\lambda \in \Lambda_m} \left(b_\lambda^2 \sum_{1 \leq i \neq j \leq r_\lambda} m_{2q,\lambda,i}^2 m_{2q,\lambda,j}^2 \right)} \quad .$$

Lemma 6 (Lemma 5.4 of (1)). *Assume that $\min_{\lambda \in \Lambda_m} \{np_\lambda\} \geq B_n \geq 1$ and $T \in (0, 1]$. Define $c_1 = 0.184$, $c_2 = 0.28$, $c_3 = 9.6$, $c_4 = 0.09$, $c_5 = 10.5$, and for every $t \geq 0$, $\varphi_1(t) = \max(t, 1)e^{-\max(t,1)}$.*

1. *Lower deviations: for every $x \geq 0$, with probability at least $1 - e^{-x}$,*

$$\mathbb{E}[Z_{m,1}] - Z_{m,1} \leq \frac{\varphi_1(c_1 B_n)}{c_1} \sum_{\lambda \in \Lambda_m} \frac{a_\lambda}{np_\lambda} + 3\sqrt{2} \sqrt{\sum_{\lambda \in \Lambda_m} \frac{a_\lambda^2}{(np_\lambda)^2}} \sqrt{4D_m \exp(-c_1 B_n) + x} \quad (6)$$

2. *Upper deviations: for every $x \geq 0$, with probability at least $1 - e^{-x}$,*

$$Z_{m,T} - \mathbb{E}[Z_{m,T}] \leq \frac{\varphi_1(c_2 B_n)}{c_2} \sum_{\lambda \in \Lambda_m} \left(\frac{a_\lambda}{np_\lambda} \right) + \sqrt{\sum_{\lambda \in \Lambda_m} \left(\frac{a_\lambda}{np_\lambda} \right)^2 (D_m e^{-c_4 B_n} + x)} \times c_3 \vee \left[\frac{c_5 T \sqrt{x + e^{-c_4 B_n}}}{n \min_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{a_\lambda} \right\} \sqrt{\sum_{\lambda \in \Lambda_m} \left(\frac{a_\lambda}{np_\lambda} \right)^2}} \right] \quad (7)$$

In the proof of Lemma 14, we need the following two consequences of Sect. 5.3.5 of Massart (7).

Lemma 7 (Lemma 8.17 of (1)). *Let (X_1, \dots, X_n) be n independent random variables, f a measurable function $\mathbb{R}^n \mapsto \mathbb{R}$ and*

$$Z = f(X_1, \dots, X_n) .$$

Then, there exists $\kappa \leq 1.271$ such that for every $q \geq 2$,

$$\|Z - \mathbb{E}[Z]\|_q \leq 2\sqrt{\kappa} \sqrt{q \left\| \sum_{i=1}^n (Z - \mathbb{E}[Z | (X_j)_{j \neq i}])^2 \right\|_{q/2}} . \quad (8)$$

proof of Lemma 7. Making references to (7), we use (5.58) in Thm. 5.10, with V defined by (5.12) and $Z_i = \mathbb{E}[Z | (X_j)_{j \neq i}]$. Hence,

$$\|(Z - \mathbb{E}[Z])_+\|_q \leq \sqrt{\kappa q \|V\|_{q/2}} .$$

It is then sufficient to apply it to both Z and $-Z$, and finally use the triangular inequality, in order to bound the moments of $(Z - \mathbb{E}[Z])$. \square

Lemma 8 (Lemma 8.18 of (1)). *Let (X_1, \dots, X_n) be n independent random variables admitting q -th moments for some $q \geq 2$: $m_{i,q} = \mathbb{E}[|X_i|^q]^{1/q}$. Let $S = \sum_{i=1}^s X_i$. Then,*

$$\|S\|_q \leq 2\sqrt{\kappa} \sqrt{q} \sqrt{\sum_{i=1}^s m_{i,q}^2} .$$

proof of Lemma 8. Apply Lemma 7 to S :

$$\begin{aligned} \|S - \mathbb{E}[S]\|_q &\leq 2\sqrt{\kappa} \sqrt{q \left\| \sum_{i=1}^s \mathbb{E}[(S - \mathbb{E}[S | X_i])^2 | X_1, \dots, X_n] \right\|_{q/2}} \\ &= 2\sqrt{\kappa} \sqrt{q \left\| \sum_{i=1}^s X_i^2 \right\|_{q/2}} \leq 2\sqrt{\kappa} \sqrt{q \sum_{i=1}^s \|X_i\|_q^2} . \end{aligned}$$

\square

5. Proofs of the technical lemmas

We here give the complete proofs of the three technical lemmas stated in App. **9.10**. First, we add to Lemma **15** two inequalities that we used in Sect. 4.

Lemma 9 (Lemma 12 of (3)). *Let $(p_\lambda)_{\lambda \in \Lambda_m}$ be non-negative real numbers of sum 1, $(n\widehat{p}_\lambda)_{\lambda \in \Lambda_m}$ a multinomial vector of parameters $(n; (p_\lambda)_{\lambda \in \Lambda_m})$, $\gamma > 0$. Assume that $\text{Card}(\Lambda_m) \leq n$ and $\min_{\lambda \in \Lambda_m} \{np_\lambda\} \geq B_n > 0$. There is an event of probability at least $1 - Ln^{-\gamma}$ on which the following three inequalities hold.*

$$\max_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{\widehat{p}_\lambda} \mathbb{1}_{\widehat{p}_\lambda > 0} \right\} \leq L \times (\gamma + 1) \ln(n) \quad (9)$$

$$\min_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{\widehat{p}_\lambda} \right\} \geq \frac{1}{2 + (\gamma + 1)B_n^{-1} \ln(n)} \quad (10)$$

$$\min_{\lambda \in \Lambda_m} \{n\widehat{p}_\lambda\} \geq \frac{\min_{\lambda \in \Lambda_m} \{np_\lambda\}}{2} - 2(\gamma + 1) \ln(n) \quad (11)$$

proof of Lemma 9. Those three results come from Bernstein's inequality (e.g. Prop. 2.9 of (7)) applied to $n\widehat{p}_\lambda$: for every $\lambda \in \Lambda_m$, there is a set of probability $1 - 2n^{-(\gamma+1)}$ on which

$$np_\lambda - \sqrt{2np_\lambda(\gamma + 1) \ln(n)} - \frac{(\gamma + 1) \ln(n)}{3} \leq n\widehat{p}_\lambda \leq np_\lambda + \sqrt{2np_\lambda(\gamma + 1) \ln(n)} + \frac{(\gamma + 1) \ln(n)}{3} .$$

For (9), if $np_\lambda \geq 8(\gamma + 1) \ln(n)$, the lower bound gives the result. Otherwise, remark only that $(p_\lambda/\widehat{p}_\lambda) \mathbb{1}_{\widehat{p}_\lambda > 0} \leq np_\lambda \leq 8(\gamma + 1) \ln(n)$. For (10), use the upper bound and remark that $np_\lambda(\gamma + 1) \ln(n) B_n^{-1} \geq (\gamma + 1) \ln(n)$. For (11), use the lower bound and remark that $\sqrt{2np_\lambda(\gamma + 1) \ln(n)} \leq (np_\lambda)/2 + (\gamma + 1) \ln(n)$. Finally, the union bound gives the result since $\text{Card}(\Lambda_m) \leq n$. \square

proof of Lemma 16. Remark that for every $\lambda \in \Lambda_m$ such that σ does not jump on I_λ ,

$$(\sigma_\lambda^r)^2 \geq \min_{x \in I_\lambda} \{ \sigma(x)^2 \} \geq \frac{1}{2 \text{Leb}(I_\lambda)} \int_{I_\lambda} \sigma(x)^2 \text{Leb}(dx) - (K_\sigma \text{diam}(I_\lambda))^2$$

since $(a - b)^2 = a^2 - 2ab + b^2 \geq \frac{a^2}{2} - b^2$ and σ is K_σ Lipschitz. There is at most J_σ other λ , for which

$$(\sigma_\lambda^r)^2 \geq 0 \geq \frac{1}{2 \text{Leb}(I_\lambda)} \int_{I_\lambda} \sigma(x)^2 \text{Leb}(dx) - \frac{\|\sigma(X)\|_\infty^2}{2} .$$

This implies

$$\sum_{\lambda \in \Lambda_m} (\sigma_\lambda^r)^2 \geq \frac{\text{Leb}(\mathcal{X})}{2 \max_{\lambda \in \Lambda_m} \{ \text{Leb}(I_\lambda) \}} \|\sigma\|_{L^2(\text{Leb})}^2 - D_m K_\sigma^2 \max_{\lambda \in \Lambda_m} \{ \text{diam}(I_\lambda)^2 \} - \frac{J_\sigma \|\sigma(X)\|_\infty^2}{2} .$$

\square

proof of Lemma 17. From **(A1)** and the upper bound in **(Ar_{ℓ,u})**,

$$\|s - s_m\|_\infty \leq B \max_{\lambda \in \Lambda_m} \{\text{diam}(I_\lambda)\} = B \max_{\lambda \in \Lambda_m} \{\text{Leb}(I_\lambda)\} \leq B c_{r,u} \text{Leb}(\mathcal{X}) D_m^{-1} . \quad (12)$$

For the lower bound, let Λ_m^J be the set of $\lambda \in \Lambda_m$ such that $I_\lambda \subset J$,

$$s_{\lambda, \text{Leb}} = \text{Leb}(I_\lambda)^{-1} \int_{I_\lambda} s(x) \text{Leb}(dx)$$

and $\mu = \mathcal{L}(X)$. Then, using **(Ad_ℓ)**,

$$\begin{aligned} \|s - s_m\|_{L^2(\mu)}^2 &\geq \sum_{\lambda \in \Lambda_m^J} \int_{I_\lambda} (s(x) - s_m(x))^2 c_X^{\min} \text{Leb}(\mathcal{X})^{-1} \text{Leb}(dx) \\ &\geq c_X^{\min} \text{Leb}(\mathcal{X})^{-1} \sum_{\lambda \in \Lambda_m^J} \int_{I_\lambda} (s(x) - s_{\lambda, \text{Leb}})^2 \text{Leb}(dx) . \end{aligned}$$

For any $\lambda \in \Lambda_m^J$, since s is continuous on I_λ , there is some $x_\lambda \in I_\lambda$ such that $s_{\lambda, \text{Leb}} = s(x_\lambda)$. By **(A1)**, for every $x \in I_\lambda$,

$$(s(x) - s(x_\lambda))^2 \geq B_0^2 (x - x_\lambda)^2$$

so that

$$\begin{aligned} \|s - s_m\|_{L^2(\mu)}^2 &\geq c_X^{\min} \text{Leb}(\mathcal{X})^{-1} \sum_{\lambda \in \Lambda_m^J} \frac{B_0^2 \text{Leb}(I_\lambda)^3}{12} \geq \frac{c_X^{\min} B_0^2 c_{r,\ell}^3 \text{Leb}(\mathcal{X})^2}{12 D_m^3} \text{Card}(\Lambda_m^J) \\ &\geq \frac{c_X^{\min} B_0^2 c_{r,\ell}^3 \text{Leb}(\mathcal{X})^2}{12 D_m^3} \left(\frac{c_J D_m}{c_{r,u}} - 2 \right)_+ . \end{aligned}$$

Combined with (12), this gives the result. \square

References

- [1] ARLOT, S. (2007). Resampling and model selection. Ph.D. thesis, University Paris-Sud 11. Available online at <http://tel.archives-ouvertes.fr/tel-00198803/en/>.
- [2] ARLOT, S. (2008a). Model selection by resampling penalization.
- [3] ARLOT, S. (2008b). V -fold cross-validation improved: V -fold penalization. arXiv:0802.0566.
- [4] BURCHARD, H. G. AND HALE, D. F. (1975). Piecewise polynomial approximation on optimal meshes. *J. Approximation Theory* **14**, 2, 128–147. MRMR0374761 (51 #10957)
- [5] DEVORE, R. A. AND LORENTZ, G. G. (1993). *Constructive approximation*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. **303**. Springer-Verlag, Berlin. MRMR1261635 (95f:41001)

- [6] DONOHO, D. L. AND JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 432, 1200–1224. MRMR1379464 (96k:62093)
- [7] MASSART, P. (2007). *Concentration inequalities and model selection*. Lecture Notes in Mathematics, Vol. **1896**. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MRMR2319879