Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Technical appendix to "Model selection by resampling penalization"

Sylvain Arlot*

Sylvain Arlot
CNRS ; Willow Project-Team
Laboratoire d'Informatique de l'Ecole Normale Superieure
(CNRS/ENS/INRIA UMR 8548)
45, rue d'Ulm, 75230 Paris, France
e-mail: sylvain.arlot@ens.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, 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, regressogram.

Contents

*The author was financed in part by Univ Paris-Sud (Laboratoire de Mathematiques, CNRS - UMR 8628).

Throughout this appendix, we use the notation of the main paper [2]. References within the appendix are denoted by (1) or 1, whereas references to the main paper [2] are denoted by **(1)** or **1**.

Following the ordering of [2], we first complete Section **3.3.2** with other alternative assumptions to Theorem **1** (Section 1). Then, we state some approximation theory results for histogram models that give sufficient condition for assumption (**Ap**) on the bias of the models (Section 2). An extensive simulation experiment (completing the one of Section **5**) is provided in Section 3. Finally, Proposition **10** is proved in Section 4, and Lemmas **8** and **9** are proved in Section 5.

1. Assumption sets for oracle inequalities for RP

In this section, we consider alternative assumption sets for the results of Section **3** about Resampling Penalization. Since Theorem **1** relies on a general result (Lemma **7**), giving alternative assumptions for Theorem **1** remains to give sufficient conditions for (**Bg**) or (**Ug**).

1.1. Bounded case

One way of removing (**An**) was suggested in Section **3.3.2**. Actually, (**An**) can also be replaced in the assumptions of Theorem **1** by

1. ($\mathbf{A}_{\mathrm{gauss}}$) the noise is sub-Gaussian
   ($\mathbf{Ar}_{\mathbf{u}}^{\mathbf{X}}$) the partition is "upper-regular" for $\mathcal{D}(X)$, that is $\exists c_{\mathrm{r,u}}^{X} > 0$ such that
   $$D_m \max_{\lambda \in \Lambda_m} p_\lambda \leq c_{\mathrm{r,u}}^{X} \enspace .$$
2. $X \subset \mathbb{R}^k$,
   ($\mathbf{A}_{\mathrm{gauss}}$) the noise is sub-Gaussian,
   ($\mathbf{Ar}_{\mathbf{u}}$) the partition is "upper-regular" for Leb, that is $\exists c_{\mathrm{r,u}} > 0$ such that
   $$D_m \max_{\lambda \in \Lambda_m} \mathrm{Leb}(I_\lambda) \leq c_{\mathrm{r,u}} \,\mathrm{Leb}(\mathcal{X})$$
   ($\mathbf{Ad}_{\mathbf{u}}$) the density of X with respect to Leb is bounded from above, that is $\exists c_X^{\max} > 0$ such that
   $$\forall I \subset \mathcal{X}, \quad \mathbb{P}(X \in I) \leq c_X^{\max} \frac{\mathrm{Leb}(I)}{\mathrm{Leb}(\mathcal{X})} \enspace .$$

Proof. Following the proof given in Section **8.4**, 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

$$Q_m^{(p)} \geq \frac{1}{D_m} \sum_{\lambda \in \Lambda_m} \left(\sigma_\lambda^r \right)^2 = \sum_{\lambda \in \Lambda_m} \frac{p_\lambda \left(\sigma_\lambda^r \right)^2}{D_m p_\lambda}$$
$$\geq \sum_{\lambda \in \Lambda_m} \frac{p_\lambda \left(\sigma_\lambda^r \right)^2}{\max_{\lambda \in \Lambda_m} \left\{ D_m p_\lambda \right\}} \geq \frac{\left\| \sigma(X) \right\|_2^2}{c_{\mathrm{r,u}}^X} > 0 \enspace .$$

The second case is a consequence of the first one since

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

□

Moreover, in all the assumption sets above and in those of Section **3**, the sub-Gaussian assumption on the noise ($\mathbf{A}_{\mathrm{gauss}}$) can be replaced by a general moment inequality:

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

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

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

1.2. Unbounded case

In Section **3.3.2**, we also give a set of assumptions for Theorem **1** in the unbounded case. The boundedness assumption (**Ab**) and the lower bound on the noise (**An**) can be removed simultaneously from the assumptions of Theorem **1** at the price of adding

($\mathbf{A}_{\mathrm{gauss}}$) The noise is sub-Gaussian: there exists $c_{\mathrm{gauss}} > 0$ such that

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

($\mathbf{A}\delta$) Global moment assumption for the bias: there exist constants $c_{\Delta,m}^g, D_0 > 0$ such that for every $m \in \mathcal{M}_n$ of dimension $D_m \geq D_0$,

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

($\mathbf{A}\sigma_{\max}$) Noise-level bounded from above: $\sigma^2(X) \leq \sigma_{\max}^2 < +\infty$ a.s.
($\mathbf{As}_{\max}$) Bound on the target function: $\left\| s \right\|_\infty \leq A$.

and one among the following

1. ($\mathbf{Ar_u^X}$) the partition is "upper-regular" for $\mathcal{D}(X)$, that is $\exists c_{\mathrm{r,u}}^X > 0$ such that
$$D_m \max_{\lambda \in \Lambda_m} p_\lambda \leq c_{\mathrm{r,u}}^X \enspace .$$
2. $X \subset \mathbb{R}^k$,
($\mathbf{Ar_u}$) the partition is "upper-regular" for Leb, that is $\exists c_{\mathrm{r,u}} > 0$ such that
$$D_m \max_{\lambda \in \Lambda_m} \mathrm{Leb}(I_\lambda) \leq c_{\mathrm{r,u}} \mathrm{Leb}(\mathcal{X})$$
($\mathbf{Ad_u}$) the density of X with respect to Leb is bounded from above, that is $\exists c_X^{\max} > 0$ such that
$$\forall I \subset \mathcal{X}, \quad \mathbb{P}(X \in I) \leq c_X^{\max} \frac{\mathrm{Leb}(I)}{\mathrm{Leb}(\mathcal{X})} \enspace .$$
3. $X \subset \mathbb{R}^k$ is bounded, equipped with $\left\| \cdot \right\|_\infty$,
($\mathbf{Ar_u^d}$) the partition is "upper-regular", that is $\exists c_{\mathrm{r,u}}^d, \alpha_d > 0$ such that
$$\max_{\lambda \in \Lambda_m} \left\{ \mathrm{diam}(I_\lambda) \right\} \leq c_{\mathrm{r,u}}^d D_m^{-\alpha_d} \mathrm{diam}(X) \enspace ,$$
($\mathbf{Ar_u}$) the partition is "upper-regular" for Leb, that is $\exists c_{\mathrm{r,u}} > 0$ such that
$$\max_{\lambda \in \Lambda_m} \left\{ \mathrm{Leb}(I_\lambda) \right\} \leq c_{\mathrm{r,u}} D_m^{-1} \mathrm{Leb}(X) \enspace ,$$
and ($\mathbf{A}\sigma$) σ is piecewise K_σ-Lipschitz with at most J_σ jumps.

Proof. In Section **8.4**, (**An**) was only used to prove a lower bound on $Q_m^{(p)}$. The two first cases thus follow from the proof given in Section 1.1 above. The last one follows from Lemma **8**. □

As in the bounded case, the sub-Gaussian assumption on the noise ($\mathbf{A}_{\mathrm{gauss}}$) can be replaced everywhere by the more general moment assumption ($\mathbf{A}\epsilon$).

Sufficient conditions for ($\mathbf{A}\delta$) can be derived either from Lemma **9** or from Lemma 1 below.

1.3. An alternative to Lemma 9

In Section **8.4**, we give a sufficient condition for ($\mathbf{A}\delta$) that relies on the regularity of s, a lower bound on the density of X with respect to Leb and the regularity of the partition (Lemma **9**). We state below a more precise lemma when $\mathcal{X} \subset \mathbb{R}$ is bounded.

Lemma 1. *Let s be a B-Lipschitz function on $\mathcal{X} \subset \mathbb{R}$ and μ a probability measure on $\mathcal{X}$. 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 $\mathcal{X}$ such that their sizes D_{m_k} are tending to infinity when k tends to infinity and* $(\mathbf{Ar_u^d})$*:*

$$\forall k \in \mathbb{N}, \qquad \max_{\lambda \in \Lambda_{m_k}} \{ \operatorname{diam}(I_\lambda) \} \leq c_{\mathrm{r,u}}^d D_{m_k}^{-1} \operatorname{diam}(\mathcal{X}) \enspace .$$

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

$$\left\| s - s_{m_k} \right\|_\infty \leq c_{\Delta,m}^g \left\| s - s_{m_k} \right\|_{L^2(\mu)} \enspace .$$

Proof. If s is constant, the result is obvious. Otherwise, both $\left\| s - s_{m_k} \right\|_\infty$ and $\left\| s - s_{m_k} \right\|_2$ are positive.

Since s is Lipschitz with constant B,

$$\left\| s - s_{m_k} \right\|_\infty \leq B \max_{\lambda \in \Lambda_{m_k}} \{ \operatorname{diam}(I_\lambda) \} \leq D_{m_k}^{-1} \operatorname{diam}(\mathcal{X}) c_{\mathrm{r,u}}^d B \enspace .$$

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

Remark 1. If further regularity conditions on s are assumed, the difference between $\left\| s - s_{m_k} \right\|_{L^2(\mu)}^2$ and its limit when $k \to \infty$ can be controlled. Then, the constant $c_{\Delta,m}^g$ only depends on s through B and $\left\| s' \right\|_{L^2(\mu)}^2$ 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 Theorem **1**, we use the following assumption:

(**Ap**) Polynomially decreasing bias: there exist $\beta_1 \geq \beta_2 > 0$ and $C_{\mathrm{b}}^+, C_{\mathrm{b}}^- > 0$ such that

$$C_{\mathrm{b}}^- D_m^{-\beta_1} \leq \left\| s - s_m \right\|_{L^2(\mu)} \leq C_{\mathrm{b}}^+ D_m^{-\beta_2} \enspace .$$

where $\mu = \mathcal{D}(X)$ and s_m is the $L^2(\mu)$ projection onto some histogram model S_m. (**Ap**) 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 piecewise constant functions. Lemma 2 below shows for any $\alpha \in (0,1]$, that all other α-Hölderian functions satisfy (**Ap**). On approximation theory, we refer to the book of DeVore and Lorentz [5]. All the proofs of the following results can be found in [1, Chapter 8.10], where they were initially stated.

Let $(\mathcal{X}, d)$ be a metric space. For every $\alpha \in (0;1]$, $\delta, \epsilon, R > 0$, $\mathcal{H}_{\delta,\epsilon}(\alpha, R)$ denotes the set of α-Hölderian functions f on $\mathcal{X}$, that is

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

for which $x_1, x_2 \in \mathcal{X}$ exist such that

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

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

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

2.1. Regular histograms in [0;1]

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

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

For every $D \in \mathbb{N}$, s_D denotes the $L^2(\mathrm{Leb})$ projection of s on the space of regular histograms with D pieces. Then, there exist constants

$$C_1 = L(\alpha) R^{-\alpha^{-1}} \epsilon^{2+\alpha^{-1}} |x_1 - x_2|^{-1-\alpha^{-1}} > 0 \enspace ,$$
$$C_2 = R^2 \enspace , \quad \beta_1 = 1 + \frac{1}{\alpha} \quad \textit{and} \quad \beta_2 = 2\alpha$$

such that for every $D > 0$,

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

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

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

If $(\mathbf{Ad}_\ell)$ holds, then

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

and thus (1) implies the lower bound in $(\mathbf{Ap})$.

Remark 3. The lower bound in (1) cannot be improved, as showed in Section 2.3: for every $\alpha, R, \delta, \epsilon > 0$, there exists C_1' such that for every $D > 0$,

$$\inf_{s \in \mathcal{H}_{\delta,\epsilon}(\alpha,R)} \left\{ \|s - s_D\|^2_{L^2(\mathrm{Leb})} \right\} \leq \frac{C_1'}{D^{1+\alpha^{-1}}} \enspace .$$

2.2. Regular histograms in $\mathbb{R}^k$

We now generalize the previous result to subsets of $\mathbb{R}^k$. For the sake of simplicity, $\mathcal{X}$ is assumed to be a ball of $(\mathbb{R}^k, \|\cdot\|_{\infty})$. Otherwise, if $\overset{o}{\mathcal{X}}$ is connex and nonempty, 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 $B(s)$ and derive $(\mathbf{Ap})$. 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" partition[1] $(I_\lambda)_{\lambda\in\Lambda_d}$ of $\mathcal{X}$ of pace D^{-1}. Let s_D be the piecewise constant function, defined on each piece I_λ of this partition as*

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

Then,

$$\int_{\mathcal{X}} \left(s(t) - s_D \right)^2 dt \geq L_{k,\alpha} \epsilon^{2+k\alpha^{-1}} \delta^{-1-k\alpha^{-1}} R^{1-k(1+\alpha^{-1})} \times \left(D \vee \delta^{-1} \right)^{-1-k\alpha^{-1}+(k-1)\alpha} \enspace . \tag{2}$$

Remark 4.
1. The number of pieces in the partition is not D but approximatively $\mathrm{Leb}(\mathcal{X})D^k$, the exact value depending on the precise shape of $\mathcal{X}$. Then, if X has a lower bounded density with respect to Leb on $\mathcal{X}$, under the assumptions of Lemma 3, the lower bound in (**Ap**) holds with $\beta_1 = k^{-1} + \alpha^{-1} - (k-1)k^{-1}\alpha$.
2. The following upper bound on the bias is straightforward:
$$\frac{1}{\mathrm{Leb}(\mathcal{X})} \int_{\mathcal{X}} \left(s(t) - s_D \right)^2 dt \leq \|s - s_D\|_\infty^2 \leq R^2 D^{-2\alpha} \enspace .$$
3. When $\mathcal{X}$ is not a ball of $(\mathbb{R}^k, \|\cdot\|_\infty)$, we can use a general argument assuming only that for every $x_1, x_2 \in \overset{o}{\mathcal{X}}$, there exists a path from x_1 to x_2 that has an η-enlargement in $\mathcal{X}$ for some $\eta > 0$. We then obtain a lower bound weaker than (2) 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 $\mathcal{X} = [0; 1]$, $R > 0$, $\alpha \in (0; 1]$, $\eta > 0$, $1 \geq \delta \geq (1 + \eta)D^{-1}$ and $L(\alpha)R\lfloor D\delta\rfloor D^{-1} \geq \epsilon > 0$. Then,*

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

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

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.

Hence, the exponent $\beta_1 = 1 + \alpha^{-1}$ cannot be improved as long as we look for a lower bound uniform over $\mathcal{H}_{\delta,\epsilon}(\alpha, R)$. However, there does not necessarily exist 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) can be found in [4]. See also the book by DeVore and Lorentz [5], in particular Chapter 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 Section **5** with the same benchmarks, but with only $N = 250$ samples for each experiment.

Data are generated according to

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

where $(X_i)_{1 \leq i \leq n}$ are independent with uniform distribution over $\mathcal{X} = [0; 1]$ and $(\epsilon_i)_{1 \leq i \leq n}$ are independent standard Gaussian variables independent of $(X_i)_{1 \leq i \leq n}$. The experiments differ from

- the regression function s:
  - S1000, S$\sqrt{0.1}$, S0.1 and Svar2 have the same smooth function $s(x) = \sin(\pi x)$ as in S1 and S2, see Figure 1.
  - Sqrt has $s(x) = \sqrt{x}$, which is smooth except around 0, see Figure 6.
  - His6 has a regular histogram with 5 jumps (hence it belongs to the regular histogram model of dimension 6), see Figure 8.
  - DopReg and Dop2bin have the Doppler function, as defined by Donoho and Johnstone [6], see Figure 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.
  - $\sigma(x) = \mathbf{1}_{x \geq 1/2}$ 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 collection of models $(S_m)_{m \in \mathcal{M}_n}$: with the notation introduced in Section **5**,
  - for S1000, S$\sqrt{0.1}$, S0.1, Sqrt and His6, we use the "regular" collection, as for S1:
$$\mathcal{M}_n = \left\{ 1, \ldots, \left\lfloor \frac{n}{\ln(n)} \right\rfloor \right\} \enspace .$$
  - for Svar2, we use the "regular with two bin sizes" collection, as for S2:
$$\mathcal{M}_n = \{1\} \cup \left\{ 1, \ldots, \left\lfloor \frac{n}{2\ln(n)} \right\rfloor \right\}^2 \enspace .$$
  - for DopReg, we use the "regular dyadic" collection, as for HSd1:
$$\mathcal{M}_n = \left\{ 2^k \text{ s.t. } 0 \leq k \leq \ln_2(n) - 1 \right\} \enspace .$$
  - for Dop2bin, we use the "regular dyadic with two bin sizes" collection, as for HSd2:
$$\mathcal{M}_n = \{1\} \cup \left\{ 2^k \text{ s.t. } 0 \leq k \leq \ln_2(n) - 2 \right\}^2 \enspace .$$

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

Instances of data sets for each experiment are given in Figures 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).

We consider for each of these experiments the same procedures as in Section **5** and we add to them:

1. Mal*, which is Mallows' C_p penalty with the true value of the variance: $\operatorname{pen}(m) = 2\mathbb{E}\left[\sigma^2(X)\right] D_m n^{-1}$. Although it cannot be used on real data sets, it is an interesting point of comparison, which does not have possible weaknesses coming from the variance estimator $\widehat{\sigma}^2$.
2. penPoi, which is RP with Poisson(1) weights and $C = C_W = 1$.
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. pen$V$-F for $V \in \{2, 5, 10, 20\}$ which is the V-fold subsampling version of RP, called V-fold penalization. These penalties are properly defined and studied in [3].

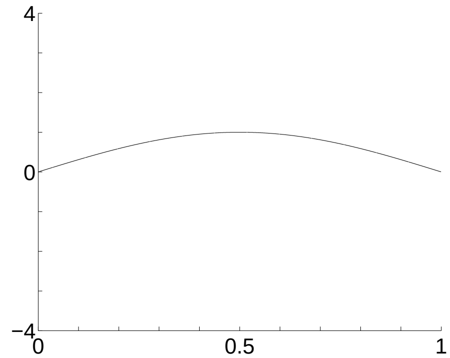

Fig 1. $s(x) = \sin(\pi x)$

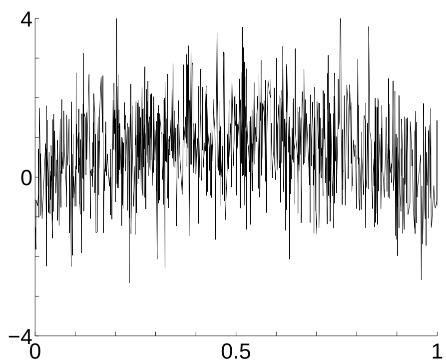

Fig 2. *Data sample for S1000*

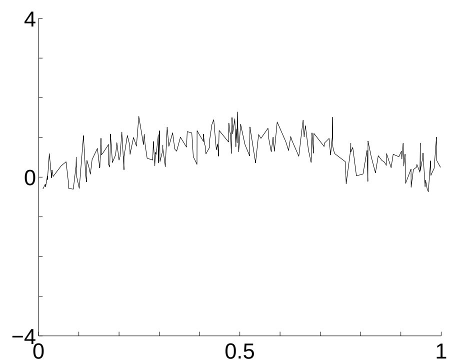

Fig 3. *Data sample for* $S\sqrt{0.1}$

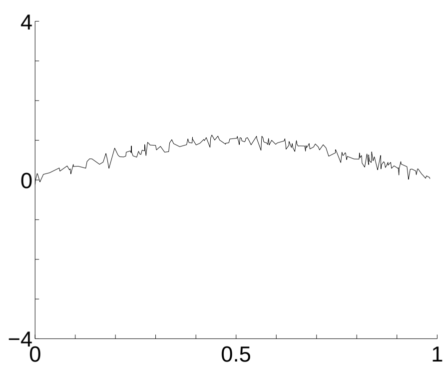

Fig 4. *Data sample for S0.1*

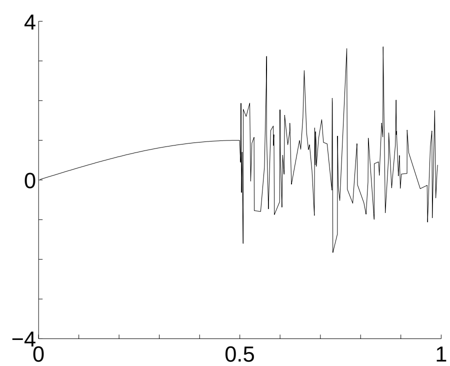

Fig 5. *Data sample for Svar2*

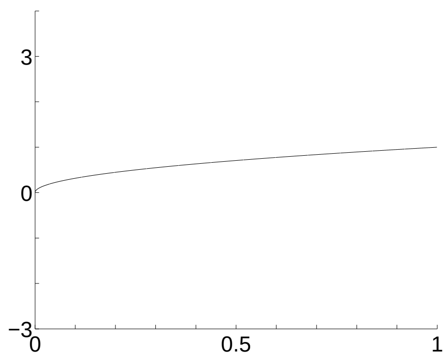

Fig 6. $s(x) = \sqrt{x}$

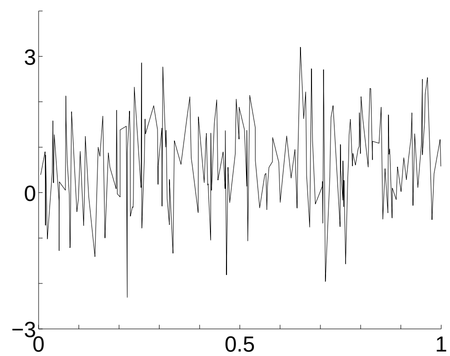

Fig 7. *Data sample for Sqrt*

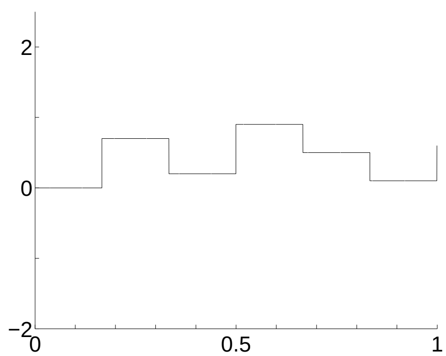

Fig 8. $s(x) = \mathrm{His}_6(x)$

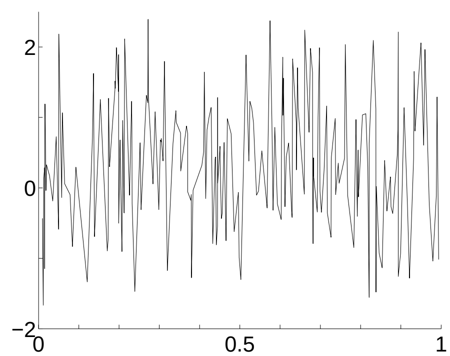

Fig 9. *Data sample for His6*

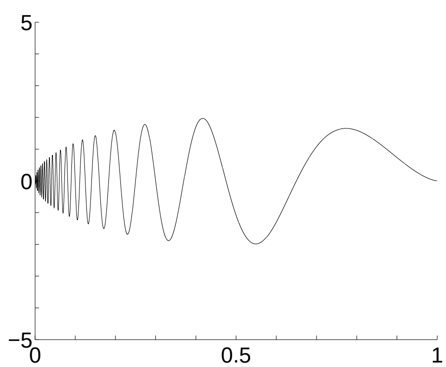

Fig 10. $s(x) = \mathrm{Doppler}(x)$ *(see [6])*

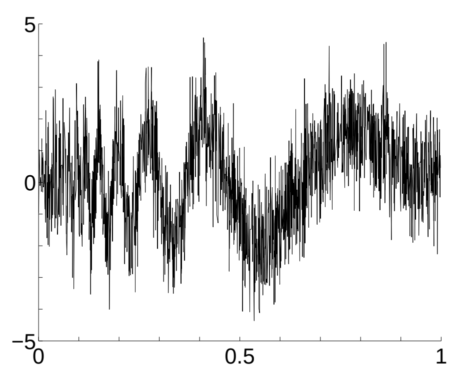

Fig 11. *Data sample for DopReg and Dop2bin*

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

Estimations of C_{or} (and uncertainties for these estimations) for all the procedures are reported in Tables 1–3 (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

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

is the leading constant which appear in oracle inequalities of the form

$$\mathbb{E}\left[\ell\left(s, \widehat{s}_{\widehat{m}}\right)\right] \leq c_{\mathrm{or}} \inf_{m\in\mathcal{M}_n}\left\{\mathbb{E}\left[\ell\left(s, \widehat{s}_m\right)\right]\right\} + \mathcal{O}(n^{-1}) \enspace ,$$

which are the most commonly proved oracle inequalities in the model selection literature. Note that C'_{or} is always smaller than C_{or}.

The comparison between Mallows', VFCV and Resampling Penalization is quite the same as in Section **5**: in "easy" homoscedastic frameworks (S1000, S$\sqrt{0.1}$, S0.1), their model selection performances are similar. A harder problem such as Svar2 (which is heteroscedastic but different from S2) makes Mallows' C_p fail whereas RP and VFCV still work.

As expected, taking n larger (S1000) or σ smaller (S$\sqrt{0.1}$ and S0.1) makes the constant C_{or} closer to 1. Note also that overpenalization often improves the quality of the procedure (but not always: see DopReg and S0.1). We have for instance C_{or}(penLoo) $<$ C_{or}(penRho) $<$ C_{or}(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 and with V-fold penalties. Exact formulas with exchangeable schemes (Rad, Rho) yield better model selection performance than Monte-Carlo approximations or V-fold penalties in most of the experiments (and sometimes significantly, for instance in S2 or HSd1). But the difference of model selection performance is not so large, making for instance 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.

Nevertheless, it is quite unclear whether one should prefer Monte-Carlo approximations or V-fold subsampling schemes (taking $V = 20$ to make computational complexities comparable). According to the experiments, pen20-F can be better (S2), worse (HSd1) or not significantly[2] different (S1,HSd2) from penRad(20).

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

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

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

in Tables 4–6. As noted in [2], they are quite similar to the previous ones.

[2]Given the number of experiments made here, we should also mention that we do not take into account multiplicity when saying 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 indices 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 procedures (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]$ and $\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]+$ 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+	$\mathbf{1.800 \pm 0.03}$	3.173 ± 0.07	$\mathbf{1.002 \pm 0.003}$	1.411 ± 0.008
Mal*	2.028 ± 0.04	2.657 ± 0.06	1.044 ± 0.004	1.513 ± 0.005
Mal*+	$\mathbf{1.827 \pm 0.03}$	2.437 ± 0.05	$\mathbf{1.004 \pm 0.003}$	1.548 ± 0.003
$\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]$	1.919 ± 0.03	2.296 ± 0.05	1.028 ± 0.004	$\mathbf{1.102 \pm 0.004}$
$\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]+$	$\mathbf{1.792 \pm 0.03}$	$\mathbf{2.028 \pm 0.04}$	$\mathbf{1.003 \pm 0.003}$	$\mathbf{1.089 \pm 0.004}$
2-FCV	2.078 ± 0.04	2.542 ± 0.05	$\mathbf{1.002 \pm 0.003}$	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	$\mathbf{1.102 \pm 0.004}$
penRho	1.982 ± 0.04	2.502 ± 0.06	1.018 ± 0.003	$\mathbf{1.103 \pm 0.004}$
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	$\mathbf{1.103 \pm 0.004}$
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+	$\mathbf{1.799 \pm 0.03}$	$\mathbf{2.137 \pm 0.05}$	$\mathbf{1.002 \pm 0.003}$	$\mathbf{1.095 \pm 0.004}$
penRho+	$\mathbf{1.798 \pm 0.03}$	$\mathbf{2.142 \pm 0.05}$	$\mathbf{1.002 \pm 0.003}$	$\mathbf{1.095 \pm 0.004}$
penLoo+	$\mathbf{1.844 \pm 0.03}$	$\mathbf{2.215 \pm 0.05}$	$\mathbf{1.004 \pm 0.003}$	$\mathbf{1.096 \pm 0.004}$
penEfr+	2.016 ± 0.05	2.605 ± 0.06	1.011 ± 0.003	$\mathbf{1.097 \pm 0.004}$
penPoi+	2.039 ± 0.05	2.620 ± 0.06	1.011 ± 0.003	$\mathbf{1.097 \pm 0.004}$
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	$\mathbf{1.006 \pm 0.003}$	$\mathbf{1.102 \pm 0.004}$
pen10-F+	1.872 ± 0.03	2.285 ± 0.05	$\mathbf{1.005 \pm 0.003}$	$\mathbf{1.098 \pm 0.004}$
pen20-F+	1.898 ± 0.03	2.254 ± 0.05	$\mathbf{1.004 \pm 0.003}$	$\mathbf{1.098 \pm 0.004}$
$C'_{\mathrm{or}}/C_{\mathrm{or}}$	0.768	0.753	0.999	0.854

TABLE 2

Accuracy indices 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 procedures (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]$ and $\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]+$ 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	$\mathbf{1}_{x\geq 1/2}$
n (sample size)	1000	200	200	200
$\mathcal{M}_n$	regular	regular	regular	2 bin sizes
Mal	$\mathbf{1.667 \pm 0.04}$	$\mathbf{1.611 \pm 0.03}$	$\mathbf{1.400 \pm 0.02}$	5.643 ± 0.22
Mal+	$\mathbf{1.619 \pm 0.03}$	$\mathbf{1.593 \pm 0.03}$	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*+	$\mathbf{1.617 \pm 0.03}$	2.073 ± 0.04	3.641 ± 0.07	3.544 ± 0.17
$\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]$	1.745 ± 0.05	$\mathbf{1.571 \pm 0.03}$	$\mathbf{1.373 \pm 0.02}$	$\mathbf{2.409 \pm 0.13}$
$\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]+$	$\mathbf{1.617 \pm 0.03}$	$\mathbf{1.554 \pm 0.03}$	$\mathbf{1.392 \pm 0.02}$	$\mathbf{2.005 \pm 0.10}$
2-FCV	$\mathbf{1.668 \pm 0.04}$	1.663 ± 0.04	$\mathbf{1.394 \pm 0.02}$	2.960 ± 0.15
5-FCV	1.756 ± 0.07	1.693 ± 0.04	$\mathbf{1.393 \pm 0.02}$	2.950 ± 0.16
10-FCV	1.746 ± 0.04	1.684 ± 0.04	$\mathbf{1.385 \pm 0.02}$	2.681 ± 0.14
20-FCV	1.774 ± 0.05	1.645 ± 0.03	$\mathbf{1.382 \pm 0.02}$	2.742 ± 0.16
LOO	1.768 ± 0.05	1.639 ± 0.04	$\mathbf{1.379 \pm 0.02}$	2.641 ± 0.15
penRad	1.748 ± 0.05	$\mathbf{1.609 \pm 0.03}$	$\mathbf{1.405 \pm 0.02}$	2.510 ± 0.15
penRho	1.748 ± 0.05	$\mathbf{1.619 \pm 0.03}$	$\mathbf{1.404 \pm 0.02}$	2.518 ± 0.15
penLoo	1.776 ± 0.05	1.641 ± 0.04	$\mathbf{1.379 \pm 0.02}$	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	$\mathbf{1.636 \pm 0.04}$	1.415 ± 0.02	2.966 ± 0.17
penRho(20)	1.725 ± 0.05	1.641 ± 0.04	$\mathbf{1.397 \pm 0.02}$	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	$\mathbf{1.390 \pm 0.02}$	3.209 ± 0.18
pen5-F	1.816 ± 0.05	$\mathbf{1.638 \pm 0.04}$	$\mathbf{1.400 \pm 0.02}$	2.749 ± 0.15
pen10-F	1.783 ± 0.05	1.706 ± 0.04	$\mathbf{1.374 \pm 0.02}$	2.598 ± 0.15
pen20-F	1.801 ± 0.05	1.657 ± 0.03	$\mathbf{1.385 \pm 0.02}$	2.684 ± 0.15
penRad+	$\mathbf{1.619 \pm 0.03}$	$\mathbf{1.574 \pm 0.03}$	1.417 ± 0.02	$\mathbf{2.232 \pm 0.12}$
penRho+	$\mathbf{1.619 \pm 0.03}$	$\mathbf{1.578 \pm 0.03}$	1.417 ± 0.02	$\mathbf{2.243 \pm 0.12}$
penLoo+	$\mathbf{1.626 \pm 0.03}$	$\mathbf{1.587 \pm 0.03}$	$\mathbf{1.401 \pm 0.02}$	$\mathbf{2.349 \pm 0.13}$
penEfr+	$\mathbf{1.636 \pm 0.03}$	1.670 ± 0.04	$\mathbf{1.407 \pm 0.02}$	2.614 ± 0.16
penPoi+	$\mathbf{1.636 \pm 0.03}$	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+	$\mathbf{1.683 \pm 0.04}$	$\mathbf{1.616 \pm 0.03}$	$\mathbf{1.399 \pm 0.02}$	$\mathbf{2.460 \pm 0.14}$
pen10-F+	$\mathbf{1.627 \pm 0.04}$	$\mathbf{1.613 \pm 0.03}$	$\mathbf{1.385 \pm 0.02}$	$\mathbf{2.398 \pm 0.14}$
pen20-F+	$\mathbf{1.644 \pm 0.04}$	$\mathbf{1.583 \pm 0.03}$	$\mathbf{1.390 \pm 0.02}$	$\mathbf{2.316 \pm 0.13}$
$C'_{\mathrm{or}}/C_{\mathrm{or}}$	0.8	0.801	0.816	0.779

Table 3

Accuracy indices 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 procedures (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]$ and $\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]+$ when they have better or comparable performances.

Experiment	Sqrt	His6	DopReg	Dop2bin
s	$\sqrt{\cdot}$	His_6	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	$\mathbf{1.052 \pm 0.01}$
Mal+	$\mathbf{1.989 \pm 0.08}$	$\mathbf{1.799 \pm 0.09}$	1.090 ± 0.00	$\mathbf{1.047 \pm 0.01}$
Mal*	2.483 ± 0.12	2.021 ± 0.11	$\mathbf{1.013 \pm 0.01}$	1.061 ± 0.01
Mal*+	$\mathbf{2.075 \pm 0.09}$	$\mathbf{1.836 \pm 0.10}$	1.070 ± 0.00	$\mathbf{1.041 \pm 0.01}$
$\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]$	2.365 ± 0.11	$\mathbf{1.805 \pm 0.10}$	1.025 ± 0.01	1.056 ± 0.01
$\mathbb{E}[\mathrm{pen}_{\mathrm{id}}]+$	$\mathbf{2.012 \pm 0.09}$	$\mathbf{1.632 \pm 0.08}$	1.083 ± 0.00	$\mathbf{1.040 \pm 0.01}$
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	$\mathbf{1.049 \pm 0.01}$
10-FCV	2.571 ± 0.13	2.074 ± 0.11	1.043 ± 0.01	$\mathbf{1.051 \pm 0.01}$
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	$\mathbf{1.884 \pm 0.10}$	1.043 ± 0.01	1.055 ± 0.01
penRho	2.448 ± 0.12	$\mathbf{1.907 \pm 0.11}$	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+	$\mathbf{2.036 \pm 0.09}$	$\mathbf{1.746 \pm 0.09}$	1.092 ± 0.00	1.058 ± 0.01
penRho+	$\mathbf{2.053 \pm 0.09}$	$\mathbf{1.747 \pm 0.09}$	1.091 ± 0.00	1.059 ± 0.01
penLoo+	$\mathbf{2.152 \pm 0.10}$	$\mathbf{1.858 \pm 0.10}$	1.082 ± 0.00	$\mathbf{1.048 \pm 0.01}$
penEfr+	2.205 ± 0.11	$\mathbf{1.924 \pm 0.11}$	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	$\mathbf{1.869 \pm 0.09}$	1.082 ± 0.00	$\mathbf{1.050 \pm 0.01}$
pen20-F+	2.180 ± 0.10	$\mathbf{1.832 \pm 0.09}$	1.079 ± 0.00	$\mathbf{1.052 \pm 0.01}$
$C'_{\mathrm{or}}/C_{\mathrm{or}}$	0.795	0.996	0.998	0.977

Table 4
Accuracy indices $C_{\text{path}-\text{or}}$ 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 procedures (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+	$\mathbf{1.921 \pm 0.03}$	3.500 ± 0.09	$\mathbf{1.002 \pm 0.001}$	1.354 ± 0.008
Mal⋆	2.168 ± 0.04	2.907 ± 0.07	1.045 ± 0.003	1.453 ± 0.006
Mal⋆+	$\mathbf{1.941 \pm 0.03}$	2.645 ± 0.06	$\mathbf{1.004 \pm 0.001}$	1.487 ± 0.005
$\mathbb{E}[\text{pen}_{\text{id}}]$	2.053 ± 0.04	2.458 ± 0.06	1.029 ± 0.003	$\mathbf{1.050 \pm 0.002}$
$\mathbb{E}[\text{pen}_{\text{id}}]+$	$\mathbf{1.903 \pm 0.03}$	$\mathbf{2.142 \pm 0.04}$	$\mathbf{1.003 \pm 0.001}$	$\mathbf{1.038 \pm 0.002}$
2-FCV	2.230 ± 0.05	2.755 ± 0.06	$\mathbf{1.002 \pm 0.001}$	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	$\mathbf{1.051 \pm 0.002}$
penRho	2.111 ± 0.04	2.726 ± 0.07	1.018 ± 0.002	$\mathbf{1.051 \pm 0.002}$
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	$\mathbf{1.052 \pm 0.003}$
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+	$\mathbf{1.917 \pm 0.03}$	$\mathbf{2.304 \pm 0.06}$	$\mathbf{1.002 \pm 0.001}$	$\mathbf{1.045 \pm 0.002}$
penRho+	$\mathbf{1.915 \pm 0.03}$	$\mathbf{2.308 \pm 0.06}$	$\mathbf{1.002 \pm 0.001}$	$\mathbf{1.045 \pm 0.002}$
penLoo+	$\mathbf{1.959 \pm 0.03}$	$\mathbf{2.397 \pm 0.06}$	$\mathbf{1.004 \pm 0.001}$	$\mathbf{1.045 \pm 0.002}$
penEfr+	2.155 ± 0.05	2.841 ± 0.08	1.011 ± 0.002	$\mathbf{1.046 \pm 0.002}$
penPoi+	2.179 ± 0.05	2.855 ± 0.08	1.011 ± 0.002	$\mathbf{1.045 \pm 0.002}$
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	$\mathbf{1.006 \pm 0.001}$	$\mathbf{1.052 \pm 0.002}$
pen10-F+	1.997 ± 0.03	2.436 ± 0.05	$\mathbf{1.005 \pm 0.001}$	$\mathbf{1.048 \pm 0.002}$
pen20-F+	2.018 ± 0.04	2.416 ± 0.06	$\mathbf{1.004 \pm 0.001}$	$\mathbf{1.047 \pm 0.002}$

Table 5

Accuracy indices $C_{\text{path-or}}$ 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 procedures (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	$\mathbf{1}_{x\geq 1/2}$
n (sample size)	1000	200	200	200
$\mathcal{M}_n$	regular	regular	regular	2 bin sizes
Mal	$\mathbf{1.704 \pm 0.04}$	$\mathbf{1.654 \pm 0.03}$	$\mathbf{1.407 \pm 0.02}$	7.212 ± 0.40
Mal+	$\mathbf{1.670 \pm 0.03}$	$\mathbf{1.636 \pm 0.03}$	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*+	$\mathbf{1.664 \pm 0.03}$	2.175 ± 0.05	3.719 ± 0.08	4.284 ± 0.25
$\mathbb{E}[\text{pen}_{\text{id}}]$	1.793 ± 0.04	$\mathbf{1.611 \pm 0.03}$	$\mathbf{1.378 \pm 0.01}$	$\mathbf{2.785 \pm 0.19}$
$\mathbb{E}[\text{pen}_{\text{id}}]+$	$\mathbf{1.194 \pm 0.02}$	$\mathbf{1.177 \pm 0.02}$	$\mathbf{1.128 \pm 0.01}$	$\mathbf{1.337 \pm 0.07}$
2-FCV	$\mathbf{1.721 \pm 0.04}$	1.723 ± 0.04	$\mathbf{1.400 \pm 0.02}$	3.507 ± 0.19
5-FCV	1.801 ± 0.06	1.740 ± 0.04	$\mathbf{1.399 \pm 0.02}$	3.486 ± 0.24
10-FCV	1.802 ± 0.05	1.735 ± 0.04	$\mathbf{1.388 \pm 0.02}$	3.149 ± 0.20
20-FCV	1.832 ± 0.05	1.687 ± 0.03	$\mathbf{1.388 \pm 0.02}$	3.257 ± 0.23
LOO	1.815 ± 0.05	1.685 ± 0.04	$\mathbf{1.385 \pm 0.01}$	3.127 ± 0.24
penRad	1.796 ± 0.05	$\mathbf{1.655 \pm 0.04}$	$\mathbf{1.411 \pm 0.02}$	2.932 ± 0.22
penRho	1.796 ± 0.05	$\mathbf{1.666 \pm 0.04}$	$\mathbf{1.409 \pm 0.02}$	2.951 ± 0.23
penLoo	1.825 ± 0.05	1.687 ± 0.04	$\mathbf{1.385 \pm 0.01}$	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	$\mathbf{1.675 \pm 0.04}$	1.423 ± 0.02	3.650 ± 0.29
penRho(20)	1.768 ± 0.05	1.689 ± 0.04	$\mathbf{1.403 \pm 0.02}$	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	$\mathbf{1.394 \pm 0.02}$	3.839 ± 0.27
pen5-F	1.852 ± 0.05	$\mathbf{1.675 \pm 0.04}$	$\mathbf{1.404 \pm 0.02}$	3.237 ± 0.23
pen10-F	1.812 ± 0.05	1.767 ± 0.04	$\mathbf{1.381 \pm 0.01}$	3.093 ± 0.23
pen20-F	1.839 ± 0.05	1.706 ± 0.03	$\mathbf{1.391 \pm 0.01}$	3.123 ± 0.23
penRad+	$\mathbf{1.665 \pm 0.03}$	$\mathbf{1.615 \pm 0.03}$	1.427 ± 0.02	$\mathbf{2.502 \pm 0.15}$
penRho+	$\mathbf{1.665 \pm 0.03}$	$\mathbf{1.619 \pm 0.03}$	1.428 ± 0.02	$\mathbf{2.511 \pm 0.15}$
penLoo+	$\mathbf{1.673 \pm 0.03}$	$\mathbf{1.624 \pm 0.03}$	$\mathbf{1.409 \pm 0.02}$	$\mathbf{2.659 \pm 0.18}$
penEfr+	$\mathbf{1.683 \pm 0.03}$	1.730 ± 0.04	$\mathbf{1.413 \pm 0.02}$	3.098 ± 0.25
penPoi+	$\mathbf{1.683 \pm 0.03}$	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+	$\mathbf{1.732 \pm 0.04}$	$\mathbf{1.664 \pm 0.03}$	$\mathbf{1.408 \pm 0.02}$	$\mathbf{2.890 \pm 0.22}$
pen10-F+	$\mathbf{1.663 \pm 0.04}$	$\mathbf{1.657 \pm 0.03}$	$\mathbf{1.394 \pm 0.02}$	$\mathbf{2.810 \pm 0.21}$
pen20-F+	$\mathbf{1.680 \pm 0.04}$	$\mathbf{1.623 \pm 0.03}$	$\mathbf{1.397 \pm 0.01}$	$\mathbf{2.657 \pm 0.19}$

Table 6

Accuracy indices $C_{\mathrm{path-or}}$ 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 procedures (taking the uncertainty into account) are bolded, as well as $\mathbb{E}[\mathrm{pen_{id}}]$ and $\mathbb{E}[\mathrm{pen_{id}}]+$ when they have better or comparable performances.

Experiment	Sqrt	His6	DopReg	Dop2bin
s	$\sqrt{\cdot}$	His_6	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	$\mathbf{1.049 \pm 0.00}$
Mal+	$\mathbf{2.232 \pm 0.10}$	$\mathbf{2.041 \pm 0.12}$	1.094 ± 0.00	$\mathbf{1.045 \pm 0.01}$
Mal⋆	2.838 ± 0.15	2.533 ± 0.21	$\mathbf{1.013 \pm 0.00}$	1.057 ± 0.00
Mal⋆+	$\mathbf{2.349 \pm 0.11}$	$\mathbf{2.168 \pm 0.16}$	1.073 ± 0.00	$\mathbf{1.038 \pm 0.00}$
$\mathbb{E}[\mathrm{pen_{id}}]$	2.678 ± 0.14	$\mathbf{2.182 \pm 0.17}$	1.026 ± 0.00	1.053 ± 0.00
$\mathbb{E}[\mathrm{pen_{id}}]+$	$\mathbf{1.348 \pm 0.07}$	$\mathbf{1.230 \pm 0.06}$	1.050 ± 0.00	$\mathbf{1.038 \pm 0.00}$
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	$\mathbf{1.046 \pm 0.00}$
10-FCV	2.912 ± 0.16	2.639 ± 0.21	1.045 ± 0.00	$\mathbf{1.047 \pm 0.00}$
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	$\mathbf{2.272 \pm 0.19}$	1.044 ± 0.00	1.052 ± 0.00
penRho	2.755 ± 0.14	$\mathbf{2.291 \pm 0.19}$	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+	$\mathbf{2.291 \pm 0.11}$	$\mathbf{2.018 \pm 0.14}$	1.095 ± 0.00	1.056 ± 0.01
penRho+	$\mathbf{2.317 \pm 0.11}$	$\mathbf{2.019 \pm 0.14}$	1.095 ± 0.00	1.057 ± 0.01
penLoo+	$\mathbf{2.437 \pm 0.12}$	$\mathbf{2.218 \pm 0.18}$	1.085 ± 0.00	$\mathbf{1.045 \pm 0.00}$
penEfr+	2.495 ± 0.13	$\mathbf{2.348 \pm 0.19}$	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	$\mathbf{2.364 \pm 0.20}$	1.085 ± 0.00	$\mathbf{1.047 \pm 0.01}$
pen20-F+	2.497 ± 0.13	$\mathbf{2.318 \pm 0.20}$	1.082 ± 0.00	$\mathbf{1.049 \pm 0.01}$

4. Complete proofs of some concentration inequalities

We here give complete proofs and key results of several concentration inequalities stated in [2] which come from [3].

*complete proof of Proposition **10**.* According to the explicit expressions **(29)** and **(30)**, $\widetilde{p_1}(m)$ and $p_2(m)$ are both U-statistics of order 2 conditionally on $(\mathbf{1}_{X_i \in I_\lambda})_{(i,\lambda)}$. Then, we use Lemma 5 with $\xi_{i,\lambda} = Y_i - \beta_\lambda$, $a_\lambda = 0$, $b_\lambda = p_\lambda (n \widehat{p}_\lambda)^{-2}$ for $\widetilde{p_1}$ and $b_\lambda = (n^2 \widehat{p}_\lambda)^{-1}$ for p_2. This proves, for every $q \geq 2$,

$$\left\| \widetilde{p_1}(m) - \mathbb{E}^{\Lambda_m}[\widetilde{p_1}(m)] \right\|_q^{(\Lambda_m)} \leq \max_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{\widehat{p}_\lambda} \mathbf{1}_{\widehat{p}_\lambda > 0} \right\} L_{a_\ell, \xi_\ell} D_m^{-1/2} q \mathbb{E}\left[p_2(m) \right] \quad (3)$$

$$\left\| p_2(m) - \mathbb{E}[p_2(m)] \right\|_q^{(\Lambda_m)} \leq L_{a_\ell, \xi_\ell} D_m^{-1/2} q \mathbb{E}\left[p_2(m) \right] \quad . \quad (4)$$

We deduce conditional concentration inequalities from these moment inequalities [1, Lemma 8.9] 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 $\widetilde{p_1}$, we use (9) in Lemma 9.

We now have to control the distance between $\mathbb{E}^{\Lambda_m}\left[\widetilde{p_1} \right]$ and $\mathbb{E}\left[\widetilde{p_1} \right]$. First, if $B_n \geq 1$, we can use Lemma 6: taking $X_\lambda = n \widehat{p}_\lambda$ and $a_\lambda = p_\lambda \left(\sigma_\lambda \right)^2$, according to **(29)**, we have $\widetilde{p_1}(m) = Z_{m,1}$ and the concentration inequality for $\widetilde{p_1}$ follows. Second, if we only know that $B_n > 0$, instead of using Lemma 6, we remark that

$$\mathbb{E}^{\Lambda_m}\left[\widetilde{p_1}(m) \right] \geq \min_{\lambda \in \Lambda_m} \left\{ \frac{p_\lambda}{\widehat{p}_\lambda} \right\} \mathbb{E}^{\Lambda_m}\left[p_2(m) \right] \quad ,$$

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

Lemma 5 (Lemma 5 of [3])**.** *Let $(a_\lambda)_{\lambda \in \Lambda_m}$, $(b_\lambda)_{\lambda \in \Lambda_m}$ be two families of real numbers and $(r_\lambda)_{\lambda \in \Lambda_m}$ be a family of integers. For every $\lambda \in \Lambda_m$, 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 $q \geq 2$. Let $S_{\lambda,1}$, $S_{\lambda,2}$ and Z be defined as follows:*

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

Then, an absolute constant $\kappa \leq 1.271$ exists such that for every $q \geq 2$,

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

Lemma 6 (Lemma 4 of [3]). *Let $(a_\lambda)_{\lambda\in\Lambda_m}$ be a family of real numbers, $(p_\lambda)_{\lambda\in\Lambda_m}$ be a family of non-negative real numbers such that $\sum_{\lambda\in\Lambda_m} p_\lambda = 1$, $n \in \mathbb{N}$, $(X_\lambda)_{\lambda\in\Lambda_m}$ be a multinomial random vector with parameters $(n;(p_\lambda)_{\lambda\in\Lambda_m})$ and define*

$$Z_{m,T} := \sum_{\lambda\in\Lambda_m} \left(a_\lambda \min\left(T, X_\lambda^{-1}\right)\right) \qquad \text{for any } T \in (0,1] \enspace .$$

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 $t \geq 0$, $\varphi_1(t) = \max(t,1)e^{-\max(t,1)}$. Then, the two following deviation inequalities hold.

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

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

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

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

In the proof of Lemma **12**, we need the following two consequences of [7, Section 5.3.5] which can be found in [1].

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

$$Z = f(X_1, \ldots, X_n) \enspace .$$

Then, an absolute constant $\kappa \leq 1.271$ exists such that for every $q \geq 2$,

$$\left\lVert Z - \mathbb{E}[Z]\right\rVert_q \leq 2\sqrt{\kappa}\sqrt{q\left\lVert\sum_{i=1}^n \left(Z - \mathbb{E}\left[Z \mid (X_j)_{j\neq i}\right]\right)^2\right\rVert_{q/2}} \enspace . \tag{8}$$

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

$$\left\lVert\left(Z - \mathbb{E}\left[Z\right]\right)_+\right\rVert_q \leq \sqrt{\kappa q \left\lVert V\right\rVert_{q/2}} \enspace .$$

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

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

$$\begin{aligned}\|S - \mathbb{E}[S]\|_q &\leq 2\sqrt{\kappa}\sqrt{q\left\|\sum_{i=1}^s \mathbb{E}\left[(S - \mathbb{E}[S \mid X_i])^2 \mid X_{1\ldots n}\right]\right\|_{q/2}} \\ &= 2\sqrt{\kappa}\sqrt{q\left\|\sum_{i=1}^s X_i^2\right\|_{q/2}} \leq 2\sqrt{\kappa}\sqrt{q\sum_{i=1}^s \|X_i\|_q^2} \enspace .\end{aligned}$$

□

5. Proofs of some technical lemmas

We here give the complete proofs of the technical lemmas stated in Section **8.4**. First, we add to Lemma **13** two inequalities that we used in Section 4.

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

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

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

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

proof of Lemma 9. The three inequalities come from Bernstein's inequality (for instance [7, Proposition 2.9]) applied to $n\widehat{p}_\lambda$: for every $\lambda \in \Lambda_m$, an event Ω_λ of probability $1 - 2n^{-(\gamma+1)}$ exists on which

$$n p_\lambda - \sqrt{2 n p_\lambda (\gamma + 1) \ln(n)} - \frac{(\gamma + 1) \ln(n)}{3} \leq n\widehat{p}_\lambda \tag{12}$$

$$\leq n p_\lambda + \sqrt{2 n p_\lambda (\gamma + 1) \ln(n)} + \frac{(\gamma + 1) \ln(n)}{3} \enspace . \tag{13}$$

For proving (9), for any $\lambda \in \Lambda_m$ such that $np_\lambda \geq 8(\gamma+1)\ln(n)$, (12) gives the result. Otherwise, remark only that $(p_\lambda / \widehat{p}_\lambda) \mathbf{1}_{\widehat{p}_\lambda > 0} \leq np_\lambda \leq 8(\gamma+1)\ln(n)$.

For proving (10), use (13) and remark that $np_\lambda (\gamma+1)\ln(n) B_n^{-1} \geq (\gamma+1)\ln(n)$.

For proving (11), use (12) and remark that $\sqrt{2np_\lambda(\gamma+1)\ln(n)} \leq (np_\lambda)/2 + (\gamma+1)\ln(n)$.

To conclude, (9)–(11) hold on $\Omega = \bigcap_{\lambda\in\Lambda_m} \Omega_\lambda$; the union bound proves yields that $\mathbb{P}(\Omega) \geq 1 - \operatorname{Card}(\Lambda_m) 2n^{-(\gamma+1)} \geq 1 - 2n^{-\gamma}$. □

*proof of Lemma **8**.* Remark that for every $\lambda \in \Lambda_m$ such that σ does not jump on I_λ,

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

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

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

Therefore,

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

□

*proof of Lemma **9**.* From (**Al**) and the upper bound in ($\mathbf{Ar}_{\ell,\mathbf{u}}$),

$$\|s - s_m\|_\infty \leq B \max_{\lambda\in\Lambda_m} \{\operatorname{diam}(I_\lambda)\} = B \max_{\lambda\in\Lambda_m} \{\operatorname{Leb}(I_\lambda)\} \leq B c_{\mathrm{r,u}} D_m^{-1} \enspace . \tag{14}$$

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

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

and $\mu = \mathcal{D}(X)$. Then, using ($\mathbf{Ad}_\ell$),

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

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

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

so that

$$\|s - s_m\|^2_{L^2(\mu)} \geq c_X^{\min} \sum_{\lambda \in \Lambda_m^J} \frac{B_0^2 \operatorname{Leb}(I_\lambda)^3}{12} \geq \frac{c_X^{\min} B_0^2 c_{\mathrm{r},\ell}^3}{12 D_m^3} \operatorname{Card}\left(\Lambda_m^J\right)$$

$$\geq \frac{c_X^{\min} B_0^2 c_{\mathrm{r},\ell}^3}{12 D_m^3} \left(\frac{c_J D_m}{c_{\mathrm{r},\mathrm{u}}} - 2 \right)_+ \enspace .$$

Combined with (14), the result follows. □

References

[1] Sylvain Arlot. *Resampling and Model Selection*. PhD thesis, University Paris-Sud 11, December 2007. oai:tel.archives-ouvertes.fr:tel-00198803_v1.

[2] Sylvain Arlot. Model selection by resampling penalization. *Electron. J. Stat.*, 3:557–624, 2009.

[3] Sylvain Arlot. V-fold cross-validation improved: V-fold penalization, February 2008. arXiv:0802.0566v2.

[4] H. G. Burchard and D. F. Hale. Piecewise polynomial approximation on optimal meshes. *J. Approximation Theory*, 14(2):128–147, 1975. MR0374761

[5] Ronald A. DeVore and George G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. MR1261635

[6] David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90(432):1200–1224, 1995. MR1379464

[7] Pascal 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. MR2319879