Optimal model selection for stationary data under various mixing conditions

Matthieu Lerasle

Optimal model selection for stationary data under various mixing conditions.

Matthieu Lerasle*

Abstract:

We build penalized least-squares estimators of the marginal density of a stationary process, using the slope algorithm and resampling penalties. When the data are β or τ-mixing, these estimators satisfy oracle inequalities with leading constant asymptotically equal to 1.

Key words: Density estimation, optimal model selection, resampling methods, slope heuristic, weak dependence.
2000 Mathematics Subject Classification: 62G07, 62G09, 62M99.

1 Introduction

The history of statistical model selection goes back at least to Akaike [Aka70], [Aka73] and Mallows [Mal73]. They proposed to select among a collection of parametric models the one which minimizes an empirical loss plus some penalty term proportional to the dimension of the models. Birgé & Massart [BM97] and Barron, Birgé & Massart [BBM99] generalize this approach, making the link between model selection and adaptive estimation. They also proved that several estimation procedures as cross-validation (Rudemo [Rud82]) or hard thresholding (Donoho *et.al.* [DJKP96]) can be interpreted in terms of model selection.
More recently, Birgé & Massart [BM07], Arlot & Massart [AM09] and Arlot [Arl07], [Arl09] arised the problem of optimal model selection. Basically, the aim is to select an estimator satisfying an oracle inequality with leading constant asymptotically equal to 1.
Two totally data driven procedures are known to achieve this goal: the slope algorithm, introduced by Birgé & Massart [BM07] and the resampling penalties defined by Arlot [Arl09]. Arlot & Massart [AM09] and Arlot [Arl09] proved that these estimators are efficient to select the best histogram in a general regression framework. In [Ler09b], we proved that these procedures are also optimal in density estimation, when the data are independent.
There exists a lot of statistical frameworks where the data are not independent. The previous results may therefore not hold. Baraud *et.al.* [BCV01] proved that penalties proportional to the dimension can also be used when the data are β-mixing (for a definition of the coefficient β, see Rozanov & Volkonskii [VR59] or Section 2). They worked in a regression framework and Comte & Merlevède [CM02] extended the result to density estimation. In [Ler09a], we proved that the same penalties can also be used with τ-mixing data (the coefficient τ has been introduced by Dedecker & Prieur [DP05], see Section 2).
The main problem of the algorithm proposed by Comte & Merlevède [CM02] is that the penalty term involves a constant depending on the mixing coefficients (both in the β and

*Institut de Mathématiques (UMR 5219), INSA de Toulouse, Université de Toulouse, France

τ-mixing cases) which is typically unknown in practice.
As in the independent case, we prove that a resampling estimator catches the shape of the ideal penalty with great generality as it "learns" part of the mixing structure of the data (Künsch [Kün89], Liu & Singh [LS92]). We will also prove that the slope algorithm can be used to calibrate in an optimal way the constant in front of the penalty term. The new penalization procedure is totally data driven.
Let us now explain more precisely the problem that we will consider.

1.1 Least-squares estimators

We observe n real valued, identically distributed random variables $X_1, ..., X_n$, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with common law P. We assume that a measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is given. We denote by $L^2(\mu)$ the Hilbert space of square integrable real valued functions and by $\|.\|$ the associated L^2-norm. The parameter of interest is the density s of P with respect to μ, we assume that it belongs to $L^2(\mu)$. For all function g in $L^1(P)$, we define

$$Pg = \int_{\mathbb{R}} gsd\mu = \mathbb{E}\left(g(X)\right),\ P_n g = \frac{1}{n}\sum_{i=1}^{n} g(X_i),$$

where X is a copy of X_1, independent of $(X_1, ..., X_n)$. s minimizes the integrated contrast $t \mapsto \|t\|^2 - 2Pt$ over $L^2(\mu)$. The risk of an estimator $\hat{s}$ of s is measured with the L^2-loss, that is $\|s - \hat{s}\|^2$, which is random when $\hat{s}$ is.
The problem of density estimation is a problem of M-estimation. These problems are classically solved in two steps when the data are independent. First, we choose a "model" S_m close to the parameter s, which means that $\inf_{t \in S_m} \|s - t\|^2$ is "small". Then, we minimize over S_m the empirical version of the integrated contrast, that is, we choose

$$\hat{s}_m \in \arg\min_{t \in S_m} \|t\|^2 - 2P_n t.$$

When the data are mixing, the coupling method is a very powerful tool to extend the methods developed in the independent case. It can be summarized as follows.

Coupling method: Let $I_0, J_0, ..., I_{p-1}, J_{p-1}$ be a partition of $\{1, ..., n\}$ satisfying $q = \min_{k=0,...,p-1} \min(I_{k+1}) - \max(I_k) > 0$ (for a proper definition of this partition see Section 2). For all $k = 0, ..., p-1$, let $A_k = (X_l)_{l \in I_k}$ and let l_k be the length of I_k. A coupling lemma associates to the sequence $(A_k)_{k=0,...,p-1}$ independent random variables (A_k^*) such that $\mathbb{E}\left(d(A_k, A_k^*)\right) \leq \gamma(q)$, where γ is the mixing coefficient of the data, d is a distance on $\mathbb{R}^{l_k}$. Let $I = \cup_{k=0}^{p-1} I_k$ and let P_A be the empirical process based on the data $(X_i,\ i \in I)$, that is $P_A = \sum_{i \in I} \delta_{X_i}/|I|$. To bound quantities of the form $F(P_n)$, built with the empirical process, we first use algebraic inequalities to obtain

$$F(P_n) \leq CF(P_A). \tag{1}$$

Then we have

$$F(P_A) \leq F(P_{A^*}) + |F(P_A) - F(P_{A^*})|.$$

We can now use the results available for independent random variables to bound $F(P_{A^*})$ and the mixing properties to bound $|F(P_A) - F(P_{A^*})|$.

Up to our knowledge, all the model selection procedures proposed for mixing data use the coupling methods. In this scheme, the bounds given on $F(P_n)$ are the same as those given for $F(P_A)$ and the only essentially suboptimal bound is the first one: $F(P_n) \leq CF(P_A)$.

We extend the procedures developed in the independent case in [Ler09b] through the coupling method. As we are looking for optimal results, we will work with the process P_A instead of P_n, avoiding the lost (1). The counterpart of this choice is that we do not use all the data to build our estimator. In particular, the variance of an oracle built only with the variables $(X_i)_{i\in I}$ is bigger than the one of an oracle built with all the sample when the data are independent. However, we will see in Section 4 that our final estimator improves the previous procedures proposed in a mixing setting.
Let us now define the least-squares estimators by $\hat{s}_{A,m} \in \arg\min_{t\in S_m} P_A Q(t)$. The minimization problem defining $\hat{s}_{A,m}$ can be computationally untractable for general sets S_m, leading to untractable procedures. However, in density estimation, it can be easily solved when S_m is a linear subspace of $L^2(\mu)$ since, for any orthonormal basis $(\psi_\lambda)_{\lambda\in m}$ of S_m,

$$\hat{s}_{A,m} = \sum_{\lambda\in m} (P_A\psi_\lambda)\psi_\lambda.$$

The risk of $\hat{s}_{A,m}$ is decomposed in the classical bias and variance terms thanks to Pythagoras relation. Let s_m be the orthogonal projection of s onto S_m, then

$$\|s - \hat{s}_{A,m}\|^2 = \|s - s_m\|^2 + \|s_m - \hat{s}_{A,m}\|^2. \tag{2}$$

The space S_m should be chosen in order to realize a trade-off between those quantities. In [Ler09b], we proved a concentration inequality for $\|s_m - \hat{s}_{A,m}\|^2$ around its expectation when the data are independent. It proves that $D^*_{A,m} = n\mathbb{E}(\|s_m - \hat{s}_{A,m}\|^2)$ is a natural complexity measure of S_m and, when the models S_m are sufficiently regular, we recovered that the dimension d_m of S_m has the same order as $D^*_{A,m}$. However, this is not true in general, because there exist simple models (histograms with a small d_m) where $D^*_{A,m} >> d_m$ and model of infinite dimension where $D^*_{A,m}$ behaves nicely (see Birgé [Bir08] or Section 4).

1.2 Model selection

The choice of a "good" model S_m is impossible without strong assumptions on s, for example that we have precise information on its regularity. However, if we only assume that s is regular, it is possible to choose a collection of models $(S_m)_{m\in\mathcal{M}_n}$ such that one of them realizes an optimal trade-off (see for example Birgé & Massart [BM97] or Barron, Birgé & Massart [BBM99]). Given the projection estimators $(\hat{s}_{A,m})_{m\in\mathcal{M}_n}$ associated to this collection, the aim is then to build an estimator $\hat{m}$ such that the final estimator, $\tilde{s} = \hat{s}_{A,\hat{m}}$ behaves almost as well as any model m_o in the set of oracles

$$\mathcal{M}^*_n = \{m_o \in \mathcal{M}_n,\ \|\hat{s}_{A,m_o} - s\|^2 = \inf_{m\in\mathcal{M}_n} \|\hat{s}_{A,m} - s\|^2\}.$$

This is the problem of model selection. More precisely, we want the final estimator $\tilde{s} = \hat{s}_{A,\hat{m}}$ to satisfy one of the following type of oracle inequalities

$$\exists K > 0, C_n > 0, \gamma > 1,\ \mathbb{P}\left(\|\tilde{s} - s\|^2 > C_n \inf_{m\in\mathcal{M}_n} \left\{\|s - \hat{s}_{A,m}\|^2\right\}\right) \le \frac{K}{n^\gamma}. \tag{3}$$

$$\exists K > 0, C_n > 0,\ \mathbb{E}\left(\|\tilde{s} - s\|^2\right) \le C_n \mathbb{E}\left(\inf_{m\in\mathcal{M}_n} \left\{\|s - \hat{s}_{A,m}\|^2\right\}\right) + \frac{K}{n}. \tag{4}$$

In both cases, the leading constant C_n should be as close as possible to 1. In order to build $\hat{m}$, remark that, for all m in $\mathcal{M}_n$,

$$\|s - \hat{s}_{A,m}\|^2 - \|s\|^2 = \|\hat{s}_{A,m}\|^2 - 2P_A\hat{s}_{A,m} + 2\nu_A(\hat{s}_{A,m}),$$

where $\nu_A = P_A - P$. An oracle minimizes $\|s - \hat{s}_{A,m}\|^2 - \|s\|^2$ over $\mathcal{M}_n$. As we want to imitate the oracle, we will design a map pen : $\mathcal{M}_n \to \mathbb{R}^+$ and choose

$$\hat{m} \in \arg\min_{m\in\mathcal{M}_n} \|\hat{s}_{A,m}\|^2 - 2P_A\hat{s}_{A,m} + \mathrm{pen}(m),\ \ \tilde{s} = \hat{s}_{A,\hat{m}}. \tag{5}$$

It is clear that the ideal penalty is $\mathrm{pen}_{id}(m) = 2\nu_A(\hat{s}_{A,m})$ and our goal is to design sharp estimators of this quantity as penalty functions.
The key point to obtain oracle inequalities is the following decomposition of the risk of $\tilde{s}$. For all m in $\mathcal{M}_n$, let

$$p(m) = \nu_A(\hat{s}_{A,m} - s_m) = \|\hat{s}_{A,m} - s_m\|^2.$$

For all m in $\mathcal{M}_n$,

$$\begin{aligned}\|s-\tilde{s}\|^2 &= \|\tilde{s}\|^2 - 2P\tilde{s} + \|s\|^2 = \|\tilde{s}\|^2 - 2P_A\tilde{s} + 2\nu_A\tilde{s} + \|s\|^2\\ &\le \|\hat{s}_{A,m}\|^2 - 2P_A\hat{s}_{A,m} + \mathrm{pen}(m) + (2\nu_A(\tilde{s}) - \mathrm{pen}(\hat{m})) + \|s\|^2\\ &= \|s - \hat{s}_{A,m}\|^2 + (\mathrm{pen}(m) - 2\nu_A(\hat{s}_{A,m})) + (2\nu_A(\tilde{s}) - \mathrm{pen}(\hat{m}))\end{aligned}$$

Thus, for all m in $\mathcal{M}_n$,

$$\|s-\tilde{s}\|^2 \le \|s - \hat{s}_{A,m}\|^2 + (\mathrm{pen}(m) - 2p(m)) + (2p(\hat{m}) - \mathrm{pen}(\hat{m})) + 2\nu_A(s_{\hat{m}} - s_m). \tag{6}$$

1.3 Optimal model selection

Let us now precise the definition of the methods that we will use to calibrate the penalty.

1.3.1 The slope algorithm

The "slope heuristic" was introduced by Birgé & Massart [BM07] in the Gaussian regression framework. It states that there exists a complexity measure Δ_m of S_m and a constant $K_{\min}$ such that

1. if $\mathrm{pen}(m) < K_{\min}\Delta_m$, $\Delta_{\hat{m}}$ is too large, typically $\Delta_{\hat{m}} \ge C\sup_{m\in\mathcal{M}_n}\Delta_m$, where C is a constant independent of n.
2. if $\mathrm{pen}(m) \simeq K\Delta_m$ for some $K > K_{\min}$, then $\Delta_{\hat{m}}$ is "much smaller",
3. if $\mathrm{pen}(m) \simeq 2K_{\min}\Delta_m$, then the risk of the selected estimator satisfies

   $$\|\tilde{s} - s\|^2 \le C_n \inf_{m\in\mathcal{M}_n}\left\{\|s - \hat{s}_{A,m}\|^2\right\},\ \text{with } C_n \to 1,\ \text{when } n\to\infty$$

   in expectation and with large probability.

When both Δ_m and the associated $K_{\min}$ are known, point 3 in this heuristic says that $\mathrm{pen}(m) \simeq 2K_{\min}\Delta_m$ is an optimal penalty. This heuristic is classically used when Δ_m is known and $K_{\min}$ is unknown. Arlot & Massart [AM09] introduced the following algorithm to calibrate the penalty term in this situation.

Slope algorithm

- For all $K > 0$, compute the selected model $\hat{m}(K)$ given by (5) with the penalty $\mathrm{pen}(m) = K\Delta_m$ and the associated complexity $\Delta_{\hat{m}(K)}$.

- Find a constant K_o such that $\Delta_{\hat{m}(K)}$ is large when $K < K_o$, and "much smaller" when $K > K_o$.
- Take the final $\hat{m} = \hat{m}(2K_o)$.

In [Ler09b], we justified the slope heuristic in density estimation with independent data for $\Delta_m = \mathbb{E}(\|s_m - \hat{s}_{A,m}\|^2)$, $K_{\min} = 1$. This complexity is unknown in practice and has to be estimated. We proposed a resampling estimator and proved that it works without extra assumptions on our collection of models. In this paper, we will extend these results to mixing processes.

1.3.2 Resampling penalties

Data-driven penalties have been studied in density estimation, in particular, cross-validation methods as in Stone [Sto74], Rudemo [Rud82] or Celisse [Cél08]. We extend the approach of [Ler09b] based on the resampling penalties introduced by Arlot [Arl09]. We prove that it provides optimal model selection procedures. An important ingredient in the proofs is the coupling properties of mixing processes. The coupling result proved in Viennet [Vie97] for β-mixing processes allows a straightforward extension of the results of [Ler09b]. The coupling lemma available for τ-mixing sequences is not so powerful and in that case, we have to develop new methods of proofs.

The paper is organized as follows. In Section 2, we introduce our new estimation procedure and describe our main assumptions. In Section 3, we state our main results, we prove the efficiency of the penalized least-squares estimators based on the slope heuristic and on resampling methods. In Section 4, we compare our new estimators with those given in [Ler09b]. The proofs of the main theorems are postponed to Section 5. Section 6 is an Appendix where we recall some probabilistic lemmas proved in [Ler09b].

2 New estimation procedures

2.1 Blockwise decomposition of the data

Assume that n is even and let p and q be two integers such that $2pq = n$. For all $k = 0, ..., p-1$, let $I_k = (2kq+1, ..., (2k+1)q)$, $A_k = (X_l)_{l \in I_k}$ and $I = \cup_{k=0}^{p-1} I_k$. For all functions t in $L^2(\mu)$ and all $x_1, ..., x_q$ in $\mathbb{R}$, let

$$L_q(t)(x_1, ..., x_q) = \frac{1}{q}\sum_{i=1}^{q} t(x_i),\ P_A t = \frac{1}{p}\sum_{k=0}^{p-1} L_q(t)(A_k) = \frac{2}{n}\sum_{i \in I} t(X_i),$$

$$\nu_A(t) = (P_A - P)(t).$$

Let S_m be a linear space. The estimator $\hat{s}_{A,m}$ associated to S_m, is defined by

$$\hat{s}_{A,m} \in \arg\min_{t \in S_m} \|t\|^2 - 2P_A t. \tag{7}$$

Given an orthonormal basis $(\psi_\lambda)_{\lambda \in m}$ of S_m, classical computations prove that

$$\hat{s}_{A,m} = \sum_{\lambda \in m} (P_A \psi_\lambda)\psi_\lambda,\ \|s_m - \hat{s}_{A,m}\|^2 = \sum_{\lambda \in m} (\nu_A(\psi_\lambda))^2 = \sup_{t \in B_m} (\nu_A(t))^2.$$

2.2 Resampling penalties

The first penalization procedure is based on the resampling penalties introduced by Arlot [Arl09]. The resampling algorithm is slightly modified in order to keep the dependence structure inside the blocks (see Künsh [Kün89], Liu & Singh [LS92] or Radulovic [Rad02]). Let $W_0, ..., W_{p-1}$ be a resampling scheme, that is, a vector of random variables, independent of $X_1, ..., X_n$ and exchangeable, i.e., for all permutation ξ of $\{0, ...p-1\}$,

$$(W_{\xi(0)}, ..., W_{\xi(p-1)}) \text{ has the same law as } (W_0, ..., W_{p-1}).$$

Let P_A^W and ν_A^W be the associated resampling empirical processes defined, for all t in $L^2(\mu)$, by

$$P_A^W(t) = \frac{1}{p}\sum_{k=0}^{p-1} W_k L_q(t)(A_k),$$

$$\nu_A^W(t) = (P_A^W - \bar{W}_p P_A)(t) = \frac{1}{p}\sum_{k=0}^{p-1}(W_k - \bar{W}_p)L_q(t)(A_k), \text{ where } \bar{W}_p = \frac{1}{p}\sum_{k=0}^{p-1} W_k.$$

For all m in $\mathcal{M}_n$, let

$$\hat{s}_{A,m}^W = \arg\min_{t \in S_m} \|t\|^2 - 2P_A^W(t) = \sum_{\lambda \in m} (P_A^W \psi_\lambda)\psi_\lambda.$$

Setting $v_W^2 = \text{Var}(W_1 - \bar{W}_p)$ and $C_W = v_W^{-2}$, the resampling penalty is defined by

$$\text{pen}(m) = 2C_W \mathbb{E}^W \left(\sup_{t \in B_m} (\nu_A^W(t))^2 \right) = 2C_W \sum_{\lambda \in m} \mathbb{E}^W \left((\nu_A^W(\psi_\lambda))^2 \right). \quad (8)$$

Hereafter, for all m in $\mathcal{M}_n$ and for all function pen, the final estimator is always denoted by

$$\tilde{s} = \hat{s}_{A,\hat{m}}, \text{ where } \hat{m} = \arg\min_{m \in \mathcal{M}_n} \|\hat{s}_m\|^2 - 2P_A \hat{s}_{A,m} + \text{pen}(m). \quad (9)$$

2.3 Some measures of dependence

2.3.1 β-mixing data

The coefficient β was introduced by Rozanov & Volkonskii [VR59]. For a random variable Y defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a σ-algebra $\mathcal{M}$ in $\mathcal{A}$, let

$$\beta(\mathcal{M}, \sigma(Y)) = \mathbb{E}\left(\sup_{A \in \mathcal{B}} |\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)| \right).$$

For all stationary sequence of random variables $(X_n)_{n \in \mathbb{Z}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$, let

$$\beta_k = \beta(\sigma(X_i, i \le 0), \sigma(X_i, i \ge k)).$$

The process $(X_n)_{n\in\mathbb{Z}}$ is said to be β-mixing when $\beta_k \to 0$ as $k \to \infty$. Examples of β-mixing processes can be found in the books of Doukhan [Dou94] and Bradley [Bra07]. One of the most important is the following: a stationary, irreducible, aperiodic and positively recurent Markov chain $(X_i)_{i \ge 1}$ is β-mixing.
Let us recall Lemma 5.1 in Viennet [Vie97].

Lemma: *(Viennet 1997) Assume that the process $(X_1, ..., X_n)$ is β-mixing and let p, q and $A_0, ..., A_{p-1}$ be respectively the integers and the random variables defined in Section 2.1. There exist random variables $A_0^*, ..., A_{p-1}^*$ such that:*

1. *for all* $k = 0, ..., p-1$, $A_k^* = (X_{2kq+1}^*, ..., X_{(2k+1)q}^*)$ *has the same law as* A_k,

2. *for all* $k = 0, ..., p-1$, A_k^* *is independent of* $A_0, ..., A_{k-1}, A_0^*, ..., A_{k-1}^*$,

3. *for all* $k = 0, ..., p-1$, $\mathbb{P}(A_k \neq A_k^*) \leq \beta_q$.

2.3.2 τ-mixing data

The coefficient τ was introduced by Dedecker & Prieur [DP05]. For all l in $\mathbb{N}^*$, for all x, y in $\mathbb{R}^l$, let $d_l(x, y) = \sum_{i=1}^l |x_i - y_i|$. For all l in $\mathbb{N}^*$, for all function t defined on $\mathbb{R}^l$, the Lipschitz semi-norm of t is defined by

$$\mathrm{Lip}_l(t) = \sup_{x \neq y \in \mathbb{R}^l} \frac{|t(x) - t(y)|}{d_l(x, y)}.$$

For all functions t defined on $\mathbb{R}$, we will denote for short by $\mathrm{Lip}(t) = \mathrm{Lip}_1(t)$. Let λ_1 be the set of all functions $t : \mathbb{R}^l \to \mathbb{R}$ such that $\mathrm{Lip}_l(t) \leq 1$. For all integrable, $\mathbb{R}^l$-valued, random variables Y defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and all σ-algebra $\mathcal{M}$ in $\mathcal{A}$, let

$$\tau(\mathcal{M}, Y) = \mathbb{E}\left(\sup_{t \in \lambda_1} |\mathbb{P}_{Y|\mathcal{M}}(t) - \mathbb{P}_Y(t)| \right).$$

For all stationary sequences of integrable random variables $(X_n)_{n \in \mathbb{Z}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$, for all integers k, r, let

$$\tau_{k,r} = \max_{1 \leq l \leq r} \frac{1}{l} \sup_{k \leq i_1 < .. < i_l} \{\tau(\sigma(X_p, p \leq 0), (X_{i_1}, ..., X_{i_l}))\}, \ \tau_k = \sup_{r \in \mathbb{N}^*} \tau_{k,r}.$$

The process $(X_n)_{n \in \mathbb{Z}}$ is said to be τ-mixing when $\tau_k \to 0$ as $k \to \infty$. Examples of τ-mixing processes can be found in the book of Dedecker *et. al* [DDL+07] or the articles of Dedecker & Prieur [DP05] and Comte *et. al* [CDT08].
The following result has been obtained in Claim 1 in the proof of Theorem 4.1 of [Ler09a]. This is a consequence of a coupling lemma proved by Dedecker & Prieur [DP05].

Lemma: [τ-coupling, Claim 1 p17 in [Ler09a]] *Assume that the process* $(X_1, ..., X_n)$ *is* τ*-mixing and let* p, q *and* $A_0, ..., A_{p-1}$ *be respectively the integers and the random variables defined in Section 2.1. There exist random variables* $A_0^*, ..., A_{p-1}^*$ *such that:*

1. *for all* $k = 0, ..., p-1$, $A_k^* = (X_{2kq+1}^*, ..., X_{(2k+1)q}^*)$ *has the same law as* A_k,

2. *for all* $k = 0, ..., p-1$, A_k^* *is independent of* $A_0, ..., A_{k-1}, A_0^*, ..., A_{k-1}^*$,

3. *for all* $k = 0, ..., p-1$, $\mathbb{E}(d_q(A_k, A_k^*)) \leq q\tau_q$.

2.4 Main assumptions

Let p, q and $A_0, ..., A_{p-1}$ be respectively the integers and the random variables defined in Section 2.1. For all m, m' in $\mathcal{M}_n$, let

$$v_{A,m,m'}^2 = \sup_{t \in S_m + S_{m'}, \|t\| \leq 1} q\mathrm{Var}(L_q(t)(A_0)), \ D_{A,m} = q \sum_{\lambda \in m} \mathrm{Var}(L_q(\psi_\lambda)(A_0)),$$

$$b_{m,m'} = \sup_{t \in S_m + S_{m'}, \|t\| \leq 1} \|t\|_\infty.$$

For all m in $\mathcal{M}_n$, let

$$R_{A,m} = n\|s - s_m\|^2 + 2D_{A,m},\ e_{A,m,m'} = \frac{q}{p} b^2_{m,m'}.$$

We denote by $e_{A,m} = e_{A,m,m}$, $v_{A,m} = v_{A,m,m}$. For all $k \in \mathbb{N}$, let $\mathcal{M}_n^k = \{m \in \mathcal{M}_n,\ R_{A,m} \in [k, k+1)\}$. For all n in $\mathbb{N}^*$, for all $k > 0$, $k' > 0$, for all $\gamma \geq 0$, let $[k]$ denote the integer part of k and let

$$l_{n,\gamma}(k, k') = \ln\left((1 + \mathrm{Card}(\mathcal{M}_n^{[k]}))(1 + \mathrm{Card}(\mathcal{M}_n^{[k']}))(k+1)(k'+1)\right) + (\ln n)^\gamma \tag{10}$$

The following assumptions generalize Assumptions **[V]** and **[BR]** made in [Ler09b].

[V']: *There exist $\gamma > 1$ and a sequence $(\epsilon_n)_{n\in\mathbb{N}}$, with $\epsilon_n \to 0$ such that, for all n in $\mathbb{N}$,*

$$\sup_{(m,m')\in(\mathcal{M}_n)^2} \left\{ \left(\left(\frac{v^2_{A,m,m'}}{R_{A,m} \vee R_{A,m'}} \right)^2 \vee \frac{e_{A,m,m'}}{R_{A,m} \vee R_{A,m'}} \right) l_{m,m'} \right\} \leq \epsilon_n^4,$$

where, for all m, m' in $\mathcal{M}_n$, $l_{m,m'} = l_{n,\gamma}(R_{A,m}, R_{A,m'})$.

[BR'] *There exist two sequences $(h_n^*)_{n\in\mathbb{N}^*}$ and $(h_n^o)_{n\in\mathbb{N}^*}$ with $(h_n^o \vee h_n^*) \to 0$ as $n \to \infty$ such that, for all n in $\mathbb{N}^*$, for all $m_o \in \arg\min_{m\in\mathcal{M}_n} R_{A,m}$ and all $m^* \in \arg\max_{m\in\mathcal{M}_n} D_{A,m}$, we have*

$$\frac{R_{A,m_o}}{D_{A,m^*}} \leq h_n^o,\ \frac{n\|s - s_{m^*}\|^2}{D_{A,m^*}} \leq h_n^*.$$

3 Main results

3.1 Resampling penalties

The first theorem justifies the use of resampling penalties for β-mixing data.

Theorem 3.1 *Let $X_1, ..., X_n$ be a strictly stationary sequence of random variables with common density s and let $(S_m)_{m\in\mathcal{M}_n}$ be a collection of linear subspaces of $L^2(\mu)$ satisfying Assumption **[V']**. Let $\tilde{s}$ be the estimator defined in (9) with pen(m) defined in (8). Assume that $X_1, ..., X_n$ are β-mixing, then, there exists a constant $C > 0$ such that*

$$\mathbb{P}\left(\|s - \tilde{s}\|^2 > (1 + 110\epsilon_n) \inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|^2 \right) \leq Ce^{-\frac{1}{2}(\ln n)^\gamma} + p\beta_q. \tag{11}$$

The second theorem justifies the use of resampling penalties for τ-mixing data.

Theorem 3.2 *Let $X_1, ..., X_n$ be a strictly stationary sequence of random variables with common density s and let $(S_m)_{m\in\mathcal{M}_n}$ be a collection of linear subspaces of $L^2(\mu)$ satisfying Assumption **[V']**. Let $\tilde{s}$ be the estimator defined in (9) with pen(m) defined in (8). Assume that $X_1, ..., X_n$ are real valued and τ-mixing, then, there exists an absolute constant $C > 0$ such that we have*

$$\mathbb{E}\left(\|s - \tilde{s}\|^2\right) \leq (1 + 160\epsilon_n)\mathbb{E}\left(\inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|^2 \right) + C\left(e^{-\frac{1}{2}(\ln n)^\gamma} + \tau_q MC_n \right), \tag{12}$$

where the mixing complexity MC_n is defined by the following formula:

$$MC_n = \sum_{m\in\mathcal{M}_n} \left(\left\| \sum_{\lambda\in m} |\psi_\lambda| \right\|_\infty \sup_{\lambda\in m} Lip(\psi_\lambda) + \|s\| |\mathcal{M}_n| \sup_{t\in B_m} Lip(t) \right).$$

Comments:

- Theorems 3.1 and 3.2 can be compared with Theorem 2.5 in [Ler09b]. An extra term $p\beta_q$ appears in the control of the deviation probability when the data are β-mixing. In Section 4, it is proved that p and q can be chosen in order to have $p\beta_q \leq Cn^{-\alpha}$ for some $\alpha > 1$ under classical assumptions on the mixing coefficients.

- When the data are τ-mixing, the mixing coefficient τ_q must control the mixing complexity MC_n. It is clear that $MC_n = \infty$ for many collections of linear spaces $(S_m)_{m\in\mathcal{M}_n}$ (as histogram spaces for example). Therefore, the collection $\mathcal{M}_n$ should be chosen carefully when we deal with τ-mixing data. In Section 4, it is proved that, on wavelet spaces, p and q can be chosen in order to have $\tau_q MC_n \leq Cn^{-1}$ under classical assumptions on the mixing coefficient.

- Up to our knowledge, inequalities (11) and (12) are the first oracle inequalities obtained for totally data driven PLSE of the density s when the data are mixing. Moreover, this is the first time that the risk of the selected estimator is compared with the risk of an oracle and not with an upper bound.

3.2 Slope heuristic

We will now justify the use of the slope heuristic when the data are mixing. The following theorems give point 1 in this heuristic, respectively for β and τ-mixing sequences. In both cases, the complexity $\Delta_m = D_{A,m}/n$ can be used with the constant $K_{\min} = 2$.

Theorem 3.3 *Let $X_1, ..., X_n$ be a strictly stationary sequence of random variables, with common density s. Let $\mathcal{M}_n$ be a collection of models satisfying Assumptions* **[V']**, **[BR']** *and let $\epsilon_n^* = \epsilon_n \vee h_n^*$.*
Assume that there exists a constant $0 < \delta < 1$ such that, for all m in $\mathcal{M}_n$,

$$0 \leq pen(m) \leq \frac{(2-\delta)D_{A,m}}{n}.$$

Let $\hat{m}, \tilde{s}$ be the random variables defined in (9). Assume that $X_1, ..., X_n$ are β-mixing and let

$$c_n = \frac{\delta - 75\epsilon_n^*}{2(1+27\epsilon_n)}.$$

There exists a constant $C > 0$, such that, with probability larger than $1 - Ce^{-\frac{1}{2}(\ln n)^\gamma} - p\beta_q$,

$$D_{A,\hat{m}} \geq c_n D_{A,m^*}, \; \|s - \tilde{s}\|^2 \geq \frac{c_n}{h_n^o} \inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|^2. \tag{13}$$

Theorem 3.4 *Let $X_1, ..., X_n$ be a strictly stationary sequence of random variables, with common density s. Let $\mathcal{M}_n$ be a collection of models satisfying Assumptions* **[V']**, **[BR']**.
Assume that there exists a constant $0 < \delta < 1$ such that, for all m in $\mathcal{M}_n$,

$$0 \leq pen(m) \leq \frac{(2-\delta)D_{A,m}}{n}.$$

Let $\hat{m}, \tilde{s}$ be the random variables defined in (9). Assume that $X_1, ..., X_n$ are τ-mixing, let MC_n be the mixing complexity defined in Theorem 3.2 and let

$$c_n' = \frac{\delta - h_n^*}{2(1+35\epsilon_n)}.$$

There exists an absolute constant $C > 0$ such that

$$\mathbb{E}(D_{A,\hat{m}}) \geq c_n' D_{A,m^*} - Cn\left(e^{-\frac{1}{2}(\ln n)^\gamma} + \tau_q MC_n\right). \tag{14}$$

$$\mathbb{E}\left(\|s - \tilde{s}\|^2\right) \geq 2\frac{c_n'}{h_n^o}\mathbb{E}\left(\inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|^2\right) - C\left(e^{-\frac{1}{2}(\ln n)^\gamma} + \tau_q MC_n\right). \tag{15}$$

Comment: When n is sufficiently large, $c_n \geq \delta/4$, $c_n' \geq \delta/4$. Hence, when $\mathrm{pen}(m)$ is not larger than $2D_{A,m}/n$, inequalities (13) and (14) ensure that with high probability or in expectation $D_{A,\hat{m}} \geq cD_{A,m^*}$, which is as large as possible. Inequalities (13) and (15) show that no optimal oracle inequality can hold. This proves point 1 of the slope heuristic.
The following theorems justify the remaining points.

Theorem 3.5 *Let $X_1, ..., X_n$ be a stationary sequence of random variables with common density s. Let $(S_m)_{m\in\mathcal{M}_n}$ be a collection of models satisfying* **[V']**. *For all m in $\mathcal{M}_n$, let $pen(m)$ be a penalty function and let $\tilde{s}$ be the estimator defined in (9).*
Assume that $X_1, ..., X_n$ are β-mixing and that there exist constants $\bar{\delta} \geq \underline{\delta} > -1$ and $0 \leq p' < 1$ such that, with probability at least $1 - p'$, for all m in $\mathcal{M}_n$,

$$\frac{4D_{A,m}}{n} + \underline{\delta}\frac{R_{A,m}}{n} \leq pen(m) \leq \frac{4D_{A,m}}{n} + \bar{\delta}\frac{R_{A,m}}{n}.$$

Let

$$c_n = \begin{pmatrix} \frac{1+\bar{\delta}+37\epsilon_n}{2(1+\underline{\delta}-27\epsilon_n)} & \text{if } 1+\underline{\delta}-27\epsilon_n > 0 \\ +\infty & \text{if } 1+\underline{\delta}-27\epsilon_n \leq 0 \end{pmatrix}.$$

There exists a constant $C > 0$, such that, with probability at least $1 - Ce^{-\frac{1}{2}(\ln n)^\gamma} - p\beta_q - p'$,

$$D_{A,\hat{m}} \leq c_n R_{A,m_o},\ \|s - \tilde{s}\|^2 \leq 2c_n \inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|^2, \tag{16}$$

Theorem 3.6 *Let $X_1, ..., X_n$ be a stationary sequence of random variables with common density s. Let $(S_m)_{m\in\mathcal{M}_n}$ be a collection of models satisfying* **[V']**. *For all m in $\mathcal{M}_n$, let $pen(m)$ be a penalty function and let $\tilde{s}$ be the estimator defined in (9).*
Assume that $X_1, ..., X_n$ are τ-mixing and that there exist constants $\bar{\delta} \geq \underline{\delta} > -1$ and a sequence $(e_n)_{n\in\mathbb{N}}$, with $\sum_{n\in\mathbb{N}} e_n < \infty$ such that

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(\frac{4D_{A,m}}{n} + \underline{\delta}\frac{R_{A,m}}{n} - pen(m)\right)_+\right) \leq e_n,$$

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(pen(m) - \frac{4D_{A,m}}{n} - \bar{\delta}\frac{R_{A,m}}{n}\right)_+\right) \leq e_n.$$

Let MC_n be the mixing complexity defined in Theorem 3.2 and let

$$c_n = \begin{pmatrix} \frac{1+\bar{\delta}+55\epsilon_n}{2(1+\underline{\delta}-85\epsilon_n)} & \text{if } 1+\underline{\delta}-85\epsilon_n > 0 \\ +\infty & \text{if } 1+\underline{\delta}-85\epsilon_n \leq 0 \end{pmatrix}.$$

There exists a constant $C > 0$, such that,

$$\mathbb{E}\left(D_{A,\hat{m}}\right) \leq c_n\left(R_{A,m_o} + n(C\tau_q MC_n + e_n)\right). \tag{17}$$

$$\mathbb{E}\left(\|s - \tilde{s}\|^2\right) \leq c_n\left(\mathbb{E}\left(\inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|^2\right) + C\left(\tau_q MC_n + e_n\right)\right) \tag{18}$$

Comments:

- $D_{A,\hat{m}}$ jumps from D_{A,m^*} (Theorem 3.3 and 3.4) to R_{A,m_o} when $\mathrm{pen}(m)$ is around $2D_{A,m}/n$. R_{A,m_o} is much smaller than D_{A,m^*} under Assumption **[BR']**. This justifies point 2 of the slope heuristic. Point 3 comes from inequalities (16) and (18) applied with $\underline{\delta} = \overline{\delta} = 0$.

- It may be useful to overpenalize a little from a non asymptotic point of view. Imagine that $1 - 67\epsilon_n$ is very close to 0, then c_n is much smaller if $\underline{\delta} > 0$ than if we take its asymptotic optimal value 0.

- The practical implementation of these algorithms is discussed in general in Arlot & Massart [AM09], see also the discussion for density estimation in [Ler09b]. The slope heuristic is very fast to compute and shall be prefered when a shape of the ideal penalty is available. The resampling-based estimators give this shape for more general collections.

4 Comparison with previous results

In this section, we compare the estimator given by the resampling penalty with those given in [Ler09a]. Recall that the estimator was chosen among the collection of least-squares estimators $(\hat{s}_m)_{m\in\mathcal{M}_n}$, where $\hat{s}_m = \arg\min_{t\in S_m} \|t\|^2 - 2P_n t$, by a penalization procedure

$$\tilde{s} = \hat{s}_{\hat{m}}, \text{ where } \hat{m} = \arg\min_{m\in\mathcal{M}_n} \|\hat{s}_m\|^2 - 2P_n\hat{s}_m + \mathrm{pen}(m). \tag{19}$$

Mixing assumptions In [Ler09a], we considered two kinds of rates of convergence to 0 of the mixing coefficients. Let $\gamma = \beta$ or τ.
[AR(θ)**]** arithmetical γ-mixing with rate θ: there exists $C > 0$ such that, for all k in $\mathbb{N}$, $\gamma_k \leq C(1+k)^{-(1+\theta)}$,
[GEO(θ)**]** geometrical γ-mixing with rate θ: there exists $C > 0$ such that, for all k in $\mathbb{N}$, $\gamma_k \leq Ce^{-\theta k}$.

4.1 β-mixing processes

In Comte & Merlevède [CM02], as well as in [Ler09a], the collection of models was assumed to satisfy the following assumptions:

$[M_1]$ *For all* $m \in \mathcal{M}_n$, S_m *is a linear space with finite dimension* $d_m \geq 2$ *and* $N_n = \max_{m\in\mathcal{M}_n} d_m$ *satisfies* $N_n \leq n$.
$[M_2]$ *There exists a constant* Φ *such that*

$$\forall m, m' \in \mathcal{M}_n, \forall t \in S_m, \forall t' \in S_{m'}, \|t + t'\|_\infty \leq \Phi\sqrt{\dim(S_m + S_{m'})}\|t + t'\|_2.$$

$[M_3]$ $d_m \leq d_{m'}$ *implies that* $S_m \subset S_{m'}$.

From $[M_1]$, for all $k > n$, $\mathcal{M}_n^k = \emptyset$ and, from $[M_3]$, for all $k \leq n$, $\mathrm{Card}(\mathcal{M}_n^k) \leq 1$. Hence, there exists a constant c_V such that, for all $\gamma > 1$,

$$l_{m,m'} \leq c_V(\ln n)^{2\gamma}$$

In order to verify **[V']**, we need two other assumptions.

$[M_4]$ *There exists* $c'_D > 0$ *such that, for all* n *in* $\mathbb{N}^*$, *for all* m *in* $\mathcal{M}_n$, $D_{A,m} \geq c'_D d_m$.

[M_5] *There exist* $\gamma > 1$ *and a sequence* $r_n \to \infty$ *such that* $R_n(\ln n)^{-4\gamma} \geq r_n$*, where* $R_n = \inf_{m \in \mathcal{M}_n} R_{A,m}$.

Under these assumption, the following result holds:

Corollary 4.1 *Let* $\mathcal{M}_n$ *be a collection of models satisfying* [M_1]- [M_5]. *Assume that the process* $(X_n)_{n\in\mathbb{Z}}$ *is strictly stationary and arithmetically* [$\mathbf{AR}(\theta)$] β*-mixing with mixing rate* $\theta > 1$. *Let* $\tilde{s}$ *be the estimator defined in (9) with a resampling penalty (8). Let* $\epsilon_n^* = (\ln n)^{-1/4} \wedge r_n^{-1/8}$.
There exist constants $C > 0$ *and* $\kappa > 0$ *such that*

$$\mathbb{P}\left(\|\tilde{s} - s\|_2^2 > (1 + \kappa\epsilon_n^*) \inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|_2^2 \right) \leq C \frac{(\log n)^{2(\theta+2)}}{n^{\theta/2}}.$$

Comments:

- This result can be compared with Corollary 3.1 in [Ler09b]. In the independent case, the rate of convergence of the leading constant is always given by $(r_n)^{-1/4}$ and this rate is often polynomial in n. It is not faster than $(\ln n)^{-1/4}$ in the β-mixing case. The deviation probability was upper bounded by $Ce^{-\frac{1}{2}(\ln n)^\gamma}$ for some constants $C > 0$ and $\gamma > 1$, it is now polynomial in n.
- [M_5] is hard to check in general. Let $c_d^{-1} = 2\sup_{x\geq 1}(\ln x)^8 x^{-1}$. [$M_5$] is satisfied for example, if there is no model in $\mathcal{M}_n$ that have dimension $d_m \leq c_d(\ln n)^8$ and if [M_4] is satisfied. In this case, [M_5] holds with $r_n = (\ln n)^2$ and we deduce from our previous computations the following result.

Corollary 4.2 *Let* $\mathcal{M}_n$ *be a collection of models satisfying* [M_1]- [M_4]. *Assume that the process* $(X_n)_{n\in\mathbb{Z}}$ *is strictly stationary and arithmetically* [$\mathbf{AR}(\theta)$] β*-mixing with mixing rate* $\theta > 1$. *Let* $\tilde{s}$ *be the estimator defined in (9) with a resampling penalty (8). Then, there exist constants* $\kappa > 0$, $C > 0$ *such that, with probability larger than* $1 - Cn^{-\theta/2}(\log n)^{2(\theta+2)}$,

$$\|\tilde{s} - s\|_2^2 \leq \left(1 + \frac{\kappa}{(\ln n)^{1/4}}\right) \inf_{m\in\mathcal{M}_n, d_m \geq c_d(\ln n)^8} \|s - \hat{s}_{A,m}\|_2^2 .$$

Comments: Corollary 4.2 can be compared with Theorem 3.1 in [Ler09a].

- Both procedures lead to trajectorial oracle inequalities of type (3).
- The penalty term in [Ler09a] depends on a constant c_D, which is in general unknown. On the other hand, in Corollary 4.2, the selection algorithm $(X_1, ..., X_n) \mapsto \tilde{s}$ is totally computable.
- The risk of $\tilde{s}$ in Corollary 4.2 is compared with the best of the risks in the collection $\mathcal{M}_n$. It is compared with an upper bound on $\|s - s_m\|^2 + 2\mathbb{E}\left(\|s_m - \hat{s}_{A,m}\|^2\right)$ in [Ler09a].
- In [Ler09a], [M_4] was not necessary. However, our new estimator improves this previous procedure every time that [M_4] (or any other assumption ensuring [**V'**]) holds.

Corollary 4.3 *Let* $\mathcal{M}_n$ *be a collection of models satisfying* [M_1]-[M_5]. *Assume that the process* $(X_n)_{n\in\mathbb{Z}}$ *is strictly stationary and geometrically* [$\mathbf{GEO}(\theta)$] β*-mixing with mixing rate* $\theta > 0$. *Let* $\tilde{s}$ *be the estimator defined in (9) with a resampling penalty (8). Let*

$\epsilon_n^* = \left(r_n^{-1/8} \vee n^{-1/4}(\ln n)^{1+\gamma/2}\right)$ *and* $\theta_1 = \theta \wedge 1$.
There exist constants $C > 0$ *and* $\kappa > 0$ *such that*

$$\mathbb{P}\left(\|\tilde{s} - s\|_2^2 > (1 + \kappa\epsilon_n^*) \inf_{m \in \mathcal{M}_n} \|s - \hat{s}_{A,m}\|_2^2\right) \leq C\frac{n}{(\ln n)^2} e^{-\frac{\theta_1}{2}(\ln n)^2}$$

Comments : Under the stronger assumption that the process is geometrically β-mixing, we almost recover the same results as in the independent case. The rate of convergence is now essentially given by $r_n^{-1/8}$ (it was $r_n^{-1/4}$ in the independent case) and the deviation probability is upper bounded by $Cn(\ln n)^{-2}e^{-\frac{\theta_1}{2}(\ln n)^2}$ (instead of $Ce^{-\frac{1}{2}(\ln n)^2}$ in the independent case).

4.2 τ-mixing processes

Our results for τ-mixing processes do not apply to general collections of models as mentioned before. We give in this section a classical collection where they might be used.
Dyadic Wavelet spaces:
This collection was the one of [Ler09a]. Wavelet spaces are classically considered because the oracle is adaptive over Besov spaces (see for example Birgé & Massart [BM97] or [Ler09a]). Hereafter, r is a real number, $r \geq 1$ and we work with an r-regular orthonormal multiresolution analysis of $L^2(\mu)$, associated with a compactly supported scaling function ϕ and a compactly supported mother wavelet ψ. Without loss of generality, we suppose that the support of the functions ϕ and ψ is included in an interval $[A_1, A_2)$ where A_1 and A_2 are integers such that $A_2 - A_1 = A \geq 1$.
For all functions t in $L^2(\mu)$, we denote by $\|t\|_{BV}$ its bounded variation semi-norm, that is

$$\|t\|_{BV} = \sup_{l \in \mathbb{N}^*} \sup_{-\infty < a_1 < \ldots < a_l < +\infty} \sum_{j=1}^{l-1} |t(a_{j+1}) - t(a_j)|.$$

For all k in $\mathbb{Z}$ and j in $\mathbb{N}^*$, let $\psi_{0,k} : \ x \to \sqrt{2}\phi(2x - k)$ and $\psi_{j,k} : \ x \to 2^{j/2}\psi(2^j x - k)$. The family $\{(\psi_{j,k})_{j\geq 0, k\in\mathbb{Z}}\}$ is an orthonormal basis of $L^2(\mu)$. Let us recall the following inequalities: let $K_\infty = (\sqrt{2}\|\phi\|_\infty) \vee \|\psi\|_\infty$, $K_L = (2\sqrt{2}\mathrm{Lip}(\phi)) \vee \mathrm{Lip}(\psi)$, $K_{BV} = AK_L$. Then for all $j \geq 0$, we have $\|\psi_{j,k}\|_\infty \leq K_\infty 2^{j/2}$,

$$\left\|\sum_{k\in\mathbb{Z}} |\psi_{j,k}|\right\|_\infty \leq AK_\infty 2^{j/2} \quad (20)$$

$$\mathrm{Lip}(\psi_{j,k}) \leq K_L 2^{3j/2}, \quad (21)$$

$$\|\psi_{j,k}\|_{BV} \leq K_{BV} 2^{j/2}, . \quad (22)$$

We assume that $\mathcal{M}_n$ is the following collection.

[W] *dyadic wavelet generated spaces: let* $J_n = [\ln(n)/\ln(2)]$, *for all* $J_m = 1, ..., J_n$, *let*

$$m = \{(j,k), \ 0 \leq j \leq J_m, \ k \in \mathbb{Z}\}$$

and let S_m *be the linear span of* $\{\psi_{j,k}\}_{(j,k)\in m}$.

Hereafter, u denotes the following real number

$$u = \frac{3}{1+\theta} \wedge 1.$$

As in the previous section, we add extra assumptions to prove **[V']**.

[T4] *There exists a constant* $c'_D > 0$ *such that, for all* $n \in \mathbb{N}^*$, *for all* m *in* $\mathcal{M}_n$,

$$D_{A,m} \geq c'_D 2^{J_m}.$$

[T5] *There exist a sequence* $r_n \to \infty$ *and a constant* $\gamma > 1$ *such that,*

$$R_n (\ln n)^{-\frac{2\gamma}{1-u}} \geq r_n.$$

As in the β-mixing case, we deduce the following corollary.

Corollary 4.4 *Assume that the process* $(X_n)_{n\in\mathbb{Z}}$ *is strictly stationary and arithmetically* **[AR**(θ)**]** τ*-mixing with mixing rate* $\theta > 2$. *Let* $\mathcal{M}_n$ *be a collection of regular wavelet spaces* **[W]** *and assume moreover that* **[T4]**, **[T5]** *hold. Let* $\tilde{s}$ *be the estimator defined in (9) with a resampling penalty (8). Let* $\epsilon_n^* = \left(\ln n \wedge r_n^{1-u}\right)^{-1/4}$.
There exist constants $C > 0$ *and* $\kappa > 0$ *such that*

$$\mathbb{E}\left(\|\tilde{s} - s\|_2^2\right) \leq (1 + \kappa\epsilon_n^*)\mathbb{E}\left(\inf_{m\in\mathcal{M}_n} \|s - \hat{s}_{A,m}\|_2^2\right) + C\frac{(\ln n)^{2(1+\theta)}}{n^{(\theta-3)/2}}.$$

Comments:

- With a mixing rate $\theta > 5$, the estimator selected by a resampling penalty satisfies an oracle inequality (4). This result can be compared with Corollary 4.1. When the data are τ-mixing, we do not obtain a trajectorial oracle inequality (3) and the condition on the mixing rate is stronger than in the β-mixing case. However, as mentioned in the introduction, this result is very interesting because there is a lot of examples of processes that are τ-mixing and not β-mixing.
- Assumption **[T5]** is hard to check in practice but it can be removed as in the β-mixing, provided that we only consider models with dimension larger than $c_M(\ln n)^\eta$ for some well chosen constants c_M and η.
- We can get better rates of convergence if we assume that the process is geometrically τ-mixing and if we choose p and q as in Corollary 4.3.

This result can also be compared with Theorem 4.1 in [Ler09a].

- As in the β-mixing case, the main improvement of Corollary 4.4 is that the new procedure is totally data driven.
- The risk of $\tilde{s}$ is compared with the oracle in Corollary 4.4 whereas it is compared with an upper bound on $\inf_{m\in\mathcal{M}_n}\left\{\|s - s_m\|^2 + 2\mathbb{E}\left(\|s_m - \hat{s}_{A,m}\|^2\right)\right\}$ in [Ler09a].
- **[T4]** was not necessary in [Ler09a] but our new procedure improves the one given in [Ler09a] every time that **[T4]** or any other Assumption ensuring **[V']** holds.

5 Proofs

5.1 Notations

Let us give some notations that we will use repeatedly all along the proofs.
Recall that p and q are integers such that $2pq = n$. For all $k = 0, ..., p-1$, $I_k = (2kq +$

$1,...,(2k+1)q)$, $A_k=(X_i)_{i\in I_k}$ and $I=\cup_{k=0}^{p-1}I_k$. For all functions t in $L^2(\mu)$ and all $x_1,...,x_q$ in $\mathbb{R}$,

$$L_q(t)(x_1,...,x_q)=\frac{1}{q}\sum_{i=1}^{q}t(x_i),\ P_At=\frac{1}{p}\sum_{k=0}^{p-1}L_q(t)(A_k)=\frac{2}{n}\sum_{i\in I}t(X_i),$$
$$\nu_A(t)=(P_A-P)(t).$$

The estimator $\hat{s}_{A,m}$ associated to the model S_m, is defined as

$$\hat{s}_{A,m}\in\arg\min_{t\in S_m}P_AQ(t).$$

For all m, m' in $\mathcal{M}_n$, let

$$T_m=\sum_{\lambda\in m}(L_q(\psi_\lambda)-P\psi_\lambda)^2,$$
$$U_m=\frac{1}{p(p-1)}\sum_{i\neq j=0}^{p-1}\sum_{\lambda\in m}(L_q(\psi_\lambda)(A_i)-P\psi_\lambda)(L_q(\psi_\lambda)(A_j)-P\psi_\lambda),$$
$$p(m)=\|s_m-\hat{s}_{A,m}\|^2=\sup_{t\in B_m}(\nu_A(t))^2=\sum_{\lambda\in m}(\nu_A(\psi_\lambda))^2.$$
$$p_W(m)=\frac{1}{v_W^2}\sum_{\lambda\in m}\mathbb{E}^W\left((\nu_A^W(\psi_\lambda))^2\right),\ \delta(m,m')=2\nu_A(s_m-s_{m'}).$$

Lemma 6.2 applied with $n=p$, $\Lambda=m$, $t_\lambda=L_q(\psi_\lambda)$, $X_i=A_{i-1}$, gives

$$p_W(m)=\tfrac{1}{p}(P_A(T_m)-U_m) \tag{23}$$
$$p(m)-p_W(m)=U_m, \tag{24}$$

where $P_A(T_m)=\sum_{k=0}^{p-1}T_m(A_k)/p$.
For all functional $T=F(A_0,...,A_{p-1})$, let $T^*=F(A_0^*,...,A_{p-1}^*)$, where the random variables (A_k^*) are given by the coupling Lemmas given in Section 2.3. In particular, we will use repeatedly the notations P_A^*, ν_A^*, U_m^*, $p^*(m)$, $p_W^*(m)$, $\delta^*(m,m')$.
For all functions t of $L^2(\mu)$, for all r in $\mathbb{N}$ and all $x_1,...,x_r,y_1,...y_r$ in $\mathbb{R}$,

$$\begin{aligned}|L_r(t)(x_1,...,x_r)-L_r(t)(y_1,...,y_r)| &\leq \frac{1}{r}\sum_{i=1}^{r}|t(x_i)-t(y_i)|\\ &\leq \frac{1}{r}\mathrm{Lip}(t)d_r((x_1,...,x_r),(y_1,...,y_r)).\end{aligned}$$

Thus, for all r in $\mathbb{N}^*$, $\mathrm{Lip}_r(L_r(t))\leq\mathrm{Lip}(t)/r$.
For all $k\in\mathbb{N}$, $\mathcal{M}_n^k=\{m\in\mathcal{M}_n,\ R_{A,m}\in[k,k+1)\}$ and for all n in $\mathbb{N}$ and, for all $k>0$, $k'>0$ and $\gamma\geq0$, let

$$l_{n,\gamma}(k,k')=\ln\Big((1+\mathrm{Card}(\mathcal{M}_n^{[k]}))(1+\mathrm{Card}(\mathcal{M}_n^{[k']}))(k+1)(k'+1)\Big)+(\ln n)^\gamma.$$

For all m, m' in $\mathcal{M}_n$, let $l_{m,m'}=l_{n,\gamma}(R_{A,m},R_{A,m'})$. From Lemma 6.1 applied with $\alpha=\alpha'=0$, for all $K>1$, there exists a constant $C>0$ such that

$$\sum_{(m,m')\in(\mathcal{M}_n)^2}e^{-Kl_{m,m'}}=Ce^{-K(\ln n)^\gamma}.$$

Under **[V']**,

$$\sup_{(m,m')\in(\mathcal{M}_n)^2}\left\{\left(\left(\frac{v_{A,m,m'}^2}{R_{A,m}\vee R_{A,m'}}\right)^2\vee\frac{e_{A,m,m'}}{R_{A,m}\vee R_{A,m'}}\right)l_{m,m'}^2\right\}\leq\epsilon_n^4.$$

5.2 Technical Lemmas

Lemmas 5.1 and 5.2 are coupling lemmas. They allow to work with $p^*(m)$, $p^*_W(m)$, $\delta^*(m,m')$ instead of $p(m)$, $p_W(m)$, $\delta(m,m')$. Lemma 5.3 is a consequence of our study of the independent case. It allows to extend the proofs of [Ler09b] to the mixing case. It is the main tool of this paper.

Lemma 5.1 *Let $X_1,...,X_n$ be stationary random variables, real valued and β-mixing. Let p and q be two integers such that $2pq = n$ and let $A^*_0,...,A^*_{p-1}$ be the random variables given by Viennet's Lemma in Section 2.3.1. Let $(S_m)_{m\in\mathcal{M}_n}$ be a collection of linear spaces of functions. Let $(p(m))_{m\in\mathcal{M}_n}$, $(p_W(m))_{m\in\mathcal{M}_n}$, $(\delta(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$, $(p^*(m))_{m\in\mathcal{M}_n}$, $(p^*_W(m))_{m\in\mathcal{M}_n}$, $(\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$, be the associated collections defined in Section 5.1. There exists an event Ω_C such that $\mathbb{P}(\Omega^c_C)\le p\beta_q$ and such that, on Ω_C, for all m, m' in $\mathcal{M}_n$, we have*

$$p(m) = p^*(m),\ p_W(m) = p^*_W(m),\ \delta(m,m') = \delta^*(m,m'). \tag{25}$$

Proof :
Let $\Omega_C = \{\forall k = 0,...,p-1,\ A_k = A^*_k\}$. It comes from Viennet's Lemma that $\mathbb{P}(\Omega^c_C)\le p\beta_q$ and it is clear that, on Ω_C, (25) holds.

Lemma 5.2 *Let $X_1,...,X_n$ be stationary random variables, real valued, τ-mixing and with common density s. Let p and q be two integers such that $2pq = n$ and let $A^*_0,...,A^*_{p-1}$ be the random variables given by the τ-coupling's Lemma in Section 2.3.2. Let $\mathcal{M}_n$ be a collection of models. Let $(p(m))_{m\in\mathcal{M}_n}$, $(p_W(m))_{m\in\mathcal{M}_n}$, $(\delta(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$, $(p^*(m))_{m\in\mathcal{M}_n}$, $(p^*_W(m))_{m\in\mathcal{M}_n}$, $(\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$, be the associated collections defined in Section 5.1. Let MC_n be the mixing complexity of $\mathcal{M}_n$ defined by*

$$MC_n = \sum_{m\in\mathcal{M}_n}\left(\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m} Lip(\psi_\lambda) + \|s\||\mathcal{M}_n|\sup_{t\in B_m} Lip(t)\right).$$

For all m, m' in $\mathcal{M}_n$,

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}|p(m)-p^*(m)|\right)\le 4\tau_q MC_n \tag{26}$$

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}|p_W(m)-p^*_W(m)|\right)\le \frac{8\tau_q}{p} MC_n \tag{27}$$

$$\mathbb{E}\left(\sup_{m,m'\in\mathcal{M}_n}\delta(m,m')-\delta^*(m,m')\right)\le 4\tau_q MC_n. \tag{28}$$

Proof :
For all m in $\mathcal{M}_n$, we have

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}|p(m)-p^*(m)|\right)\le \sum_{m\in\mathcal{M}_n}\mathbb{E}\left(|p(m)-p^*(m)|\right).$$

Moreover, for all m in $\mathcal{M}_n$,

$$\begin{aligned}
|p(m)-p^*(m)| &= \left|\sum_{\lambda\in m}((P_A-P)\psi_\lambda)^2-((P_A^*-P)\psi_\lambda)^2\right| \\
&= \left|\sum_{\lambda\in m}((\nu_A+\nu_A^*)\psi_\lambda)\,((P_A-P_A^*)\psi_\lambda)\right| \\
&\leq \sum_{\lambda\in m}|(\nu_A+\nu_A^*)\psi_\lambda|\,\frac{1}{p}\sum_{k=0}^{p-1}|L_q(\psi_\lambda)(A_k)-L_q(\psi_\lambda)(A_k^*)| \\
&\leq 4\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}_q(L_q(\psi_\lambda))\frac{1}{p}\sum_{k=0}^{p-1}d_q(A_k,A_k^*) \\
&\leq \frac{4}{q}\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}(\psi_\lambda)\frac{1}{p}\sum_{k=0}^{p-1}d_q(A_k,A_k^*).
\end{aligned}$$

We take the expectation in this last inequality and we use the τ-coupling Lemma of Section 2.3.2 to obtain (26).
From (23), we have

$$|p_W(m)-p_W^*(m)| = \frac{1}{p}\left|(P_A-P_A^*)(T_m)-(U_m-U_m^*)\right|.$$

We have

$$\begin{aligned}
&(P_A-P_A^*)T_m = \\
&\sum_{\lambda\in m}\frac{1}{p}\sum_{k=0}^{p-1}\left(L_q(\psi_\lambda)(A_k)-L_q(\psi_\lambda)(A_k^*)\right)\left(L_q(\psi_\lambda)(A_k)+L_q(\psi_\lambda)(A_k^*)-2P\psi_\lambda\right),
\end{aligned}$$

thus

$$\begin{aligned}
|(P_A-P_A^*)T_m| &= 4\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}_q(L_q(\psi_\lambda))\frac{1}{p}\sum_{k=0}^{p-1}d_q(A_k,A_k^*) \\
&\leq \frac{4}{q}\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}(\psi_\lambda)\frac{1}{p}\sum_{k=0}^{p-1}d_q(A_k,A_k^*).
\end{aligned}$$

Moreover

$$\begin{aligned}
U_m-U_m^* &= \frac{1}{p(p-1)}\sum_{i\neq j=0}^{p-1}\sum_{\lambda\in m}(L_q(\psi_\lambda(A_j))-P\psi_\lambda)(L_q(\psi_\lambda(A_i))-L_q(\psi_\lambda(A_i^*))) \\
&\quad+\frac{1}{p(p-1)}\sum_{i\neq j=0}^{p-1}\sum_{\lambda\in m}(L_q(\psi_\lambda(A_i^*))-P\psi_\lambda)(L_q(\psi_\lambda(A_j))-L_q(\psi_\lambda(A_j^*))),
\end{aligned}$$

thus

$$|U_m-U_m^*| \leq \frac{4}{q}\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}(\psi_\lambda)\frac{1}{p}\sum_{k=0}^{p-1}d_q(A_k,A_k^*).$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(|p_W(m) - p_W^*(m)|\right) &\le \frac{8}{pq}\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}(\psi_\lambda)\frac{1}{p}\sum_{k=0}^{p-1}\mathbb{E}(d_q(A_k, A_k^*))\\ &\le \frac{8\tau_q}{p}\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}(\psi_\lambda).\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}|p_W(m) - p_W^*(m)|\right) &\le \sum_{m\in\mathcal{M}_n}\mathbb{E}\left(|p_W(m) - p_W^*(m)|\right)\\ &\le \frac{8\tau_q}{p}\sum_{m\in\mathcal{M}_n}\left\|\sum_{\lambda\in m}|\psi_\lambda|\right\|_\infty \sup_{\lambda\in m}\mathrm{Lip}(\psi_\lambda).\end{aligned}$$

Finally,

$$\mathbb{E}\left(\sup_{m,m'\in\mathcal{M}_n}\delta(m,m') - \delta^*(m,m')\right) \le \sum_{m,m'\in\mathcal{M}_n}\mathbb{E}\left(|\delta(m,m') - \delta^*(m,m')|\right)$$

and, for all m, m' in $\mathcal{M}_n$,

$$\begin{aligned}\mathbb{E}\left(|\delta(m,m') - \delta^*(m,m')|\right) &= 2\mathbb{E}\left(|(P_A - P_A^*)(s_m - s_{m'})|\right)\\ &\le \frac{2}{pq}\mathrm{Lip}(s_m - s_{m'})\sum_{k=0}^{p-1}\mathbb{E}\left(d_q(A_k, A_k^*)\right)\\ &\le 2\tau_q\mathrm{Lip}(s_m - s_{m'}).\end{aligned}$$

For all x, y in $\mathbb{R}$ and all m, m' in $\mathcal{M}_n$,

$$(s_m - s_{m'})(x) - (s_m - s_{m'})(y) \le \|s\|\left(\sup_{t\in B_m}\mathrm{Lip}(t) + \sup_{t\in B_{m'}}\mathrm{Lip}(t)\right)d(x,y)$$

Hence, $\mathrm{Lip}(s_m - s_{m'}) \le \|s\|\left(\sup_{t\in B_m}\mathrm{Lip}(t) + \sup_{t\in B_{m'}}\mathrm{Lip}(t)\right)$, thus

$$\mathbb{E}\left(\sup_{m,m'\in\mathcal{M}_n}\delta(m,m') - \delta^*(m,m')\right) \le 4\tau_q\|s\||\mathcal{M}_n|\sum_{m\in\mathcal{M}_n}\sup_{t\in B_m}\mathrm{Lip}(t).$$

Let us now derive some consequences of the results of [Ler09b].

Lemma 5.3 *Let A_0^*, ...,A_{p-1}^* be i.i.d random variables valued in $\mathbb{R}^q$, with $2pq = n$. Let $\mathcal{M}_n$ be a collection of models satisfying* **[V']** *and let $(p^*(m))_{m\in\mathcal{M}_n}$, $(p_W^*(m))_{m\in\mathcal{M}_n}$, $(\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$, $(D_{A,m})_{m\in\mathcal{M}_n}$, $(R_{A,m})_{m\in\mathcal{M}_n}$ be the associated collections defined in Section 5.1. There exists a constant $C > 0$ such that*

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*(m) - \frac{2D_{A,m}}{n} > 15\epsilon_n\frac{R_{A,m}}{n}\right\}\right) \le Ce^{-\frac{1}{2}(\ln n)^\gamma}, \qquad (29)$$

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*(m) - \frac{2D_{A,m}}{n} < -25\epsilon_n\frac{R_{A,m}}{n}\right\}\right) \le Ce^{-\frac{1}{2}(\ln n)^\gamma}. \qquad (30)$$

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*(m)-p_W^*(m)>15\epsilon_n\frac{R_{A,m}}{n}\right\}\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}, \tag{31}$$

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*(m)-p_W^*(m)<-25\epsilon_n\frac{R_{A,m}}{n}\right\}\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}. \tag{32}$$

$$\mathbb{P}\left(\bigcup_{m,m'\in\mathcal{M}_n}\left\{\delta^*(m,m')>12\epsilon_n\left(\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\right\}\right)\leq Ce^{-(\ln n)^\gamma}. \tag{33}$$

There exists an absolute constant $C>0$ such that

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(p^*(m)-\frac{2D_{A,m}}{n}-15\epsilon_n\frac{R_{A,m}}{n}\right)_+\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}. \tag{34}$$

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(-p^*(m)+\frac{2D_{A,m}}{n}-35\epsilon_n\frac{R_{A,m}}{n}\right)_+>\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}. \tag{35}$$

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(p^*(m)-p_W^*(m)-20\epsilon_n\frac{R_{A,m}}{n}\right)_+\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}. \tag{36}$$

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(-p^*(m)+p_W^*(m)-35\epsilon_n\frac{R_{A,m}}{n}\right)_+\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}. \tag{37}$$

$$\mathbb{E}\left(\sup_{m,m'\in\mathcal{M}_n}\left(\delta^*(m,m')-20\epsilon_n\left(\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\right)_+\right)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma}. \tag{38}$$

Proof of the concentration inequalities :
$p^*(m)=\sup_{t\in B_m}((\nu_A^*)(t))^2$ and $A_0^*,...,A_{p-1}^*$ are independent. Thus

$$\mathbb{E}(p^*(m))=\sum_{\lambda\in m}\frac{\mathrm{Var}\left(L_q(\psi_\lambda)(A_0)\right)}{p}=\frac{2D_{A,m}}{n},$$

$$\sup_{t\in B_m}\mathrm{Var}\left(L_q(t)(A_0)\right)=\frac{v_{A,m}^2}{q},\ \frac{\sup_{t\in B_m}\|L_q(t)\|_\infty^2}{p}\leq\frac{e_{A,m}}{q}.$$

We apply Proposition 6.3 in the Appendix with $B=\{L_q(t),t\in B_m\}$, $D=D_{A,m}/q$, $v^2=v_{A,m}^2/q$, $\epsilon=e_{A,m}/q$ and $n=p$. For all $x>0$ and all m in $\mathcal{M}_n$, with probability larger than $1-e^{-x}$

$$p^*(m)-\frac{2D_{A,m}}{n}\leq\frac{2D_{A,m}^{3/4}(e_{A,m}(19x)^2)^{1/4}+6\sqrt{D_{A,m}v_{A,m}^2x}+6v_{A,m}^2x+2e_{A,m}(19x)^2}{n}$$

and, with probability larger than $1-2.8e^{-x}$

$$\frac{2D_{A,m}}{n}-p^*(m)\leq\frac{16D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4}+15.22\sqrt{D_{A,m}v_{A,m}^2x}+2e_{A,m}(40.25x)^2}{n}. \tag{39}$$

Let $K>0$ be a constant to be chosen later, let $l_m=l_{n,\gamma}(R_m,R_m)$, and let $x=K^2l_m$. From **[V']** applied with $m=m'$, since $D_{A,m}\leq R_{A,m}$,

$$v_{A,m}^2x\leq(K\epsilon_n)^2R_{A,m},\ e_{A,m}x^2\leq(K\epsilon_n)^4R_{A,m},$$

$$D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} \leq K\epsilon_n R_{A,m}, \ \sqrt{D_{A,m}v_{A,m}^2 x} \leq K\epsilon_n R_m. \qquad (40)$$

Let $e_n(K) = (2\sqrt{19}+6)K + 6K^2\epsilon_n + 2(19)^2K^4\epsilon_n^3$, from (40),

$$\frac{2D_{A,m}^{3/4}(e_{A,m}(19x)^2)^{1/4} + 6\sqrt{D_{A,m}v_{A,m}^2 x} + 6v_{A,m}^2 x + 2e_{A,m}(19x)^2}{n} \leq e_n(K)\epsilon_n \frac{R_m}{n}.$$

Thus, from Lemma 6.1, for all $K > 1/\sqrt{2}$, there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{\frac{2D_{A,m}}{n} - p^*(m) > e_n(K)\epsilon_n \frac{R_{A,m}}{n}\right\}\right) \leq$$
$$\sum_{m\in\mathcal{M}_n}\mathbb{P}\left(p^*(m) - \frac{2D_{A,m}}{n} > e_n(K)\epsilon_n\frac{R_{A,m}}{n}\right) \leq \sum_{m\in\mathcal{M}_n} e^{-K^2 l_m} \leq Ce^{-K^2(\ln n)^\gamma}.$$

Let $K = 11/(2\sqrt{19}+6) > 1/\sqrt{2}$ and choose n sufficiently large such that $6K^2\epsilon_n + 2(19)^2K^4\epsilon_n^3 \leq 4$, then $e_n(K) \leq 15$ and (29) holds for all n sufficiently large. It holds for all n provided that we enlarge C if necessary.
Let $e_n^{(2)}(K) = 31,22K + 2(40,25)^2K^4\epsilon_n^3$, from (40),

$$\frac{16D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} + 15.22\sqrt{D_{A,m}v_{A,m}^2 x} + 2e_{A,m}(40.25x)^2}{n} \leq e_n^{(2)}(K)\epsilon_n\frac{R_{A,m}}{n}.$$

We apply inequality (39) with $x = K^2 l_m$. For all $K > 1/\sqrt{2}$, there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*(m) - \frac{2D_{A,m}}{n} < -e_n^{(2)}(K)\epsilon_n\frac{R_{A,m}}{n}\right\}\right) \leq$$
$$\sum_{m\in\mathcal{M}_n}\mathbb{P}\left(p^*(m) - \frac{2D_{A,m}}{n} < -e_n^{(2)}(K)\epsilon_n\frac{R_{A,m}}{n}\right) \leq 2.8\sum_{m\in\mathcal{M}_n} e^{-K^2 l_m} \leq Ce^{-K^2(\ln n)^\gamma}.$$

Take $K = 23/31.22 > 1/\sqrt{2}$ and n sufficiently large to have $2(40,25)^2K^4\epsilon_n^3 \leq 2$, then $e_n^{(2)}(K) \leq 25$ and (30) holds for sufficiently large n. It holds then in general, provided that we enlarge the constant C if necessary.
From (24), $p^*(m) - p_W^*(m) = U_m^*$. Therefore, from Lemma 6.4 in the appendix, for all m in $\mathcal{M}_n$ and all $x > 0$, with probability larger than $1 - 2e^{-x}$,

$$p^*(m) - p_W^*(m) \leq \frac{10.62D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} + 6\sqrt{v_{A,m}^2 D_{A,m}x} + 6v_{A,m}^2 x + 2e_{A,m}(19.1x)^2}{n-1}, \qquad (41)$$

and, with probability larger than $1 - 3.8e^{-x}$,

$$p_W^*(m) - p^*(m) > \frac{18D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} + 15.22\sqrt{v_{A,m}^2 D_{A,m}x} + 2e_{A,m}(40.3x)^2}{n-1}. \qquad (42)$$

Let $K > 0$, $e_n^{(3)}(K) = (16.62K + 6K^2\epsilon_n + 2(19.1)^2K^4\epsilon_n^3)n/(n-1)$ and $x = K^2 l_m$, from (40),

$$\frac{10.62D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} + 6\sqrt{v_{A,m}^2 D_{A,m}x} + 6v_{A,m}^2 x + 2e_{A,m}(19.1x)^2}{n-1} \leq e_n^{(3)}(K)\epsilon_n\frac{R_{A,m}}{n}.$$

We apply (41) with $x = K^2 l_m$. From Lemma 6.1, for all $K > 1/\sqrt{2}$, there exists a constant C such that

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*(m) - p^*_W(m) > e_n^{(3)}(K)\epsilon_n\frac{R_{A,m}}{n}\right\}\right) \leq$$

$$\sum_{m\in\mathcal{M}_n}\mathbb{P}\left(p^*(m) - p^*_W(m) > e_n^{(3)}(K)\epsilon_n\frac{R_{A,m}}{n}\right) \leq 2\sum_{m\in\mathcal{M}_n} e^{-K^2 l_m} \leq Ce^{-K^2(\ln n)^\gamma}.$$

Take $K = 12/16.62 > 1/\sqrt{2}$ and $n \geq 15$ such that $6K^2\epsilon_n + 2(19.1)^2 K^4\epsilon_n^3 \leq 2$, then $e_n^{(3)}(K) \leq 15$ and (31) holds for sufficiently large n. It holds in general provided that we enlarge C if necessary.
Let $K > 0$, $e_n^{(4)}(K) = (33.22K + 2(40.3)^2 K^4\epsilon_n^3)n/(n-1)$ and $x = K^2 l_m$. From (40),

$$\frac{18D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} + 15.22\sqrt{v^2_{A,m}D_{A,m}x} + 2e_{A,m}(40.3x)^2}{n-1} \leq e_n^{(4)}(K)\epsilon_n\frac{R_{A,m}}{n}.$$

We apply (41) with $x = K^2 l_m$. From Lemma 6.1, for all $K > 1/\sqrt{2}$, there exists a constant C such that

$$\mathbb{P}\left(\bigcup_{m\in\mathcal{M}_n}\left\{p^*_W(m) - p^*(m) > e_n^{(4)}(K)\epsilon_n\frac{R_{A,m}}{n}\right\}\right) \leq$$

$$\sum_{m\in\mathcal{M}_n}\mathbb{P}\left(p^*_W(m) - p^*(m) > e_n^{(4)}(K)\epsilon_n\frac{R_{A,m}}{n}\right) \leq 3.8\sum_{m\in\mathcal{M}_n} e^{-K^2 l_m} \leq Ce^{-K^2(\ln n)^\gamma}.$$

Take $K = 23.5/33.22 > 1/\sqrt{2}$ and $n \geq 25$ such that $2(40.3)^2K^4\epsilon_n^3 \leq 0.5$, then $e_n^{(4)}(K) \leq 25$ and (32) holds for sufficiently large n. It holds in general provided that we enlarge C if necessary.
Finally, we apply Lemma 6.5 in the appendix to the functions $s_m - s_{m'}$, with $L = L_q$ and $\nu_n = \nu_A$, we have $v^2 \leq v^2_{A,m,m'}/q$ and $\epsilon \leq e_{A,m,m'}/q$. For all m, m' in $\mathcal{M}_n$,

$$\|s_m - s_{m'}\|^2 \leq 2(\|s_m - s\|^2 + \|s_{m'} - s\|^2) \leq 4\frac{R_{A,m}\vee R_{A,m'}}{n},$$

thus, for all $\eta > 0$, for all $x > 0$,

$$\mathbb{P}\left(\delta^*(m,m') > 4\eta\left(\frac{R_{A,m}\vee R_{A,m'}}{n}\right) + \frac{8v^2_{A,m,m'}x + 4e_{A,m,m'}x^2/9}{\eta n}\right) \leq e^{-x}. \tag{43}$$

Let $K > 0$, $l_{m,m'} = l_{n,\gamma}(R_{A,m}, R_{A,m'})$, $x = K^2 l_{m,m'}$ and $e_n^{(5)}(K) = \sqrt{2K^2 + K^4\epsilon_n^2/9}$. From (40),

$$8v^2_{A,m,m'}x + 4e_{A,m,m'}x^2/9 \leq 4(e_n^{(5)}(K))\epsilon_n)^2 R_{A,m}\vee R_{A,m'},$$

thus, for $\eta = e_n^{(5)}(K))\epsilon_n$,

$$4\eta\left(\frac{R_{A,m}\vee R_{A,m'}}{n}\right) + \frac{8v^2_{A,m,m'}x + 4e_{A,m,m'}x^2/9}{\eta n} \leq 8e_n^{(5)}(K))\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}.$$

Hence, for all $K>1$, there exists a constant $C>0$ such that

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{m,m'\in\mathcal{M}_n}\left\{\delta^*(m,m')>8e_n^{(5)}(K))\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right\}\right)\\
&\leq\sum_{m,m'\in\mathcal{M}_n}\mathbb{P}\left(\delta^*(m,m')>8e_n^{(5)}(K))\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\\
&\leq\sum_{m,m'\in\mathcal{M}_n}e^{-K^2l_{m,m'}}\leq Ce^{-K^2(\ln n)^\gamma}.
\end{aligned}$$

Take $K=11.4/(8\sqrt{2})>1$ and n sufficiently large to have $8\sqrt{K^4\epsilon_n^2/9}\leq 0.6$, then $8e_n^{(5)}(K))\leq 12$ and (33) holds for sufficiently large n. It holds in general provided that we increase C if necessary.

Proof of the results in expectation

Let $K>0$, $z>0$, $l_m=l_{n,\gamma}(R_m,R_m)$, $x=K^2l_m(1+z)$ and

$$e_n^{(6)}(K,z)=(2\sqrt{19}+6)K\sqrt{x}+6K^2\epsilon_nx+4(19)^2K^4\epsilon_n^3x^2.$$

From (40),

$$\begin{aligned}
&\frac{2D_{A,m}^{3/4}(e_{A,m}(19x)^2)^{1/4}+6\sqrt{D_{A,m}v_{A,m}^2x}+6v_{A,m}^2x+2e_{A,m}(19x)^2}{n}\\
&\leq(e_n^{(6)}(K,1)+e_n^{(6)}(K,z))\epsilon_n\frac{R_{A,m}}{n}.
\end{aligned}$$

Thus, from Proposition 6.3 in the Appendix, for all $z>0$ and all m in $\mathcal{M}_n$,

$$\mathbb{P}\left(p^*(m)-\frac{2D_{A,m}}{n}-e_n^{(6)}(K,1)\epsilon_n\frac{R_{A,m}}{n}>e_n^{(6)}(K,z))\epsilon_n\frac{R_{A,m}}{n}\right)\leq e^{-K^2l_m(1+z)}.$$

Let us now briefly explain how to deduce from this concentration inequalities the results in expectation.

[MI]: Integration of the concentration inequality

Let $\epsilon_m=\epsilon_nR_{A,m}/n$ and $f(m)=p^*(m)-2D_{A,m}/n-e_n^{(6)}(K,1)\epsilon_m$, we have

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}(f(m))_+\right)\quad\leq\quad\sum_{m\in\mathcal{M}_n}\mathbb{E}\left((f(m))_+\right)=\sum_{m\in\mathcal{M}_n}\int_0^\infty\mathbb{P}\left(f(m)>y\right)dy.$$

Since $z\mapsto g(z)=e_n^{(6)}(K,z))$ is clearly a C^1-diffeomorphism of $\mathbf{R}_+^*$, this last integral is equal to

$$\int_0^\infty\mathbb{P}\left(f(m)>\epsilon_mg(y)\right)\epsilon_mg'(y)dy$$

For all $K>0$, there exists a constant $C>0$ such that $g'(z)\leq C(z^{-1/2}+1+z)$. From Lemma 6.1, for all $K>1$, $n\geq 2$, there exists a constant $C>0$ such that

$$\begin{aligned}
&\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(p^*(m)-\frac{2D_{A,m}}{n}-e_n^{(6)}(K,1)\epsilon_n\frac{R_{A,m}}{n}\right)_+\right)\leq\\
&C\sum_{m\in\mathcal{M}_n}\epsilon_nR_{A,m}e^{-K^2l_m}\left(\int_0^\infty(z^{-1/2}+1+z)e^{-K^2l_mz}dz\right)\leq Ce^{-K^2(\ln n)^\gamma}.
\end{aligned}$$

The last inequality comes from the fact that ϵ_n is bounded and $K^2 l_m \geq c > 0$ for all $n \geq 2$, $K > 1$. Take $K = 14.75/(2\sqrt{19}+6) > 1$ and choose n sufficiently large such that $6K^2\epsilon_n + 4(19)^2K^4\epsilon_n^3 \leq 0.25$, then $e_n^{(6)}(K) \leq 15$ and (34) holds for all n sufficiently large. It holds for all n provided that we enlarge C if necessary.
We obtain (35) with the same arguments.
Let us now turn to the result on the resampling estimator of $p(m)$. Let $K > 0$, $z > 0$, $l_m = l_{n,\gamma}(R_m, R_m)$, $x = K^2 l_m(1+z)$,

$$e_n^{(7)}(K,z) = \frac{n}{n-1}\left(16,62K\sqrt{x} + 6K^2\epsilon_n x + 4(19.1)^2K^4\epsilon_n^2 x^2\right),$$

From inequalities (40), we have

$$\frac{10.62D_{A,m}^{3/4}(e_{A,m}x^2)^{1/4} + 6\sqrt{v_{A,m}^2 D_{A,m}x} + 6v_{A,m}^2 x + 2e_{A,m}(19.1x)^2}{n-1}$$
$$\leq (e_n^{(7)}(K,1) + e_n^{(7)}(K,z))\epsilon_n \frac{R_{A,m}}{n}$$

From inequalities (41) with $x = K^2 l_m(1+z)$ and for all $z > 0$, for all m in $\mathcal{M}_n$ and all $z > 0$

$$\mathbb{P}\left(p^*(m) - p_W^*(m) - e_n^{(7)}(K,1)\epsilon_n\frac{R_{A,m}}{n} > e_n^{(7)}(K,z)\epsilon_n\frac{R_{A,m}}{n}\right) \leq 2e^{-K^2 l_m(1+z)}.$$

We use again the method of integration **[MI]** to prove that, for all $K > 1$, there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(p^*(m) - p_W^*(m) - e_n^{(7)}(K,1)\epsilon_n\frac{R_{A,m}}{n}\right)_+\right) \leq Ce^{-K^2(\ln n)^\gamma}.$$

Take $K = 17/16.62 > 1$ and $n \geq 20$ such that $6K^2\epsilon_n + 4(19.1)^2K^4\epsilon_n^3 \leq 2$, then $e_n^{(7)}(K,1) \leq 20$ and (36) holds for sufficiently large n. It holds in general provided that we enlarge C if necessary.
We obtain (37) with the same arguments.
Let $K > 0$, $l_{m,m'} = l_{n,\gamma}(R_m, R_{m'})$, $z > 0$, $x = K^2 l_{m,m'}(1+z)$,

$$e_n^{(8)}(K,z) = \sqrt{2K^2 z + 2K^4\epsilon_n^2 x^2/9},$$

$e_n^{(8)}(K) = e_n^{(8)}(K,1)$, $g_K(z) = (e_n^{(8)}(K,z))^2/e_n^{(8)}(K)$ and $\eta = e_n^{(8)}(K,1)\epsilon_n$.

$$4\eta\frac{R_{A,m}\vee R_{A,m'}}{n} + \frac{8v_{A,m,m'}^2 x + 4e_{A,m,m'}x^2/9}{\eta n}$$
$$\leq 4\left(2e_n^{(8)}(K) + g_K(z)\right)\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}.$$

Thus from (43), for all $z > 0$, for all m, m' in $\mathcal{M}_n$ and all $K > 0$,

$$\mathbb{P}\left(\delta(m,m') - 8e_n^{(8)}(K)\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n} > \epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}g_K(z)\right) \leq e^{-K^2 l_{m,m'}(1+z)}.$$

Thus

$$\mathbb{E}\left(\sup_{(m,m')\in\mathcal{M}_n^2}\left(\delta^*(m,m') - 8e_n^{(8)}(K)\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right)_+\right)$$
$$\leq \sum_{(m,m')\in\mathcal{M}_n^2}\mathbb{E}\left(\left(\delta^*(m,m') - 8e_n^{(8)}(K)\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right)_+\right)$$
$$= \sum_{(m,m')\in\mathcal{M}_n^2}\int_0^\infty \mathbb{P}\left(\delta^*(m,m') - 8e_n^{(8)}(K)\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n} > x\right)dx$$

Let $x = \epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}g_K(z)$. For all $K>0$, for all $n\geq 2$, there exists a constant $C>0$ such that $g_K(z)'\leq C(1+z)$. Thus, from Lemma 6.1, for all $K>\sqrt{2}$, there exists a constant $C>0$ such that

$$\mathbb{E}\left(\sup_{(m,m')\in\mathcal{M}_n^2}\left(\delta^*(m,m') - 8e_n^{(8)}(K)\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right)_+\right)$$
$$\leq C\sum_{(m,m')\in\mathcal{M}_n^2}\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}e^{-K^2 l_{m,m'}}\int_0^\infty(1+z)e^{-K^2 l_{m,m'}z}dz \leq Ce^{-K^2(\ln n)^\gamma}.$$

Take $K = 17/(8\sqrt{2}) > \sqrt{2}$ and n sufficiently large to have $8\sqrt{2K^4\epsilon_n^2/9}\leq 3$, then $8e_n^{(8)}(K,1))\leq 20$ and (38) holds for sufficiently large n. It holds in general provided that we increase C if necessary. We can now turn to the proofs of the main results of this part.

5.3 Proof of Theorem 3.1

Let us first assume that $X_1,...,X_n$ are β-mixing. Let A_0^*, ..., A_{p-1}^* be the random variables given by Viennet's Lemma in Section 2.3.1. Let $(p(m)$, $p_W(m)$, $\delta(m,m')$, $p^*(m)$, $p_W^*(m)$, $D_{A,m}$, $R_{A,m}$, $\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$ be the quantities defined in Section 5.1. Let us define the events

$$\Omega_p = \bigcap_{m\in\mathcal{M}_n}\left\{-25\epsilon_n\frac{R_{A,m}}{n}\leq p^*(m) - \frac{2D_{A,m}}{n}\leq 15\epsilon_n\frac{R_{A,m}}{n}\right\}, \tag{44}$$

$$\tilde{\Omega}_p = \bigcap_{m\in\mathcal{M}_n}\left\{-25\epsilon_n\frac{R_{A,m}}{n}\leq p^*(m) - p_W^*(m)\leq 15\epsilon_n\frac{R_{A,m}}{n}\right\}$$

$$\Omega_d = \bigcap_{(m,m')\in\mathcal{M}_n}\left\{\delta(m,m')\leq 12\epsilon_n\left(\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\right\}, \tag{45}$$

$$\Omega_C = \left(\bigcap_{m\in\mathcal{M}_n}\{p(m)=p^*(m)\}\right)\cap\left(\bigcap_{m\in\mathcal{M}_n}\{p_W(m)=p_W^*(m)\}\right)$$
$$\cap\left(\bigcap_{(m,m')\in\mathcal{M}_n}\left\{\delta(m,m')=\delta^*(m,m')\right\}\right). \tag{46}$$

From Lemmas 5.1 and 5.3, there exists a constant $C>0$ such that

$$\mathbb{P}(\Omega_p^c)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma},\ \mathbb{P}(\tilde{\Omega}_p^c)\leq Ce^{-\frac{1}{2}(\ln n)^\gamma},\ \mathbb{P}(\Omega_d^c)\leq Ce^{-(\ln n)^\gamma},\ \mathbb{P}(\Omega_C^c)\leq p\beta_q.$$

Let $\Omega = \Omega_p \cap \tilde{\Omega}_p \cap \Omega_d \cap \Omega_C$. Recall that $\text{pen}(m) = 2p_W(m)$. On Ω, from inequality (6), for all m in $\mathcal{M}_n$,

$$\begin{aligned}\|s-\tilde{s}\|^2 &\leq \|s-\hat{s}_{A,m}\|^2 + 2\left(p_W(m)-p(m)\right) - 2\left(p_W(\hat{m})-p(\hat{m})\right) + \delta(m,\hat{m})\\ &= \|s-\hat{s}_{A,m}\|^2 + 2\left(p_W^*(m)-p^*(m)\right) - 2\left(p_W^*(\hat{m})-p^*(\hat{m})\right) + \delta^*(m,\hat{m})\\ &\leq \|s-\hat{s}_{A,m}\|^2 + 62\epsilon_n\frac{R_{A,m}}{n} + 42\epsilon_n\frac{R_{A,\hat{m}}}{n}.\end{aligned}$$

On Ω,

$$\frac{R_{A,m}}{n} = \|s-\hat{s}_{A,m}\|^2 + \frac{2D_{A,m}}{n} - p^*(m) \leq \|s-\hat{s}_{A,m}\|^2 + 25\epsilon_n\frac{R_{A,m}}{n}.$$

If $25\epsilon_n < 1$, on Ω,

$$\|s-\tilde{s}\|^2 \leq \frac{1+37\epsilon_n}{1-25\epsilon_n}\|s-\hat{s}_{A,m}\|^2 + \frac{42\epsilon_n}{1-25\epsilon_n}\|s-\tilde{s}\|^2.$$

Hence, if $67\epsilon_n < 1$,

$$\mathbb{P}\left(\|s-\tilde{s}\|^2 > \frac{1+37\epsilon_n}{1-67\epsilon_n}\inf_{m\in\mathcal{M}_n}\|s-\hat{s}_{A,m}\|^2\right) \leq Ce^{-\frac{1}{2}(\ln n)^\gamma} + p\beta_q.$$

Take n sufficiently large to have $67\epsilon_n < 1$ and $104/(1-67\epsilon_n) \leq 110$. Then,

$$\frac{1+37\epsilon_n}{1-67\epsilon_n} = 1 + \frac{104}{1-67\epsilon_n}\epsilon_n \leq 1 + 110\epsilon_n$$

and (11) holds for sufficiently large n. It holds in general provided that we increase the constant C if necessary.

5.4 Proof of Theorem 3.2

Let us now assume that $X_1, ..., X_n$ are τ-mixing. Let A_0^*, ..., A_{p-1}^* be the random variables given by the τ-couling Lemma in Section 2.3.2. Let $(p(m)$, $p_W(m)$, $\delta(m,m')$, $p^*(m)$, $p_W^*(m)$, $D_{A,m}$, $R_{A,m}$, $\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$ be the quantities defined in Section 5.1. From inequality (6), for all m in $\mathcal{M}_n$,

$$\begin{aligned}\|s-\tilde{s}\|^2 &\leq \|s-\hat{s}_{A,m}\|^2 + 2\left(p_W(m)-p(m)\right) - 2\left(p_W(\hat{m})-p(\hat{m})\right) + \delta(m,\hat{m})\\ &= \|s-\hat{s}_{A,m}\|^2 + 2\left(p_W^*(m)-p^*(m)-35\epsilon_n\frac{R_{A,m}}{n}\right) + 90\epsilon_n\frac{R_{A,m}}{n}\\ &\quad +2\left(p^*(\hat{m})-p_W^*(\hat{m})-20\epsilon_n\frac{R_{A,\hat{m}}}{n}\right) + 60\epsilon_n\frac{R_{A,\hat{m}}}{n}\\ &\quad +\delta^*(m,\hat{m}) - 20\epsilon_n\frac{R_{A,m}\vee R_{A,\hat{m}}}{n} + 2(p_W(m)-p_W^*(m))\\ &\quad +2(p^*(m)-p(m)+p_W^*(\hat{m})-p_W(\hat{m})+p(\hat{m})-p^*(\hat{m}))\\ &\quad +\delta(m,\hat{m})-\delta^*(m,\hat{m}).\end{aligned}$$

For all m in $\mathcal{M}_n$,

$$\begin{aligned}\frac{R_{A,m}}{n} &= \frac{\|s-\hat{s}_{A,m}\|^2}{1-35\epsilon_n} + \frac{(1-35\epsilon_n)R_{A,m}/n - \|s-\hat{s}_{A,m}\|^2}{1-35\epsilon_n}\\ &= \frac{\|s-\hat{s}_{A,m}\|^2}{1-35\epsilon_n} + \frac{2D_{A,m}/n - 35\epsilon_n R_{A,m}/n - \|s_m-\hat{s}_{A,m}\|^2}{1-35\epsilon_n}.\end{aligned} \quad (47)$$

In the control of $\|s-\tilde{s}\|^2$, we replace $R_{A,m}/n$ and $R_{A,\hat{m}}/n$ by the expressions obtained in (47) in the terms $90\epsilon_n R_{A,m}/n$ and $60\epsilon_n R_{A,\hat{m}}/n$. Assume that $35\epsilon_n<1$,

$$\frac{1-95\epsilon_n}{1-35\epsilon_n}\|s-\tilde{s}\|^2 \le \frac{1+55\epsilon_n}{1-35\epsilon_n}\inf_{m\in\mathcal{M}_n}\|s-\hat{s}_{A,m}\|^2$$
$$+\frac{150\epsilon_n}{1-35\epsilon_n}\sup_{m\in\mathcal{M}_n}\left(\frac{2D_{A,m}}{n}-p^*(m)-35\epsilon_n\frac{R_m}{n}\right)+$$
$$\frac{4+10\epsilon_n}{1-35\epsilon_n}\sup_{m\in\mathcal{M}_n}|p^*(m)-p(m))|+\sup_{m,m'\in\mathcal{M}_n}\left(\delta^*(m,m')-20\epsilon_n\left(\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\right)$$
$$+2\sup_{m\in\mathcal{M}_n}\left(p_W^*(m)-p^*(m)-35\epsilon_n\frac{R_m}{n}\right)+4\sup_{m\in\mathcal{M}_n}|p_W(m)-p_W^*(m)|$$
$$+2\sup_{m\in\mathcal{M}_n}(p^*(m)-p_W^*(m)-15\epsilon_n fracR_m n)+\sup_{m,m'\in(\mathcal{M}_n)^2}\delta(m,m')-\delta^*(m,m').$$

We take the expectation in this last inequality and we use inequalities (26), (27), (28), (35), (36), (37) and (38) to obtain that, when $95\epsilon_n<1$, there exists a constant $C>0$ such that

$$\|s-\tilde{s}\|^2\le\frac{1+55\epsilon_n}{1-95\epsilon_n}\inf_{m\in\mathcal{M}_n}\|s-\hat{s}_{A,m}\|^2+C\left(\tau_q MC_n+e^{-\frac{1}{2}(\ln n)^\gamma}\right)$$

Take n sufficiently large to have $95\epsilon_n<1$ and $150/(1-95\epsilon_n)\le 160$. Then,

$$\frac{1+55\epsilon_n}{1-95\epsilon_n}=1+\frac{150}{1-95\epsilon_n}\epsilon_n\le 1+160\epsilon_n$$

and (12) holds for sufficiently large n. It holds in general provided that we increase the constant C if necessary.

5.5 Proof of Theorem 3.3

Assume that $X_1,...,X_n$ are β-mixing and let $A_0^*,...,A_{p-1}^*$ be the random variables built with Viennet's Lemma in Section 2.3.1. Let $(p(m),\ p_W(m),\ \delta(m,m'),\ p^*(m),\ p_W^*(m),\ D_{A,m},\ R_{A,m},\ \delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$ be the quantities defined in Section 5.1. Let $\Omega_T=\Omega_p\cap\Omega_d\cap\Omega_C$ where Ω_p, Ω_d and Ω_C are defined respectively in (44), (45) and (46). Recall that there exists a constant $C>0$ such that

$$\mathbb{P}(\Omega_p^c)\le Ce^{-\frac{1}{2}(\ln n)^\gamma},\ \mathbb{P}(\Omega_d^c)\le Ce^{-(\ln n)^\gamma},\ \mathbb{P}(\Omega_C^c)\le p\beta_q.$$

If $c_n\le 0$, there is nothing to prove, hence, we can assume that $c_n>0$ and thus that $75\epsilon_n^*<\delta<1$.
$\hat{m}$ minimizes by definition the following criterion

$$\begin{aligned}\mathrm{Crit}(m) &= \|\hat{s}_{A,m}\|^2-2P_A(\hat{s}_{A,m})+\mathrm{pen}(m)+\|s\|^2+2\nu_A(s_{m_o})\\ &= \|\hat{s}_{A,m}\|^2-2P(\hat{s}_{A,m})+\|s\|^2-2\nu_A(\hat{s}_{A,m})+2\nu_A(s_{m_o})+\mathrm{pen}(m)\\ &= \|\hat{s}_{A,m}-s\|^2-2\nu_A(\hat{s}_{A,m}-s_m)+2\nu_A(s_{m_o}-s_m)+\mathrm{pen}(m)\\ &= \|\hat{s}_{A,m}-s\|^2-2\|\hat{s}_{A,m}-s_m\|^2+\delta(m_o,m)+\mathrm{pen}(m)\\ &= \|s-s_m\|^2-p(m)+\delta(m_o,m)+\mathrm{pen}(m)\end{aligned}$$

since $p(m)=\|s_m-\hat{s}_{A,m}\|^2=\nu_A(\hat{s}_{A,m}-s_m)$. Thus, on Ω_T, $\hat{m}$ minimizes the following criterion

$$\mathrm{Crit}(m)=\|s-s_m\|^2-p^*(m)+\delta^*(m,m_o)+\mathrm{pen}(m)$$

For all m in $\mathcal{M}_n$, we have $0 \leq \text{pen}(m) < (2-\delta)D_{A,m}/n$ and $R_{A,m_o} \leq R_{A,m}$. Thus, for all m in $\mathcal{M}_n$, on Ω_T

$$\begin{aligned}
\text{Crit}(m) &\geq \|s-s_m\|^2 - \frac{2D_{A,m}}{n} + \left(\frac{2D_{A,m}}{n} - p^*(m)\right) + \delta^*(m,m_o)\\
&\geq (1-27\epsilon_n)\|s-s_m\|^2 - (1+27\epsilon_n)\frac{2D_{A,m}}{n} \geq -(1+27\epsilon_n)\frac{2D_{A,m}}{n}\\
\text{Crit}(m) &\leq \|s-s_m\|^2 - (\delta - 74\epsilon_n)\frac{D_{A,m}}{n}.
\end{aligned}$$

If $D_{A,m} < c_n D_{A,m*}$, then

$$\begin{aligned}
\text{Crit}(m) &\geq -(1+27\epsilon_n)\frac{2D_{A,m}}{n} > -(1+27\epsilon_n)c_n\frac{2D_{A,m^*}}{n}\\
&\geq -(\delta - 74\epsilon_n - h_n^*)\frac{D_{A,m^*}}{n} \geq \text{Crit}(m^*).
\end{aligned}$$

Since $\text{Crit}(\hat{m}) \leq \text{Crit}(m^*)$, $D_{A,\hat{m}} \geq c_n D_{A,m*}$.
It follows that, on Ω_T,

$$\begin{aligned}
\|s-\tilde{s}\|^2 &= \frac{R_{A,\hat{m}}}{n} + \left(p(\hat{m}) - \frac{2D_{A,\hat{m}}}{n}\right) \geq (1-25\epsilon_n)\frac{R_{A,\hat{m}}}{n}\\
&\geq (1-25\epsilon_n)\frac{2D_{A,\hat{m}}}{n} \geq (1-25\epsilon_n)c_n\frac{2D_{A,m^*}}{n}.
\end{aligned}$$

Moreover, on Ω_T,

$$\inf_{m\in\mathcal{M}_n}\|s-\hat{s}_{A,m}\|^2 \leq \inf_{m\in\mathcal{M}_n}\frac{R_{A,m}}{n}(1+15\epsilon_n) \leq \frac{R_{A,m_o}}{n}(1+15\epsilon_n).$$

Thus

$$\|s-\tilde{s}\|^2 \geq (1-25\epsilon_n)c_n\frac{2D_{A,m^*}}{n} \geq 2c_n\left(\frac{1-25\epsilon_n}{1+15\epsilon_n}\right)\frac{D_{A,m^*}}{R_{A,m_o}}\inf_{m\in\mathcal{M}_n}\|s-\hat{s}_{A,m}\|^2.$$

Since $\epsilon_n < 1/75$, we have $2(1-25\epsilon_n)(1+15\epsilon_n) \geq 2(1-1/3)(1+1/5) \geq 1$. This conclude the proof of (13).

5.6 Proof of Theorem 3.4

Assume that $X_1, ..., X_n$ are τ-mixing. Let A_0^*, ..., A_{p-1}^* be the random variables given by the τ-couling Lemma in Section 2.3.2. Let $(p(m)$, $p_W(m)$, $\delta(m,m')$, $p^*(m)$, $p_W^*(m)$, $D_{A,m}$, $R_{A,m}$, $\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$ be the quantities defined in Section 5.1. In order to prove inequality (14) observe that, for all m in $\mathcal{M}_n$, since $\text{pen}(m) \geq 0$, and $\|s-s_m\|^2 - 35\epsilon_n R_{A,m}/n \geq -35\epsilon_n D_{A,m}/n$

$$\begin{aligned}
\text{Crit}(m) \geq\ & \|s-s_m\|^2 + \left(-p^*(m) + 15\epsilon_n\frac{R_{A,m}}{n}\right) + (p^*(m) - p(m)) - 35\epsilon_n\frac{R_{A,m}}{n}\\
&+ \left(\delta^*(m,m_o) + 20\epsilon_n\frac{R_{A,m}}{n}\right) + (\delta(m,m_o) - \delta^*(m,m_o))\\
=\ & -(1+35\epsilon_n)\frac{2D_{A,m}}{n} + \left(\frac{2D_{A,m}}{n} - p^*(m) + 15\epsilon_n\frac{R_{A,m}}{n}\right) + (p^*(m) - p(m))\\
&+ \left(\delta^*(m,m_o) + 20\epsilon_n\frac{R_{A,m}}{n}\right) + (\delta(m,m_o) - \delta^*(m,m_o)).
\end{aligned}$$

Therefore,

$$\begin{aligned}
-\frac{2D_{A,\hat{m}}}{n}(1+35\epsilon_n) &\leq \mathrm{Crit}(\hat{m})+\sup_{m\in\mathcal{M}_n}\left(p^*(m)-\frac{2D_{A,m}}{n}-15\epsilon_n\frac{R_{A,m}}{n}\right)\\
&+\sup_{(m,m')\in\mathcal{M}_n^2}\left(\delta^*(m,m')-20\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\\
&+\sup_{m\in\mathcal{M}_n}(p(m)-p^*(m))+\sup_{m,m'\in\mathcal{M}_n}(\delta(m,m')-\delta^*(m,m'))
\end{aligned} \quad (48)$$

Since, for all m in $\mathcal{M}_n$, $\mathrm{pen}(m)\leq(2-\delta)D_{A,m}/n$,

$$\mathrm{Crit}(m)\leq\|s-s_m\|^2+\left(\frac{2D_{A,m}}{n}-p^*(m)\right)-\delta\frac{D_{A,m}}{n}+(p^*(m)-p(m))+\delta(m,m_o). \quad (49)$$

Since $\mathrm{Crit}(\hat{m})\leq\mathrm{Crit}(m^*)$, from (49) and (26),

$$\begin{aligned}
\mathbb{E}\left(\mathrm{Crit}(\hat{m})\right) &\leq \mathbb{E}\left(\mathrm{Crit}(m^*)\right)\leq\|s-s_{m^*}\|^2-\delta\frac{D_{A,m^*}}{n}+4\tau_qMC_n\\
&\leq -(\delta-h_n^*)\frac{D_{A,m^*}}{n}+4\tau_qMC_n\leq-c_n'(1+35\epsilon_n)\frac{2D_{A,m^*}}{n}+4\tau_qMC_n.
\end{aligned}$$

Take the expectation in (48) and use inequalities (26), (28), (34) and (38) to obtain (14). We deduce from (14) that there exists a constant $C>0$ such that

$$\begin{aligned}
\mathbb{E}\left(\|s-\tilde{s}\|^2\right) &\geq \frac{2}{n}\mathbb{E}\left(D_{A,\hat{m}}\right)\geq 2c_n'\frac{D_{A,m^*}}{n}-C(e^{-\frac{1}{2}(\ln n)^\gamma}+\tau_qMC_n)\\
&\geq 2\frac{c_n'}{h_n^o}\frac{R_{m_o}}{n}-C(e^{-\frac{1}{2}(\ln n)^\gamma}+\tau_qMC_n).
\end{aligned}$$

The proof of (15) is conclude since

$$\mathbb{E}\left(\inf_{m\in\mathcal{M}_n}\|s-\hat{s}_{A,m}\|^2\right)\leq\inf_{m\in\mathcal{M}_n}\mathbb{E}\left(\|s-\hat{s}_{A,m}\|^2\right)=\frac{R_{m_o}}{n}+C\tau_qMC_n.$$

5.7 Proof of Theorem 3.5

If $c_n=\infty$, there is nothing to prove. Thus we can assume that $c_n<\infty$ and thus that $1+\underline{\delta}-27\epsilon_n>0$. Let us first assume that $X_1,...,X_n$ are β-mixing and let $A_0^*,...A_{p-1}^*$ be the random variables given by Viennet's Lemma in Section 2.3.1. Let $(p(m)$, $p_W(m)$, $\delta(m,m')$, $p^*(m)$, $p_W^*(m)$, $D_{A,m}$, $R_{A,m}$, $\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$ be the quantities defined in Section 5.1. Recall that $\hat{m}$ minimizes over $\mathcal{M}_n$ the following criterion.

$$\mathrm{Crit}(m) = \|s-s_m\|^2-p(m)+\delta(m,m_o)+\mathrm{pen}(m).$$

We keep the notations Ω_p, Ω_d and Ω_C defined by (44), (45), (46). We introduce the event

$$\Omega_{\mathrm{pen}}=\bigcap_{m\in\mathcal{M}_n}\left\{\frac{4D_{A,m}}{n}+\underline{\delta}\frac{R_{A,m}}{n}\leq\mathrm{pen}(m)\leq\frac{4D_{A,m}}{n}+\bar{\delta}\frac{R_{A,m}}{n}\right\}$$

and let $\Omega=\Omega_p\cap\Omega_d\cap\Omega_C\cap\Omega_{\mathrm{pen}}$. Since $R_{A,m_o}\leq R_{A,m}$, on Ω,

$$\begin{aligned}
\mathrm{Crit}(m) &\geq (1+\underline{\delta}-12\epsilon_n)\frac{R_{A,m}}{n}+\left(\frac{2D_{A,m}}{n}-p^*(m)\right)\\
&\geq (1+\underline{\delta}-27\epsilon_n)\frac{R_{A,m}}{n}\geq(1+\underline{\delta}-27\epsilon_n)\frac{2D_{A,m}}{n}.\\
\mathrm{Crit}(m) &\leq (1+\bar{\delta}+37\epsilon_n)\frac{R_{A,m}}{n}.
\end{aligned}$$

If $D_{A,m} > c_n R_{A,m_o}$,

$$\begin{aligned} \mathrm{Crit}(m) &\geq (1+\underline{\delta}-27\epsilon_n)\frac{2D_{A,m}}{n} > 2(1+\underline{\delta}-27\epsilon_n)c_n\frac{R_{A,m_o}}{n} \\ &\geq (1+\bar{\delta}+37\epsilon_n)\frac{R_{A,m_o}}{n} \geq \mathrm{Crit}(m_o) \end{aligned}$$

Since $\mathrm{Crit}(\hat{m}) \leq \mathrm{Crit}(m_o)$, this implies that $D_{A,\hat{m}} \leq c_n R_{A,m_o}$. Moreover, from (6), for all m in $\mathcal{M}_n$

$$\begin{aligned} \|s-\tilde{s}\|^2 &\leq \|s-\hat{s}_{A,m}\|^2 + (\mathrm{pen}(m)-2p^*(m)) + (2p^*(\hat{m})-\mathrm{pen}(\hat{m})) + \delta^*(m,\hat{m}) \\ &\leq \|s-\hat{s}_{A,m}\|^2 + 2\left(\frac{2D_{A,m}}{n}-p^*(m)\right) + (\bar{\delta}+12\epsilon_n)\frac{R_{A,m}}{n} \\ &\quad +2\left(p^*(\hat{m})-\frac{2D_{A,\hat{m}}}{n}\right) + (-\underline{\delta}+12\epsilon_n)\frac{R_{A,\hat{m}}}{n} \\ &\leq \|s-\hat{s}_{A,m}\|^2 + (37\epsilon_n+\bar{\delta})\frac{R_{A,m}}{n} + (27\epsilon_n-\underline{\delta})\frac{R_{A,\hat{m}}}{n}. \end{aligned}$$

For all m in $\mathcal{M}_n$, on Ω,

$$\|s-\hat{s}_{A,m}\|^2 = \frac{R_{A,m}}{n} + \left(p^*(m)-\frac{2D_{A,m}}{n}\right) \geq (1-25\epsilon_n)\frac{R_{A,m}}{n}.$$

Assume that $25\epsilon_n < 1$, then, for all $m \in \mathcal{M}_n$,

$$\|s-\tilde{s}\|^2 \leq \|s-\hat{s}_{A,m}\|^2\left(1+\frac{37\epsilon_n+\bar{\delta}}{1-25\epsilon_n}\right) + \frac{27\epsilon_n-\underline{\delta}}{1-25\epsilon_n}\|s-\tilde{s}\|^2.$$

This proves (16) for sufficiently large n. (16) holds in general provided that we increase the constant C if necessary.

5.8 Proof of Theorem 3.6

Assume that $X_1,...,X_n$ τ-mixing and let $A_0^*,...,A_{p-1}^*$, be the random variables given by the τ-mixing Lemma in Section 2.3.2. Let $(p(m)$, $p_W(m)$, $\delta(m,m')$, $p^*(m)$, $p_W^*(m)$, $D_{A,m}$, $R_{A,m}$, $\delta^*(m,m'))_{(m,m')\in(\mathcal{M}_n)^2}$ be the quantities defined in Section 5.1. Recall that

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(\frac{4D_{A,m}}{n}+\underline{\delta}\frac{R_{A,m}}{n}-\mathrm{pen}(m)\right)_+\right) \leq e_n,$$

$$\mathbb{E}\left(\sup_{m\in\mathcal{M}_n}\left(\mathrm{pen}(m)-\frac{4D_{A,m}}{n}-\bar{\delta}\frac{R_{A,m}}{n}\right)_+\right) \leq e_n.$$

For all m in $\mathcal{M}_n$, we have,

$$\begin{aligned} \frac{R_{A,m}}{n} &= \mathrm{Crit}(m) + \left(p^*(m)-\frac{2D_{A,m}}{n}-15\epsilon_n\frac{R_{A,m}}{n}\right) + \left(\frac{4D_{A,m}}{n}-\mathrm{pen}(m)+\underline{\delta}\frac{R_{A,m}}{n}\right) \\ &\quad -\left(\delta^*(m,m_o)+20\epsilon_n\frac{R_{A,m}}{n}\right) + (p(m)-p^*(m)) \\ &\quad +(35\epsilon_n-\underline{\delta})\frac{R_{A,m}}{n} + (\delta^*(m,m_o)-\delta(m,m_o)). \end{aligned}$$

Therefore

$$\begin{aligned}(1+\underline{\delta}-35\epsilon_n)\frac{R_{A,\hat{m}}}{n} &\leq \mathrm{Crit}(m_o)+\sup_{m\in\mathcal{M}_n}\left(p^*(m)-\frac{2D_{A,m}}{n}-15\epsilon_n\frac{R_{A,m}}{n}\right)\\ &+\sup_{m\in\mathcal{M}_n}\left(\frac{4D_{A,m}}{n}-\mathrm{pen}(m)+\underline{\delta}\frac{R_{A,m}}{n}\right)\\ &+\sup_{(m,m')\in\mathcal{M}_n^2}\left(\delta(m,m')-20\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right)\\ &+\sup_{(m,m')\in\mathcal{M}_n^2}\left(\delta^*(m,m')-\delta(m,m')\right)\\ &+\sup_{m\in\mathcal{M}_n}|p(m)-p^*(m)|. \qquad (50)\end{aligned}$$

On the other hand, for all m in $\mathcal{M}_n$, $\mathrm{Crit}(m)=\|s-s_m\|^2-p(m)+\delta(m_o,m)+\mathrm{pen}(m)$, thus

$$\begin{aligned}\mathrm{Crit}(m_o) &\leq (1+\bar{\delta})R_{A,m_o}+\left(\frac{2D_{A,m_o}}{n}-p^*(m_o)\right)+(p^*(m_o)-p(m_o))\\ &+\mathrm{pen}(m_o)-\frac{4D_{A,m_o}}{n}-\bar{\delta}\frac{R_{A,m_o}}{n}.\end{aligned}$$

Since $\mathbb{E}\left(\mathrm{pen}(m_o)-4D_{A,m_o}/n-\bar{\delta}R_{A,m_o}/n\right)\leq e_n$ and $2D_{A,m_o}/n=\mathbb{E}(p^*(m_o))$, from inequality (26), there exists a constant $C>0$ such that

$$\mathbb{E}\left(\mathrm{Crit}(m_o)\right)\leq(1+\bar{\delta})\frac{R_{A,m_o}}{n}+C\tau_q MC_n+e_n.$$

For all m in $\mathcal{M}_n$, $2D_{A,m}\leq R_{A,m}$. Take the expectation in (50), from inequalities (26), (28), (34) and (38), there exists an absolut constant $C>0$ such that

$$\mathbb{E}\left(D_{\hat{m}}\right)\leq c_n\left(R_{m_o}+Cn\left[\tau_q MC_n+e^{-\frac{1}{2}(\ln n)^\gamma}+e_n\right]\right).$$

This proves inequality (17).

From (6), for all m in $\mathcal{M}_n$,

$$\begin{aligned}\|s-\tilde{s}\|^2 &\leq \|s-\hat{s}_{A,m}\|^2+2\left(\frac{2D_{A,m}}{n}-p^*(m)-35\epsilon_n\frac{R_{A,m}}{n}\right)\\ &+\left(\delta^*(m,\hat{m})-20\epsilon_n\frac{R_{A,m}\vee R_{A,\hat{m}}}{n}\right)\\ &+2\left(-\frac{2D_{A,\hat{m}}}{n}+p^*(\hat{m})-15\epsilon_n\frac{R_{\hat{m}}}{n})\right)\\ &+\left(-\mathrm{pen}(\hat{m})+2\frac{2D_{\hat{m}}}{n}+\underline{\delta}\frac{R_{A,\hat{m}}}{n}\right)+\left(\mathrm{pen}(m)-2\frac{2D_{A,m}}{n}-\bar{\delta}\frac{R_{A,m}}{n}\right)\\ &+(90\epsilon_n+\bar{\delta})\frac{R_{A,m}}{n}+(50\epsilon_n-\underline{\delta})\frac{R_{A,\hat{m}}}{n}\\ &+4\sup_{m\in\mathcal{M}_n}|p(m)-p^*(m)|+\sup_{(m,m')\in\mathcal{M}_n^2}(\delta(m,m')-\delta^*(m,m')). \qquad (51)\end{aligned}$$

Assume that $35\epsilon_n < 1$, for all m in $\mathcal{M}_n$, we have

$$
\begin{aligned}
\frac{R_{A,m}}{n} &= \frac{(1-35\epsilon_n)R_{A,m}}{n} - \frac{\|s-\hat{s}_{A,m}\|^2}{1-35\epsilon_n} + \frac{\|s-\hat{s}_{A,m}\|^2}{1-35\epsilon_n} \\
&\leq \frac{1}{1-35\epsilon_n}\left(\|s-\hat{s}_{A,m}\|^2 + \frac{2D_{A,m}}{n} - p(m) - 35\epsilon_n \frac{R_{A,m}}{n}\right) \\
&\leq \frac{1}{1-35\epsilon_n}\left(\|s-\hat{s}_{A,m}\|^2 + \frac{2D_{A,m}}{n} - p^*(m) - 35\epsilon_n \frac{R_{A,m}}{n} + p(m) - p^*(m)\right)
\end{aligned}
$$

We use this expression in the terms $(90\epsilon_n + \bar{\delta})R_{A,m}/n$ and $(50\epsilon_n - \underline{\delta})R_{A,\hat{m}}/n$ of inequality (51). We deduce that, for all m in $\mathcal{M}_n$,

$$
\begin{aligned}
&\frac{1+\underline{\delta}-85\epsilon_n}{1-35\epsilon_n}\|s-\tilde{s}\|^2 \leq \frac{1+\bar{\delta}+55\epsilon_n}{1-35\epsilon_n} \inf_{m\in\mathcal{M}_n} \|s-\hat{s}_{A,m}\|^2 \\
&+\frac{2+70\epsilon_n+\bar{\delta}-\underline{\delta}}{1-35\epsilon_n} \sup_{m\in\mathcal{M}_n}\left(\frac{2D_{A,m}}{n} - p^*(m) - 35\epsilon_n\frac{R_{A,m}}{n}\right) \\
&+\sup_{m\in\mathcal{M}_n}\left(\mathrm{pen}(m) - \frac{4D_{A,m}}{n} - \bar{\delta}\frac{R_{A,m}}{n}\right) + \sup_{m\in\mathcal{M}_n}\left(\left(\frac{4D_{A,m}}{n} + \underline{\delta}\right)\frac{R_{A,\hat{m}}}{n} - \mathrm{pen}(m)\right) \\
&+2\sup_{m\in\mathcal{M}_n}\left(p^*(m) - \frac{2D_{A,m}}{n} - 15\epsilon_n\frac{R_{A,m}}{n})\right) \\
&+\sup_{(m,m')\in\mathcal{M}_n^2}\left(\delta^*(m,m') - 20\epsilon_n\frac{R_{A,m}\vee R_{A,m'}}{n}\right) \\
&+\frac{4+\bar{\delta}-\underline{\delta}}{1-35\epsilon_n}\sup_{m\in\mathcal{M}_n}|p(m)-p^*(m)| + \sup_{(m,m')\in\mathcal{M}_n^2}(\delta(m,m')-\delta^*(m,m')).
\end{aligned}
$$

We take the expectation in this last inequality and we deduce that, for sufficiently large n, (18) comes from (26), (28), (34), (35) and (38). It holds in general provided that we enlarge the constant C if necessary.

5.9 Proof of Corollaries 4.1 and 4.3.

In [Ler09a], we obtained the following inequalities. If there exists $\theta > 1$ such that $X_1, ..., X_n$ are arithmetically $[\mathbf{AR}(\theta)]$, β-mixing, there exists constants c_v, c_e, c_D, c_M such that, for all m, m' in $\mathcal{M}_n$,

$$
v^2_{A,m,m'} \leq c_v(d_m \vee d_{m'})^{3/4},\ b^2_{A,m,m'} \leq c_e(d_m \vee d_{m'}), D_{A,m} \leq c_D d_m.
$$

The constants c_v and c_D depend on the mixing coefficients and are unknown in practice. Without loss of generality, assume that $\gamma \leq 3/2$ in $[M_5]$.
Let us first assume that $(X_1, ..., X_n)$ are arithmetically β-mixing. Choose $p \geq \sqrt{n}(\ln n)^2/2$, $q \geq \sqrt{n}(\ln n)^{-2}/2$ such that $2pq = n$. Hence, there exists a constant c_M such that $p\beta_q \leq c_M(\log n)^{2(\theta+2)}n^{-\theta/2}$. For all m, m' in $\mathcal{M}_n$,

$$
e_{A,m,m'} \leq 2c_e\frac{d_m \vee d_{m'}}{(\ln n)^4} \leq \frac{2c_e}{c'_D}\frac{D_{A,m}\vee D_{A,m'}}{(\ln n)^4} \leq \frac{2c_e}{c'_D}(\ln n)^{-1}\frac{R_{A,m}\vee R_{A,m'}}{(\ln n)^{2\gamma}}.
$$

When $d_m \vee d_{m'} \leq r_n(\ln n)^{4\gamma}$, then

$$
v^2_{A,m,m'} \leq c_v(d_m \vee d_{m'})^{3/4} \leq c_v(r_n)^{-1/4}\frac{R_n}{(\ln n)^\gamma} \leq c_v(r_n)^{-1/4}\frac{R_{A,m}\vee R_{A,m'}}{(\ln n)^\gamma}.
$$

When $d_m \vee d_{m'} \geq r_n (\ln n)^{4\gamma}$, then

$$v_{A,m,m'}^2 \leq c_v (d_m \vee d_{m'})^{3/4} \leq \frac{c_v}{c'_D} \frac{D_{A,m} \vee D_{A,m'}}{(d_m \vee d_{m'})^{1/4}} \leq \frac{c_v}{c'_D} \frac{R_{A,m} \vee R_{A,m'}}{r_n^{1/4} (\ln n)^\gamma}.$$

Therefore, $[M_1]$-$[M_5]$ and $[\mathbf{AR}(\theta)]$ with $\theta > 1$ imply $[\mathbf{V'}]$ with

$$\epsilon_n^* = C \left((\ln n)^{-1/4} \wedge r_n^{-1/8} \right).$$

Let us now assume that there exists $\theta > 0$ such that the data $X_1, ..., X_n$ are geometrically $[\mathbf{GEO}(\theta)]$ β-mixing. We still assume $[M_1]$-$[M_5]$ on the models. Let $p \geq n(\ln n)^{-2}/2$, $q \geq (\ln n)^2/2$ such that $2pq = n$. Then there exist constants c_e, c_M such that, for all m, m' in $\mathcal{M}_n$,

$$p\beta_q \leq c_M \frac{n}{(\ln n)^2} e^{-\frac{\theta}{2}(\ln n)^2},$$

$$e_{A,m,m'} \leq c_e \frac{(\ln n)^4}{n} (d_m \vee d_{m'}) \leq \frac{c_e}{c'_D} \frac{(\ln n)^{4+2\gamma}}{n} \frac{R_{A,m} \vee R_{A,m'}}{(\ln n)^{2\gamma}}.$$

5.10 Proof of Corollary 4.4.

It is a classical result (see for example Birgé & Massart [BM97]) that the collection of wavelet spaces $\mathcal{M}_n$ satisfies the following assumptions

[T1] *for all* $m \in \mathcal{M}_n$, $2^{J_m} \leq n$*;*
[T2] *there exists a constant* Φ *such that*

$$\forall m, m' \in \mathcal{M}_n, \forall t \in S_m, \forall t' \in S_{m'}, \|t + t'\|_\infty \leq \Phi 2^{(J_m \vee J_{m'})/2} \|t + t'\|_2;$$

[T3] $|\mathcal{M}_n| \leq \ln n / \ln 2$.

Under these assumptions, the following lemma hold.

Lemma 5.4 *Let* $\theta > 2$ *and assume that* $X_1, ..., X_n$ *are arithmetically* $[\mathbf{AR}(\theta)]$ τ*-mixing and let* $u = 3/(1+\theta) \wedge 1$*. Let* $\mathcal{M}_n$ *be a collection of regular wavelet spaces* $[\mathbf{W}]$*. There exist constants* c_D*,* c_v*,* c_b *such that, for all* m*,* m' *in* $\mathcal{M}_n$*,*

$$D_{A,m} \leq c_D 2^{J_m}, \; v_{A,m,m'}^2 \leq c_v \left(2^{J_m \vee J_{m'}}\right)^{\frac{1}{2}(1+u)}, \; b_{A,m,m'}^2 \leq c_b 2^{J_m \vee J_{m'}}.$$

Moreover, $MC_n \leq c_T n^2$*.*

Without loss of generality, assume that $\gamma \leq 3/2$ in **[T5]** and recall that there exists $\theta > 2$ such that $X_1, ..., X_n$ are arithmetically $[\mathbf{AR}(\theta)]$ τ-mixing. Choose $p \geq \sqrt{n}(\ln n)^2/2$, $q \geq \sqrt{n}(\ln n)^{-2}/2$ such that $2pq = n$. Then, $u < 1$ and there exists constants $c_T^{(2)}$, c_e such that

$$\tau_q MC_n \leq c_T^{(2)} \frac{(\ln n)^{2(1+\theta)}}{n^{(\theta-3)/2}}, e_{A,m,m'} \leq \frac{c_e}{\ln n} \frac{R_{A,m} \vee R_{A,m'}}{(\ln n)^{2\gamma}}.$$

When $2^{J_m \vee J_{m'}} \leq r_n (\ln n)^{\frac{2\gamma}{1-u}}$,

$$v_{A,m,m'}^2 \leq c_v \left(r_n (\ln n)^{\frac{2\gamma}{1-u}} \right)^{\frac{1}{2}(1+u)} \leq c_v r_n^{-\frac{1-u}{2}} \frac{R_n}{(\ln n)^\gamma} \leq c_v \frac{R_{A,m} \vee R_{A,m'}}{r_n^{\frac{1-u}{2}} (\ln n)^\gamma}.$$

When $2^{J_m \vee J_{m'}} \geq r_n (\ln n)^{\frac{2\gamma}{1-u}}$,

$$v_{A,m,m'}^2 \leq \frac{c_v}{c'_D} \frac{D_{A,m} \vee D_{A,m'}}{\left(r_n (\ln n)^{\frac{2\gamma}{1-u}} \right)^{\frac{1-u}{2}}} \leq \frac{c_v}{c'_D} r_n^{-\frac{1-u}{2}} \frac{R_{A,m} \vee R_{A,m'}}{(\ln n)^\gamma}.$$

We conclude the proof as in the β-mixing case.

5.11 Proof of Lemma 5.4.

Let us first recall the following lemma, obtained in [Ler09a] as a consequence of the covariance inequality proved by Dedecker & Prieur [DP05] for τ-mixing sequences.

Lemma 5.5 *Let X, Y be two identically distributed real valued random variables, with common density s in $L^2(\mu)$. There exists a constant c_τ and a random variable $b(\sigma(X), Y)$ such that $\mathbb{E}(b(\sigma(X), Y)) = c_\tau\, (\tau(\sigma(X), Y))^{1/3}$ such that, for all Lipschitz functions f and all h in BV*

$$|Cov(f(X), h(Y))| \leq \|h\|_{BV}\, \mathbb{E}\left(|f(X)|b(\sigma(X), Y)\right) \leq c_\tau \|h\|_{BV} \|f\|_\infty \left(\tau(\sigma(X), Y)\right)^{1/3}. \tag{52}$$

It comes from this Lemma and inequalities (20, 21, 22) that

$$\begin{aligned} D_{A,m} &= \frac{1}{q} \sum_{(j,k)\in m} \mathrm{Var}\left(\sum_{i=1}^{q} \psi_{j,k}(X_i)\right) \leq 2 \sum_{(j,k)\in m} \sum_{l=1}^{q} (q+1-l)|\mathrm{Cov}(\psi_{j,k}(X_1), \psi_{j,k}(X_l))| \\ &\leq \frac{2}{q} \sum_{j=0}^{J_m} \sum_{k\in\mathbb{Z}} \sum_{l=1}^{q} \|\psi_{j,k}\|_{BV}\, \mathbb{E}\left(|\psi_{j,k}(X_1)|b(\sigma(X_1), X_l)\right) \\ &\leq 2c_\tau K_{BV} \sum_{j=0}^{J_m} 2^{j/2} \left\|\sum_{k\in\mathbb{Z}} |\psi_{j,k}|\right\|_\infty \sum_{l=1}^{q} \tau_{l-1}^{1/3} \\ &\leq 4\left(c_\tau A K_\infty K_{BV} \sum_{l=0}^{\infty} \tau_l^{1/3}\right) 2^{J_m}. \end{aligned}$$

When $\theta > 2$, the series $\sum_{l=0}^{\infty} \tau_l^{1/3}$ is convergent and we obtain the inequality on $D_{A,m}$ with $c_D = 4\left(c_\tau A K_\infty K_{BV} \sum_{l=0}^{\infty} \tau_l^{1/3}\right)$.
As the models are nested, we only have to compare, for all m in $\mathcal{M}_n$, $b^2_{A,m}$ and $v^2_{A,m}$ with 2^{J_m}. From **[T2]**, $b^2_m \leq \Phi^2 2^{J_m}$, this proves the inequality on $b^2_{A,m,m'}$ with $c_b = \Phi^2$.
For all t in B_m,

$$q\mathrm{Var}(L_q(t)(A_0)) \leq 2 \sum_{l=1}^{q} |\mathrm{Cov}(t(X_1), t(X_l))|. \tag{53}$$

Let X_l^* be a random variable, independent of X_1, with law P, such that

$$\mathbb{E}\left(|X_l - X_l^*|\right) \leq \tau_{l-1}.$$

This random variable can be defined thanks to the coupling lemma of Dedecker & Prieur [DP05] (section 7.1).

$$\begin{aligned} |\mathrm{Cov}(t(X_1), t(X_l))| &= |\mathrm{Cov}(t(X_1), t(X_l) - t(X_l^*))| \\ &\leq \sqrt{\mathrm{Var}(t(X_1))\mathbb{E}\left((t(X_l) - t(X_l^*))^2\right)} \\ &\leq \sqrt{2\mathrm{Var}(t(X_1))\, \|t\|_\infty\, \mathbb{E}\left(|t(X_l) - t(X_l^*)|\right)} \\ &\leq \sqrt{2\|s\|\, \|t\|_\infty^2\, \mathrm{Lip}(t)\tau_{l-1}}. \end{aligned}$$

Since t belongs to B_m, $\|t\|_\infty^2 \leq \Phi^2 2^{J_m}$. Moreover, let $a_{j,k} = \int_{\mathbb{R}} t\psi_{j,k} d\mu$, then

$$\begin{aligned} \mathrm{Lip}(t) &= \sup_{x\neq y\in\mathbb{R}} \frac{|t(x)-t(y)|}{|x-y|} \leq \sum_{j=0}^{J_m} \sup_{x\neq y\in\mathbb{R}} \sum_{k\in\mathbb{Z}} |a_{j,k}| \frac{|\psi_{j,k}(x)-\psi_{j,k}(y)|}{|x-y|} \\ &\leq 2AK_L \sum_{j=0}^{J_m} 2^{3j/2} \sup_{k\in\mathbb{Z}} |a_{j,k}|. \end{aligned} \tag{54}$$

The last inequality holds since, for all x, y in $\mathbb{R}$ there is less than $2A$ indices k in $\mathbb{Z}$ such that $|\psi_{j,k}(x) - \psi_{j,k}(y)| \neq 0$. Since t belongs to B_m, $\sum_{(j,k)\in m} a_{j,k}^2 \leq 1$, in particular, for all j, $\sup_{k\in\mathbb{Z}} |a_{j,k}| \leq 1$. Thus, there exists a constant c such that $\mathrm{Lip}(t) \leq c2^{3J_m/2}$. Hence, there exists a constant c such that, for all t in B_m and all l in $\mathbb{N}^*$

$$|\mathrm{Cov}(t(X_1), t(X_l))| \leq c2^{5J_m/4}\sqrt{\tau_{l-1}}.$$

Remark that we also have

$$|\mathrm{Cov}(t(X_1), t(X_l))| \leq \|t\|_\infty \|t\| \|s\| \leq c2^{J_m/2}.$$

Recall that $u = 3/(1+\theta)$, there exist constants c, which may vary from line to line such that

$$\begin{aligned} \sum_{l=1}^{q} |\mathrm{Cov}(t(X_1), t(X_l))| &\leq c2^{J_m/2} \sum_{l=1}^{\infty} (2^{3J_m/4}\sqrt{\tau_{l-1}} \wedge 1) \\ &\leq c2^{J_m/2} \sum_{l=1}^{\infty} (2^{3J_m/4} l^{-(1+\theta)/2} \wedge 1) \\ &\leq c2^{J_m/2} \left(\sum_{l=1}^{2^{uJ_m/2}} 1 + \sum_{l=2^{uJ_m/2}}^{\infty} 2^{3J_m/4} l^{-(1+\theta)/2} \right) \\ &\leq c2^{\frac{J_m}{2}(1+u)}. \end{aligned}$$

We deduce the inequality on $v_{A,m,m'}^2$ from (53) and this last inequality.
It remains to control MC_n, recall that

$$MC_n = \sum_{m\in\mathcal{M}_n} \left(\left\| \sum_{\lambda\in m} |\psi_\lambda| \right\|_\infty \sup_{\lambda\in m} \mathrm{Lip}_d(\psi_\lambda) + \|s\| |\mathcal{M}_n| \sup_{t\in B_m} \mathrm{Lip}(t) \right).$$

From (20, 21), for all m in $\mathcal{M}_n$,

$$\left\| \sum_{\lambda\in m} |\psi_\lambda| \right\|_\infty \leq \frac{\sqrt{2}}{\sqrt{2}-1} AK_\infty 2^{J_m/2}, \ \sup_{\lambda\in m} \mathrm{Lip}_d(\psi_\lambda) \leq K_L 2^{3J_m/2}.$$

From (54), for all m in $\mathcal{M}_n$, there exists a constant c such that $\sup_{t\in B_m} \mathrm{Lip}(t) \leq c2^{3J_m/2}$. Since $\mathrm{Card}(\mathcal{M}_n) \leq \ln n / \ln 2$, and $2^{\max_{m\in\mathcal{M}_n} J_m} \leq n$, there exists a constant c_M such that $MC_n \leq c_M n^2$.

6 Appendix

In this section, we recall some technical lemmas proved in [Ler09b].

Lemma 6.1 *For all $\alpha \geq 0$, $K > \alpha + 1$,*

$$\Sigma(K, \alpha) = \sum_{k \in \mathbb{N}} \sum_{m \in \mathcal{M}_n^k} (1+k)^\alpha e^{-K[\ln(1+\mathrm{Card}(\mathcal{M}_n^k))+\ln(1+k)]} < \infty.$$

For all m in $\mathcal{M}_n$, let $l_m = l_{n,\gamma}(R_{A,m}, R_{A,m})$. Then, for all $\alpha \geq 0$, for all $K > \sqrt{(1+\alpha)/2}$,

$$\sum_{m \in \mathcal{M}_n} R_{A,m}^\alpha e^{K^2 l_m} \leq \Sigma(K^2, \alpha) e^{-K^2 (\ln n)^\gamma}.$$

For all m in $\mathcal{M}_n$, let $l_{m,m'} = l_{n,\gamma}(R_{A,m}, R_{A,m'})$. Then, for all $\alpha \geq 0$, $\alpha' \geq 0$ and all $K > \sqrt{1 + \alpha \vee \alpha'}$,

$$\sum_{(m,m') \in (\mathcal{M}_n)^2} R_{A,m}^\alpha R_{A,m'}^{\alpha'} e^{-K^2 l_{m,m'}} = \Sigma(K^2, \alpha) \Sigma(K^2, \alpha') e^{-K^2 (\ln n)^\gamma}.$$

Lemma 6.2 *Let n be an integer and let $X_1, ..., X_n$ be real valued, identically distributed random variables with common law P. Let $(t_\lambda)_{\lambda \in \Lambda}$ be a collection of functions in $L^2(\mu)$. Let $p(\Lambda) = \sum_{\lambda \in \Lambda} (\nu_n(t_\lambda))^2$. Let $(W_1, ..., W_n)$ be a resampling scheme, let $\bar{W}_n = \sum_{i=1}^n W_i/n$ and let $v_W^2 = Var(W_1 - \bar{W}_n)$. Let*

$$p^W(\Lambda) = (v_W^2)^{-1} \sum_{\lambda \in \Lambda} \mathbb{E}^W \left((\nu_n^W(t_\lambda))^2 \right),$$

$T = \sum_{\lambda \in \Lambda} (t_\lambda - P t_\lambda)^2$ and

$$U = \frac{1}{n(n-1)} \sum_{i \neq j = 1}^n \sum_{\lambda \in \Lambda} (t_\lambda(X_i) - P t_\lambda)(t_\lambda(X_j) - P t_\lambda).$$

Then

$$p(\Lambda) = \frac{1}{n} P_n T + \frac{n-1}{n} U, \; p^W(\Lambda) = \frac{1}{n} P_n T - \frac{1}{n} U, \; p(\Lambda) - p^W(\Lambda) = U.$$

Proposition 6.3 *Let $X, X_1, ..., X_n$ be i.i.d random variables taking value in a measurable space $(\mathbb{X}, \mathcal{X})$ with common law P. Let B be a symetric class of functions bounded by b. Let $Z = \sup_{t \in B} (\nu_n t)$, $\epsilon = b^2/n$, $v^2 = \sup_{t \in B} Var(t(X))$, $D = \mathbb{E} \left(\sup_{t \in B} (t(X) - Pt)^2 \right)$. For all $x > 0$, we have*

$$\mathbb{P} \left(Z^2 - \frac{D}{n} > \frac{D^{3/4} (\epsilon (19x)^2)^{1/4} + 3\sqrt{D v^2 x} + 3 v^2 x + \epsilon (19x)^2}{n} \right) \leq e^{-x}.$$

$$\mathbb{P} \left(Z^2 - \frac{D}{n} < - \frac{8 D^{3/4} (\epsilon x^2)^{1/4} + 7.61 \sqrt{v^2 D x} + \epsilon (40.25x)^2}{n} \right) \leq 2.8 e^{-x}.$$

Lemma 6.4 *Let $X, X_1, ..., X_n$ be i.i.d random variables taking value in a measurable space $(\mathbb{X}, \mathcal{X})$ with common law P. Let μ be a measure on $(\mathbb{X}, \mathcal{X})$ and let $(t_\lambda)_{\lambda \in \Lambda}$ be a set of functions in $L^2(\mu)$. Let $B = \{ t = \sum_{\lambda \in \Lambda} a_\lambda t_\lambda, \; \sum_{\lambda \in \Lambda} a_\lambda^2 \leq 1 \}$, $D = \mathbb{E} \left(\sup_{t \in B} (t(X) - Pt)^2 \right)$, $v^2 = \sup_{t \in B} Var(t(X))$, $b = \sup_{t \in B} \|t\|_\infty$ and $\epsilon = b^2/n$. Let*

$$U = \frac{1}{n(n-1)} \sum_{i \neq j = 1}^n \sum_{\lambda \in \Lambda} (t_\lambda(X_i) - P t_\lambda)(t_\lambda(X_j) - P t_\lambda).$$

Then the following inequality holds

$$\forall x > 0,\ \mathbb{P}\left(U > \frac{5.31 D^{3/4}(\epsilon x^2)^{1/4} + 3\sqrt{v^2 D x} + 3v^2 x + \epsilon(19.1x)^2}{n-1}\right) \leq 2e^{-x}.$$

$$\forall x > 0,\ \mathbb{P}\left(U < -\frac{9 D^{3/4}(\epsilon x^2)^{1/4} + 7.61\sqrt{v^2 D x} + \epsilon(40.3x)^2}{n-1}\right) \leq 3.8e^{-x}.$$

Lemma 6.5 *Let $X, X_1, ..., X_n$ be i.i.d random variables taking value in a measurable space $(\mathbb{X}, \mathcal{X})$ with common law P. Let μ be a measure on $(\mathbb{X}, \mathcal{X})$ and let $(\psi_\lambda)_{\lambda\in\Lambda}$ be an orthonormal system in $L^2(\mu)$. Let L be a linear functional in $L^2(\mu)$ and let $B = \{t = \sum_{\lambda\in\Lambda} a_\lambda L(\psi_\lambda),\ \sum_{\lambda\in\Lambda} a_\lambda^2 \leq 1\}$, $v^2 = \sup_{t\in B} Var(t(X))$, $b = \sup_{t\in B} \|t\|_\infty$ and $\epsilon = b^2/n$. Let s be a function in S, the linear space spanned by the functions $(t_\lambda)_{\lambda\in\Lambda}$ and let $\eta > 0$. Then the following inequality holds*

$$\forall x > 0,\ \mathbb{P}\left(\nu_n(L(s)) > \frac{\eta}{2}\|s\|^2 + \frac{2v^2 x + \epsilon x^2/9}{\eta n}\right) \leq e^{-x}.$$

References

[Aka70] H. Akaike. Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22:203–217, 1970.

[Aka73] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest, 1973.

[AM09] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *Journal of Machine learning research*, 10:245–279, 2009.

[Arl07] S. Arlot. *Resampling and model selection.* PhD thesis, Université Paris-Sud 11, 2007.

[Arl09] S. Arlot. Model selection by resampling penalization. *Electron. J. Statist.*, 3:557–624, 2009.

[BBM99] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

[BCV01] Y. Baraud, F. Comte, and G. Viennet. Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.*, 29(3):839–875, 2001.

[Bir08] L. Birgé. Model selection for density estimation with l^2-loss. *Preprint*, 2008.

[BM97] L. Birgé and P. Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York, 1997.

[BM07] L. Birgé and P. Massart. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2):33–73, 2007.

[Bra07] R. C. Bradley. *Introduction to strong mixing conditions. Vol. 1.* Kendrick Press, Heber City, UT, 2007.

[CDT08] F. Comte, J. Dedecker, and M. L. Taupin. Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.*, 17(2):87–112, 2008.

[Cél08] A. Célisse. Density estimation via cross validation: Model selection point of view. *Preprint, downloadable on arXiv.org : 08110802*, 2008.

[CM02] F. Comte and F. Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238 (electronic), 2002. New directions in time series analysis (Luminy, 2001).

[DDL+07] J. Dedecker, P. Doukhan, G. Lang, J. R. León R., S. Louhichi, and C. Prieur. *Weak dependence: with examples and applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York, 2007.

[DJKP96] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 1996.

[Dou94] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

[DP05] J. Dedecker and C. Prieur. New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, 132(2):203–236, 2005.

[Kün89] H.R. Künsch. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, 17:1217–1241, 1989.

[Ler09a] M. Lerasle. Adaptive density estimation of stationary β-mixing and τ-mixing sequences. *Mathematical Methods of statistics*, 18(1):59–83, 2009.

[Ler09b] M Lerasle. Optimal model selection in density estimation. *arXiv:0910.1654*, 2009.

[LS92] R. Y. Liu and K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 225–248. Wiley, New York, 1992.

[Mal73] C.L. Mallows. Some comments on c_p. *Technometrics*, 15:661–675, 1973.

[Rad02] D. Radulović. On the bootstrap and empirical processes for dependent sequences. In *Empirical process techniques for dependent data*, pages 345–364. Birkhäuser Boston, Boston, MA, 2002.

[Rud82] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, 9(2):65–78, 1982.

[Sto74] M. Stone. Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. Ser. B*, 36:111–147, 1974. With discussion by G. A. Barnard, A. C. Atkinson, L. K. Chan, A. P. Dawid, F. Downton, J. Dickey, A. G. Baker, O. Barndorff-Nielsen, D. R. Cox, S. Giesser, D. Hinkley, R. R. Hocking, and A. S. Young, and with a reply by the authors.

[Vie97] G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492, 1997.

[VR59] V. A. Volkonskiĭ and Y. A. Rozanov. Some limit theorems for random functions. I. *Teor. Veroyatnost. i Primenen*, 4:186–207, 1959.