Submitted to Bernoulli

Rademacher complexity for Markov chains: Applications to kernel smoothing and Metropolis-Hasting

PATRICE BERTAIL[*] and FRANÇOIS PORTIER[†]

[*] *Modal'X, UPL, University of Paris-Nanterre*
[†] *Télécom ParisTech, University of Paris-Saclay*

The concept of Rademacher complexity for independent sequences of random variables is extended to Markov chains. The proposed notion of "regenerative block Rademacher complexity" (of a class of functions) follows from renewal theory and allows to control the expected values of suprema (over the class of functions) of empirical processes based on Harris Markov chains as well as the excess probability. For classes of Vapnik-Chervonenkis type, bounds on the "regenerative block Rademacher complexity" are established. These bounds depend essentially on the sample size and the probability tails of the regeneration times. The proposed approach is employed to obtain convergence rates for the kernel density estimator of the stationary measure and to derive concentration inequalities for the Metropolis-Hasting algorithm.

MSC 2010 subject classifications: Primary 62M05; secondary 62G07, 60J22.
Keywords: Markov chains, Concentration inequalities, Rademacher complexity, Kernel smoothing, Metropolis Hasting.

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose that $X = (X_i)_{i\in\mathbb{N}}$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ valued in $(E, \mathcal{E})$. Let $\mathcal{F}$ denote a countable class of real-valued measurable functions defined on E. Let $n \geq 1$, define

$$Z = \sup_{f\in\mathcal{F}} \left| \sum_{i=1}^{n} (f(X_i) - \mathbb{E}[f(X_i)]) \right|.$$

The random variable Z plays a crucial role in machine learning and statistics: it can be used to bound the risk of an algorithm [63, 14] as well as to study *M and Z estimators* [60]; it serves to describe the (uniform) accuracy of function estimates such as the cumulative distribution function, the quantile function or the cumulative hazard functions [56] or *kernel smoothing estimates* of the probability density function [21, 24]. Depending on

[*]E-mail: patrice.bertail@gmail.com
[†]E-mail: francois.portier@gmail.com

the class $\mathcal{F}$ many different bounds are known when X is formed by independent and identically distributed (i.i.d.) random variables.. Reference textbooks include [61, 17, 14, 13, 26].

When X is a Markov chain, many approaches have already been investigated. For a single function f, exponential-type bounds on the excess probability of Z, i.e., $\mathbb{P}(Z > t)$, are obtained in [10] based on the regenerative approach; in [31] using a curvature assumption; in [48] using spectral methods. For general classes $\mathcal{F}$, the concentration of Z around its expected value, i.e., bounding $\mathbb{P}(|Z - \mathbb{E}Z| > t)$, is studied in [1] where a Berstein-type bound is established; and in [18] and [48] where the technique of *bounded differences* [38] is employed to derive Hoeffding-type inequalities (see also [66] which extends [18] to the case of unbounded chains). For instance, the inequalities of [48] hold under uniform ergodicity conditions but are explicit in term or the mixing time of the chain, making them operational in many applications. Similarly, the exponential bounds obtained in [43], based on chaining arguments, also hold under uniformly ergodic conditions. Since aperiodic Harris recurrent chains are β-mixing processes (see for instance [15] and the references therein), the McDiarmid types of inequalities obtained by [42] and the notion of Rademacher complexity introduced and studied in [41] can be applied when X is Markovian. Notice however that their bounds are not exactly of exponential type because they are affected by the rate of convergence of the β-mixing coefficients (see Remark 3 for more details).

The overall goal of the paper is to establish new bounds on the expected value of Z as well as on the excess probability of Z when X is a Markov chain and the class $\mathcal{F}$ has a complexity of Vapnik-Chervonenkis (VC) type. In contrast with previous works, we study directly the tail of Z and our bounds involve the variance of the class $\mathcal{F}$ and do not require uniform geometric ergodicity.

The approach taken in this paper is based on renewal theory and is known as the *regenerative method*, see [57, 45, 3]. Indeed it is well known that sample paths of a Harris chain may be divided into i.i.d. *regeneration blocks*. These blocks are defined as data cycles between random times called *regeneration times* at which the chain forgets its past. Hence, most of the results established in the i.i.d. setup may be extended to the Markovian framework by applying the latter to (functionals of) the regeneration blocks. Refer to [40] for the *strong law of large numbers* and the *central limit theorem*, to [35] for functional CLT, as well as [12, 36, 37, 9, 7, 19] for refinements of the central limit theorem.

Following the seminal approach by Vapnik and Cèrvonenkis [64], we introduce a new notion of complexity that we call the *regenerative block Rademacher complexity* which extends the classical Rademacher complexity for independent sequences of random variables to Markov chains. Refer to the books [33, 13, 26] for nice accounts and applications of Rademacher complexity in the i.i.d cases. As in the independent case, the *regenerative block Rademacher complexity* is useful to bound the expected values of empirical processes (over some classes of functions) and intervenes as well to control the excess probability. Depending on the probability tails of the regeneration times, which are considered to be either exponential or polynomial, we derive bounds on the regenerative block Rademacher complexity of classes of VC type. Interestingly, the obtained bounds bears resemblance to the ones provided in [21, 24] (for independent X) as they depend on the variance of the underlying class of functions $\mathcal{F}$ allowing to take advantage of classes $\mathcal{F}$ having small fluctuations.

Kernel density estimator. Kernel density estimators, as well as their variations, *Nadaraya-Watson*, *nearest neighbors* or delta-sequences estimators [65], are local averaging techniques forming the basis of nonparametric estimation. They are at the core of many semi-parametric statistical procedures [2, 50] in which controlling Z-type quantities permits to take advantage of the tightness of the empirical process [62]. The asymptotic properties of kernel density estimators, based on independent and identically distributed data, are well understood since the seventies-eighties [58]. However finite sample properties were only studied more recently [21, 24]. The function class of interest (taking the role of $\mathcal{F}$) in this problem is given by

$$\mathcal{K}_n = \{x \mapsto K((x-y)/h_n) \, : \, y \in \mathbb{R}^d\},$$

where $K : \mathbb{R}^d \to \mathbb{R}$ is called the kernel and $(h_n)_{n\in\mathbb{N}}$ is a positive sequence converging to 0 called the bandwidth. Based on the property that $\mathcal{K}_n$ is included on some VC class [44], some master results have been obtained by [21, 22, 23, 24] who proved some concentration inequalities, based on the seminal work of [59], allowing to establish precisely the rate of uniform convergence of kernel density estimators. Kernel density estimates are particular because the variance of each element in $\mathcal{K}_n$ goes to 0 as $n \to \infty$. This needs to be considered to derive accurate bounds, e.g., the one presented in [24]. The proposed approach takes care of this phenomenon as, under reasonable conditions, our bound for Markov chains scales at the same rate as the ones obtained in the independent case. Note that our results extend the ones given in [4] where under similar assumptions the consistency is established.

The study of kernel estimators for dependent data has only recently received special attention in the statistical literature. To the best of our knowledge, uniform results are limited to the alpha and beta mixing cases when dependency occurs [49, 28] by using *coupling techniques*.

Metropolis-Hasting algorithm. Metropolis-Hasting (MH) algorithm is one of the state of the art method in computational statistics and is frequently used to compute Bayesian estimators [53]. Theoretical results for MH are often deduced from the analysis of geometrically ergodic Markov chains as presented for instance in [39, 54, 30, 55, 20]. Whereas many results on the asymptotic behavior of MH are known, e.g. central limit theorem or convergence in total variation, only few non-asymptotic results are available for such Markov chains; see for instance [34] where the estimation error is controlled *via* a Rosenthal-type inequality. We consider the popular *random walk MH*, which is at the heart of the adaptive MH version introduced in [27]. Building upon the pioneer works [54, 30] where the geometric ergodicity is established for the random walk MH, we show that whenever the class $\mathcal{F}$ is VC, the expected value of $\sup_{f\in\mathcal{F}} |\sum_{i=1}^n (f(X_i) - \int f d\pi)|$ is bounded by $D\sqrt{n(1 \vee \log(\log(n)))}$, where π stands for the stationary measure and $D > 0$

depends notably on the distribution of the regeneration times. By further applying this to the quantile function, we obtain a concentration inequality for Bayesian credible intervals.

Outline. The paper is organized as follows. In section 2, the notations and main assumptions are first set out. Conceptual background related to the renewal properties of Harris chains and the regenerative method are also briefly exposed. In section 3, the notion of block Rademacher complexity for Markov chains is introduced and some basic results on block VC classes are presented. Section 4 provides the main result of the paper: a bound on the Rademacher complexity. Our methodology is illustrated in section 5 on kernel density estimation and MH.

2. Regenerative Markov chains

2.1. Basic definitions

In this section, for seek of completeness we recall the following important basic definitions and properties of regenerative Markov chains. An interested reader may look into [46] or [40] for detailed survey of regeneration theory.

Consider an homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}}$ defined on $(\Omega, \mathcal{F}, \mathbb{P}_\nu)$ valued in a countably generated state space $(E, \mathcal{E})$ with transition probability $P(.,.)$, and initial probability ν. The assumption that $\mathcal{E}$ is countably generated allows to avoid measurability problems. For any $x \in E$ and any probability measure μ, the notation $\mathbb{P}_x$ (resp. $\mathbb{P}_\mu$) stands for the probability measure such that $X_0 = x$ (resp. $X_0 \sim \mu$), and $\mathbb{E}_x(\cdot)$ (resp. $\mathbb{E}_\mu(\cdot)$) stands for the associated expectation. For any $n \geq 1$, let P^n denote the n-th iterate of the transition probability P. Given a set $B \in E$, define τ_B as the first time the chain enters B.

Definition 1 (irreducibility). *The chain is ψ-irreducible if there exists a σ-finite measure ψ such that, for all set $B \in \mathcal{E}$ satisfying $\psi(B) > 0$, for any $x \in E$ there exists $n \geq 1$ such that $P^n(x, B) > 0$. With words, regardless of the starting point, the chain visits B with strictly positive probability.*

Definition 2 (aperiodicity). *Assuming ψ-irreducibility, there exists $d' \in \mathbb{N}^*$ disjoint sets $D_1, ..., D_{d'}$ (set $D_{d'+1} = D_1$) positively weighted by ψ such that $\psi(E\backslash \cup_{1\leqslant i\leqslant d'} D_i) = 0$ and $\forall x \in D_i$, $P(x, D_{i+1}) = 1$. The period of the chain is the g.c.d. d of such integers, it is said to be aperiodic if $d = 1$.*

Definition 3 (Harris recurrence). *A ψ-irreducible Markov chain is said to be positive Harris recurrent if for all $B \in E$ with $\psi(B) > 0$, we have $\mathbb{E}_x\tau_B < \infty$ for all $x \in B$.*

Recall that a chain is positive Harris recurrent and aperiodic if and only if it is ergodic [46, Proposition 6.3], i.e., there exists a probability measure π, called the stationary distribution, such that $\lim_{n\to+\infty} \|P^n(x, \cdot) - \pi\|_{\mathrm{tv}} = 0$. The Nummelin splitting technique (presented in the forthcoming section) depends heavily on the notion of small set. Such sets exist for positive Harris recurrent chain [29].

Definition 4 (small sets)**.** *A set $S \in \mathcal{E}$ is said to be Ψ-small if there exists $\delta > 0$, a positive probability measure Ψ supported by S and an integer $m \in \mathbb{N}^*$ such that*

$$\forall x \in S,\ B \in \mathcal{E} \ \ P^m(x, B) \geq \delta\ \Psi(B). \tag{2.1}$$

In the whole paper, we work under the following generic hypothesis in which the chain is supposed to be Harris recurrent.

(H) The chain $(X_n)_{n\in\mathbb{N}}$ is a positive Harris recurrent aperiodic Markov chain defined on $(\Omega, \mathcal{F}, \mathbb{P}_\nu)$ valued in a countably generated state space $(E, \mathcal{E})$ with transition kernel $P(x, dy)$ and initial measure ν. Let S be Ψ-small with $m = 1$ and suppose that the hitting time τ_S satisfies

$$\sup_{x\in S} \mathbb{E}_x[\tau_S] < \infty, \qquad \text{and} \qquad \mathbb{E}_\nu[\tau_S] < \infty.$$

This is only for clarity reasons that we assume that $m = 1$. As explained in Remark 9 below, the study of sums over general Harris chain, i.e., when $m \geq 1$, can easily be derived from the case $m = 1$.

2.2. The Nummelin splitting technique

The Nummelin splitting technique [45, 3] allows to retrieve all regeneration properties for general Harris Markov chains. It consists in extending the probabilistic structure of the chain in order to construct an artificial atom [47]. Start by recalling the definition of regenerative chains.

Definition 5 (regenerative chain)**.** *We say that a ψ-irreducible, aperiodic chain is regenerative or atomic if there exists a measurable set A called an atom, such that $\psi(A) > 0$ and for all $(x, y) \in A^2$ we have $P(x, \cdot) = P(y, \cdot)$. Roughly speaking, an atom is a set on which the transition probabilities are the same.*

Assume that the chain X satisfies the generic hypothesis (H). Then the sample space is expanded in order to define a sequence $(Y_n)_{n\in\mathbb{N}}$ of independent Bernoulli random variables with parameter δ. The construction relies on the mixture representation of P on S, namely $P(x, A) = \delta\Psi(A) + (1 - \delta)(P(x, A) - \delta\Psi(A))/(1 - \delta)$, with two components, one of which not depending on the starting point (implying regeneration when this component is picked up in the mixture). The regeneration structure can be retrieved by the following randomization of the transition probability P each time the chain X visits the set S:

- If $X_n \in S$ and $Y_n = 1$ (which happens with probability $\delta \in]0, 1[$), then X_{n+1} is distributed according to the probability measure Ψ,

- If $X_n \in S$ and $Y_n = 0$ (that happens with probability $1-\delta$), then X_{n+1} is distributed according to the probability measure $(1-\delta)^{-1}(P(X_n,\cdot)-\delta\Psi(\cdot))$.

The bivariate Markov chain $Z = (X_n, Y_n)_{n\in\mathbb{N}}$ is called the *split chain*. It takes its values in $E \times \{0,1\}$ and is atomic with atom given by $A = S \times \{1\}$. Define the sequence of *regeneration times* $(\tau_A(j))_{j\geq 1}$, i.e.

$$\tau_A = \tau_A(1) = \inf\{n \geq 1 \,:\, Z_n \in A\}$$

and, for $j \geq 2$,

$$\tau_A(j) = \inf\{n > \tau_A(j-1) \,:\, Z_n \in A\}.$$

It is well known that the bivariate chain Z inherits all the stability and communication properties of the chain X, as aperiodicity and ψ-irreducibility. For instance, the regeneration time has a finite expectation (by recurrence property). More precisely, it holds that [4, Lemma ∗A1]

$$\sup_{x\in A}\mathbb{E}_x[\tau_A] < \infty \qquad \text{and} \qquad \mathbb{E}_\nu[\tau_A] < \infty.$$

It is known from regeneration theory [40] that given the sequence $(\tau_A(j))_{j\geq 1}$, we can divide the chain into block segments or cycles defined by

$$B_j = (X_{1+\tau_A(j)}, \cdots, X_{\tau_A(j+1)}),\ j \geq 1$$

according to the consecutive visits of the chain to the regeneration set A. The strong Markov property implies that $(\tau_A(j))_{j\geq 1}$ and $(B_j)_{j\geq 1}$ are i.i.d. [7, Lemma 3.1]. Denote by $\mathbb{P}_A$ the probability measure such that $Z_0 \in A$. The stationary distribution is given by the Pitman's occupation measure:

$$\pi(B) = \frac{1}{\mathbb{E}_A(\tau_A)}\mathbb{E}_A\left(\sum_{i=1}^{\tau_A}\mathbb{I}_B(X_i)\right),\ \forall B \in \mathcal{E},$$

where $\mathbb{I}_B$ is the indicator function of the event B.

Remark 1 (small set or atom). *The Nummelin splitting technique is useless in the case of countable state space chains for which any state is an atom. For sake of generality, we choose to focus on the general framework of Harris chains.*

3. Regenerative Block Rademacher complexity

3.1. The independent case

Let $\xi = (\xi_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ valued in $(E, \mathcal{E})$ with common distribution P on $(E, \mathcal{E})$. Let $\mathcal{F}$ be a countable class of real-valued measurable functions defined on E. The Rademacher complexity associated to $\mathcal{F}$ is given by

$$R_{n,\xi}(\mathcal{F}) = \mathbb{E} \sup_{f\in\mathcal{F}} \left| \sum_{i=1}^{n} \epsilon_i f(\xi_i) \right|,$$

where the $(\epsilon_i)_{i\in\mathbb{N}}$ are i.i.d. Rademacher random variables, i.e., taking values $+1$ and -1, with probability $1/2$, independent from ξ.

The notion of VC class is powerful because it covers many interesting classes of functions and ensures suitable properties on the Rademacher complexity. The function F is an envelope for the class $\mathcal{F}$ if $|f(x)| \le F(x)$ for all $x \in E$ and all $f \in \mathcal{F}$. For a metric space $(\mathcal{F}, d)$, the covering number $\mathcal{N}(\epsilon, \mathcal{F}, d)$ is the minimal number of balls of size ϵ needed to cover $\mathcal{F}$. The metric of interest is the $L_2(Q)$-norm denoted by $\|.\|_{L_2(Q)}$ and given by $\|f\|_{L_2(Q)} = \{\int f^2 dQ\}^{1/2}$.

Definition 6 (VC class)**.** *A class $\mathcal{F}$ of measurable functions $E \to \mathbb{R}$ is said to be of VC-type (or Vapnik-Chervonenkis type) for an envelope F and admissible characteristic (C, v) (positive constants such that $C \ge (3\sqrt{e})^v$ and $v \ge 1$), if for all probability measure Q on $(E, \mathcal{E})$ with $0 < \|F\|_{L_2(Q)} < \infty$ and every $0 < \epsilon < 1$,*

$$\mathcal{N}\left(\epsilon\|F\|_{L_2(Q)},\, \mathcal{F},\, \|.\|_{L_2(Q)}\right) \le C\epsilon^{-v}.$$

We also assume that the class is countable to avoid measurability issues (but the non-countable case may be handled similarly by using outer probability and additional measurability assumptions, see [62]).

The next theorem is taken from [23], Proposition 2.1, and has been successfully applied to kernel density estimators in [24]. A similar approach is provided in [22], Proposition 1.

Theorem 1 ([23], Proposition 2.1)**.** *Let $\mathcal{F}$ be a measurable uniformly bounded VC class of functions defined on E with envelop F and characteristic (C, v). Let $U > 0$ such that $|f(x)| \le U$ for all $x \in E$ and $f \in \mathcal{F}$. Let σ^2 be such that $\mathbb{E}[f(\xi)^2] \le \sigma^2$ for all $f \in \mathcal{F}$. Then, whenever $0 < \sigma \le U$, it holds*

$$R_{n,\xi}(\mathcal{F}) \le M \left[vU \log\frac{CU}{\sigma} + \sqrt{vn\sigma^2 \log\frac{CU}{\sigma}} \right],$$

where M is a universal constant.

3.2. The Harris case

To extend the previous approach to any Harris chain X, we decompose the chain X according to the independent blocks $(B_j)_{j\ge1}$ introduced in section 2.2. Let $n \ge 1$ and define

$$l_n = \sum_{i=1}^{n} \mathbb{I}_A(X_i),$$

the total number of renewals before n. Assuming that $l_n > 1$, we thus observe $l_n - 1$ (complete) i.i.d. blocks, namely $B_1, \ldots, B_{l_{n-1}}$. The first block $B_0 = (X_1, \ldots, X_{\tau_A(1)})$ and the last block $B_{l_n} = (X_{\tau_A(l_n)}, \ldots, X_n)$, often called incomplete blocks, are not part of the i.i.d. sequence simply because they have a different distribution than B_1. Those incomplete blocks will be treated separately. We have

$$\left|\sum_{i=1}^{n}(f(X_i) - \mathbb{E}_\pi[f])\right| \leq \left|\sum_{i=\tau_A(1)+1}^{\tau_A(l_n)}(f(X_i) - \mathbb{E}_\pi[f])\right| + \left|\sum_{i=1}^{\tau_A}(f(X_i) - \mathbb{E}_\pi[f])\right| + \left|\sum_{i=\tau_A(l_n)+1}^{n}(f(X_i) - \mathbb{E}_\pi[f])\right|, \tag{3.1}$$

with the convention that empty sums are 0. Because $\tau_A(l_n) - \tau_A(1) = \sum_{k=1}^{l_n-1} \ell(B_k)$, where $\ell(B_k)$ denote the size of block k, it holds that

$$\left|\sum_{i=\tau_A(1)+1}^{\tau_A(l_n)}(f(X_i) - \mathbb{E}_\pi[f])\right| = \left|\sum_{k=1}^{l_n-1}(f'(B_k) - \ell(B_k)\mathbb{E}_\pi[f])\right|,$$

$$f'(B_k) = \sum_{i=\tau_A(k)+1}^{\tau_A(k+1)} f(X_i).$$

Hence this term is a (random) summation over complete blocks. Recall that, under (H), $l_n/n \to 1/\mathbb{E}_A\tau_A$, $\mathbb{P}_\nu$-almost surely. Thus, aiming to reproduce the Rademacher approach in the i.i.d. setting, we introduce the following block Rademacher complexity of the class $\mathcal{F}$.

Definition 7 (regenerative block Rademacher complexity)**.** *The regenerative block Rademacher complexity of a class $\mathcal{F}$ associated to a Harris chain X with atom A is given by*

$$R_{n,B}(\mathcal{F}) = \mathbb{E}_A \sup_{f\in\mathcal{F}} \left|\sum_{k=1}^{n} \epsilon_k f'(B_k)\right|,$$

where $(\epsilon_k)_{k\in\mathbb{N}}$ are Rademacher random variables independent from the blocks $(B_k)_{k\in\mathbb{N}}$.

Remark 2 (random number of blocks)**.** *The number of blocks $l_n - 1$ is random and correlated to the blocks itself. This causes a major difficulty when deriving second order asymptotic results as well as non-asymptotic results for regenerative Markov chains. In* *the subsequent development, Lemma 1.2.6 in [17] (see also Theorem 3.1.21 in [26] for a refinement) plays a major role as it will relate the expected behavior of the empirical sums over $l_n - 1$ blocks to simple Rademacher sums over n blocks, i.e., $R_{n,B}(\mathcal{F})$ (see Theorem 5 for more details).*

Remark 3 (comparison to other Rademacher complexities)**.** *In [41] the authors consider the notion of Rademacher for β-mixing stationary processes. To control this Radema--cher complexity, the authors make use of Berbee's coupling techniques [8] which consist in replacing fixed length dependent blocks by independent ones up to an error depending on the mixing coefficient $\beta(.)$. If we denote by b_n the block length, then we pay the price of dependence by an additional term of order $n\beta(b_n)/b_n$ which deteriorates the convergence rate. In our case, because the regenerative blocking techniques is based on small blocks (of random length) with average size $E_A\tau_A$, their is no loss in term of rate of convergence. For sake of completeness, we also mention that [52] introduced a notion of sequential Rademacher complexity adapted to special martingale structures on trees. However, they cannot be used in our framework.*

Remark 4. *The notion of Rademacher complexity is at the heart of many generalisation bounds in machine learning [5] and in model selection [32]. We refer to the books [33, 26] for a nice account of this field. In the i.i.d case, this quantity may itself be estimated by the empirical Rademacher complexity and controlled by some exponential inequality (using for instance McDiarmid inequality), even for very large classes of functions with infinite Vapnik dimension [6, 5]. However extending such results in the Markovian case are far from begin direct at least for two reasons: first, the functionals on blocks are not bounded, second, the blocks themselves in the general Harris recurrent case may be unknown (depending on the true transition kernel of the chain) and should be estimated as done in [11] with pseudo-regeneration techniques. This will be the subject of future researches.*

3.3. Block VC classes

Even if the blocks $(B_k)_{k\geq 1}$ form an independent sequence, we cannot apply directly concentration results for empirical processes over bounded classes, e.g., Theorem 1, simply because the class of functions formed by the f' is not bounded. To solve this problem we will show that it is possible by an adequate probability transformation to bound the covering number of the f' functions by the one of the original class $\mathcal{F}$. In particular, we show that the class formed by the f' functions has a similar size, in terms of covering number, as the class $\mathcal{F}$. This in turn will help to extend existing concentration inequalities on $\mathcal{F}$ for i.i.d. sequences to concentration inequalities on $\mathcal{F}'$ for i.i.d. sequence of blocks.

Recall that E denotes for the state space of X. Define $E' = \cup_{k=1}^{\infty} E^k$ and let the *occupation measure* M be given by

$$M(B, dy) = \sum_{x \in B} \delta_x(y), \qquad \text{for every } B \in E'.$$

The function that gives the size of the blocks ℓ is $\ell : E' \to \mathbb{N}^*$, defined by,

$$\ell(B) = \int M(B, \mathrm{d}y), \qquad \text{for every } B \in E'.$$

Let $\mathcal{E}'$ denote the smallest σ-algebra formed by the elements of the σ-algebras $\mathcal{E}^k$, $k \geq 1$, where $\mathcal{E}^k$ stands for the classical product σ-algebra. Let Q' denote a probability measure on $(E', \mathcal{E}')$. If $B(\omega)$ is a random variable with distribution Q', then $M(B(\omega), \mathrm{d}y)$ is a random measure, i.e., $M(B(\omega), \mathrm{d}y)$ is a (counting) measure on $(E, \mathcal{E})$, almost surely, and for every $A \in \mathcal{E}$, $M(B(\omega), A) = \int_A M(B(\omega), \mathrm{d}y)$ is a measurable random variable (valued in $\mathbb{N}$). Henceforth $\ell(B(\omega)) \times \int f(y)\, M(B(\omega), \mathrm{d}y)$ is a random variable and, provided that $Q'(\ell^2) < \infty$, the map Q, defined by

$$Q(A) = E_{Q'}\left(\ell(B) \times \int_A M(B, \mathrm{d}y)\right) / E_{Q'}(\ell^2), \qquad \text{for every } A \in \mathcal{E}, \tag{3.2}$$

is a probability measure on $(E, \mathcal{E})$. The notation E_Q stands for the expectation with respect to the underlying measure Q. Introduce the following notations: for any function $f : E \to \mathbb{R}$, let $f' : E' \to \mathbb{R}$ be given by

$$f'(B) = \int f(y)\, M(B, \mathrm{d}y) = \sum_{x \in B} f(x),$$

and for any class $\mathcal{F}$ of real-valued functions defined on E, denote by $\mathcal{F}' = \{f' : f \in \mathcal{F}\}$.

Lemma 2. *Let Q' be a probability measure on $(E', \mathcal{E}')$ such that $0 < \|\ell\|_{L_2(Q')} < \infty$ and $\mathcal{F}$ be a class of measurable real-valued functions defined on $(E, \mathcal{E})$. Then we have, for every $0 < \epsilon < \infty$,*

$$\mathcal{N}(\epsilon \|\ell\|_{L_2(Q')},\, \mathcal{F}',\, L_2(Q')) \leq \mathcal{N}(\epsilon,\, \mathcal{F},\, L_2(Q)),$$

where Q is given in (3.2). Moreover if $\mathcal{F}$ is VC with constant envelope U and characteristic (C, v), then $\mathcal{F}'$ is VC with envelope $U\ell$ and characteristic (C, v).

Proof. The proof is inspired from the proof of Lemma 4.2 presented in [35]. Let $f' \in \mathcal{F}'$, i.e., there exists $f \in \mathcal{F}$ such that $f'(B) = \int f(y)\, M(B, \mathrm{d}y)$. Then, using Jensen's inequality,

$$\begin{aligned} E_{Q'}(f'^2) &= \mathbb{E}_{Q'}\left(\left(\int f(y)\, M(B, dy)\right)^2\right) \\ &\leq \mathbb{E}_{Q'}\left(\ell(B)\left(\int f(y)^2\, M(B, dy)\right)\right) \\ &= E_Q(f^2) E_{Q'}(\ell^2). \end{aligned}$$

Applying this to the function

$$f'(B) - f'_k(B) = \int (f(y) - f_k(y))\, M(B, dy),$$

when each f_k is the center of an ϵ-cover of the space $\mathcal{F}$ and $\|f - f_k\|_{L_2(Q)} \leq \epsilon$ gives the first assertion of the lemma. To obtain the second assertion, note that $F' = U\ell$ is an envelope for $\mathcal{F}'$. In addition, we have that

$$\|F'\|_{L_2(Q')} = U\|\ell\|_{L_2(Q')}.$$

From this we derive that, for every $0 < \epsilon < 1$,

$$\mathcal{N}(\epsilon\|F'\|_{L_2(Q')},\, \mathcal{F}',\, L_2(Q')) = \mathcal{N}(\epsilon U\|\ell\|_{L_2(Q')},\, \mathcal{F}',\, L_2(Q')).$$

Then using the first assertion of the lemma, we obtain for every $0 < \epsilon < 1$,

$$\mathcal{N}(\epsilon\|F'\|_{L_2(Q')},\, \mathcal{F}',\, L_2(Q')) \leq \mathcal{N}(\epsilon U,\, \mathcal{F},\, L_2(Q)),$$

which implies the second assertion whenever the class $\mathcal{F}$ is VC for the envelope F. □

Now that we know that any bounded VC class $\mathcal{F}$ can be extended to a VC class $\mathcal{F}'$ unbounded defined over the blocks, we consider the bounded case $\mathcal{F}'1_{\{\ell\leq L\}} = \{f'1_{\{\ell\leq L\}} : f \in \mathcal{F}\}$ which, unsurprisingly, is shown to remain VC.

Lemma 3. *Let Q' be a probability measure on $(E', \mathcal{E}')$ and $\mathcal{F}$ be a class of measurable real-valued functions defined on $(E, \mathcal{E})$. Then we have, for every $0 < \epsilon < \infty$,*

$$\mathcal{N}(\epsilon L,\, \mathcal{F}'1_{\{\ell\leq L\}},\, L_2(Q')) \leq \mathcal{N}(\epsilon,\, \mathcal{F},\, L_2(\tilde{Q})),$$

where $\tilde{Q} = E_{Q'}\left(\ell(B)1_{\{\ell(B)\leq L\}} \times \int_A M(B, \mathrm{d}y)\right) / E_{Q'}(\ell(B)^2 1_{\{\ell(B)\leq L\}})$. Moreover if $\mathcal{F}$ is VC with constant envelope U and characteristic (C, v), then $\mathcal{F}'1_{\{\ell\leq L\}}$ is VC with envelope LU and characteristic (C, v).

Proof. The proof follows the same lines as the proof of Lemma 2, replacing ℓ by $\ell 1_{\{\ell\leq L\}}$. □

4. Main results

The main results of the paper are now stated. They extend concentration inequalities for empirical processes over independent random variables [22, 23, 24], e.g., Theorem 1, to Markov chains. We shall distinguish between two assumptions on the regeneration time τ_A. We say that τ_A has polynomial moments, whenever

(PM) there exists $p > 1$ such that $\mathbb{E}_A[\tau_A^p] < \infty$, and that τ_A has some exponential moments, as soon as

(EM) there exists $\lambda > 0$ such that $\mathbb{E}_A[\exp(\tau_A \lambda)] < \infty$.

Theorem 4 (regenerative block Rademacher complexity). *Assume that the chain X satisfies the generic hypothesis (H). Let $\mathcal{F}$ be VC with constant envelope U and characteristic (C, v). Let σ'^2 be such that*

$$\mathbb{E}_A\left[\left(\sum_{i=1}^{\tau_A} f(X_i)\right)^2\right] \leq \sigma'^2, \qquad \text{for all } f \in \mathcal{F}.$$

For some universal constant $M > 0$, and any L such that $0 < \sigma' \leq LU$,

(i) if (PM) holds, then

$$R_{n,B}(\mathcal{F}) \leq M\left[vLU \log\frac{CLU}{\sigma'} + \sqrt{vn\sigma'^2 \log\frac{CLU}{\sigma'}}\right] + \frac{nU\mathbb{E}_A[\tau_A^p]}{L^{p-1}},$$

(ii) if (EM) holds, then

$$R_{n,B}(\mathcal{F}) \leq M\left[vLU \log\frac{CLU}{\sigma'} + \sqrt{vn\sigma'^2 \log\frac{CLU}{\sigma'}}\right] + nU\exp(-L\lambda/2)C_\lambda,$$

where $C_\lambda = 2\mathbb{E}_A[\exp(\tau_A\lambda)]/\lambda$.

Proof. First we show that, regardless of (i) and (ii),

$$R_{n,B}(\mathcal{F}) \leq M\left[vLU \log\frac{CLU}{\sigma'} + \sqrt{vn\sigma'^2 \log\frac{CLU}{\sigma'}}\right] + nU\mathbb{E}_A[\tau_A 1_{\{\tau_A > L\}}], \tag{4.1}$$

for some universal constant $M > 0$. Then we consider the two cases (i) and (ii) to bound $\mathbb{E}_A[\tau_A 1_{\{\tau_A > L\}}]$ accordingly.
Use the decomposition

$$\sum_{k=1}^n \epsilon_k f'(B_k) = \sum_{k=1}^n \epsilon_k \underline{f}'_L(B_k) + \sum_{k=1}^n \epsilon_k \overline{f}'_L(B_k), \tag{4.2}$$

where, for any $B \in E'$,

$$\underline{f}'_L(B) = f'(B)1_{\{\ell(B) \leq L\}} \qquad \text{and} \qquad \overline{f}'_L(B) = f'(B)1_{\{\ell(B) > L\}}.$$

The first term in (4.2) represents a classical Rademacher complexity over a bounded class: $\mathcal{F}'1_{\{\ell(B)\leq L\}}$. It follows from Lemma 3 that the product class $\mathcal{F}'1_{\{\ell(B)\leq L\}}$ is VC with constant envelop LU. As by assumption, $0 < \sigma' \leq LU$, we deduce from applying Theorem 1 (with LU in place of U), that

$$\mathbb{E}_A \sup_{f \in \mathcal{F}} \left| \sum_{k=1}^{n} \epsilon_k \, \underline{f}'_L(B_k) \right| \leq M \left[vLU \log \frac{CLU}{\sigma'} + \sqrt{vn\sigma'^2 \log \frac{CLU}{\sigma'}} \right].$$

For the second term in (4.2), we find

$$\mathbb{E}_A \sup_{f \in \mathcal{F}} \left| \sum_{k=1}^{n} \epsilon_k \overline{f}'_n(B_k) \right| \leq nU\mathbb{E}_A[\ell(B_1)1_{\{\ell(B_1)>L\}}] = nU\mathbb{E}_A[\tau_A 1_{\{\tau_A > L\}}]$$

Hence (4.1) is established. To obtain point (i), simply use Markov's inequality. To obtain point (ii), note that

$$\mathbb{E}_A[\tau_A 1_{\{\tau_A>L\}}] \leq \exp(-L\lambda/2)\mathbb{E}_A[\tau_A \exp(\tau_A\lambda/2)] \leq \frac{2}{\lambda}\exp(-L\lambda/2)\mathbb{E}_A[\exp(\tau_A\lambda)],$$

where the last inequality follows from $t \leq \exp(t)$ with $t = \tau_A\lambda/2$. □

Remark 5 (geometric ergodicity and (EM)). *Condition (EM) is equivalent to each of the following assertions: (i) the geometric ergodicity of the chain X, (ii) the (uniform) Doeblin condition, (iii) the Foster-Lyapunov drift condition (see Theorem 16.0.2 in [40] for the details). Under this assumption, most classical convergence results (for instance, the law of the iterated logarithm or the central limit theorem) are valid [40, Chapter 17].*

Remark 6 (mixing and (PM)). *We point out that the relationship between (PM) and the rate of decay of mixing coefficients has been investigated in Bolthausen (1982): (PM) is typically fulfilled as soon as the strong mixing coefficients sequence decreases as an arithmetic rate n^{-s}, for some $s > p-1$.*

The two following results show that the regenerative block Rademacher complexity, previously introduced, is useful to control the expected values as well as the excess probability of suprema over classes of functions.

Theorem 5 (expectation bound). *Assume that the chain X satisfies the generic hypothesis (H). Let $\mathcal{F}$ be a countable class of measurable functions bounded by U. It holds that*

$$\begin{aligned}\mathbb{E}_\nu \left[\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} (f(X_i) - \mathbb{E}_\pi[f]) \right| \right] &\leq 4R_{n,B}(\mathcal{F}_c) + 2U(\mathbb{E}_\nu[\tau_A] + \mathbb{E}_A[\tau_A]) \\ &\leq 16R_{n,B}(\mathcal{F}) + 2U(\mathbb{E}_\nu[\tau_A] + \mathbb{E}_A[\tau_A]),\end{aligned}$$

where $\mathcal{F}_c$ denote the class formed by $\{f - \mathbb{E}_\pi[f],\ f \in \mathcal{F}\}$ and ν stands for the initial measure.

Proof. Start by establishing the first inequality. We rely on the block decomposition given in (3.1). First, we apply Lemma 1.2.6 in [17] (or Theorem 3.1.21 in [26]) to treat the term formed by complete blocks. We obtain

$$\begin{aligned}
\mathbb{E}_\nu\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=\tau_A(1)+1}^{\tau_A(l_n)}(f(X_i)-\mathbb{E}_\pi[f])\right|\right] &\le \mathbb{E}_A\left[\max_{1\le l\le n}\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{l_n-1}\{f'(B_k)-\ell(B_k)\mathbb{E}_\pi[f]\}\right|\right]\\
&\le 4\mathbb{E}_A\left[\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}\epsilon_k\{f'(B_k)-\ell(B_k)\mathbb{E}_\pi[f]\}\right|\right]\\
&= 4R_{n,B}(\mathcal{F}_c).
\end{aligned}$$

The terms corresponding to incomplete blocks are treated as follows. We have

$$\mathbb{E}_\nu\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{\tau_A(1)}(f(X_i)-\mathbb{E}_\pi[f])\right|\le 2U\mathbb{E}_\nu[\tau_A],$$
$$\mathbb{E}_\nu\sup_{f\in\mathcal{F}}\left|\sum_{i=\tau_A(l_n)}^{n}(f(X_i)-\mathbb{E}_\pi[f])\right|\le 2U\mathbb{E}_A[\tau_A].$$

To obtain the second inequality, apply Theorem 3.1.21 in [26]. □

Using Theorem 5, we now apply [1, Theorem 7] to obtain a concentration bound for the empirical process involving the Rademacher complexity $R_{n,B}(\mathcal{F})$ and a variance term.

Theorem 6 (concentration bound, [1])**.** *Assume that the chain X satisfies the generic hypothesis (H), (EM) and there exists $\lambda>0$ such that $\mathbb{E}_\nu[\exp(\lambda\tau_A)]<\infty$. Let $\mathcal{F}$ be a countable class of measurable functions bounded by U. Let*

$$R_n\ge 16R_{n,B}(\mathcal{F})+2U(\mathbb{E}_\nu[\tau_A]+\mathbb{E}_A[\tau_A]),$$
$$\sigma'^2\ge\sup_{f\in\mathcal{F}}\mathbb{E}_A\left[\left(\sum_{i=1}^{\tau_A}f(X_i)\right)^2\right].$$

Then, for some universal constant $K>0$, and for $\tau>0$ depending on the tails of the regeneration time, we have, for all $t\ge1$,

$$\mathbb{P}_\nu\left(\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}(f(X_i)-\mathbb{E}_\pi(f))\right|\ge t+KR_n\right)\le K\exp\left[-\frac{\mathbb{E}_A[\tau_A]}{K}\min\left(\frac{t^2}{n\sigma'^2},\frac{t}{\tau^3Ulog\ n}\right)\right],$$

yielding alternatively, that for any $n/\log(n)\ge\tau^3U/\sigma'^2$ with probability $1-\delta$ we have,

$$\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}(f(X_i)-\mathbb{E}_\pi(f))\right|\le KR_n+\max\left(\sqrt{n}\sigma'\sqrt{K\log\left(\frac{K}{\delta}\right)},\log\left(\frac{K}{\delta}\right)\frac{\tau^3U\log(n)}{\mathbb{E}_A[\tau_A]}\right).$$

Remark 7 (on Theorem 6)**.** *An explicit value for the constant K is difficult to obtain from the results of [1] but would be of great interest in practical applications. Notice that for n large the second member of the inequality reduces to the bound $KR_n + \sqrt{n}\sigma'\sqrt{K\log(K/\delta)}$, which is similar to the rate in the i.i.d. case.*

Remark 8. *Paulin has obtained in [48] powerful concentration inequalities for uniformly ergodic Markov chains using Marton coupling techniques. An interesting feature of these inequalities is that the constants can be made explicit as a function of the mixing time (see his definition 1.3). His main result is a McDiarmid type inequality (theorem 2.1 and corollary 2.10) which may be applied to empirical processes. It is easy to see how to combine our results with [48] to get bounds on empirical processes depending on the mixing time. In this paper we are rather interested in exponential control depending on a variance term (which may be small) as in our application to kernel density estimation below. Moreover, notice that the results in [1] (which also make use of regeneration techniques) holds for more general chain, that may not be uniformly geometrically ergodic.*

Remark 9 (m different from 1)**.** *We have reduced our analysis to the case $m = 1$, however it is very easy to see now how the case $m > 1$ can be handled up to a modified constant in the bound. Recall that when $m > 1$ then the blocks $(B_k)_{k\geq 1}$ are 1-dependent (see for instance [16] Corollary 2.3). We can split the sum as follows*

$$\sum_{k=1}^{l_n-1} f(B_k) = \sum_{k=1,\, k\ even}^{l_n-1} f(B_k) + \sum_{k=1,\, k\ odd}^{l_n-1} f(B_k).$$

Then notice that, because of the 1*-dependence property, in each sums the blocks are independent and we now have two sums of at most $n/2$ independent blocks that can be treated separately based on the presented results.*

5. Applications

In this section, we consider two applications of our results. The first one is dealing with a particular class of kernel functions (useful in nonparametric estimation) and the second one is focusing on some Markov chains called Metropolis-Hasting chains (useful in simulation methods). All the proofs of the section are postponed to the Appendix.

5.1. Kernel density estimator

Given $n \geq 1$ observations of a Markov chain $X \subset \mathbb{R}^d$ satisfying the generic assumption (H), the kernel density estimator of the stationary measure π is given by

$$\hat{\pi}_n(x) = (nh_n^d)^{-1} \sum_{i=1}^{n} K((x - X_i)/h_n),$$

where $K : \mathbb{R}^d \to \mathbb{R}$, called kernel, is such that $\int K(x)dx = 1$ and $(h_n)_{n\geq 1}$ is a positive sequence of bandwidths.

The analysis of the asymptotic behaviour of $\hat{\pi}_n - \pi$ is traditionally executed by studying two terms: the bias term, $\mathbb{E}_\nu \hat{\pi}_n - \pi$, which can be handled using functional analysis [25, section 4.1.1] and the variance term, $\hat{\pi}_n - \mathbb{E}_\nu \hat{\pi}_n$, which follows from empirical process theory (for independent random variables). In the next, we provide some results on the asymptotic behaviour of the variance term.

We shall consider kernel functions $K : \mathbb{R}^d \to \mathbb{R}$ taking one of the two following forms,

$$(i) \quad K(x) = K^{(0)}(|x|), \qquad \text{or} \qquad (ii) \quad K(x) = \prod_{k=1}^{d} K^{(0)}(x_k), \tag{5.1}$$

where $K^{(0)}$ is a bounded function of bounded variation with support $[-1, 1]$. Note that more general (but less simple) conditions on K [24, Assumption (K_1)] could have been used in place of (5.1). From [44], the class of functions

$$\mathcal{K} = \{y \mapsto K((x-y)/h) \,:\, h > 0,\, x \in \mathbb{R}^d\} \quad \text{is a uniformly bounded VC class.}$$

Theorem 7. *Assume that the chain $X \subset \mathbb{R}^d$ satisfies the generic hypothesis (H) ,the stationary density π is supposed to be bounded, the kernel K is given by (5.1) and $K(x) \leq U$, for all $x \in \mathbb{R}^d$. Suppose that $h_n \to 0$ and there exists $\beta > 0$ such that $h_n \geq n^{-\beta}$.*

(i) If (PM) holds for $p > 2$ and $0 < \beta(p/(p-1)) < 1/d$, we have

$$\mathbb{E}_\nu \left[\sup_{x \in \mathbb{R}^d} |\hat{\pi}_n(x) - \mathbb{E}_\pi[\hat{\pi}_n(x)]| \right] = O\left(\sqrt{\frac{\log{(n)}}{n h_n^{dp/(p-1)}}} \right).$$

(ii) If (EM) holds and $0 < \beta < 1/d$, we have

$$\mathbb{E}_\nu \left[\sup_{x \in \mathbb{R}^d} |\hat{\pi}_n(x) - \mathbb{E}_\pi[\hat{\pi}_n(x)]| \right] = O\left(\sqrt{\frac{\log(n)^2}{n h_n^{d}}} \right).$$

Comparing the rate given in Theorem 7 with the usual rate $\sqrt{\log(n)/(nh_n^d)}$ corresponding to the independent case [21, 24], we see that the rate obtained for the Markovian setting is slightly poorer. Even when the regeneration time has exponential moments, a loss of a factor $\log(n)^{1/2}$ is observed with respect to the independent case. This loss is due to the variance term that scales differently due to the block size. To fill this gap, we provide in the following theorem an additional assumption on the chain X that ensures the same rate as in the independent case.

Theorem 8. *Assume that the chain $X \subset \mathbb{R}^d$ satisfies the generic hypothesis (H) and (EM), the stationary density π is supposed to be bounded, the kernel K is given by (5.1)* *and $K(x) \leq U$, for all $x \in \mathbb{R}^d$. Suppose that $h_n \to 0$ and that $nh_n^d/|\log(h_n)| \to +\infty$, if there exist $p > 2$ and $C > 0$ such that for all $x \in E$, $\pi(x)\mathbb{E}_x[\tau_A^p] \leq C$, then we have*

$$\mathbb{E}_\nu \left[\sup_{x \in \mathbb{R}^d} |\hat{\pi}_n(x) - \mathbb{E}_\pi[\hat{\pi}_n(x)]| \right] = O\left(\sqrt{\frac{|\log(h_n)|}{nh_n^d}} \right).$$

Remark 10 (on the bandwidth)**.** *In the independent case, given $x \in \mathbb{R}$, the variance of $\hat{\pi}_n(x)$ is ensured to vanish whenever $nh_n^d \to +\infty$, and asking for $nh_n^d/|\log(h_n)| \to +\infty$ is a slight additional requirement to guarantee the convergence to hold uniformly over $\mathbb{R}$. In Theorem 8, the assumptions on the bandwidth are the same as in the independent case whereas in Theorem 7, the fact that $h_n \geq n^{-\beta}$ is slightly stronger.*

Remark 11 (on the additional assumption on X)**.** *The additional assumption, namely, for all $x \in E$, $\pi(x)\mathbb{E}_x[\tau_A^p] \leq C$, can be understood as a tail condition. In fact when $x \mapsto \pi(x)\mathbb{E}_x[\tau_A^p]$ is continuous, this condition reduces to $\lim_{\|x\|\to\infty} \pi(x)\mathbb{E}_x[\tau_A^p]$ exists. In other words, the return time when departing from x should not increase faster than the decrease of $\pi(x)$ when $x \to \infty$.*

5.2. Metropolis-Hasting algorithm

Bayesian estimation requires to compute moments of the so called *posterior distribution* whose probability density function π is given by

$$\pi(\theta) = \frac{\mathcal{L}(\theta)}{\int \mathcal{L}(\theta) d\theta} \qquad \theta \in \mathbb{R}^d,$$

where $\mathcal{L}$ is a positive function which stands for the likelihood of the observed data. The (unknown) quantities of interest writes as $\int g d\pi$, for some given measurable functions $g : \mathbb{R}^d \to \mathbb{R}$. A particular feature in this framework is that the integral at the denominator of π is unknown and difficult to compute making impossible to generate observations directly from π. Markov chains Monte Carlo (MCMC) methods aim to produce samples $X_1, \ldots, X_n$ in $\mathbb{R}^d$ that are approximately distributed according to π. Then $\int g d\pi$ is classically approximated by the empirical average over the chain :

$$n^{-1} \sum_{i=1}^n g(X_i).$$

For inference, Bayesian credible intervals are usually computed using the quantiles the coordinate chains (see below). We refer to [53] for a complete description of MCMC methods. In what follows, we focus on the special MCMC method called Metropolis-Hasting (MH). Aim is to derive new concentration inequalities for suprema of $\sum_{i=1}^n g(X_i)$ over g in some VC classes, and to apply these results to measure the accuracy of Bayesian credible intervals.

Let us introduce the MH algorithm with target density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ and proposal $Q(x, dy) = q(x, y)dy$, where q is a positive function defined on $\mathbb{R}^d \times \mathbb{R}^d$ satisfying $\int q(x, y)dy = 1$. Define for any $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\rho(x, y) = \begin{cases} \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) & \text{if } \pi(x)q(x, y) > 0, \\ 1 & \text{if } \pi(x)q(x, y) = 0. \end{cases}$$

The MH chain starts at $X_0 \sim \nu$ and moves from X_n to X_{n+1} according to the following rule:

(i) Generate

$$Y \sim Q(X_n, dy) \qquad \text{and} \qquad W \sim \mathcal{B}(\rho(X_n, Y)).$$

(ii) Set

$$X_{n+1} = \begin{cases} Y & \text{if} \quad W = 1, \\ X_n & \text{if} \quad W = 0. \end{cases}$$

In the particular case that $q(x, y) = q_0(x - y)$, the previous algorithm is refereed to as the random walk MH.

The asymptotic behavior of the random walk MH chain has been studied in [54, 30] where central limit theorems are established based on the geometric ergodicity of the chain. From Remark 5, the results in [54, 30] imply that (EM) is satisfied. This allows to apply Theorem 4 almost directly for the random walk MH. For the sake of completeness, we provide the following alternative development, in which we verify (EM) via the (uniform) Doeblin condition. Contrasting with [54, 30], we focus on π with bounded support.

Denote by $B(x, \epsilon)$ (resp. $B(\epsilon)$) the open ball with centre x (resp. 0) and radius ϵ with respect to the Euclidean norm $\|\cdot\|$. We consider the following ball condition on the proposal q_0 associated to the random walk MH.

(BC) Let π be a bounded probability density supported by $E \subset \mathbb{R}^d$, a bounded and convex set with non-empty interior. Suppose that there exists $b > 0$ and $\epsilon > 0$ such that $\forall x \in \mathbb{R}^d \times \mathbb{R}^d$, $q_0(x) \geq b 1_{B(\epsilon)}(x)$.

Proposition 9. *Under (BC), the random walk MH chain satisfies (H) and (EM).*

Based on Proposition 9, we are in position to apply point (ii) of Proposition 4 to the random walk MH.

Proposition 10. *Let $\mathcal{G}$ be a countable VC class of measurable functions on S bounded by U with characteristics (C, v). Under (BC), for all $n \geq 1$, it holds that*

$$\mathbb{E}_\nu \left[\sup_{g \in \mathcal{G}} \left| \sum_{i=1}^n (g(X_i) - \pi(g)) \right| \right] \leq D\sqrt{n(1 \vee \log(\log(n)))},$$

where D depends only on (v, C, U) and on the tails of the regeneration time. Moreover, for any $n/\log(n) \geq \tau^3/(U\mathbb{E}_A[\tau_A^2])$, we have with probability $1-\delta$,

$$\sup_{g\in\mathcal{G}} \left| \sum_{i=1}^{n} (g(X_i) - \pi(g)) \right| \leq$$
$$KD\sqrt{n(1 \vee \log(\log(n)))} + \max\left(\sqrt{nU^2\mathbb{E}_A[\tau_A^2]K\log\left(\frac{K}{\delta}\right)}, \log\left(\frac{K}{\delta}\right)\frac{\tau^3 U\log(n)}{\mathbb{E}_A[\tau_A]} \right),$$

where $K > 0$ is a universal constant.

Let $k \in \{1, \dots, d\}$ and denote by $X_i^{(k)}$ the k-th coordinate of X_i. Define the associated empirical cumulative distribution function for any $t \in \mathbb{R}$,

$$\hat{\Pi}_k(t) = n^{-1}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i^{(k)}).$$

and the quantile function, for any $u \in (0, 1)$,

$$\hat{Q}_k(u) = \inf\{x \in \mathbb{R} \,:\, \hat{\Pi}_k(x) \geq u\}.$$

As a corollary of the previous result, we obtain an upper bound for the estimation error of Bayesian credible intervals defined as $[\hat{Q}_k(u), \hat{Q}_k(1-u)]$, for $k = 1, \dots, d$. The targeted interval is $[Q_k(u), Q_k(1-u)]$, where Q_k is the true quantile function of the posterior marginal distribution Π_k whose associated density is denoted by π_k.

Proposition 11. *Under (BC), for all $0 < \gamma < 1/4$ and $k \in \{1, \dots, p\}$,*

$$\sup_{u\in[\gamma,1-\gamma]} \left|\hat{Q}_k(u) - Q_k(u)\right| = O_{\mathbb{P}_\nu}\left(\sqrt{\frac{\log\log n}{n}}\right).$$

Remark 12. *In contrast with the study of kernel density estimator given in section 5.1, the approach taken in this section cannot take advantage of classes with small variance (e.g., going to 0 with n). In particular, in Theorem 10, the variance over the class $\mathcal{G}$ has been crudely bounded by $U\mathbb{E}_A[\tau_A^2]$. This is in line with the functions of interest, $1_{(-\infty,t]}$, in Proposition 11. The stationary variance is given by $\Pi(t)(1-\Pi(t))$ whose maximum is $1/4$.*

Appendix A Proofs of the results of section 5.1

A.1 Proof of Proposition 7

In virtue of Theorem 5, it suffices to provide for both cases a sufficiently tight bound on $R_{n,B}(\mathcal{K}_n)$ with $\mathcal{K}_n = \{y \mapsto K((x-y)/h_n) \,:\, x \in \mathbb{R}^d\}$, which in virtue of [44] is included in $\mathcal{K}$ a VC class of functions. First we consider (i). By Jensen inequality we have

$$\left(\frac{1}{\ell(B_1)}\sum_{X_i\in B_1} K((x-X_i)/h_n)\right)^2 \leq \frac{1}{\ell(B_1)}\sum_{X_i\in B_1} K((x-X_i)/h_n)^2$$

and for any $\tilde{L}$, it holds that

$$\begin{aligned}
\sigma'^2 &= \mathbb{E}_A\left[\left(\sum_{X_i\in B_1} K((x-X_i)/h_n)\right)^2\right]\\
&\leq \mathbb{E}_A\left[\ell(B_1)\sum_{X_i\in B_1} K((x-X_i)/h_n)^2\right]\\
&\leq \tilde{L}\mathbb{E}_A[\ell(B_1)]\mathbb{E}_\pi\left[K((x-X)/h_n)^2\right] + U^2\mathbb{E}_A\left[\ell(B_1)^2 1_{\{\ell(B_1)>\tilde{L}\}}\right].
\end{aligned}$$

Use Markov inequality and the expression of Pitman's occupation measure to get

$$\sigma'^2 \leq \tilde{L}\mathbb{E}_A[\tau_A]h_n^d\|\pi\|_\infty v_K + \frac{U^2\mathbb{E}_A\left[\tau_A^p\right]}{\tilde{L}^{p-2}},$$

where $v_K = \int K(u)^2du$ and $\|\pi\|_\infty = \sup_{x\in\mathbb{R}^d}|\pi(x)|$. Equilibrating between the first and second term gives $\tilde{L} = h_n^{-d/(p-1)}$, which implies that $\sigma'^2 \leq A_1 h_n^r$, $A_1 > 0$, $r = d(p-2)/(p-1)$. Applying Theorem 4, we get

$$\begin{aligned}
&R_{n,B}(\mathcal{K}_n)\\
&\leq A_2\left(\frac{L\log(CLUA_1h_n^{-r})}{nh_n^d} + \sqrt{\frac{h_n^r\log(CLUA_1h_n^{-r})}{nh_n^{2d}}} + \frac{1}{L^{p-1}h_n^d}\right),\\
&= A_2\left(\frac{L\log(CLUA_1h_n^{-r})}{nh_n^d} + \sqrt{\frac{\log(CLUA_1h_n^{-r})}{nh_n^{dp/(p-1)}}} + \frac{1}{L^{p-1}h_n^d}\right),
\end{aligned}$$

with $A_2 > 0$. Choose L by (almost) equilibrating the first and last term of the preceding decomposition

$$L_n = \left(\frac{n}{\log(\alpha_n)}\right)^{1/p},$$

with $\alpha_n = n^{1/p}h_n^{-r}$. Take n large enough to get that

$$\begin{aligned}
R_{n,B}(\mathcal{K}_n) &\leq A_2\sqrt{\frac{\log(CUA_1\alpha_n/(\log\alpha_n)^{1/p})}{nh_n^{dp/(p-1)}}} + 2A_2\left(\frac{\log(\alpha_n)}{nh_n^{dp/(p-1)}}\right)^{(p-1)/p}\\
&= (1+o(1))\sqrt{\frac{\log(\alpha_n)}{nh_n^{dp/(p-1)}}},
\end{aligned}$$

where the last equality is because $p > 2$. Because α_n is smaller than some power of n and using Theorem 5 leads to the result.

In the second case, namely (ii), a similar bound is valid for σ'^2, we have

$$\begin{aligned}\sigma'^2 &\le \tilde{L}\mathbb{E}_A[\tau_A] h_n^d \|\pi\|_\infty v_K + U^2 \mathbb{E}_A\left[\tau_A^2 \exp(\lambda\tau_A/2)\right]\exp(-\lambda\tilde{L}/2)\\ &\le \tilde{L}\mathbb{E}_A[\tau_A] h_n^d \|\pi\|_\infty v_K + 8\left(\frac{U}{\lambda}\right)^2 \mathbb{E}_A\left[\exp(\tau_A\lambda)\right]\exp(-\lambda\tilde{L}/2),\end{aligned}$$

where for the second inequality, we use that $t^2 \le 2\exp(t)$ with $t = \lambda\tau_A/2$. Taking $\tilde{L} = 2\log(h_n^{-d})/\lambda$ gives $\sigma'^2 \le A_3 h_n^d \log(h_n^{-d})$, $A_3 > 0$. Then using (ii) in Theorem 4, we find, for n large enough,

$$R_{n,B}(\mathcal{K}_n) \le A_2 \times \\ \left(\frac{L\log(CLUA_3^{-1}h_n^{-d}/\log(h_n^{-d}))}{nh_n^d} + \sqrt{\frac{\log(h_n^{-d})\log(CLUA_3^{-1}h_n^{-d}/\log(h_n^{-d}))}{nh_n^d}} + \frac{\exp(-L\lambda/2)}{h_n^d}\right).$$

Choosing $L_n = 2\log(n)/\lambda$, noticing that the term in the middle is the leading term and applying Theorem 5, we obtain the stated result.

□

A.2 Proof of Proposition 8

The main step is to show that there exists $c > 0$ such that

$$\mathbb{E}_A\left[\left(\sum_{i=1}^{\tau_A} K((x - X_i)/h_n)\right)^2\right] \le ch_n^d, \qquad \text{for all } x \in E, \tag{A.1}$$

then the conclusion will follow straightforwardly. The fact that (A.1) holds true follows from Lemma ∗A.3 in the supplementary file of [4], which gives that, for any measurable function f,

$$\mathbb{E}_A\left[\left(\sum_{i=1}^{\tau_A} f(X_i)\right)^2\right] \le A_4(\pi(f^2) + \mathbb{E}_\pi[f(X_0)^2\tau_A^p]),$$

with $A_4 > 0$. Whenever $f(X) = K((x - X)/h_n)$, we get that $\pi(f^2) \le v_K\|\pi\|_\infty h_n^d$, and defining $g(x) = \pi(x)\mathbb{E}_x[\tau_A^p]$ we get

$$\mathbb{E}_\pi[f(X_0)^2\tau_A^p] = \int K((x-y)/h_n)g(y)dy \le Cv_K h_n^d.$$

Hence we have obtained (A.1). It follows from Theorem 4 that

$$R_{n,B}(\mathcal{K}_n) \le A_2\left(\frac{L\log(CLUc^{-1}h_n^{-d})}{nh_n^d} + \sqrt{\frac{\log(CLUc^{-1}h_n^{-d})}{nh_n^d}} + \frac{\exp(-L\lambda/2)}{h_n^d}\right). \tag{A.2}$$

Setting $L_n = 2\log(n)/\lambda$ we obtain the desired result by applying Theorem 5.

Appendix B Proofs of the results of section 5.2

B.1 Proof of Proposition 9

The proof follows from the following Lemma in which some conditions are given to ensure the uniform Doeblin condition.

Proposition 12. *Let P be a transition kernel. Let Φ be a positive measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Suppose that $E = \operatorname{supp}(\Phi)$ is bounded and convex with non-empty interior. Suppose that there exists $\epsilon > 0$ such that $\forall x \in E$, $P(x, dy) \geq 1_{B(x,\epsilon)}(y)\Phi(dy)$. Then there exists $C > 0$ and $n \geq 1$ such that for any $x \in E$ and any measurable set $A \subset E$,*

$$P^n(x, A) \geq C\Phi(A). \tag{B.1}$$

Proof. We decompose the proof according to 4 steps.
First step: Let $0 < \gamma \leq \eta$. There exists $c > 0$ such that for any $(x, y) \in E \times E$, it holds that

$$\int 1_{B(x,\eta)}(x_1)1_{B(x_1,\gamma)}(y)\,\Phi(dx_1) \geq c1_{B(x,\eta+\gamma/4)}(y). \tag{B.2}$$

To obtain the previous statement, we can restrict our attention to the case when $y \in B(x, \eta+\gamma/4)$. Else the inequality is trivial. Note that there exists a point m lying strictly in the line segment between x and y such that

$$B(m, \gamma/4) \subset \{B(x, \eta) \cap B(y, \gamma)\}.$$

By convexity of E, $m \in E \times E$. Hence

$$\int 1_{B(x,\eta)}(x_1)1_{B(y,\gamma)}(x_1)\Phi(dx_1) \geq \Phi\{B(m, \gamma/4)\} \geq \inf_{m\in E} \Phi\{B(m, \gamma/4)\}.$$

But he function $m \mapsto \Phi(B(m, \gamma/4) \cap E)$ is continuous on E and positive for each $m \in E$, by definition of the support and the fact that m is an interior point of E by convexity.
Second step: We iterate (B.2) to obtain the following statement. For any $n \geq 1$, there exists $C_n > 0$ such that for any $(x, y) \in E$, it holds that

$$\int \ldots \int 1_{B(x,\epsilon)}(x_1)1_{B(x_1,\epsilon)}(x_2)\ldots 1_{B(x_{n-1},\epsilon)}(x_n)1_{B(x_n,\epsilon)}(y)\,\Phi(dx_1)\ldots\Phi(dx_n)$$
$$\geq C_n 1_{B(x,\epsilon(1+n/4))}(y).$$

Third step: Take n such that $\epsilon(1 + n/4) > \sup_{(x,y)\in E} \|x - y\|$. Then for any $x \in E$ and $y \in E$, $y \in B(x, \epsilon(1 + n/4))$. It follows that there exists $C_n > 0$ such that for all $(x, y) \in E$,

$$\int \ldots \int 1_{B(x,\epsilon)}(x_1)1_{B(x_1,\epsilon)}(x_2)\ldots 1_{B(x_{n-1},\epsilon)}(x_n)1_{B(x_n,\epsilon)}(y)\,\Phi(dx_1)\ldots\Phi(dx_n) \geq C_n.$$

Fourth step: Using the last step and the assumption on P, it holds that for any $x \in E$ and any measurable set $A \subset E$,

$$\begin{aligned}
&P^n(x, A) \\
&\geq \int 1_{B(x,\epsilon)}(x_1) \dots 1_{B(x_{n-1},\epsilon)}(x_n) 1_{B(x_n,\epsilon)}(y) 1_{\{y \in A\}} \, \Phi(dx_1) \dots \Phi(dx_n) \Phi(dy) \\
&\geq C_n \Phi(A).
\end{aligned}$$

□

Now we can conclude the proof of Proposition 9. Because $\rho(x, y) \geq \pi(y)/\|\pi\|_\infty$, the Markov kernel P of the MH chain verifies, for any $x \in E$,

$$P(x, dy) \geq \rho(x, y) Q(x, dy) \geq \|\pi\|_\infty^{-1} 1_{B(x,\epsilon)}(y) \pi(y) dy. \tag{B.3}$$

Applying Proposition 12 with $\Phi(dy) = \|\pi\|_\infty^{-1} \pi(y) dy$, we deduce that whenever $\pi(A) > 0$, there exists $n \geq 1$ such that $P^n(x, A) > 0$. This is π-irreducibly. Let $z \in E$. From (B.3), whenever $x \in B(z, \epsilon/2)$ and $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P(x, A) \geq \|\pi\|_\infty^{-1} \pi(A \cap B(z, \epsilon/2)).$$

This means that any ball with positive radius is $\pi_{|B(z,\epsilon/2)}$-small with $m = 1$. From [54, proof of Theorem 2.2], this implies the aperiodicity of the chain. Applying Proposition 12 again with $\Phi(dy) = \|\pi\|_\infty^{-1} \pi(y) dy$, we obtain (B.1) which implies (EM) in virtue of Theorem 16.0.2 in [40]. More precisely, in their Theorem 16.0.2, (B.1) implies point (iv) which is equivalent to point (vii). That is, we have shown that whenever $\psi(B) > 0$, there is $\lambda_B > 0$ such that $\sup_{x \in E} \mathbb{E}_x[\exp(\lambda_B \tau_B)] < \infty$. This is stronger than positive Harris recurrence. Finally, the latter is true with B equal to the atom A (of the extended chain). This means that the moment conditions in (EM) are satisfied.

□

B.2 Proof of Proposition 10

Set $\sigma'^2 = U^2 \mathbb{E}_A[\tau_A^2]$ and apply Theorem 4 to get that

$$R_{n,B}(\mathcal{F}) \leq M \left[vLU \log \frac{CL}{\mathbb{E}_A[\tau_A^2]^{1/2}} + \sqrt{vnU^2 \mathbb{E}_A[\tau_A^2] \log \frac{CL}{\mathbb{E}_A[\tau_A^2]^{1/2}}} \right] + nU \exp(-L\lambda/2) C_\lambda.$$

Take $L = 2\log(n)/\lambda$ to obtain

$$R_{n,B}(\mathcal{F}) \leq M \left[2\log(n) vU \log(A \log(n))/\lambda + \sqrt{vnU^2 \mathbb{E}_A[\tau_A^2] \log(A \log(n))} \right] + UC_\lambda$$

with $A = 2C/(\lambda \mathbb{E}_A[\tau_A^2]^{1/2})$. We obtain the first stated result by straightforward manipulations. The second result is a direct consequence of Theorem 6. □

B.3 Proof of Proposition 11

Let $k \in \{1, \dots, d\}$ and $0 < \gamma < 1/4$. Let $b_{\gamma,k} = \inf_{u \in [\gamma, 1-\gamma]} \pi_k(Q_k(u))$ which by assumption is positive. We will need the following classical lemma describing the behavior of the quantile function based on the associated cumulative distribution function.

Lemma 13. *Let $\gamma < 1/4$. Suppose that F and G are cumulative distribution functions such that G has a density g verifying $b_\gamma = \inf_{u \in [\gamma, 1-\gamma]} g(G^-(u)) > 0$. If $\sup_{t \in \mathbb{R}} |F(t) - G(t)| \leq \gamma$, then $\sup_{u \in [2\gamma, 1-2\gamma]} |F^-(u) - G^-(u)| \leq b_\gamma^{-1} \sup_{t \in \mathbb{R}} |F(t) - G(t)|$.*

Proof. From the mean-value theorem we have that for any $(u, v) \in [\gamma, 1-\gamma]^2$,

$$|G^-(u) - G^-(v)| \leq b_\gamma^{-1} |u - v|. \tag{B.4}$$

Recalling the classical result; see e.g., [51, Lemma 12]; that whenever F and G are two cumulative distribution functions and $\epsilon = \sup_{t \in \mathbb{R}} |F(t) - G(t)|$, then for any $u \in [\epsilon, 1-\epsilon]$,

$$|F^-(u) - G^-(u)| \leq \sup_{|\delta| \leq \epsilon} |G^-(u + \delta) - G^-(u)|.$$

Because $\epsilon \leq \gamma$, it holds that $\gamma \leq u + \delta \leq 1 - \gamma$ whenever $2\gamma \leq u \leq 1 - 2\gamma$. Using (B.4), we get that, for any $u \in [2\gamma, 1-2\gamma]$,

$$|F^-(u) - G^-(u)| \leq b_\gamma^{-1} \epsilon.$$

□

Lemma 13 is applied with $F = \hat{\Pi}_k$, $G = \Pi_k$ and the bound on $\sup_{t \in \mathbb{R}} |\hat{\Pi}_k(t) - \Pi_k(t)|$ is obtained (with high probability) from Proposition 10. In virtue of Example 2.5.4 in [61], we have that

$$\mathcal{N}(\epsilon, \{1_{(-\infty,t]} : t \in \mathbb{R}\}, \|\cdot\|_{L_2(Q)}) \leq 2\epsilon^{-2},$$

which allows us to apply Proposition 10 with $U = 1$, $C = v = 2$. The second bound in Proposition 10 implies that $\sup_{t \in \mathbb{R}} |\hat{\Pi}_k(t) - \Pi_k(t)| \leq \gamma$ holds with probability going to 1 when $n \to \infty$. Hence with probability going to 1, the stated bound holds true.

Acknowledgement The authors are grateful to Gabriela Ciolek for helpful comments on an earlier version of the paper. The authors would like to thank two referees and an associate editor for their insightful comments and references. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electronic Journal of Probability 13*, 1000–1034.

[2] Akritas, M. G. and I. Van Keilegom (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics 28*(3), 549–567.

[3] Athreya, K. B. and P. Ney (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc. 245*, 493–501.

[4] Azaïs, R., B. Delyon, and F. Portier (2018). Integral estimation based on markovian design. *Advances in Applied Probability 50*(3), 833–857.

[5] Bartlett, P. L., O. Bousquet, and S. Mendelson (2005, 08). Local rademacher complexities. *Ann. Statist. 33*(4), 1497–1537.

[6] Bartlett, P. L. and S. Mendelson (2002). Rademacher and gaussian complexities: Risk bounds and structural results. *JMLR 3*, 463–482.

[7] Bednorz, W., K. Latuszynski, and R. Latala (2008). A regeneration proof of the central limit theorem for uniformly ergodic markov chains. *Electronic Communications in Probability 13*, 85–98.

[8] Berbee, H. (1979). *Random walks with stationary increments and renewal theory.* Math. Cent. Tracts, Amsterdam.

[9] Bertail, P. and S. Clémençon (2004). Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. *Probab. Relat. Fields 130*(3), 388–414.

[10] Bertail, P. and S. Clémençon (2010). Sharp bounds for the tails of functionals of Markov chains. *Th. Prob. Appl. 54*(3), 505–515.

[11] Bertail, P. and S. Clémençon (2006). Regenerative block bootstrap for markov chains. *Bernoulli 12*(4), 689–712.

[12] Bolthausen, E. (1980). The Berry-Esseen theorem for functionals of discrete Markov chains. *Z. Wahrsch. Verw. Geb. 54*(1), 59–73.

[13] Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

[14] Bousquet, O., S. Boucheron, and G. Lugosi (2004). Introduction to statistical learning theory. In *Advanced lectures on machine learning*, pp. 169–207. Springer.

[15] Bradley, R. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability survey 2*, 107–144.

[16] Chen, X. (1999). *Limit theorems for functionals of ergodic Markov chains with general state space*, Volume 664. American Mathematical Soc.

[17] de la Peña, V. c. H. and E. Giné (1999). *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

[18] Dedecker, J. and S. Gouëzel (2015). Subgaussian concentration inequalities for geometrically ergodic markov chains. *Electronic Communications in Probability 20*.

[19] Douc, R., A. Guillin, and E. Moulines (2008). Bounds on regeneration times and limit theorems for subgeometric markov chains. *Ann. Inst. H. Poincaré Probab. Statist. 44*(2), 239–257.

[20] Douc, R., E. Moulines, and J. S. Rosenthal (2004). Quantitative bounds on convergence of time-inhomogeneous markov chains. *Annals of Applied Probability*, 1643–1665.

[21] Einmahl, U. and D. M. Mason (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab. 13*(1), 1–37.

[22] Einmahl, U. and D. M. Mason (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist. 33*(3), 1380–1403.

[23] Giné, E. and A. Guillou (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist. 37*(4), 503–522.

[24] Giné, E. and A. Guillou (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist. 38*(6), 907–921. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

[25] Giné, E. and R. Nickl (2008). Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields 141*(3-4), 333–387.

[26] Gin, E. and R. Nickl (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[27] Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive metropolis algorithm. *Bernoulli 7*(2), 223–242.

[28] Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*(03), 726–748.

[29] Jain, J. and B. Jamison (1967). Contributions to Doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb. 8*, 19–40.

[30] Jarner, S. F. and E. Hansen (2000). Geometric ergodicity of metropolis algorithms. *Stochastic processes and their applications 85*(2), 341–361.

[31] Joulin, A. and Y. Ollivier (2010). Curvature, concentration and error estimates for markov chain monte carlo. *The Annals of Probability 38*(6), 2418–2442.

[32] Koltchinskii, V. (2006). Local rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist. 34*(6), 2593–2656.

[33] Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. Lecture Notes in Mathematics. Springer-Verlag Berlin Heidelberg.

[34] Łatuszyński, K., B. Miasojedow, and W. Niemiro (2013). Nonasymptotic bounds on the estimation error of mcmc algorithms. *Bernoulli 19*(5A), 2033–2066.

[35] Levental, S. (1988). Uniform limit theorems for harris recurrent markov chains. *Probability theory and related fields 80*(1), 101–118.

[36] Malinovskiĭ, V. (1987). Limit theorems for Harris Markov chains I. *Theory Prob. Appl. 31*, 269–285.

[37] Malinovskiĭ, V. (1989). Limit theorems for Harris Markov chains II. *Theory Prob. Appl. 34*, 252–265.

[38] McDiarmid, C. (1989). On the method of bounded differences. *Surveys in combinatorics 141*(1), 148–188.

[39] Mengersen, K. L. and R. L. Tweedie (1996). Rates of convergence of the hastings and metropolis algorithms. *The annals of Statistics 24*(1), 101–121.

[40] Meyn, S. and R. L. Tweedie (2009). *Markov chains and stochastic stability* (Second ed.). Cambridge University Press. With a prologue by Peter W. Glynn.

[41] Mohri, M. and A. Rostamizadeh (2010a). Rademacher complexity for non i.i.d. processes. *Advances in Neural Information Processing Systems*, 1097–1104.

[42] Mohri, M. and A. Rostamizadeh (2010b). Stability bounds for stationary ϕ-mixing and β-mixing processes. *Journal of Machine Learning Research 11*, 789–814.

[43] Nickl, R. and J. Shl (2017, 08). Nonparametric bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist. 45*(4), 1664–1693.

[44] Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *The Annals of Statistics 15*(2), 780–799.

[45] Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete 43*(4), 309–318.

[46] Nummelin, E. (1984a). *General irreducible Markov chains and nonnegative operators*, Volume 83 of *Cambridge Tracts in Mathematics*. Cambridge University Press.

[47] Nummelin, E. (1984b). *General Irreducible Markov Chains and Non-Negative Operators*. New York: Cambridge University Press.

[48] Paulin, D. (2015). Concentration inequalities for markov chains by marton couplings and spectral methods. *Electron. J. Probab. 20*, 32 pp.

[49] Peligrad, M. (1992). Properties of uniform consistency of the kernel estimators of density and regression functions under dependence assumptions. *Stochastics and Stochastic Reports 40*(3-4), 147–168.

[50] Portier, F. (2016). On the asymptotics of z-estimators indexed by the objective functions. *Electronic Journal of Statistics 10*(1), 464–494.

[51] Portier, F. and J. Segers (2018). On the weak convergence of the empirical conditional copula under a simplifying assumption. *Journal of Multivariate Analysis 166*, 160–181.

[52] Rakhlin, A. and K. Sridharan (2015). In *Measures of Complexity*, Festschrift for Alexey Chervonenkis, pp. 171–185. Vovk, V., Papadopoulos, H., Gammerman, A. editors, Springer Verlag.

[53] Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods* (Second ed.). Springer Texts in Statistics. Springer-Verlag, New York.

[54] Roberts, G. and R. Tweedie (1996). Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *Biometrika*, 95–110.

[55] Roberts, G. O. and J. S. Rosenthal (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv. 1*, 20–71.

[56] Shorack, G. R. and J. A. Wellner (2009). *Empirical processes with applications to statistics*, Volume 59. Siam.

[57] Smith, W. L. (1955). Regenerative stochastic processes. *Proc. Royal Stat. Soc. 232*, 6–31.

[58] Stute, W. (1982). A law of the logarithm for kernel density estimators. *The annals of Probability*, 414–422.

[59] Talagrand, M. (1994). Sharper bounds for gaussian and empirical processes. *The Annals of Probability*, 28–76.

[60] van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

[61] van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes. With applications to statistics*. Springer Series in Statistics. New York: Springer-Verlag.

[62] van der Vaart, A. W. and J. A. Wellner (2007). Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, 234–252.

[63] Vapnik, V. N. (1998). *Statistical learning theory. 1998.* Wiley, New York.

[64] Vapnik, V. N. and A. Y. Cèrvonenkis (1971). On the uniform convergence of frequencies of occurrence of events to their probabilities. *Probability Theory Appl. 1*(16), 264–280.

[65] Walter, G., J. Blum, et al. (1979). Probability density estimation using delta sequences. *the Annals of Statistics 7*(2), 328–340.

[66] Wintenberger, O. (2017). Exponential inequalities for unbounded functions of geometrically ergodic markov chains: applications to quantitative error bounds for regenerative metropolis algorithms. *Statistics 51*(1), 222–234.