



**HAL**  
open science

# Concentration inequality for U-statistics of order two for uniformly ergodic Markov chains

Quentin Duchemin, Yohann de Castro, Claire Lacour

► **To cite this version:**

Quentin Duchemin, Yohann de Castro, Claire Lacour. Concentration inequality for U-statistics of order two for uniformly ergodic Markov chains. *Bernoulli*, 2022, 29 (2), pp.929-956. 10.3150/22-BEJ1485 . hal-03014763v4

**HAL Id: hal-03014763**

**<https://hal.science/hal-03014763v4>**

Submitted on 16 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Concentration inequality for U-statistics of order two for uniformly ergodic Markov chains

Quentin Duchemin

LAMA, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France.

quentin.duchemin@univ-eiffel.fr

&

Yohann De Castro

Institut Camille Jordan, École Centrale de Lyon, Lyon, France

yohann.de-castro@ec-lyon.fr

&

Claire Lacour

LAMA, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France.

claire.lacour@univ-eiffel.fr

## Abstract

We prove a new concentration inequality for U-statistics of order two for uniformly ergodic Markov chains. Working with bounded and  $\pi$ -canonical kernels, we show that we can recover the convergence rate of Arcones and Giné who proved a concentration result for U-statistics of independent random variables and canonical kernels. Our result allows for a dependence of the kernels  $h_{i,j}$  with the indexes in the sums, which prevents the use of standard blocking tools. Our proof relies on an inductive analysis where we use martingale techniques, uniform ergodicity, Nummelin splitting and Bernstein's type inequality.

Assuming further that the Markov chain starts from its invariant distribution, we prove a Bernstein-type concentration inequality that provides sharper convergence rate for small variance terms.

## 1 Introduction

Concentration of measure has been intensely studied during the last decades since it finds application in large span of topics such as model selection (see [30] and [28]), statistical learning (see [8]), online learning (see [41]) or random graphs (see [9] and [14]). Important contributions in this field are those concerning U-statistics. A U-statistic of order  $m$  is a sum of the form

$$\sum_{1 \leq i_1 < \dots < i_m \leq n} h_{i_1, \dots, i_m}(X_{i_1}, \dots, X_{i_m}),$$

where  $X_1, \dots, X_n$  are independent random variables taking values in a measurable space  $(E, \Sigma)$  (with  $E$  Polish) and with respective laws  $P_i$  and where  $h_{i_1, \dots, i_m}$  are measurable functions of  $m$  variables  $h_{i_1, \dots, i_m} : E^m \rightarrow \mathbb{R}$ .

One important exponential inequality for U-statistics was provided by [4] using a Rademacher chaos approach. Their result holds for bounded and canonical (or degenerate) kernels, namely satisfying for all  $i_1, \dots, i_m \in [n] := \{1, \dots, n\}$  with  $i_1 < \dots < i_m$  and for all  $x_1, \dots, x_m \in E$ ,

$$\|h_{i_1, \dots, i_m}\|_\infty < \infty \quad \text{and} \quad \forall j \in [1, n], \mathbb{E}_{X_j} [h_{i_1, \dots, i_m}(x_1, \dots, x_{j-1}, X_j, x_{j+1}, \dots, x_m)] = 0.$$

They proved that in the degenerate case, the convergence rates for U-statistics are expected to be  $n^{m/2}$ . Relying on precise moment inequalities of Rosenthal type, Giné, Latala and Zinn in [20] improved the

---

This work was supported by grants from Région Ile-de-France.

result from [4] by providing the optimal four regimes of the tail, namely Gaussian, exponential, Weibull of orders  $2/3$  and  $1/2$ . In the specific case of order 2 U-statistics, Houdré and Reynaud-Bouret in [24] recovered the result from [20] by replacing the moment estimates by martingales type inequalities, giving as a by-product explicit constants. When the kernels are unbounded, it was shown that some results can be extended provided that the random variables  $h_{i_1, \dots, i_m}(X_{i_1}, \dots, X_{i_m})$  have sufficiently light tails. One can mention [16, Theorem 3.26] where an exponential inequality for U-statistics with a single Banach-space valued, unbounded and canonical kernel is proved. Their approach is based on a decoupling argument originally obtained by [10] and the tail behavior of the summands is controlled by assuming that the kernel satisfies the so-called weak Cramér condition. It is now well-known that with heavy-tailed distribution for  $h_{i_1, \dots, i_m}(X_{i_1}, \dots, X_{i_m})$  we cannot expect to get exponential inequalities anymore. Nevertheless working with kernels that have finite  $p$ -th moment for some  $p \in (1, 2]$ , Joly and Lugosi in [26] construct an estimator of the mean of the U-process using the median-of-means technique that performs as well as the classical U-statistic with bounded kernels.

All the above mentioned results consider that the random variables  $(X_i)_{i \geq 1}$  are independent. This condition can be prohibitive for practical applications since modelization of real phenomena often involves some dependence structure. The simplest and the most widely used tool to incorporate such dependence is Markov chain. One can give the example of Reinforcement Learning (see [39]) or Biology (see [38]). Recent works provide extensions of the classical concentration results to the Markovian settings as [17, 25, 32, 1, 8]. The asymptotic behaviour of U-statistics in the Markovian setup has already been investigated by several papers. We refer to [6] where the authors proved a Strong Law of Large Numbers and a Central Limit Theorem proved for U-statistics of order 2 using the *renewal approach* based on the splitting technique. One can also mention [15] regarding large deviation principles. However, there are only few results for the non-asymptotic behaviour of tails of U-statistics in a dependent framework. The first results were provided in [7] and [22] where exponential inequalities for U-statistics of order  $m \geq 2$  of time series under mixing conditions are proved. Those works were improved by [36] where a Hoeffding-type inequality for V and U-statistics is provided under conditions on the time dependent process that are easier to check in practice. In Section 3.4, we describe in details the result of [36] and the differences with our work. Let us point out that all the above mentioned works regarding non-asymptotic tail bound for U-statistics in a dependent framework consider a fixed kernel, namely  $h \equiv h_{i_1, \dots, i_m}$  for all  $i_1, \dots, i_m$ . Our work is the first to consider time dependent kernel functions which makes the theoretical analysis more challenging since the standard splitting method can be unworkable (cf. Section 2.5). In Section 3.4.1 and 3.4.2, we stress the importance of working with index-dependent kernels for practical applications and we show on a specific example that one can reach significantly faster convergence rates with this approach.

For the first time, we provide in this paper a Bernstein-type concentration inequality for U-statistics of order 2 in a dependent framework with kernels that may depend on the indexes of the sum and that are not assumed to be symmetric or smooth. We work on a general state space with bounded kernels that are  $\pi$ -canonical. This latter notion was first introduced in [19] who proved a variance inequality for U-statistics of ergodic Markov chains. Our Bernstein bound holds for stationary chains but we provide a Hoeffding-type inequality without any assumption on the initial distribution of the Markov chain.

## 1.1 Outline

In Section 2, we present and comment the assumptions under which our main results hold. In Section 3.1, we define and comment the key quantities involved in our results and we present our exponential inequalities with Theorems 1 and 2 in Section 3.2. Section 3.3 is dedicated to discussions where we give examples of Markov chains satisfying our assumptions and where we compare our results with the independent case. The proofs of both Theorems are presented in Section 4. In the Appendix, we provide the proof of some technical lemmas.

## 2 Assumptions and notations

We consider a Markov chain  $(X_i)_{i \geq 1}$  with transition kernel  $P : E \times E \rightarrow \mathbb{R}$  taking values in a measurable space  $(E, \Sigma)$ , and we introduce bounded functions  $h_{i,j} : E^2 \rightarrow \mathbb{R}$ . In this section, we describe the different assumptions on the Markov chain  $(X_i)_{i \geq 1}$  and on the functions  $h_{i,j}$  that we will consider in Theorems 1 and 2 presented in the next section.

### 2.1 Uniform ergodicity

**Assumption 1.** *The Markov chain  $(X_i)_{i \geq 1}$  is  $\psi$ -irreducible for some maximal irreducibility measure  $\psi$  on  $\Sigma$  (see [31, Section 4.2]). Moreover, there exist an integer  $m \geq 1$ ,  $\delta_m > 0$  and some probability measure  $\mu$  such that*

$$\forall x \in E, \forall A \in \Sigma, \quad \delta_m \mu(A) \leq P^m(x, A).$$

We denote by  $\pi$  the unique invariant distribution of the Markov chain  $(X_i)_{i \geq 1}$ .

For the reader familiar with the theory of Markov chains, Assumption 1 states that the whole space  $E$  is a small set which is equivalent to the uniform ergodicity of the Markov chain  $(X_i)_{i \geq 1}$  (see [31, Theorem 16.0.2]), namely there exist constants  $0 < \rho < 1$  and  $L > 0$  such that

$$\|P^n(x, \cdot) - \pi\|_{\text{TV}} \leq L\rho^n, \quad \forall n \geq 0, \pi\text{-a.e } x \in E,$$

where  $\pi$  is the unique invariant distribution of the chain  $(X_i)_{i \geq 1}$  and where for any measure  $\omega$  on  $(E, \Sigma)$ ,  $\|\omega\|_{\text{TV}} := \sup_{A \in \Sigma} |\omega(A)|$  is the total variation norm of  $\omega$ . From [18, section 2.3]), we also know that the Markov chain  $(X_i)_{i \geq 1}$  admits an absolute spectral gap  $1 - \lambda > 0$  with  $\lambda \in [0, 1)$  (thanks to uniform ergodicity). We refer to [17, Section 3.1] for a reminder on the spectral gap of Markov chains.

### 2.2 Upper-bounded Markov kernel

Assumption 2 can be read as a reverse Doeblin's condition and allows us to achieve a change of measure in expectations in our proof to work with i.i.d. random variables with distribution  $\nu$ . As a result, Assumption 2 is the cornerstone of our approach since it allows to decouple the U-statistic in the proof.

**Assumption 2.** *There exist  $\delta_M > 0$  and some probability measure  $\nu$  such that*

$$\forall x \in E, \forall A \in \Sigma, \quad P(x, A) \leq \delta_M \nu(A).$$

Assumption 2 has already been used in the literature (see [29, Section 4.2]) and was introduced in [12]. This condition can typically require the state space to be compact as highlighted in [29].

Let us describe another situation where Assumption 2 holds. Consider that  $(E, \|\cdot\|)$  is a normed space and that for all  $x \in E$ ,  $P(x, dy)$  has density  $p(x, \cdot)$  with respect to some measure  $\eta$  on  $(E, \Sigma)$ . We further assume that there exists an integrable function  $u : E \rightarrow \mathbb{R}_+$  such that  $\forall x, y \in E, \quad p(x, y) \leq u(y)$ . Then considering for  $\nu$  the probability measure with density  $u/\|u\|_1$  with respect to  $\eta$  and  $\delta_M = \|u\|_1$ , Assumption 2 holds.

### 2.3 Exponential integrability of the regeneration time

We introduce some additional notations which will be useful to apply Talagrand concentration result from [34]. Note that this section is inspired from [1] and [31, Theorem 17.3.1]. We assume that Assumption 1 is satisfied and we extend the Markov chain  $(X_i)_{i \geq 1}$  to a new (so called *split*) chain  $(\tilde{X}_n, R_n) \in E \times \{0, 1\}$  (see [31, Section 5.1] for a reminder on the splitting technique), satisfying the following properties.

- $(\tilde{X}_n)_n$  is again a Markov chain with transition kernel  $P$  with the same initial distribution as  $(X_n)_n$ . We recall that  $\pi$  is the invariant distribution on the  $E$ .
- if we define  $T_1 = \inf\{n > 0 : R_{nm} = 1\}$ ,

$$T_{i+1} = \inf\{n > 0 : R_{(T_1+\dots+T_i+n)m} = 1\},$$

then  $T_1, T_2, \dots$  are well defined and independent. Moreover  $T_2, T_3, \dots$  are i.i.d.

- if we define  $S_i = T_1 + \dots + T_i$ , then the “blocks”

$$Y_0 = (\tilde{X}_1, \dots, \tilde{X}_{mT_1+m-1}), \quad \text{and} \quad Y_i = (\tilde{X}_{m(S_i+1)}, \dots, \tilde{X}_{m(S_{i+1}+1)-1}), \quad i > 0,$$

form a one-dependent sequence (i.e. for all  $i$ ,  $\sigma((Y_j)_{j < i})$  and  $\sigma((Y_j)_{j > i})$  are independent). Moreover, the sequence  $Y_1, Y_2, \dots$  is stationary and if  $m = 1$  the variables  $Y_0, Y_1, \dots$  are independent. In consequence, for any measurable space  $(S, \mathcal{B})$  and measurable functions  $f : S \rightarrow \mathbb{R}$ , the variables

$$Z_i = Z_i(f) = \sum_{j=m(S_i+1)}^{m(S_{i+1}+1)-1} f(\tilde{X}_j), \quad i \geq 1,$$

constitute a one-dependent sequence (an i.i.d. sequence if  $m = 1$ ). Additionally, if  $f$  is  $\pi$ -integrable (recall that  $\pi$  is the unique stationary measure for the chain), then

$$\mathbb{E}[Z_i] = \delta_m^{-1} m \int f d\pi.$$

- the distribution of  $T_1$  depends only on  $\pi, P, \delta_m, \mu$ , whereas the law of  $T_2$  only on  $P, \delta_m$  and  $\mu$ .

**Remark** Let us highlight that  $(\tilde{X}_n)_n$  is a Markov chain with transition kernel  $P$  and same initial distribution as  $(X_n)_n$ . Hence for our purposes of estimating the tail probabilities, we will identify  $(X_n)_n$  and  $(\tilde{X}_n)_n$ .

To derive a concentration inequality, we use the exponential integrability of the regeneration times which is ensured if the chain is uniformly ergodic as stated by Proposition 1. A proof can be found in Section G of the Appendix.

**Definition 1.** For  $\alpha > 0$ , define the function  $\psi_\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with the formula  $\psi_\alpha(x) = \exp(x\alpha) - 1$ . Then for a random variable  $X$ , the  $\alpha$ -Orlicz norm is given by

$$\|X\|_{\psi_\alpha} = \inf \{ \gamma > 0 : \mathbb{E}[\psi_\alpha(|X|/\gamma)] \leq 1 \}.$$

**Proposition 1.** If Assumption 1 holds, then

$$\|T_1\|_{\psi_1} < \infty \quad \text{and} \quad \|T_2\|_{\psi_1} < \infty, \tag{1}$$

where  $\|\cdot\|_{\psi_1}$  is the 1-Orlicz norm introduced in Definition 1. We denote  $\tau := \max(\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1})$ .

## 2.4 $\pi$ -canonical and bounded kernels

With Assumption 3, we introduce the notion of  $\pi$ -canonical kernel which is the counterpart of the canonical property from [21].

**Assumption 3.** Let us denote  $\mathcal{B}(\mathbb{R})$  the Borel algebra on  $\mathbb{R}$ . For all  $i, j \in [n]$ , we assume that  $h_{i,j} : (E^2, \Sigma \otimes \Sigma) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is measurable and is  $\pi$ -canonical, namely

$$\forall x, y \in E, \quad \mathbb{E}_\pi[h_{i,j}(X, x)] = \mathbb{E}_\pi[h_{i,j}(X, y)] = \mathbb{E}_\pi[h_{i,j}(x, X)] = \mathbb{E}_\pi[h_{i,j}(y, X)].$$

This common expectation will be denoted  $\mathbb{E}_\pi[h_{i,j}]$ .

Moreover, we assume that for all  $i, j \in [n]$ ,  $\|h_{i,j}\|_\infty < \infty$ .

### Remarks

- A large span of kernels are  $\pi$ -canonical. This is the case of translation-invariant kernels which have been widely studied in the Machine Learning community. Another example of  $\pi$ -canonical kernel is a rotation invariant kernel when  $E = \mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$  with  $\pi$  also rotation invariant (see [9] or [14]).
- The notion of  $\pi$ -canonical kernels is the counterpart of canonical kernels in the i.i.d. framework (see for example [24]). Note that we are not the first to introduce the notion of  $\pi$ -canonical kernels working with Markov chains. In [19], Fort and al. provide a variance inequality for U-statistics whose underlying

sequence of random variables is an ergodic Markov Chain. Their results holds for  $\pi$ -canonical kernels as stated with [19, Assumption A2].

• Note that if the kernels  $h_{i,j}$  are not  $\pi$ -canonical, the U-statistic decomposes into a linear term and a  $\pi$ -canonical U-statistic. This is called the *Hoeffding decomposition* (see [21, p.176]) and takes the following form

$$\begin{aligned} \sum_{i \neq j} (h_{i,j}(X_i, X_j) - \mathbb{E}_{(X,Y) \sim \pi \otimes \pi} [h_{i,j}(X, Y)]) &= \sum_{i \neq j} \tilde{h}_{i,j}(X_i, X_j) - \mathbb{E}_\pi [\tilde{h}_{i,j}] \\ &+ \sum_{i \neq j} (\mathbb{E}_{X \sim \pi} [h_{i,j}(X, X_j)] - \mathbb{E}_{(X,Y) \sim \pi \otimes \pi} [h_{i,j}(X, Y)]) \\ &+ \sum_{i \neq j} (\mathbb{E}_{X \sim \pi} [h_{i,j}(X_i, X)] - \mathbb{E}_{(X,Y) \sim \pi \otimes \pi} [h_{i,j}(X, Y)]), \end{aligned}$$

where for all  $j$ , the kernel  $\tilde{h}_{i,j}$  is  $\pi$ -canonical with

$$\forall x, y \in E, \quad \tilde{h}_{i,j}(x, y) = h_{i,j}(x, y) - \mathbb{E}_{X \sim \pi} [h_{i,j}(x, X)] - \mathbb{E}_{X \sim \pi} [h_{i,j}(X, y)].$$

## 2.5 Additional technical assumption

In the case where the kernels  $h_{i,j}$  depend on both  $i$  and  $j$ , we need Assumption 4.(ii) to prove Theorem 1. Assumption 4.(ii) is a mild condition on the initial distribution of the Markov chain that is used when we apply Bernstein's inequality for Markov chains from Proposition 4 (see Section C of the Appendix).

**Assumption 4.** *At least one of the following conditions holds.*

- (i) *For all  $i, j \in [n]$ ,  $h_{i,j} \equiv h_{1,j}$ , i.e. the kernel function  $h_{i,j}$  does not depend on  $i$ .*
- (ii) *The initial distribution of the Markov chain  $(X_i)_{i \geq 1}$ , denoted  $\chi$ , is absolutely continuous with respect to the invariant measure  $\pi$  and its density  $\frac{d\chi}{d\pi}$  has finite  $p$ -moment for some  $p \in (1, \infty]$ , i.e*

$$\infty > \left\| \frac{d\chi}{d\pi} \right\|_{\pi, p} := \begin{cases} \left[ \int \left| \frac{d\chi}{d\pi} \right|^p d\pi \right]^{1/p} & \text{if } p < \infty, \\ \text{ess sup} \left| \frac{d\chi}{d\pi} \right| & \text{if } p = \infty. \end{cases}$$

In the following, we will denote  $q = \frac{p}{p-1} \in [1, \infty)$  (with  $q = 1$  if  $p = +\infty$ ) which satisfies  $\frac{1}{p} + \frac{1}{q} = 1$ .

Assumption 4 is needed at one specific step of our proof where we need to bound with high probability

$$\sum_{j=2}^n \mathbb{E} \left[ \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j) \right|^k \right], \quad \text{with } \forall i, j, \quad \forall x, y \in E, \quad p_{i,j}(x, y) := h_{i,j}(x, y) - \mathbb{E}_\pi [h_{i,j}],$$

and where  $(X'_j)_{j \geq 1}$  are i.i.d. random variables with distribution  $\nu$  from Assumption 2. In the case where Assumption 4.(i) holds, we can use for any fixed  $j \in \{2, \dots, n\}$  the splitting method to decompose the sum  $\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)$  in different blocks whose lengths are given by the regeneration times of the split chain. Thanks to Assumption 4.(i), those blocks are independent and we can use standard concentration tools for sums of independent random variables. This approach is valid for any initial distribution of the chain. However, if Assumption 4.(i) is not satisfied, the blocks used to decompose  $\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)$  are not independent and the splitting method can no longer be used. To bypass this issue, we need a Bernstein-type concentration inequality for additive functionals of Markov chains with time-dependent functions (see Proposition 4 in Section C of the Appendix). Proposition 4 is a straightforward corollary of [25, Theorem 1] and requires Assumption 4.(ii) to be satisfied. We refer to Section 4.2 and in particular to Section 4.2.1 for further details.

### 3 Main results

#### 3.1 Preliminary comments

Under the assumptions presented in Section 2, Theorem 1 and 2 provided in Section 3.2 give exponential inequalities for the U-statistic

$$U_{\text{stat}}(n) = \sum_{1 \leq i < j \leq n} (h_{i,j}(X_i, X_j) - \mathbb{E}[h_{i,j}(X_i, X_j)]).$$

Theorem 1 provides an Hoeffding-type concentration result that holds without any (or mild) condition on the initial distribution of the chain. By assuming that the chain  $(X_i)_{i \geq 1}$  is stationary (meaning that  $X_1$  is distributed according to  $\pi$ ), Theorem 2 gives a Bernstein-type concentration inequality and leads to a better convergence rate compared to Theorem 1.

The proof of our main results relies on a martingale technique conducted by induction at depth  $t_n := \lfloor r \log n \rfloor$  with  $r > 2(\log(1/\rho))^{-1}$  (see the remark following Assumption 1 for the definition of  $\rho$ ). With the notations of Section 2, our concentration inequalities involve the following quantities

$$A := 2 \max_{i,j} \|h_{i,j}\|_{\infty}, \quad C_n^2 := \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbb{E} \left[ \mathbb{E}_{X' \sim \nu} [p_{i,j}^2(X_i, X')] \right], \quad (2)$$

$$B_n^2 := \max \left[ \max_{0 \leq k \leq t_n} \max_i \sup_x \sum_{j=i+1}^n \mathbb{E}_{X' \sim \nu} \left( \mathbb{E}_{X \sim P^k(X', \cdot)} p_{i,j}(x, X) \right)^2, \right. \\ \left. \max_{0 \leq k \leq t_n} \max_j \sup_y \sum_{i=1}^{j-1} \mathbb{E}_{\tilde{X} \sim \pi} \left( \mathbb{E}_{X \sim P^k(y, \cdot)} p_{i,j}(\tilde{X}, X) \right)^2 \right], \quad (3)$$

with the convention that  $P^0(y, \cdot)$  is the Dirac measure at point  $y \in E$ . Let us comment those terms.

- Understanding of the origin of  $B_n$ .  $B_n$  involves supremums over  $k$  ranging from 0 to  $t_n$ . The terms in the supremum corresponding to some specific value of  $k$  arise in our proof at the  $k$ -th step of our induction procedure (and will be denoted  $\mathfrak{B}_k$  in Section 4, so that  $B_n = \sup_{0 \leq k \leq t_n} \mathfrak{B}_k$ ).
- Bounding  $B_n$  with uniform ergodicity. The uniform ergodicity of the Markov chain ensured by Assumption 1 can allow to bound  $B_n$  since for all  $x, y \in E$  and for all  $k \geq 0$ ,

$$\left| \mathbb{E}_{X \sim P^k(y, \cdot)} p_{i,j}(x, X) \right| \leq \sup_z |h_{i,j}(x, z)| \times \|P^k(y, \cdot) - \pi\|_{\text{TV}}.$$

- The case where  $\nu = \pi$  and the independent setting

In the specific case where  $\nu = \pi$  (which includes the independent setting), we get that

$$C_n^2 = \sum_{i < j} \mathbb{E} \left\{ \text{Var}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X}) | X_i] \right\},$$

and using Jensen inequality that

$$B_n^2 \leq \max \left[ \sup_{x,i} \sum_{j=i+1}^n \text{Var}_{\tilde{X} \sim \pi} [h_{i,j}(x, \tilde{X})], \sup_{y,j} \sum_{i=1}^{j-1} \text{Var}_{\tilde{X} \sim \pi} [h_{i,j}(\tilde{X}, y)] \right].$$

Hence,  $C_n^2$  and  $B_n^2$  can be understood as variance terms that would tend to be larger as  $\nu$  moves away from  $\pi$ . Let us point out that in the independent setting, all terms for  $k$  ranging from 1 to  $t_n$  in the definition of  $B_n^2$  vanish but the term corresponding to  $k = 0$  does not since  $P^0(y, \cdot)$  is the Dirac measure at  $y$ . We provide a detailed comparison of our results with known exponential inequalities in the independent setting in Section 3.3.2.

- Bounding  $B_n$  and  $C_n$ : A way to read immediately the convergence rates in our main results

Using coarse bounds, one immediately gets that  $B_n \leq A\sqrt{n}$  and  $C_n \leq An$ . We prompt the reader to keep in mind these bounds in order to directly see the rate of convergence and the dominant terms in the inequalities from Section 3.2. These bounds can be significantly improved for particular cases as done in the example presented in Section 3.4.2.

### 3.2 Exponential inequalities

We now state our first result Theorem 1 whose proof can be found in Section 4.1.1.

**Theorem 1.** *Let  $n \geq 2$ . We suppose Assumptions 1, 2 and 3 described in Section 2. There exist two constants  $\beta, \kappa > 0$  such that for any  $u > 0$ ,*

a) *if Assumption 4.(i) is satisfied, it holds with probability at least  $1 - \beta e^{-u} \log(n)$ ,*

$$U_{\text{stat}}(n) \leq \kappa \log(n) \left( [A \log(n) \sqrt{n}] \sqrt{u} + [A + B_n \sqrt{n}] u + [2A \sqrt{n}] u^{3/2} + A[u^2 + n] \right),$$

b) *if Assumption 4.(ii) is satisfied, it holds with probability at least  $1 - \beta e^{-u} \log(n)$ ,*

$$U_{\text{stat}}(n) \leq \kappa \log(n) \left( [C_n + A \log(n) \sqrt{n}] \sqrt{u} + [A + B_n \sqrt{n}] u + [2A \sqrt{n}] u^{3/2} + A[u^2 + n] \right).$$

Note that the kernels  $h_{i,j}$  do not need to be symmetric and that we do not consider any assumption on the initial measure of the Markov chain  $(X_i)_{i \geq 1}$  if the kernels  $h_{i,j}$  do not depend on  $i$  (see Assumption 4). By bounding coarsely  $B_n$  and  $C_n$  in Theorem 1 (respectively by  $\sqrt{n}A$  and  $nA$ ), we get that there exist constants  $\beta, \kappa > 0$  such that for any  $u \geq 1$ , it holds with probability at least  $1 - \beta e^{-u} \log n$ ,

$$\frac{2}{n(n-1)} U_{\text{stat}}(n) \leq \kappa \max_{i,j} \|h_{i,j}\|_{\infty} \log n \left\{ \frac{u}{n} + \left[ \frac{u}{n} \right]^2 \right\}. \quad (4)$$

In particular it holds

$$\frac{2}{n(n-1)} U_{\text{stat}}(n) = \mathcal{O}_{\mathbb{P}} \left( \frac{\log(n) \log \log n}{n} \right),$$

where  $\mathcal{O}_{\mathbb{P}}$  denotes stochastic boundedness. Up to a  $\log(n) \log \log n$  multiplicative term, we uncover the optimal rate of Hoeffding's inequality for canonical U-statistics of order 2, see [26]. Taking a close look at the proof of Theorem 1 (and more specifically at Section 4.3), one can remark that the same results hold if the U-statistic is centered with the expectations  $\mathbb{E}_{\pi}[h_{i,j}]$ , namely for

$$\sum_{1 \leq i < j \leq n} (h_{i,j}(X_i, X_j) - \mathbb{E}_{\pi}[h_{i,j}]).$$

It is well-known that one can expect a better convergence rate when variance terms are small with a Bernstein bound. The main limitation in Theorem 1 that prevents us from taking advantage of small variances is the term at the extreme right on the concentration inequality of Theorem 1, namely  $An \log n$ . Working with the additional assumption that the Markov chain  $(X_i)_{i \geq 1}$  is stationary – meaning that the initial distribution of the chain is the invariant distribution  $\pi$  – we are able to prove a Bernstein-type concentration inequality as stated with Theorem 2. The proof of Theorem 2 is provided in Section 4.1.2. Stationarity is only used to bound the remaining terms that were not already considered in the  $t_n$  steps of our induction procedure (see Section 3.1 for the definition of  $t_n$ ). We refer to the proof of Proposition 3.b) in Section 4.3 for details.

**Theorem 2.** *We suppose Assumptions 1, 2 and 3 described in Section 2. We further assume that the Markov chain  $(X_i)_{i \geq 1}$  is stationary. Then there exist two constants  $\beta, \kappa > 0$  such that for any  $u > 0$ , it holds with probability at least  $1 - \beta e^{-u} \log n$ ,*

$$U_{\text{stat}}(n) \leq \kappa \log(n) \left( [C_n + A \log(n) \sqrt{n}] \sqrt{u} + [A + B_n \sqrt{n}] u + [2A \sqrt{n}] u^{3/2} + A[u^2 + \log n] \right).$$

*In case where Assumption 4.(i) holds, one can remove  $C_n$  in the previous inequality.*



### 3.3 Discussion

#### 3.3.1 Examples of Markov chains satisfying the Assumptions

**Example 1: Finite state space.** For Markov chains with finite state space, Assumption 2 holds trivially. Hence, in such framework the result of Theorem 1 holds for any uniformly ergodic Markov chain. In particular, this is true for any aperiodic and irreducible Markov chains using [5, Lemma 7.3.(ii)].

**Example 2: AR(1) process.** Let us consider the process  $(X_n)_{n \in \mathbb{N}}$  on  $\mathbb{R}^k$  defined by

$$X_0 \in \mathbb{R}^k \text{ and for all } n \in \mathbb{N}, \quad X_{n+1} = H(X_n) + Z_n,$$

where  $(Z_n)_{n \in \mathbb{N}}$  are i.i.d random variables in  $\mathbb{R}^k$  and  $H : \mathbb{R}^k \rightarrow \mathbb{R}^k$  is an application. Such a process is called an auto-regressive process of order 1, noted AR(1). Assuming that the distribution of  $Z_1$  has density  $f_Z$  with respect to the Lebesgue measure on  $\mathbb{R}^k$  (denoted  $\lambda_{Leb}$ ), it is well-known that mild regularity conditions on  $H$  and  $f_Z$  ensure that the Markov chain  $(X_i)_{i \geq 1}$  is uniformly ergodic. These conditions require in particular that both  $f_Z$  and  $H$  are continuously differentiable with  $H$  bounded. We refer to [13] for the full statement.

We denote  $B_H := B(0, \|H\|_\infty)$  the euclidean ball in  $\mathbb{R}^k$  with radius  $\|H\|_\infty$  centered at 0. Assuming that  $y \mapsto \sup_{\{z \in B_H\}} f_Z(y-z)$  is integrable on  $\mathbb{R}^k$  with respect to  $\lambda_{Leb}$ , we get that Assumption 2 holds (see the remark following Assumption 2). The previous condition on  $f_Z$  is for example satisfied for Gaussian distributions. We deduce that Theorems 1 and 2 can be applied in such settings that are typically found in nonlinear filtering problem (see [12, Section 4]).

**Example 3: ARCH process.** Let us consider  $E = \mathbb{R}$ . The ARCH model is

$$X_{n+1} = H(X_n) + G(X_n)Z_{n+1},$$

where  $H$  and  $G$  are continuous functions, and  $(Z_n)_n$  are i.i.d. centered normal random variables with variance  $\sigma^2 > 0$ . Assuming that  $\inf_x |G(x)| \geq a > 0$ , we know that the Markov chain  $(X_n)_n$  is irreducible and aperiodic (see [3, Lemma 1]). Assuming further that  $\|H\|_\infty \leq b < \infty$  and that  $\|G\|_\infty \leq c$ , we can show that Assumptions 1 and 2 hold. Let us first remark that the transition kernel  $P$  of the Markov chain  $(X_n)_n$  is such that for any  $x \in \mathbb{R}$ ,  $P(x, dy)$  has density  $p(x, \cdot)$  with respect to the Lebesgue measure with

$$p(x, y) = (2\pi\sigma^2)^{-1} \exp\left(-\frac{(y-H(x))^2}{2\sigma^2 G(x)^2}\right).$$

Defining for any  $y \in \mathbb{R}$ ,

$$g_m(y) := \frac{1}{2\pi\sigma^2} \times \begin{cases} \exp\left(-\frac{(y-B)^2}{2\sigma^2 a^2}\right) & \text{if } y < -b \\ \exp\left(-\frac{2B^2}{\sigma^2 a^2}\right) & \text{if } |y| \leq b \\ \exp\left(-\frac{(y+B)^2}{2\sigma^2 a^2}\right) & \text{if } y > b \end{cases} \quad \text{and} \quad g_M(y) := (2\pi\sigma^2)^{-1} \times \begin{cases} \exp\left(-\frac{(y+b)^2}{2\sigma^2 c^2}\right) & \text{if } y < -b \\ 1 & \text{if } |y| \leq b \\ \exp\left(-\frac{(y-b)^2}{2\sigma^2 c^2}\right) & \text{if } y > b \end{cases},$$

it holds  $g_m(y) \leq p(x, y) \leq g_M(y)$  for any  $x, y \in \mathbb{R}$ . We deduce that considering  $\delta_m = \|g_m\|_1$ ,  $\delta_M = \|g_M\|_1$  and  $\mu$  (resp.  $\nu$ ) with density  $g_m/\|g_m\|_1$  (resp.  $g_M/\|g_M\|_1$ ) with respect to the Lebesgue measure on  $\mathbb{R}$ , Assumptions 1 and 2 hold.

#### 3.3.2 The independent setting

In the case where the random variables  $(X_i)_{i \geq 1}$  are independent, the terms  $B_n^2$  and  $C_n^2$  involved in our exponential inequality take the following form

$$B_n^2 = \max \left[ \sup_{i,x} \sum_{j=i+1}^n \mathbb{E} [p_{i,j}^2(x, X_j)], \sup_{j,y} \sum_{i=1}^{j-1} \mathbb{E} [p_{i,j}^2(X_i, y)] \right] \quad (5)$$

and  $C_n^2 = \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbb{E} [p_{i,j}(X_i, X_j)^2],$

where we remind that all terms for  $k \in \{1, \dots, t_n\}$  in the definition of  $B_n^2$  in Eq.(3) vanish and it only remains the contribution of terms for  $k = 0$ . In the independent setting, [24, Theorem 1] proved that for any  $u > 0$ , it holds with probability at least  $1 - 3e^{-u}$ ,

$$U_{\text{stat}}(n) \leq C_n \sqrt{u} + (D_n + F_n)u + B_n u^{3/2} + Au^2,$$

where  $A$ ,  $B_n$  and  $C_n$  coincide with the terms of this paper (see Eq.(5)). Let us comment the quantities involved in the different regimes of the tail behaviour.

- Sub-Gaussian. In Theorem 2, we recover the term  $C_n$  from [24] and we suffer an additional  $A\sqrt{n} \log n$  term.
- Sub-Exponential.  $D_n$  and  $F_n$  come from duality arguments in the proof of [24]. We do not recover the counterpart of these terms in Theorem 2 since working with dependent variables bring additional technical difficulties and the use for example of a decoupling argument.  $D_n + F_n$  is replaced by  $A + B_n \sqrt{n}$  in our result.
- Sub-Weibull with parameter 2/3. While [24] finds the quantity  $B_n$  for the term  $u^{3/2}$ , the counterpart in Theorem 2 is the worst case scenario since it always holds  $B_n \leq A\sqrt{n}$ .
- Sub-Weibull with parameter 1/2. We obtain the same behaviour for the sub-Weibull (with parameter 1/2) regime of the tail behaviour.

Let us also mention that Theorem 2 has an additive term  $A \log^2 n$  (that will not be dominant for standard choice of  $u$ ). This term can be understood as a proof artefact and arises when we bound the remaining terms in the U-statistic that were not considered in our induction procedure. We finally point out that our result involves additive  $\log n$  factors (both in the tail bound and in the probability).

### 3.4 Connections with the literature

In this section, we describe the concentration inequality obtained in [36] for U-statistics in a dependent framework and we explain the differences with our work. We consider an integer  $n \in \mathbb{N} \setminus \{0\}$  and a geometrically  $\alpha$ -mixing sequence  $(X_i)_{i \in [n]}$  (see [36, Section 2]) with coefficient

$$\alpha(i) \leq \gamma_1 \exp(-\gamma_2 i), \quad \text{for all } i \geq 1,$$

where  $\gamma_1, \gamma_2$  are two positive absolute constants. We consider a kernel  $h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  degenerate, symmetric, continuous, integrable and satisfying for some  $q \geq 1$ ,  $\int_{\mathbb{R}^{2d}} |\mathcal{F}h(u)| \|u\|_2^q du < \infty$ , where  $\mathcal{F}h$  denotes the Fourier-transform of  $h$ . Then Eq.(2.4) from [36] states that there exists a constant  $c > 0$  such that for any  $u > 0$ , it holds with probability at least  $1 - 6e^{-u}$

$$\frac{2}{n(n-1)} U_{\text{stat}}(n) \leq 4c \|\mathcal{F}h\|_{L^1} \left\{ A_n^{1/2} \frac{u}{n} + c \log^4(n) \left[ \frac{u}{n} \right]^2 \right\}, \quad (6)$$

where  $A_n^{1/2} = 4 \left( \frac{64\gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{\log^4(n)}{n} \right)$  and  $U_{\text{stat}}(n) = \sum_{1 \leq i < j \leq n} (h(X_i, X_j) - \mathbb{E}_\pi[h])$ .

[36] has the merit of working with geometrically  $\alpha$ -mixing stationary sequences which includes in particular geometrically (and hence uniformly) ergodic Markov chains (see [27, p.6]). For the sake of simplicity, we presented the result of [36] for U-statistics of order 2, but their result holds for U-statistics of arbitrary order  $m \geq 2$ . Nevertheless, they only consider state spaces like  $\mathbb{R}^d$  with  $d \geq 1$  and they work with a unique kernel  $h$  (i.e.  $h_{i_1, \dots, i_m} = h$  for any  $i_1, \dots, i_m$ ) which is assumed to be symmetric continuous, integrable and that satisfies some smoothness assumption. On the contrary, we consider general state spaces and we allow different kernels  $h_{i,j}$  that are not assumed to be symmetric or smooth. In addition, Theorem 2 is a Bernstein-type exponential inequality where we can benefit from small variance terms, which is not the case for [36]. We provide a specific example in Section 3.4.2.

### 3.4.1 Motivations for the study of time dependent kernels

In this section, we want to stress the importance of working with weighted U-statistics for practical applications. In the following, we detail two specific examples borrowed from the fields of information retrieval and of homogeneity tests. Note that one could find other applications such as in genetic association (cf. [42]) or for independence tests (cf. [37]).

**Average-Precision Correlation** When we search the Internet, the browser computes a numeric score on how well each object in the database matches the query, and rank the objects according to this value. In order to evaluate the quality of this browser, a standard approach in the field of information retrieval consists in comparing the ranking provided by the web search engine and the ranking obtained from human labels (cf. [23]). One way to measure how well both rankings are aligned is to report the correlation between them. One of the most commonly used rank correlation statistic is the Kendall's  $\tau$ . Considering a dataset of size  $n \in \mathbb{N}$  ordered according to the human labels and denoting  $X_i$  the rank the browser gives to the  $i$ -th element, the Kendall's  $\tau$  is defined by

$$\tau^{\text{Ken}} := \frac{2}{n(n-1)} \sum_{i \neq j} \left\{ \mathbb{1}_{X_i > X_j} \mathbb{1}_{i > j} + \mathbb{1}_{X_i < X_j} \mathbb{1}_{i < j} \right\} - 1.$$

Since only the top ranking objects are shown to the user, it would be legitimate to penalize heavier errors made on items having high rankings. The Kendall's  $\tau$  does not make such distinctions and new correlation measurements have been popularized to address this issue. One of them is the so-called Average-Precision Correlation (cf. [43]) which is defined by

$$\tau^{\text{AP}} := \frac{2}{n-1} \sum_{j=2}^n \frac{\sum_{i=1}^{j-1} \mathbb{1}_{X_i < X_j}}{j-1} - 1.$$

Note that  $\tau^{\text{AP}}$  is a U-statistic where the kernels  $h_{i,j}(x, y) := \frac{\mathbb{1}_{x < y}}{j-1}$  depend on  $j$ . Let us point out that  $h_{i,j}$  do not depend on  $i$  so that Assumption 4.(i) holds.

**Accounting for confounding covariates** U-statistics are powerful tools to compare the distributions of random variables across two groups (say with labels 0 and 1) from samples  $X_1, \dots, X_n$  and  $X_{n+1}, \dots, X_{n+m}$ . The typical example is the Wilcoxon Rank Sum Test (WRST) based on the following U-statistic

$$\sum_{i=1}^n \sum_{j=1}^m h(X_i, X_{n+j}) \quad \text{where} \quad h(x, y) := \frac{1}{2} \mathbb{1}_{x < y} + \frac{1}{2} \mathbb{1}_{x \leq y}.$$

The WRST relies on the following idea: if the data is pooled and then ranked, the average rank of observations from each group should be the same. For any  $i \in [n+m]$ , let  $G_i$  be the random variable valued in  $\{0, 1\}$  allocating the  $i$ -th individual to one of the two groups. Note that the observed allocation  $(g_i)_{i \in [n+m]}$  is given by  $g_i = 0$  if and only if  $i \leq n$ . When group membership is not assigned through randomization, there may be confounding covariates  $Z$  (assumed to be observed) that can cause a spurious association between outcome and group membership. In that case, we wish rather to test the null hypothesis  $\mathbb{P}(X \leq t | G = 0, Z = z) = \mathbb{P}(X \leq t | Z = z)$ . In [35], the authors developed such a test by working with the following adjusted U-statistics involving index-dependent kernels

$$\left( \sum_{i: g_i=0} w(z_i, g_i) \sum_{j: g_j=1} w(z_j, g_j) \right)^{-1} \sum_{i: g_i=0} \sum_{j: g_j=1} h(X_i, X_j) w(z_i, g_i) w(z_j, g_j),$$

where the weights  $w(z_i, g_i) = (\mathbb{P}(G = g_i | Z = z_i))^{-1}$  can be estimated with a logistic regression.

### 3.4.2 Time dependent kernels and convergence rate

In this section, we consider a stationary Markov chain  $(X_i)_{i \geq 1}$  satisfying Assumptions 1, 2 and 3. We study the case where there exist reals  $(a_{i,j})_{i,j \in \mathbb{N}}$  such that for all  $i, j \in \mathbb{N}$ ,  $h_{i,j} = a_{i,j} h$  for some  $\pi$ -canonical kernel

$h : E^2 \rightarrow \mathbb{R}$ . For simplicity, we consider that  $\mathbb{E}_\pi h = 0$  leading to  $p_{i,j} = h_{i,j}$ . Let us consider the specific example where  $a_{i,j} = |j-i|^{-1}$  for  $i \neq j$ . In such setting, the coefficients  $a_{i,j}$ 's are weighting each summand in the U-statistic: the larger  $|j-i|$ , the smaller is the contribution of the term indexed by  $(i, j)$  in the sum. As a result, interpreting indexes as time steps, the  $a_{i,j}$ 's can be understood as forgetting factors. Since

$$B_n^2 \leq A^2 \max \left\{ \max_i \sum_{j=i+1}^n |j-i|^{-2}, \max_j \sum_{i=1}^{j-1} |j-i|^{-2} \right\} \leq A^2 \sum_{j=2}^n |j-1|^{-2} \leq A^2 \frac{\pi^2}{6},$$

$$\text{and } C_n^2 \leq A^2 \sum_{j=2}^n \sum_{i=1}^{j-1} |j-i|^{-2} \leq A^2 \sum_{s=1}^n \frac{s}{s^2} \leq A^2 \left(1 + \int_1^n \frac{1}{x} dx\right) \leq A^2(1 + \log n),$$

Theorem 2 ensures that there exist constants  $\beta, \kappa > 0$  such that for any  $u \geq 1$  it holds with probability at least  $1 - \beta e^{-u} \log n$ ,

$$\frac{2}{n(n-1)} U_{\text{stat}}(n) \leq \kappa A \log n \left( \log(n) \frac{\sqrt{u}}{n^{3/2}} + \left[ \frac{u}{n} \right]^{3/2} + \left[ \frac{u}{n} \right]^2 \right).$$

In particular, with probability at least  $1 - \beta \frac{\log n}{n}$  we have  $\frac{2}{n(n-1)} U_{\text{stat}}(n) \leq 3\kappa A \frac{\log^{5/2} n}{n^{3/2}}$ . This convergence rate improves significantly the one obtained from an Hoeffding-type concentration inequality like Eq. (4) that would lead to  $U_{\text{stat}}(n) \leq 2\kappa A \frac{\log^{3/2} n}{n}$  with probability at least  $1 - \beta \frac{\log n}{n}$ .

## 4 Proofs

### 4.1 Proofs of Theorems 1 and 2

Our proof is inspired from [24] where a Bernstein-type inequality is shown for U-statistics of order 2 in the independent setting. Their proof relies on the *canonical* property of the kernel functions which endowed the U-statistic with a martingale structure. We want to use a similar argument and we decompose  $U_{\text{stat}}(n)$  to recover the martingale property for each term (except for the last one). Considering for any  $l \geq 1$  the  $\sigma$ -algebra  $G_l = \sigma(X_1, \dots, X_l)$ , the notation  $\mathbb{E}_l$  refers to the conditional expectation with respect to  $G_l$ . Then we decompose  $U_{\text{stat}}(n)$  as follows,

$$U_{\text{stat}}(n) = M_{\text{stat}}^{(t_n)}(n) + R_{\text{stat}}^{(t_n)}(n), \tag{7}$$

$$\text{with } M_{\text{stat}}^{(t_n)}(n) = \sum_{k=1}^{t_n} \sum_{i < j} \left( \mathbb{E}_{j-k+1} [h_{i,j}(X_i, X_j)] - \mathbb{E}_{j-k} [h_{i,j}(X_i, X_j)] \right),$$

$$R_{\text{stat}}^{(t_n)}(n) = \sum_{i < j} \left( \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right),$$

and where  $t_n$  is an integer that scales logarithmically with  $n$ . We recall that  $t_n := \lfloor r \log n \rfloor$  with  $r > 2(\log(1/\rho))^{-1}$  where  $\rho \in (0, 1)$  is a constant characterizing the uniform ergodicity of the Markov chain (see Assumption 1). By convention, we assume here that for all  $k < 1$ ,  $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot]$ . Hence the first term that we will consider is given by

$$U_n = \sum_{1 \leq i < j \leq n} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j),$$

where for all  $x, y, z \in E$ ,  $h_{i,j}^{(0)}(x, y, z) = h_{i,j}(x, z) - \int_w h_{i,j}(x, w) P(y, dw)$ .

We provide a detailed proof of a concentration result for  $U_n$  by taking advantage of its martingale structure. Reasoning by induction, we show that the  $t_n - 1$  following terms involved in the decomposition (7) of  $U_{\text{stat}}(n)$  can be handled using a similar approach. Since the last term  $R_{\text{stat}}^{(t_n)}(n)$  of the decomposition (7) has not a martingale property, another argument is required. We deal with  $R_{\text{stat}}^{(t_n)}(n)$  exploiting the uniform ergodicity of the Markov chain  $(X_i)_{i \geq 1}$  which is guaranteed by Assumption 1 (see [33, Theorem 8]).

The cornerstones of our approach are the following two propositions whose proofs are postponed to Section 4.2 and Section 4.3 respectively.

**Proposition 2.** Let  $n \geq 2$ . We keep the notations of Sections 2 and 3.1. We suppose Assumptions 1, 2 and 3 described in Section 2. There exist two constants  $\beta, \kappa > 0$  such that for any  $u > 0$ ,

a) if Assumption 4.(i) is satisfied, it holds with probability at least  $1 - \beta e^{-u} \log(n)$ ,

$$M_{\text{stat}}^{(t_n)}(n) \leq \kappa \log(n) \left( [A\sqrt{n} \log n] \sqrt{u} + [A + B_n \sqrt{n}] u + [2A\sqrt{n}] u^{3/2} + Au^2 \right).$$

b) if Assumption 4.(ii) is satisfied, it holds with probability at least  $1 - \beta e^{-u} \log(n)$ ,

$$M_{\text{stat}}^{(t_n)}(n) \leq \kappa \log(n) \left( [C_n + A\sqrt{n} \log n] \sqrt{u} + [A + B_n \sqrt{n}] u + [2A\sqrt{n}] u^{3/2} + Au^2 \right).$$

**Proposition 3.** Let  $n \geq 2$ . We keep the notations of Sections 2 and 3.1. We suppose Assumptions 1, 2 and 3. Then

a)  $R_{\text{stat}}^{(t_n)}(n) \leq A(2L + nt_n)$ .

b) if the Markov chain  $(X_i)_{i \geq 1}$  is stationary,  $R_{\text{stat}}^{(t_n)}(n) \leq 2LA(1 + t_n + t_n^2)$ .

#### 4.1.1 Proof of Theorem 1

We suppose Assumptions 1, 2, 3 and 4.(i) (respectively 4.(ii)). From the decomposition (7) coupled with Proposition 2.a) (respectively Proposition 2.b)) and Proposition 3.a) (respectively Theorem 1.a) (respectively Theorem 1.b)) is straightforward.

#### 4.1.2 Proof of Theorem 2

We suppose Assumptions 1, 2 and 3. We assume in addition that the Markov chain is stationary which implies in particular that Assumption 4.(ii) holds. From the decomposition (7) coupled with Proposition 2.b) and Proposition 3.b), the result of Theorem 2 is straightforward. Note that in case Assumption 4.(i) holds, the quantity  $C_n$  (involved in the sub-Gaussian regime of the tail) can be removed from the inequality by simply using Proposition 2.a) rather than Proposition 2.b).

### 4.2 Proof of Proposition 2

Let us recall that Proposition 2 requires either a mild condition on the initial distribution of the Markov chain or the fact that the kernels  $h_{i,j}$  do not depend on  $i$  (see Assumption 4). One only needs to consider different Bernstein concentration inequalities for sums of functions of Markov chains to go from one result to the other. In this section, we give the proof of Proposition 2 in the case where Assumption 4.(i) holds. We specify the part of the proof that should be changed to get the result when  $h_{i,j}$  may depend on both  $i$  and  $j$  and when Assumption 4.(ii) holds. We make this easily identifiable using the symbol  $\hat{\otimes}$ .

#### 4.2.1 Concentration of the first term of the decomposition of the U-statistic

**Martingale structure of the U-statistic** Defining  $Y_j = \sum_{i=1}^{j-1} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j)$ ,  $U_n$  can be written as  $U_n = \sum_{j=2}^n Y_j$ . Since

$$\mathbb{E}_{j-1}[Y_j] = \mathbb{E}[Y_j | X_1, \dots, X_{j-1}] = 0,$$

we know that  $(U_k)_{k \geq 2}$  is a martingale relative to the  $\sigma$ -algebras  $G_l$ ,  $l \geq 2$ . This martingale can be extended to  $n = 0$  and  $n = 1$  by taking  $U_0 = U_1 = 0$ ,  $G_0 = \{\emptyset, E\}$ ,  $G_1 = \sigma(X_1)$ . We will use the martingale structure of  $(U_n)_n$  through the following Lemma.

**Lemma 1.** (cf. [21, Lemma 3.4.6])

Let  $(U_m, G_m)$ ,  $m \in \mathbb{N}$ , be a martingale with respect to a filtration  $G_m$  such that  $U_0 = U_1 = 0$ . For each  $m \geq 1$  and  $k \geq 2$ , define the angle brackets  $A_m^k = A_m^k(U)$  of the martingale  $U$  by

$$A_m^k = \sum_{i=1}^m \mathbb{E}_{i-1}[(U_i - U_{i-1})^k]$$

(and note  $A_1^k = 0$  for all  $k$ ). Suppose that for  $\alpha > 0$  and all  $i \geq 1$ ,  $\mathbb{E}[e^{\alpha|U_i - U_{i-1}|}] < \infty$ . Then

$$\left( \varepsilon_m := e^{\alpha U_m - \sum_{k \geq 2} \alpha^k A_m^k / k!}, G_m \right), \quad m \in \mathbb{N},$$

is a supermartingale. In particular,  $\mathbb{E}[\varepsilon_m] \leq \mathbb{E}[\varepsilon_1] = 1$ , so that, if  $A_m^k \leq w_m^k$  for constants  $w_m^k \geq 0$ ; then

$$\mathbb{E}[e^{\alpha U_m}] \leq e^{\sum_{k \geq 2} \alpha^k w_m^k / k!}.$$

We will also use the following convexity result several times.

**Lemma 2.** [21, page 179] For all  $\theta_1, \theta_2, \varepsilon \geq 0$ , and for all integer  $k \geq 1$ ,

$$(\theta_1 + \theta_2)^k \leq (1 + \varepsilon)^{k-1} \theta_1^k + (1 + \varepsilon^{-1})^{k-1} \theta_2^k.$$

For all  $k \geq 2$  and  $n \geq 1$ , we have using Assumption 3:

$$\begin{aligned} A_n^k &= \sum_{j=2}^n \mathbb{E}_{j-1} \left[ \left| \sum_{i=1}^{j-1} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j) \right|^k \right] \leq V_n^k := \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j) \right|^k \\ &= \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} (h_{i,j}(X_i, X_j) - \mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] + \mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] - \mathbb{E}_{j-1} [h_{i,j}(X_i, X_j)]) \right|^k \\ &= \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} (p_{i,j}(X_i, X_j) + m_{i,j}(X_i, X_{j-1})) \right|^k, \end{aligned}$$

$$\text{where } p_{i,j}(x, z) = h_{i,j}(x, z) - \mathbb{E}_{\pi} [h_{i,j}] \quad \text{and} \quad m_{i,j}(x, y) = \mathbb{E}_{\pi} [h_{i,j}] - \int_z h_{i,j}(x, z) P(y, dz).$$

Using Lemma 2 with  $\varepsilon = 1/2$ , we deduce that

$$\begin{aligned} V_n^k &\leq \sum_{j=2}^n \mathbb{E}_{j-1} \left( \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X_j) \right|^k + \left| \sum_{i=1}^{j-1} m_{i,j}(X_i, X_{j-1}) \right|^k \right) \\ &\leq \left( \frac{3}{2} \right)^{k-1} \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X_j) \right|^k + 3^{k-1} \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} m_{i,j}(X_i, X_{j-1}) \right|^k. \end{aligned}$$

Let us remark that

$$\begin{aligned} \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} m_{i,j}(X_i, X_{j-1}) \right|^k &= \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} (\mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] - \mathbb{E}_{j-1} [h_{i,j}(X_i, X_j)]) \right|^k \\ &= \sum_{j=2}^n \left| \sum_{i=1}^{j-1} (\mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] - \mathbb{E}_{j-1} [h_{i,j}(X_i, X_j)]) \right|^k \\ &= \sum_{j=2}^n \left| \mathbb{E}_{j-1} \left[ \sum_{i=1}^{j-1} (\mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] - h_{i,j}(X_i, X_j)) \right] \right|^k \\ &\leq \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} (\mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] - h_{i,j}(X_i, X_j)) \right|^k, \end{aligned}$$

where the last inequality comes from Jensen's inequality. We obtain the following upper-bound for  $V_n^k$ ,

$$V_n^k \leq 2 \times 3^{k-1} \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X_j) \right|^k \leq 2 \times 3^{k-1} \delta_M \sum_{j=2}^n \mathbb{E}_{X'_j} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j) \right|^k,$$

where the random variables  $(X'_j)_j$  are i.i.d. with distribution  $\nu$  (see Assumption 2).  $\mathbb{E}_{X'_j}$  denotes the expectation on the random variable  $X'_j$ .

**Lemma 3.** (cf. [21, Ex.1 Section 3.4]) Let  $Z_j$  be independent random variables with respective probability laws  $P_j$ . Let  $k > 1$ , and consider functions  $f_1, \dots, f_N$  where for all  $j \in [N]$ ,  $f_j \in L^k(P_j)$ . Then the duality of  $L^p$  spaces and the independence of the variables  $Z_j$  imply that

$$\left( \sum_{j=1}^N \mathbb{E}[|f_j(Z_j)|^k] \right)^{1/k} = \sup_{\sum_{j=1}^N \mathbb{E}|\xi_j(Z_j)|^{k/(k-1)}=1} \sum_{j=1}^N \mathbb{E}[f_j(Z_j)\xi_j(Z_j)],$$

where the sup runs over  $\xi_j \in L^{k/(k-1)}(P_j)$ .

Then by the duality result of Lemma 3,

$$\begin{aligned} (V_n^k)^{1/k} &\leq \left( 2\delta_M \times 3^{k-1} \sum_{j=2}^n \mathbb{E}_{X'_j} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j) \right|^k \right)^{1/k} \\ &\leq (2\delta_M)^{1/k} \sup_{\xi \in \mathcal{L}_k} \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbb{E}_{X'_j} \left[ p_{i,j}(X_i, X'_j) \xi_j(X'_j) \right] \\ \text{where } \mathcal{L}_k &= \left\{ \xi = (\xi_2, \dots, \xi_n) \text{ s.t. } \forall 2 \leq j \leq n, \xi_j \in L^{k/(k-1)}(\nu) \text{ with } \sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)} = 1 \right\}. \\ &= (2\delta_M)^{1/k} \sup_{\xi \in \mathcal{L}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}_{X'_j} \left[ p_{i,j}(X_i, X'_j) \xi_j(X'_j) \right] \end{aligned}$$

Let us denote by  $F$  the subset of the set  $\mathcal{F}(E, \mathbb{R})$  of all measurable functions from  $(E, \Sigma)$  to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  that are bounded by  $A$ . We set  $S := E \times F^{n-1}$ . For all  $i \in [n]$ , we define  $W_i$  by

$$W_i := \left( X_i, \underbrace{0, \dots, 0}_{(i-1) \text{ times}}, p_{i,i+1}(X_i, \cdot), p_{i,i+2}(X_i, \cdot), \dots, p_{i,n}(X_i, \cdot) \right) \in S.$$

Hence for all  $i \in [n]$ ,  $W_i$  is  $\sigma(X_i)$ -measurable. We define for any  $\xi = (\xi_2, \dots, \xi_n) \in \prod_{i=2}^n L^{k/(k-1)}(\nu)$  the function

$$\forall w = (x, p_2, \dots, p_n) \in S, \quad f_\xi(w) = \sum_{j=2}^n \int p_j(y) \xi_j(y) d\nu(y).$$

Then setting  $\mathcal{F} = \{f_\xi : \sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)} = 1\}$ , we have

$$(V_n^k)^{1/k} \leq (2\delta_M)^{1/k} \sup_{f_\xi \in \mathcal{F}} \sum_{i=1}^{n-1} f_\xi(W_i).$$

By the separability of the  $L^p$  spaces of finite measures,  $\mathcal{F}$  can be replaced by a countable subset  $\mathcal{F}_0$ . To upper-bound the tail probabilities of  $U_n$ , we will bound the variable  $V_n^k$  on sets of large probability using Talagrand's inequality. Then we will use Lemma 1 on these sets by means of optional stopping.

**Application of Talagrand's inequality for Markov chains** The proof of Lemma 4 is provided in Section A of the Appendix (and relies mainly on the Talagrand concentration result from [34]).

**Lemma 4.** Let us denote

$$Z = \sup_{f_\xi \in \mathcal{F}} \sum_{i=1}^{n-1} f_\xi(W_i), \quad \sigma_k^2 = \mathbb{E} \left[ \sum_{i=1}^{n-1} \sup_{f_\xi \in \mathcal{F}} f_\xi(W_i)^2 \right] \quad \text{and} \quad b_k = \sup_{w \in S} \sup_{f_\xi \in \mathcal{F}} |f_\xi(w)|.$$

Then it holds for any  $t > 0$ ,

$$\mathbb{P}(Z > \mathbb{E}[Z] + t) \leq \exp \left( -\frac{1}{8\|\Gamma\|^2} \min \left( \frac{t^2}{4\sigma_k^2}, \frac{t}{b_k} \right) \right),$$

where  $\Gamma$  is a  $n \times n$  matrix defined in Section A which satisfies  $\|\Gamma\| \leq \frac{2L}{1-\rho}$ .

Using Lemma 4, we deduce that for any  $t > 0$ ,

$$\mathbb{P}\left((V_n^k)^{1/k} \geq (2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k}t\right) \leq \exp\left(-\frac{1}{8\|\Gamma\|^2} \min\left(\frac{t^2}{4\sigma_k^2}, \frac{t}{b_k}\right)\right),$$

which implies that for any  $x \geq 0$ ,

$$\mathbb{P}\left((V_n^k)^{1/k} \geq (2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k}2\sigma_k\sqrt{x} + (2\delta_M)^{1/k}b_kx\right) \leq \exp\left(-\frac{x}{8\|\Gamma\|^2}\right).$$

Using the change of variable  $x = k8\|\Gamma\|^2u$  with  $u \geq 0$  in the previous inequality leads to

$$\mathbb{P}\left(\bigcup_{k=2}^{\infty} (V_n^k)^{1/k} \geq (2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k}\sigma_k3\|\Gamma\|\sqrt{ku} + (2\delta_M)^{1/k}k8\|\Gamma\|^2b_ku\right) \leq 1.62e^{-u},$$

because

$$1 \wedge \sum_{k=2}^{\infty} \exp(-ku) \leq 1 \wedge \frac{1}{e^u(e^u - 1)} = \left(e^u \wedge \frac{1}{e^u - 1}\right) e^{-u} \leq \frac{1 + \sqrt{5}}{2} e^{-u} \leq 1.62e^{-u}.$$

Using Lemma 2 twice and using Holder inequality to bound  $b_k$  and  $\sigma_k^2$ , we obtain (9) from Lemma 5. The proof of Lemma 5 is postponed to Section B.

**Lemma 5.** For any  $u > 0$ , we denote

$$w_n^k := ((1 + \varepsilon)^{k-1}2\delta_M(\mathbb{E}[Z])^k + 2\delta_M(1 + \varepsilon^{-1})^{2k-2}(8\|\Gamma\|^2)^k(nA^2)A^{k-2}(ku)^k + (1 + \varepsilon)^{k-1}(1 + \varepsilon^{-1})^{k-1}2\delta_M(3\|\Gamma\|)^k\mathfrak{B}_0^2A^{k-2}(nku)^{k/2},$$

$$\text{with } \mathfrak{B}_0^2 := \max\left[\max_i \left\| \sum_{j=i+1}^n \mathbb{E}_{X \sim \nu} [p_{i,j}^2(\cdot, X)] \right\|_{\infty}, \max_j \left\| \sum_{i=1}^{j-1} \mathbb{E}_{X \sim \pi} [p_{i,j}^2(X, \cdot)] \right\|_{\infty}\right] \leq B_n^2, \quad (8)$$

where the dependence in  $u$  of  $w_n^k$  is leaved implicit. Then it holds

$$\mathbb{P}(V_n^k \leq w_n^k \quad \forall k \geq 2) \geq 1 - 1.62e^{-u}. \quad (9)$$

**Bounding  $(\mathbb{E}[Z])^k$ .**



The way we bound  $(\mathbb{E}[Z])^k$  is the only part of the proof that needs to be modified to get the concentration result when Assumption 4.(i) or Assumption 4.(ii) holds. This is where we can use different Bernstein concentration inequalities according to whether the splitting method is applicable or not (see Section 2.5 for details). Here we present the approach when  $h_{i,j} \equiv h_{1,j}$ ,  $\forall i, j$  (i.e. when Assumption 4.(i) is satisfied). We refer to Section C of the Appendix for the details regarding the way we bound  $(\mathbb{E}Z)^k$  when Assumption 4.(ii) holds.

Using Jensen inequality and Lemma 3, we obtain

$$\begin{aligned} (\mathbb{E}[Z])^k &\leq \mathbb{E}[Z^k] = \mathbb{E}\left[\left(\sup_{\xi \in \mathcal{L}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}_{X'_j} [p_{i,j}(X_i, X'_j)\xi_j(X'_j)]\right)^k\right] \\ &= \mathbb{E}\left[\sum_{j=2}^n \mathbb{E}_{X'_j} \left[\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right]\right] = \sum_{j=2}^n \mathbb{E}\left[\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right], \end{aligned}$$

where we recall that  $\mathbb{E}_{X'_j}$  denotes the expectation on the random variable  $X'_j$ . One can remark that conditionally to  $X'_j$ , the quantity  $\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)$  is a sum of function of the Markov chain  $(X_i)_{i \geq 1}$ . Hence to control this term, we apply a Bernstein inequality for Markov chains.



Let us consider some  $j \in [n]$  and some  $x \in E$ . Using the notations of Section 2.3, we define

$$\forall l \in \{0, \dots, n\}, \quad Z_l^j(x) = \sum_{i=m(S_l+1)}^{m(S_{l+1}+1)-1} p_{i,j}(X_i, x).$$

By convention, we set  $p_{i,j} \equiv 0$  for any  $i \geq j$ . Let us consider  $N_j = \sup\{i \in \mathbb{N} : mS_{i+1} + m - 1 \leq j - 1\}$ . Then using twice Lemma 2, we have

$$\begin{aligned} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, x) \right|^k &= \left| \sum_{l=0}^{N_j} Z_l^j(x) + \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right|^k \\ &\leq \left(\frac{3}{2}\right)^{k-1} \left| \sum_{l=1}^{N_j} Z_l^j(x) \right|^k + 3^{k-1} \left| \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right|^k \\ &\leq \left(\frac{9}{4}\right)^{k-1} \left| \sum_{l=0}^{\lfloor N_j/2 \rfloor} Z_{2l}^j(x) \right|^k + \left(\frac{9}{2}\right)^{k-1} \left| \sum_{l=0}^{\lfloor (N_j-1)/2 \rfloor} Z_{2l+1}^j(x) \right|^k + 3^{k-1} \left| \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right|^k. \end{aligned} \quad (10)$$

We have  $\left| \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right| \leq AmT_{N_j+1}$ . So using the definition of the Orlicz norm and the fact that the random variables  $(T_i)_{i \geq 2}$  are i.i.d., it holds for any  $t \geq 0$ ,

$$\begin{aligned} \mathbb{P} \left( \left| \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right| \geq t \right) &\leq \mathbb{P}(T_{N_j+1} \geq \frac{t}{Am}) \leq \mathbb{P}(\max(T_1, T_2) \geq \frac{t}{Am}) \\ &\leq \mathbb{P}(T_1 \geq \frac{t}{Am}) + \mathbb{P}(T_2 \geq \frac{t}{Am}) \leq 4 \exp\left(-\frac{t}{Am\tau}\right). \end{aligned}$$

Hence, using that for an exponential random variable  $G$  with parameter 1,  $\mathbb{E}[G^p] = p! \quad \forall p \geq 0$ ,

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right|^k \right] &= 4 \int_0^{+\infty} \mathbb{P} \left( \left| \sum_{i=m(S_{N_j+1})}^{j-1} p_{i,j}(X_i, x) \right| \geq t \right) dt \\ &\leq 4 \int_0^{+\infty} \exp\left(-\frac{t^{1/k}}{Am\tau}\right) dt \leq 4(Am\tau)^k \int_0^{+\infty} \exp(-v) k v^{k-1} dv = 4(Am\tau)^k k!, \end{aligned}$$

The random variable  $Z_{2l}^j(x)$  is  $\sigma(X_{m(S_{2l}+1)}, \dots, X_{m(S_{2l+1}+1)-1})$ -measurable. Let us insist that this holds because we consider that  $h_{i,j} \equiv h_{1,j}$ ,  $\forall i, j$  which implies that  $p_{i,j} \equiv p_{1,j}$ ,  $\forall i, j$ . Hence for any  $x \in E$ , the random variables  $(Z_{2l}^j(x))_l$  are independent (see Section 2.3). Moreover, one has that for any  $l$ ,  $\mathbb{E}[Z_{2l}^j(x)] = 0$ . This is due to [31, Eq.(17.23) Theorem 17.3.1] together with Assumption 3 which gives that  $\forall x' \in E$ ,  $\mathbb{E}_{X \sim \pi}[p_{i,j}(X, x')] = 0$ . Let us finally notice that for any  $x \in E$  and any  $l \geq 0$ ,  $|Z_{2l}^j(x)| \leq AmT_{2l+1}$ , so  $\|Z_{2l}^j(x)\|_{\psi_1} \leq Am \max(\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1}) \leq Am\tau$ . First, we use Lemma 6 to obtain that

$$\mathbb{E} \left[ \left| \sum_{l=0}^{\lfloor N_j/2 \rfloor} Z_{2l}^j(x) \right|^k \right] \leq \mathbb{E} \max_{0 \leq s \leq n-1} \left| \sum_{l=0}^s Z_{2l}^j(x) \right|^k \leq 2 \times 4^k \mathbb{E} \left[ \left| \sum_{l=0}^{n-1} Z_{2l}^j(x) \right|^k \right],$$

where for the last inequality we gathered (12) with the left hand side of (11) from Lemma 6.

**Lemma 6.** (cf. [11, Lemma 1.2.6])

Let us consider some separable Banach space  $B$  endowed with the norm  $\|\cdot\|$ . Let  $X_i$ ,  $i \leq n$ , be independent centered  $B$ -valued random variables with norms  $L_p$  for some  $p \geq 1$  and let  $\varepsilon_i$  be independent Rademacher random variables independent of the variables  $X_i$ . Then

$$2^{-p} \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i X_i \right\|^p \leq \mathbb{E} \left\| \sum_{i=1}^n X_i \right\|^p \leq 2^p \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i X_i \right\|^p, \quad (11)$$

$$\text{and } \mathbb{E} \max_{k \leq n} \left\| \sum_{i=1}^k X_i \right\|^p \leq 2^{p+1} \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i X_i \right\|^p \quad (12)$$

Similarly, the random variables  $(Z_{2l+1}^j(x))_l$  are independent and satisfy for any  $l$ ,  $\mathbb{E}[Z_{2l+1}^j(x)] = 0$ . With an analogous approach, we get that

$$\mathbb{E} \left| \sum_{l=0}^{\lfloor (N_j-1)/2 \rfloor} Z_{2l+1}^j(x) \right|^k \leq \mathbb{E} \max_{0 \leq s \leq n-1} \left| \sum_{l=0}^s Z_{2l+1}^j(x) \right|^k \leq 2 \times 4^k \mathbb{E} \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(x) \right|^k.$$

Let us denote for any  $j \in [n]$ ,  $\mathbb{E}_{|X'_j}$  the conditional expectation with respect to the  $\sigma$ -algebra  $\sigma(X'_j)$ . Coming back to (10), we proved that

$$\begin{aligned} \mathbb{E}_{|X'_j} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j) \right|^k &\leq \left(\frac{9}{4}\right)^{k-1} \mathbb{E}_{|X'_j} \left| \sum_{l=0}^{\lfloor N_j/2 \rfloor} Z_{2l}^j(X'_j) \right|^k \\ &+ \left(\frac{9}{2}\right)^{k-1} \mathbb{E}_{|X'_j} \left| \sum_{l=0}^{\lfloor (N_j-1)/2 \rfloor} Z_{2l+1}^j(X'_j) \right|^k + 3^{k-1} \mathbb{E}_{|X'_j} \left| \sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i, X'_j) \right|^k \\ &\leq 2 \times 9^k \mathbb{E}_{|X'_j} \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(X'_j) \right|^k + 2 \times 18^k \mathbb{E}_{|X'_j} \left| \sum_{l=0}^{n-1} Z_{2l}^j(X'_j) \right|^k + 4(3Am\tau)^k k!. \end{aligned} \quad (13)$$

It remains to bound the two expectations in (13). The two latter expectations will be controlled similarly and we give the details for the first one. We use the following Bernstein's inequality with the sequence of random variables  $(Z_{2l+1}^j(x))_l$ .

**Lemma 7.** (Bernstein's  $\psi_1$  inequality, [40, Lemma 2.2.11] and the subsequent remark).

If  $Y_1, \dots, Y_n$  are independent random variables such that  $\mathbb{E}Y_i = 0$  and  $\|Y_i\|_{\psi_1} \leq \tau$ , then for every  $t > 0$ ,

$$\mathbb{P} \left( \left| \sum_{i=1}^n Y_i \right| > t \right) \leq 2 \exp \left( -\frac{1}{K} \min \left( \frac{t^2}{n\tau^2}, \frac{t}{\tau} \right) \right),$$

for some universal constant  $K > 0$  ( $K = 8$  fits).

We obtain

$$\mathbb{P} \left( \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(x) \right| > t \right) \leq 2 \exp \left( -\frac{1}{K} \min \left( \frac{t^2}{nA^2m^2\tau^2}, \frac{t}{Am\tau} \right) \right).$$

We deduce that for any  $x \in E$ , any  $j \in [n]$  and any  $t \geq 0$ ,

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(x) \right|^k \right] &= \int_0^\infty \mathbb{P} \left( \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(x) \right| > t \right) dt \\ &= 2 \int_0^\infty \exp \left( -\frac{1}{K} \min \left( \frac{t^{2/k}}{nA^2m^2\tau^2}, \frac{t^{1/k}}{Am\tau} \right) \right) dt. \end{aligned}$$

Let us remark that  $\frac{t^{2/k}}{A^2m^2n\tau^2} \leq \frac{t^{1/k}}{Am\tau} \Leftrightarrow t \leq (nA\tau m)^k$ . Hence for any  $j \in [n]$ ,

$$\begin{aligned} &\mathbb{E} \left[ \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(X'_j) \right|^k \right] \\ &\leq 2 \int_0^{(nA\tau m)^k} \exp \left( -\frac{t^{2/k}}{KnA^2m^2\tau^2} \right) dt + 2 \int_0^\infty \exp \left( -\frac{t^{1/k}}{KAm\tau} \right) dt. \\ &\leq 2 \int_0^{n/K} \exp(-v) \frac{k}{2} v^{k/2-1} (\sqrt{K}n^{1/2}A\tau m)^k dv + 2 \int_0^\infty \exp(-v) kv^{k-1} (KAm\tau)^k dv. \end{aligned}$$

$$\begin{aligned}
&\leq 2 \int_0^{n/K} \exp(-v) \frac{k}{2} v^{k/2-1} (\sqrt{K} n^{1/2} A \tau m)^k dv + 2k \times (k-1)! (K A m \tau)^k \\
&\leq k (\sqrt{K} n^{1/2} A \tau m)^k \int_0^{n/K} \exp(-v) v^{k/2-1} dv + 2k! (K A m \tau)^k,
\end{aligned}$$

where we used again that if  $G$  is an exponential random variable with parameter 1, then for any  $p \in \mathbb{N}$ ,  $\mathbb{E}[G^p] = p!$ . Since for any real  $l \geq 1$ ,

$$\int_0^{\frac{n}{K}} e^{-v} v^{l-1} dv = \sum_{r=0}^{+\infty} \frac{(-1)^r}{r!} \int_0^{\frac{n}{K}} v^{r+l-1} dv = \sum_{r=0}^{+\infty} \frac{(-1)^r (\frac{n}{K})^{r+l}}{r!(r+l)} \leq \sum_{r=0}^{+\infty} \frac{(-1)^r (\frac{n}{K})^{r+l}}{r! \times l} \leq \frac{(\frac{n}{K})^l}{l} e^{-\frac{n}{K}},$$

we get that

$$k (\sqrt{K} n^{1/2} A \tau m)^k \int_0^{\frac{n}{K}} e^{-v} v^{k/2-1} dv \leq 2 (\sqrt{K} n^{1/2} A \tau m)^k e^{-\frac{n}{K}} \left(\frac{n}{K}\right)^{k/2} = 2 (n A \tau m)^k e^{-\frac{n}{K}}.$$

Hence we proved that for some universal constant  $K > 1$ ,

$$\mathbb{E} \left[ \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(x) \right|^k \right] \leq 2 (n A \tau m)^k e^{-n/K} + 2k! (K A m \tau)^k \leq 4k! (K A m \tau)^k,$$

since for all  $k \geq 2$ ,  $e^{-n/K} (n/K)^k / (k!) \leq 1$ . Using a similar approach, one can show the same bound for the second expectation in (13). We proved that for some universal constant  $K > 1$ ,

$$\begin{aligned}
(\mathbb{E}[Z])^k &\leq \sum_{j=2}^n \mathbb{E} \left[ \mathbb{E}_{|X'_j} \left| \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j) \right|^k \right] \leq 2 \times 9^k \sum_{j=2}^n \mathbb{E} \left[ \mathbb{E}_{|X'_j} \left| \sum_{l=0}^{n-1} Z_{2l+1}^j(X'_j) \right|^k \right] \\
&\quad + 2 \times 18^k \sum_{j=2}^n \mathbb{E} \left[ \mathbb{E}_{|X'_j} \left| \sum_{l=0}^{n-1} Z_{2l}^j(X'_j) \right|^k \right] + 4 \sum_{j=2}^n (3 A m \tau)^k k! \\
&\leq 2n \times 18^k \times 4k! (K A m \tau)^k + 4n (3 A m \tau)^k k! = 16n \times k! (K A m \tau)^k,
\end{aligned} \tag{14}$$

where in the last inequality, we still call  $K$  the universal constant defined by  $18K$ .

**Upper-bounding  $U_n$  using the martingale structure** Let

$$T + 1 := \inf\{l \in \mathbb{N} : V_l^k \geq w_n^k \text{ for some } k \geq 2\}.$$

Then, the event  $\{T \leq l\}$  depends only on  $X_1, \dots, X_l$  for all  $l \geq 1$ . Hence,  $T$  is a stopping time for the filtration  $(\mathcal{G}_l)_l$  where  $\mathcal{G}_l = \sigma((X_i)_{i \in [l]})$  and we deduce that  $U_l^T := U_{l \wedge T}$  for  $l = 0, \dots, n$  is a martingale with respect to  $(\mathcal{G}_l)_l$  with  $U_0^T = U_0 = 0$  and  $U_1^T = U_1 = 0$ . We remark that  $U_j^T - U_{j-1}^T = U_j - U_{j-1}$  if  $T \geq j$  and zero otherwise, and that  $\{T \geq j\}$  is  $\mathcal{G}_{j-1}$  measurable. Then, the angle brackets of this martingale admit the following bound:

$$\begin{aligned}
A_n^k(U^T) &= \sum_{j=2}^n \mathbb{E}_{j-1} [(U_j^T - U_{j-1}^T)^k] \\
&\leq \sum_{j=2}^n \mathbb{E}_{j-1} |U_j - U_{j-1}|^k \mathbf{1}_{T \geq j} = \sum_{j=2}^n \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} h^{(0)}(X_i, X_{j-1}, X_j) \right|^k \mathbf{1}_{T \geq j} \\
&= \sum_{j=2}^{n-1} V_j^k \mathbf{1}_{T=j} + V_n^k \mathbf{1}_{T \geq n} \leq w_n^k \left( \sum_{j=2}^{n-1} \mathbf{1}_{T=j} + \mathbf{1}_{T \geq n} \right) \leq w_n^k,
\end{aligned}$$

since, by definition of  $T$ ,  $V_j^k \leq w_n^k$  for all  $k$  on  $\{T \geq j\}$ . Hence, Lemma 1 applied to the martingale  $U_n^T$  implies

$$\mathbb{E}e^{\alpha U_n^T} \leq \exp\left(\sum_{k \geq 2} \frac{\alpha^k}{k!} w_n^k\right).$$

Also, since  $V_n^k$  is nondecreasing in  $n$  for each  $k$ , inequality (9) implies that

$$\mathbb{P}(T < n) \leq \mathbb{P}(V_n^k \geq w_n^k \text{ for some } k \geq 2) \leq 1.62e^{-u}.$$

Thus we deduce that for all  $s \geq 0$ ,

$$\mathbb{P}(U_n \geq s) \leq \mathbb{P}(U_n^T \geq s, T \geq n) + \mathbb{P}(T < n) \leq e^{-\alpha s} \exp\left(\sum_{k \geq 2} \frac{\alpha^k}{k!} w_n^k\right) + 1.62e^{-u}. \quad (15)$$

The final step of the proof consists in simplifying  $\exp\left(\sum_{k \geq 2} \frac{\alpha^k}{k!} w_n^k\right)$ .

$$\begin{aligned} \sum_{k \geq 2} \frac{\alpha^k}{k!} w_n^k &= 2\delta_M \sum_{k \geq 2} \frac{\alpha^k}{k!} (1 + \varepsilon)^{k-1} (\mathbb{E}[Z])^k \\ &\quad + 2\delta_M \sum_{k \geq 2} \frac{\alpha^k}{k!} (2 + \varepsilon + \varepsilon^{-1})^{k-1} (3\|\Gamma\|)^k \mathfrak{B}_0^2 A^{k-2} (nku)^{k/2} \\ &\quad + 2\delta_M \sum_{k \geq 2} \frac{\alpha^k}{k!} (1 + \varepsilon^{-1})^{2k-2} (8\|\Gamma\|^2)^k (nA^2) A^{k-2} (ku)^k =: a_1 + a_2 + a_3. \end{aligned} \quad (16)$$

Using the bound Eq.(14) obtained on  $(\mathbb{E}Z)^k$ , Lemma 8 bounds the three sums  $a_1, a_2$  and  $a_3$ .

**Lemma 8.**  $\exp\left(\sum_{k \geq 2} \frac{\alpha^k}{k!} w_n^k\right) \leq \exp\left(\frac{\alpha^2 W^2}{1 - \alpha c}\right)$  where

$$\begin{aligned} W &= 6\sqrt{\delta_M} (1 + \varepsilon)^{1/2} n^{1/2} K A \tau m \\ &\quad + \sqrt{2\delta_M} (2 + \varepsilon + \varepsilon^{-1})^{1/2} 3\|\Gamma\| \mathfrak{B}_0 \sqrt{nu} + \sqrt{2\delta_M} A (1 + \varepsilon^{-1}) 8\|\Gamma\|^2 \sqrt{neu}, \\ \text{and } c &= \max\left[(1 + \varepsilon) K A \tau m, (2 + \varepsilon + \varepsilon^{-1}) (3\|\Gamma\|) A (nu)^{1/2}, (1 + \varepsilon^{-1})^2 (8\|\Gamma\|^2) A e u\right]. \end{aligned}$$

The proof of Lemma 8 can be found in Section D Using the result from Lemma 8 in (15) and taking  $s = 2W\sqrt{u} + cu$  and  $\alpha = \sqrt{u}/(W + c\sqrt{u})$  in this inequality yields

$$\mathbb{P}(U_n \geq 2W\sqrt{u} + cu) \leq e^{-u} + 1.62e^{-u} \leq (1 + e)e^{-u}.$$

By taking  $\varepsilon = 1/2$ , we deduce that for any  $u \geq 0$ , it holds with probability at least  $1 - (1 + e)e^{-u}$

$$\begin{aligned} \sum_{i < j} h_j^{(0)}(X_i, X_{j-1}, X_j) &\leq 12\sqrt{\delta_M} K A \tau m \sqrt{nu} + 18\sqrt{\delta_M} \|\Gamma\| \mathfrak{B}_0 \sqrt{nu} \\ &\quad + 100\sqrt{\delta_M} \|\Gamma\|^2 A \sqrt{neu}^{3/2} + 3K A \tau m u + 27A \|\Gamma\| \sqrt{nu}^{3/2} + 72A \|\Gamma\|^2 e u^2, \end{aligned}$$

Denoting  $\kappa := \max(12\sqrt{\delta_M} K A \tau m, 18\sqrt{\delta_M} \|\Gamma\|, 100\sqrt{\delta_M} \|\Gamma\|^2 e, 3K A \tau m, 72\|\Gamma\|^2 e)$ , we have with probability at least  $1 - (1 + e)e^{-u}$

$$\sum_{i < j} h_j^{(0)}(X_i, X_{j-1}, X_j) \leq \kappa (A\sqrt{n}\sqrt{u} + (A + \mathfrak{B}_0\sqrt{n})u + 2A\sqrt{nu}^{3/2} + Au^2).$$

### 4.2.2 Reasoning by descending induction with a logarithmic depth

As previously explained, we apply a proof similar to the one of the previous subsection on the  $t_n := \lfloor r \log n \rfloor$  terms in the sum  $M_{\text{stat}}^{(t_n)}(n)$  (see (7)), with  $r > 2(\log(1/\rho))^{-1}$ . Let us give the key elements to justify such approach by considering the second term of the sum  $M_{\text{stat}}^{(t_n)}(n)$ , namely

$$\begin{aligned} & \sum_{i < j} (\mathbb{E}_{j-1} [h_{i,j}(X_i, X_j)] - \mathbb{E}_{j-2} [h_{i,j}(X_i, X_j)]) \\ &= \underbrace{\sum_{i=1}^{n-2} \sum_{j=i+2}^n h_{i,j}^{(1)}(X_i, X_{j-2}, X_{j-1})}_{=: U_{n-1}^{(1)}} + \underbrace{\sum_{i=1}^{n-1} \{ \mathbb{E}_i [h_{i,i+1}(X_i, X_{i+1})] - \mathbb{E}_{i-1} [h_{i,i+1}(X_i, X_{i+1})] \}}_{=: (*)} \end{aligned} \quad (17)$$

where  $h_{i,j}^{(1)}(x, y, z) = \int_w h_{i,j}(x, w)P(z, dw) - \int_w h_{i,j}(x, w)P^2(y, dw)$ . Using McDiarmid's inequality for Markov chain (see [32, Corollary 2.10 and Remark 2.11]), we obtain Lemma 9.

**Lemma 9.** *Let us consider  $l \in \{1, \dots, t_n\}$ . For any  $u > 0$ , it holds with probability at least  $1 - 2e^{-u}$ ,*

$$\left| \sum_{i=1}^{n-1} \sum_{j=i+1}^{(i+l) \wedge n} (\mathbb{E}_{j-l} [h_{i,j}(X_i, X_j)] - \mathbb{E}_{j-l-1} [h_{i,j}(X_i, X_j)]) \right| \leq 3At_n \sqrt{t_{\text{mix}} nu},$$

where  $t_{\text{mix}}$  is the mixing time of the Markov chain and is given by

$$t_{\text{mix}} := \min \left\{ t \geq 0 : \sup_x \|P^t(x, \cdot) - \pi\|_{\text{TV}} < \frac{1}{4} \right\}.$$

Lemma 9 allows to bound (\*) in (17) (by choosing  $l = 1$ ). Now we aim at proving a concentration result for the term

$$U_{n-1}^{(1)} = \sum_{j=2}^{n-1} \sum_{i=1}^{j-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j),$$

using an approach similar to the one of the previous subsection. We state here the concentration result for  $U_{n-1}^{(1)}$  with Lemma 10 and a proof can be found in Section E.

**Lemma 10.** *For any  $u > 0$ , it holds with probability at least  $1 - (1 + e)e^{-u}$ ,*

$$U_{n-1}^{(1)} = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j) \leq \kappa (A\sqrt{n}\sqrt{u} + (A + B_n\sqrt{n})u + 2A\sqrt{nu}^{3/2} + Au^2)$$

Going back to (17) and using both Lemmas 9 and 10, we get that for any  $u > 0$ , it holds with probability at least  $1 - (1 + e + 2)e^{-u}$ ,

$$\begin{aligned} & \sum_{i < j} (\mathbb{E}_{j-1} [h_{i,j}(X_i, X_j)] - \mathbb{E}_{j-2} [h_{i,j}(X_i, X_j)]) \\ & \leq \kappa (A\sqrt{n}\sqrt{u} + (A + B_n\sqrt{n})u + 2A\sqrt{nu}^{3/2} + Au^2) + 3At_n \sqrt{t_{\text{mix}} nu} \end{aligned} \quad (18)$$

One can do the same analysis for the  $t_n$  first terms in the decomposition (7). Still denoting  $\kappa$  the constant  $\kappa + 3\sqrt{t_{\text{mix}}}$ , we get that for any  $u > 0$  it holds with probability at least  $1 - (3 + e)e^{-u}t_n$ ,

$$M_{\text{stat}}^{(t_n)}(n) \leq \kappa t_n (At_n \sqrt{n}\sqrt{u} + (A + B_n\sqrt{n})u + 2A\sqrt{nu}^{3/2} + Au^2).$$

### 4.3 Proof of Proposition 3

In the following, we assume that  $t_n \leq n$ , otherwise  $R_{\text{stat}}^{(t_n)}(n)$  is an empty sum. Using our convention which states that for all  $k < 1$ ,  $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot]$ , we need to control

$$\left| R_{\text{stat}}^{(t_n)}(n) \right| = \left| \sum_{i < j} (\mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)]) \right| \leq (1) + (2), \quad (19)$$

with denoting  $H_{i,j} = \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)]$ ,

$$(1) := \left| \sum_{i=1}^{n-t_n} \sum_{j=t_n+1}^n H_{i,j} \right| = \left| \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} H_{i,j} \right|$$

and  $(2) := \left| \sum_{i=1}^{n-1} \sum_{j=i+1}^{(i+t_n-1) \wedge n} H_{i,j} \right| = \left| \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-1} H_{i,j} \right|.$

We start by bounding the term (1) regardless of the initial distribution of the chain. We will bound in different ways the term (2) depending on whether the Markov chain is stationary or not. Let us first bound the term (1) splitting it into two terms,

$$(1) = \left| \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right| \leq (1a) + (1b).$$

Using Assumption 3, it holds  $\mathbb{E}_\pi [h_{i,j}] = \mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] = \int_x h_{i,j}(X_i, x) d\pi(x)$ . Hence we get that

$$(1a) := \left| \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E}_\pi [h_{i,j}] \right|$$

$$\leq \sum_{j=t_n+1}^n \left| \int_{x_j} \sum_{i=1}^{j-t_n} h_{i,j}(X_i, x_j) (P^{t_n}(X_{j-t_n}, dx_j) - d\pi(x_j)) \right|$$

$$\leq \sum_{j=t_n+1}^n \sup_{x_j} \left| \sum_{i=1}^{j-t_n} h_{i,j}(X_i, x_j) \right| \sup_z \|P^{t_n}(z, \cdot) - \pi\|_{\text{TV}}$$

$$\leq \sum_{j=t_n+1}^n \sup_{x_j} \left| \sum_{i=1}^{j-t_n} h_{i,j}(X_i, x_j) \right| L \rho^{t_n} \leq \sum_{j=t_n+1}^n \sup_{x_j} \left| \sum_{i=1}^{j-t_n} h_{i,j}(X_i, x_j) \right| L \frac{1}{n^2} \leq LA,$$

where in the penultimate inequality we used that  $\rho^{t_n} \leq \rho^{r \log(n)} = n^{r \log(\rho)} \leq n^{-2}$ . Indeed  $2+r \log(\rho) < 0$  because we choose  $r$  such that  $r > 2(\log(1/\rho))^{-1}$ .

Using again Assumption 3, it holds  $\mathbb{E}_\pi [h_{i,j}] = \int_{x_i} \chi P^i(dx_i) \int_x h_{i,j}(x_i, x) d\pi(x)$  where  $\chi$  is the initial distribution of the Markov chain  $(X_i)_{i \geq 1}$ . We get that

$$(1b) := \left| \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \mathbb{E}_\pi [h_{i,j}] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right|$$

$$\leq \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \left| \int_{x_i} \int_{x_j} h_{i,j}(x_i, x_j) \chi P^i(dx_i) (P^{j-i}(x_i, dx_j) - d\pi(x_j)) \right|$$

$$\leq \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty \underbrace{\int_{x_i} \chi P^i(dx_i)}_{=1} \sup_z \int_{x_j} |P^{j-i}(z, dx_j) - d\pi(x_j)|$$

$$\leq \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty L \rho^{j-i} \leq \sum_{j=t_n+1}^n \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty L \rho^{t_n} \leq LA,$$

where in the penultimate inequality we used that  $\rho^{t_n} \leq \rho^{r \log(n)} = n^{r \log(\rho)} \leq n^{-2}$ .

#### 4.3.1 Bounding (2) without stationarity

Without assuming that the Markov chain is stationary, we bound coarsely (2) as follows

$$(2) = \left| \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-1} \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right| \leq Ant_n.$$

This concludes the proof of Proposition 3.a) since we obtain,  $R_{\text{stat}}^{(t_n)}(n) \leq A(2L + nt_n)$ .

### 4.3.2 Bounding (2) with stationarity

Considering now that the chain is stationary, we split (2) into three different contributions.

$$\begin{aligned}
(2) &= \left| \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{j-1} \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right| \leq (2a) + (2b) + (2c), \\
\text{with } (2a) &:= \left| \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E}_\pi [h_{i,j}] \right|, \\
(2b) &:= \left| \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \mathbb{E}_\pi [h_{i,j}] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right|, \\
\text{and } (2c) &:= \left| \sum_{j=2}^n \sum_{i=(\lfloor \frac{j-t_n}{2} \rfloor + 1)\vee 1}^{j-1} \mathbb{E}_{j-t_n} [h_{i,j}(X_i, X_j)] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right|.
\end{aligned}$$

⚡ The only place where we use the stationarity of the chain is to bound the terms (2b) and (2c) by writing that  $\mathbb{E}[h_{i,j}(X_i, X_j)] = \int_{x_i} \int_{x_j} d\pi(x_i) P^{j-i}(x_i, dx_j)$ . Both are bounded using similar ideas, that is why we show here how to deal with (2b) and we postpone the proof of Lemma 11 to Section F.

**Lemma 11.** *It holds  $(2a) \leq LAt_n$  and  $(2c) \leq 2LAt_n^2$ .*

Let us now bound the term (2b),

$$\begin{aligned}
(2b) &:= \left| \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \mathbb{E}_\pi [h_{i,j}] - \mathbb{E} [h_{i,j}(X_i, X_j)] \right| \\
&\leq \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \left| \int_{x_i} \int_{x_j} h_{i,j}(x_i, x_j) d\pi(x_i) (d\pi(x_j) - P^{j-i}(x_i, dx_j)) \right| \\
&\leq \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \|h_{i,j}\|_\infty \underbrace{\int_{x_i} d\pi(x_i) \sup_y \int_{x_j} |d\pi(x_j) - P^{j-i}(y, dx_j)|}_{=1 \quad =\sup_y \|P^{j-i}(y, \cdot) - \pi\|_{\text{TV}}} \\
&\leq \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \|h_{i,j}\|_\infty L\rho^{j-i} \leq \sum_{j=2}^n \sum_{i=(j-t_n+1)\vee 1}^{\lfloor \frac{j-t_n}{2} \rfloor} \|h_{i,j}\|_\infty L\rho^{t_n/2} \leq LAt_n,
\end{aligned}$$

where we used that  $\rho^{t_n/2} \leq \rho^{r \log(n)/2} = n^{r \log(\rho)/2} \leq n^{-1}$ . Indeed  $1 + r \log(\rho)/2 < 0$  because we choose  $r$  such that  $r > 2(\log(1/\rho))^{-1}$ . Coming back to Eq.(19), we deduce that  $R_{\text{stat}}^{(t_n)}(n) \leq AL(2 + 2t_n + 2t_n^2)$  which concludes the proof of Proposition 3.

## Acknowledgements

This work was supported by a grant from Région Ile de France. The authors gratefully acknowledge the constructive comments and recommendations of the editor and of the anonymous reviewers which definitely help to improve the readability and quality of the paper.

## References

- [1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13(34):1000–1034, 2008.
- [2] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006.
- [3] P. Ango Nze. Critères d’ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation Markovienne. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(3):371–376, 1998.
- [4] M. A. Arcones and E. Giné. Limit theorems for  $U$ -processes. *Ann. Probab.*, 21(3):1494–1542, 1993.
- [5] E. Behrends. *Introduction to Markov chains*. Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig, 2000.
- [6] P. Bertail and S. Cléménçon. A renewal approach to Markovian  $U$ -statistics. *Math. Methods Statist.*, 20(2):79–105, 2011.
- [7] I. S. Borisov and N. V. Volodko. A note on exponential inequalities for the distribution tails of canonical von Mises’ statistics of dependent observations. *Statist. Probab. Lett.*, 96:287–291, 2015.
- [8] S. Cléménçon, P. Bertail, and G. Cioł ek. Statistical learning based on Markovian data maximal deviation inequalities and learning rates. *Ann. Math. Artif. Intell.*, 88(7):735–757, 2020.
- [9] Y. De Castro, C. Lacour, and T. M. Pham Ngoc. Adaptive estimation of nonparametric geometric graphs. *Math. Stat. Learn.*, 2(3):217–274, 2019.
- [10] V. H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate  $U$ -statistics. *Ann. Probab.*, 23(2):806–816, 1995.
- [11] V. de la Peña and E. Giné. Decoupling, from dependence to independence, randomly stopped processes,  $U$ -statistics and processes, martingales and beyond. *Journal of the American Statistical Association*, 95, 09 2000.
- [12] P. Del Moral and A. Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.*, 9(2):275–297, 1999.
- [13] P. Doukhan and M. Ghindès. Estimations dans le processus “ $X_{n+1} = f(X_n) + \varepsilon_n''$ ”. *C. R. Acad. Sci. Paris Sér. A-B*, 291(1):61–64, 1980.
- [14] Q. Duchemin and Y. De Castro. Markov random geometric graph, MRGG: A growth model for temporal dynamic networks. *Electron. J. Stat.*, 16(1):671–699, 2022.
- [15] P. Eichelsbacher and U. Schmock. Large deviations for products of empirical measures of dependent sequences. *Markov Process. Related Fields*, 7(3):435–468, 2001.
- [16] P. Eichelsbacher and U. Schmock. Rank-dependent moderate deviations of  $U$ -empirical measures in strong topologies. *Probab. Theory Related Fields*, 126(1):61–90, 2003.
- [17] J. Fan, B. Jiang, and Q. Sun. Hoeffding’s inequality for general Markov chains and its applications to statistical learning. *J. Mach. Learn. Res.*, 22(139):1–35, 2021.
- [18] D. Ferré, L. Hervé, and J. Ledoux. Limit theorems for stationary Markov processes with  $L^2$ -spectral gap. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(2):396–423, 2012.
- [19] G. Fort, E. Moulines, P. Priouret, and P. Vandekerkhove. A simple variance inequality for  $U$ -statistics of a Markov chain with applications. *Statist. Probab. Lett.*, 82(6):1193–1201, 2012.
- [20] E. Giné, R. Latał a, and J. Zinn. Exponential and moment inequalities for  $U$ -statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA, 2000.



- [21] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016.
- [22] F. Han. An exponential inequality for U-statistics under mixing conditions. *J. Theoret. Probab.*, 31(1):556–578, 2018.
- [23] F. Han and T. Qian. On inference validity of weighted U-statistics under data heterogeneity. *Electron. J. Stat.*, 12(2):2637–2708, 2018.
- [24] C. Houdré and P. Reynaud-Bouret. Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 55–69. Birkhäuser, Basel, 2003.
- [25] B. Jiang, Q. Sun, and J. Fan. Bernstein’s inequality for general Markov chains. *arXiv preprint arXiv:1805.10721*, 2018.
- [26] E. Joly and G. Lugosi. Robust estimation of U-statistics. *Stochastic Process. Appl.*, 126(12):3760–3773, 2016.
- [27] G. L. Jones. On the Markov chain central limit theorem. *Probab. Surveys*, 1:299–320, 2004.
- [28] M. Lerasle, N. M. Magalhães, and P. Reynaud-Bouret. Optimal kernel selection for density estimation. In *High dimensional probability VII*, volume 71 of *Progr. Probab.*, pages 425–460. Springer, [Cham], 2016.
- [29] F. Lindsten, R. Douc, and E. Moulines. Uniform ergodicity of the particle Gibbs sampler. *Scand. J. Stat.*, 42(3):775–797, 2015.
- [30] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, 2007.
- [31] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.
- [32] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 32, 2015.
- [33] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004.
- [34] P.-M. Samson. Concentration of measure inequalities for Markov chains and  $\Phi$ -mixing processes. *Ann. Probab.*, 28(1):416–461, 2000.
- [35] G. A. Satten, M. Kong, and S. Datta. Multisample adjusted U-statistics that account for confounding covariates. *Stat. Med.*, 37(23):3357–3372, 2018.
- [36] Y. Shen, F. Han, and D. Witten. Exponential inequalities for dependent V-statistics via random Fourier features. *Electron. J. Probab.*, 25:1–18, 2020.
- [37] G. S. Shieh, R. A. Johnson, and E. W. Frees. Testing independence of bivariate circular data and weighted degenerate U-statistics. *Statist. Sinica*, 4(2):729–747, 1994.
- [38] M. A. Suchard, R. E. Weiss, and J. S. Sinsheimer. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular biology and evolution*, 18(6):1001–1013, 2001.
- [39] R. S. Sutton and A. G. Barto. *Reinforcement learning: an introduction*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2018.
- [40] A. Van Der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 2013.

- [41] Y. Wang, R. Khardon, D. Pechyony, and R. Jones. Generalization Bounds for Online Learning Algorithms with Pairwise Loss Functions. In S. Mannor, N. Srebro, and R. C. Williamson, editors, *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 13.1–13.22, Edinburgh, Scotland, 2012. PMLR.
- [42] C. Wei, R. C. Elston, and Q. Lu. A weighted U statistic for association analyses considering genetic heterogeneity. *Stat. Med.*, 35(16):2802–2814, 2016.
- [43] E. Yilmaz, J. A. Aslam, and S. Robertson. A new rank correlation coefficient for information retrieval. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, page 587–594, New York, NY, USA, 2008. Association for Computing Machinery.

## A Proof of Lemma 4

In the section, we show that in the proof of Proposition 2, we can use the concentration inequality for the supremum of an empirical process of [34, Theorem 3].

Let us consider the sequence of random variables  $W = (W_1, \dots, W_n)$  on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  taking values in the measurable space  $S = E \times F^{n-1}$  where  $F$  is the subset of the set  $\mathcal{F}(E, \mathbb{R})$  of all measurable functions from  $(E, \Sigma)$  to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  that are bounded by  $A$ . Note that

$$\{0_{\mathcal{F}(E, \mathbb{R})}\} \cup \{p_{i,j}(x, \cdot) : x \in E, i, j \in [n]\} \subset F.$$

We define  $\mathcal{D} := \left\{ D \in \mathcal{P}(F) : \forall i \in [n-1], \forall j \in \{i+1, \dots, n\}, f_{i,j}^{-1}(D) \in \Sigma \right\}$  where  $\mathcal{P}(F)$  is the powerset of  $F$  and where  $\forall i \in [n-1], \forall j \in \{i+1, \dots, n\}$ ,

$$f_{i,j} : (E, \Sigma) \rightarrow (F, \mathcal{D}(F)) \quad x \mapsto p_{i,j}(x, \cdot).$$

Then we have the following straightforward result.

**Lemma 12.**  $\mathcal{D}$  is a  $\sigma$ -algebra on  $F$ .

In the following, we endow the space  $F$  with the  $\sigma$ -algebra  $\mathcal{D}$  and we consider on  $S$  the product  $\sigma$ -algebra given by  $\mathcal{S} := \sigma(\{C \times D_2 \times \dots \times D_n : C \in \Sigma, D_j \in \mathcal{D} \forall j \in \{2, \dots, n\}\})$ .

For all  $i \in [n]$ , we define  $W_i$  by

$$W_i := \left( X_i, \underbrace{0, \dots, 0}_{(i-1) \text{ times}}, p_{i,i+1}(X_i, \cdot), p_{i,i+2}(X_i, \cdot), \dots, p_{i,n}(X_i, \cdot) \right).$$

Hence for all  $i \in [n]$ ,  $W_i$  is  $\sigma(X_i)$ -measurable. Let us consider for any  $i \in [n-1]$ ,

$$\Phi_i : (E, \Sigma) \rightarrow (S, \mathcal{S}) \quad \text{such that} \quad \forall x \in E, \Phi_i(x) = \left( x, \underbrace{0_F, \dots, 0_F}_{(i-1) \text{ times}}, p_{i,i+1}(x, \cdot), \dots, p_{i,n}(x, \cdot) \right).$$

Then, one can directly see that for all  $i \in [n-1]$ ,  $W_i = \Phi_i(X_i)$  and by construction of  $\mathcal{D}$  and  $\mathcal{S}$ ,  $\Phi_i$  is measurable. Indeed, each coordinate of  $\Phi_i$  is measurable by construction of  $\mathcal{D}$  and this ensures that  $\Phi_i$  is measurable thanks to the following Lemma.

**Lemma 13.** (See [2, Lemma 4.49]) Let  $(X, \Sigma), (X_1, \Sigma_1)$  and  $(X_2, \Sigma_2)$  be measurable spaces, and let  $f_1 : X \rightarrow X_1$  and  $f_2 : X \rightarrow X_2$ . Define  $f : X \rightarrow X_1 \times X_2$  by  $f(x) = (f_1(x), f_2(x))$ . Then  $f : (X, \Sigma) \rightarrow (X_1 \times X_2, \Sigma_1 \otimes \Sigma_2)$  is measurable if and only if the two functions  $f_1 : (X, \Sigma) \rightarrow (X_1, \Sigma_1)$  and  $f_2 : (X, \Sigma) \rightarrow (X_2, \Sigma_2)$  are both measurable.

Then it holds for any  $i \in \{2, \dots, n-1\}$  and any  $G \in \mathcal{S}$ ,

$$\begin{aligned} \mathbb{P}(W_i \in G \mid W_{i-1}) &= \mathbb{P}(\Phi_i(X_i) \in G \mid W_{i-1}) = \mathbb{P}(\Phi_i(X_i) \in G \mid X_{i-1}) \\ &= \mathbb{P}(X_i \in \Phi_i^{-1}(G) \mid X_{i-1}) = P(X_{i-1}, \Phi_i^{-1}(G)) = [(\Phi_i)_\# P(X_{i-1}, \cdot)](G), \end{aligned} \quad (20)$$

where  $(\Phi_i)_\# P(X_{i-1}, \cdot)$  denotes the pushforward measure of the measure  $P(X_{i-1}, \cdot)$  by the measurable map  $\Phi_i$ . We deduce that  $W_i$  is non-homogeneous Markov chain. Moreover, (20) proves that the transition kernel of the Markov chain  $(W_k)_k$  from state  $i-1$  to state  $i$  is given by  $K^{(i-1, i)}$  where for all  $(x, p_2, \dots, p_n) \in S$  and for all  $G \in \mathcal{S}$ ,

$$K^{(i-1, i)}((x, p_2, \dots, p_n), G) = [(\Phi_i)_\# P(x, \cdot)](G).$$

One can easily generalize this notation. Let us consider some  $i, j \in [n]$  with  $i < j$  and let us denote by  $K^{(i, j)}$  the transition kernel of the Markov chain  $(W_k)_k$  from state  $i$  to state  $j$ . Then for all  $x \in E$ , for all  $p_2, \dots, p_n \in F$  and for all  $G \in \mathcal{S}$ ,

$$K^{(i, j)}((x, p_2, \dots, p_n), G) = [(\Phi_j)_\# P^{j-i}(x, \cdot)](G),$$

We introduce the mixing matrix  $\Gamma = (\gamma_{i,j})_{1 \leq i, j \leq n-1}$  where coefficients are defined by

$$\gamma_{i,j} := \sup_{w_i \in S} \sup_{z_i \in S} \|\mathcal{L}(W_j \mid W_i = w_i) - \mathcal{L}(W_j \mid W_i = z_i)\|_{\text{TV}}.$$

For any  $w \in S = E \times F^{n-1}$ , we denote by  $w^{(1)}$  the first coordinate of the vector  $w$ . Hence,  $w^{(1)}$  is an element of  $E$ . Then

$$\begin{aligned}
\gamma_{i,j} &= \sup_{w_i \in S} \sup_{z_i \in S} \sup_{G \in \mathcal{S}} \left| \left[ (\Phi_j)_{\#} P^{j-i}(w_i^{(1)}, \cdot) \right] (G) - \left[ (\Phi_j)_{\#} P^{j-i}(z_i^{(1)}, \cdot) \right] (G) \right| \\
&= \sup_{w_i \in S} \sup_{z_i \in S} \sup_{G \in \mathcal{S}} \left| P^{j-i}(w_i^{(1)}, \Phi_j^{-1}(G)) - P^{j-i}(z_i^{(1)}, \Phi_j^{-1}(G)) \right| \\
&\leq \sup_{w_i \in S} \sup_{z_i \in S} \sup_{C \in \Sigma} \left| P^{j-i}(w_i^{(1)}, C) - P^{j-i}(z_i^{(1)}, C) \right| \\
&= \sup_{x_i \in E} \sup_{x'_i \in E} \sup_{C \in \Sigma} \left| P^{j-i}(x_i, C) - P^{j-i}(x'_i, C) \right| = \sup_{x_i \in E} \sup_{x'_i \in E} \|P^{j-i}(x_i, \cdot) - \pi(\cdot) + \pi(\cdot) - P^{j-i}(x'_i, \cdot)\|_{\text{TV}} \\
&\leq \sup_{x_i \in E} \|P^{j-i}(x_i, \cdot) - \pi(\cdot)\|_{\text{TV}} + \sup_{x'_i \in E} \|P^{j-i}(x'_i, \cdot) - \pi(\cdot)\|_{\text{TV}} \leq 2L\rho^{j-i},
\end{aligned}$$

where in the first inequality we used that  $\Phi_j : (E, \Sigma) \rightarrow (S, \mathcal{S})$  is measurable and in the last inequality we used the uniform ergodicity of the Markov chain  $(X_i)_{i \geq 1}$ . We deduce that

$$\|\Gamma\| \leq 2L \left\| \text{Id} + \sum_{l=1}^{n-1} \rho^l N_l \right\|, \quad N_l = \left( n_{i,j}^{(l)} \right)_{1 \leq i,j \leq n-1} \text{ with } n_{i,j}^{(l)} = \begin{cases} 1 & \text{if } j-i=l \\ 0 & \text{otherwise.} \end{cases}$$

Note that  $N_l$  is a nilpotent matrix of order  $l$ . Since for each  $1 \leq l \leq n-1$ ,  $\|N_l\| \leq 1$ , it follows from the triangular inequality that

$$\|\Gamma\| \leq 2L \sum_{l=0}^{n-1} \rho^l \leq \frac{2L}{1-\rho}.$$

To conclude the proof and get the concentration result stated in Lemma 4, one only needs to apply [34, Theorem 3] with the class of functions  $\mathcal{F}$  and with the Markov chain  $(W_k)_k$ . Let us recall that  $\mathcal{F}$  is defined by  $\mathcal{F} = \{f_{\xi} : \sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)} = 1\}$  where for any  $\xi = (\xi_2, \dots, \xi_n) \in \prod_{i=2}^n L^{k/(k-1)}(\nu)$ ,

$$\forall w = (x, p_2, \dots, p_n) \in E \times F^{n-1}, \quad f_{\xi}(w) = \sum_{j=2}^n \int p_j(y) \xi_j(y) d\nu(y).$$

## B Proof of Lemma 5

**Bounding  $b_k$ .** Using Hölder's inequality we have,

$$\begin{aligned}
b_k &= \sup_{w \in S} \sup_{f_{\xi} \in \mathcal{F}} |f_{\xi}(w)| = \sup_{(p_2, \dots, p_n) \in F^{n-1}} \sup_{\xi \in \mathcal{L}_k} \sum_{j=2}^n \mathbb{E}[p_j(X'_j) \xi_j(X'_j)] \\
&\leq \sup_{(p_2, \dots, p_n) \in F^{n-1}} \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)} = 1} \sum_{j=2}^n \left( \mathbb{E} |p_j(X'_j)|^k \right)^{1/k} \left( \mathbb{E} |\xi_j(X'_j)|^{k/(k-1)} \right)^{(k-1)/k} \\
&\leq \sup_{(p_2, \dots, p_n) \in F^{n-1}} \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)} = 1} \left( \sum_{j=2}^n \mathbb{E} |p_j(X'_j)|^k \right)^{1/k} \left( \sum_{j=2}^n \mathbb{E} |\xi_j(X'_j)|^{k/(k-1)} \right)^{(k-1)/k} \\
&\leq \sup_{(p_2, \dots, p_n) \in F^{n-1}} \left( \sum_{j=2}^n \mathbb{E} |p_j(X'_j)|^k \right)^{1/k} \leq ((nA^2)A^{k-2})^{1/k},
\end{aligned}$$

where  $A := 2 \max_{i,j} \|h_{i,j}\|_{\infty}$  which satisfies  $\max_{i,j} \|p_{i,j}\|_{\infty} \leq A$ . Here, we used that  $F$  is the set of measurable functions from  $(E, \Sigma)$  to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  bounded by  $A$ .

**Bounding the variance.**

$$\begin{aligned}
\sigma_k^2 &= \mathbb{E} \left[ \sum_{i=1}^{n-1} \sup_{f_\xi \in \mathcal{F}} f_\xi(W_i)^2 \right] = \sum_{i=1}^{n-1} \mathbb{E} \left[ \sup_{\xi \in \mathcal{L}_k} \left( \sum_{j=i+1}^n \mathbb{E}_{X'_j} [p_{i,j}(X_i, X'_j) \xi_j(X'_j)] \right)^2 \right] \\
&= \sum_{i=1}^{n-1} \mathbb{E} \left[ \left( \sup_{\xi \in \mathcal{L}_k} \left| \sum_{j=i+1}^n \mathbb{E}_{X'_j} [p_{i,j}(X_i, X'_j) \xi_j(X'_j)] \right| \right)^2 \right] \leq n (\mathfrak{B}_0^2 A^{k-2})^{2/k},
\end{aligned}$$

where the last inequality comes from the following (where we use twice Holder's inequality),

$$\begin{aligned}
& \sup_{\xi \in \mathcal{L}_k} \left| \sum_{j=i+1}^n \mathbb{E}_{X'_j} [p_{i,j}(X_i, X'_j) \xi_j(X'_j)] \right| \\
& \leq \sup_{\xi \in \mathcal{L}_k} \sum_{j=i+1}^n \left( \mathbb{E}_{X'_j} |p_{i,j}(X_i, X'_j)|^k \right)^{1/k} \left( \mathbb{E} |\xi_j(X'_j)|^{k/(k-1)} \right)^{(k-1)/k} \\
& \leq \sup_{\sum_{j=2}^n \mathbb{E} |\xi_j(X'_j)|^{k/(k-1)} = 1} \left( \sum_{j=i+1}^n \mathbb{E}_{X'_j} |p_{i,j}(X_i, X'_j)|^k \right)^{1/k} \left( \sum_{j=i+1}^n \mathbb{E}_{X'_j} |\xi_j(X'_j)|^{k/(k-1)} \right)^{(k-1)/k} \\
& \leq \left( \sum_{j=i+1}^n \mathbb{E}_{X'_j} |p_{i,j}(X_i, X'_j)|^k \right)^{1/k} \leq (\mathfrak{B}_0^2 A^{k-2})^{1/k}, \quad \text{where } \mathfrak{B}_0 \text{ is defined in (8)}.
\end{aligned}$$

Using Lemma 2 twice and the bounds obtained on  $b_k$  and  $\sigma_k^2$  gives for  $u > 0$ ,

$$\begin{aligned}
& \left[ (2\delta_M)^{1/k} \mathbb{E}[Z] + (2\delta_M)^{1/k} \sigma_k 3 \|\Gamma\| \sqrt{ku} + (2\delta_M)^{1/k} k 8 \|\Gamma\|^2 b_k u \right]^k \\
& \leq \left[ (2\delta_M)^{1/k} \mathbb{E}[Z] + (2\delta_M)^{1/k} 3 \|\Gamma\| (\mathfrak{B}_0^2 A^{k-2})^{1/k} \sqrt{nk u} + (2\delta_M)^{1/k} 8 \|\Gamma\|^2 ((nA^2) A^{k-2})^{1/k} k u \right]^k \\
& \leq (1 + \varepsilon)^{k-1} 2\delta_M (\mathbb{E}[Z])^k + (1 + \varepsilon^{-1})^{k-1} \left[ (2\delta_M)^{1/k} 8 \|\Gamma\|^2 ((nA^2) A^{k-2})^{1/k} k u \right. \\
& \quad \left. + (2\delta_M)^{1/k} 3 \|\Gamma\| (\mathfrak{B}_0^2 A^{k-2})^{1/k} \sqrt{nk u} \right]^k \\
& \leq (1 + \varepsilon)^{k-1} 2\delta_M (\mathbb{E}[Z])^k + 2\delta_M (1 + \varepsilon^{-1})^{2k-2} (8 \|\Gamma\|^2)^k (nA^2) A^{k-2} (ku)^k \\
& \quad + (1 + \varepsilon)^{k-1} (1 + \varepsilon^{-1})^{k-1} 2\delta_M (3 \|\Gamma\|)^k \mathfrak{B}_0^2 A^{k-2} (nk u)^{k/2}.
\end{aligned}$$

## C Bounding $(\mathbb{E}[Z])^k$ under Assumption 4.(ii)

In this section, we only provide the part of the proof of Proposition 2 that needs to be modified to get the result when the kernels  $h_{i,j}$  depend on both  $i$  and  $j$  and when Assumption 4.(ii) holds. Keeping the notations of the proof of Proposition 2, we only want to bound  $(\mathbb{E}[Z])^k$  (and thus  $a_1$ ) using a different concentration result that can allow to deal with kernel functions  $h_{i,j}$  that might depend on  $i$ . We will use Proposition 4.

**Proposition 4.** (cf. [25, Theorem 1]) *Suppose that the sequence  $(X_i)_{i \geq 1}$  is a Markov chain satisfying Assumptions 1 and 4.(ii) with invariant distribution  $\pi$  and with an absolute spectral gap  $1 - \lambda > 0$ . Let us consider some  $n \in \mathbb{N}^*$  and bounded real valued functions  $(f_i)_{1 \leq i \leq n}$  such that for any  $i \in \{1, \dots, n\}$ ,  $\int f_i(x) d\pi(x) = 0$  and  $\|f_i\|_\infty \leq c$  for some  $c > 0$ . Let  $\sigma^2 = \sum_{i=1}^n \int f_i^2(x) d\pi(x)/n$ . Then for any  $0 \leq t < (1 - \lambda)/(5cq)$ ,*

$$\mathbb{E}_\chi \left[ e^{t \sum_{i=1}^n f_i(X_i)} \right] \leq \left\| \frac{d\chi}{d\pi} \right\|_{\pi, p} \exp \left( \frac{n\sigma^2}{qc^2} (e^{tqc} - tqc - 1) + \frac{n\sigma^2 \lambda q t^2}{1 - \lambda - 5cqt} \right), \quad (21)$$

where  $q$  is the constant introduced in Assumption 4.(ii). Moreover for any  $u \geq 0$  it holds

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n f_i(X_i) > \frac{2quA_1c}{n} + \sqrt{\frac{2quA_2\sigma^2}{n}}\right) \leq \left\|\frac{d\chi}{d\pi}\right\|_{\pi,p} e^{-u}.$$

where  $A_2 := \frac{1+\lambda}{1-\lambda}$  and  $A_1 := \frac{1}{3}\mathbb{1}_{\lambda=0} + \frac{5}{1-\lambda}\mathbb{1}_{\lambda>0}$ .

Note that the Bernstein inequality [25, Theorem 1] is given for stationary chains but one can easily extend this result to obtain Proposition 4 by working under the milder assumption Assumption 4.(ii) following for example the approach used in [17, Theorem 2.3]. Let us recall that

$$\begin{aligned} (\mathbb{E}[Z])^k &\leq \mathbb{E}[Z^k] \quad (\text{Using Jensen's inequality}) \\ &= \mathbb{E}\left[\left(\sup_{f_\xi \in \mathcal{F}} \sum_{i=1}^{n-1} f_\xi(X_i)\right)^k\right] = \mathbb{E}\left[\left(\sup_{f_\xi \in \mathcal{F}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}_{j-1}[p_{i,j}(X_i, X'_j)] \xi_j(X'_j)\right)^k\right] \\ &= \mathbb{E}\left[\sum_{j=2}^n \mathbb{E}_{j-1} \left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right] \quad (\text{Using Lemma 3}) \\ &= \mathbb{E}\left[\sum_{j=2}^n \mathbb{E}_{|X'} \left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right]. \end{aligned}$$

Thus we have

$$\alpha_1 = \frac{2\delta_M}{1+\varepsilon} \mathbb{E} \sum_{j=2}^n \left( \mathbb{E}_{|X'} [e^{\alpha(1+\varepsilon)K|C^{(j)}|}] - \alpha(1+\varepsilon)K \mathbb{E}_{|X'} [|C^{(j)}|] - 1 \right),$$

where  $C^{(j)} = \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)$  and where the notation  $\mathbb{E}_{|X'}$  refers to the expectation conditionally to the  $\sigma$ -algebra  $\sigma(X'_2, \dots, X'_n)$ .

Now we use a symmetrization trick: since  $e^x - x - 1 \geq 0$  for all  $x$  and since  $e^{a|x|} + e^{-a|x|} = e^{ax} + e^{-ax}$ , adding  $\mathbb{E}_{|X'}[\exp(-\alpha(1+\varepsilon)K|C^{(j)}|)] + \alpha(1+\varepsilon)K \mathbb{E}_{|X'} [|C^{(j)}|] - 1$  to  $\alpha_1$  gives

$$\alpha_1 \leq \frac{2\delta_M}{1+\varepsilon} \mathbb{E} \sum_{j=2}^n \left( \mathbb{E}_{|X'} [e^{\alpha(1+\varepsilon)KC^{(j)}}] - 1 + \mathbb{E}_{|X'} [e^{-\alpha(1+\varepsilon)KC^{(j)}}] - 1 \right). \quad (22)$$

Let us consider some  $j \in \{2, \dots, n\}$ . Conditionally on  $\sigma(X'_2, \dots, X'_n)$ ,  $C^{(j)}$  is a sum of bounded functions (by A) depending on the Markov chain. We denote

$$v_j(X'_j) = \sum_{i=1}^{j-1} \mathbb{E}_{X_i \sim \pi} [p_{i,j}^2(X_i, X'_j) | X'_j] \leq \mathfrak{B}_0^2$$

and  $V = \sum_{j=2}^n \mathbb{E} v_j^k(X'_j) \leq C_n^2 \mathfrak{B}_0^{2(k-1)}$  (with  $C_n^2 = \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbb{E} [p_{i,j}^2(X_i, X'_j)]$ ).

Remark that

$$\begin{aligned} \mathbb{E}_{X_i \sim \pi} [p_{i,j}(X_i, X'_j) | X'_j] &= \mathbb{E}_{X_i \sim \pi} [h_{i,j}(X_i, X'_j) - \mathbb{E}_{\tilde{X} \sim \pi} [h_{i,j}(X_i, \tilde{X})] | X'_j] \\ &= \int_{x'} \left( \int_{x_i} (h_{i,j}(x_i, X'_j) - h_{i,j}(x_i, \tilde{x})) d\pi(x_i) \right) d\pi(\tilde{x}) = 0, \end{aligned}$$

where the last equality comes from Assumption 3. We use the Bernstein inequality for Markov chain (see Proposition 4). Notice from Taylor expansion that  $(1-p/3)(e^p - p - 1) \leq p^2/2$  for all  $p \geq 0$ . Applying (21) with  $t = \alpha(1+\varepsilon)K$  and  $c = A$ , we get that for  $\alpha < [(1+\varepsilon)K\sqrt{q}(A\sqrt{q}/3 + \mathfrak{B}_0\sqrt{3/2})]^{-1} \wedge [(1-\lambda)^{-1/2}(1+\varepsilon)K\sqrt{q}(5A\sqrt{q}(1-\lambda)^{-1/2} + \sqrt{3\lambda}\mathfrak{B}_0)]^{-1}$ ,

$$\mathbb{E}_{|X'} [e^{\alpha(1+\varepsilon)K|C^{(j)}|}] \leq 2 \left\|\frac{d\chi}{d\pi}\right\|_{\pi,p} \times \mathbb{E}_{|X'} \left[ \exp\left(\frac{\alpha^2(1+\varepsilon)^2 K^2 q v_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3} + \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2 K^2 q}{1-\lambda-5\alpha(1+\varepsilon)KAq}\right)\right].$$

Considering  $\alpha < [(1+\varepsilon)K\sqrt{q}(A\sqrt{q}/3+\mathfrak{B}_0\sqrt{3/2})]^{-1} \wedge [(1-\lambda)^{-1/2}(1+\varepsilon)K\sqrt{q}(5A\sqrt{q}(1-\lambda)^{-1/2} + \sqrt{3\lambda}\mathfrak{B}_0)]^{-1}$ ,  $\varepsilon < 1$  and using (22), this leads to

$$\begin{aligned}
\frac{\alpha_1}{2\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}} &\leq \frac{2\delta_M}{1+\varepsilon} \sum_{j=2}^n \mathbb{E} \left[ \exp \left( \frac{\alpha^2(1+\varepsilon)^2 K^2 q v_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3} + \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2 K^2 q}{1-\lambda-5\alpha(1+\varepsilon)KAq} \right) - 1 \right] \\
&= \frac{2\delta_M}{1+\varepsilon} \sum_{j=2}^n \sum_{k=1}^{\infty} \frac{1}{k!} \left( \frac{\alpha^2(1+\varepsilon)^2 K^2 q v_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3} + \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2 K^2 q}{1-\lambda-5\alpha(1+\varepsilon)KAq} \right)^k \\
&= \frac{2\delta_M}{1+\varepsilon} \sum_{j=2}^n \sum_{k=1}^{\infty} \frac{1}{k!} \left( \frac{3}{2} \right)^{k-1} \left( \frac{\alpha^2(1+\varepsilon)^2 K^2 q v_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3} \right)^k \\
&\quad + \frac{2\delta_M}{1+\varepsilon} \sum_{j=2}^n \sum_{k=1}^{\infty} \frac{1}{k!} 3^{k-1} \left( \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2 K^2 q}{1-\lambda-5\alpha(1+\varepsilon)KAq} \right)^k \quad (\text{Using Lemma 2}) \\
&\leq \frac{\delta_M}{3(1+\varepsilon)} \sum_{k=1}^{\infty} \frac{3^k \alpha^{2k} (1+\varepsilon)^{2k} K^{2k} q^k V}{(4-4Aq\alpha(1+\varepsilon)K/3)^k} + \frac{2\delta_M}{3(1+\varepsilon)} \sum_{k=1}^{\infty} \frac{3^k V \lambda^k \alpha^{2k} (1+\varepsilon)^{2k} K^{2k} q^k}{(1-\lambda-5\alpha(1+\varepsilon)KAq)^k} \\
&\leq \frac{\delta_M}{3(1+\varepsilon)} \sum_{k=1}^{\infty} \frac{3^k \alpha^{2k} (1+\varepsilon)^{2k} K^{2k} q^k C_n^2 \mathfrak{B}_0^{2(k-1)}}{(2-2Aq\alpha(1+\varepsilon)K/3)^k} + \frac{2\delta_M}{3(1+\varepsilon)} \sum_{k=1}^{\infty} \frac{3^k C_n^2 \mathfrak{B}_0^{2(k-1)} \lambda^k \alpha^{2k} (1+\varepsilon)^{2k} K^{2k} q^k}{(1-\lambda-5\alpha(1+\varepsilon)KAq)^k} \\
&= \frac{(1+\varepsilon)C_n^2 \alpha^2 K^2 \delta_M q}{2-2Aq\alpha(1+\varepsilon)K/3-3\alpha^2(1+\varepsilon)^2 K^2 \mathfrak{B}_0^2 q} + \frac{2\delta_M C_n^2 \lambda \alpha^2 (1+\varepsilon) K^2 q}{1-\lambda-5\alpha(1+\varepsilon)KAq-3\mathfrak{B}_0^2 \lambda \alpha^2 (1+\varepsilon)^2 K^2 q} \\
&= \frac{(1+\varepsilon)C_n^2 \alpha^2 K^2 \delta_M q / 2}{1-Aq\alpha(1+\varepsilon)K/3-3\alpha^2(1+\varepsilon)^2 K^2 \mathfrak{B}_0^2 q / 2} + \frac{2\delta_M C_n^2 \lambda \alpha^2 (1+\varepsilon) K^2 q (1-\lambda)^{-1}}{1-5(1-\lambda)^{-1}\alpha(1+\varepsilon)KAq-3\mathfrak{B}_0^2 \lambda (1-\lambda)^{-1} \alpha^2 (1+\varepsilon)^2 K^2 q} \\
&\leq \frac{(1+\varepsilon)C_n^2 \alpha^2 K^2 \delta_M q / 2}{1-\alpha(1+\varepsilon)K\sqrt{q}(A\sqrt{q}/3+\mathfrak{B}_0\sqrt{3/2})} + \frac{2\delta_M C_n^2 \lambda \alpha^2 (1+\varepsilon) K^2 q (1-\lambda)^{-1}}{1-\alpha(1-\lambda)^{-1/2}(1+\varepsilon)K\sqrt{q}(5A\sqrt{q}(1-\lambda)^{-1/2} + \sqrt{3\lambda}\mathfrak{B}_0)}.
\end{aligned}$$

From this bound on  $\alpha_1$ , one can follow the steps of the proof of Proposition 2 to conclude.

## D Proof of Lemma 8

**Bounding  $\alpha_3$ .** Using the inequality  $k! \geq (k/e)^k$ , we have,

$$\begin{aligned}
\alpha_3 &\leq 2\delta_M \sum_{k \geq 2} \alpha^k (1+\varepsilon^{-1})^{2k-2} (8\|\Gamma\|^2)^k (nA^2)A^{k-2}(eu)^k \\
&= 2\delta_M \alpha^2 [\sqrt{n}A(1+\varepsilon^{-1})8\|\Gamma\|^2 eu]^2 \sum_{k \geq 2} \alpha^{k-2} (1+\varepsilon^{-1})^{2(k-2)} (8\|\Gamma\|^2)^{k-2} A^{k-2} (eu)^{k-2} \\
&= \frac{2\delta_M \alpha^2 [\sqrt{n}A(1+\varepsilon^{-1})8\|\Gamma\|^2 eu]^2}{1-\alpha(1+\varepsilon^{-1})^2(8\|\Gamma\|^2)Aeu}, \quad \text{for } \alpha < ((1+\varepsilon^{-1})^2(8\|\Gamma\|^2)Aeu)^{-1}.
\end{aligned}$$

**Bounding  $\alpha_2$ .** We use the inequality  $k! \geq k^{k/2}$  because  $(k/e)^k > k^{k/2}$  for  $k \geq e^2$  and for  $k$  smaller, the inequality follows by direct verification. Hence,

$$\begin{aligned}
\alpha_2 &\leq 2\delta_M \sum_{k \geq 2} \alpha^k (2+\varepsilon+\varepsilon^{-1})^{k-1} (3\|\Gamma\|)^k \mathfrak{B}_0^2 A^{k-2} (nu)^{k/2} \\
&= 2\delta_M (2+\varepsilon+\varepsilon^{-1}) \alpha^2 [3\|\Gamma\|\mathfrak{B}_0\sqrt{nu}]^2 \sum_{k \geq 2} \alpha^{k-2} (2+\varepsilon+\varepsilon^{-1})^{k-2} (3\|\Gamma\|)^{k-2} A^{k-2} (nu)^{(k-2)/2} \\
&= \frac{2\delta_M (2+\varepsilon+\varepsilon^{-1}) \alpha^2 [3\|\Gamma\|\mathfrak{B}_0\sqrt{nu}]^2}{1-\alpha(2+\varepsilon+\varepsilon^{-1})(3\|\Gamma\|)A(nu)^{1/2}}, \quad \text{for } \alpha < ((2+\varepsilon+\varepsilon^{-1})(3\|\Gamma\|)A(nu)^{1/2})^{-1}.
\end{aligned}$$

**Bounding  $a_1$ .** Using the bound previously obtained for  $(\mathbb{E}[Z])^k$  we get,

$$\begin{aligned} a_1 &= 2\delta_M \sum_{k \geq 2} \frac{\alpha^k}{k!} (1 + \varepsilon)^{k-1} (\mathbb{E}[Z])^k \leq 32\delta_M n \sum_{k \geq 2} \alpha^k (1 + \varepsilon)^{k-1} (KAm\tau)^k \\ &\leq 32\delta_M n \alpha^2 (1 + \varepsilon) [KAm\tau]^2 \sum_{k \geq 2} \alpha^{k-2} (1 + \varepsilon)^{k-2} (KAm\tau)^{k-2} \\ &\leq \frac{32\delta_M n \alpha^2 (1 + \varepsilon) [KAm\tau]^2}{1 - \alpha(1 + \varepsilon)KAm\tau}, \quad \text{for } 0 < \alpha < ((1 + \varepsilon)KAm\tau)^{-1}. \end{aligned}$$

## E Proof of Lemma 10

We prove a concentration result for the term

$$U_{n-1}^{(1)} = \sum_{j=2}^{n-1} \sum_{i=1}^{j-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j),$$

where  $h_{i,j}^{(1)}(x, y, z) = \int_w h_{i,j}(x, w)P(z, dw) - \int_w h_{i,j}(x, w)P^2(y, dw)$ , using an approach similar to the one from Section 4.2.1.

- Martingale structure

Denoting  $Y_j^{(1)} = \sum_{i=1}^{j-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j)$ , we have  $U_{n-1}^{(1)} = \sum_{j=2}^{n-1} Y_j^{(1)}$  which shows that  $(U_n^{(1)})_n$  is a martingale with respect to the  $\sigma$ -algebras  $(G_l)_l$ . Indeed, we have  $\mathbb{E}_{j-1}[Y_j^{(1)}] = 0$ .

- Talagrand's inequality To upper-bound  $(V_n^k)_n$ , we split it as previously namely

$$\begin{aligned} V_n^k &:= \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j) \right|^k \\ &= \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left( I_{i,j}^{(1)}(X_i, X_j) - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k, \end{aligned}$$

where  $I_{i,j}^{(1)}(x, z) = \int_w h_{i,j}(x, w)P(z, dw)$ . Using as previously Lemma 2 with  $\varepsilon = 1/2$ , we get

$$\begin{aligned} V_n^k &= \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left( I_{i,j}^{(1)}(X_i, X_j) - \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] \right. \right. \\ &\quad \left. \left. + \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k \\ &\leq (3/2)^{k-1} \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left( I_{i,j}^{(1)}(X_i, X_j) - \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] \right) \right|^k \\ &\quad + 3^{k-1} \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left( \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k. \end{aligned}$$

Again, basic computations and Jensen's inequality lead to

$$\begin{aligned} &\sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left( \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k \\ &= \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} p_{i,j}^{(1)}(X_i, X_j) \right|^k, \end{aligned}$$



where  $p_{i,j}^{(1)}(x, z) := I_{i,j}^{(1)}(x, z) - \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(x, \tilde{X})]$ . Hence, using Assumption 1 and Lemma 6 exactly like in the previous section, we get (for  $(X'_j)_j$  i.i.d. with distribution  $\nu$ )

$$V_n^k = 2 \times 3^{k-1} \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} p_{i,j}^{(1)}(X_i, X_j) \right|^k \leq 2 \times 3^{k-1} \delta_M \sum_{j=2}^{n-1} \mathbb{E}_{X'_j} \left| \sum_{i=1}^{j-1} p_{i,j}^{(1)}(X_i, X'_j) \right|^k.$$

Then, one can use the same duality trick to show that the  $V_n^k$  can be controlled using the supremum of a sum of functions of the Markov chain  $(X_i)_{i \geq 1}$  using [34, Theorem 3].

- Bounding  $\exp(w_n^k \alpha^k / k!)$

The terms  $a_2$  and  $a_3$  (see Eq.(16)) can be bounded in a similar way. For the term  $a_1$ , we only need to show that  $p_{i,j}^{(1)}$  satisfies  $\mathbb{E}_{X_i \sim \pi}[p_{i,j}^{(1)}(X_i, z)] = 0$ ,  $\forall z \in E$  in order to apply as previously a Bernstein's type inequality.

$$\begin{aligned} \mathbb{E}_{X_i \sim \pi}[p_{i,j}^{(1)}(X_i, z)] &= \int_{x_i} d\pi(x_i) \int_w h_{i,j}(x_i, w) P(z, dw) - \mathbb{E}_{\tilde{X} \sim \pi} \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(X, \tilde{X})] \\ &= \mathbb{E}_\pi[h_{i,j}] - \mathbb{E}_\pi[h_{i,j}] \quad (\text{Using Assumption 3}) \\ &= 0. \end{aligned}$$

- Conclusion of the proof

Let us consider the quantities  $A_1$  and  $\mathfrak{B}_1$  defined as the counterparts of  $A$  and  $\mathfrak{B}_0$  (see (8)) by replacing the functions  $(p_{i,j})_{i,j}$  by  $(p_{i,j}^{(1)})_{i,j}$ . One can easily see that  $A_1 = A$ . Let us give details about  $\mathfrak{B}_1$ . For any  $x \in E$ ,

$$\begin{aligned} \mathbb{E}_{X' \sim \nu} [(p_{i,j}^{(1)})^2(x, X')] &= \int_z \left( I_{i,j}^{(1)}(x, z) - \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(x, \tilde{X})] \right)^2 d\nu(z) \\ &= \int_z \left( \int_w h_{i,j}(x, w) P(z, dw) - \int_w h_{i,j}(x, w) \underbrace{\int_a P(a, dw) d\pi(a)}_{=d\pi(w)} \right)^2 d\nu(z) \\ &= \mathbb{E}_{X' \sim \nu} \left[ \mathbb{E}_{X \sim P(X', \cdot)} h_{i,j}(x, X) - \mathbb{E}_\pi[h_{i,j}] \right]^2, \end{aligned}$$

and for any  $y \in E$ ,

$$\begin{aligned} \mathbb{E}_{\tilde{X} \sim \pi} [(p_{i,j}^{(1)})^2(\tilde{X}, y)] &= \int_x \left( I_{i,j}^{(1)}(x, y) - \mathbb{E}_{\tilde{X} \sim \pi}[I_{i,j}^{(1)}(x, \tilde{X})] \right)^2 d\pi(x) \\ &= \int_x \left( \int_w h_{i,j}(x, w) P(y, dw) - \int_w h_{i,j}(x, w) \underbrace{\int_a P(a, dw) d\pi(a)}_{=d\pi(w)} \right)^2 d\pi(x) \\ &= \mathbb{E}_{\tilde{X} \sim \pi} \left[ \mathbb{E}_{X \sim P(y, \cdot)} h_{i,j}(\tilde{X}, X) - \mathbb{E}_\pi[h_{i,j}] \right]^2. \end{aligned}$$

Hence we get that

$$\begin{aligned} \mathfrak{B}_1^2 &:= \max \left[ \max_i \left\| \sum_{j=i+1}^n \mathbb{E}_{X \sim \nu} [(p_{i,j}^{(1)})^2(\cdot, X)] \right\|_\infty, \max_j \left\| \sum_{i=1}^{j-1} \mathbb{E}_{X \sim \pi} [(p_{i,j}^{(1)})^2(X, \cdot)] \right\|_\infty \right] \\ &\leq B_n^2, \end{aligned} \tag{23}$$

where we recall that

$$\begin{aligned} B_n^2 &= \max \left[ \sup_{0 \leq k \leq t_n} \max_i \sup_x \sum_{j=i+1}^n \mathbb{E}_{X' \sim \nu} \left[ \mathbb{E}_{X \sim P^k(X', \cdot)} h_{i,j}(x, X) - \mathbb{E}_\pi[h_{i,j}] \right]^2, \right. \\ &\quad \left. \sup_{0 \leq k \leq t_n} \max_j \sup_y \sum_{i=1}^{j-1} \mathbb{E}_{\tilde{X} \sim \pi} \left[ \mathbb{E}_{X \sim P^k(y, \cdot)} h_{i,j}(\tilde{X}, X) - \mathbb{E}_\pi[h_{i,j}] \right]^2 \right]. \end{aligned}$$

This allows us to get a concentration inequality similar to the one of the previous subsection, namely for any  $u > 0$ , it holds with probability at least  $1 - (1 + e)e^{-u}$ ,

$$\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j) \leq \kappa (A\sqrt{n}\sqrt{u} + (A + B_n\sqrt{n})u + 2A\sqrt{nu}^{3/2} + Au^2)$$

## F Proof of Lemma 11

Using Assumption 3, we have that  $\mathbb{E}_\pi[h_{i,j}] = \int_{x_i} P^{i-j+t_n}(X_{j-t_n}, dx_i) \int_{x_j} h_{i,j}(x_i, x_j) d\pi(x_j)$ . Hence we get,

$$\begin{aligned} (2a) &:= \left| \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-\lfloor \frac{tn}{2} \rfloor} \mathbb{E}_{j-t_n}[h_{i,j}(X_i, X_j)] - \mathbb{E}_\pi[h_{i,j}] \right| \\ &\leq \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-\lfloor \frac{tn}{2} \rfloor} \left| \int_{x_i} \int_{x_j} h_{i,j}(x_i, x_j) P^{i-j+t_n}(X_{j-t_n}, dx_i) (P^{j-i}(x_i, dx_j) - d\pi(x_j)) \right| \\ &\leq \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-\lfloor \frac{tn}{2} \rfloor} \|h_{i,j}\|_\infty \underbrace{\int_{x_i} P^{i-j+t_n}(X_{j-t_n}, dx_i)}_{=1} \sup_y \int_{x_j} |P^{j-i}(y, dx_j) - d\pi(x_j)| \\ &\leq \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-\lfloor \frac{tn}{2} \rfloor} \|h_{i,j}\|_\infty L \rho^{j-i} \leq \sum_{j=2}^n \sum_{i=(j-t_n+1) \vee 1}^{j-\lfloor \frac{tn}{2} \rfloor} \|h_{i,j}\|_\infty L \rho^{tn/2} \leq LAt_n, \end{aligned}$$

where we used that  $\rho^{tn/2} \leq \rho^{r \log(n)/2} = n^{r \log(\rho)/2} \leq n^{-1}$ . Indeed  $1 + r \log(\rho)/2 < 0$  because we choose  $r$  such that  $r > 2(\log(1/\rho))^{-1}$ . With an analogous approach, we bound the term (2c) as follows.

$$\begin{aligned} (2c) &:= \left| \sum_{j=2}^n \sum_{i=(j-\lfloor \frac{tn}{2} \rfloor+1) \vee 1}^{j-1} \mathbb{E}_{j-t_n}[h_{i,j}(X_i, X_j)] - \mathbb{E}[h_{i,j}(X_i, X_j)] \right| \\ &\leq \sum_{j=2}^n \sum_{i=(j-\lfloor \frac{tn}{2} \rfloor+1) \vee 1}^{j-1} \left| \int_{x_j} \int_{x_i} P^{j-i}(x_i, dx_j) h_{i,j}(x_i, x_j) (P^{i-j+t_n}(X_{j-t_n}, dx_i) - d\pi(x_i)) \right| \\ &\leq \sum_{j=2}^n \sum_{i=(j-\lfloor \frac{tn}{2} \rfloor+1) \vee 1}^{j-1} \|h_{i,j}\|_\infty \sup_z \|P^{i-j+t_n}(z, \cdot) - \pi\|_{TV} \\ &\leq \sum_{j=\lfloor \frac{tn}{2} \rfloor}^n \sum_{i=(j-\lfloor \frac{tn}{2} \rfloor+1)}^{j-1} \|h_{i,j}\|_\infty L \rho^{i-j+t_n} + \sum_{j=2}^{\lfloor \frac{tn}{2} \rfloor} \sum_{i=(j-\lfloor \frac{tn}{2} \rfloor+1) \vee 1}^{j-1} \|h_{i,j}\|_\infty L \rho^{i-j+t_n} \\ &\leq \sum_{j=\lfloor \frac{tn}{2} \rfloor}^n \sum_{i=(j-\lfloor \frac{tn}{2} \rfloor+1)}^{j-1} \|h_{i,j}\|_\infty L \rho^{tn/2} + t_n^2 \|h_{i,j}\|_\infty L \leq LA(t_n^2 + nt_n \rho^{tn/2}) \leq 2LAT_n^2, \end{aligned}$$

where we used that  $\rho^{tn/2} \leq \rho^{r \log(n)/2} = n^{r \log(\rho)/2} \leq n^{-1}$ .

## G Proof of Proposition 1

Since the split chain has the same distribution as the original Markov chain, we get that  $(\tilde{X}_i)_i$  is  $\psi$ -irreducible for some measure  $\psi$  and uniformly ergodic. From [31, Theorem 16.0.2], Assumption 1 ensures that for every measurable set  $A \subset E \times \{0, 1\}$  such that  $\psi(A) > 0$ , there exists some  $\kappa_A > 1$  such that

$$\sup_x \mathbb{E}[\kappa_A^{\tau_A} | \tilde{X}_1 = x] < \infty,$$

where  $\tau_A := \inf\{n \geq 1 : \tilde{X}_n \in A\}$  is the first hitting time of the set  $A$ . Let us recall that  $T_1$  and  $T_2$  are defined as hitting times of the atom of the split chain  $E \times \{1\}$  which is accessible (i.e. the atom has a positive  $\psi$ -measure). Hence, there exist  $C > 0$  and  $\kappa > 1$  such that,

$$\sup_x \mathbb{E}[\kappa^{\tau_{E \times \{1\}}} | \tilde{X}_1 = x] = \sup_x \mathbb{E}[\exp(\tau_{E \times \{1\}} \log(\kappa)) | \tilde{X}_1 = x] \leq C.$$

Considering  $k \geq 1$  such that  $C^{1/k} \leq 2$ , a straight forward application of Jensen inequality gives that  $\max(\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1}) \leq k/\log(\kappa)$ .