



HAL
open science

Deviation inequalities, Moderate deviations and some limit theorems for bifurcating Markov chains with application.

Valère Bitseki Penda, Hacène Djellout, Arnaud Guillin

► **To cite this version:**

Valère Bitseki Penda, Hacène Djellout, Arnaud Guillin. Deviation inequalities, Moderate deviations and some limit theorems for bifurcating Markov chains with application.. The Annals of Applied Probability, 2014, 24 (1), pp.235-291. 10.1214/13-AAP921 . hal-00646588v2

HAL Id: hal-00646588

<https://hal.science/hal-00646588v2>

Submitted on 17 Jan 2014

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.

DEVIATION INEQUALITIES, MODERATE DEVIATIONS AND SOME LIMIT THEOREMS FOR BIFURCATING MARKOV CHAINS WITH APPLICATION

BY S. VALÈRE BITSEKI PENDA, HACÈNE DJELLOUT AND
ARNAUD GUILLIN

Université Blaise Pascal

First, under a geometric ergodicity assumption, we provide some limit theorems and some probability inequalities for the bifurcating Markov chains (BMC). The BMC model was introduced by Guyon to detect cellular aging from cell lineage, and our aim is thus to complete his asymptotic results. The deviation inequalities are then applied to derive first result on the moderate deviation principle (MDP) for a functional of the BMC with a restricted range of speed, but with a function which can be unbounded. Next, under a uniform geometric ergodicity assumption, we provide deviation inequalities for the BMC and apply them to derive a second result on the MDP for a bounded functional of the BMC with a larger range of speed. As statistical applications, we provide superexponential convergence in probability and deviation inequalities (for either the Gaussian setting or the bounded setting), and the MDP for least square estimators of the parameters of a first-order bifurcating autoregressive process.

1. Introduction. Bifurcating Markov chains (BMC) are an adaptation of (usual) Markov chains to the data of a regular binary tree; see below for a more precise definition. In other terms, it is a Markov chain for which the index set is a regular binary tree. They are appropriate, for example, in the modeling of cell lineage data when each cell in one generation gives birth to two offspring in the next. Recently, they have received a great deal of attention because of the experiments of biologists on aging of *Escherichia Coli*; see [15, 20]. *E. Coli* is a rod-shaped bacterium which reproduces by dividing in the middle, thus producing two cells, one which already existed,

Received November 2011; revised January 2013.

AMS 2000 subject classifications. Primary 60F05, 60F10, 60F15, 60E15; secondary 60G42, 60J05, 62M02, 62M05, 62P10.

Key words and phrases. Bifurcating Markov chains, limit theorems, ergodicity, deviation inequalities, moderate deviation, martingale, first-order bifurcating autoregressive process, cellular aging.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Applied Probability*, 2014, Vol. 24, No. 1, 235–291. This reprint differs from the original in pagination and typographic detail.

that we call old pole progeny, and the other which is new, that we call new pole progeny. The aim of their experiments was to look for evidence of aging in *E. Coli*. In this section, we will introduce the model that allowed the authors of [15] to study the aging of *E. Coli* and we refer to their works for further motivations and insights on the data leading to the model studied here. This model is a typical example of bifurcating Markovian dynamics, and it has been the motivation for the rigorous mathematical study of BMC in [14]. This also motivates Sections 2 and 3 in the sequel, where we give a rigorous asymptotic (and nonasymptotic) study of BMC under geometric ergodicity and uniform geometric ergodicity assumptions.

1.1. *The model.* Let \mathbb{T} be a binary regular tree in which each vertex is seen as a positive integer different from 0; see Figure 1. For $r \in \mathbb{N}$, let

$$\mathbb{G}_r = \{2^r, 2^r + 1, \dots, 2^{r+1} - 1\}, \quad \mathbb{T}_r = \bigcup_{q=0}^r \mathbb{G}_q,$$

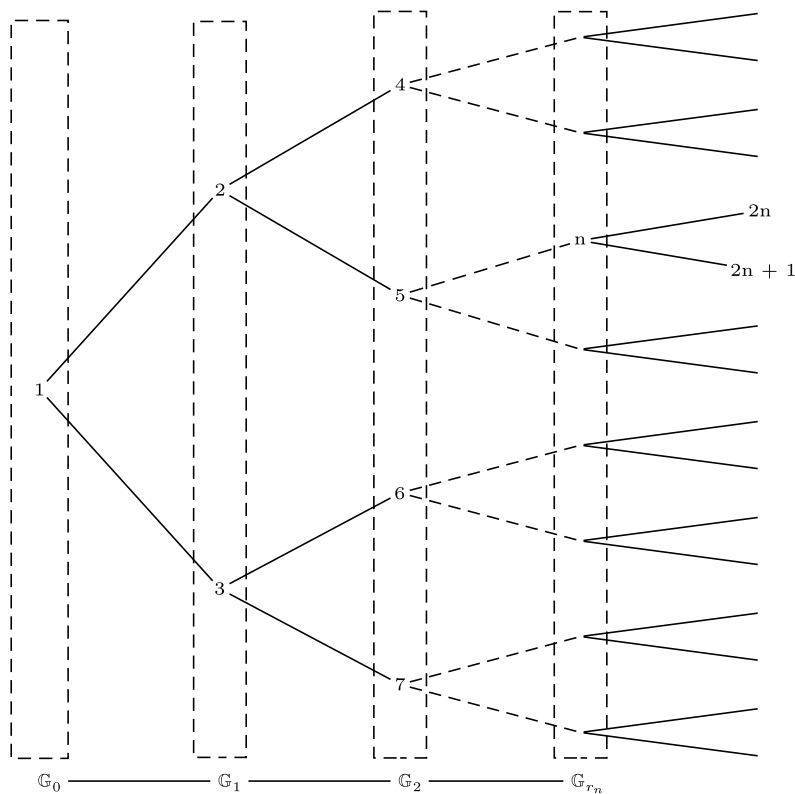


FIG. 1. *The binary tree \mathbb{T} .*

which denote, respectively, the r th column and the first $(r + 1)$ columns of the tree. Then, the cardinality $|\mathbb{G}_r|$ of \mathbb{G}_r is 2^r and that of \mathbb{T}_r is $|\mathbb{T}_r| = 2^{r+1} - 1$. A column of a given integer n is \mathbb{G}_{r_n} with $r_n = \lfloor \log_2 n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of the real number x .

The genealogy of the cells is described by this tree. In the sequel we will thus see \mathbb{T} as a given population. Then the vertex n , the column \mathbb{G}_r and the first $(r + 1)$ columns \mathbb{T}_r designate, respectively, individual n , the r th generation and the first $(r + 1)$ generations. The initial individual is denoted 1.

Guyon et al. [14, 15] proposed the following linear Gaussian model to describe the evolution of the growth rate of the population of cells derived from an initial individual:

$$(1.1) \quad \mathcal{L}(X_1) = \nu \quad \text{and} \quad \forall n \geq 1 \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases}$$

where X_n is the growth rate of individual n , n is the mother of $2n$ (the new pole progeny cell) and $2n + 1$ (the old pole progeny cell), ν is a distribution probability on \mathbb{R} , $\alpha_0, \alpha_1 \in (-1, 1)$; $\beta_0, \beta_1 \in \mathbb{R}$ and $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of i.i.d. bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$, where

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1).$$

The processes (X_n) defined by (1.1) are typical examples of BMC which are called the first-order bifurcating autoregressive processes [BAR(1)]. The BAR(1) processes are an adaptation of autoregressive processes, when the data have a binary tree structure. They were first introduced by Cowan and Staudte [6] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation. We will not discuss here extensions to m -ary tree, which follow more or less from the same method, or Markov chains on Galton–Watson trees that are left for an other study.

In [14], Guyon, after establishing the first results on the theory of BMC, proves laws of large numbers and central limit theorem for the least-square estimators $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$ of the 4-dimensional parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$; see Section 4 for a more precise definition. He also gives some statistical tests which allow to check if the model is symmetric or not (roughly $\alpha_0 = \alpha_1$ or not), and if the new pole and the old pole populations are even distinct in mean, which allows him to conclude a statistical evidence in aging in *E. Coli*. Let us also mention [4], where Bercu et al., using the martingale approach, give asymptotic analysis of the least squares estimators of the unknown parameters of a general asymmetric p th-order BAR processes.

In this paper, we will give moderate deviation principle (MDP) for this estimator and the statistical tests done by Guyon. We will also give deviation

inequalities for $\hat{\theta}^r - \theta$, which are important for a rigorous (nonasymptotic) statistical study. This will be done in two cases: the Gaussian case as described above and the case where the noise and the initial state X_1 are assumed to take values in a compact set. Note that the latter case implies that the BAR(1) process defined by (1.1) valued in compact set.

We are now going to give a rigorous definition of BMC. We refer to [14] for more detail.

1.2. *Definitions.* For an individual $n \in \mathbb{T}$, we are interested in the quantity X_n (it may be the weight, the growth rate, ...) with values in the metric space S endowed with its Borel σ -field \mathcal{S} .

DEFINITION 1.1 (\mathbb{T} -transition probability, see [14]). We call \mathbb{T} -transition probability any mapping $P: S \times \mathcal{S}^2 \rightarrow [0, 1]$ such that:

- $P(\cdot, A)$ is measurable for all $A \in \mathcal{S}^2$;
- $P(x, \cdot)$ is a probability measure on $(\mathcal{S}^2, \mathcal{S}^2)$ for all $x \in S$.

For a \mathbb{T} -transition probability P on $S \times \mathcal{S}^2$, we denote by P_0 , P_1 and Q , respectively, the first and the second marginal of P , and the mean of P_0 and P_1 , that is, $P_0(x, B) = P(x, B \times S)$, $P_1(x, B) = P(x, S \times B)$ for all $x \in S$ and $B \in \mathcal{S}$ and $Q = \frac{P_0 + P_1}{2}$.

For $p \geq 1$, we denote by $\mathcal{B}(\mathcal{S}^p)$ [resp., $\mathcal{B}_b(\mathcal{S}^p)$], the set of all \mathcal{S}^p -measurable (resp., \mathcal{S}^p -measurable and bounded) mappings $f: \mathcal{S}^p \rightarrow \mathbb{R}$. For $f \in \mathcal{B}(\mathcal{S}^3)$, we denote by $Pf \in \mathcal{B}(S)$ the function

$$x \mapsto Pf(x) = \int_{\mathcal{S}^2} f(x, y, z) P(x, dy, dz) \quad \text{when it is defined.}$$

DEFINITION 1.2 (Bifurcating Markov chains; see [14]). Let $(X_n, n \in \mathbb{T})$ be a family of S -valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$. Let ν be a probability on (S, \mathcal{S}) and P be a \mathbb{T} -transition probability. We say that $(X_n, n \in \mathbb{T})$ is a (\mathcal{F}_r) -bifurcating Markov chain with initial distribution ν and \mathbb{T} -transition probability P if:

- X_n is \mathcal{F}_{r_n} -measurable for all $n \in \mathbb{T}$;
- $\mathcal{L}(X_1) = \nu$;
- for all $r \in \mathbb{N}$ and for all family $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(\mathcal{S}^3)$

$$\mathbb{E} \left[\prod_{n \in \mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1}) \middle/ \mathcal{F}_r \right] = \prod_{n \in \mathbb{G}_r} Pf_n(X_n).$$

In the following, when unspecified, the filtration implicitly used will be $\mathcal{F}_r = \sigma(X_i, i \in \mathbb{T}_r)$. We denote by $(Y_r, r \in \mathbb{N})$ the Markov chain on S with

$Y_0 = X_1$ and transition probability Q . The chain $(Y_r, r \in \mathbb{N})$ corresponds to a random lineage taken in the population.

We denote by \mathfrak{G} the set of all permutations of \mathbb{N}^* that leaves each \mathbb{G}_r invariant. We draw a permutation Π uniformly on \mathfrak{G} , independently of $X = (X_n, n \in \mathbb{T})$. Drawing Π “uniformly” on \mathfrak{G} means drawing the restriction of Π on \mathbb{G}_r uniformly among the $(2^r)!$ permutations of \mathbb{G}_r . In particular, $(\Pi(2^r), \Pi(2^r + 1), \dots, \Pi(2^{r+1} - 1))$ can be viewed as a random drawing of all the elements of \mathbb{G}_r without replacement. Notice that Π allows one to define a random order on \mathbb{T} which preserves the genealogical order. For example, $(\Pi(i), 1 \leq i \leq n)$ denotes the set of the “first” n individuals of \mathbb{T} . Π was introduced by Guyon in order to sample over the “first” n individuals. As mentioned in [14], this choice of Π allows one to preserve the same asymptotic behavior for the empirical means resulting from the sampling over (say) the r th generation, the first $(r + 1)$ generations or the “first” n individuals. In general, the choice of another permutation does not preserve the asymptotic behavior of these empirical means. We refer to [14], Section 2.2, for more detail.

Throughout the paper, we will denote by:

- $f \otimes g$ the mapping $(x, y) \mapsto f(x)g(y)$.
- Q^p the p th iterated of Q recursively defined by the formulas $Q^0(x, \cdot) = \delta_x$ and $Q^{p+1}(x, B) = \int_S Q(s, dy)Q^p(y, B)$ for all $B \in \mathcal{S}$; Q^p is a transition probability in (S, \mathcal{S}) .
- νQ the distribution on (S, \mathcal{S}) defined by $\nu Q(B) = \int_S \nu(dx)Q(x, B)$; νQ^p is the law of Y_p .
- $(Qf)(x) = \int_S f(y)Q(x, dy)$ when it is defined.
- (νf) or (ν, f) the integral $\int_S f d\nu$ when it is defined.

For all $i \in \mathbb{T}$, we set $\Delta_i = (X_i, X_{2i}, X_{2i+1})$. We introduce the following empirical quantities:

$$(1.2) \quad \begin{cases} \overline{M}_{\mathbb{G}_r}(f) = \frac{1}{|\mathbb{G}_r|} \sum_{i \in \mathbb{G}_r} f(\tilde{\Delta}_i), \\ \overline{M}_{\mathbb{T}_r}(f) = \frac{1}{|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} f(\tilde{\Delta}_i), \\ \overline{M}_n^\Pi(f) = \frac{1}{n} \sum_{i=1}^n f(\tilde{\Delta}_{\Pi(i)}), \end{cases}$$

where $f(\tilde{\Delta}_i) = f(\Delta_i) = f(X_i, X_{2i}, X_{2i+1})$ if $f \in \mathcal{B}(S^3)$ and $f(\tilde{\Delta}_i) = f(X_i)$ if $f \in \mathcal{B}(S)$.

Guyon in [14] studied limit theorems of the empirical means (1.2), namely the law of large numbers (L^2 and almost sure versions) and the central limit theorems for (1.2) when $f \in \mathcal{B}(S^3)$, but centered by the conditional expectation rather than by the limit mean. An extension of the BMC has been

proposed in [8], in which the authors studied a model of BMC with missing data. To take into account the possibility for a cell to die, the authors of [8] use Galton–Watson tree instead of a regular tree. And they give a weak law of large numbers, an invariance principle and the central limit result for the average over one generation or up to one generation. As previously mentioned, this setting will be considered in incoming works. One can also mention the work of De Saporta et al. [7] dealing with bifurcating autoregressive processes with missing data in the estimation procedure of the parameters of the asymmetric BAR process. They use a two type Galton–Watson process to model the genealogy and give convergence and asymptotic normality of their estimators. It is important to remark that the nonasymptotic study of deviation inequalities has not been considered at all in these works, despite their practical interest.

1.3. *Objectives.* Our objectives in this paper are:

- to give some limit theorems for BMC that complete those done in [14] (LLN, LIL, ...);
- to give probability inequalities and deviation inequalities for the empirical means (1.2), that is, for $f \in \mathcal{B}(S)$ and all $x > 0$

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) - (\mu, f) \geq x) \leq e^{-C(x,r)},$$

where $C(x, r)$ will crucially depend on our set of assumptions on f and on the ergodic property of Q but valid for (nearly) all r ;

- to study moderate deviation principle (MDP) for BMC, that is, for some range of speed $\sqrt{r} \ll b_r \ll r$ (depending on assumptions) and for $f \in \mathcal{B}_b(S^3)$ with $Pf = 0$

$$\frac{b_{|\mathbb{T}_r|}^2}{|\mathbb{T}_r|} \log \mathbb{P}\left(\frac{1}{b_{|\mathbb{T}_r|}} M_{\mathbb{T}_r}(f) \geq x\right) \sim -\frac{x^2}{2\sigma^2};$$

- to obtain the MDP and deviation inequalities for the estimator of bifurcating autoregressive process, which are important for a rigorous statistical study.

All these results will be obtained under hypothesis of geometric ergodicity or uniform geometric ergodicity, meaning that Q^r converges (uniformly) exponentially fast to a limiting measure.

The limit theorems, proved in this paper, include strong law of large numbers for the empirical average $\overline{M}_n^{\text{II}}(f)$ with $f \in \mathcal{B}(S)$ (this case is not studied in [14]), the law of the iterated logarithm and the almost sure functional central limit theorem. A strong law of large numbers will be obtained via control of 4th order moments. We thus generalize the computation of 2nd order moments made by Guyon in [14]. It will be noted that the technique

we will use can be applied to compute the other higher-order moments, but at the price of huge and tedious computations.

Deviation inequalities will be obtained in the setting of unbounded functions, by using the classical Markov inequality and under geometric ergodicity assumption. The results are, however, at this point quite restrictive.

Exponential deviation inequalities will be shown for bounded functions and under a uniform geometric ergodicity assumption. Their proof intensively uses the Azuma–Bennett–Hoeffding inequality [1, 3, 16], which requires bounded random variables. Extension to unbounded functions and weaker ergodicity assumptions will be done in a further work, using transportation inequalities in the spirit of [12].

The MDP will be mainly deduced from these inequalities and general results on moderate deviations of martingales; see [11], recalled in the Appendix B. Their speed will depend on whether uniform geometric ergodicity or only geometric ergodicity is satisfied.

Before presenting the plan of our paper, let us recall the definition of a moderate deviation principle (MDP): let $(b_n)_{n \geq 0}$ be a positive sequence such that

$$\frac{b_n}{n} \xrightarrow{n \rightarrow \infty} 0 \quad \text{and} \quad \frac{b_n^2}{n} \xrightarrow{n \rightarrow \infty} \infty.$$

We say that a sequence of centered random variables $(M_n)_n$ with topological state space (S, \mathcal{S}) satisfies a MDP with speed b_n^2/n and rate function $I: S \rightarrow \mathbb{R}_+^*$ if for each $A \in \mathcal{S}$,

$$\begin{aligned} - \inf_{x \in A^o} I(x) &\leq \liminf_{n \rightarrow \infty} \frac{n}{b_n^2} \log \mathbb{P} \left(\frac{n}{b_n} M_n \in A \right) \leq \limsup_{n \rightarrow \infty} \frac{n}{b_n^2} \log \mathbb{P} \left(\frac{n}{b_n} M_n \in A \right) \\ &\leq - \inf_{x \in \bar{A}} I(x); \end{aligned}$$

here A^o and \bar{A} denote the interior and closure of A , respectively.

The MDP can thus be seen as an intermediate behavior between the central limit theorem ($b_n = b\sqrt{n}$) and large deviation ($b_n = bn$). Usually, the MDP exhibits a simpler rate function inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the large deviation principle.

Our paper is organized as follows. Section 2 states the moments control inequalities and their consequences. We shall state in this section a first result on the MDP for BMC in a general framework, but with a very restricted range of speed. Section 3 deals with the exponential inequalities and their consequences. In this section, we shall generalize the MDP done in Section 2, allowing for a larger range of speed, but under more stringent assumptions. In Section 4, we will focus particularly on the first order bifurcating autoregressive processes. The proofs of some inequalities are technical so

postponed in Appendix A. Appendix B is devoted to definitions and limit theorems for martingales used intensively in the paper, and are included here for completeness.

2. Moments control and consequences. Let F be a vector subspace of $\mathcal{B}(S)$ such that:

- (i) F contains the constants;
- (ii) $F^2 \subset F$;
- (iii) $F \otimes F \subset L^1(P(x, \cdot))$ for all $x \in S$, and $P(F \otimes F) \subset F$;
- (iv) there exists a probability μ on (S, \mathcal{S}) such that $F \subset L^1(\mu)$ and

$$\lim_{r \rightarrow \infty} \mathbb{E}_x[f(Y_r)] = (\mu, f)$$

for all $x \in S$ and $f \in F$;

- (v) for all $f \in F$, there exists $g \in F$ such that for all $r \in \mathbb{N}$, $|Q^r f| \leq g$;
- (vi) $F \subset L^1(\nu)$,

where we have used the notation $F^2 = \{f^2/f \in F\}$, $F \otimes F = \{f \otimes g/f, g \in F\}$ and $PE = \{Pf/f \in E\}$ whenever an operator P acts on a set E .

The following hypothesis is about the geometric ergodicity of Q :

(H1) Assume that for all $f \in F$ such that $(\mu, f) = 0$, there exists $g \in F$ such that for all $r \in \mathbb{N}$ and for all $x \in S$, $|Q^r f(x)| \leq \alpha^r g(x)$ for some $\alpha \in (0, 1)$; that is, the Markov chain $(Y_r, r \in \mathbb{N})$ is geometrically ergodic.

Recall that under this hypothesis, Guyon [14] has shown the weak law of large numbers for the three empirical average $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\text{II}}(f)$ (see [14], Theorem 11 when $f \in F$ and Theorem 12 when $f \in \mathcal{B}(S^3)$) and the strong law of large numbers only for $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$; see [14], Theorem 14 and Corollary 15 when $f \in F$ and Theorem 18 when $f \in \mathcal{B}(S^3)$.

When $f \in \mathcal{B}(S^3)$ and under the additional hypothesis Pf^2 and Pf^4 exist and belong to F , he proved the central limit theorem for $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\text{II}}(f)$; see [14], Theorem 19 and Corollary 21. Recall that the central limit theorem for the three empirical means (1.2) when $f \in \mathcal{B}(S)$ is still an open question; see [8] for more precision.

In this section, we complete these results by showing the strong law of large numbers for $\overline{M}_n^{\text{II}}(f)$, when $f \in F$. We prove also the law of the iterated logarithm (LIL) and almost sure functional central limit theorem (ASFCLT) for $\overline{M}_n^{\text{II}}(f)$ when $f \in \mathcal{B}(S^3)$.

2.1. Control of the 4th order moments. In order to establish limit theorems below, let us state the following:

THEOREM 2.1. *Let F satisfy (i)–(vi). Let $f \in F$ such that $(\mu, f) = 0$. We assume hypothesis (H1). Then for all $r \in \mathbb{N}$,*

$$(2.1) \quad \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4] \leq \begin{cases} c(\frac{1}{4})^r, & \text{if } \alpha^2 < \frac{1}{2}, \\ cr^2(\frac{1}{4})^r, & \text{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{4r}, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

where the positive constant c depends on α and f (and may differ line by line).

PROOF. First note that $f(X_i) \in L^4$ for all $i \in \mathbb{G}_r$. Indeed, let $(z_1, \dots, z_r) \in \{0, 1\}^r$ the unique path in the binary tree from the root 1 to i . Then,

$$\mathbb{E}[f^4(X_i)] = \nu P_{z_1} \cdots P_{z_r} f^4,$$

and from hypotheses (ii), (iii) and (vi) we conclude that $\nu P_{z_1} \cdots P_{z_r} f^4 < \infty$.

Now, the proof divides into two parts.

Part 1. Computation of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$. Independently of X , let us draw four independent indices I_r, J_r, K_r and L_r uniformly from \mathbb{G}_r . Then

$$\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4] = \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})].$$

For all $p \in \{0, \dots, r\}$, let us define the following events:

- E_0^p : The ancestors of I_r, J_r, K_r and L_r are different in \mathbb{G}_p .
- E_1^p : Exactly two of I_r, J_r, K_r and L_r have the same ancestor in \mathbb{G}_p .
- E_2^p : I_r, J_r, K_r and L_r have the same ancestor two by two in \mathbb{G}_p .
- E_3^p : Exactly three of I_r, J_r, K_r and L_r have the same ancestor in \mathbb{G}_p .
- E_4^p : I_r, J_r, K_r and L_r have the same ancestor in \mathbb{G}_p .

We also consider the following events whose for each fixed $p \leq r$, probability depend only on p .

- $E'_0{}^p$: Draw uniformly four independent indices from \mathbb{G}_p which are different.
- $E'_1{}^p$: Draw uniformly four independent indices from \mathbb{G}_p such that two are the same, and the others are different.
- $E'_2{}^p$: Draw uniformly four independent indices from \mathbb{G}_p which are the same, two by two.
- $E'_3{}^p$: Draw uniformly four independent indices from \mathbb{G}_p such that exactly three are the same.
- $E'_4{}^p$: Draw uniformly four independent indices from \mathbb{G}_p which are all the same.

In the sequel we do the convention that E_0^{r+1} is a certain event. Then after successive conditioning by events E_i^p for $p \in \{0, \dots, r\}$ and $i \in \{0, \dots, 4\}$, we have

$$\begin{aligned}
& \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})] \\
&= \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^2] \times \mathbb{P}(E_0^2) \\
&+ \sum_{p=2}^r \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_1^p] \times \mathbb{P}(E_1^p \cap E_0^{p+1}) \\
(2.2) \quad &+ \sum_{p=2}^r \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_2^p] \times \mathbb{P}(E_2^p \cap E_0^{p+1}) \\
&+ \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_3^r] \times \mathbb{P}(E_3^r) \\
&+ \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_4^r] \times \mathbb{P}(E_4^r).
\end{aligned}$$

Let us notice that

- for all $i \in \{1, 2, 3, 4\}$, E_i^r and $E_i'^r$ have the same probability;
- the realization of “ $E_1^p \cap E_0^{p+1}$ ” can be seen as “draw uniformly four independent indices from \mathbb{G}_p such that two are the same and others are different, and the two indices which are the same take different paths at \mathbb{G}_{p+1} .” Thus “ $E_1^p \cap E_0^{p+1}$ ” has the same probability that “ $E_1'^p \cap A_{p,p+1}$,” where “ $A_{p,p+1}$ ” is the event, “the indices which are the same in \mathbb{G}_p take different paths at \mathbb{G}_{p+1} ”;
- similarly, the realization of “ $E_2^p \cap E_0^{p+1}$ ” may be interpreted as, “draw uniformly four independent indices from \mathbb{G}_p which are the same two by two, and all the indices take different paths at \mathbb{G}_{p+1} .” Thus “ $E_2^p \cap E_0^{p+1}$ ” has the same probability that “ $E_2'^p \cap A_{p,p+1}$,” where “ $A_{p,p+1}$ ” is the event, “the indices which are the same in \mathbb{G}_p take different paths at \mathbb{G}_{p+1} ”;
- for all $p \in \{0, \dots, r\}$, we have

$$\begin{aligned}
\mathbb{P}(E_1'^p) &= \frac{6(2^p - 1)(2^p - 2)}{2^{3p}}, & \mathbb{P}(E_2'^p) &= \frac{3(2^p - 1)}{2^{3p}}, \\
\mathbb{P}(E_3'^p) &= \frac{4(2^p - 1)}{2^{3p}}, & \mathbb{P}(E_4'^p) &= \frac{1}{2^{3p}}.
\end{aligned}$$

We may then deduce that

$$\mathbb{P}(E_0^2) = \frac{3}{32}, \quad \mathbb{P}(E_3^r) = \frac{4(2^r - 1)}{2^{3r}}, \quad \mathbb{P}(E_4^r) = \frac{1}{2^{3r}}$$

and for $p \in \{2, \dots, r-1\}$,

$$\mathbb{P}(E_1^p \cap E_0^{p+1}) = \mathbb{P}(E_1'^p) \mathbb{P}(A_{p,p+1}/E_1'^p) = \frac{3(2^p - 1)(2^p - 2)}{2^{3p}}$$

and

$$\mathbb{P}(E_2^p \cap E_0^{p+1}) = \mathbb{P}(E_2^p) \mathbb{P}(A_{p,p+1}/E_2^p) = \frac{3 \cdot 2^p - 1}{4 \cdot 2^{3p}}.$$

We are now going to compute each term which appears in (2.2). We have the following convention: $P(Q^{-1}f \otimes Q^{-1}f) = f^2$. In the sequel, we will use intensively, with a slight modification, the calculations made by Guyon [14] in order to compute conditional expectations related to the event, “draw uniformly two independent indices from \mathbb{G}_p ,” for $p \in \{0, \dots, r\}$.

(a) We have that

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_4^r] = \nu Q^r f^4.$$

(b) Conditionally on E_3^r , we may assume that the indices I_r , K_r and L_r are the same. We then have, using the calculations made by Guyon [14],

$$\begin{aligned} & \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_3^r] \\ &= \mathbb{E}[f^3(X_{I_r})f(X_{J_r})/E_3^r] \\ &= \frac{2^r}{2^r - 1} \left\{ \sum_{p=0}^{r-1} 2^{-p-2} \nu Q^p P(Q^{r-p-1} f^3 \otimes Q^{r-p-1} f \right. \\ & \qquad \qquad \qquad \left. + Q^{r-p-1} f \otimes Q^{r-p-1} f^3) \right\}. \end{aligned}$$

(c) Let $p \in \{2, \dots, r\}$. Conditionally on E_2^p and E_0^{p+1} we may assume that I_r and J_r have the same ancestor at \mathbb{G}_p , and K_r and L_r have the same ancestor at \mathbb{G}_p . For simplification, we will use the following notation:

$$(2.3) \quad Q_{\otimes}^k f := Q^k f \otimes Q^k f,$$

and we thus have

$$\begin{aligned} & \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_2^p] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/\mathcal{F}_{p+1}]/\mathcal{F}_p]/E_0^{p+1}, E_2^p] \\ &= \mathbb{E}[P(Q_{\otimes}^{r-p-1} f)(X_{I_r \wedge_p J_r})P(Q_{\otimes}^{r-p-1} f)(X_{K_r \wedge_p L_r})/E_0^{p+1}, E_2^p] \\ &= \frac{2^p}{2^p - 1} \sum_{l=0}^{p-1} 2^{-l-1} \nu Q^l P((Q^{p-l-1} P(Q_{\otimes}^{r-p-1} f)) \\ & \qquad \qquad \qquad \otimes (Q^{p-l-1} P(Q_{\otimes}^{r-p-1} f))), \end{aligned}$$

where $I_r \wedge_p J_r$ (resp., $K_r \wedge_p L_r$) denotes the common ancestor of I_r and J_r which is in \mathbb{G}_p (resp., the common ancestor of K_r and L_r which is in \mathbb{G}_p).

(d) Let $p \in \{2, \dots, r\}$. Now conditionally on E_1^p and E_0^{p+1} we may assume that it is K_r and L_r which have the same ancestor in \mathbb{G}_p . We denote by $p(I_r)$ and $p(J_r)$, respectively, the ancestor of I_r and J_r which are in \mathbb{G}_p . As before, the common ancestor of K_r and L_r , which are in \mathbb{G}_p , is denoted by $K_r \wedge_p L_r$. At this step, we may repeat the successive conditioning that we have done in the beginning but this time for indices $p(I_r)$, $p(J_r)$ and $K_r \wedge_p L_r$. This leads us to

$$\begin{aligned}
& \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_1^p] \\
&= \mathbb{E}[Q^{r-p}f(X_{p(I_r)})Q^{r-p}f(X_{p(J_r)})P(Q_{\otimes}^{r-p-1}f)(X_{K_r \wedge_p L_r})/E_0^{p+1}, E_1^p] \\
&= \frac{2^{2p}}{(2^p - 1)(2^p - 2)} \sum_{l=1}^{p-1} \frac{1}{2^{l+1}} \frac{1}{2} \\
&\quad \times \sum_{m=0}^{l-1} 2^{-m-1} \{ \nu Q^m P((Q^{l-m-1}P(Q_{\otimes}^{r-l-1}f)) \otimes Q^{p-m-1}P(Q_{\otimes}^{r-p-1}f)) \\
&\quad + \nu Q^m P((Q^{p-m-1}P(Q_{\otimes}^{r-p-1}f)) \otimes (Q^{l-m-1}P(Q_{\otimes}^{r-l-1}f))) \\
&\quad + \nu Q^m P((Q^{l-m-1}P(Q^{r-l-1}f \otimes Q^{p-l-1}P(Q_{\otimes}^{r-p-1}f))) \otimes (Q^{r-m-1}f)) \\
&\quad + \nu Q^m P(Q^{r-m-1}f \otimes (Q^{l-m-1}P(Q^{r-l-1}f \otimes Q^{p-l-1}P(Q_{\otimes}^{r-p-1}f)))) \\
&\quad + \nu Q^m P((Q^{l-m-1}P(Q^{p-l-1}P(Q_{\otimes}^{r-p-1}f) \otimes Q^{r-l-1}f)) \otimes (Q^{r-m-1}f)) \\
&\quad + \nu Q^m P((Q^{r-m-1}f) \otimes (Q^{l-m-1}P(Q^{p-l-1}P(Q_{\otimes}^{r-p-1}f) \otimes Q^{r-l-1}f))) \}.
\end{aligned}$$

(e) Finally,

$$\begin{aligned}
& \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^2] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/\mathcal{F}_2]/\mathcal{F}_1]/E_0^2] \\
&= \mathbb{E}[P(Q_{\otimes}^{r-2}f)(X_2)P(Q_{\otimes}^{r-2}f)(X_3)/E_0^2] \\
&= \nu P(P(Q_{\otimes}^{r-2}f) \otimes P(Q_{\otimes}^{r-2}f)).
\end{aligned}$$

Gathering together all of these terms, each multiplied by their respective probability, we obtain an explicit expression for $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$.

Part 2. Rate. We are now going to give some rates for the different terms that appear in the expression of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$.

Throughout this part, we will use intensively the following to bound quantities which appear in the expression of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$:

- Let $f \in F$ such that $(\mu, f) = 0$. Then from (i)–(vi) and hypothesis (H1), there exists a positive constant c such that $\forall l, m, n \in \mathbb{N}$,

$$\nu Q^l P(Q^m f \otimes Q^n f) \leq \alpha^{m+n} \nu Q^l P(g \otimes g) \leq c \alpha^{m+n},$$

where g is given in hypothesis (H1).

In the sequel, c denotes a positive constant which depends on f , and c_1 denotes a positive constant which depends on α . The constants c and c_1 may vary from one line to another and from one expression to another.

(a) For the first term appearing in (2.2), we have

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^2] \times \mathbb{P}(E_0^2) \leq c_1 c \alpha^{4r}.$$

(b) For the fifth term appearing in (2.2), we have

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_4^r] \times \mathbb{P}(E_4^r) \leq c \left(\frac{1}{2}\right)^{3r},$$

where, from (ii), (v) and (vi), c is such that $\nu Q^r f^4 < c$.

(c) For the fourth term appearing in (2.2), we have

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_3^r] \times \mathbb{P}(E_3^r) \leq c c_1 \alpha^r \left(\frac{1}{4}\right)^r \sum_{p=0}^{r-1} \left(\frac{1}{2\alpha}\right)^p,$$

where, from (ii), (iii), (v) and (vi), c is such that for all $p, q \in \mathbb{N}$

$$\max(\nu Q^p P(Q^q f^3 \otimes g), \nu Q^p P(g \otimes Q^q f^3)) < c,$$

and from hypothesis (H1), g is such that for all $p \in \{1, \dots, r-1\}$

$$(2.4) \quad Q^{r-p-1} f \leq \alpha^{r-p-1} g.$$

Now depending on the value of α , we obtain that

$$\begin{aligned} & \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_3^r] \times \mathbb{P}(E_3^r) \\ & \leq \begin{cases} c_1 c \left(\left(\frac{\alpha}{4}\right)^r + \left(\frac{1}{2^3}\right)^r \right), & \text{if } \alpha \neq \frac{1}{2}, \\ c_1 c r \left(\frac{1}{2^3}\right)^r, & \text{if } \alpha = \frac{1}{2}. \end{cases} \end{aligned}$$

(d) Let us denote the third term appearing in (2.2) by

$$A_r := \sum_{p=2}^r \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_2^p] \times \mathbb{P}(E_2^p \cap E_0^{p+1}).$$

So we have

$$A_r \leq c_1 c \left(\left(\frac{1}{4}\right)^r + \alpha^{4r} \sum_{p=2}^{r-1} \left(\frac{1}{4\alpha^4}\right)^p \right),$$

where, from (ii), (iii), (v) and (vi), c is such that for all $p \in \{2, \dots, r-1\}$, $q \in \{0, \dots, r-1\}$, $l \in \{0, \dots, p-1\}$

$$\max(\nu Q^q P(Q_\otimes^{r-q-1} f^2), \nu Q^l P(Q_\otimes^{p-l-1} P(g \otimes g))) < c,$$

and g is defined as before (2.4) and the notation Q_\otimes is given in (2.3).

Now depending on the value of α , we obtain that:

- if $\alpha^2 \neq \frac{1}{2}$, then $A_r \leq c_1 c((\frac{1}{4})^r + \alpha^{4r})$;
- if $\alpha^2 = \frac{1}{2}$, then $A_r \leq c_1 c(r-1)(\frac{1}{4})^r$.

(e) For the second term appearing in (2.2), we have when $p = r$:

- if $\alpha = \frac{1}{2}$, then

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_1^r] \times \mathbb{P}(E_1^r) \leq c_1 c(\frac{1}{4})^r;$$

- if $\alpha \neq \frac{1}{2}$:
 - if $\alpha^2 = \frac{1}{2}$, then

$$\mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_1^r] \times \mathbb{P}(E_1^r) \leq c_1 (r-1)(\frac{1}{4})^r;$$

- if $\alpha^2 \neq \frac{1}{2}$, then

$$\begin{aligned} & \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_1^r] \times \mathbb{P}(E_1^r) \\ & \leq c_1 c \left(\left(\frac{\alpha^2}{2} \right)^r + \left(\frac{1}{4} \right)^r \right), \end{aligned}$$

where, from (ii), (iii), (v) and (vi), c is such that for all $l \in \{2, \dots, r-1\}$, $q \in \{0, \dots, l-1\}$

$$\begin{aligned} & \max(\nu Q^q P(Q^{l-q-1} P(g \otimes g) \otimes Q^{r-q-1} f^2), \\ & \nu Q^q P(Q^{l-q-1} P(g \otimes Q^{r-l-1} f^2) \otimes g)) < c \end{aligned}$$

and g is defined as before (2.4).

(f) For the second terms appearing in (2.2), and for the remaining term in the sum ($p \neq r$), let us denote by

$$B_r := \sum_{p=2}^{r-1} \mathbb{E}[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})/E_0^{p+1}, E_1^p] \times \mathbb{P}(E_1^p \cap E_0^{p+1}).$$

So we have:

- if $\alpha = \frac{1}{2}$, then $B_r \leq c_1 c(\frac{1}{4})^r$;
- if $\alpha \neq \frac{1}{2}$:
 - if $\alpha^2 = \frac{1}{2}$, then $B_r \leq c_1 c r^2 (\frac{1}{4})^r$;
 - if $\alpha^2 \neq \frac{1}{2}$, then $B_r \leq c_1 c(\alpha^{4r} + (\frac{\alpha^2}{2})^r + (\frac{1}{4})^r)$,

where c is defined in the same way as before.

Now the results of the Theorem 2.1 follow from (a)–(f) of part 2. \square

It leads us to an extension of Theorem 2.1 to the two empirical averages $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\text{II}}(f)$.

COROLLARY 2.2. *Let F satisfy (i)–(vi). Let $f \in F$ such that $(\mu, f) = 0$. We assume that hypothesis (H1) is fulfilled. Then for all $r \in \mathbb{N}$ and $n \in \mathbb{N}$,*

$$(2.5) \quad \mathbb{E}[(\overline{M}_{\mathbb{T}_r}(f))^4] \leq \begin{cases} c(\frac{1}{4})^{r+1}, & \text{if } \alpha^2 < \frac{1}{2}, \\ cr^2(\frac{1}{4})^{r+1}, & \text{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{4(r+1)}, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

and

$$(2.6) \quad \mathbb{E}[(\overline{M}_n^\Pi(f))^4] \leq \begin{cases} c(\frac{1}{4})^{r_n+1}, & \text{if } \alpha^2 < \frac{1}{2}, \\ cr_n^2(\frac{1}{4})^{r_n+1}, & \text{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{4(r_n+1)}, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

where the positive constant c depends on α and f and may differ line by line.

PROOF. The proof follows the same steps as in the proof of parts 2 and 3 of Theorem 2.11, and uses the results of the proof of Theorem 2.5 to get the control of the 4th order moment in incomplete generation. See Sections 2.2 and A.1 for more detail. \square

REMARK 2.3. If $f \in \mathcal{B}(S^3)$ is such that Pf^2 and Pf^4 exist and belong to F , with $Pf = 0$, then we have for all $r \in \mathbb{N}$ and for some positive constant c ,

$$(2.7) \quad \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4] \leq \frac{c}{|\mathbb{G}_r|^2}.$$

Indeed, let $M_{\mathbb{G}_r}(f) = \sum_{i \in \mathbb{G}_r} f(\Delta_i)$. We have

$$\begin{aligned} \mathbb{E}[(M_{\mathbb{G}_r}(f))^4] &= \mathbb{E}[M_{\mathbb{G}_r}(f^4)] + 6\mathbb{E}\left[\sum_{i \neq j \in \mathbb{G}_r} f^2(\Delta_i)f^2(\Delta_j)\right] \\ &\quad + 4\mathbb{E}\left[\sum_{i \neq j \in \mathbb{G}_r} f^3(\Delta_i)f(\Delta_j)\right] \\ &\quad + 12\mathbb{E}\left[\sum_{i \neq j \neq k \in \mathbb{G}_r} f^2(\Delta_i)f(\Delta_j)f(\Delta_k)\right] \\ &\quad + 24\mathbb{E}\left[\sum_{i \neq j \neq k \neq l \in \mathbb{G}_r} f(\Delta_i)f(\Delta_j)f(\Delta_k)f(\Delta_l)\right] \\ &= \mathbb{E}\left[\sum_{i \in \mathbb{G}_r} Pf^4(X_i)\right] + 6\mathbb{E}\left[\sum_{i \neq j \in \mathbb{G}_r} Pf^2(X_i)Pf^2(X_j)\right], \end{aligned}$$

where the last equality was obtained after conditioning by \mathcal{F}_r and using the fact that $Pf = 0$. Now, dividing by $|\mathbb{G}_r|^4$ leads us to

$$\begin{aligned} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4] &= \frac{6}{|\mathbb{G}_r|^2} \mathbb{E} \left[\frac{1}{|\mathbb{G}_r|^2} \sum_{i \neq j \in \mathbb{G}_r} Pf^2(X_i) Pf^2(X_j) \right] \\ &\quad + \frac{1}{|\mathbb{G}_r|^3} \mathbb{E} \left[\frac{1}{|\mathbb{G}_r|} \sum_{i \in \mathbb{G}_r} Pf^4(X_i) \right] \\ &\leq \frac{6}{|\mathbb{G}_r|^2} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf^2))^2] \\ &\quad + \frac{1}{|\mathbb{G}_r|^3} \mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^4)], \end{aligned}$$

and (2.7) then follows from the control of

$$(\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf^2))^2])_r \quad \text{and} \quad (\mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^4)])_r;$$

see [14].

REMARK 2.4. From Remark 2.3, we deduce that if $f \in \mathcal{B}(S^3)$ is such that Pf^2 and Pf^4 exist and belong to F , with $Pf = 0$, then we have for all $r \in \mathbb{N}$ and for some positive constant c ,

$$(2.8) \quad \mathbb{E}[(\overline{M}_{\mathbb{T}_r}(f))^4] \leq c \left(\frac{1}{4}\right)^{r+1}.$$

Indeed, from the equality

$$\overline{M}_{\mathbb{T}_r}(f) = \sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \overline{M}_{\mathbb{G}_q}(f),$$

we deduce that

$$\mathbb{E}[(\overline{M}_{\mathbb{T}_r}(f))^4] \leq \left(\sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \|\overline{M}_{\mathbb{G}_q}(f)\|_4 \right)^4,$$

where $\|\cdot\|_4$ stands for the L^4 -norm. We then infer from (2.7) that

$$\mathbb{E}[(\overline{M}_{\mathbb{T}_r}(f))^4] \leq c \left(\sum_{q=0}^r \frac{(\sqrt{2})^q}{2^{r+1}} \right)^4$$

for some positive constant c . (2.8) then follows from the last inequality.

2.2. *Strong law of large numbers on incomplete subtree.* We now turn to prove the strong law of large numbers for $\overline{M}_n^{\text{II}}(f)$, completing the work of Guyon [14], where the LLN was proved only for the two averages $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_{\mathbb{G}_r}(f)$.

THEOREM 2.5. *Let F satisfy (i)–(vi). Let $f \in F$ such that $(\mu, f) = 0$. We assume that hypothesis (H1) is fulfilled with $\alpha \in (0, \frac{\sqrt{8}}{2})$. Then $\overline{M}_n^\Pi(f)$ almost surely converges to 0 as n goes to ∞ .*

PROOF. From the decomposition

$$\overline{M}_n^\Pi(f) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{G_q}(f) + \frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}),$$

it is enough to check that

$$\sum_{n=1}^{\infty} \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right)^4 \right] < \infty.$$

Indeed, since $\overline{M}_{G_q}(f)$ almost surely converges to 0 (Corollary 15 in [14]), we deduce that the first term on the right-hand side of the previous decomposition almost surely converges to 0 (Lemma 13 in [14]). We have

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right)^4 \right] \\ (2.9) \quad &= \frac{1}{n^4} \mathbb{E} \left[\sum_{i=2^{r_n}}^n f^4(X_{\Pi(i)}) \right] + \frac{6}{n^4} \mathbb{E} \left[\sum_{i,j=2^{r_n}; i \neq j}^n f^2(X_{\Pi(i)}) f^2(X_{\Pi(j)}) \right] \\ &+ \frac{4}{n^4} \mathbb{E} \left[\sum_{i,j=2^{r_n}; i \neq j}^n f^3(X_{\Pi(i)}) f(X_{\Pi(j)}) \right] \\ &+ \frac{12}{n^4} \mathbb{E} \left[\sum_{i,j,k=2^{r_n}; i \neq j \neq k}^n f^2(X_{\Pi(i)}) f(X_{\Pi(j)}) f(X_{\Pi(k)}) \right] \\ &+ \frac{24}{n^4} \mathbb{E} \left[\sum_{i,j,k,l=2^{r_n}; i \neq j \neq k \neq l}^n f(X_{\Pi(i)}) f(X_{\Pi(j)}) f(X_{\Pi(k)}) f(X_{\Pi(l)}) \right]. \end{aligned}$$

We will control each term appearing in decomposition (2.9). For the first term on the right-hand side of (2.9), using (ii), (v) and (vi) we have for some positive constant c ,

$$\mathbb{E} \left[\sum_{i=2^{r_n}}^n f^4(X_{\Pi(i)}) \right] = (n - 2^{r_n} + 1) \nu Q^{r_n} f^4 \leq c(n - 2^{r_n} + 1),$$

which implies that

$$(2.10) \quad \frac{1}{n^4} \mathbb{E} \left[\sum_{i=2^{r_n}}^n f^4(X_{\Pi(i)}) \right] = O\left(\frac{1}{n^3}\right).$$

Recall the following: for i, j, k and $l \in \{2^{r_n}, \dots, n\}$:

- If $i \neq j$, then $r_n \geq 1$. Independently on (X, Π) , draw two independent indices I_{r_n} and J_{r_n} uniformly from \mathbb{G}_{r_n} . Then the law of $(\Pi(i), \Pi(j))$ is the conditional law of (I_{r_n}, J_{r_n}) given $\{I_{r_n} \neq J_{r_n}\}$.
- If $i \neq j \neq k$, then $r_n \geq 2$. Independently on (X, Π) , draw three independent indices I_{r_n}, J_{r_n} and K_{r_n} uniformly from \mathbb{G}_{r_n} . Then the law of $(\Pi(i), \Pi(j), \Pi(k))$ is the conditional law of $(I_{r_n}, J_{r_n}, K_{r_n})$ given $\{I_{r_n} \neq J_{r_n} \neq K_{r_n}\}$.
- If $i \neq j \neq k \neq l$, then $r_n \geq 2$. Independently on (X, Π) , draw four independent indices $I_{r_n}, J_{r_n}, K_{r_n}$ and L_{r_n} uniformly from \mathbb{G}_{r_n} . Then the law of $(\Pi(i), \Pi(j), \Pi(k), \Pi(l))$ is the conditional law of $(I_{r_n}, J_{r_n}, K_{r_n}, L_{r_n})$ given $\{I_{r_n} \neq J_{r_n} \neq K_{r_n} \neq L_{r_n}\}$.

Now we have to control the second and third terms of (2.9). We have to check that

$$(2.11) \quad \frac{1}{n^4} \mathbb{E} \left[\sum_{i, j=2^{r_n}; i \neq j}^n f^2(X_{\Pi(i)}) f^2(X_{\Pi(j)}) \right] = O\left(\frac{1}{n^2}\right)$$

and

$$(2.12) \quad \frac{1}{n^4} \mathbb{E} \left[\sum_{i, j=2^{r_n}; i \neq j}^n f^3(X_{\Pi(i)}) f(X_{\Pi(j)}) \right] = o\left(\frac{1}{n^2}\right).$$

Indeed, from the previous reminder and (i)–(vi), we have for some positive constant c ,

$$\begin{aligned} & \mathbb{E} \left[\sum_{i, j=2^{r_n}; i \neq j}^n f^2(X_{\Pi(i)}) f^2(X_{\Pi(j)}) \right] \\ &= \frac{(n - 2^{r_n})(n - 2^{r_n} + 1)}{(1 - 2^{-r_n})} \\ & \quad \times \sum_{p=0}^{r_n-1} 2^{-p-1} \nu Q^p P(Q_{\otimes}^{r_n-p-1} f^2) \\ & \leq c(n - 2^{r_n})(n - 2^{r_n} + 1), \end{aligned}$$

which implies (2.11). In the same way and using in addition hypothesis (H1), we obtain that

$$\begin{aligned} & \mathbb{E} \left[\sum_{i, j=2^{r_n}; i \neq j}^n f^3(X_{\Pi(i)}) f(X_{\Pi(j)}) \right] \\ &= \frac{(n - 2^{r_n})(n - 2^{r_n} + 1)}{(1 - 2^{-r_n})} \end{aligned}$$

$$\begin{aligned}
 & \times \sum_{p=0}^{r_n-1} 2^{-p-2} \nu Q^p P(Q^{r_n-p-1} f^3 \otimes Q^{r_n-p-1} f \\
 & \qquad \qquad \qquad + Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f^3) \\
 & \leq \begin{cases} c2^{-r_n}(n-2^{r_n})(n-2^{r_n}+1), & \text{if } \alpha < \frac{1}{2}, \\ cr_n 2^{-r_n}(n-2^{r_n})(n-2^{r_n}+1), & \text{if } \alpha = \frac{1}{2}, \\ c\alpha^{r_n}(n-2^{r_n})(n-2^{r_n}+1), & \text{if } \alpha > \frac{1}{2}, \end{cases}
 \end{aligned}$$

which implies (2.12).

Let us deal with the remaining term of (2.9):

$$\begin{aligned}
 & \frac{1}{n^4} \mathbb{E} \left[\sum_{i,j,k=2^{r_n}; i \neq j \neq k}^n f^2(X_{\Pi(i)}) f(X_{\Pi(j)}) f(X_{\Pi(k)}) \right] \\
 & = \frac{(n-2^{r_n}-1)(n-2^{r_n})(n-2^{r_n}+1)}{\mathbb{P}(I_{r_n} \neq J_{r_n} \neq K_{r_n}) \times n^4} \\
 & \quad \times \mathbb{E}[f^2(X_{I_{r_n}}) f(X_{J_{r_n}}) f(X_{K_{r_n}}) \mathbf{1}_{\{I_{r_n} \neq J_{r_n} \neq K_{r_n}\}}].
 \end{aligned}$$

Then, we get an explicit expression for the last expectation similar to that obtained in part (d) of the calculus of $\mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^4]$ with a slight modification of the functions. Calculating the rate of this expression, we obtain

$$\begin{aligned}
 & \sum_{n=4}^{\infty} \frac{1}{n^4} \mathbb{E} \left[\sum_{i,j,k=2^{r_n}; i \neq j \neq k}^n f^2(X_{\Pi(i)}) f(X_{\Pi(j)}) f(X_{\Pi(k)}) \right] \\
 & \leq c \sum_{n=1}^{\infty} \frac{1}{n} \alpha^{2r_n} + c \sum_{n=1}^{\infty} \sum_{p=2}^{r_n-1} \sum_{l=0}^{p-1} \frac{1}{n} \frac{1}{2^p} \frac{1}{2^{l+1}} \alpha^{2r_n-2p} \\
 & \quad + c \sum_{n=1}^{\infty} \sum_{p=2}^{r_n-1} \sum_{l=0}^{p-1} \frac{1}{n} \frac{1}{2^p} \frac{1}{2^{l+1}} \alpha^{2r_n-p-l}
 \end{aligned}$$

for some positive c . Now it is not hard to see that the right-hand side is finite.

Finally, to check that the series of general term

$$\frac{1}{n^4} \mathbb{E} \left[\sum_{i,j,k,l=2^{r_n}; i \neq j \neq k \neq l}^n f(X_{\Pi(i)}) f(X_{\Pi(j)}) f(X_{\Pi(k)}) f(X_{\Pi(l)}) \right]$$

is finite, it is enough, according to the calculation of rates we have done in part 2 of the proof of Theorem 2.1, to check that $\sum_{n=1}^{\infty} \alpha^{4r_n} < \infty$, which is the case if $\alpha \in (0, \frac{\sqrt[4]{8}}{2})$, and this completes the proof of Theorem 2.5. \square

REMARK 2.6. Note that this theorem can be improved, but the price to pay is enormous computations related to the calculation of higher moments. If f is bounded, this result is true for every $\alpha \in (0, 1)$, as we will see in Section 3.

2.3. *Law of the iterated logarithm (LIL).* Using the LIL for martingales (see Theorem B.3 of Stout in Appendix B), we are going to prove a LIL for the BMC. This will be done when f depends on the mother-daughters triangle (Δ_i) . We use the notation $M_n^\Pi(f) = \sum_{i=1}^n f(\Delta_{\Pi(i)})$ and $M_{\mathbb{T}_r}(f) = \sum_{i \in \mathbb{T}_r} f(\Delta_i)$.

THEOREM 2.7. *Let F satisfy (i)–(vi). Let $f \in \mathcal{B}(\mathcal{S}^3)$ such that $Pf = 0$, Pf^2 and Pf^4 exist and belong to F . We assume that hypothesis (H1) is fulfilled. Then*

$$\limsup_{n \rightarrow \infty} \frac{M_n^\Pi(f)}{\sqrt{2 \langle M^\Pi(f) \rangle_n \log \log \langle M^\Pi(f) \rangle_n}} = 1 \quad a.s.$$

And in particular,

$$\limsup_{r \rightarrow \infty} \frac{M_{\mathbb{T}_r}(f)}{\sqrt{2 |\mathbb{T}_r| \log \log |\mathbb{T}_r|}} = \sqrt{(\mu, Pf^2)} \quad a.s.$$

PROOF. We will check the hypothesis of Stout Theorem's B.3. Let $f \in \mathcal{B}(\mathcal{S}^3)$. We introduce the filtration $(\mathcal{H}_n)_{n \geq 0}$ defined by $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \leq i \leq n)$. Let $(M_n^\Pi(f))_{n \geq 0}$ defined by $M_0^\Pi(f) = 0$ and $M_n^\Pi(f) = \sum_{i=1}^n f(\Delta_{\Pi(i)})$. Then since $Pf = 0$, $(M_n^\Pi(f))$ is a \mathcal{H}_n -martingale with $\mathbb{E}[M_1^\Pi(f)] = 0$. The bracket of the above martingale is given by

$$\langle M^\Pi(f) \rangle_n = \sum_{i=0}^n Pf^2(X_{\Pi(i)}) = M_n^\Pi(Pf^2).$$

We have the following decomposition:

$$\frac{\langle M^\Pi(f) \rangle_n}{n} = \overline{M}_n^\Pi(Pf^2) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(Pf^2) + \frac{1}{n} \sum_{i=2^{r_n}}^n Pf^2(X_{\Pi(i)}).$$

Since

$$\forall q \leq r_n - 1 \quad \frac{2^q}{2^{r_n+1}} \leq \frac{2^q}{n} \leq \frac{2^q}{2^{r_n}} \quad \text{and} \quad \frac{1}{n} \sum_{i=2^{r_n}}^n Pf^2(X_{\Pi(i)}) \leq \overline{M}_{\mathbb{G}_{r_n}}(Pf^2),$$

we deduce that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{2^{r_n+1}} \overline{M}_{\mathbb{G}_q}(Pf^2) \leq \overline{M}_n^\Pi(Pf^2) \leq \sum_{q=0}^{r_n} \frac{2^q}{2^{r_n}} \overline{M}_{\mathbb{G}_q}(Pf^2).$$

From the strong law of large numbers of $\overline{M}_{\mathbb{G}_q}(Pf^2)$ (see [14], Corollary 15) and from Lemma 5.2 of [7], we infer that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{2^{r_n+1}} \overline{M}_{\mathbb{G}_q}(Pf^2) \xrightarrow{\text{a.s.}} \frac{(\mu, Pf^2)}{2} \quad \text{and} \quad \sum_{q=0}^{r_n} \frac{2^q}{2^{r_n}} \overline{M}_{\mathbb{G}_q}(Pf^2) \xrightarrow{\text{a.s.}} 2(\mu, Pf^2).$$

Using these results, we thus deduce that $\langle M^\Pi(f) \rangle_n = O(n)$ and $n = O(\langle M^\Pi(f) \rangle_n)$ a.s. This implies in particular that $\langle M^\Pi(f) \rangle_n \xrightarrow[n \rightarrow \infty]{} \infty$ a.s.

Now let $K_n = \frac{\sqrt{2}}{\sqrt{\log \log(n)}}$ in Theorem B.3, and we have

$$\begin{aligned} R &:= \sum_{n=1}^{\infty} \frac{2 \log \log \langle M^\Pi(f) \rangle_n}{K_n^2 \langle M^\Pi(f) \rangle_n} \\ &\quad \times \mathbb{E}[f^2(\Delta_{\Pi(n)}) \mathbf{1}_{\{f^2(\Delta_{\Pi(n)}) > K_n^2 \langle M^\Pi(f) \rangle_n / (2 \log \log \langle M^\Pi(f) \rangle_n)\}}] / \mathcal{H}_{n-1} \\ &\leq \sum_{n=1}^{\infty} \frac{4(\log \log \langle M^\Pi(f) \rangle_n)^2}{K_n^4 (\langle M^\Pi(f) \rangle_n)^2} Pf^4(X_{\Pi(n)}) \quad \text{a.s.}, \end{aligned}$$

since $\langle M^\Pi(f) \rangle_n = O(n)$ a.s., so that for $R < \infty$ a.s., it is enough to check that

$$(2.13) \quad \sum_{n=1}^{\infty} \frac{Pf^4(X_{\Pi(n)})}{n^\delta} < \infty \quad \text{a.s. with any } 1 < \delta < 2.$$

Now, according to (v) and (vi), there exists a positive constant c such that for all $n \geq 1$, $\mathbb{E}[Pf^4(X_{\Pi(n)})] = \nu Q^{r_n} Pf^4 \leq c$, and (2.13) follows. Applying Theorem B.3, we have

$$\limsup_{n \rightarrow \infty} \frac{M_n^\Pi(f)}{\sqrt{2 \langle M^\Pi(f) \rangle_n \log \log \langle M^\Pi(f) \rangle_n}} = 1 \quad \text{a.s.}$$

Now, for $n = |\mathbb{T}_r|$, we have the following:

$$\begin{aligned} &\frac{M_{\mathbb{T}_r}(f)}{\sqrt{2 \langle M^\Pi(f) \rangle_{|\mathbb{T}_r|} \log \log \langle M^\Pi(f) \rangle_{|\mathbb{T}_r|}}} \\ &= \sqrt{\frac{|\mathbb{T}_r| \langle M^\Pi(f) \rangle_{|\mathbb{T}_r|} / |\mathbb{T}_r|}{2 \log \log \langle M^\Pi(f) \rangle_{|\mathbb{T}_r|}}} \times \frac{M_{\mathbb{T}_r}(f)}{|\mathbb{T}_r| \langle M^\Pi(f) \rangle_{|\mathbb{T}_r|} / |\mathbb{T}_r|} \end{aligned}$$

and since $\frac{\langle M^\Pi(f) \rangle_{|\mathbb{T}_r|}}{|\mathbb{T}_r|} = \overline{M}_{\mathbb{T}_r}(Pf^2) \xrightarrow[r \rightarrow \infty]{} (\mu, Pf^2)$ a.s. (see Theorem 18 in [14]), we get

$$\limsup_{r \rightarrow \infty} \frac{M_{\mathbb{T}_r}(f)}{\sqrt{2 |\mathbb{T}_r| \log \log |\mathbb{T}_r|}} = \sqrt{(\mu, Pf^2)} \quad \text{a.s.},$$

which completes the proof. \square

REMARK 2.8. Let us note that using Theorem 2.5, we can prove that if hypothesis (H1) is fulfilled with $\alpha \in (0, \frac{\sqrt[3]{8}}{2})$, then

$$\limsup_{n \rightarrow \infty} \frac{M_n^\Pi(f)}{\sqrt{2n \log \log n}} = \sqrt{(\mu, Pf^2)} \quad \text{a.s.},$$

and via the computation of $2k$ th order moments of $\overline{M}_{G_r}(g)$, with $k > 2$ and $g \in \mathcal{B}(S)$, it is possible to prove the latter for all $\alpha \in (0, 1)$. But, as already emphasized, this comes at the price of enormous computations.

2.4. *Almost-sure functional central limit theorem (ASFCLT).* We are now going to prove an ASFCLT theorem for the BMC $(X_n, n \in \mathbb{T})$. Here again, this will be done when f depends on the mother-daughters triangle by using the ASFCLT for discrete time martingale. We refer to Chaabane, Theorem B.4, Appendix B, for the definition of an ASFCLT.

THEOREM 2.9. *Let F satisfy (i)–(vi). Let $f \in \mathcal{B}(\mathcal{S}^3)$ such that $Pf = 0$, Pf^2 and Pf^4 exist and belong to F . We assume that hypothesis (H1) is fulfilled with $\alpha \in (0, \frac{\sqrt[3]{8}}{2})$. Then $M_n^\Pi(f)$ verifies an ASFCLT, when n goes to ∞ .*

PROOF. We use Theorem B.4. Let $(\mathcal{H}_n)_{n \in \mathbb{N}}$ be the filtration defined as in Section 2.3. Then $(M_n^\Pi(f))$ is a \mathcal{H}_n martingale. We have to check the hypotheses of Theorem B.4. For all $n \geq 1$, let $V_n = s\sqrt{n}$ where $s^2 = (\mu, Pf^2)$. Then according to Theorem 2.5,

$$\frac{\langle M^\Pi(f) \rangle_n}{V_n^2} = V_n^{-2} M_n^\Pi(Pf^2) \xrightarrow[n \rightarrow \infty]{} 1 \quad \text{a.s.}$$

Let $\varepsilon > 0$. We have

$$\begin{aligned} & \sum_{n \geq 1} \frac{1}{V_n^2} \mathbb{E}[f^2(\Delta_{\Pi(n)}) \mathbf{1}_{\{|f(\Delta_{\Pi(n)})| > \varepsilon V_n\}} / \mathcal{H}_{n-1}] \\ & \leq \frac{1}{\varepsilon^2 s^4} \sum_{n \geq 1} \frac{Pf^4(X_{\Pi(n)})}{n^2} \quad \text{a.s.} \end{aligned}$$

According to (v) and (vi), there exists a positive constant c such that for all $n \geq 1$, $\mathbb{E}[Pf^4(X_{\Pi(n)})] = \nu Q^{r_n} Pf^4 \leq c$, and therefore, $\forall \varepsilon > 0$

$$\sum_{n \geq 1} \frac{1}{V_n^2} \mathbb{E}[f^2(\Delta_{\Pi(n)}) \mathbf{1}_{\{|f(\Delta_{\Pi(n)})| > \varepsilon V_n\}} / \mathcal{H}_{n-1}] < \infty \quad \text{a.s.}$$

Finally, we have

$$\sum_{n \geq 1} \frac{1}{V_n^4} \mathbb{E}[f^4(\Delta_{\Pi(n)}) \mathbf{1}_{\{|f(\Delta_{\Pi(n)})| \leq V_n\}} / \mathcal{H}_{n-1}] \leq \frac{1}{s^4} \sum_{n \geq 1} \frac{Pf^4(X_{\Pi(n)})}{n^2} \quad \text{a.s.},$$

which as before is a.s. finite, and the proof is then complete. \square

REMARK 2.10. As before, let us note that this result can be extended to the general case $\alpha \in (0, 1)$, but at the price of enormous computation related to the computation of $2k$ -order moments, $k > 2$, for $\overline{M}_{G_r}(g)$, $g \in \mathcal{B}(S)$.

2.5. *Deviation inequalities for BMC.* We are now going to give some deviation inequalities under (i)–(vi) and (H1) for the empirical means (1.2) when $f \in \mathcal{B}(S)$ with $(\mu, f) = 0$ and when $f \in \mathcal{B}(S^3)$ with $(\mu, Pf) = 0$. This will help us in the sequel to obtain a MDP result in a general framework, that is, for functional of BMC with unbounded test functions. Let us recall that the main disadvantage of this “weak” set of assumptions is that the range of speed for the MDP is very restricted. However, we still work under geometric ergodicity assumption and general test function, which will not be the case when we would want to extend the MDP; see Section 3. Note that we postpone to Appendix A nearly all the proofs of this section, these proofs being quite long and technical.

THEOREM 2.11. *Let F satisfy conditions (i)–(vi). We assume that (H1) is fulfilled. Let $f \in F$ such that $(\mu, f) = 0$. Then we have for all $\delta > 0$ and all $r \in \mathbb{N}$ and all $n \in \mathbb{N}$,*

$$(2.14) \quad \mathbb{P}(|\overline{M}_{G_r}(f)| > \delta) \leq \begin{cases} \frac{c}{\delta^2} \left(\frac{1}{2}\right)^r, & \text{if } \alpha^2 < \frac{1}{2}; \\ \frac{c}{\delta^2} r \left(\frac{1}{2}\right)^r, & \text{if } \alpha^2 = \frac{1}{2}; \\ \frac{c}{\delta^2} \alpha^{2r}, & \text{if } \alpha^2 > \frac{1}{2}; \end{cases}$$

$$(2.15) \quad \mathbb{P}(|\overline{M}_n^{\Pi}(f)| > \delta) \leq \begin{cases} \frac{c}{\delta^2} \left(\frac{1}{2}\right)^{r_n+1}, & \text{if } \alpha^2 < \frac{1}{2}; \\ \frac{c}{\delta^2} r_n \left(\frac{1}{2}\right)^{r_n+1}, & \text{if } \alpha^2 = \frac{1}{2}; \\ \frac{c}{\delta^2} \alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \frac{1}{2}; \end{cases}$$

and

$$(2.16) \quad \mathbb{P}(|\overline{M}_{T_r}(f)| > \delta) \leq \begin{cases} \frac{c}{\delta^2} \left(\frac{1}{2}\right)^{r+1}, & \text{if } \alpha^2 < \frac{1}{2}; \\ \frac{c}{\delta^2} r \left(\frac{1}{2}\right)^{r+1}, & \text{if } \alpha^2 = \frac{1}{2}; \\ \frac{c}{\delta^2} \alpha^{2(r+1)}, & \text{if } \alpha^2 > \frac{1}{2}; \end{cases}$$

where the positive constant c depends on f and α and may differ term by term.

PROOF. See Section A.1 in Appendix A. \square

We shall also need an extension of Theorem 2.11 to the case when f does not only depend on an individual X_i , but on the mother-daughters triangle (Δ_i) .

THEOREM 2.12. *Let F satisfy conditions (i)–(vi). We assume that (H1) is fulfilled. Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. Then we have the same conclusion as in Theorem 2.11 for the three empirical averages given in (1.2): $\overline{M}_{G_r}(f)$, $\overline{M}_{T_r}(f)$ and $\overline{M}_n^{\text{II}}(f)$.*

PROOF. See Section A.2 in Appendix A. \square

We thus have the following first result on the superexponential convergence in probability, whose definition we present now:

DEFINITION 2.13. Let (E, d) a metric space. Let (Z_n) be a sequence of random variables valued in E , Z be a random variable valued in E and (v_n) be a rate. We say that Z_n converges v_n -superexponentially fast in probability to Z if for all $\delta > 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mathbb{P}(d(Z_n, Z) > \delta) = -\infty.$$

This “exponential convergence” with speed v_n will be shortened as

$$Z_n \xrightarrow[v_n]{\text{superexp}} Z.$$

We may now set:

PROPOSITION 2.14. *Let F satisfy conditions (i)–(vi). Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. We assume that (H1) is fulfilled. Let (b_n) be a sequence of increasing positive real numbers such that*

$$(2.17) \quad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \quad \frac{b_n}{\sqrt{n \log n}} \longrightarrow 0, \quad \frac{n}{b_n} \text{ is nondecreasing.}$$

Then

$$\overline{M}_n^{\text{II}}(f) \xrightarrow[b_n^2/n]{\text{superexp}} 0.$$

PROOF. The proof is a direct consequence of Theorem 2.12. \square

2.6. *Moderate deviations for BMC.* Now, using the MDP for martingale (see, e.g., [11, 24]), we are going to prove a MDP for BMC. We will use Proposition B.5, in Appendix B.

THEOREM 2.15. *Let F satisfy conditions (i)–(vi). We assume that (H1) is satisfied. Let $f \in \mathcal{B}(S^3)$ such that Pf^2 and Pf^4 exist and belong to F . Assume that $Pf = 0$. Let (b_n) be a sequence of increasing positive real numbers satisfying (2.17). If*

$$(2.18) \quad \limsup_{n \rightarrow \infty} \frac{n}{b_n^2} \log(n \operatorname{ess\,sup}_{1 \leq k \leq c^{-1}(b_{n+1})} \mathbb{P}(|f(\Delta_{\Pi(k)})| > b_n/\mathcal{H}_{k-1})) = -\infty,$$

where $c^{-1}(b_{n+1}) := \inf\{k \in \mathbb{N} : \frac{k}{b_k} \geq b_{n+1}\}$, then $(M_n^\Pi(f)/b_n)$ satisfies a MDP in \mathbb{R} with the speed b_n^2/n and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.

PROOF. First, note that under the hypothesis, $M_n^\Pi(f)$ is a \mathcal{H}_n -martingale, with $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \leq i \leq n)$. From Proposition B.5 in Appendix B, we only have to check conditions (C1) and (C3).

On one hand, (2.15) applied to $Pf^4 - (\mu, Pf^4)$ implies that for all $\delta > 0$,

$$\limsup_{n \rightarrow \infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n Pf^4(X_{\Pi(i)}) > (\mu, Pf^4) + \delta\right) = -\infty,$$

and this implies the exponential Lindeberg condition (see, e.g., [24]), that is, condition (C3).

On the other hand, we have $\langle M^\Pi(f) \rangle_n = M_n^\Pi(Pf^2)$ and (2.15) applied to $Pf^2 - (\mu, Pf^2)$ implies that

$$\overline{M}_n^\Pi(Pf^2 - (\mu, Pf^2)) \xrightarrow[b_n^2/n]{\text{superexp}} 0,$$

that is, condition (C1). \square

REMARK 2.16. One of the main difficulties in the application of this Theorem lies in the verification of (2.18). Note, however, that in the range of speed considered it is sufficient to have some uniform control in X_i of some moment of $f(X_i, X_{2i}, X_{2i+1})$ conditionally on X_i , which leads to condition of the type $P|f|^k$ bounded for some $k \geq 2$. It is, of course, the case if f is bounded.

REMARK 2.17. In the special case of model (1.1), we have (see Section 4), for f such that $Pf = 0$ and for all k ,

$$\mathbb{E}\left[\exp\left(\lambda \frac{b_n}{n} f(\Delta_{\Pi(k)})\right) / \mathcal{H}_{k-1}\right] = \exp\left(\frac{b_n^2}{n} \left(\frac{\lambda^2 Pf^2}{2n}\right)(X_{\Pi(k)})\right).$$

This condition implies that a MDP is satisfied for $(M_n^\Pi(f)/b_n)$. Indeed, if this relation is satisfied, we then have that for $\lambda \in \mathbb{R}$ the quantity

$$G_n(\lambda) = \frac{\lambda^2}{2n} \sum_{k=1}^n Pf^2(X_{\Pi(k)}) = \frac{\lambda^2}{2} \overline{M}_n^\Pi(Pf^2)$$

is an upper and lower cumulant (see, e.g., [24]), and we may apply Gärtner–Ellis-type methodology. In addition, due to (2.15) applied to $Pf^2 - (\mu, Pf^2)$, we have for $\lambda \in \mathbb{R}$,

$$G_n(\lambda) \xrightarrow[b_n^2/n]{\text{superexp}} \frac{\lambda^2(\mu, Pf^2)}{2},$$

which implies that $(M_n^\Pi(f)/b_n)$ satisfies a MDP in \mathbb{R} with the speed b_n^2/n and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.

3. Exponential deviation inequalities for BMC and consequences. We give here stronger deviation inequalities than the one obtained in Section 2, namely exponential deviation inequalities. Of course, it requires more stringent assumptions.

3.1. *Exponential deviation inequalities.* Let us consider the following hypothesis.

(H2) There exists a probability μ on (S, \mathcal{S}) such that, for all $f \in \mathcal{B}_b(S)$ with $(\mu, f) = 0$, there exists a positive constant c such that

$$|Q^r f(x)| \leq c\alpha^r \quad \text{for some } \alpha \in (0, 1) \text{ and for all } x \in S.$$

One can easily check that, under hypothesis (H2), $\mathcal{B}_b(S)$ fulfills hypothesis (i)–(vi) of the previous section.

Under this assumption, we will prove exponential deviation inequalities for $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^\Pi(f)$ when $f \in \mathcal{B}_b(S)$ with $(\mu, f) = 0$ [resp., $f \in \mathcal{B}_b(S^3)$ with $(\mu, Pf) = 0$].

THEOREM 3.1. *Assume that (H2) is satisfied. Let $f \in \mathcal{B}_b(S)$ such that $(\mu, f) = 0$. Then we have for all $\delta > 0$,*

$$(3.1) \quad \mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \begin{cases} \exp(c''\delta) \exp(-c'\delta^2|\mathbb{G}_r|), & \text{if } \alpha \leq \frac{1}{2}, \\ \exp(-c'\delta^2|\mathbb{G}_r|), & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{|\mathbb{G}_r|}{r}\right), & \text{if } \alpha^2 = \frac{1}{2}, \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^{2r}}\right), & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

$$(3.2) \quad \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \begin{cases} \exp(c''\delta) \exp(-c'\delta^2|\mathbb{T}_r|), \\ \quad \forall r \in \mathbb{N}, & \text{if } \alpha < \frac{1}{2}, \\ \exp(2c'\delta(r+1)) \exp(-c'\delta^2|\mathbb{T}_r|), \\ \quad \forall r \in \mathbb{N}, & \text{if } \alpha = \frac{1}{2}, \\ \exp(-c'\delta^2|\mathbb{T}_r|), \\ \quad \forall r \in \mathbb{N} \text{ such that } r > r_0 - 1, & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{|\mathbb{T}_r|}{r+1}\right), \\ \quad \forall r \in \mathbb{N} \text{ such that } r > r_0 - 1, & \text{if } \alpha = \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^{2(r+1)}}\right), \\ \quad \forall r \in \mathbb{N}^* \text{ such that } r > r_0 - 3, & \text{if } \alpha > \frac{\sqrt{2}}{2}, \end{cases}$$

and

$$(3.3) \quad \mathbb{P}(\overline{M}_n^{\Pi}(f) > \delta) \leq \begin{cases} \exp(c''\delta) \exp(-c'\delta^2n), \\ \quad \forall n \in \mathbb{N}, & \text{if } \alpha < \frac{1}{2}, \\ \exp(2c'\delta(r_n+1)) \exp(-c'\delta^2n), \\ \quad \forall n \in \mathbb{N}, & \text{if } \alpha = \frac{1}{2}, \\ \exp(-c'\delta^2n), \\ \quad \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{n}{r_n+1}\right), \\ \quad \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \alpha = \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^{2(r_n+1)}}\right), \\ \quad \forall n \in \mathbb{N}^* \text{ such that } r_n > r_0 - 2, & \text{if } \alpha > \frac{\sqrt{2}}{2}, \end{cases}$$

where $r_0 := \log(\frac{\delta}{c_0})/\log(\alpha)$, and c_0 , c' and c'' are positive constants which depend on α and f , and differ line by line; see the proofs for the dependence.

PROOF. The details of the proof are in Section A.3 in Appendix A. It relies mainly on successive conditioning, using carefully the uniform geometric ergodicity assumption to get rid of the conditioning. \square

The condition about α less than $1/2$ or greater is of course linked to the binary structure of the tree. The extension to m -ary tree will follow from the same ideas.

THEOREM 3.2. *Assume that (H2) is satisfied. Let $f \in \mathcal{B}_b(S^3)$ such that $(\mu, Pf) = 0$. Then we have the same conclusions, for the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$, as in the Theorem 3.1.*

PROOF. See Section A.4 in Appendix A. \square

Now, using the Borel–Cantelli Theorem and (3.3), we state easily the following:

COROLLARY 3.3. *Assume that (H2) is satisfied. Let $f \in \mathcal{B}_b(S)$ such that $(\mu, f) = 0$ [resp., $f \in \mathcal{B}_b(S^3)$ and $(\mu, Pf) = 0$]. Then $\overline{M}_n^{\Pi}(f)$ almost surely converges to 0 as n goes to ∞ .*

REMARK 3.4. Of course uniform ergodicity and bounded test functions are surely a very strong set of assumptions, but it is not so difficult to verify if the Markov chain’s daughters lie in a compact set. We are convinced that it is possible to consider the geometric ergodic case and bounded test functions, but for the price of tedious calculations that we will pursue in an other work. We will also investigate the use of transportation inequalities, leading to deviation inequality for Lipschitz test functions under some Wasserstein contraction property for the kernel P , in the spirit of the Theorems 2.5 or 2.11 in [12].

3.2. *Moderate deviation principle for BMC.* We introduce the following assumption on the speed of the MDP.

ASSUMPTION 1. Let (b_n) be an increasing sequence of positive real numbers such that

$$\frac{b_n}{\sqrt{n}} \longrightarrow +\infty$$

and:

- if $\alpha^2 < \frac{1}{2}$, the sequence (b_n) is such that $b_n/n \longrightarrow 0$;
- if $\alpha^2 = \frac{1}{2}$, the sequence (b_n) is such that $(b_n \log n)/n \longrightarrow 0$;
- if $\alpha^2 > \frac{1}{2}$, the sequence (b_n) is such that $(b_n \alpha^{r_n+1})/\sqrt{n} \longrightarrow 0$.

Using the MDP for martingale with bounded jumps (see, e.g., [9, 11]), we can now state the following:

THEOREM 3.5. *Assume that (H2) is satisfied. Let $f \in \mathcal{B}_b(S^3)$ such that $Pf = 0$. Let (b_n) be a sequence of real numbers satisfying the Assumption 1; then $(M_n^\Pi(f)/b_n)$ satisfies a MDP in S with the speed b_n^2/n and rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.*

PROOF. The proof easily follows from the previous exponential probability inequalities and the MDP for martingale with bounded jumps; see, for example, [9, 11, 24]. \square

REMARK 3.6. Taking particularly $n = |\mathbb{T}_r|$ and (b_n) as a sequence of real numbers satisfying Assumption 1, we get that for all $f \in \mathcal{B}_b(S^3)$, $(M_{|\mathbb{T}_r|}(f)/b_{|\mathbb{T}_r|})$ satisfies a MDP in \mathbb{R} with the speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.

4. Application: First order Bifurcating autoregressive processes. In this section, we seek to apply the results of the previous sections to the following bifurcating autoregressive process with memory 1 defined by

$$(4.1) \quad \mathcal{L}(X_1) = \nu \quad \text{and} \quad \forall n \geq 1 \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases}$$

where $\alpha_0, \alpha_1 \in (-1, 1)$; $\beta_0, \beta_1 \in \mathbb{R}$, $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of i.i.d. bivariate random variables and ν a probability measure on \mathbb{R} .

Several extensions of the model have been proposed and various estimators are studied in the literature for the unknown parameters; see, for instance, [2, 17–19, 25, 26]. See [4] for a relevant references.

Throughout this section, we assume that the distribution ν has finite moments of all orders.

In the sequel, we will study (4.1) in two settings:

- the Gaussian setting which corresponds to the case where $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of i.i.d. bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$ with

$$(4.2) \quad \Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1);$$

- the bounded setting which corresponds to the case where X_1 and $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$, which forms a sequence of centered i.i.d. bivariate random variables, take their values in a compact set. Let us note that in this case, $(X_n, n \in \mathbb{T})$ takes its values in a compact set.

Our main goal is to give deviation inequalities and MDP for the estimator of the 4-dimensional unknown parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ and for the statistical test defined in [14].

To estimate the 4-parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$, as well as σ^2 and ρ , assume we observe a complete subtree \mathbb{T}_{r+1} . The least square estimator $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$ of θ is given by (see [14]), for $\eta \in \{0, 1\}$,

$$(4.3) \quad \begin{cases} \hat{\alpha}_\eta^r = \frac{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i X_{2i+\eta} - (|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i)(|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\eta})}{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i^2 - (|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i)^2}, \\ \hat{\beta}_\eta^r = |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\eta} - \hat{\alpha}_\eta^r |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i. \end{cases}$$

Notice that in the Gaussian case, this least square estimator corresponds to the maximum likelihood estimator.

We also need to introduce the estimators of the conditional variance σ^2 and the conditional sister–sister correlation ρ . These estimators are naturally given by

$$(4.4) \quad \begin{cases} \hat{\sigma}_r^2 = \frac{1}{2|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} (\hat{\varepsilon}_{2i}^2 + \hat{\varepsilon}_{2i+1}^2), \\ \hat{\rho}_r = \frac{1}{\hat{\sigma}_r^2} \sum_{i \in \mathbb{T}_r} \hat{\varepsilon}_{2i} \hat{\varepsilon}_{2i+1}, \end{cases}$$

where the residues are defined by $\hat{\varepsilon}_{2i+\eta} = X_{2i+\eta} - \hat{\alpha}_\eta^r X_i - \hat{\beta}_\eta^r$, with $\eta \in \{0, 1\}$.

Let us denote by $\mathcal{C}_{\text{pol}}(\mathbb{R})$ [resp., $\mathcal{C}_{\text{pol}}(\mathbb{R}^3)$] the set of all continuous functions $f: \mathbb{R} \rightarrow \mathbb{R}$ (resp., $f: \mathbb{R}^3 \rightarrow \mathbb{R}$) such that $|f|$ is bounded above by a polynomial. From [14], we know that $\mathcal{C}_{\text{pol}}(\mathbb{R})$ fulfills hypotheses (i)–(vi).

We will take $F = \mathcal{C}_{\text{pol}}^1(\mathbb{R})$ the set of all \mathcal{C}^1 functions $f: \mathbb{R} \rightarrow \mathbb{R}$ such that $|f| + |f'|$ is bounded above by a polynomial. Then, one can check that F fulfills hypotheses (i)–(vi). Moreover, for all $f \in F$, hypothesis (H1) holds with $\alpha = \max(|\alpha_0|, |\alpha_1|)$. Let μ be the unique stationary distribution of the induced Markov chain $(Y_r, r \in \mathbb{N})$; see [14] for more details.

Let us denote by $\mathcal{C}_{\text{pol}}^1(\mathbb{R}^3)$ the set of all \mathcal{C}^1 functions $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ such that $|f| + |f'|$ is bounded above by a polynomial. We shall denote by \mathbf{x} (resp., $\mathbf{x}^2, \mathbf{xy}, \mathbf{y}, \dots$) the element of $\mathcal{C}_{\text{pol}}^1(\mathbb{R}^3)$ defined by $(x, y, z) \mapsto x$ (resp., x^2, xy, y, \dots).

We define two continuous functions $\mu_1: \Theta \rightarrow \mathbb{R}$ and $\mu_2: \Theta \times \mathbb{R}_+^* \rightarrow \mathbb{R}$ by writing

$$(4.5) \quad (\mu, \mathbf{x}) = \mu_1(\theta) \quad \text{and} \quad (\mu, \mathbf{x}^2) = \mu_2(\theta, \sigma^2),$$

where $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1) \in \Theta = (-1, 1) \times \mathbb{R} \times (-1, 1) \times \mathbb{R}$.

To segregate between $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ and its alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$, we shall use the test statistic

$$\chi_r^{(1)} = \frac{|\mathbb{T}_r|}{2\hat{\sigma}_r^2(1 - \hat{\rho}_r)} \{(\hat{\alpha}_0^r - \hat{\alpha}_1^r)^2(\hat{\mu}_{2,r} - \hat{\mu}_{1,r}^2) + ((\hat{\alpha}_0^r - \hat{\alpha}_1^r)\hat{\mu}_{1,r} + \hat{\beta}_0^r - \hat{\beta}_1^r)^2\},$$

where we write $\hat{\mu}_{1,r} = \mu_1(\hat{\theta}^r)$ and $\hat{\mu}_{2,r} = \mu_2(\hat{\theta}_r, \hat{\sigma}_r)$.

As usual the Gaussian setting has specific properties that allow easier calculations and more general assumptions.

4.1. *The Gaussian setting.* We introduce the following assumption on the speed of the MDP. Let (b_n) be an increasing sequence of positive real numbers such that

$$(4.6) \quad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty \quad \text{and} \quad \frac{b_n}{\sqrt{n \log n}} \rightarrow 0.$$

PROPOSITION 4.1. *Let (b_n) be a sequence of real numbers satisfying (4.6). Then*

$$\hat{\theta}^r \xrightarrow[b_{|\mathbb{T}_r|^2/|\mathbb{T}_r|}^{\text{superexp}} \theta.$$

PROOF. We will treat the case of $\hat{\alpha}_0^r$ given in (4.3). The others, $\hat{\beta}_0^r, \hat{\alpha}_1^r$ and $\hat{\beta}_1^r$, given in (4.3), may be treated in a similar way. Note that $\hat{\alpha}_0^r = \frac{C_r}{B_r}$, where

$$C_r = \overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})\overline{M}_{\mathbb{T}_r}(\mathbf{y}) \quad \text{and} \quad B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2.$$

Now, using Lemma B.2 and Proposition 2.14, it follows that

$$\hat{\alpha}_0^r \xrightarrow[b_{|\mathbb{T}_r|^2/|\mathbb{T}_r|}^{\text{superexp}} \alpha_0. \quad \square$$

We recall that in the BAR model (4.1), we use $\alpha = \max\{|\alpha_0|, |\alpha_1|\}$, and $b := \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2$, where μ_1 and μ_2 are given in (4.5), so we have the following deviation inequality:

PROPOSITION 4.2. *For all $\delta > 0$, for all $r \in \mathbb{N}$ and for all $\gamma < \min(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}})$, where c_1 is a positive constant which depends on μ_1 , we have*

$$(4.7) \quad \mathbb{P}(\|\hat{\theta}^r - \theta\| > \delta) \leq \begin{cases} \frac{c}{\gamma^{4q}\delta^{4-p}} \left(\frac{1}{4}\right)^{r+1}, & \text{if } \alpha^2 < \frac{1}{2}, \\ \frac{c}{\gamma^{4q}\delta^{4-p}} r^2 \left(\frac{1}{4}\right)^{r+1}, & \text{if } \alpha^2 = \frac{1}{2}, \\ \frac{c}{\gamma^{4q}\delta^{4-p}} \alpha^{4(r+1)}, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

where the constant c depends on α , μ_1 , μ_2 and differs line by line, $p = p(\delta) \in \{0, 2, 4\}$ and $q = q(\delta) \in \{0, 1\}$.

REMARK 4.3. The values of p and q in Proposition 4.2 depend on the order of δ . For example, if δ is small enough, we have $p = 0$ and $q = 0$.

PROOF. See Section A.5 in Appendix A. \square

REMARK 4.4. Proposition 4.2 can be improved by calculating the $2k$ th order moments, with $k > 2$, as in the proof of Theorem 2.1. But, as we have said, this comes at the price of enormous computation.

PROPOSITION 4.5. Let (b_n) be a sequence of real numbers satisfying (4.6). Then

$$(\hat{\sigma}_r^2, \hat{\rho}_r) \xrightarrow[b_{|\mathbb{T}_r|^2}/|\mathbb{T}_r|]{\text{superexp}} (\sigma^2, \rho).$$

PROOF. Let us first deal with $\hat{\sigma}_r^2$ given in (4.4). We have (see, e.g., [14])

$$\hat{\sigma}_r^2 = \frac{1}{2} \overline{M}_{\mathbb{T}_r}(f(\cdot, \theta)) + D_r,$$

where $f(x, y, z, \theta) = (y - \alpha_0 x - \beta_0)^2 + (z - \alpha_1 x - \beta_1)^2$ and

$$D_r = \frac{1}{2|\mathbb{T}_r|} \sum_{i \in \mathbb{T}_r} (f(\Delta_i, \hat{\theta}^r) - f(\Delta_i, \theta)).$$

By the Taylor–Lagrange formula, we can find $g \in \mathcal{C}_{\text{pol}}(\mathbb{R}^3)$ such that (see [14])

$$|D_r| \leq \frac{1}{2} \|\hat{\theta}^r - \theta\| (1 + \|\theta\| + \|\hat{\theta}^r - \theta\|) \overline{M}_{\mathbb{T}_r}(g).$$

Now, Propositions 2.14 and 4.1 lead us to

$$\hat{\sigma}_r^2 \xrightarrow[b_{|\mathbb{T}_r|^2}/|\mathbb{T}_r|]{\text{superexp}} \sigma^2.$$

The proof for $\hat{\rho}_r$ given in (4.4) is similar. \square

PROPOSITION 4.6. Let (b_n) be a sequence of real numbers satisfying (4.6). Then the sequence $(|\mathbb{T}_r|(\hat{\theta}^r - \theta)/b_{|\mathbb{T}_r|})$ satisfies the MDP on \mathbb{R}^4 with the speed $b_{|\mathbb{T}_r|^2}/|\mathbb{T}_r|$ and the rate function I given by

$$I(x) = \frac{1}{2} x^t (\Sigma')^{-1} x,$$

where

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix}$$

with

$$K = \frac{1}{\mu_2(\theta, \sigma^2) - \mu_1(\theta)^2} \begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta, \sigma^2) \end{pmatrix}.$$

PROOF. We first observe that

$$\frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}}(\hat{\theta}^r - \theta) = M(A_r, B_r) \cdot \frac{U^r(f)}{b_{|\mathbb{T}_r|}},$$

where $f = (f_1, f_2, f_3, f_4)^t = (\mathbf{x}\mathbf{y}, \mathbf{y}, \mathbf{x}\mathbf{z}, \mathbf{z})^t$, $U^r(f) = M_{\mathbb{T}_r}(f - Pf)$, $A_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x})$, $B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2$ and

$$M(A_r, B_r) = \begin{pmatrix} \frac{1}{B_r} & \frac{-A_r}{B_r} & 0 & 0 \\ \frac{-A_r}{B_r} & \frac{B_r + A_r^2}{B_r} & 0 & 0 \\ 0 & 0 & \frac{1}{B_r} & \frac{-A_r}{B_r} \\ 0 & 0 & \frac{-A_r}{B_r} & \frac{B_r + A_r^2}{B_r} \end{pmatrix}.$$

For the sake of simplicity we wrote $Pf = (Pf_1, Pf_2, Pf_3, Pf_4)^t$, where P denotes the \mathbb{T} -transition probability associated to BAR(1) process in the Gaussian case, which is given by

$$P(x, dy, dz) = \frac{1}{2\pi\sigma^2(1-\rho^2)} \times \exp\left(-\frac{1}{2} \begin{pmatrix} y - \alpha_0 x - \beta_0 \\ z - \alpha_1 x - \beta_1 \end{pmatrix}^t \Gamma^{-1} \begin{pmatrix} y - \alpha_0 x - \beta_0 \\ z - \alpha_1 x - \beta_1 \end{pmatrix}\right) dy dz,$$

where Γ is the covariance matrix defined in (4.2).

On one hand, from Proposition 2.14,

$$A_r \xrightarrow[b_{|\mathbb{T}_r|/|\mathbb{T}_r|}^{\text{superexp}} a := \mu_1(\theta) \quad \text{and} \quad B_r \xrightarrow[b_{|\mathbb{T}_r|/|\mathbb{T}_r|}^{\text{superexp}} b := \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2,$$

so that by Lemma B.2, we obtain

$$M(A_r, B_r) \xrightarrow[b_{|\mathbb{T}_r|/|\mathbb{T}_r|}^{\text{superexp}} M(a, b) := \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}.$$

On the other hand, let $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^t \in \mathbb{R}^4$. For all $x \in \mathbb{R}$, we have that

$$\begin{aligned} & P \exp(\lambda^t(f - Pf))(x) \\ &= \int_{\mathbb{R}^2} \exp\left(\sum_{i=1}^4 \lambda_i(f_i - Pf_i)\right)(x, y, z) P(x, dy, dz) \\ &= \int_{\mathbb{R}^2} \exp\left(\lambda^t \begin{pmatrix} xy - x(\alpha_0 x + \beta_0) \\ y - \alpha_0 x - \beta_0 \\ xz - x(\alpha_1 x + \beta_1) \\ z - \alpha_1 x - \beta_1 \end{pmatrix}\right) P(x, dy, dz) \end{aligned}$$

$$\begin{aligned}
&= \exp\left(-\begin{pmatrix} \alpha_0 x + \beta_0 \\ \alpha_1 x + \beta_1 \end{pmatrix}^t \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right) \\
&\quad \times \int_{\mathbb{R}^2} \exp\left(\begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \begin{pmatrix} y \\ z \end{pmatrix}\right) P(x, dy, dz).
\end{aligned}$$

We know that

$$\begin{aligned}
&\int_{\mathbb{R}^2} \exp\left(\begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \begin{pmatrix} y \\ z \end{pmatrix}\right) P(x, dy, dz) \\
&= \exp\left(\begin{pmatrix} \alpha_0 x + \beta_0 \\ \alpha_1 x + \beta_1 \end{pmatrix}^t \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right) \\
&\quad \times \exp\left(\frac{1}{2} \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \Gamma \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right).
\end{aligned}$$

Let $\Xi(x)$ denote the square matrix with entries $(Pf_i f_j - Pf_i P f_j)(x)$, for $1 \leq i, j \leq 4$. So we obtain that

$$\begin{aligned}
P \exp(\lambda^t(f - Pf))(x) &= \exp\left(\frac{1}{2} \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}^t \Gamma \begin{pmatrix} \lambda_1 x + \lambda_2 \\ \lambda_3 x + \lambda_4 \end{pmatrix}\right) \\
&= \exp\left(\frac{1}{2} \sum_{i,j=1}^4 \lambda_i \lambda_j (Pf_i f_j - Pf_i P f_j)(x)\right) \\
&= \exp\left(\frac{1}{2} \lambda^t \Xi(x) \lambda\right).
\end{aligned}$$

Recall that the filtration $(\mathcal{H}_n)_{n \geq 0}$ is defined by $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \leq i \leq n)$. Therefore, from the previous calculations, we deduce that for all $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}[\exp(\lambda^t(f - Pf)(\Delta_{\Pi(k)})) / \mathcal{H}_{k-1}] &= P(\exp(\lambda^t(f - Pf)))(X_{\Pi(k)}) \\
&= \exp\left(\frac{1}{2} \lambda^t \Xi(X_{\Pi(k)}) \lambda\right).
\end{aligned}$$

Now, recall that $(M_n^\Pi(f - Pf))_{n \in \mathbb{N}}$ is a (\mathcal{H}_n) -martingale and by straightforward calculations, its increasing process is given by $\langle M^\Pi(f - Pf) \rangle_n = \sum_{k=1}^n \Xi(X_{\Pi(k)})$. From the foregoing, we infer that

$$\left(\exp\left(\lambda^t M_n^\Pi(f - Pf) - \frac{\lambda^t \langle M^\Pi(f - Pf) \rangle_n \lambda}{2} \right) \right)_{n \in \mathbb{N}}$$

is a (\mathcal{H}_n) -martingale. It then follows that for all $\lambda \in \mathbb{R}^4$, $G_n(\lambda) = \frac{1}{2^n} \lambda^t \langle M^\Pi(f - Pf) \rangle_n \lambda$ is an upper and lower cumulant. Moreover, from Proposition 2.14 and Lemma B.2,

$$G_n(\lambda) \xrightarrow[b_{|\mathbb{T}_r|}^{\text{superexp}}]{|\mathbb{T}_r|} \frac{1}{2} \lambda^t \Sigma \lambda \quad \text{where } \Sigma = \sigma^2 \begin{pmatrix} K^{-1} & \rho K^{-1} \\ \rho K^{-1} & K^{-1} \end{pmatrix}.$$

We thus deduce that (see, e.g., [24]) $(M_n^\Pi(f)/b_n)$ satisfies a MDP on \mathbb{R}^4 with speed b_n^2/n and the rate function

$$(4.8) \quad J(x) = \frac{1}{2}x^t \Sigma^{-1}x.$$

Taking $n = |\mathbb{T}_r|$, it follows that $(U^r(f)/b_{|\mathbb{T}_r|})$ satisfies a MDP with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function J given in (4.8). Finally, using the contraction principle (see, e.g., [10]) as in [23], we get the result. \square

Let us now consider the test statistic.

PROPOSITION 4.7. *Let (b_n) a sequence of real numbers satisfying (4.6). Then under the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$, $\frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}}(\chi_r^{(1)})^{1/2}$ satisfies a MDP on \mathbb{R} with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function*

$$I'(y) = \begin{cases} \frac{y^2}{2}, & \text{if } y \in \mathbb{R}_+, \\ +\infty, & \text{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1 of H_0 , we have for all $A > 0$,

$$\limsup_{r \rightarrow \infty} \frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}^2} \log \mathbb{P}(\chi_r^{(1)} < A) = -\infty.$$

PROOF. We have

$$H_0 = \{g(\theta) = 0\} \quad \text{where } g(\theta) = (\alpha_0 - \alpha_1, \beta_0 - \beta_1)^t.$$

From Proposition 4.6, $(|\mathbb{T}_r|(\hat{\theta}^r - \theta)/b_{|\mathbb{T}_r|})$ satisfies a MDP on \mathbb{R}^4 with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function $I(x) = \frac{1}{2}x^t(\Sigma')^{-1}x$. So that, using the delta method for the MDP (see, e.g., [13], Theorem 3.1) we conclude that $(|\mathbb{T}_r|(g(\hat{\theta}^r) - g(\theta))/b_{|\mathbb{T}_r|})$ satisfies a MDP on \mathbb{R}^2 with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function

$$J(y) = \inf\{I(x); y = g'(\theta)x\}.$$

Identification of this rate function by usual optimization argument leads us to

$$(4.9) \quad J(x) = \frac{1}{2}x^t(\Sigma'')^{-1}x \quad \text{where } \Sigma'' = 2\sigma^2(1 - \rho)K.$$

Under the null hypothesis H_0 , we have $g(\theta) = 0$, so that $(|\mathbb{T}_r|g(\hat{\theta}^r)/b_{|\mathbb{T}_r|})$ satisfies a MDP on \mathbb{R}^2 with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and rate function J given in (4.9).

Now, since $K = K(\theta, \sigma)$ is a continuous function of (θ, σ) (see [14]), so that, letting $\hat{K}_r = K(\hat{\theta}^r, \hat{\sigma}_r)$, Lemma B.2, Propositions 4.6 and 4.5 entail that

$$\hat{\Sigma}_r'' = 2\hat{\sigma}_r^2(1 - \hat{\rho}_r)\hat{K}_r \xrightarrow[b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|]{\text{superexp}} \Sigma''.$$

It follows using the contraction principle (see, e.g., [23]) that

$$(|\mathbb{T}_r| \hat{\Sigma}_r''^{-1/2} g(\hat{\theta}^r) / b_{|\mathbb{T}_r|})$$

satisfies a MDP on \mathbb{R}^2 with speed $b_{|\mathbb{T}_r|}^2 / |\mathbb{T}_r|$ and the rate function $J'(y) = \frac{\|y\|^2}{2}$.

In particular,

$$\left\| \frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}} \hat{\Sigma}_r''^{-1/2} g(\hat{\theta}^r) \right\| = \frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}} \sqrt{\chi_r^{(1)}}$$

satisfies a MDP with speed $b_{|\mathbb{T}_r|}^2 / |\mathbb{T}_r|$ and the rate function I' given in the Proposition 4.7.

Now, under the alternative hypothesis H_1 ,

$$\frac{\chi_r^{(1)}}{|\mathbb{T}_r|} = g(\hat{\theta}^r)^t \hat{\Sigma}_r''^{-1} g(\hat{\theta}^r) \xrightarrow[b_{|\mathbb{T}_r|}^2 / |\mathbb{T}_r|]{\text{superexp}} g(\theta)^t (\Sigma'')^{-1} g(\theta) > 0,$$

so that $\chi_r^{(1)}$ converges $\frac{b_{|\mathbb{T}_r|}^2}{|\mathbb{T}_r|}$ -superexponentially fast to $+\infty$. This concludes the proof of the Proposition 4.7. \square

4.2. *Compact case: The uniformly ergodic setting.* We recall that the model under study in this section is the model (4.1) where we assume that the noise and initial state X_1 take their values in a compact set. The results will be given without proofs, since the proofs are similar to those done in the previous section. The novelty here is that the range of speed is improved in comparison to the previous section. However, we suppose that the process takes its values in a compact set, which is not the case in the previous section.

We take $F = \mathcal{C}_b^1(\mathbb{R})$ the set of all \mathcal{C}^1 functions bounded on \mathbb{R} . Therefore, one can easily check (as in [14], proof of Proposition 28) that hypothesis (H2) is satisfied with $\alpha = \max(|\alpha_0|, |\alpha_1|)$. We use the same notation as in the previous section.

Let us begin by the fact that the estimator of θ converges super exponentially fast to the true parameter.

PROPOSITION 4.8. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then we have*

$$\hat{\theta}^r \xrightarrow[b_{|\mathbb{T}_r|}^2 / |\mathbb{T}_r|]{\text{superexp}} \theta.$$

We may now refine this result by proving deviation inequality.

PROPOSITION 4.9. *For all $\delta > 0$ and for all $\gamma < \min(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}})$, where c_1 is a positive constant which depends on μ_1 , and for $r_0 :=$*

$\frac{\log(\gamma^q \delta^{1-p/2}/c_0)}{\log \alpha}$, we have

$$(4.10) \quad \mathbb{P}(\|\hat{\theta}^r - \theta\| > \delta) \leq \begin{cases} c_2 \exp(c'' \gamma^q \delta^{1-p/2}) \exp(-c' \gamma^{2q} \delta^{2-p} |\mathbb{T}_r|), \\ \quad \forall r \in \mathbb{N}, \quad \text{if } \alpha < \frac{1}{2}, \\ c_2 \exp(c' \gamma^q \delta^{1-p/2} (r+1) - c' \gamma^{2q} \delta^{2-p} |\mathbb{T}_r|), \\ \quad \forall r \in \mathbb{N}, \quad \text{if } \alpha = \frac{1}{2}, \\ c_2 \exp(-c' \gamma^{2q} \delta^{2-p} |\mathbb{T}_r|), \\ \quad \forall r > r_0, \quad \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ c_2 \exp\left(-c' \gamma^q \delta^{2-p} \frac{|\mathbb{T}_r|}{r+1}\right), \\ \quad \forall r > r_0, \quad \text{if } \alpha = \frac{\sqrt{2}}{2}, \\ c_2 \exp\left(-c' \gamma^{2q} \delta^{2-p} \frac{1}{\alpha^{2(r+1)}}\right), \\ \quad \forall r > r_0, \quad \text{if } \alpha > \frac{\sqrt{2}}{2}, \end{cases}$$

where c_2 is a positive constant, c' and c'' depend on α , and c and may differ line by line, c_0 depends on α , c and γ , and may differ line by line, $p \in \{0, 1, 3/2\}$ and $q \in \{0, 1\}$.

We have now to consider super exponential convergence of the estimators of the other parameters.

PROPOSITION 4.10. *Let (b_n) a sequence of real numbers satisfying Assumption 1. Then we have*

$$(\hat{\sigma}_r^2, \hat{\rho}_r) \xrightarrow[b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|]{\text{superexp}} (\sigma^2, \rho).$$

As previously we may now prove MDP for the estimator of θ .

PROPOSITION 4.11. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then $(|\mathbb{T}_r|(\hat{\theta}^r - \theta)/b_{|\mathbb{T}_r|})$ satisfies the MDP on \mathbb{R}^4 with the speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and rate function*

$$I(x) = \frac{1}{2} x^t (\Sigma')^{-1} x,$$

where

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix}$$

with

$$K = \frac{1}{\mu_2(\theta, \sigma^2) - \mu_1(\theta)^2} \begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta, \sigma^2) \end{pmatrix}.$$

REMARK 4.12. Notice that the proof of Proposition 4.11 does not need the cumulant method as in the proof of Proposition 4.6. Indeed, since we are in the bounded case, from MDP of martingale with bounded jumps (see [9]), we need only to prove the superexponential convergence of increasing process of the martingale. This convergence is easily obtained from Theorem 3.2.

Let us give us our last result by considering a MDP for the test statistic.

PROPOSITION 4.13. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then under the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$, $\frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}}(\chi_r^{(1)})^{1/2}$ satisfies a MDP on \mathbb{R} with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function*

$$I'(y) = \begin{cases} \frac{y^2}{2}, & \text{if } y \in \mathbb{R}_+, \\ +\infty, & \text{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1 of H_0 , we have for all $A > 0$,

$$\limsup_{r \rightarrow \infty} \frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}^2} \log \mathbb{P}(\chi_r^{(1)} < A) = -\infty.$$

APPENDIX A: PROOF OF THE EXPONENTIAL INEQUALITIES

This section is devoted to the proofs of Theorems 2.11, 2.12, 3.1, 3.2 and Proposition 4.2.

A.1. Proof of Theorem 2.11. Let $f \in F$ such that $(\mu, f) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\text{II}}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. By the Markov inequality, we get, for all $\delta > 0$,

$$\mathbb{P}(|\overline{M}_{\mathbb{G}_r}(f)| > \delta) = \mathbb{P}(|\overline{M}_{\mathbb{G}_r}(f)|^2 > \delta^2) \leq \frac{1}{\delta^2} \mathbb{E}[|\overline{M}_{\mathbb{G}_r}(f)|^2].$$

By Guyon (see [14]), we have

$$\mathbb{E}[|\overline{M}_{\mathbb{G}_r}(f)|^2] = \sum_{p=0}^r 2^{-p-1} \nu_{p < r} \nu Q^p P(Q^{r-p-1} f \otimes Q^{r-p-1} f).$$

Hypothesis (H1) implies that there exists $g \in F$ and $\alpha \in (0, 1)$ such that for all $p \in \{0, 1, \dots, r\}$,

$$\nu Q^p P(Q^{r-p-1} f \otimes Q^{r-p-1} f) \leq \alpha^{2(r-p-1)} \nu Q^p P(g \otimes g).$$

Next, hypotheses (iii), (v) and (vi) imply that there is a positive constant c such that for all $p \in \{0, 1, \dots, r\}$,

$$\alpha^{2(r-p-1)} \nu Q^p P(g \otimes g) \leq c \alpha^{2(r-p-1)}.$$

This leads us to

$$(A.1) \quad \begin{aligned} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^2] &\leq c \sum_{p=0}^r 2^{-p-1} \alpha^{2(r-p-1)} \\ &= \begin{cases} c \left(\frac{1}{2}\right)^r + c \frac{\alpha^{2r} - (1/2)^r}{2\alpha^2 - 1}, & \text{if } \alpha^2 \neq \frac{1}{2}, \\ c^r \left(\frac{1}{2}\right)^r, & \text{if } \alpha^2 = \frac{1}{2}, \end{cases} \end{aligned}$$

and therefore (2.14) follows.

Part 2. Let us now consider $\overline{M}_n^\Pi(f)$. By the Markov inequality and the triangle inequality, we get, for all $\delta > 0$,

$$(A.2) \quad \begin{aligned} &\mathbb{P}(|\overline{M}_n^\Pi(f)| > \delta) \\ &= \mathbb{P}(|\overline{M}_n^\Pi(f)|^2 > \delta^2) \leq \frac{1}{\delta^2} \mathbb{E}[(\overline{M}_n^\Pi(f))^2] \\ &\leq \frac{2}{\delta^2} \mathbb{E} \left[\left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) \right)^2 \right] + \frac{2}{\delta^2} \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right)^2 \right]. \end{aligned}$$

In the last inequality (A.2), we have used the decomposition

$$\overline{M}_n^\Pi(f) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}).$$

In what follows, the constant c may be slightly different from that of part 1 and may differ term by term. For the first term appearing in (A.2), we have

$$\mathbb{E} \left[\left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) \right)^2 \right] = \left\| \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) \right\|_2^2 \leq \left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \|\overline{M}_{\mathbb{G}_q}(f)\|_2 \right)^2.$$

Using (A.1), we get that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{n} \|\overline{M}_{\mathbb{G}_q}(f)\|_2 \leq \begin{cases} \frac{c}{n} \sum_{q=0}^{r_n-1} (\sqrt{2})^q \leq c \frac{\sqrt{2}^{r_n}}{n}, & \text{if } \alpha^2 < \frac{1}{2}, \\ \frac{c}{n} \sum_{q=0}^{r_n} q^{1/2} \sqrt{2}^q \leq c \frac{r_n^{1/2} \sqrt{2}^{r_n}}{n}, & \text{if } \alpha^2 = \frac{1}{2}, \\ \frac{c}{n} \sum_{q=0}^{r_n-1} (2\alpha)^q \leq c\alpha^{r_n}, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

which implies that

$$(A.3) \quad \mathbb{E} \left[\left(\sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) \right)^2 \right] \leq \begin{cases} c \frac{2^{r_n}}{n^2} \leq c \left(\frac{1}{2} \right)^{r_n+1}, & \text{if } \alpha^2 < \frac{1}{2}, \\ c \frac{r_n}{2^{r_n+1}}, & \text{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \frac{1}{2}. \end{cases}$$

Now, we have to control the second term in (A.2). As in Guyon [14], we have that

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right)^2 \right] \\ & \leq \frac{n - 2^{r_n} + 1}{n^2} \nu Q^{r_n} f^2 \\ & \quad + \frac{(n - 2^{r_n})(n - 2^{r_n} + 1)}{n^2(1 - 2^{-r_n})} \sum_{p=0}^{r_n-1} 2^{-p-1} \nu Q^p P(Q^{r_n-p-1} f \otimes Q^{r_n-p-1} f) \\ & \leq \frac{c}{n} + c \sum_{p=0}^{r_n-1} 2^{-p-1} \alpha^{2r_n-2p-2}. \end{aligned}$$

Discussing following the value of α , we obtain that

$$(A.4) \quad \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right)^2 \right] \leq \begin{cases} c \frac{1}{2^{r_n+1}}, & \text{if } \alpha^2 < \frac{1}{2}, \\ c \frac{r_n}{2^{r_n+1}}, & \text{if } \alpha^2 = \frac{1}{2}, \\ c\alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \frac{1}{2}. \end{cases}$$

Inequality (2.15) then follows from (A.3) and (A.4).

Part 3. The case of $\overline{M}_{\mathbb{T}_r}(f)$ can be deduced from the previous by taking $n = |\mathbb{T}_r|$.

A.2. Proof of Theorem 2.12. Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. By the Markov inequality, we get for all $\delta > 0$,

$$\begin{aligned} \mathbb{P}(|\overline{M}_{\mathbb{G}_r}(f)| > \delta) &\leq \frac{1}{\delta^2} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(f))^2] \\ &= \frac{1}{\delta^2} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf))^2] + \frac{1}{\delta^2} \frac{1}{|\mathbb{G}_r|} \mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^2 - (Pf)^2)] \\ &\leq \frac{1}{\delta^2} \mathbb{E}[(\overline{M}_{\mathbb{G}_r}(Pf))^2] + \frac{c}{\delta^2} \left(\frac{1}{2}\right)^r. \end{aligned}$$

The last inequality follows from the convergence of the sequence $(\mathbb{E}[\overline{M}_{\mathbb{G}_r}(Pf^2 - (Pf)^2)])_r$ (see [14]).

Now, using part 1 of the proof of Theorem 2.11 with Pf instead of f leads us to a similar inequality (2.14) in Theorem 2.12 for $f \in \mathcal{B}(S^3)$.

Part 2. Let us now treat $\overline{M}_n^{\Pi}(f)$. Using the two equalities

$$\begin{aligned} \overline{M}_n^{\Pi}(f) &= \sum_{q=0}^{r_n-1} \frac{|\mathbb{G}_q|}{n} \overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n} \sum_{i=2^{r_n}}^n f(\Delta_{\Pi(i)}), \\ \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(\Delta_{\Pi(i)}) \right)^2 \right] &= \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^n Pf(X_{\Pi(i)}) \right)^2 \right] \\ &\quad + \frac{1}{n} \mathbb{E} \left[\frac{1}{n} \sum_{i=2^{r_n}}^n (Pf^2 - (Pf)^2)(X_{\Pi(i)}) \right], \end{aligned}$$

and part 2 of the proof of Theorem 2.11 with Pf instead of f leads us to a similar inequality (2.15) in Theorem 2.12 for $f \in \mathcal{B}$.

Part 3. The case of $\overline{M}_{\mathbb{T}_r}(f)$ can be deduced from the previous by taking $n = |\mathbb{T}_r|$.

A.3. Proof of Theorem 3.1. Let $f \in \mathcal{B}_b(S)$ such that $(\mu, f) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. We have for all $\lambda > 0$ and for all $\delta > 0$

$$(A.5) \quad \mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp(-\lambda\delta|\mathbb{G}_r|) \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right].$$

By subtracting and adding terms, we get

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i))) \right. \right. \\ & \qquad \qquad \qquad \left. \left. \times \prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i)) / \mathcal{F}_{r-1} \right] \right]. \end{aligned}$$

Now using the fact that conditionally to the $(r-1)$ first generations the sequence $\{\Delta_i, i \in \mathbb{G}_{r-1}\}$ is a sequence of independent random variables, we have that

$$\begin{aligned} & \mathbb{E} \left[\mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i))) \right. \right. \\ & \qquad \qquad \qquad \left. \left. \times \prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i)) / \mathcal{F}_{r-1} \right] \right] \\ &= \mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i)) \right. \\ & \qquad \left. \times \prod_{i \in \mathbb{G}_{r-1}} \mathbb{E}[\exp(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i))) / \mathcal{F}_{r-1}] \right]. \end{aligned}$$

Using the Azuma–Bennett–Hoeffding inequalities [1, 3, 16] (see Lemma B.1 for more detail), we get according to (H2), for all $i \in \mathbb{G}_{r-1}$,

$$\mathbb{E}[\exp(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i))) / \mathcal{F}_{r-1}] \leq \exp(2\lambda^2 c^2 (1 + \alpha)^2).$$

This leads us to

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right] \leq \exp(\lambda^2 c^2 (1 + \alpha)^2 |\mathbb{G}_r|) \mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i)) \right].$$

Doing the same thing for $\mathbb{E}[\prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i))]$ with Qf replacing f , we get

$$\begin{aligned} & \mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i)) \right] \\ & \leq \exp(2\lambda^2 c^2 (\alpha + \alpha^2)^2 |\mathbb{G}_r|) \mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-2}} \exp(2^2 \lambda Q^2 f(X_i)) \right]. \end{aligned}$$

Iterating this procedure, we get

$$\begin{aligned} \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right] &\leq \mathbb{E} [\exp(2^r \lambda Q^r f(X_1))] \\ &\quad \times \prod_{k=1}^r \exp(2^{k-1} \lambda^2 c^2 (\alpha^{k-1} + \alpha^k)^2 |\mathbb{G}_r|). \end{aligned}$$

Once again, according to (H2), we have

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right] \leq \exp(\lambda c \alpha^r |\mathbb{G}_r|) \times \exp \left(\lambda^2 c^2 (1 + \alpha)^2 |\mathbb{G}_r| \sum_{k=1}^r (2\alpha^2)^{k-1} \right).$$

Hence:

- if $\alpha^2 \neq \frac{1}{2}$, then

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right] \leq \exp \left(\lambda^2 c^2 (1 + \alpha)^2 \frac{1 - (2\alpha^2)^r}{1 - 2\alpha^2} |\mathbb{G}_r| \right) \times \exp(\lambda c \alpha^r |\mathbb{G}_r|);$$

- if $\alpha^2 = \frac{1}{2}$, then

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(X_i) \right) \right] \leq \exp(\lambda^2 c^2 (1 + \alpha)^2 r |\mathbb{G}_r|) \times \exp \left(\lambda c \left(\frac{\sqrt{2}}{2} \right)^r |\mathbb{G}_r| \right).$$

We then consider three cases:

- (a) If $\alpha^2 < \frac{1}{2}$, then $\frac{1 - (2\alpha^2)^r}{1 - 2\alpha^2} < \frac{1}{1 - 2\alpha^2}$ for all r . Taking $\lambda = \frac{(1 - 2\alpha^2)\delta}{2c^2(1 + \alpha)^2}$ in (A.5) leads us to

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp \left(- \left(\frac{(1 - 2\alpha^2)\delta^2}{4c^2(1 + \alpha)^2} - \alpha^r \frac{(1 - 2\alpha^2)\delta}{2c(1 + \alpha)^2} \right) |\mathbb{G}_r| \right).$$

- If $\alpha \leq \frac{1}{2}$, then $(2\alpha)^r \leq 1$ for all $r \in \mathbb{N}$. We then have for all $r \in \mathbb{N}$,

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp \left(\frac{(1 - 2\alpha^2)\delta}{2c(1 + \alpha)^2} \right) \exp \left(- \frac{(1 - 2\alpha^2)\delta^2 |\mathbb{G}_r|}{4c^2(1 + \alpha)^2} \right).$$

- If $\frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}$, then for all $r \in \mathbb{N}$ such that $r > \log(\frac{\delta}{4c}) / \log \alpha$, we have $(\delta - 2c\alpha^r) > \frac{\delta}{2}$, and it then follows that

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp \left(- \frac{(1 - 2\alpha^2)\delta^2 |\mathbb{G}_r|}{8c^2(1 + \alpha)^2} \right).$$

- (b) If $\alpha^2 = \frac{1}{2}$, then for all $\lambda > 0$,

$$\begin{aligned} \mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) &\leq \exp((- \delta \lambda + c^2(1 + \alpha)^2 r \lambda^2) |\mathbb{G}_r|) \\ &\quad \times \exp \left(\lambda c \left(\frac{\sqrt{2}}{2} \right)^r |\mathbb{G}_r| \right). \end{aligned}$$

Taking $\lambda = \frac{\delta}{2c^2(1+\alpha)^{2r}}$, we are led to

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{\delta|\mathbb{G}_r|}{4c^2(1+\alpha)^{2r}}\left(\delta - 2c\left(\frac{\sqrt{2}}{2}\right)^r\right)\right).$$

For all $r \in \mathbb{N}$ such that $r > \log(\frac{\delta}{4c})/\log(\frac{\sqrt{2}}{2})$, we have $(\delta - 2c(\frac{\sqrt{2}}{2})^r) > \frac{\delta}{2}$ and for such r , it follows that

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{\delta^2|\mathbb{G}_r|}{18c^2r}\right).$$

(c) If $\alpha^2 > \frac{1}{2}$, then for all $\lambda > 0$,

$$\begin{aligned} \mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) &\leq \exp(-\lambda\delta|\mathbb{G}_r|) \times \exp\left(\lambda^2c^2(1+\alpha)^2\frac{(2\alpha^2)^r - 1}{2\alpha^2 - 1}|\mathbb{G}_r|\right) \\ &\quad \times \exp(\lambda c\alpha^r|\mathbb{G}_r|) \\ &\leq \exp\left(-|\mathbb{G}_r|\left(\lambda\delta - \frac{\lambda^2c^2(1+\alpha)^2}{2\alpha^2 - 1}(2\alpha^2)^r\right)\right) \\ &\quad \times \exp(\lambda c\alpha^r|\mathbb{G}_r|). \end{aligned}$$

Taking $\lambda = \frac{(2\alpha^2-1)\delta}{2c^2(1+\alpha)^2(2\alpha^2)^r}$ leads us to

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp\left(-\frac{(2\alpha^2-1)\delta}{4c^2(1+\alpha)^2\alpha^{2r}}(\delta - 2c\alpha^r)\right).$$

Now for all $r \in \mathbb{N}$ such that $r > \log(\frac{\delta}{4c})/\log \alpha$, we have

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f)) \leq \exp\left(-\frac{(2\alpha^2-1)\delta^2}{8c^2(1+\alpha)^2\alpha^{2r}}\right).$$

Part 2. Let us now deal with $\overline{M}_{\mathbb{T}_r}(f)$. We have for all $\lambda > 0$ and all $\delta > 0$,

$$(A.6) \quad \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \exp(-\lambda\delta|\mathbb{T}_r|)\mathbb{E}\left[\exp\left(\lambda\sum_{i \in \mathbb{T}_r} f(X_i)\right)\right].$$

By subtracting and adding terms, we get

$$\begin{aligned} &\mathbb{E}\left[\exp\left(\lambda\sum_{i \in \mathbb{T}_r} f(X_i)\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)))\right.\right. \\ &\quad \left.\left.\times \prod_{i \in \mathbb{G}_{r-1}} \exp(2\lambda Qf(X_i)) \times \prod_{i \in \mathbb{T}_{r-1}} \exp(\lambda f(X_i))/\mathcal{F}_{r-1}\right)\right] \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{E} \left[\mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i))) \right. \right. \\
 &\quad \left. \left. \times \prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f + 2Qf)(X_i)) \times \prod_{i \in \mathbb{T}_{r-2}} \exp(\lambda f(X_i)) / \mathcal{F}_{r-1} \right] \right].
 \end{aligned}$$

The fact that conditionally to the $(r - 1)$ first generations the sequence $\{\Delta_i, i \in \mathbb{G}_{r-1}\}$ is a sequence of independent random variables and Azuma–Bennett–Hoeffding inequality (see Lemma B.1) lead us according to (H2) to

$$\begin{aligned}
 &\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{T}_r} f(X_i) \right) \right] \\
 &\leq \exp(2\lambda^2 c^2 (1 + \alpha)^2 |\mathbb{G}_{r-1}|) \\
 &\quad \times \mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f + 2Qf)(X_i)) \prod_{i \in \mathbb{T}_{r-2}} \exp(\lambda f(X_i)) \right].
 \end{aligned}$$

Doing the same things for

$$\mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-1}} \exp(\lambda(f + 2Qf)(X_i)) \prod_{i \in \mathbb{T}_{r-2}} \exp(\lambda f(X_i)) \right]$$

with $f + 2Qf$ replacing f , we get

$$\begin{aligned}
 &\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{T}_r} f(X_i) \right) \right] \\
 &\leq \exp(2\lambda^2 c^2 (1 + \alpha)^2 |\mathbb{G}_{r-1}|) \times \exp(2\lambda^2 c^2 (1 + 3\alpha + 2\alpha^2)^2 |\mathbb{G}_{r-2}|) \\
 &\quad \times \mathbb{E} \left[\prod_{i \in \mathbb{G}_{r-2}} \exp(\lambda(f + 2Qf + 2^2 Q^2 f)(X_i)) \prod_{i \in \mathbb{T}_{r-3}} \exp(\lambda f(X_i)) \right].
 \end{aligned}$$

Iterating this procedure leads us to

$$\begin{aligned}
 &\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{T}_r} f(X_i) \right) \right] \\
 &\leq \exp \left(2\lambda^2 c^2 (1 + \alpha)^2 \sum_{q=1}^r \left(\sum_{k=0}^{q-1} (2\alpha)^k \right)^2 |\mathbb{G}_{r-q}| \right) \\
 &\quad \times \mathbb{E}[\exp(\lambda(f + 2Qf + 2^2 Q^2 f + \dots + 2^r Q^r f)(X_1))].
 \end{aligned}$$

Using (H2) we get

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{T}_r} f(X_i) \right) \right]$$

$$\leq \exp\left(\lambda c \sum_{k=0}^r (2\alpha)^k + 2\lambda^2 c^2 (1+\alpha)^2 \sum_{q=1}^r \left(\sum_{k=0}^{q-1} (2\alpha)^k\right)^2 |\mathbb{G}_{r-q}|\right).$$

Now for $\alpha \neq \frac{1}{2}$ and $\alpha^2 \neq \frac{1}{2}$ we have

$$\begin{aligned} & \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \\ & \leq \exp(-\lambda\delta|\mathbb{T}_r|) \exp\left(2\lambda^2 c^2 (1+\alpha)^2 \left(\frac{2^r - 1}{(1-2\alpha)^2} - \frac{\alpha(1-\alpha^r)2^{r+1}}{(1-2\alpha)^2(1-\alpha)}\right.\right. \\ & \quad \left.\left. + \frac{2\alpha^2(1-(2\alpha^2)^r)2^r}{(1-2\alpha)^2(1-2\alpha^2)}\right)\right) \\ & \quad \times \exp\left(\lambda c \frac{1-(2\alpha)^{r+1}}{1-2\alpha}\right) \\ & \leq \exp\left(-|\mathbb{T}_r| \left(\lambda\delta - \frac{\lambda^2 c^2 (1+\alpha)^2}{(1-2\alpha)^2} \left(1 + \frac{4\alpha^2(1-(2\alpha^2)^r)}{1-2\alpha^2}\right)\right)\right) \\ & \quad \times \exp\left(\lambda c \frac{1-(2\alpha)^{r+1}}{1-2\alpha}\right). \end{aligned}$$

Taking $\lambda = \frac{\delta}{(2c^2(1+\alpha)^2/(1-2\alpha)^2)(1+4\alpha^2(1-(2\alpha^2)^r)/(1-2\alpha^2))}$ leads us to

$$\begin{aligned} & \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \\ & \leq \exp\left(-|\mathbb{T}_r| \frac{(1-2\alpha)^2 \delta^2}{4c^2(1+\alpha)^2(1+4\alpha^2(1-(2\alpha^2)^r)/(1-2\alpha^2))}\right) \\ & \quad \times \exp\left(\frac{(1-2\alpha)^2 \delta}{2c(1+\alpha)^2(1+4\alpha^2(1-(2\alpha^2)^r)/(1-2\alpha^2))} \frac{1-(2\alpha)^{r+1}}{1-2\alpha}\right). \end{aligned}$$

- If $\alpha < \frac{1}{2}$, then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all $r \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) & \leq \exp\left(\frac{1-2\alpha}{2c(1+\alpha)^2} \delta\right) \\ & \quad \times \exp\left(-\frac{(1-2\alpha^2)(1-2\alpha)^2 \delta^2}{4c^2(1+\alpha)^2(1+2\alpha^2)} |\mathbb{T}_r|\right). \end{aligned}$$

- If $\frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}$, then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all $r \in \mathbb{N}$,

$$\begin{aligned} & \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \\ & \leq \exp\left(-\frac{(1-2\alpha^2)(2\alpha-1)^2 \delta |\mathbb{T}_r|}{4c^2(1+\alpha)^2(1+2\alpha^2)} \left(\delta - \frac{2c(1-2\alpha^2)\alpha^{r+1}}{(2\alpha-1)(1+2\alpha^2)}\right)\right). \end{aligned}$$

Now for all $r \in \mathbb{N}$ such that $r + 1 > \log\left(\frac{(2\alpha-1)(1+2\alpha^2)\delta}{4c(1-2\alpha^2)}\right)/\log \alpha$, we have $\delta - \frac{2c(1-2\alpha^2)\alpha^{r+1}}{(2\alpha-1)(1+2\alpha^2)} > \frac{\delta}{2}$ so that for such r , we have

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \exp\left(-\frac{(1-2\alpha^2)(2\alpha-1)^2\delta^2|\mathbb{T}_r|}{8c^2(1+\alpha)^2(1+2\alpha^2)}\right).$$

- If $\alpha^2 > \frac{1}{2}$, then for all $r \geq 1$, we have

$$\begin{aligned} & \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \\ & \leq \exp\left(-\frac{(2\alpha-1)^2(2\alpha^2-1)\delta}{32c^2(1+\alpha)^2\alpha^{2(r+1)}}\left(\delta - \frac{16\alpha^2c\alpha^{r+1}}{(2\alpha^2-1)(2\alpha-1)}\right)\right). \end{aligned}$$

For all $r \in \mathbb{N}^*$ such that $r + 3 > \log\left(\frac{(2\alpha^2-1)(2\alpha-1)\delta}{32c}\right)/\log \alpha$, we have $\delta - \frac{16\alpha^2c\alpha^{r+1}}{(2\alpha^2-1)(2\alpha-1)} > \frac{\delta}{2}$ so that

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \exp\left(-\frac{(1-2\alpha)^2(2\alpha^2-1)\delta^2}{64c^2(1+\alpha)^2}\left(\frac{1}{\alpha^2}\right)^{r+1}\right).$$

Now if $\alpha = \frac{1}{2}$, then $\sum_{q=1}^r \frac{q^2}{2^q} < \sum_{q=1}^{\infty} \frac{q^2}{2^q} = 6$. Then for all $\lambda > 0$,

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \exp(-(\lambda\delta - 27c^2\lambda^2)|\mathbb{T}_r|) \times \exp(\lambda c(r+1)).$$

Taking $\lambda = \frac{\delta}{54c^2}$ leads us to

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \exp\left(-\frac{\delta^2}{108c^2}|\mathbb{T}_r|\right) \times \exp\left(\frac{\delta}{54c}(r+1)\right).$$

Finally, if $\alpha^2 = \frac{1}{2}$, in the same way as previously, for all $r \in \mathbb{N}$ such that $r + 1 > \log\left(\frac{(\sqrt{2}-1)\delta}{4c}\right)/\log\left(\frac{\sqrt{2}}{2}\right)$, we have

$$\mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \exp\left(-\frac{(\sqrt{2}-1)^2\delta^2|\mathbb{T}_r|}{4c^2(1+\sqrt{2})^2 r+1}\right).$$

Part 3. Eventually, let us look at $\overline{M}_n^{\Pi}(f)$. We have for all $\delta > 0$

$$\mathbb{P}\left(\frac{1}{n}M_n^{\Pi}(f) > \delta\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i \in \mathbb{T}_{r_{n-1}}} f(X_i) > \frac{\delta}{2}\right) + \mathbb{P}\left(\frac{1}{n}\sum_{i=2^r n}^n f(X_{\Pi(i)}) > \frac{\delta}{2}\right).$$

On the one hand, (3.2) leads us to

$$(A.7) \quad \mathbb{P}\left(\frac{1}{n} \sum_{i \in \mathbb{T}_{r_n-1}} f(X_i) > \frac{\delta}{2}\right) \leq \begin{cases} \exp(c''\delta) \exp(-c'\delta^2 n), \\ \quad \forall n \in \mathbb{N}, & \text{if } \alpha < \frac{1}{2}, \\ \exp(2c'\delta(r_n+1)) \exp(-c'\delta^2 n), \\ \quad \forall n \in \mathbb{N}, & \text{if } \alpha = \frac{1}{2}, \\ \exp(-c'\delta^2 n), \\ \quad \forall r_n > r_0, & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{n}{r_n+1}\right), \\ \quad \forall r_n > r_0, & \text{if } \alpha = \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^{2(r_n+1)}}\right), \\ \quad \forall r_n > r_0 - 2, & \text{if } \alpha > \frac{\sqrt{2}}{2}, \end{cases}$$

where $r_0 := \log(\frac{\delta}{c_0})/\log \alpha$ and c_0 , c' and c'' are positive constants which depend on α , $\|f\|_\infty$ and c . c_0 , c' and c'' differ line by line. On the other hand, for all $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) > \frac{\delta}{2}\right) \leq \exp\left(-\frac{\lambda\delta}{2}n\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^n f(X_{\Pi(i)})\right)\right].$$

Now let:

- $\mathcal{O}_{r_n} = \{\Pi(2^{r_n}), \Pi(2^{r_n}+1), \dots, \Pi(n)\}$;
- $\mathcal{O}_{r_n-1}^1$ the set of individuals of generation \mathbb{G}_{r_n-1} which are ancestors of one individual in \mathcal{O}_{r_n} ;
- $\mathcal{O}_{r_n-1}^2$ the set of individuals of generation \mathbb{G}_{r_n-1} which are ancestors of two individuals in \mathcal{O}_{r_n} ;
- \mathcal{O}'_{r_n} the set of individuals of \mathcal{O}_{r_n} whose parents belong to $\mathcal{O}_{r_n-1}^1$;
- $\mathcal{O}_{r_n-1} = \mathcal{O}_{r_n-1}^1 \cup \mathcal{O}_{r_n-1}^2$.

We introduce the filtration $\tilde{\mathcal{F}}_r := \sigma(\mathcal{F}_r, \Pi(i), 1 \leq i \leq \mathbb{T})$. Then we have

$$\begin{aligned} & \mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^n f(X_{\Pi(i)})\right)\right] \\ &= \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathcal{O}_{r_n-1}^2} 2Qf(X_i) + \lambda \sum_{i \in \mathcal{O}_{r_n-1}^1} Qf(X_i)\right)\right] \end{aligned}$$

$$\begin{aligned} & \times \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}'_{r_n}} f(X_i) - Qf(X_{[i/2]}) \right) / \tilde{\mathcal{F}}_{r_n-1} \right] \\ & \times \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i) \right) / \tilde{\mathcal{F}}_{r_n-1} \right]. \end{aligned}$$

Using the Azuma–Bennett–Hoeffding inequality, as in part 1, we get

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}'_{r_n}} f(X_i) - Qf(X_{[i/2]}) \right) / \tilde{\mathcal{F}}_{r_n-1} \right] \leq \exp \left(\frac{\lambda^2 c^2 (1 + \alpha)^2}{2} |\mathcal{O}'_{r_n}| \right)$$

and

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i) \right) / \tilde{\mathcal{F}}_{r_n-1} \right] \\ & \leq \exp(2\lambda^2 c^2 (1 + \alpha)^2 |\mathcal{O}^2_{r_n-1}|). \end{aligned}$$

Now, we have

$$\begin{aligned} & \exp \left(\frac{\lambda^2 c^2 (1 + \alpha)^2}{2} |\mathcal{O}'_{r_n}| \right) + \exp(2\lambda^2 c^2 (1 + \alpha)^2 |\mathcal{O}^2_{r_n-1}|) \\ & = \exp \left(\lambda^2 c^2 (1 + \alpha)^2 \left(2|\mathcal{O}^2_{r_n-1}| + \frac{|\mathcal{O}'_{r_n}|}{2} \right) \right) \\ & \leq \exp(\lambda^2 c^2 (1 + \alpha)^2 n). \end{aligned}$$

This leads us to

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\lambda \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right) \right] \\ & \leq \exp(\lambda^2 c^2 (1 + \alpha)^2 n) \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}^2_{r_n-1}} 2Qf(X_i) + \lambda \sum_{i \in \mathcal{O}^1_{r_n-1}} Qf(X_i) \right) \right]. \end{aligned}$$

Now let:

- $\mathcal{O}^{1,1}_{r_n-2}$ the set of individuals of \mathbb{G}_{r_n-2} which are ancestors of one individual in \mathcal{O}_{r_n-1} and one individual in \mathcal{O}_{r_n} ;
- $\mathcal{O}^{1,2}_{r_n-2}$ the set of individuals of \mathbb{G}_{r_n-2} which are ancestors of one individual in \mathcal{O}_{r_n-1} and two individuals in \mathcal{O}_{r_n} ;
- $\mathcal{O}^{2,2}_{r_n-2}$ the set of individuals of \mathbb{G}_{r_n-2} which are ancestors of two individuals in \mathcal{O}_{r_n-1} and two individuals in \mathcal{O}_{r_n} ;
- $\mathcal{O}^{2,3}_{r_n-2}$ the set of individuals of \mathbb{G}_{r_n-2} which are ancestors of two individuals in \mathcal{O}_{r_n-1} and three individuals in \mathcal{O}_{r_n} ;

- $\mathcal{O}_{r_n-2}^{2,4}$ the set of individuals of \mathbb{G}_{r_n-2} which are ancestors of two individuals in \mathcal{O}_{r_n-1} and four individuals in \mathcal{O}_{r_n} ;
- \mathcal{O}'_{r_n-1} the set of individuals of \mathcal{O}_{r_n-1} whose parents belong to $\mathcal{O}_{r_n-2}^{1,1}$;
- \mathcal{O}''_{r_n-1} the set of individuals of \mathcal{O}_{r_n-1} whose parents belong to $\mathcal{O}_{r_n-2}^{1,2}$.

Then we have

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}_{r_n-1}^2} 2Qf(X_i) + \lambda \sum_{i \in \mathcal{O}_{r_n-1}^1} Qf(X_i) \right) \right] \\ &= \mathbb{E}[I_1 \times I_2 \times I_3 \times I_4 \times I_5 \times I_6 \times I_7], \end{aligned}$$

where

$$\begin{aligned} I_1 = \exp \left(\lambda \sum_{i \in \mathcal{O}_{r_n-2}^{1,1}} Q^2 f(X_i) + \lambda \sum_{i \in \mathcal{O}_{r_n-2}^{1,2}} 2Q^2 f(X_i) + \lambda \sum_{i \in \mathcal{O}_{r_n-2}^{2,2}} 2Q^2 f(X_i) \right. \\ \left. + \lambda \sum_{i \in \mathcal{O}_{r_n-2}^{2,3}} 3Q^2 f(X_i) + \lambda \sum_{i \in \mathcal{O}_{r_n-2}^{2,4}} 4Q^2 f(X_i) \right), \end{aligned}$$

$$I_2 = \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}'_{r_n-1}} Qf(X_i) - Q^2 f(X_{[i/2]}) \right) / \tilde{\mathcal{F}}_{r_n-2} \right],$$

$$I_3 = \mathbb{E} \left[\exp \left(2\lambda \sum_{i \in \mathcal{O}''_{r_n-1}} Qf(X_i) - Q^2 f(X_{[i/2]}) \right) / \tilde{\mathcal{F}}_{r_n-2} \right],$$

$$I_4 = \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}_{r_n-1}^{2,2}} Qf(X_{2i}) + Qf(X_{2i+1}) - 2Q^2 f(X_i) \right) / \tilde{\mathcal{F}}_{r_n-2} \right],$$

$$I_5 = \mathbb{E} \left[\exp \left(\frac{\lambda}{2} \sum_{i \in \mathcal{O}_{r_n-1}^{2,3}} 2Qf(X_{2i}) + Qf(X_{2i+1}) - 3Q^2 f(X_i) \right) / \tilde{\mathcal{F}}_{r_n-2} \right],$$

$$I_6 = \mathbb{E} \left[\exp \left(\frac{\lambda}{2} \sum_{i \in \mathcal{O}_{r_n-1}^{2,3}} Qf(X_{2i}) + 2Qf(X_{2i+1}) - 3Q^2 f(X_i) \right) / \tilde{\mathcal{F}}_{r_n-2} \right],$$

$$I_7 = \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathcal{O}_{r_n-1}^{2,4}} 2Qf(X_{2i}) + 2Qf(X_{2i+1}) - 4Q^2 f(X_i) \right) / \tilde{\mathcal{F}}_{r_n-2} \right].$$

Using the Azuma–Bennett–Hoeffding inequality, we get

$$\begin{aligned} & I_2 \times I_3 \times I_4 \times I_5 \times I_6 \times I_7 \\ & \leq \exp \left(\lambda^2 c^2 (\alpha + \alpha^2)^2 \left(\frac{|\mathcal{O}'_{r_n-1}|}{2} + 2|\mathcal{O}''_{r_n-1}| + 2|\mathcal{O}_{r_n-1}^{2,2}| \right) \right) \end{aligned}$$

$$+ \frac{9|\mathcal{O}_{r_n-1}^{2,3}|}{2} + 8|\mathcal{O}_{r_n-1}^{2,4}| \Big) \Big)$$

$$\leq \exp(2\lambda^2 c^2 (\alpha + \alpha^2)^2 n),$$

hence

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right) \right] \leq \exp(\lambda^2 c^2 (1 + \alpha)^2 n) \exp(2\lambda^2 c^2 (\alpha + \alpha^2)^2 n) \mathbb{E}[I_1].$$

Now, iterating this procedure we get

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) \right) \right] \leq \exp \left(\lambda^2 c^2 (1 + \alpha)^2 n \sum_{p=0}^{r_n} (2\alpha^2)^p \right) \exp(\lambda c \alpha^{r_n} n).$$

Then it follows as in part 1 that

$$(A.8) \quad \mathbb{P} \left(\frac{1}{n} \sum_{i=2^{r_n}}^n f(X_{\Pi(i)}) > \frac{\delta}{2} \right) \leq \begin{cases} \exp(c'' \delta) \exp(-c' \delta^2 n), & \text{if } \alpha \leq \frac{1}{2}, \\ \exp(-c' \delta^2 n), & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp \left(-c' \delta^2 \frac{n}{r_n} \right), & \text{if } \alpha^2 = \frac{1}{2}, \\ \exp \left(-c' \delta^2 \left(\frac{1}{\alpha} \right)^{2r_n} \right), & \text{if } \alpha^2 > \frac{1}{2}, \end{cases}$$

where $r_0 := \log(\frac{\delta}{c_0}) / \log(\alpha)$ and the positive constants c_0 , c' and c'' depend on α , δ , c and differ line to line. Finally (A.7) and (A.8) lead us to (3.3).

A.4. Proof of Theorem 3.2. Let $f \in \mathcal{B}_b(S^3)$ such that $(\mu, Pf) = 0$.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. We have for all $\delta > 0$ and $\lambda > 0$,

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp(-\lambda \delta |\mathbb{G}_r|) \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(\Delta_i) \right) \right].$$

Conditioning and using Bennett–Hoeffding inequality gives us

$$\mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} f(\Delta_i) \right) \right] \leq \exp(2\lambda^2 \|f\|_\infty |\mathbb{G}_r|) \mathbb{E} \left[\exp \left(\lambda \sum_{i \in \mathbb{G}_r} Pf(X_i) \right) \right].$$

Now, applying part 1 of the proof of the Theorem 3.1 to Pf , we get (3.1) for $f \in \mathcal{B}_b(S^3)$.

Part 2. Let us now treat $\overline{M}_{\mathbb{T}_r}(f)$. We have for all $\delta > 0$,

$$(A.9) \quad \mathbb{P}(\overline{M}_{\mathbb{T}_r}(f) > \delta) \leq \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f - Pf) > \frac{\delta}{2}\right) + \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(Pf) > \frac{\delta}{2}\right).$$

Now, since $(M_n^{\Pi}(f - Pf))_{n \geq 1}$ is a \mathcal{H}_n -martingale with bounded jumps, the Azuma inequality [1] gives us for some positive constant c' ,

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f - Pf) > \frac{\delta}{2}\right) \leq \exp(-c'\delta^2|\mathbb{T}_r|).$$

For the second term on the right-hand side of (A.9), we use inequalities (3.2) with Pf instead of f . Gathering these inequalities, we get (3.2) for all r large enough.

Part 3. The proof for the case $\overline{M}_n^{\Pi}(f)$ follows the same lines as the proof of part 2.

A.5. Proof of Proposition 4.2. We will prove the deviation inequality for $|\hat{\alpha}_0^r - \alpha_0|$. The other deviation inequalities for $|\hat{\beta}_0^r - \beta_0|, |\hat{\alpha}_1^r - \alpha_1|$ and $|\hat{\beta}_1^r - \beta_1|$ may be treated in a similar way.

One easily checks that

$$\hat{\alpha}_0^r - \alpha_0 = \frac{(\overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(P(\mathbf{xy}))) - (\overline{M}_{\mathbb{T}_r}(\mathbf{x}))(\overline{M}_{\mathbb{T}_r}(\mathbf{y}) - \overline{M}_{\mathbb{T}_r}(P(\mathbf{y})))}{B_r}.$$

We then have, for all $\delta > 0$,

$$\begin{aligned} & \mathbb{P}(|\hat{\alpha}_0^r - \alpha_0| > \delta) \\ & \leq \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))|}{B_r} > \frac{\delta}{2}\right) \\ & \quad + \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x})||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{2}\right). \end{aligned}$$

On one hand, for all $\gamma_1 > 0$ we have

$$(A.10) \quad \begin{aligned} & \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))|}{B_r} > \frac{\delta}{2}\right) \\ & \leq \mathbb{P}(B_r < \gamma_1) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))| > \frac{\delta\gamma_1}{2}\right). \end{aligned}$$

Now, for $b = \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2$, where μ_1 and μ_2 are given in (4.5), we have

$$\begin{aligned} \mathbb{P}(B_r < \gamma_1) &\leq \mathbb{P}\left(-\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2 - \mu_2) > \frac{b - \gamma_1}{3}\right) \\ &\quad + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\sqrt{b - \gamma_1}}{\sqrt{3}}\right) \\ &\quad + \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1) > \frac{b - \gamma_1}{6|\mu_1|}\right). \end{aligned}$$

We choose $\gamma_1 < \min\left\{\frac{2b}{2+3\delta}, \frac{-4+\sqrt{48b\delta^2+16}}{6\delta^2}, \frac{b}{1+3\delta|\mu_1|}\right\}$ so that $\frac{\delta\gamma_1}{2} < \max\left\{\frac{b-\gamma_1}{3}, \frac{\sqrt{b-\gamma_1}}{\sqrt{3}}, \frac{b-\gamma_1}{6|\mu_1|}\right\}$. Then we have

$$\mathbb{P}(B_r < \gamma_1) \leq \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mu_2 - \mathbf{x}^2) > \frac{\delta\gamma_1}{2}\right) + 2\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\delta\gamma_1}{2}\right),$$

and therefore we get

$$\begin{aligned} &\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))|}{B_r} > \frac{\delta}{2}\right) \\ &\leq 2\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\delta\gamma_1}{2}\right) + \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mu_2 - \mathbf{x}^2) > \frac{\delta\gamma_1}{2}\right) \\ &\quad + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))| > \frac{\delta\gamma_1}{2}\right). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x})||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{2}\right) &\leq \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{4}\right) \\ &\quad + \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{4|\mu_1|}\right). \end{aligned}$$

The last term of the previous inequality can be dealt with in the same way as inequality (A.10), using $\gamma_3 > 0$ such that

$$\gamma_3 < \min\left\{\frac{4b|\mu_1|}{4|\mu_1| + 3\delta}, \frac{2|\mu_1|(-4 + \sqrt{24b\delta^2/|\mu_1| + 16})}{3\delta^2}, \frac{2b}{2 + 3\delta}\right\}.$$

For the second term, we have

$$\begin{aligned} &\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)||\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{4}\right) \\ &\leq \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\sqrt{\delta}}{2}\right) + \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\sqrt{\delta}}{2}\right). \end{aligned}$$

Let $\gamma_2 > 0$ such that $\gamma_2 < \min\left\{\frac{2b}{2+3\sqrt{\delta}}, \frac{-4+\sqrt{48b\delta+16}}{b\delta}, \frac{b}{1+3\sqrt{\delta}|\mu_1|}\right\}$, in such a way that we obtain $\frac{\gamma_2\sqrt{\delta}}{2} < \max\left\{\frac{b-\gamma_2}{3}, \frac{\sqrt{b-\gamma_2}}{\sqrt{3}}, \frac{b-\gamma_2}{6|\mu_1|}\right\}$. We thus have

$$\begin{aligned} & \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| |\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))|}{B_r} > \frac{\delta}{4}\right) \\ & \leq \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\sqrt{\delta}}{2}\right) \\ & \quad + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2 - \mu_2)| > \frac{\gamma_2\sqrt{\delta}}{2}\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))| > \frac{\gamma_2\sqrt{\delta}}{2}\right) \\ & \quad + 2\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\gamma_2\sqrt{\delta}}{2}\right). \end{aligned}$$

From the foregoing, we deduce that for all $\gamma > 0$ such that $\gamma < \min(\gamma_1, \gamma_2, \gamma_3)$,

$$\begin{aligned} & \mathbb{P}(|\hat{\alpha}_0^{(r)} - \alpha_0| > \delta) \\ & \leq 2\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\delta\gamma}{2}\right) + \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mu_2 - \mathbf{x}^2) > \frac{\delta\gamma}{2}\right) \\ & \quad + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))| > \frac{\delta\gamma}{2}\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\sqrt{\delta}}{2}\right) \\ & \quad + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2 - \mu_2)| > \frac{\gamma\sqrt{\delta}}{2}\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))| > \frac{\gamma\sqrt{\delta}}{2}\right) \\ & \quad + 2\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\gamma\sqrt{\delta}}{2}\right) + 2\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)| > \frac{\delta\gamma}{4|\mu_1|}\right) \\ & \quad + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mu_2 - \mathbf{x}^2)\right| > \frac{\delta\gamma}{4|\mu_1|}\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))| > \frac{\delta\gamma}{4|\mu_1|}\right). \end{aligned}$$

Now, using (2.8) and Markov's inequality we get

$$\begin{aligned} & \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{xy} - P(\mathbf{xy}))| > \frac{\delta\gamma}{2}\right) \leq \frac{c}{\delta^4\gamma^4} \left(\frac{1}{4}\right)^{r+1}, \\ & \mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))| > \frac{\delta\gamma}{4|\mu_1|}\right) \leq \frac{c\mu_1^4}{\delta^4\gamma^4} \left(\frac{1}{4}\right)^{r+1} \end{aligned}$$

and

$$\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(\mathbf{y} - P(\mathbf{y}))| > \frac{\gamma\sqrt{\delta}}{2}\right) \leq \frac{c}{\delta^2\gamma^4} \left(\frac{1}{4}\right)^{r+1},$$

where the constant c can be found as in Remark 2.4.

Finally, the other terms, that is, the terms related to $\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2 - \mu_2)$ and $\overline{M}_{\mathbb{T}_r}(\mathbf{x} - \mu_1)$, can be bounded as in Corollary 2.2 and this completes the proof.

APPENDIX B

Let us gather here, for the convenience of the readers, various theorems useful to establish LIL, ASFCLT, deviation inequalities and MDP.

First, let us enunciate the Azuma–Bennett–Hoeffding inequality [1, 3, 16].

LEMMA B.1. *Let X be a real-valued and centered random variable such that $a \leq X \leq b$ a.s., with $a < b$. Then for all $\lambda > 0$, we have*

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

LEMMA B.2. *Let (E, d) a metric space. Let (Z_n) a sequence of random variables values in E , (v_n) a rate and $g: \mathcal{D}_E \subset E \rightarrow \mathbb{R}$ continuous. Let $z \in E$ be a deterministic value:*

$$\text{If } Z_n \xrightarrow[v_n]{\text{superexp}} z \quad \text{then } g(Z_n) \xrightarrow[v_n]{\text{superexp}} g(z).$$

PROOF. For all $\delta > 0$, there exists (see, e.g., [22], proof of Theorem 2.3) $\alpha_0(\delta) > 0$

$$(B.1) \quad \mathbb{P}(|g(Z_n) - g(z)| > \delta) \leq \mathbb{P}(d(Z_n, z) > \alpha_0(\delta)).$$

Indeed, since g is continuous, for all $\delta > 0$, there exists $\alpha_0(\delta) > 0$ such that

$$|g(x) - g(z)| \leq \delta \quad \text{whenever } d(x, z) \leq \alpha_0(\delta).$$

We then have

$$\{\omega: d(Z_n(\omega), z) \leq \alpha_0(\delta)\} \subset \{\omega: |g(Z_n(\omega)) - g(z)| \leq \delta\}$$

and therefore inequality (B.1). Now, the result of the lemma follows since $Z_n \xrightarrow[v_n]{\text{superexp}} z$. \square

Let $M = (M_n, \mathcal{H}_n, n \geq 0)$ be a centered square integrable martingale defined on a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and $(\langle M \rangle_n)$ its bracket. We recall some limit theorems for martingale used intensively in this paper.

We recall the following result due to W. F. Stout (Theorem 3 in [21]).

THEOREM B.3. *Let (M_n) such that $M_0 = 0$. If $\langle M \rangle_n \rightarrow \infty$ a.s. and*

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{2 \log \log \langle M \rangle_n}{K_n^2 \langle M \rangle_n} \mathbb{E}[(M_n - M_{n-1})^2 \mathbf{1}_{\{(M_n - M_{n-1})^2 > K_n^2 \langle M \rangle_n\}} / \mathcal{H}_{n-1}] \\ < \infty \quad \text{a.s.,} \end{aligned}$$

where K_n are \mathcal{H}_{n-1} measurable and $K_n \rightarrow 0$ a.s., then $\limsup \frac{M_n}{\sqrt{2\langle M \rangle_n \log \log \langle M \rangle_n}} = 1$ a.s.

We recall the following result due to Chaabane (Corollary 2.2 in [5]).

THEOREM B.4. *Let (V_n) be a (\mathcal{H}_n) -predictable increasing process such that:*

H-1 $V_n^{-2} \langle M \rangle_n \xrightarrow{n \rightarrow \infty} 1$, a.s.;

H-2 for all $\varepsilon > 0$, $\sum_{n \geq 1} V_n^{-2} \mathbb{E}[(M_n - M_{n-1})^2 \mathbf{1}_{|M_n - M_{n-1}| > \varepsilon V_n} / \mathcal{H}_{n-1}] < \infty$, a.s.;

H-3 for some $a > 1$, $\sum_{n \geq 1} V_n^{-2a} \mathbb{E}[(M_n - M_{n-1})^{2a} \mathbf{1}_{|M_n - M_{n-1}| \leq V_n} / \mathcal{H}_{n-1}] < \infty$, a.s.

Then M_n satisfies an ASFCLT; that is, for almost all ω , the weighted random measures

$$W_N(\omega, \bullet) = (\log V_N^2)^{-1} \sum_{n=1}^N \left(1 - \frac{V_n^2}{V_{n+1}^2}\right) \delta_{\{\psi_n(\omega) \in \bullet\}}$$

associated to the continuous processes $\Psi_n(\omega) = \{\Psi_n(\omega, t), 0 \leq t \leq 1\}$ defined by

$$\Psi_n(\omega, t) = V_n^{-1} \{M_k + (V_{k+1}^2 - V_k^2)^{-1} (tV_n^2 - V_k^2)(M_{k+1} - M_k)\},$$

when $V_k^2 \leq tV_n^2 < V_{k+1}^2$, $0 \leq k \leq n-1$, weakly converge to the Wiener measure on $\mathcal{C}([0, 1], \mathbb{R})$.

Let us enunciate the following which corresponds to the unidimensional case of Theorem 1 in [11].

PROPOSITION B.5. *Let (b_n) a sequence satisfying*

$$b_n \text{ is increasing, } \quad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \quad \frac{b_n}{n} \longrightarrow 0,$$

such that $c(n) := n/b_n$ is nondecreasing, and define the reciprocal function $c^{-1}(t)$ by

$$c^{-1}(t) := \inf\{n \in \mathbb{N} : c(n) \geq t\}.$$

Under the following conditions:

(C1) there exists $Q \in \mathbb{R}_+^*$ such that $\frac{\langle M \rangle_n}{n} \xrightarrow{\text{superexp}}_{b_n^2/n} Q$;

(C2) $\limsup_{n \rightarrow +\infty} \frac{n}{b_n^2} \log(n \operatorname{ess\,sup}_{1 \leq k \leq c^{-1}(b_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > b_n / \mathcal{H}_{k-1})) = -\infty$;

(C3) for all $a > 0$ $\frac{1}{n} \sum_{k=1}^n \mathbb{E}(|M_k - M_{k-1}|^2 \mathbf{1}_{\{|M_k - M_{k-1}| \geq an/b_n\}}) / \mathcal{H}_{k-1}) \xrightarrow[b_n^2/n]{\text{superexp}} 0;$

$(M_n/b_n)_{n \in \mathbb{N}}$ satisfies the MDP in \mathbb{R} with the speed b_n^2/n and the rate function $I(x) = \frac{x^2}{2Q}$.

Acknowledgments. Let us thank two anonymous referees for their very careful reading and useful suggestions, which have clearly improved both presentation and mathematical rigor of the present paper.

REFERENCES

- [1] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)* **19** 357–367. [MR0221571](#)
- [2] BASAWA, I. V. and ZHOU, J. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* **41A** 55–64. [MR2057565](#)
- [3] BENNETT, G. (1962). Probability inequalities for sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.
- [4] BERCU, B., DE SAPORTA, B. and GÉGOUT-PETIT, A. (2009). Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* **14** 2492–2526. [MR2563249](#)
- [5] CHAABANE, F. (1996). Version forte du théorème de la limite centrale fonctionnel pour les martingales. *C. R. Acad. Sci. Paris Sér. I Math.* **323** 195–198. [MR1402542](#)
- [6] COWAN, R. and STAUDTE, R. G. (1986). The bifurcating autoregressive model in cell lineage studies. *Biometrics* **42** 769–783.
- [7] DE SAPORTA, B., GÉGOUT-PETIT, A. and MARSALLE, L. (2011). Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* **5** 1313–1353. [MR2842907](#)
- [8] DELMAS, J.-F. and MARSALLE, L. (2010). Detection of cellular aging in a Galton–Watson process. *Stochastic Process. Appl.* **120** 2495–2519. [MR2728175](#)
- [9] DEMBO, A. (1996). Moderate deviations for martingales with bounded jumps. *Electron. Commun. Probab.* **1** 11–17 (electronic). [MR1386290](#)
- [10] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* **38**. Springer, New York. [MR1619036](#)
- [11] DJELLOUT, H. (2002). Moderate deviations for martingale differences and applications to ϕ -mixing sequences. *Stoch. Stoch. Rep.* **73** 37–63. [MR1914978](#)
- [12] DJELLOUT, H., GUILLIN, A. and WU, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** 2702–2732. [MR2078555](#)
- [13] GAO, F. and ZHAO, X. (2011). Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.* **39** 1211–1240. [MR2816352](#)
- [14] GUYON, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** 1538–1569. [MR2358633](#)
- [15] GUYON, J., BIZE, A., PAUL, G., STEWART, E., DELMAS, J.-F. and TADDÉI, F. (2005). Statistical study of cellular aging. In *CEMRACS 2004—Mathematics and Applications to Biology and Medicine. ESAIM Proceedings* **14** 100–114 (electronic). EDP Sci., Les Ulis. [MR2226805](#)

- [16] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](#)
- [17] Huggins, R. M. and Basawa, I. V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36** 1225–1233. [MR1746406](#)
- [18] Huggins, R. M. and Basawa, I. V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42** 423–432. [MR1802966](#)
- [19] Hwang, S. Y., Basawa, I. V. and Yeo, I. K. (2009). Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Stat. Methodol.* **6** 61–69. [MR2655539](#)
- [20] Stewart, E. J., Madden, R., Paul, G. and Taddéi, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* **3** e45.
- [21] Stout, W. F. (1970). A martingale analogue of Kolmogorov’s law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **15** 279–290. [MR0293701](#)
- [22] Van der Vaart, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- [23] Worms, J. (1999). Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* **4** 28 pp. (electronic). [MR1684149](#)
- [24] Worms, J. (2001). Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.* **10** 38–72. [MR1841808](#)
- [25] Zhou, J. and Basawa, I. V. (2005). Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74** 77–88. [MR2189078](#)
- [26] Zhou, J. and Basawa, I. V. (2005). Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Series Anal.* **26** 825–842. [MR2203513](#)

S. V. BITSEKI PENDA
 H. DJELLOUT
 LABORATOIRE DE MATHÉMATIQUES
 UNIVERSITÉ BLAISE PASCAL
 CNRS UMR 6620
 24 AVENUE DES LANDAIS
 BP 80026, 63177 AUBIÈRE
 FRANCE
 E-MAIL:

Valere.Bitsekipenda@math.univ-bpclermont.fr
Hacene.Djellout@math.univ-bpclermont.fr

A. GULLIN
 INSTITUT UNIVERSITAIRE DE FRANCE
 ET LABORATOIRE DE MATHÉMATIQUES
 UNIVERSITÉ BLAISE PASCAL
 CNRS UMR 6620
 24 AVENUE DES LANDAIS
 BP 80026, 63177 AUBIÈRE
 FRANCE
 E-MAIL:

Arnaud.Guillin@math.univ-bpclermont.fr