# DEVIATION INEQUALITIES FOR BIFURCATING MARKOV CHAINS ON GALTON–WATSON TREE

# S. Valère Bitseki Penda<sup>1</sup>

Abstract. We are interested in bifurcating Markov chains on Galton–Watson tree. These processes are an extension of bifurcating Markov chains, which was introduced by Guyon to detect cellular aging from cell lineage, in case the index set is a binary Galton–Watson process. First, under geometric ergodicity assumption of an embedded Markov chain, we provide polynomial deviation inequalities for properly normalized sums of bifurcating Markov chains on Galton–Watson tree. Next, under some uniformity, we derive exponential inequalities. These results allow to exhibit different regimes of convergence which correspond to a competition between the geometric ergodic speed of the underlying Markov chain and the exponential growth of the Galton–Watson tree. As application, we derive deviation inequalities (for either the Gaussian setting or the bounded setting) for the least-squares estimator of autoregressive parameters of bifurcating autoregressive processes with missing data which allow, in the case of cell division, to take into account the cell's death.

#### Mathematics Subject Classification. 60E15, 60J80, 60J10.

Received July 1, 2014. Revised December 1, 2014.

# 1. INTRODUCTION

Bifurcating Markov chains (BMC) on Galton–Watson (GW) tree are an extension of BMC to GW tree data. They were introduced by Delmas and Marsalle [15] in order to take into account the death of individuals in the *Escherichia coli's* (*E. coli*) reproduction model. *E. coli* is a rod-shaped bacterium which reproduces by dividing in the middle and producing two cells: One which has the old end of the mother (that is the one which is pre-existing) and that called old pole progeny cell; the other which has the new end of the mother (that is the one derived from a division) and that called new pole progeny cell. In fact, each daughter cell has two ends. One which already existed (old pole) and the other which is new (new pole). The age of a cell is given by the age of its old pole (*i.e.* the number of generations in the past of the cell before the old pole was produced).

Keywords and phrases. Bifurcating Markov chains, Galton–Watson processes, ergodicity, deviation inequalities, first order bifurcating autoregressive process with missing data, cellular aging.

<sup>&</sup>lt;sup>1</sup> CMAP, UMR 7641, École polytechnique CNRS, Route de Saclay, 91128 Palaiseau, France. simeon-valere.bitseki-penda@polytechnique.edu

#### S.V. BITSEKI PENDA

Guyon *et al.* [18] proposed the following linear Gaussian model to describe the evolution of the growth rate of the population of cells derived from an initial individual:

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

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),  $\nu$  is a distribution probability on  $\mathbb{R}$ ,  $\alpha_0, \alpha_1 \in (-1, 1)$ ;  $\beta_0, \beta_1 \in \mathbb{R}$  and  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \ge 1)$  forms a sequence of i.i.d bivariate random variables with law  $\mathcal{N}_2(0, \Gamma)$ , where

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1).$$
(1.2)

The processes  $(X_n)$  defined by (1.1) are typical example 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 (see Fig. 1). They were first introduced by Cowan and Staudte [11] for cell lineage data where each individual in one generation gives rise to two offsprings in the next generation.

In [17], Guyon, using BMC, gives laws of large numbers and central limit theorem for the least-squares estimator  $\hat{\theta}^n = (\hat{\alpha}_0^n, \hat{\beta}_0^n, \hat{\alpha}_1^n, \hat{\beta}_1^n)$  of the 4-dimensional parameter  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ . He also tests if the model is symmetric or not, and if the new pole and the old pole populations are even distinct in mean. This allows him to conclude a statistical evidence in aging in *E. coli*. Let us also mention [7], where Bercu *et al.* using a martingale approach give asymptotic analysis of the least squares estimator of the unknown parameters of a general asymmetric *p*th-order BAR process. The same idea was used recently by Bercu and Blandin [6] and Blandin [10] where they study the least squares estimator of the unknown parameters of a general asymmetric random coefficient BAR process. Recently, De Saporta *et al.* [14] studied the least squares estimator of the unknown parameters of a general asymmetric random coefficient BAR process with missing data. In their work, they combine a martingale approach and the BMC.

However, in the BMC model presented by Guyon, cells are assumed to never die (a death corresponds to no more division). To take into account cells's death, Delmas and Marsalle [15], instead of a regular binary tree, used a binary GW tree to label cells. In the sequel, we will introduce the model which allowed them to study the behavior of the growth rate of cells, taking into account their possible death.

#### 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

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

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  $|\mathbb{T}_r| = 2^{r+1} - 1$ . 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 rth generation and the first (r+1) generations. The initial individual is denoted 1. The model proposed by Delmas and Marsalle [15] is defined as follows (the growth rate of cell n is  $X_n$ ):

• With probability  $p_{1,0}$ , n gives birth to two cells 2n and 2n + 1 which both divide. The growth rate of the daughters  $X_{2n}$  and  $X_{2n+1}$  are then linked to the mother's one through auto-regressive equations (1.1).



FIGURE 1. The binary tree  $\mathbb{T}$ .

• With probability  $p_0$ , only the new pole 2n divides. Its growth rate  $X_{2n}$  is linked to its mother's one  $X_n$  through the relation

$$X_{2n} = \alpha'_0 X_n + \beta'_0 + \varepsilon'_{2n}, \tag{1.3}$$

where  $\alpha'_0 \in (-1,1), \beta'_0 \in \mathbb{R}$  and  $(\varepsilon'_{2n}, n \in \mathbb{T})$  is a sequence of independent centered Gaussian random variables with variance  $\sigma_0^2 > 0$ .

• With probability  $p_1$ , only the old pole 2n + 1 divides. Its growth rate  $X_{2n+1}$  is linked to its mother's one  $X_n$  through the relation

$$X_{2n+1} = \alpha'_1 X_n + \beta'_1 + \varepsilon'_{2n+1}, \tag{1.4}$$

where  $\alpha'_1 \in (-1,1)$ ,  $\beta'_1 \in \mathbb{R}$  and  $(\varepsilon'_{2n+1}, n \in \mathbb{T})$  is a sequence of independent centered Gaussian random variables with variance  $\sigma_1^2 > 0$ .

- With probability  $1 p_{1,0} p_1 p_0$ , which is supposed non-negative, n gives birth to two cells which do not divide.
- The sequences  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T}), (\varepsilon'_{2n}, n \in \mathbb{T}), (\varepsilon'_{2n+1}, n \in \mathbb{T})$  and  $X_1$  are independent.

The process  $(X_n)$  described above is a typical example of BMC on GW tree. In [12], this process is called bifurcating autoregressive process (BAR) with missing data. It is an extension of bifurcating autoregressive process with the data having a binary GW tree structure, see Figure 2 for example of binary GW tree. Indeed, one can assume that the cells which do not divide and those which do not exist are missing or dead.



FIGURE 2. A binary GW tree up to the 4 th generation. In this tree, individual 1 gives birth to two individuals which both divide, this happens with probability  $p_{1,0}$ . Individual 2 gives birth to two individuals of which only one (the new pole) divides, this happens which probability  $p_0$ . Individual 12 gives birth to two individuals which do not divide, this happens with probability  $1 - p_{1,0} - p_0 - p_1$ .

In [15], Delmas and Marsalle using their results for BMC on GW tree, gave laws of large numbers and central limit theorem for the maximum likelihood estimator of the parameter

$$\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1, \alpha'_0, \beta'_0, \alpha'_1, \beta'_1).$$
(1.5)

In this paper, we will give deviation inequalities for the least squares estimator of the parameter  $\theta$ . This will be done in two cases: the Gaussian case described above and the special case where  $\max\{|\alpha_0|, |\alpha_1|, |\alpha'_0|, |\alpha'_1|\} < 1$ and the noise sequence and the initial state  $X_1$  take their values in a compact set. Note that the latter implies that the BAR process with missing data also takes its values in compact set. These deviation inequalities are important for a rigorous non asymptotic statistical study. Indeed, when the sample size is insufficient to apply limit theorems, they allow for example to estimate the errors in the estimation of unknown parameters. Furthermore, these inequalities allow to get a rate of convergence in the laws of large numbers, and this permits, for example, to build non-asymptotic confidence intervals.

We are now going to give a rigorous definition of BMC on GW tree (see [15] for more details).

#### 1.2. Definitions

For an individual  $n \in \mathbb{T}$ , we are interested in the quantity  $X_n$  which may be the weight, the growth rate and which takes values in the metric space S endowed with its Borel  $\sigma$ -field S.

**Definition 1.1** (T-transition probability, see ([17])). We call T-transition probability any mapping  $P: S \times S^2 \rightarrow [0, 1]$  such that

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

For  $p \ge 1$ , we denote by  $\mathcal{B}(S^p)(\text{resp. } \mathcal{B}_b(S^p), \mathcal{C}(S^p), \mathcal{C}_b(S^p))$  the set of all  $\mathcal{S}^p$ -measurable (resp.  $\mathcal{S}^p$ -measurable and bounded, continuous, continuous and bounded) mappings  $f : S^p \to \mathbb{R}$ . For  $f \in \mathcal{B}(S^3)$ , when it is defined, we denote by  $Pf \in \mathcal{B}(S)$  the function

$$x \mapsto Pf(x) = \int_{S^2} f(x, y, z) P(x, \mathrm{d}y, \mathrm{d}z).$$

**Definition 1.2** (bifurcating Mmrkov chains, see ([17])). 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  $\nu$  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  $\nu$  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 families  $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$

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

Now, we add a cemetery point to S,  $\partial$ . Let  $\overline{S} = S \cup \{\partial\}$ , and  $\overline{S}$  be the  $\sigma$ -field generated by S and  $\{\partial\}$ . In the previous biological framework, S corresponds to the state space of the quantities related to living cells, and  $\partial$  is the default value for dead cells. Let  $P^*$  be a  $\mathbb{T}$ -transition probability defined on  $\overline{S} \times \overline{S}^2$  such that

$$P^*(\partial, \{(\partial, \partial)\}) = 1. \tag{1.6}$$

In the previous biological framework, equation (1.6) means that no dead cell can give birth to a living cell. We denote by  $P_0^*$  and  $P_1^*$  the restriction of the first and the second marginal of  $P^*$  to S, that is:

$$P_0^* = P^*\left(\cdot, \left(\cdot \bigcap S\right) \times \bar{S}\right) \text{ and } P_1^* = P^*\left(\cdot, \bar{S} \times \left(\cdot \bigcap S\right)\right).$$

**Definition 1.3** (BMC on GW tree, see [15]). Let  $(X_n, n \in \mathbb{T})$  be a  $P^*$ -BMC on  $(\overline{S}, \overline{S})$ , with  $P^*$  satisfying (1.6). We call  $(X_n, n \in \mathbb{T}^*)$ , with  $\mathbb{T}^* = \{n \in \mathbb{T} : X_n \neq \partial\}$ , a BMC on GW tree. The  $P^*$ -BMC is said spatially homogeneous if  $p_{1,0} = P^*(x, S \times S), p_0 = P^*(x, S \times \{\partial\})$ , and  $p_1 = P^*(x, \{\partial\} \times S)$  do not depend on  $x \in S$ . A spatially homogeneous  $P^*$ -BMC is said super-critical if m > 1, where  $m = 2p_{1,0} + p_1 + p_0$ .

We denote by  $(Y_n, n \in \mathbb{N})$  the Markov chain on S with  $Y_0 = X_1$  and transition probability  $Q = \frac{1}{m}(P_0^* + P_1^*)$ . Notice that any function f defined on S is extended to  $\overline{S}$  by setting  $f(\partial) = 0$ . In all 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  $Q^{p+1}(x, B) = \int_S Q(x, \mathrm{d}y)Q^p(y, B)$  for all  $B \in \mathcal{S}$ ;  $Q^p$  is a transition probability in  $(S, \mathcal{S})$ .
- $\nu Q$  the distribution on  $(S, \mathcal{S})$  defined by  $\nu Q(B) = \int_S \nu(\mathrm{d}x)Q(x, B); \ \nu Q^p$  is the law of  $Y_p$ ,
- $(Qf)(x) = \int_{S} f(y)Q(x, dy)$  when it is defined;
- $(\nu f)$  or  $(\nu, f)$  the integral  $\int_S f d\nu$  when it is defined.

#### S.V. BITSEKI PENDA

#### Remark 1.4.

- The name BMC on GW tree comes from the fact that condition (1.6) and spatial homogeneity imply that  $\mathbb{T}^*$  is a GW tree.
- All through this work, we shall assume that the  $P^*$ -BMC is super-critical.

Now, for any subset  $J \subset \mathbb{T}$ , let

$$J^* = J \cap \mathbb{T}^* = \{j \in J : X_j \neq \partial\}$$

be the subset of living cells among J, and |J| be the cardinal of J. The process  $(|\mathbb{G}_k^*|, k \in \mathbb{N})$ , is a GW process with the reproduction generating function

$$\psi(z) = (1 - p_0 - p_1 - p_{1,0}) + (p_0 + p_1)z + p_{1,0}z^2,$$

and the average number of daughters alive is m. It is known, see *e.g.* [2], that  $m^{-r}|\mathbb{G}_r^*|$  converges in probability to a non-negative random variable W. Moreover,  $\mathbb{P}(W > 0) = 1$  if there is no extinction almost surely which is equivalent, in our setting, to  $p_0 + p_1 + p_{1,0} = 1$ . We have for all  $r \ge 0$ ,

$$\mathbb{E}[|\mathbb{G}_{r}^{*}|] = m^{r} \quad \text{and} \quad \mathbb{E}[|\mathbb{T}_{r}^{*}|] = \sum_{q=0}^{r} \mathbb{E}[|\mathbb{G}_{q}^{*}|] = \frac{m^{r+1} - 1}{m-1} := t_{r}.$$
(1.7)

It was also established in [15] that  $t_r^{-1}|\mathbb{T}_r^*|$  converges in probability to W as well.

For  $i \in \mathbb{T}$ , set  $\Delta_i = (X_i, X_{2i}, X_{2i+1})$  the mother-daughters quantities of interest. For a finite subset  $J \subset \mathbb{T}$ , we set

$$M_J(f) = \begin{cases} \sum_{i \in J} f(X_i) & \text{for } f \in \mathcal{B}(\bar{S}), \\ \sum_{i \in J} f(\Delta_i) & \text{for } f \in \mathcal{B}(\bar{S}^3), \end{cases}$$
(1.8)

with the convention that a sum over an empty set is null. We also define the following two averages of f over J

$$\overline{M}_J(f) = \frac{1}{|J|} M_J(f) \quad \text{if } |J| > 0 \quad \text{and} \quad \widetilde{M}_J(f) = \frac{1}{\mathbb{E}[|J|]} M_J(f) \quad \text{if } \mathbb{E}[|J|] > 0.$$

$$(1.9)$$

Limit theorems for averages (1.9) have been studied in [15] for  $J = \mathbb{G}_n^*$  and  $J = \mathbb{T}_n^*$ , as n goes to infinity. Under a geometric ergodicity and a uniform geometric ergodicity type assumptions for Q, we will establish in this paper polynomial and exponential deviation inequalities for these averages. These deviation inequalities will allow to highlight three regimes for the speed of convergence of above averages, thus showing a competition between the ergodicity of the embedded Markov chain  $(Y_n, n \in \mathbb{N})$  and the size of the binary Galton–Watson tree. This new phenomenon is not observed in the asymptotic study of Delmas and Marsalle [15]. Notice that deviation inequalities were already studied in the no death case [9], that is m = 2. We will use similar approach and we will introduce some modifications in order to take into account the randomness of the index set using large deviations for branching processes [1]. This work can be seen as a true extension of the results of Bitseki *et al.* [9]. Indeed, as we will see in the sequel, when m = 2 that is in the no death case, we recover exactly the results of [9]. Our results also complete those of Delmas and Marsalle [15] in the sense that we get convergence rate for limit theorems. Let us also mention [8], where the authors establish deviation inequalities for estimators of parameters of the *p*-order bifurcating autoregressive process. Let us note that in [4], Bansaye *et al.* study a partial extension of the results of Delmas and Marsalle [15] to a continuous time setting.

The rest of the paper is organized as follows. Section 2 deals with moments control and polynomial deviation inequalities for averages (1.9), for  $J = \mathbb{G}_n^*$  and  $J = \mathbb{T}_n^*$ . This will be done under a geometric ergodicity assumption of Q. The results of this section will be obtained in the general framework of Delmas and Marsalle [15]. In Section 3, we state exponential deviation inequalities for averages (1.9), for  $J = \mathbb{G}_n^*$  and  $J = \mathbb{T}_n^*$ . This will be done under a uniform geometric ergodicity type assumption for Q, and a suitable assumption on the binary GW tree similar to that used in [20]. The results of Section 3 will be obtained for bounded functions. In Section 4, we will focus in particular on the first order bifurcating autoregressive process with missing data described in Section 1.1. Appendices A, B and C are dedicated to the proofs of our results.

# 2. Moments control and polynomial deviations inequalities

Let F be a vector subspace of  $\mathcal{B}(S)$  such that:

- (i) F contains the constants,
- (ii)  $F^2 := \{f^2; f \in F\} \subset F,$
- (iii)  $-F \otimes F \subset L^1(P(x, \cdot))$  for all  $x \in S$  and  $P(f_0 \otimes f_1) \in F$  for all  $f_0, f_1 \in F$ ,
- For  $\delta \in \{0,1\}$ ,  $F \subset L^1(P^*_{\delta}(x,\cdot))$  for all  $x \in S$  and  $P^*_{\delta}(f) \in F$  for all  $f \in F$ ,
- (iv) there exists a probability measure  $\mu$  on  $(S, \mathcal{S})$  such that  $F \subset L^1(\mu)$  and  $\lim_{n \to \infty} \mathbb{E}_x[f(Y_n)] = (\mu, f)$  for all  $x \in S$  and  $f \in F$ ,
- (v) for all  $f \in F$ , there exists  $g \in F$  such that for all  $n \in \mathbb{N}$ ,  $|Q^n f| \leq g$ ,
- (vi)  $F \subset L^1(\nu)$ ,

where we have used the notation  $F \otimes F = \{f_0 \otimes f_1/f_0, f_1 \in F\}$ . By convention a function defined on  $\overline{S}$  is said to belong to F if its restriction to S belongs to F.

The following hypothesis is about the geometric ergodicity of Q:

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

Recall that under this hypothesis, Delmas and Marsalle [15] have shown the weak law of large numbers for the empirical averages  $\widetilde{M}_{\mathbb{G}^*_{*}}(f)$ ,  $\overline{M}_{\mathbb{G}^*_{*}}(f)$ ,  $\widetilde{M}_{\mathbb{T}^*_{*}}(f)$  and  $\overline{M}_{\mathbb{T}^*_{*}}(f)$ .

In this section, we are going to control the fourth order moment of  $\widetilde{M}_{\mathbb{G}_r^*}(f)$  for all  $r \in \mathbb{N}$  and for all  $f \in F$  which verifies hypothesis (**H1**). This will help us afterwards to get (polynomial) deviation inequalities for  $\widetilde{M}_{\mathbb{G}_r^*}(f)$ ,  $\overline{M}_{\mathbb{G}_r^*}(f)$ ,  $\widetilde{M}_{\mathbb{T}_r^*}(f)$  and  $\overline{M}_{\mathbb{T}_r^*}(f)$ .

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

$$\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{G}_{r}^{*}}(f)\right)^{4}\right] \leq \begin{cases} c\left(\frac{1}{m^{2}}\right)^{r} & \text{if } m\alpha^{2} < 1\\ cr^{2}\left(\frac{1}{m^{2}}\right)^{r} & \text{if } m\alpha^{2} = 1\\ c\alpha^{4r} & \text{if } m\alpha^{2} > 1, \end{cases}$$

$$(2.1)$$

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

*Proof.* The proof of this Proposition relies on an exact computation of  $\mathbb{E}[(\widetilde{M}_{\mathbb{G}_r^*}(f))^4]$  and a computation of the rates using hypothesis (H1). These computations are very tedious. Hence, for the readability of the paper, we postpone the proof in the Appendix C.

This leads us to the following moment control for the average  $M_{\mathbb{T}^*}(f)$ .

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

$$\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{T}_{r}^{*}}(f)\right)^{4}\right] \leq \begin{cases} c\left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} < 1\\ cr^{2}\left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} = 1\\ c\alpha^{4(r+1)} & \text{if } m\alpha^{2} > 1, \end{cases}$$

$$(2.2)$$

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

*Proof.* From equality

$$\widetilde{M}_{\mathbb{T}_r^*}(f) = \sum_{q=0}^r \frac{m^q}{t_r} \widetilde{M}_{\mathbb{G}_q^*}(f),$$

we deduce that

$$\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{T}_r^*}(f)\right)^4\right] \le \left(\sum_{q=0}^r \frac{m^q}{t_r} \|\widetilde{M}_{\mathbb{G}_q^*}(f)\|_4\right)^4$$

where  $\|\cdot\|_4$  stands for the  $L^4$ -norm. Now the proof of (2.2) follows easily from (2.1).

**Remark 2.3.** Let us stress that we can improve the previous inequalities. In fact, let F satisfy (i)-(vi). Let  $f \in F$  such that  $(\mu, f) = 0$ . We assume hypothesis (**H1**). Then it is possible to prove after enormous computations related to the calculation of higher moments that for all  $k \in \mathbb{N}^*$ , there is  $r_0 \in \mathbb{N}$  which depends on k (in particular, one can take  $r_0 = k$ ) such that for all  $r \geq r_0$ ,

$$\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{G}_r^*}(f)\right)^{2k}\right] \leq \begin{cases} c\left(\frac{1}{m^k}\right)^r & \text{if } m\alpha^2 < 1\\ cr^k\left(\frac{1}{m^k}\right)^r & \text{if } m\alpha^2 = 1\\ c\alpha^{2kr} & \text{if } m\alpha^2 > 1, \end{cases}$$

where the positive constant c depends on  $\alpha$ , k and f.

Nevertheless, we stress that the constant c which appears in the previous inequalities does not depend on k uniformly. As a matter of fact, we cannot use these results to get exponential inequalities.

**Remark 2.4.** If  $f \in \mathcal{B}(\overline{S}^3)$  is such that  $P^*f^2$  and  $P^*f^4$  exist and belong to F, with  $P^*f = 0$ , then we have for all  $r \in \mathbb{N}$ , for some positive constant c and recalling notations (1.8) and (1.9):

$$\mathbb{E}\Big[\left(\widetilde{M}_{\mathbb{G}_r^*}(f)\right)^4\Big] \le c\left(\frac{1}{m^2}\right)^r \quad \text{and} \quad \mathbb{E}\Big[\left(\widetilde{M}_{\mathbb{T}_r^*}(f)\right)^4\Big] \le c\left(\frac{1}{m^2}\right)^{r+1}.$$
(2.3)

Indeed, conditioning by  $\mathcal{F}_r$  and using the fact that  $P^*f = 0$ , we have

$$\mathbb{E}\Big[\left(M_{\mathbb{G}_r^*}(f)\right)^4\Big] = \mathbb{E}\left[\sum_{i\in\mathbb{G}_r^*} P^*f^4\right] + 6\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^{*2}; i\neq j} P^*f^2(X_i)P^*f^2(X_j)\right].$$

Now, dividing by  $m^{4r}$  and using the convergence of  $\left(\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{G}_r^*}\left(P^*f^2\right)\right)^2\right]\right)_r$  and  $\left(\mathbb{E}\left[\widetilde{M}_{\mathbb{G}_r^*}\left(P^*f^4\right)\right]\right)_r$  (see [15]), we get the first inequality of (2.3).

The second inequality of (2.3) follows easily from equality

$$\widetilde{M}_{\mathbb{T}_r^*}(f) = \sum_{q=0}^r \frac{m^q}{t_r} \widetilde{M}_{\mathbb{G}_q^*}(f),$$

as in the proof of corollary (2.2).

In fact, as in remark 2.3, inequalities (2.3) can be improved. Indeed, we can prove that

$$\mathbb{E}\Big[\left(\widetilde{M}_{\mathbb{G}_r^*}(f)\right)^{2k}\Big] \le c\left(\frac{1}{m^k}\right)^r \quad \text{and} \quad \mathbb{E}\Big[\left(\widetilde{M}_{\mathbb{T}_r^*}(f)\right)^{2k}\Big] \le c\left(\frac{1}{m^k}\right)^{r+1} \tag{2.4}$$

for all  $k \in \mathbb{N}^*$  and for some positive constant c.

696

We are now in position to state polynomial deviation inequalities for the averages  $\widetilde{M}_{\mathbb{G}_r^*}(f)$ ,  $\overline{M}_{\mathbb{G}_r^*}(f)$ ,  $\widetilde{M}_{\mathbb{T}_r^*}(f)$ and  $\overline{M}_{\mathbb{T}_r^*}(f)$ .

**Theorem 2.5.** Let F satisfy (i)-(vi). Let  $f \in F$  such that  $(\mu, f) = 0$ . We assume hypothesis (H1). Then for all  $r \in \mathbb{N}$  and for all  $\delta > 0$ , we have

$$\mathbb{P}\left(\left|\widetilde{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta\right) \leq \begin{cases} \frac{c}{\delta^{4}} \left(\frac{1}{m^{2}}\right)^{r} & \text{if } m\alpha^{2} < 1\\ \frac{c}{\delta^{4}} r^{2} \left(\frac{1}{m^{2}}\right)^{r} & \text{if } m\alpha^{2} = 1\\ \frac{c}{\delta^{4}} \alpha^{4r} & \text{if } m\alpha^{2} > 1 \end{cases}$$

$$(2.5)$$

and

$$\mathbb{P}\Big(\left|\widetilde{M}_{\mathbb{T}_{r}^{*}}(f)\right| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^{4}} \left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} < 1\\ \frac{c}{\delta^{4}} r^{2} \left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} = 1\\ \frac{c}{\delta^{4}} \alpha^{4(r+1)} & \text{if } m\alpha^{2} > 1, \end{cases}$$

$$(2.6)$$

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

Proof. See Appendix A.

For the other deviation inequalities, we will need an additional hypothesis. This hypothesis is related to the study of large deviations for GW processes (see for e.g. [1, 20]).

(H2)  $p_{1,0} + p_0 + p_1 = 1$ , where  $p_{1,0}$ ,  $p_0$  and  $p_1$  are defined in Section 1.1.

(H2) corresponds to the almost sure non-extinction property.

First, we establish the following result which is a direct consequence of Theorem 2.5.

**Corollary 2.6.** Let F satisfy (i)-(vi). Let  $f \in F$  such that  $(\mu, f) \neq 0$ . We assume hypotheses (H1)-(H2). Then for all  $r \in \mathbb{N}$  and for all  $\delta > 0$ , we have

$$\mathbb{P}\Big(\left|\widetilde{M}_{\mathbb{G}_r^*}(f) - (\mu, f)W\right| > \delta\Big) \le \begin{cases} \frac{c}{\delta^4} \left(\frac{1}{m^2}\right)^r & \text{if } m\alpha^2 < 1\\ \frac{c}{\delta^4} r^2 \left(\frac{1}{m^2}\right)^r & \text{if } m\alpha^2 = 1\\ \frac{c}{\delta^4} \alpha^{4r} & \text{if } m\alpha^2 > 1 \end{cases}$$
(2.7)

and

$$\mathbb{P}\Big(\left|\widetilde{M}_{\mathbb{T}_{r}^{*}}(f) - (\mu, f)W\right| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^{4}} \left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} < 1\\ \frac{c}{\delta^{4}} r^{2} \left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} = 1\\ \frac{c}{\delta^{4}} \alpha^{4(r+1)} & \text{if } m\alpha^{2} > 1, \end{cases}$$

$$(2.8)$$

where the positive constant c depends on  $\alpha$ , m, f and (and may differ line by line).

Proof. See Appendix A.

**Remark 2.7.** The results of Theorem 2.5 and Corollary 2.6 are similar. But we stress that in the Corollary 2.6, we need additional hypothesis (**H2**). Indeed, as we will see in the Appendix A, when  $(\mu, f) \neq 0$ , there is an additional term in the deviation inequalities. This term is a contribution of the binary Galton–Watson tree to the deviation of the averages  $\widetilde{M}_{\mathbb{G}_r^*}(f)$  and  $\widetilde{M}_{\mathbb{T}_r^*}(f)$  with respect to  $(\mu, f)W$ . This explains why we need additional hypothesis (**H2**) in Corollary 2.6, because we have to deal with the deviation inequalities for Galton–Watson processes. In the case of polynomial deviation inequalities, this additional term can be neglected. But, this will be not the case for the exponential deviation inequalities.

The next results can be seen as a consequence of the previous results.

**Theorem 2.8.** Let F satisfy (i)-(vi). Let  $f \in F$ . We assume hypotheses (H1)-(H2). Then for all  $r \in \mathbb{N}$ , for all  $\delta > 0$  and for all a > 0, we have

$$\mathbb{P}\Big(\left|\overline{M}_{\mathbb{G}_r^*}(f) - (\mu, f)\right| > \delta \left|W \ge a\right) \le \begin{cases} \frac{c(1+\delta)^4}{\delta^4} \left(\frac{1}{m^2}\right)^r & \text{if } m\alpha^2 < 1\\ \frac{c(1+\delta)^4}{\delta^4} r^2 \left(\frac{1}{m^2}\right)^r & \text{if } m\alpha^2 = 1\\ \frac{c(1+\delta)^4}{\delta^4} \alpha^{4r} & \text{if } m\alpha^2 > 1 \end{cases}$$
(2.9)

and

$$\mathbb{P}\Big(\left|\overline{M}_{\mathbb{T}_{r}^{*}}(f) - (\mu, f)\right| > \delta \Big| W \ge a\Big) \le \begin{cases} \frac{c(1+\delta)^{4}}{\delta^{4}} \left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} < 1\\ \frac{c(1+\delta)^{4}}{\delta^{4}} r^{2} \left(\frac{1}{m^{2}}\right)^{r+1} & \text{if } m\alpha^{2} = 1\\ \frac{c(1+\delta)^{4}}{\delta^{4}} \alpha^{4(r+1)} & \text{if } m\alpha^{2} > 1, \end{cases}$$
(2.10)

where the positive constant c depends on  $\alpha$ , m, f, a (and may differ line by line).

Proof. See Appendix A.

**Remark 2.9.** As already mention in Remark 2.3, the previous deviation inequalities can be improved. For example, in the case of Theorem 2.5, one can prove that for all  $k \in \mathbb{N}^*$ ,

$$\mathbb{P}\Big(\left|\widetilde{M}_{\mathbb{G}_r^*}(f)\right| > \delta\Big) \le \begin{cases} \frac{c}{\delta^{2k}} \left(\frac{1}{m^k}\right)^r & \text{if } m\alpha^2 < 1\\ \frac{c}{\delta^{2k}} r^k \left(\frac{1}{m^k}\right)^r & \text{if } m\alpha^2 = 1\\ \frac{c}{\delta^{2k}} \alpha^{2kr} & \text{if } m\alpha^2 > 1 \end{cases}$$

and

$$\mathbb{P}\Big(\left|\widetilde{M}_{\mathbb{T}_r^*}(f)\right| > \delta\Big) \le \begin{cases} \frac{c}{\delta^{2k}} \left(\frac{1}{m^k}\right)^{r+1} & \text{if } m\alpha^2 < 1\\ \frac{c}{\delta^{2k}} r^2 \left(\frac{1}{m^k}\right)^{r+1} & \text{if } m\alpha^2 = 1\\ \frac{c}{\delta^{2k}} \alpha^{2k(r+1)} & \text{if } m\alpha^2 > 1. \end{cases}$$

The same thing holds for Theorems 2.6 and 2.8 at least from a certain order.

**Remark 2.10.** Furthermore, using Remark 2.4, if  $f \in \mathcal{B}(\overline{S}^3)$  is such that  $P^*f^2$  and  $P^*f^4$  exist and belong to F, with  $P^*f = 0$ , then we have for all  $r \in \mathbb{N}$ , for all  $k \in \mathbb{N}^*$ , for all  $\delta > 0$ , for all a > 0, for some positive constant c and recalling notations (1.8) and (1.9):

$$\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r^*}(f)\right| > \delta \middle| W \ge a\right) \le \frac{c(1+\delta)^{2k}}{\delta^{2k}} \left(\frac{1}{m^k}\right)^{r+1}$$

**Remark 2.11.** Let us note that when m = 2, that is the no death case, we recover exactly the results of Bitseki *et al.* [9]. Consequently, for polynomial deviation inequalities, this allows us to say that our results are relevant in the sense that, they enable to extend known results [9].

# 3. EXPONENTIAL DEVIATION INEQUALITIES

We consider the following hypotheses:

(H3) For some  $0 < \alpha < 1$ , there exists a probability measure  $\mu$  on (S, S) such that for all  $f \in \mathcal{B}_b(S)$  with  $\langle \mu, f \rangle = 0$ , there is c > 0 (which depends of f) such that for all  $k \in \mathbb{N}$  and for all  $x \in S$ ,  $|Q^k f(x)| \le c\alpha^k$ . (H4)  $m > \sqrt{2}$ .

**Remark 3.1.** (H3) is a uniform geometric ergodicity type assumption. It implies in particular that the Markov chain Y is geometrically ergodic, that is for all  $f \in C_b(S)$  and for all  $x \in S$ ,  $\lim_{k \to \infty} \mathbb{E}_x[f(Y_k)] = \langle \mu, f \rangle$  with a geometric rate.

In the sequel,  $\mathbb{H}_r$  will denote one of the set  $\mathbb{G}_r$  or  $\mathbb{T}_r$ . We set  $h_r = (m^2/2)^r$ . We can now state the main results of this section.

**Theorem 3.2.** Under the hypotheses **(H3)** and **(H4)**, let  $f \in \mathcal{B}_b(S)$  such that  $\langle \mu, f \rangle = 0$ . Then we have for all  $\delta > 0$ :

• if  $m\alpha < 1$ , then  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\right) \le \exp\left(c''\delta\right) \exp\left(-c'\delta^2 h_r\right) \le$$

• if  $m\alpha = 1$ , then for  $\mathbb{H}_r = \mathbb{G}_r$  and  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \le \exp\left(c''\delta\right) \exp\left(-c'\delta^2 h_r\right);$$

• if  $m\alpha = 1$ , then for  $\mathbb{H}_r = \mathbb{T}_r$  and  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_{r}^{*}}(f) > \delta\right) \leq \exp\left(c''\delta(r+1)\right)\exp\left(-c'\delta^{2}h_{r}\right);$$

• if  $1 < m\alpha < \sqrt{2}$ , then  $\forall r \in \mathbb{N}$  such that  $r > r_0$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\right) \le \exp\left(-c'\delta^2 h_r\right);$$

• if  $m\alpha = \sqrt{2}$ , then  $\forall r \in \mathbb{N}$  such that  $r > r_0$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\right) \le \exp\left(-\frac{c'\delta^2 h_r}{r}\right);$$

• if  $m\alpha > \sqrt{2}$ , then  $\forall r \in \mathbb{N}^*$  such that  $r > r_0$ ,

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \le \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right);$$

where.

- r<sub>0</sub> := log (δ/c<sub>0</sub>) / log(α) − k<sub>0</sub>, with k<sub>0</sub> ∈ {0, 1},
  c<sub>0</sub>, c' and c'' are positive constants which depend on α, m c, and f and may differ line by line.

**Theorem 3.3.** Under the hypotheses (H2)–(H4), we have for all  $f \in \mathcal{B}_b(S)$  such that  $(\mu, f) \neq 0$  and for all  $\delta > 0$ :

• if  $m\alpha < 1$ , then  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, f)W > \delta\Big) \le \exp\left(c''\delta\right)\exp\left(-c'\delta^2h_r\right) + A_r;$$

• if  $m\alpha = 1$ , then for  $\mathbb{H}_r = \mathbb{G}_r$  and  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, f)W > \delta\right) \le \exp\left(c''\delta\right)\exp\left(-c'\delta^2 h_r\right) + A_r;$$

• if  $m\alpha = 1$ , then for  $\mathbb{H}_r = \mathbb{T}_r$  and  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, f)W > \delta\right) \le \exp\left(c''\delta(r+1)\right)\exp\left(-c'\delta^2h_r\right) + A_r$$

• if  $1 < m\alpha < \sqrt{2}$ , then  $\forall r \in \mathbb{N}$  such that  $r > r_0$ ,

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, f)W > \delta\Big) \le \exp\left(-c'\delta^2 h_r\right) + A_r;$$

• if  $m\alpha = \sqrt{2}$ , then  $\forall r \in \mathbb{N}$  such that  $r > r_0$ ,

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, f)W > \delta\Big) \le \exp\left(-\frac{c'\delta^2 h_r}{r}\right) + A_r;$$

• if  $m\alpha > \sqrt{2}$ , then  $\forall r \in \mathbb{N}^*$  such that  $r > r_0$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, f)W > \delta\right) \le \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right) + A_r$$

where,

• for all  $r \in \mathbb{N}$ ,

$$A_r = \begin{cases} c' \exp\left(-c'' \delta^{2/3} (m^{1/3})^r\right) & \text{if } \mathbb{H}_r = \mathbb{G}_r \\ c' \exp\left(-c'' \delta^{2/3} \left(t_r / (r+1)^2\right)^{1/3}\right) & \text{if } \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

- r<sub>0</sub> := log (δ/c<sub>0</sub>) / log(α) − k<sub>0</sub>, with k<sub>0</sub> ∈ {0,1},
  c<sub>0</sub>, c' and c'' are positive constants which depend on α, m, c, and f and may differ line by line.

The next results can be seen as a consequence of the previous results.

**Theorem 3.4.** We assume that the hypotheses (H2)–(H4) are satisfied. Let  $f \in \mathcal{B}_b(S)$ . For all  $\delta > 0$ , for all a > 0 and for all b > 0 such that  $b < a/(\delta + 1)$ , we have

• if  $m\alpha < 1$ , then  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, f) > \delta \Big| W \ge a\Big) \le \exp\left(c''\delta b\right) \exp\left(-c'(\delta b)^2 h_r\right) + A_r;$$

• if  $m\alpha = 1$ , then for  $\mathbb{H}_r = \mathbb{G}_r$  and  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, f) > \delta \Big| W \ge a\Big) \le \exp\left(c'' \delta b\right) \exp\left(-c'(\delta b)^2 h_r\right) + A_r;$$

• if  $m\alpha = 1$ , then for  $\mathbb{H}_r = \mathbb{T}_r$  and  $\forall r \in \mathbb{N}$ ,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, f) > \delta \Big| W \ge a\Big) \le \exp\left(c'' \delta b(r+1)\right) \exp\left(-c'(\delta b)^2 h_r\right) + A_r;$$

• if  $1 < m\alpha < \sqrt{2}$ , then  $\forall r \in \mathbb{N}$  such that  $r > r_0$ .

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, f) > \delta \Big| W \ge a\Big) \le \exp\left(-c'(\delta b)^2 h_r\right) + A_r;$$

• if  $m\alpha = \sqrt{2}$ , then  $\forall r \in \mathbb{N}$  such that  $r > r_0$ ,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, f) > \delta \Big| W \ge a\Big) \le \exp\left(-\frac{c'(\delta b)^2 h_r}{r}\right) + A_r;$$

• if  $m\alpha > \sqrt{2}$ , then  $\forall r \in \mathbb{N}^*$  such that  $r > r_0$ ,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, f) > \delta \Big| W \ge a\Big) \le \exp\left(-\frac{c'(\delta b)^2}{\alpha^{2r}}\right) + A_r;$$

where.

• for all  $r \in \mathbb{N}$ ,

$$A_r = \begin{cases} c' \exp\left(-c''(\delta b)^{2/3} (m^{1/3})^r\right) & \text{if } \mathbb{H}_r = \mathbb{G}_r \\ c' \exp\left(-c''(\delta b)^{2/3} \left(t_r/(r+1)^2\right)^{1/3}\right) & \text{if } \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

- r<sub>0</sub> := log (δb/c<sub>0</sub>) / log(α) − k<sub>0</sub>, with k<sub>0</sub> ∈ {0,1},
  c<sub>0</sub>, c' and c'' are positive constants which depend on α, m, a, and f and may differ line by line.

We have the following extension of above theorems when f does not only depend on an individual  $X_i$ , but on the mother-daughters triangle  $\Delta_i$ .

**Theorem 3.5.** Let  $f \in \mathcal{B}_b(S^3)$ . If  $(\mu, P^*f) = 0$ , then, under the hypotheses (H3) and (H4), we have deviation inequalities of Theorem 3.2 for  $\widetilde{M}_{\mathbb{H}_r^*}(f)$ . If  $(\mu, P^*f) \neq 0$ , under additional hypothesis (H2), we have deviation inequalities of Theorem 3.3 for  $\widetilde{M}_{\mathbb{H}_r^*}(f) - (\mu, P^*f)W$  and of Theorem 3.4 for  $\overline{M}_{\mathbb{H}_r^*}(f) - (\mu, P^*f)$ .

**Remark 3.6.** Let us stress that by tedious, but straightforward calculations, the constants which appear in the previous inequalities can be made explicit.

We also stress that it is possible to prove two-sided exponential inequalities in the previous results.

Let us recall the following definition.

**Definition 3.7.** Let (E, d) be 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  $\delta > 0$ ,

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

This "exponential convergence" with speed  $v_n$  will be shortened as

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

**Remark 3.8.** Let  $(b_n)$  be a sequence of increasing positive real numbers such that

 $b_n \to +\infty$ 

and

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

From the previous deviation inequalities, we can deduce easily that

$$\widetilde{M}_{\mathbb{H}_r^*}(f) \xrightarrow[b_{\lfloor h_r \rfloor}^2]{}^{\text{superexp}} 0 \quad \text{if } \langle \mu, f \rangle = 0,$$

and if  $\langle \mu, f \rangle \neq 0$ , we have for  $m < 2^{3/5}$ 

$$\widetilde{M}_{\mathbb{H}_r^*}(f) \xrightarrow[[h_r]]{\text{superexp}} \langle \mu, f \rangle W$$

and  $\forall a > 0$ ,

$$\limsup_{r \to +\infty} \frac{1}{b_{\lfloor h_r \rfloor}^2} \log \mathbb{P}\left( \left| \overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f \rangle \right| > \delta \right| W \ge a \right) = -\infty.$$

So, for the exponential convergence of averages  $\widetilde{M}_{\mathbb{H}_{r}^{*}}(f)$  and  $\overline{M}_{\mathbb{H}_{r}^{*}}(f)$ , there are three regimes according to the value of  $m\alpha$  compared to  $\sqrt{2}$ . This phenomenon is not observed in the limit theorems of Delmas and Marsalle [15].

So, our deviation inequalities highlight a competition between the ergodicity of the embedded Markov chain with transition probability Q and the Galton–Watson binary tree.

**Remark 3.9.** Once again, let us note that when m = 2, that is the no death case, we recover exactly the results of Bitseki *et al.* [9]. Consequently, for exponential deviation inequalities, this allows us to say that our results are relevant in the sense that, they enable to extend known results [9].

# 4. Application: First order bifurcating autoregressive processes with missing Data

We consider the asymmetric autoregressive processes given in Section 1.1. Notice that the process  $(X_n, n \in \mathbb{T})$  defined in Section 1.1, with the convention that  $X_n = \partial$  if the cell n is missing, is a spatially homogeneous BMC on a GW tree. In all this section, we assume that the distribution of the ancestor  $X_1$ ,  $\nu$ , has finite moments of all orders. In the sequel, we will study the process  $(X_n, n \in \mathbb{T})$  in two settings:

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

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

and  $(\varepsilon'_{2n}, n \in \mathbb{T})$  (resp.  $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ ) forms a sequence of i.i.d. random variables with law  $\mathcal{N}(0, \sigma_0^2)$  (resp.  $\mathcal{N}(0, \sigma_1^2)$ ) with  $\sigma_0^2 > 0$  and  $\sigma_1^2 > 0$ .

• the bounded setting which corresponds to the case where the noise sequences  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T}),$  $(\varepsilon'_{2n}, n \in \mathbb{T})$  and  $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ , and the initial state  $X_1$  take their values in a compact set. The latter implies that the process  $(X_n, n \in \mathbb{T})$  is bounded. In the bounded setting, we will assume hypothesis (**H4**), that is  $2p_{1,0} + p_1 + p_0 > \sqrt{2}$ . This implies in particular that the BMC on GW is super-critical. We denote by S the state space of  $(X_n, n \in \mathbb{T})$ . In the bounded setting, we will assume without loss of generality that S is a compact subset of  $\mathbb{R}$ .

Let  $\mathbb{T}_n^{0,1}$  be the subset of cells in  $\mathbb{T}_n^*$  with two living daughters,  $\mathbb{T}_n^0$  (resp.  $\mathbb{T}_n^1$ ) be the set of cells of  $\mathbb{T}_n^*$  with only the new (resp. old) pole daughter alive:

$$\mathbb{T}_n^{1,0} = \left\{ i \in \mathbb{T}_n^* : \Delta_i \in S^3 \right\}, \quad \mathbb{T}_n^0 = \left\{ i \in \mathbb{T}_n^* : \Delta_i \in S^2 \times \{\partial\} \right\}$$
  
and  $\mathbb{T}_n^1 = \{ i \in \mathbb{T}_n^* : \Delta_i \in S \times \{\partial\} \times S \}.$ 

We compute the least-squares estimator (LSE)

$$\widehat{\theta}_n = (\widehat{\alpha}_0^n, \widehat{\beta}_0^n, \widehat{\alpha}_1^n, \widehat{\beta}_1^n, \widehat{\alpha}_0^{'n}, \widehat{\beta}_0^{'n}, \widehat{\alpha}_1^{'n}, \widehat{\beta}_1^{'n})$$

of  $\theta$  given by (1.5), based on the observation of a sub-tree  $\mathbb{T}_{n+1}^*$ . Consequently, we obviously have for  $\eta \in \{0, 1\}$ ,

$$\begin{split} \widehat{\alpha}_{\eta}^{n} &= \frac{|\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{1,0}} X_{i} X_{2i+\eta} - \left(|\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{1,0}} X_{i}\right) \left(|\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{1,0}} X_{2i+\eta}\right)}{|\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{1,0}} X_{i}^{2} - \left(|\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{1,0}} X_{i}\right)^{2}} \\ \widehat{\beta}_{\eta}^{n} &= |\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{2i+\eta} - \widehat{\alpha}_{\eta}^{n} |\mathbb{T}_{n}^{1,0}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{i}, \\ \\ \widehat{\alpha}_{\eta}^{\prime n} &= \frac{|\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{i} X_{2i+\eta} - \left(|\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{i}\right) \left(|\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{2i+\eta}\right)}{|\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{i}^{2} - \left(|\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{i}\right)^{2}}, \\ \widehat{\beta}_{\eta}^{\prime n} &= |\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{2i+\eta} - \widehat{\alpha}_{\eta}^{\prime n} |\mathbb{T}_{n}^{\eta}|^{-1} \sum_{i \in \mathbb{T}_{n}^{n}} X_{i}. \end{split}$$

Let us stress that in the Gaussian setting, those LSE correspond to the maximum likelihood estimators.

Notice that those LSE are based on polynomial functions of the observations. So, in the bounded setting, since the latter are bounded, we are in the functional setting of the results of Section 3.

In the Gaussian setting, we will take  $F = C_{pol}^1(\mathbb{R})$  the set of all  $C^1$  functions  $f : \mathbb{R} \to \mathbb{R}$  such that |f| + |f'| is bounded above by a polynomial. Following the proof of Lemma 26 in [17], one can check that F fulfills hypothesis (i)-(vi), then we are in the functional setting of the results of Section 2.

Recalling the Markov chain  $(Y_n, n \in \mathbb{N})$ , notice that  $Y_n$  is distributed as  $Z_n = a_1 a_2 \dots a_{n-1} a_n Y_0 + \sum_{k=1}^n a_1 a_2 \dots a_{k-1} b_k$ , where  $b_n = b'_n + s_n e_n$ ,  $((a_n, b'_n, s_n), n \ge 1)$  is a sequence of independent identically distributed random variables, whose common distribution is given by, for  $\eta \in \{0, 1\}$ ,

$$\mathbb{P}(a_1 = \alpha_\eta, b'_1 = \beta_\eta, s_1 = \sigma) = \frac{p_{1,0}}{m}$$
 and  $\mathbb{P}(a_1 = \alpha'_\eta, b'_1 = \beta'_\eta, s_1 = \sigma_\eta) = \frac{p_\eta}{m}$ ,

 $(e_n, n \ge 1)$  is a sequence of independent  $\mathcal{N}(0, 1)$  random variables, and is independent of  $((a_n, b'_n, s_n), n \ge 1)$ , and both sequences are independent of  $Y_0$ . Moreover, it is easy to check that the sequence  $(Z_n, n \in \mathbb{N})$  converges a.s. to a limit Z, which implies that the Markov chain  $(Y_n, n \in \mathbb{N})$  converges in distribution to Z. We refer to [15], Section 6, for more details.

Following the proof of Proposition 28, Step 1 in [17], we check hypothesis (H1) and (H3) with  $\alpha = \max(|\alpha_0|, |\alpha_1|, |\alpha_0'|, |\alpha_1'|) < 1$  and with  $\mu$  the distribution of Z.

Let  $\mu_1 = \mathbb{E}[Z]$  and  $\mu_2 = \mathbb{E}[Z^2]$ . We have (see [15])

$$\mu_1 = \frac{\overline{\beta}}{1 - \overline{\alpha}}$$
 and  $\mu_2 = \frac{2\overline{\alpha\beta\beta}/(1 - \overline{\alpha}) + \overline{\beta^2} + \overline{\alpha^2}}{1 - \overline{\alpha^2}},$ 

where  $\overline{\alpha} = \mathbb{E}[a_1], \overline{\alpha^2} = \mathbb{E}[a_1^2], \overline{\beta} = \mathbb{E}[b_1], \overline{\beta^2} = \mathbb{E}[b_1^2], \overline{\alpha\beta} = \mathbb{E}[a_1b_1] \text{ and } \overline{\sigma^2} = \mathbb{E}[s_1^2].$ As the case may be, we have the following deviation inequalities for  $\widehat{\theta}_n - \theta$ .

#### 4.1. The Gaussian setting

**Proposition 4.1.** For all  $\delta > 0$ , for all a > 0 and for all  $\gamma > 0$  such that

$$\gamma < \min\left\{c_1/(1+\delta), c_1/\left(1+\sqrt{\delta}\right), c_1/\left(1+\sqrt[4]{\delta}\right)\right\},\$$

where  $c_1$  is a positive constant which depends on  $p_{1,0}$ ,  $p_0$ ,  $p_1$ ,  $\mu_1$  and  $\mu_2$ , we have

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta|W \ge a\right) \le \begin{cases} \frac{c}{\gamma^{4q}\delta^{4-p}} \left(\frac{1}{m^2}\right)^{n+1} & \text{if } m\alpha^2 < 1\\ \frac{c}{\gamma^{4q}\delta^{4-p}} r^2 \left(\frac{1}{m^2}\right)^{n+1} & \text{if } m\alpha^2 = 1\\ \frac{c}{\gamma^{4q}\delta^{4-p}} \alpha^{4(n+1)} & \text{if } m\alpha^2 > 1, \end{cases}$$

where  $p = p(\delta) \in \{0, 2, 4\}$ ,  $q = q(\delta) \in \{0, 1\}$  and the constants c, c' and c'' depend on  $p_{1,0}$ ,  $p_0$ ,  $p_1 \alpha$ , m, W,  $\mu_1$ ,  $\mu_2$  and differ line by line.

**Remark 4.2.** The values of p and q in Proposition 4.1 depend on the order of  $\delta$ . For example, if  $\delta$  is small enough, we have p = 0 and q = 0.

**Remark 4.3.** Using Remarks 2.9 and 2.10, the previous inequalities can be improved. This improvement is related to the calculation of the 2k-order moments with  $k \in \mathbb{N}^*$ . But, as we have said, this comes at the price of enormous computations.

# 4.2. Compact case: The uniformly ergodic setting.

**Proposition 4.4.** For all  $\delta > 0$ , for all a > 0, for all b > 0 and for all  $\gamma > 0$  such that  $b < a/(\delta + 1)$  and  $\gamma < \min\left\{c_1/(1+\delta), c_1/(1+\sqrt{\delta})\right\}$ , where  $c_1$  is a positive constant which depends on  $p_{1,0}$ ,  $p_0$ ,  $p_1$ ,  $\mu_1$  and  $\mu_2$ , and for  $n_0 := (\log(\gamma^q \delta^p b/c_0)/\log \alpha) - 1$ , we have

• if  $m\alpha < 1$ , then  $\forall n \in \mathbb{N}$ ,

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta|W \ge a\right) \le c_2 \exp\left(c''\gamma^q \delta^p b\right) \exp\left(-c'\left(\gamma^q \delta^p b\right)^2 h_n\right) + A_n;$$

• if  $m\alpha = 1$ , then  $\forall n \in \mathbb{N}$ ,

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta|W \ge a\right) \le c_2 \exp\left(c''\gamma^q \delta^p b(n+1)\right) \exp\left(-c'\left(\gamma^q \delta^p b\right)^2 h_n\right) + A_n$$

• if  $1 < m\alpha < \sqrt{2}$ , then  $\forall n \in \mathbb{N}$  such that  $n > n_0$ ,

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \ge a\right) \le c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 h_n\right) + A_n;$$

• if  $m\alpha = \sqrt{2}$ , then  $\forall n \in \mathbb{N}$  such that  $n > n_0$ ,

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \ge a\right) \le c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 (1/n) h_n\right) + A_n;$$

• if  $m\alpha > \sqrt{2}$ , then  $\forall n \in \mathbb{N}^*$  such that  $n > n_0$ ,

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \ge a\right) \le c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 \alpha^{-2n}\right) + A_n;$$

where

where  $A_n = c_3 \exp\left(c' \left(\gamma^q \delta^p b\right)^{2/3}\right) \exp\left(-c'' \left(\gamma^q \delta^p b\right)^{2/3} \left(t_n/(n+1)^2\right)^{1/3}\right), \ p \in \{1/2, 1\}, \ q \in \{0, 1/2, 1\}, \ c_2, \ c_3, \ c_4, \ c'$ and c'' are positive constants which depend on  $c, \ m, \ \alpha \ p_{1,0}, \ p_0, \ p_1, \ \mu_1 \ and \ \mu_2.$ 

**Remark 4.5.** Note that the constants  $c_2$ ,  $c_3$ ,  $c_4$ , c' and c'' which appear in Proposition 4.4 may differ term by term. The values of p and q depend on the magnitude of  $\delta$  and  $\gamma$ . For example, for  $\delta$  and  $\gamma$  small enough, we have p = 1 and q = 1. We also stress that all these constants can be made explicit by tedious calculations.

#### APPENDIX A. PROOFS OF THE POLYNOMIAL DEVIATION INEQUALITIES

This section is devoted to the proofs of the Theorems 2.5 and 2.8, and Proposition 4.1.

#### A.1. Proof of Theorem 2.5

Let  $f \in F$  such that  $(\mu, f) = 0$ . We shall study the two empirical averages  $\widetilde{M}_{\mathbb{G}_r^*}(f)$  and  $\widetilde{M}_{\mathbb{T}_r^*}(f)$  successively. **Part 1.** Let us first deal with  $\widetilde{M}_{\mathbb{G}_{*}^{*}}(f)$ . By Markov inequality, we get, for all  $\delta > 0$ 

$$\mathbb{P}\Big(\big|\widetilde{M}_{\mathbb{G}_r}(f)\big| > \delta\Big) = \mathbb{P}\Big(\big|\widetilde{M}_{\mathbb{G}_r}(f)\big|^4 > \delta^4\Big) \le \frac{1}{\delta^4} \mathbb{E}\Big[\big(\widetilde{M}_{\mathbb{G}_r}(f)\big)^4\Big].$$

Now, (2.5) follows from (2.1).

**Part 2.** Let us now consider  $\widetilde{M}_{\mathbb{T}^*_n}(f)$ . In the same way as before, (2.6) follows from (2.2) after one application of Markov inequality.

## A.2. Proof of Corollary 2.6

Let  $f \in F$  such that  $(\mu, f) \neq 0$ . Once again, we are going to study successively  $\widetilde{M}_{\mathbb{G}_{r}^{*}}(f)$  and  $\widetilde{M}_{\mathbb{T}_{r}^{*}}(f)$ . **Part 1.** Let us first deal with  $M_{\mathbb{G}_{\pi}^*}(f)$ . Set  $g = f - (\mu, f)$ . Then,  $(\mu, g) = 0$  and

$$\widetilde{M}_{\mathbb{G}_r^*}(f) = \widetilde{M}_{\mathbb{G}_r^*}(g) + (|\mathbb{G}_r^*|/m^r)(\mu, f).$$

We have for all  $\delta > 0$ 

$$\mathbb{P}\left(\left|\widetilde{M}_{\mathbb{G}_r^*}(f) - (\mu, f)W\right| > \delta\right) \le \mathbb{P}\left(\left|\widetilde{M}_{\mathbb{G}_r^*}(g)\right| > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \frac{\delta}{2|(\mu, f)|}\right).$$
(A.1)

As  $(\mu, g) = 0$ , the previous computations (Proof of Thm. 2.5) give us some bound for the first term of right hand of (A.1), similar to those obtain in Theorem 2.5. Let us stress that since  $|\mathbb{G}_1^*| \leq 2$ , we have  $\mathbb{E} \left[ \exp \left( \theta |\mathbb{G}_1^*| \right) \right] < +\infty$ for all  $\theta > 0$ . Now, under hypothesis (H2), we deduce, from [1] Theorem 5, that

$$\mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \frac{\delta}{2|(\mu, f)|}\right) \le c \exp\left(-c'\delta^{2/3}m^{r/3}\right).$$

and this ends the proof of (2.7).

**Part 2.** Let us look at  $\widetilde{M}_{\mathbb{T}_r^*}(f)$ . For  $f \in F$ , set  $g = f - (\mu, f)$ . Then,  $(\mu, g) = 0$  and

$$\widetilde{M}_{\mathbb{T}_r^*}(f) = \widetilde{M}_{\mathbb{T}_r^*}(g) + (|\mathbb{T}_r^*|/t_r)(\mu, f).$$

We have for all  $\delta > 0$ 

$$\mathbb{P}\left(\left|\widetilde{M}_{\mathbb{T}_{r}^{*}}(f) - (\mu, f)W\right| > \delta\right) \le \mathbb{P}\left(\left|\widetilde{M}_{\mathbb{T}_{r}^{*}}(g)\right| > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{T}_{r}^{*}|}{t_{r}} - W\right| > \frac{\delta}{2|(\mu, f)|}\right).$$
(A.2)

Since  $(\mu, g) = 0$ , the first term of the right hand of (A.2) can be bounded as in the previous computations (Proof of Thm. 2.5). Under additional hypothesis (H2), we have, from [1] Theorem 5,

$$\begin{split} \mathbb{P}\left(\left|\frac{|\mathbb{T}_r^*|}{t_r} - W\right| > \frac{\delta}{2|(\mu, f)|}\right) &\leq \sum_{q=0}^r \mathbb{P}\left(\frac{m^q}{t_r} \left|\frac{|\mathbb{G}_q^*|}{m^q} - W\right| > \frac{\delta}{2(r+1)|(\mu, f)|}\right) \\ &= \sum_{q=0}^r \mathbb{P}\left(\left|\frac{|\mathbb{G}_q^*|}{m^q} - W\right| > \frac{\delta t_r}{2(r+1)|(\mu, f)|m^q}\right) \\ &\leq \sum_{q=0}^r c \exp\left(-c'\delta^{2/3}\left(\frac{t_r^2}{(r+1)^2m^q}\right)^{1/3}\right) \\ &\leq c \exp\left(-c'\delta^{2/3}\left(\frac{t_r}{(r+1)^2}\right)^{1/3}\right), \end{split}$$

where the last inequality comes from the fact up to a multiplicative constant c which does not depend on r (and which may differ term by term), we have

$$\sum_{q=0}^{r-1} \exp\left(-c'\delta^{2/3}\left(\left(\frac{t_r^2}{(r+1)^2m^q}\right)^{1/3} - \left(\frac{t_r^2}{(r+1)^2m^r}\right)^{1/3}\right)\right) \le cr \exp\left(-c\delta^{2/3}\left(m^{1/3} - 1\right)\frac{m^{r/3}}{(r+1)^{2/3}}\right)$$
$$\lim r \exp\left(-c\delta^{2/3}\left(m^{1/3} - 1\right)\frac{m^{r/3}}{(r+1)^{2/3}}\right) = 0.$$

and

$$\lim_{r \to +\infty} r \exp\left(-c\delta^{2/3} \left(m^{1/3} - 1\right) \frac{m^{r/3}}{(r+1)^{2/3}}\right) = 0.$$

We thus conclude that the latter sum is bounded and this ends the proof of (2.8).

# A.3. Proof of Theorem 2.8

Let  $f \in F$ . Without loss of generality, we assume that  $(\mu, f) = 0$ . Otherwise, we take  $f - (\mu, f)$ . We first deal with  $\overline{M}_{\mathbb{G}_n^*}(f)$ . Let  $\delta > 0$  and a > 0. We assume that  $\mathbb{P}(W > a) \neq 0$ . We have

$$\begin{split} \mathbb{P}\left(\left|\overline{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta|W \ge a\right) &= \mathbb{P}\left(\left|\overline{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta, \frac{|\mathbb{G}_{r}^{*}|}{m^{r}} > b|W \ge a\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta, \frac{|\mathbb{G}_{r}^{*}|}{m^{r}} \le b|W \ge a\right) \\ &= \frac{1}{\mathbb{P}\left(W \ge a\right)} \left(\mathbb{P}\left(\left|\overline{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta, \frac{|\mathbb{G}_{r}^{*}|}{m^{r}} > b, W \ge a\right) \right) \\ &+ \mathbb{P}\left(\left|\overline{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta, \frac{|\mathbb{G}_{r}^{*}|}{m^{r}} \le b, W \ge a\right)\right) \\ &\leq p_{a} \mathbb{P}\left(\left|\widetilde{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta b\right) + p_{a} \mathbb{P}\left(\left|\frac{|\mathbb{G}_{r}^{*}|}{m^{r}} - W\right| > W - b, W \ge a\right) \\ &\leq p_{a} \mathbb{P}\left(\left|\widetilde{M}_{\mathbb{G}_{r}^{*}}(f)\right| > \delta b\right) + p_{a} \mathbb{P}\left(\left|\frac{|\mathbb{G}_{r}^{*}|}{m^{r}} - W\right| > \delta b\right), \end{split}$$

where  $p_a = \mathbb{P}(W \ge a)^{-1}$  and  $b = a/(\delta + 1)$ . Now, the first term of the last inequality can be bounded as in Theorem 2.5, and the second term is bounded as in the **Part 1** of Corollary 2.6. This ends the proof (2.9).

The proof of (2.10) can be dealt with in the same way, with  $\mathbb{T}_r^*$  (resp.  $t_r$ ) taking the place of  $\mathbb{G}_r^*$  (resp.  $m^r$ ).

#### A.4. Proof of Proposition 4.1

We are going to treat  $\hat{\alpha}_{0}^{n} - \alpha_{0}$ . Deviation inequalities for  $\hat{\alpha}_{1}^{n} - \alpha_{1}$ ,  $\hat{\beta}_{\eta}^{n} - \beta_{\eta}$ ,  $\hat{\alpha}_{\eta}^{'n} - \alpha_{\eta}^{'}$ ,  $\hat{\beta}_{\eta}^{'n} - \beta_{\eta}^{'}$ ,  $\eta \in \{0, 1\}$ , can be treated in the same way. Recalling that we take  $F = C_{pol}^{1}(\mathbb{R})$ . Let  $g_{1}, g_{2}, h_{1}$  and  $h_{2}$  be the functions defined on  $S^{3}$  respectively by

$$g_{1}(x, y, z) = (xy - x(\alpha_{0}x + \beta_{0})) \mathbf{1}_{S^{3}}(x, y, z)$$
$$g_{2}(x, y, z) = (y - \alpha_{0}x - \beta_{0})) \mathbf{1}_{S^{3}}(x, y, z),$$
$$h_{1}(x, y, z) = x\mathbf{1}_{S^{3}}(x, y, z),$$
$$h_{2}(x, y, z) = x^{2}\mathbf{1}_{S^{3}}(x, y, z).$$

It is easy to see that  $P^*g_1(x) = 0$ ,  $P^*g_2(x) = 0$ ,  $P^*h_1(x) = p_{1,0}x$  and  $P^*h_2(x) = p_{1,0}x^2$  where  $P^*$  denote the transition kernel associated to the BAR(1) process with missing data. With these notations, we can rewrite  $\widehat{\alpha}_0^n - \alpha_0$  as

$$\widehat{\alpha}_{0}^{n} - \alpha_{0} = \frac{|\mathbb{T}_{n}^{*}|^{-1} |\mathbb{T}_{n}^{1,0}| \left( |\mathbb{T}_{n}^{*}|^{-1} \sum_{i \in \mathbb{T}_{n}^{*}} g_{1}(\varDelta_{i}) \right)}{B_{n}} - \frac{\left( |\mathbb{T}_{n}^{*}|^{-1} \sum_{i \in \mathbb{T}_{n}^{*}} h_{1}(\varDelta_{i}) \right) \left( |\mathbb{T}_{n}^{*}|^{-1} \sum_{i \in \mathbb{T}_{n}^{*}} g_{2}(\varDelta_{i}) \right)}{B_{n}},$$
where  $B_{n} = |\mathbb{T}_{n}^{*}|^{-1} |\mathbb{T}_{n}^{1,0}| \left( |\mathbb{T}_{n}^{*}|^{-1} \sum_{i \in \mathbb{T}_{n}^{*}} h_{2}(\varDelta_{i}) \right) - \left( |\mathbb{T}_{n}^{*}|^{-1} \sum_{i \in \mathbb{T}_{n}^{*}} h_{1}(\varDelta_{i}) \right)^{2}.$ 
Recalling (1.8), we then have for all  $\delta > 0$  and  $a > 0$ 

$$\mathbb{P}\left(\left|\widehat{\alpha}_{0}^{n}-\alpha_{0}\right| > \delta|W \ge a\right) \le \mathbb{P}\left(\frac{\left|\mathbb{T}_{n}^{*}\right|^{-1}\left|\mathbb{T}_{n}^{1,0}\right|\left|\overline{M}_{\mathbb{T}_{n}^{*}}(g_{1})\right|}{B_{n}} > \frac{\delta}{2}\left|W \ge a\right) + \mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_{n}^{*}}(h_{1})\right|\left|\overline{M}_{\mathbb{T}_{n}^{*}}(g_{2})\right|}{B_{n}} > \frac{\delta}{2}\left|W \ge a\right).$$
(A.3)

For the first term of the right hand of (A.3), since  $|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}| \leq 1$ , we have for all  $\gamma > 0$ 

$$\mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{B_n} > \frac{\delta}{2} \Big| W \ge a\right) \le \mathbb{P}\left(B_n < \gamma | W \ge a\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_n^*}(g_1)\right| > \frac{\delta\gamma}{2} \Big| W \ge a\right).$$

Notice that

$$B_n - (p_{1,0}^2 \mu_2 - p_{1,0}^2 \mu_1^2) = p_{1,0} \mu_2 \left( \frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0} \right) + \frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} \overline{M}_{\mathbb{T}_n^*} (h_2 - p_{1,0} \mu_2) - \left( \overline{M}_{\mathbb{T}_n^*} (h_1 - p_{1,0} \mu_1) \right)^2 - 2p_{1,0} \mu_1 \overline{M}_{\mathbb{T}_n^*} (h_1 - p_{1,0} \mu_1)$$

We then have for all  $0 < \gamma < \frac{2|p_{1,0}^2 \mu_2 - p_{1,0}^2 \mu_1^2|}{2 + \delta}$ ,

$$\begin{split} & \mathbb{P}\left(\frac{|\mathbb{T}_{n}^{*}|^{-1}|\mathbb{T}_{n}^{1,0}||\overline{M}_{\mathbb{T}_{n}^{*}}(g_{1})|}{B_{n}} > \frac{\delta}{2}\Big|W \ge a\right) \\ & \leq \mathbb{P}\left(\left(p_{1,0}^{2}\mu_{2} - p_{1,0}^{2}\mu_{1}^{2}\right) - B_{n} > \frac{\gamma\delta}{2}\Big|W \ge a\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_{n}^{*}}(g_{1})\right| > \frac{\delta\gamma}{2}\Big|W \ge a\right) \\ & \leq \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_{n}^{*}}(g_{1})\right| > \frac{\delta\gamma}{2}\Big|W \ge a\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_{n}^{*}}(h_{2} - p_{1,0}\mu_{2})\right| > \frac{\gamma\delta}{8}\Big|W \ge a\right) \\ & + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_{n}^{*}}(h_{1} - p_{1,0}\mu_{1})\right| > \frac{\sqrt{\gamma\delta}}{2\sqrt{2}}\Big|W \ge a\right) + \mathbb{P}\left(\left|\frac{|\mathbb{T}_{n}^{1,0}|}{|\mathbb{T}_{n}^{*}|} - p_{1,0}\right| > \frac{\delta\gamma}{8p_{1,0}\mu_{2}}\Big|W \ge a\right) \\ & + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_{n}^{*}}(h_{1} - p_{1,0}\mu_{1})\right| > \frac{\gamma\delta}{16p_{1,0}\mu_{1}}\Big|W \ge a\right). \end{split}$$

From [16], Section 5, we have

$$\mathbb{P}\left(\left|\frac{|\mathbb{T}_{n}^{1,0}|}{|\mathbb{T}_{n}^{*}|} - p_{1,0}\right| > \frac{\delta\gamma}{8p_{1,0}\mu_{2}} \Big| W \ge a\right) = \mathbb{P}\left(\left|\frac{1}{|\mathbb{T}_{n}^{*}|} \sum_{j=1}^{|\mathbb{T}_{n}^{*}|} (T_{j} - p_{1,0})\right| > \frac{\delta\gamma}{8p_{1,0}\mu_{2}} \Big| W \ge a\right),\tag{A.4}$$

where  $(T_j)_{j>1}$  is a sequence of i.i.d. Bernoulli random variables such that

$$p_{1,0} = \mathbb{P}(T_j = 1) = 1 - \mathbb{P}(T_j = 0).$$

Using the same ideas as in the proof of Theorem 2.8, we get

$$\mathbb{P}\left(\left|\frac{1}{|\mathbb{T}_n^*|}\sum_{j=1}^{|\mathbb{T}_n^*|} (T_j - p_{1,0})\right| > \frac{\delta\gamma}{8p_{1,0}\mu_2} |W \ge a\right) \\
\leq p_a \mathbb{P}\left(\frac{1}{t_n}\left|\sum_{j=1}^{|\mathbb{T}_n^*|} (T_j - p_{1,0})\right| > \frac{\delta\gamma b}{8p_{1,0}\mu_2}\right) + p_a \mathbb{P}\left(\left|\frac{|T_n^*|}{t_n} - W\right| > \frac{\delta\gamma b}{8p_{1,0}\mu_2}\right),$$

where b is taken in the same way as in the proof of Theorem 2.8. Now, the majoration of the right hand side of (A.4) follows as in the **Part 2** of the proof of Theorem 2.6.

For the second term of the right hand of (A.3), we have

$$\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_{n}^{*}}(h_{1})||\overline{M}_{\mathbb{T}_{n}^{*}}(g_{2})|}{|B_{n}|} > \frac{\delta}{2}\Big|W \ge a\right) \le \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_{n}^{*}}(g_{2})|}{|B_{n}|} > \frac{\delta}{4p_{1,0}\mu_{1}}\Big|W \ge a\right) + \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_{n}^{*}}(g_{2})|}{|B_{n}|} > \frac{\sqrt{\delta}}{2}\Big|W \ge a\right) + \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_{n}^{*}}(h_{1}-p_{1,0}\mu_{1})| > \frac{\sqrt{\delta}}{2}\Big|W \ge a\right)$$

Now, the first and the second term of the right hand of the last inequality can be treated as the first term of the right hand of (A.3).

Finally, to get the result, just apply Theorem 2.8 and Remark 2.10 to functions  $g_1$ ,  $g_2$ ,  $h_1$  and  $h_2$ .

# Appendix B. Proofs of the exponential deviation inequalities

This section is devoted to the proof of the Theorems 3.2, 3.3, 3.4 and Proposition 4.4.

For the convenience of the readers, let us first state Azuma–Bennett–Hoeffding's inequality [3,5,19,21].

**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

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

## B.1. Proof of Theorem 3.2

Let  $f \in \mathcal{B}_b(S)$  such that  $(\mu, f) = 0$ . We are going to study successively  $\widetilde{M}_{\mathbb{H}_r^*}(f)$  for  $\mathbb{H}_r = \mathbb{G}_r$  and  $\mathbb{H}_r = \mathbb{T}_r$ . **Step 1.** Let us first deal with  $\widetilde{M}_{\mathbb{G}_r^*}(f)$ . By Chernoff's inequality, we have for all  $\delta > 0$  and for all  $\lambda > 0$ 

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta m^r\right) \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r^*} f(X_i)\right)\right].$$
(B.1)

Recall that for all  $i \in \mathbb{G}_{r-1}^*$ ,

$$\mathbb{E}\left[f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}|\mathcal{F}_{r-1}\right] = mQf(X_i)$$

By substracting and adding terms in expectation of the right hand of (B.1), and conditioning with respect to  $\mathcal{F}_{r-1}$ , we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}f(X_{i})\right)\right] = \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^{*}}mQf(X_{i})\right) \times \mathbb{E}\left[\exp\left(\sum_{i\in\mathbb{G}_{r-1}^{*}}\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^{*}\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^{*}\}} - mQf(X_{i})\right)\right)\middle|\mathcal{F}_{r-1}\right]\right].$$
 (B.2)

Observing that  $\mathbb{G}_{r-1}^*$  is  $\mathcal{F}_{r-1}$  measurable, and using the fact that conditionally to  $\mathcal{F}_{r-1}$ , the triplets  $\{(\Delta_i), i \in \mathbb{G}_{r-1}\}$  are independent (this is due to the Markov's property), we have

$$\mathbb{E}\left[\exp\left(\sum_{i\in\mathbb{G}_{r-1}^{*}}\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^{*}\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^{*}\}}-mQf(X_{i})\right)\right)\middle|\mathcal{F}_{r-1}\right]$$
$$=\prod_{i\in\mathbb{G}_{r-1}^{*}}\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^{*}\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^{*}\}}-mQf(X_{i})\right)\right)\middle|\mathcal{F}_{r-1}\right].$$
(B.3)

Using Azuma–Bennett–Hoeffding's inequality (see Lem. B.1) and according to (H3) and recalling that  $Q^0 f = f$ , we get for all  $i \in \mathbb{G}_{r-1}^*$ ,

$$\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right) \middle| \mathcal{F}_{r-1}\right] \le \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2}{2}\right)$$

From (B.3), this implies that

$$\mathbb{E}\left[\exp\left(\sum_{i\in\mathbb{G}_{r-1}^*}\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\right)\right)\Big|\mathcal{F}_{r-1}\right]$$
$$\leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}^*|}{2}\right).$$

Recalling (B.2), we are led to

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

Let  $\{\mathcal{G}_r, r \in \mathbb{N}\}$  the filtration defined by  $\mathcal{G}_r = \sigma\{|\mathbb{G}_0^*|, \ldots, |\mathbb{G}_r^*|\}$ . By conditioning with respect to  $\mathcal{G}_{r-1}$ , we are led to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}f(X_{i})\right)\right] \leq \mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}|\mathbb{G}_{r-1}^{*}|}{2}\right) \times \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^{*}}mQf(X_{i})\right)\Big|\mathcal{G}_{r-1}\right]\right]$$

For all  $q \in \{1, \ldots, r\}$ , let  $\widetilde{\mathcal{F}}_{r-q}$  the  $\sigma$ -algebras defined by  $\widetilde{\mathcal{F}}_{r-q} = \sigma(\mathcal{F}_{r-q}, \mathcal{G}_{r-1})$ . We have

$$\mathbb{E}\left[\exp\left(\lambda m \sum_{i \in \mathbb{G}_{r-1}^*} Qf(X_i)\right) \middle| \mathcal{G}_{r-1}\right] = \mathbb{E}\left[\exp\left(\lambda m^2 \sum_{i \in \mathbb{G}_{r-2}^*} Q^2 f(X_i)\right) \\ \times \prod_{i \in \mathbb{G}_{r-2}^*} \mathbb{E}\left[\exp\left(\lambda m \left(Qf(X_{2i})\mathbf{1}_{\{2i \in \mathbb{T}^*\}}\right) + Qf(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQ^2 f(X_i)\right)\right) \middle| \widetilde{\mathcal{F}}_{r-2}\right] \middle| \mathcal{G}_{r-1}\right].$$

Since for  $i\in \mathbb{G}_{r-2}^{*}$  we have

$$\mathbb{E}\left[Qf(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + Qf(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQ^2f(X_i)\Big|\widetilde{\mathcal{F}}_{r-2}\right] = 0,$$

Azuma–Bennett–Hoeffding's inequality leads us to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}f(X_{i})\right)\right] \leq \mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}|\mathbb{G}_{r-1}^{*}|}{2}\right)\right.$$
$$\times \exp\left(\frac{c^{2}\lambda^{2}m^{2}(2\alpha+m\alpha^{2})^{2}|\mathbb{G}_{r-2}^{*}|}{2}\right)$$
$$\times \mathbb{E}\left[\exp\left(\lambda m^{2}\sum_{i\in\mathbb{G}_{r-2}^{*}}Q^{2}f(X_{i})\right)\left|\mathcal{G}_{r-1}\right]\right].$$

Iterating this procedure, we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}f(X_{i})\right)\right] \leq \mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}}{2}\sum_{q=0}^{r-1}\left(2\alpha^{q}+m\alpha^{q+1}\right)^{2}m^{2q}\left|\mathbb{G}_{r-q-1}^{*}\right|\right)\right]$$
$$\times \mathbb{E}\left[\exp\left(\lambda m^{r}Q^{r}f(X_{1})\right)|\mathcal{G}_{r-1}\right]\right]$$
$$\leq \exp\left(\lambda c(\alpha m)^{r}\right) \times \mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}}{2}\sum_{q=0}^{r-1}\left(\alpha^{2}m^{2}\right)^{r-q-1}\left|\mathbb{G}_{q}^{*}\right|\right)\right],$$

where the last inequality was obtained from (H3) and after a change of indices.

Now we are going to control the expectation which appears in the last inequality. By conditioning with respect to  $\mathcal{G}_{r-2}$ , we are led to

$$\mathbb{E}\left[\exp\left(\frac{c^2\lambda^2(2+m\alpha)^2}{2}\sum_{q=0}^{r-1}\left(\alpha^2m^2\right)^{r-q-1}\left|\mathbb{G}_q^*\right|\right)\right] = \mathbb{E}\left[\exp\left(\frac{c^2\lambda^2(2+m\alpha)^2}{2}\sum_{q=0}^{r-2}\left(\alpha^2m^2\right)^{r-q-1}\left|\mathbb{G}_q^*\right|\right) \times \mathbb{E}\left[\exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2}\left|\mathbb{G}_{r-1}^*\right|\right)\left|\left|\mathbb{G}_{r-2}^*\right|\right]\right].$$

Now, recalling that the process  $(|\mathbb{G}_k^*|, k \in \mathbb{N})$  is a GW process with the reproduction generating function  $\psi(z) = (p_0 + p_1)z + p_{1,0}z^2$  and using the convexity of the power functions with exponent greater than 1, hypothesis **(H2)** and the properties of Galton–Watson's processes, we are led to

$$\mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}}{2}\sum_{q=0}^{r-1}(\alpha^{2}m^{2})^{r-q-1}|\mathbb{G}_{q}^{*}|\right)\right]$$

$$\leq \mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}}{2}\sum_{q=0}^{r-2}(\alpha^{2}m^{2})^{r-q-1}|\mathbb{G}_{q}^{*}|\right)$$

$$\times\left((p_{0}+p_{1})\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}}{2}|\mathbb{G}_{r-2}^{*}|\right)+p_{1,0}\exp\left(c^{2}\lambda^{2}(2+m\alpha)^{2}|\mathbb{G}_{r-2}^{*}|\right)\right)\right]$$

$$\leq \mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}}{2}\sum_{q=0}^{r-2}(\alpha^{2}m^{2})^{r-q-1}|\mathbb{G}_{q}^{*}|\right)\times\exp\left(c^{2}\lambda^{2}(2+m\alpha)^{2}|\mathbb{G}_{r-2}^{*}|\right)\right].$$

Iterating this procedure, studious but straightforward computations lead us to

$$\mathbb{E}\left[\exp\left(\frac{c^{2}\lambda^{2}(2+m\alpha)^{2}}{2}\sum_{q=0}^{r-1}\left(\alpha^{2}m^{2}\right)^{r-q-1}|\mathbb{G}_{q}^{*}|\right)\right] \le \exp\left(c^{2}\lambda^{2}(2+m\alpha)^{2}\left(2^{r-2}+2^{r-3}(m\alpha)^{2}+\ldots+(m\alpha)^{2(r-2)}+\frac{(m\alpha)^{2(r-1)}}{2}\right)\right).$$

From the foregoing and from (B.1), we deduce that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \begin{cases} \exp\left(-\lambda\delta m^r + \frac{c^2\lambda^2(2+m\alpha)^2\left(2^r - (\alpha^2m^2)^r\right)}{2(2-\alpha^2m^2)}\right) \times \exp\left(\lambda c(\alpha m)^r\right) & \text{if } \alpha^2m^2 \neq 2, \\ \exp\left(-\lambda\delta m^r + c^2\lambda^2(2+\sqrt{2})^2r2^{r-2}\right) \exp\left(\lambda c(\sqrt{2})^r\right) & \text{if } \alpha^2m^2 = 2. \end{cases}$$

Now, the rest divides into four cases. In the sequel  $c_1$  and  $c_2$  will denote positive constants which depend on c, m, and  $\alpha$  and may vary from line to line.

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

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(c\lambda\right) \exp\left(-\lambda\delta m^r + \lambda^2 c_1 2^r\right).$$

Taking  $\lambda = (\delta m^r)/(2^{r+1}c_1)$ , we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(c_1\delta\right) \exp\left(-\delta^2 c_1\left(\frac{m^2}{2}\right)^r\right).$$

• If  $1 < m\alpha < \sqrt{2}$ , then, since  $2^r - (\alpha^2 m^2)^r < 2^r$ , we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta m^r + \lambda^2 c_1 2^r\right) \exp\left(\lambda c(m\alpha)^r\right).$$

Taking  $\lambda = (\delta m^r)/(2^{r+1}c_1)$ , we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-c_2\delta(m^2/2)^r(\delta - 2c\alpha^r)\right).$$

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

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-c_2\delta^2(m^2/2)^r\right).$$

• If  $m\alpha = \sqrt{2}$ , then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta m^r + \lambda^2 c_1 r 2^{r-2}\right) \exp\left(\lambda c \left(\sqrt{2}\right)^r\right)$$

Taking  $\lambda = (\delta m^r)/(c_1 r 2^{r-1})$ , we have for all  $r > \log(\delta/4c)/\log(\sqrt{2}/m)$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-c_2\delta^2(1/r)(m^2/2)^r\right).$$

• If  $m\alpha > \sqrt{2}$ , then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta m^r + \lambda^2 c_1(m^2\alpha^2)^r\right) \exp\left(\lambda c(m\alpha)^r\right).$$

Taking  $\lambda = \delta/(2c_1(m\alpha^2)^r)$ , we have for all  $r > \log(\delta/4c)/\log \alpha$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-c_3\delta^2\alpha^{-2r}\right).$$

This ends the proof for  $\mathbb{H}_r = \mathbb{G}_r$ .

**Step 2.** Let us look at  $\widetilde{M}_{\mathbb{T}^*_r}(f)$ . By Chernoff's inequality, we have for all  $\delta > 0$  and for all  $\lambda > 0$ 

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_{r}^{*}}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_{r}\right) \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r}^{*}}f(X_{i})\right)\right].$$
(B.4)

Expectation which appears in the right hand of (B.4) can be written as

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r}^{*}}f(X_{i})\right)\right] = \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-2}^{*}}f(X_{i})\right)\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^{*}}(f+mQf)(X_{i})\right)\right] \times \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^{*}}\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^{*}\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^{*}\}}-mQf(X_{i})\right)\right)\right|\mathcal{F}_{r-1}\right]\right].$$
(B.5)

Iterating this procedure and using the same computations as those made in Step 1, we are led to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r}^{*}}f(X_{i})\right)\right] \leq \exp\left(\frac{c^{2}(2+m\alpha)^{2}\lambda^{2}}{2}\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(m\alpha)^{k}\right)^{2}2^{r-q}\right)$$
$$\times \mathbb{E}\left[\exp\left(\lambda\sum_{q=0}^{r}m^{q}Q^{q}f(X_{1})\right)\right]$$
$$\leq \exp\left(\frac{c^{2}(2+m\alpha)^{2}\lambda^{2}}{2}\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(m\alpha)^{k}\right)^{2}2^{r-q}\right)\exp\left(\lambda c\sum_{q=0}^{r}(m\alpha)^{q}\right),$$

where the last inequality was obtained using hypothesis (H3). In the sequel,  $c_0$ ,  $c_1$  and  $c_2$  will denote some positive constants which depend on  $\alpha$ , m, and c. They may differ from one line to another. For  $m\alpha \neq 1$  and  $m\alpha \neq \sqrt{2}$ , we deduce from the foregoing and from (B.4) that

$$\begin{split} \mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) &\leq \exp\left(-\lambda \delta t_r + \frac{c^2(2+m\alpha)^2 \lambda^2}{2(m\alpha-1)^2} \left((2^r-1) + \frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)\right) \\ &\qquad \times \exp\left(\frac{\lambda c(1-(m\alpha)^{r+1})}{1-m\alpha}\right). \end{split}$$

Taking 
$$\lambda = \frac{\delta t_r (m\alpha - 1)^2}{c^2 (2 + m\alpha)^2 \left( (2^r - 1) + \frac{(m\alpha)^2 (2^r - (m^2 \alpha^2)^r)}{2 - (m\alpha)^2} \right)}$$
, we are led to  
 $\mathbb{P}\left( \widetilde{M}_{\mathbb{T}_r^*}(f) > \delta \right) \le \exp\left( -\frac{\delta^2 (1 - m\alpha)^2 t_r^2}{2c^2 (2 + m\alpha)^2 \left( 2^r - 1 + \frac{(m\alpha)^2 (2^r - (m^2 \alpha^2)^r)}{2 - (m\alpha)^2} \right)} \right)$ 

$$\times \exp\left( \frac{\delta (1 - m\alpha)^2 t_r}{c(2 + m\alpha)^2 \left( 2^r - 1 + \frac{(m\alpha)^2 (2^r - (m^2 \alpha^2)^r)}{2 - (m\alpha)^2} \right)} \times \frac{1 - (m\alpha)^{r+1}}{1 - m\alpha} \right).$$

Now, the rest of the proof divides into five cases.

• If  $m\alpha < 1$ , then, for all  $r \in \mathbb{N}$ ,  $(m\alpha)^{r+1} - 1 \le m\alpha - 1$  and  $2^r - (m\alpha)^{2r} < 2^r$ . We then deduce that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(c_2\delta\right) \exp\left(-c_2\delta^2(m^2/2)^{r+1}\right).$$

• If  $1 < m\alpha < \sqrt{2}$ , then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-c_1\delta^2(m^2/2)^{r+1}\right)\exp\left(c_2\delta\frac{(m^2\alpha)^{r+1}-1}{2^r(m\alpha-1)}\right)$$
$$\le \exp\left(-\delta c_2(m^2/2)^{r+1}(\delta-c_0\alpha^{r+1})\right).$$

Now, for all  $r \in \mathbb{N}$  such that  $r+1 > \log(\delta/2c_0)/\log(\alpha)$ , we have  $\delta - c_0\alpha^{r+1} > \delta/2$ , in such a way that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\delta^2 c_2 (m^2/2)^{r+1}\right).$$

• If  $m\alpha > \sqrt{2}$ , then for all  $r \in \mathbb{N}$ ,  $(m^2 \alpha^2)^r > 2^r$ . We then have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-c_2\delta\alpha^{-2r}(\delta - c_0\alpha^{r+1})\right).$$

Now for all  $r \in \mathbb{N}$  such that  $r+1 > \log(\delta/2c_0)/\log(\alpha)$ , we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\frac{c_2\delta^2}{\alpha^{2r}}\right).$$

- If  $m\alpha = 1$ , then
- $\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta t_r + c_1 2^r \lambda^2\right) \exp\left(\lambda c(r+1)\right)$

Taking  $\lambda = \delta t_r / c_1 2^{r+1}$ , we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(c_1 \delta \frac{(r+1)t_r}{2^{r+1}}\right) \exp\left(-c_2 \delta^2 (m^2/2)^{r+1}\right).$$

• If  $m\alpha = \sqrt{2}$ , then

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta t_r + \lambda^2 c_1(r+1)2^r\right) \exp\left(\lambda c_1(\sqrt{2})^{r+1}\right).$$

Taking  $\lambda = \delta t_r / (2c_1(r+1)2^r)$ , we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\frac{c_2\delta}{r+1}\left(\frac{m^2}{2}\right)^{r+1}\left(\delta - c_0\left(\frac{\sqrt{2}}{m}\right)^{r+1}\right)\right).$$

Now, for all  $r \in \mathbb{N}$  such that  $r+1 > \log(\delta/2c_0)/\log(\sqrt{2}/m)$ , we get

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\frac{c_2\delta^2}{r+1}\left(\frac{m^2}{2}\right)^{r+1}\right)$$

This ends the proof for  $\mathbb{H}_r = \mathbb{T}_r$ .

# B.2. Proof of Theorem 3.3

The proof follows the same ideas that the proof of Corollary 2.6 and uses the computations of the proof of Theorem 3.3 instead of Theorem 2.5.

## B.3. Proof of Theorem 3.4

The proof follows the same ideas that the proof of Theorem 2.8 and uses Theorems 3.2 and 3.3 instead of Theorem 2.5 and Corollary 2.6.

#### B.4. Proof of Theorem 3.5

Let  $f \in \mathcal{B}_b(S^3)$ .

Step 1. Let us first deal with  $\widetilde{M}_{\mathbb{G}_r^*}(f)$ . Assume that  $(\mu, P^*f) = 0$ . By Chernoff's inequality, we have for all  $\delta > 0$  and for all  $\lambda > 0$ ,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \le \exp\left(-\lambda \delta m^r\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(\Delta_i)\right)\right]$$

Conditioning by  $\mathcal{F}_r$ , and using, conditional independence of triplets  $\{\Delta_i, i \in \mathbb{G}_r\}$  with respect to  $\mathcal{F}_r$  and Azuma–Bennett–Hoeffding's inequality, we obtain

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}f(\Delta_{i})\right)\right] = \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}P^{*}f(X_{i})\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}(f(\Delta_{i})-P^{*}f(X_{i}))\right)\Big|\mathcal{F}_{r}\right]\right]$$
$$= \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}P^{*}f(X_{i})\right)\prod_{i\in\mathbb{G}_{r}^{*}}\mathbb{E}\left[\exp\left(\lambda(f(\Delta_{i})-P^{*}f(X_{i}))\right)\Big|\mathcal{F}_{r}\right]\right]$$
$$\leq \exp\left(2\lambda^{2}\|f\|_{\infty}c_{1}\|\mathbb{G}_{r}\|\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}P^{*}f(X_{i})\right)\right]. \quad (B.6)$$

We control the last expectation as in the Step 1 of the proof of Theorem 3.2, apply to  $P^*f$ . Next, we get the result discussing as in the proof of Theorem 3.2.

If  $(\mu, P^*f) \neq 0$ , we set  $g = f - (\mu, P^*f)$ . Then, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) - (\mu, P^*f)W > \delta\right) \le \mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \delta/2|(\mu, P^*f)|\right).$$
(B.7)

The first term of the right hand of (B.7) can be bounded as previously since  $(\mu, P^*g) = 0$ . The second term can be bounded as in **Part 1** of the proof of Corollary 2.6. This ends the proof for  $\widetilde{M}_{\mathbb{G}^*_r}(f)$ .

**Step 2.** Let us now treat  $\widetilde{M}_{\mathbb{T}_r^*}(f)$ . First, we assume that  $(\mu, P^*f) = 0$ . For all  $\delta > 0$ , we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \mathbb{P}\left(\frac{1}{t_r} \sum_{i \in \mathbb{T}_r^*} (f(\Delta_i) - P^*f(X_i)) > \delta/2\right) + \mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(P^*f) > \delta/2\right)$$

By Chernoff's inequality, we have for all  $\lambda > 0$ ,

$$\mathbb{P}\left(\frac{1}{t_r}\sum_{i\in\mathbb{T}_r^*}(f(\Delta_i) - P^*f(X_i)) > \delta/2\right) \le \exp\left(-\frac{\lambda\delta t_r}{2}\right) \times \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*}(f(\Delta_i) - P^*f(X_i))\right)\right]$$

Conditioning successively with respect to  $(\mathcal{F}_q)_{0 \leq q \leq r}$ , using conditional independence of triplets  $\{\Delta_i, i \in \mathbb{G}_q\}$  with respect to  $\mathcal{F}_q$  and applying successively Azuma–Bennett–Hoeffding's inequality and the fact that  $|\mathbb{G}_q^*| \leq |\mathbb{G}_q|$  for all  $q \in \{0, \ldots, r\}$ , we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r}^{*}}(f(\Delta_{i})-P^{*}f(X_{i}))\right)\right] \\
=\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-1}^{*}}(f(\Delta_{i})-P^{*}f(X_{i}))\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r}^{*}}(f(\Delta_{i})-P^{*}f(X_{i}))\right)\left|\mathcal{F}_{r}\right]\right]\right] \\
=\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-1}^{*}}(f(\Delta_{i})-P^{*}f(X_{i}))\right)\prod_{i\in\mathbb{G}_{r}^{*}}\mathbb{E}\left[\exp\left(\lambda(f(\Delta_{i})-P^{*}f(X_{i}))\right)\left|\mathcal{F}_{r}\right]\right] \\
\leq\exp\left(2\lambda^{2}||f||_{\infty}^{2}|\mathbb{G}_{r}|\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-1}^{*}}(f(\Delta_{i})-P^{*}f(X_{i}))\right)\right] \\
\vdots \\
\leq\exp\left(2\lambda^{2}||f||_{\infty}^{2}|\mathbb{T}_{r}|\right).$$

Next, optimizing on  $\lambda$ , we obtain

$$\mathbb{P}\left(\frac{1}{t_r}\sum_{i\in\mathbb{T}_r^*}(f(\Delta_i)-P^*f(X_i))>\delta/2\right)\leq \exp\left(-c_1\delta^2\left(\frac{m^2}{2}\right)^{r+1}\right),$$

for some positive constant  $c_1$ . The term  $\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(P^*f) > \delta/2\right)$  can be bounded as in the proof of Theorem 3.2, and this ends the proof when  $(\mu, P^*f) = 0$ . On the other hand, if  $(\mu, P^*f) \neq 0$ , we have

$$\widetilde{M}_{\mathbb{T}_r^*}(f) - (\mu, P^*f)W = \widetilde{M}_{\mathbb{T}_r^*}(g) + \left(\frac{|\mathbb{T}_r^*|}{t_r} - W\right)(\mu, P^*f).$$

We then proceed as for (B.7), and this ends the proof for  $\widetilde{M}_{\mathbb{T}^*_r}(f)$ .

**Step 3.** Eventually, we bound  $\mathbb{P}(\overline{M}_{\mathbb{H}_r^*}(f) > \delta - (\mu, P^*f))$ , using Steps 1 and 2, as in the proof of Theorem 3.4.

#### B.5. Proof of Proposition 4.4

The proof follows the same ideas that the proof of Proposition 4.1 and uses Theorem 3.5 apply to the functions  $g_1, g_2, h_1$  and  $h_2$ .

# Appendix C. Proof of Proposition 2.1

The proof divides into two parts.

# Part 1. Computation of $\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{G}_r^*}(f)\right)^4\right]$ .

We have

$$\mathbb{E}\left[\left(M_{\mathbb{G}_{r}^{*}}(f)\right)^{4}\right] = \mathbb{E}\left[\left(\sum_{i\in\mathbb{G}_{r}}f(X_{i})\mathbf{1}_{\{i\in\mathbb{T}^{*}\}}\right)^{4}\right]$$

$$= \mathbb{E}\left[\sum_{i\in\mathbb{G}_{r}}f^{4}(X_{i})\mathbf{1}_{\{i\in\mathbb{T}^{*}\}}\right] + 3\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_{r}^{2}, i\neq j}f^{2}(X_{i})f^{2}(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right]$$

$$+ 4\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_{r}^{2}, i\neq j}f^{3}(X_{i})f(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right]$$

$$+ 6\mathbb{E}\left[\sum_{(i,j,k)\in\mathbb{G}_{r}^{3}, i\neq j\neq k}f^{2}(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right]$$

$$+ \mathbb{E}\left[\sum_{(i,j,k,l)\in\mathbb{G}_{r}^{4}, i\neq j\neq k\neq l}f(X_{i})f(X_{j})f(X_{k})f(X_{l})\mathbf{1}_{\{(i,j,k,l)\in\mathbb{T}^{*4}\}}\right].$$
(C.1)

We are now going to compute each term which appears in (C.1). In the sequel, we will use intensively, with a slight modification, the calculations made by Delmas and Marsalle [15] in order to compute the expectations related to two indices. For simplification, we will use the following notations for two functions  $f, g \in F$  and for  $k \in \mathbb{N}$ 

$$Q^{k}_{\otimes}f = Q^{k}f \otimes Q^{k}f \quad \text{and} \quad Q^{k}_{\otimes}(f,g) = Q^{k}f \otimes Q^{k}g.$$
(C.2)
(a) Computation of  $\mathbb{E}\left[\sum_{i \in \mathbb{G}_{r}} f^{4}(X_{i})\mathbf{1}_{\{i \in \mathbb{T}^{*}\}}\right].$ 

From [15] Lemma 2.1, we have

$$\mathbb{E}\left[\sum_{i\in\mathbb{G}_r}f^4(X_i)\mathbf{1}_{\{i\in\mathbb{T}^*\}}\right] = m^r\nu Q^r f^4.$$

(b) Computation of  $3\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^2, i\neq j} f^2(X_i)f^2(X_j)\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right].$ 

We have, using the calculations made by Delmas and Marsalle [15]:

$$3\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^2, i\neq j} f^2(X_i)f^2(X_j)\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right] = 6\sum_{q=0}^{r-1} m^{2r-q-2}\nu Q^q P(Q^{r-q-1}_{\otimes}f^2).$$

(c) Computation of  $4\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^2, i\neq j} f^3(X_i)f(X_j)\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right]$ .

Once again, using the calculations made by Delmas and Marsalle [15] we get:

$$4\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_{r}^{2},i\neq j}f^{3}(X_{i})f(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right] = 4\sum_{q=0}^{r-1}m^{2r-q-2}\nu Q^{q}P\left(Q_{\otimes}^{r-q-1}(f^{3},f)\right) + 4\sum_{q=0}^{r-1}m^{2r-q-2}\nu Q^{q}P\left(Q_{\otimes}^{r-q-1}(f,f^{3})\right)$$

(d) Computation of 
$$6\mathbb{E}\left[\sum_{(i,j,k)\in\mathbb{G}_r^3, i\neq j\neq k} f^2(X_i)f(X_j)f(X_k)\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right]$$

For  $(i, j, k) \in \mathbb{G}_r^3$  such that  $i \neq j \neq k$ , let  $i \wedge j \wedge k$  the most recent common ancestor of i, j and k. We have

$$\mathbb{E}\left[\sum_{(i,j,k)\in\mathbb{G}_{r}^{3},i\neq j\neq k}f^{2}(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right]$$
$$=\mathbb{E}\left[\sum_{q=0}^{r-2}\sum_{p\in\mathbb{G}_{q}}\sum_{(i,j,k)\in\mathbb{G}_{r}^{3},i\wedge j\wedge k=p}f^{2}(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right].$$

If  $i \wedge j \wedge k = p$ , then conditionally on the states  $(X_{2p}, X_{2p+1})$  of the daughters of  $i \wedge j \wedge k$ , we have one of the following cases:

- X<sub>i</sub> is independent of {X<sub>j</sub>, X<sub>k</sub>};
  X<sub>j</sub> is independent of {X<sub>i</sub>, X<sub>k</sub>};
  X<sub>k</sub> is independent of {X<sub>i</sub>, X<sub>j</sub>}.

Using the Markov property of  $(X_i, i \in \mathbb{T}^*)$  and of the GW process, we then have

$$\begin{split} & \mathbb{E}\left[\sum_{\substack{(i,j),k)\in\mathbb{G}_{r}^{2}\\i\neq j\neq k}}f^{2}(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right] \\ &= \sum_{q=0}^{r-2}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\left[\sum_{\substack{i\in\mathbb{G}_{r-q-1}\\i\in\mathbb{G}_{r-q-1}}}\mathbb{E}_{X_{2p}}\left[f^{2}(X_{i})\mathbf{1}_{\{i\in\mathbb{T}^{*},2p\in\mathbb{T}^{*}\}}\right]\sum_{(j,k)\in\mathbb{G}_{r-q-1}^{2};j\neq k}\mathbb{E}_{X_{2p+1}}\left[f(X_{j})f(X_{k})\mathbf{1}_{\{(j,k)\in\mathbb{T}^{*2},2p+1\in\mathbb{T}^{*}\}}\right]\right] \\ &+ \sum_{q=0}^{r-2}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\left[\sum_{\substack{i\in\mathbb{G}_{r-q-1}\\i\in\mathbb{G}_{r-q-1}}}\mathbb{E}_{X_{2p+1}}\left[f^{2}(X_{i})\mathbf{1}_{\{i\in\mathbb{T}^{*},2p+1\in\mathbb{T}^{*}\}}\right]\sum_{(j,k)\in\mathbb{G}_{r-q-1}^{2};j\neq k}\mathbb{E}_{X_{2p}}\left[f(X_{j})f(X_{k})\mathbf{1}_{\{(j,k)\in\mathbb{T}^{*2},2p\in\mathbb{T}^{*}\}}\right]\right] \\ &+ 2\sum_{q=0}^{r-2}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_{r-q-1}^{2};i\neq j}\mathbb{E}_{X_{2p}}\left[f^{2}(X_{i})f(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2},2p\in\mathbb{T}^{*}\}}\right]\sum_{k\in\mathbb{G}_{r-q-1}}\mathbb{E}_{X_{2p+1}}\left[f(X_{k})\mathbf{1}_{\{k\in\mathbb{T}^{*},2p\in\mathbb{T}^{*}\}}\right] \\ &+ 2\sum_{q=0}^{r-2}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_{r-q-1}^{2};i\neq j}\mathbb{E}_{X_{2p+1}}\left[f^{2}(X_{i})f(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2},2p+1\in\mathbb{T}^{*}\}}\right]\sum_{k\in\mathbb{G}_{r-q-1}}\mathbb{E}_{X_{2p}}\left[f(X_{k})\mathbf{1}_{\{k\in\mathbb{T}^{*},2p\in\mathbb{T}^{*}\}}\right] \right]. \end{split}$$

Decomposing the sum  $\sum_{(j,k)\in\mathbb{G}^2_{r-q-1};j\neq k}\mathbb{E}_{X_{2p+1}}\left[f(X_j)f(X_k)\mathbf{1}_{\{(j,k)\in\mathbb{T}^{*2},2p+1\in\mathbb{T}^*\}}\right]$  according to the generation of  $s_1 = j \wedge k$  and using the Markov property of  $(X_i, i \in \mathbb{T}^*)$  and of the GW process, as well as the Lemma 2.1 of [15], we are led to

$$\begin{split} &\sum_{(j,k)\in\mathbb{G}_{r-q-1}^{2};j\neq k} \mathbb{E}_{X_{2p+1}} \left[ f(X_{j})f(X_{k})\mathbf{1}_{\{(j,k)\in\mathbb{T}^{*2},2p+1\in\mathbb{T}^{*}\}} \right] \\ &= \sum_{s_{1}=0}^{r-q-2} \sum_{s_{2}\in\mathbb{G}_{s_{1}}} \sum_{(j,k)\in\mathbb{G}_{r-q-1}^{2};j\wedge k=s_{2}} \mathbb{E}_{X_{2p+1}} \left[ f(X_{j})f(X_{k})\mathbf{1}_{\{(j,k)\in\mathbb{T}^{*2},2p+1\in\mathbb{T}^{*}\}} \right] \\ &= 2\sum_{s_{1}=0}^{r-q-2} \sum_{s_{2}\in\mathbb{G}_{s_{1}}} \mathbb{E}_{X_{2p+1}} \left[ m^{2(r-q-s_{1}-2)}\mathbb{E}_{X_{2s_{2}}} \left[ f(Y_{r-q-s_{1}-2}) \right] \times \mathbb{E}_{X_{2s_{2}+1}} \left[ f(Y_{r-q-s_{1}-2}) \right] \mathbf{1}_{\{(2s_{2},2s_{2}+1,2p+1)\in\mathbb{T}^{*}\}} \right] \\ &= 2\sum_{s_{1}=0}^{r-q-2} \sum_{s_{2}\in\mathbb{G}_{s_{1}}} m^{2(r-q-s_{1}-2)}\mathbb{E}_{X_{2p+1}} \left[ P\left(Q_{\otimes}^{r-q-s_{1}-2}f(X_{s_{2}})\right) \mathbf{1}_{\{s_{2}\in\mathbb{T}^{*},2p+1\in\mathbb{T}^{*}\}} \right] \\ &= 2\sum_{s_{1}=0}^{r-q-2} m^{2r-2q-s_{1}-4}Q^{s_{1}}P\left(Q_{\otimes}^{r-q-s_{1}-2}f\right) (X_{2p+1}) \mathbf{1}_{\{2p+1\in\mathbb{T}^{*}\}}. \end{split}$$

From the previous calculations, and using Lemma 2.1 of [15] and Markov property of  $(X_i, i \in \mathbb{T}^*)$  and of the GW process, we obtain

$$\begin{split} &\sum_{q=0}^{r-2} \sum_{p \in \mathbb{G}_q} \mathbb{E} \left[ \sum_{i \in \mathbb{G}_{r-q-1}} \mathbb{E}_{X_{2p}} \left[ f^2(X_i) \mathbf{1}_{\{i \in \mathbb{T}^*, 2p \in \mathbb{T}^*\}} \right] \sum_{(j,k) \in \mathbb{G}_{r-q-1}^2; j \neq k} \mathbb{E}_{X_{2p+1}} \left[ f(X_j) f(X_k) \mathbf{1}_{\{(j,k) \in \mathbb{T}^{*2}, 2p+1 \in \mathbb{T}^*\}} \right] \right] \\ &= \sum_{q=0}^{r-2} \sum_{p \in \mathbb{G}_q} \mathbb{E} \left[ 2m^{r-q-1} Q^{r-q-1} f^2(X_{2p}) \mathbf{1}_{\{2p \in \mathbb{T}^*\}} \right] \\ &\times \sum_{s_1=0}^{r-q-2} m^{2r-2q-s_1-4} Q^{s_1} P\left(Q_{\otimes}^{r-q-s_1-2} f\right) (X_{2p+1}) \mathbf{1}_{\{2p+1 \in \mathbb{T}^*\}} \right] \\ &= \sum_{q=0}^{r-2} \sum_{s_1=0}^{r-q-2} 2m^{3r-3q-s_1-5} \sum_{p \in \mathbb{G}_q} \mathbb{E} \left[ P\left(Q^{r-q-1} f^2 \otimes Q^{s_1} P\left(Q_{\otimes}^{r-q-s_1-2} f\right)\right) (X_p) \mathbf{1}_{\{p \in \mathbb{T}^*\}} \right] \\ &= \sum_{q=0}^{r-2} \sum_{s_1=0}^{r-q-2} 2m^{3r-2q-s_1-5} \nu Q^q P\left(Q^{r-q-1} f^2 \otimes Q^{s_1} P\left(Q_{\otimes}^{r-q-s_1-2} f\right)\right). \end{split}$$

In the same way as before, we have

$$\sum_{q=0}^{r-2} \sum_{p \in \mathbb{G}_q} \mathbb{E} \left[ \sum_{i \in \mathbb{G}_{r-q-1}} \mathbb{E}_{X_{2p+1}} \left[ f^2(X_i) \mathbf{1}_{\{i \in \mathbb{T}^*, 2p+1 \in \mathbb{T}^*\}} \right] \sum_{(j,k) \in \mathbb{G}_{r-q-1}^2; j \neq k} \mathbb{E}_{X_{2p}} \left[ f(X_j) f(X_k) \mathbf{1}_{\{(j,k) \in \mathbb{T}^{*2}, 2p \in \mathbb{T}^*\}} \right] \right]$$
$$= \sum_{q=0}^{r-2} \sum_{s_1=0}^{r-q-2} 2m^{3r-2q-s_1-5} \nu Q^q P \left( Q^{s_1} P \left( Q^{r-q-s_1-2}_{\otimes} f \right) \otimes Q^{r-q-1} f^2 \right).$$

Once again, decomposing the sum  $\sum_{(i,j)\in\mathbb{G}_{r-q-1}^2;i\neq j}\mathbb{E}_{X_{2p}}\left[f^2(X_i)f(X_j)\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2},2p\in\mathbb{T}^*\}}\right]$  according to the generation of  $s_1 = j \wedge k$  and using the Markov property of  $(X_i, i \in \mathbb{T}^*)$  and of the GW process, as well as the Lemma 2.1 of [15], we are led to

$$\sum_{\substack{(i,j)\in\mathbb{G}_{r-q-1}^{2};i\neq j}} \mathbb{E}_{X_{2p}} \left[ f^{2}(X_{i})f(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2},2p\in\mathbb{T}^{*}\}} \right]$$
  
= 
$$\sum_{s_{1}=0}^{r-q-2} m^{2r-2q-s_{1}-4}Q^{s_{1}}P\left(Q_{\otimes}^{r-q-s_{1}-2}(f^{2},f)\right)(X_{2p})\mathbf{1}_{\{2p\in\mathbb{T}^{*}\}}$$
  
+ 
$$\sum_{s_{1}=0}^{r-q-2} m^{2r-2q-s_{1}-4}Q^{s_{1}}P\left(Q_{\otimes}^{r-q-s_{1}-2}(f,f^{2})\right)(X_{2p})\mathbf{1}_{\{2p\in\mathbb{T}^{*}\}}.$$

Where the notation  $Q_{\otimes}^{r-q-s_1-2}$  is given in (C.2). We thus obtain

$$\begin{split} &\sum_{q=0}^{r-2} \sum_{p \in \mathbb{G}_q} \mathbb{E} \left[ \sum_{(i,j) \in \mathbb{G}_{r-q-1}^2; i \neq j} \mathbb{E}_{X_{2p}} \left[ f^2(X_i) f(X_j) \mathbf{1}_{\{(i,j) \in \mathbb{T}^{*2}, 2p \in \mathbb{T}^*\}} \right] \sum_{k \in \mathbb{G}_{r-q-1}} \mathbb{E}_{X_{2p+1}} \left[ f(X_k) \mathbf{1}_{\{k \in \mathbb{T}^*, 2p+1 \in \mathbb{T}^*\}} \right] \right] \\ &= \sum_{q=0}^{r-2} \sum_{s_1=0}^{r-q-2} m^{3r-3q-s_1-5} \sum_{p \in \mathbb{G}_q} \mathbb{E} \left[ \left( Q^{s_1} P\left( Q_{\otimes}^{r-q-s_1-2}(f^2, f) \right) (X_{2p} \right) Q^{r-q-1} f\left( X_{2p+1} \right) \right. \\ &+ \left. Q^{s_1} P\left( Q_{\otimes}^{r-q-s_1-2}(f, f^2) \right) (X_{2p}) Q^{r-q-1} f\left( X_{2p+1} \right) \right) \mathbf{1}_{\{(2p, 2p+1) \in \mathbb{T}^{*2}\}} \right] \\ &= \sum_{q=0}^{r-2} \sum_{s_1=0}^{r-q-2} m^{3r-2q-s_1-5} \nu Q^q \left( P\left( Q^{s_1} P\left( Q_{\otimes}^{r-q-s_1-2}(f^2, f) \right) \otimes Q^{r-q-1} f \right) \right. \\ &+ \left. P\left( Q^{s_1} P\left( Q_{\otimes}^{r-q-s_1-2}(f, f^2) \right) \otimes Q^{r-q-1} f \right) \right). \end{split}$$

In the same way, we have

$$\begin{split} &\sum_{q=0}^{r-2} \sum_{p \in \mathbb{G}_{q}} \mathbb{E} \left[ \sum_{(i,j) \in \mathbb{G}_{r-q-1}^{2}; i \neq j} \mathbb{E}_{X_{2p+1}} \left[ f^{2}(X_{i})f(X_{j})\mathbf{1}_{\{(i,j) \in \mathbb{T}^{*2}, 2p+1 \in \mathbb{T}^{*}\}} \right] \sum_{k \in \mathbb{G}_{r-q-1}} \mathbb{E}_{X_{2p}} \left[ f(X_{k})\mathbf{1}_{\{k \in \mathbb{T}^{*}, 2p \in \mathbb{T}^{*}\}} \right] \right] \\ &= \sum_{q=0}^{r-2} \sum_{s_{1}=0}^{r-q-2} m^{3r-2q-s_{1}-5} \nu Q^{q} \left( P\left(Q^{r-q-1}f \otimes Q^{s_{1}}P\left(Q^{r-q-s_{1}-2}(f^{2}, f)\right)\right) \right) \\ &+ P\left(Q^{r-q-1}f \otimes Q^{s_{1}}P\left(Q^{r-q-s_{1}-2}(f, f^{2})\right)\right) \right). \end{split}$$

This ends the computation of  $6\mathbb{E}\left[\sum_{(i,j,k)\in\mathbb{G}_r^3, i\neq j\neq k} f^2(X_i)f(X_j)f(X_k)\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right].$ 

(e) Computation of 
$$\mathbb{E}\left[\sum_{(i,j,k,l)\in\mathbb{G}_r^4, i\neq j\neq k\neq l} f(X_i)f(X_j)f(X_k)f(X_l)\mathbf{1}_{\{(i,j,k,l)\in\mathbb{T}^{*4}\}}\right].$$

Let  $i \wedge j \wedge k \wedge l$  the most recent common ancestor of i, j, k and l. We have

$$\mathbb{E}\left[\sum_{\substack{(i,j,k,l)\in\mathbb{G}_r^4; i\neq j\neq k\neq l}} f(X_i)f(X_j)f(X_k)f(X_l)\mathbf{1}_{\{(i,j,k,l)\in\mathbb{T}^{*4}\}}\right]$$
$$=\mathbb{E}\left[\sum_{\substack{q=0\\p\in\mathbb{G}_q}}\sum_{\substack{(i,j,k,l)\in\mathbb{G}_r^4; i\neq j\neq k\neq l\\i\wedge j\wedge k\wedge l=p}} f(X_i)f(X_j)f(X_k)f(X_l)\mathbf{1}_{\{(i,j,k,l)\in\mathbb{T}^{*4}\}}\right].$$

Now, for  $i \wedge j \wedge k \wedge l = p$ , two cases may happen:

- 2p is the ancestor of two individuals and 2p + 1 is the ancestor of two individuals;
- 2p is the ancestor of three individuals and 2p + 1 is the ancestor of one individual and vice versa.

When 2p is the ancestor of two individuals, i and j say, and 2p + 1 the ancestor of k and l, then conditionally on  $(X_{2p}, X_{2p+1})$ ,  $(X_i, X_j)$  and  $(X_k, X_l)$  are independent. The same thing for  $(X_i, X_j, X_k)$  and  $X_l$  if 2p is the common ancestor of i, j and k and 2p + 1 the ancestor of l. Then, conditioning on  $(X_{2p}, X_{2p+1})$  and applying the Markov property, we are led to

$$\begin{split} & \mathbb{E}\Bigg[\sum_{(i,j,k,l)\in\mathbb{G}_{r}^{4}; i\neq j\neq k\neq l} f(X_{i})f(X_{j})f(X_{k})f(X_{l})\mathbf{1}_{\{(i,j,k,l)\in\mathbb{T}^{*4}\}}\Bigg] \\ &= 6\sum_{q=0}^{r-2}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\Bigg[\sum_{(i,j)\in\mathbb{G}_{r-q-1}^{2}; i\neq j}\mathbb{E}_{X_{2p}}\left[f(X_{i})f(X_{j})\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}, 2p\in\mathbb{T}^{*}\}}\right] \\ & \times\sum_{(k,l)\in\mathbb{G}_{r-q-1}^{2}; k\neq l}\mathbb{E}_{X_{2p+1}}\left[f(X_{k})f(X_{l})\mathbf{1}_{\{(k,l)\in\mathbb{T}^{*2}, 2p+1\in\mathbb{T}^{*}\}}\right]\Bigg] \\ &+ 4\sum_{q=0}^{r-3}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\Bigg[\sum_{(i,j,k)\in\mathbb{G}_{r-q-1}^{3}}\mathbb{E}_{X_{2p}}\left[f(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}, 2p\in\mathbb{T}^{*}\}}\right] \\ & \times\sum_{l\in\mathbb{G}_{r-q-1}}\mathbb{E}_{X_{2p+1}}\left[f(X_{l})\mathbf{1}_{\{l\in\mathbb{T}^{*}, 2p+1\in\mathbb{T}^{*}\}}\right]\Bigg] \\ &+ 4\sum_{q=0}^{r-3}\sum_{p\in\mathbb{G}_{q}}\mathbb{E}\Bigg[\sum_{(i,j,k)\in\mathbb{G}_{r-q-1}^{3}}\mathbb{E}_{X_{2p+1}}\left[f(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}, 2p+1\in\mathbb{T}^{*}\}}\right] \\ & \times\sum_{l\in\mathbb{G}_{r-q-1}}\mathbb{E}_{X_{2p}}\left[f(X_{l})\mathbf{1}_{\{l\in\mathbb{T}^{*}, 2p\in\mathbb{T}^{*}\}}\right]. \end{split}$$

Now, the calculations similar to those done in part (d) led us to

$$\begin{split} \mathbb{E} \left[ \sum_{(i,j,k,l) \in \mathbb{G}_{*}^{4}; i \neq j \neq k \neq l} f(X_{i}) f(X_{j}) f(X_{k}) f(X_{l}) \mathbf{1}_{\{(i,j,k,l) \in \mathbb{T}^{*4}\}} \right] \\ &= 24 \sum_{q=0}^{r-2} \sum_{s_{1},s_{2}=0}^{r-q-2} m^{4r-3q-s_{1}-s_{2}-8} \nu Q^{q} P\left(Q^{s_{1}} P\left(Q^{r-q-s_{1}-2}f\right) \otimes Q^{s_{2}} P\left(Q^{r-q-s_{2}-2}f\right)\right) \\ &+ 24 \sum_{q=0}^{r-3} \sum_{s_{1}=0}^{r-q-3} \sum_{s_{2}=0}^{r-q-s_{1}-3} m^{4r-3q-2s_{1}-s_{2}-9} \nu Q^{q} P\left(Q^{s_{1}} P\left(Q^{s_{2}} P\left(Q^{r-q-s_{1}-2}f\right) \otimes Q^{r-q-1}f\right)\right) \\ &+ 24 \sum_{q=0}^{r-3} \sum_{s_{1}=0}^{r-q-s_{1}-3} \sum_{s_{2}=0}^{r-q-s_{1}-3} m^{4r-3q-2s_{1}-s_{2}-9} \nu Q^{q} P\left(Q^{s_{1}} P\left(Q^{s_{2}} P\left(Q^{r-q-s_{1}-s_{2}-3}f\right)\right) \otimes Q^{r-q-1}f\right) \\ &\otimes Q^{r-q-s_{1}-2}f\right) \otimes Q^{r-q-1}f\right) \\ &+ 24 \sum_{q=0}^{r-3} \sum_{s_{1}=0}^{r-q-s_{1}-3} \sum_{s_{2}=0}^{m^{4r-3q-2s_{1}-s_{2}-9}} \nu Q^{q} P\left(Q^{r-q-1}f \otimes Q^{s_{1}} P\left(Q^{r-q-s_{1}-s_{2}-3}f\right)\right) \\ &\otimes Q^{s_{2}} P\left(Q^{r-q-s_{1}-s_{2}-3}f\right) \\ &\otimes Q^{s_{2}} P\left(Q^{r-q-s_{1}-s_{2}-3}f\right) \\ &\otimes Q^{s_{2}} P\left(Q^{r-q-1}f \otimes Q^{s_{1}} P\left(Q^{s_{2}} P\left(Q^{r-q-s_{1}-s_{2}-3}f\right)\right)\right) \\ &+ 24 \sum_{q=0}^{r-3} \sum_{s_{1}=0}^{r-q-s_{1}-3} m^{4r-3q-2s_{1}-s_{2}-9} \nu Q^{q} P\left(Q^{r-q-1}f \otimes Q^{s_{1}} P\left(Q^{s_{2}} P\left(Q^{r-q-s_{1}-s_{2}-3}f\right)\right)\right) \\ &\otimes Q^{r-q-s_{1}-2}f\right) \\ \\ &\otimes Q^{r$$

Now dividing all the terms of (C.1) by  $m^{4r}$ , we get an explicit expression for  $\mathbb{E}\left|\left(\widetilde{M}_{\mathbb{G}_r^*}(f)\right)^4\right|$ .

#### Part 2. Computation of the rates.

We are now going to give some rate for the different terms that appear in the expression of  $\mathbb{E}\left[\left(\widetilde{M}_{\mathbb{G}_r^*}(f)\right)^4\right]$ .

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

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

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

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  $\alpha$  and m. The constants c and  $c_1$  may vary from one line to another and from one expression to another.

(a) We have

$$\frac{1}{m^{4r}} \mathbb{E}\left[\sum_{i \in \mathbb{G}_r} f^4(X_i) \mathbf{1}_{\{i \in \mathbb{T}^*\}}\right] = \left(\frac{1}{m^3}\right)^r \nu Q^r f^4 \le c \left(\frac{1}{m^3}\right)^r$$

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

(b) We have

$$\frac{3}{m^{4r}} \mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^2, i\neq j} f^2(X_i) f^2(X_j) \mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right] = 6\sum_{q=0}^{r-1} m^{-2r-q-2} \nu Q^q P\left(Q_{\otimes}^{r-q-1} f^2\right)$$
$$\leq 6c \sum_{q=0}^{r-1} m^{-2r-q-2} \leq cc_1 \left(\frac{1}{m^2}\right)^r$$

where from (ii), (iii), (v) and (vi) c is such that  $\nu Q^q P\left(Q^{r-q-1}_{\otimes}f^2\right) < c.$ 

(c) For

$$\frac{4}{m^{4r}}\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^2, i\neq j}f^3(X_i)f(X_j)\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right],\$$

we are going to treat only

$$4\sum_{q=0}^{r-1} m^{-2r-q-2} \nu Q^q P\left(Q_{\otimes}^{r-q-1}(f^3, f)\right).$$

The other terms can be treated in the same way. We have

$$4\sum_{q=0}^{r-1} m^{-2r-q-2} \nu Q^q P\left(Q_{\otimes}^{r-q-1}(f^3, f)\right) \le 4c\sum_{q=0}^{r-1} m^{-2r-q-2} \alpha^{r-q-1},$$

where from (ii), (iii), (v) and (vi) c is such that  $\nu Q^q P\left(Q^{r-q-1}f^3 \otimes g\right) < c$  and from hypothesis **(H1)**, g is such that  $Q^{r-q-1}f \leq \alpha^{r-q-1}g$ .

Now, according to the value of  $\alpha$ , we have

$$\frac{4}{m^{4r}}\mathbb{E}\left[\sum_{(i,j)\in\mathbb{G}_r^2, i\neq j} f^3(X_i)f(X_j)\mathbf{1}_{\{(i,j)\in\mathbb{T}^{*2}\}}\right] = \begin{cases} c_1cr\left(\frac{\alpha}{m^2}\right)^r & \text{if } m\alpha = 1\\ c_1c\left(\left(\frac{\alpha}{m^2}\right)^r + \left(\frac{1}{m^3}\right)^r\right) & \text{if } m\alpha \neq 1. \end{cases}$$

(d) In the same way as before, we have according to the values of  $\alpha$ 

$$\frac{6}{m^{4r}} \mathbb{E}\left[\sum_{(i,j,k)\in\mathbb{G}_{r}^{3}, i\neq j\neq k} f^{2}(X_{i})f(X_{j})f(X_{k})\mathbf{1}_{\{(i,j,k)\in\mathbb{T}^{*3}\}}\right] \\ \leq \begin{cases} c_{1}c\left(\left(\frac{\alpha^{2}}{m}\right)^{r}+\left(\frac{1}{m^{3}}\right)^{r}\right) & \text{if } m\alpha^{2}\neq 1 \text{ and } m\alpha\neq 1 \\ c_{1}c\left(r\left(\frac{1}{m^{3}}\right)^{r}+\left(\frac{1}{m^{2}}\right)^{r}\right) & \text{if } m\alpha=1 \\ c_{1}cr\left(\frac{1}{m^{2}}\right)^{r} & \text{if } m\alpha^{2}=1, \end{cases}$$

where from (ii), (iii), (v) and (vi) c is such that, for  $q \in \{0, \ldots, q-2\}$  and for  $s_1 \in \{0, \ldots, r-q-2\}$ 

$$\max\left\{\nu Q^{q} P\left(Q^{r-q-1}f^{2} \otimes Q^{s_{1}} P\left(Q^{r-q-s_{1}-2}g\right)\right); \\ \nu Q^{q} P\left(Q^{s_{1}} P\left(Q^{r-q-s_{1}-2}(f^{2},g)\right) \otimes Q^{r-q-1}g\right)\right\} < c$$

and from hypothesis (H1), g is such that  $Q^{r-q-1}f \leq \alpha^{r-q-1}g$ .

-

(e) Once again, according to the values of  $\alpha$ , we have

$$\frac{1}{m^{4r}} \mathbb{E} \left[ \sum_{(i,j,k,l) \in \mathbb{G}_r^4, i \neq j \neq k \neq l} f(X_i) f(X_j) f(X_k) f(X_l) \mathbf{1}_{\{(i,j,k,l) \in \mathbb{T}^{*4}\}} \right]$$

$$\leq \begin{cases} c_1 c \left( \alpha^{4r} + \left(\frac{1}{m^3}\right)^r + \left(\frac{\alpha^2}{m^2}\right)^r + \left(\frac{\alpha}{m^2}\right)^r \right) & \text{if } m\alpha \neq 1 \text{ and } m\alpha^2 \neq 1 \\ c_1 c \left( \left(\frac{1}{m^2}\right)^r + r^2 \left(\frac{1}{m^3}\right)^r \right) & \text{if } m\alpha = 1 \\ c_1 c r^2 \left(\frac{1}{m^2}\right)^r & \text{if } m\alpha^2 = 1, \end{cases}$$

where c is defined in the same way as in part (d). This ends the proof of Proposition 2.1.

Now the results of the Proposition 2.1 follow from (a)-(e) of Part 2.

Acknowledgements. The author thanks Pr. Arnaud GUILLIN for all its advice and suggestions during the preparation of this work. I also thanks the anonymous referee for his/her very careful reading and useful suggestions, which have clearly improved both presentation and mathematical rigor of the present paper. Finally, I thanks the Hadamard Mathematics Labex of the "Fondation Mathématiques Jacques Hadamard" for the financial support.

#### References

- [1] K.B. Athreya, Large deviations for branching processes-I. Single type case. Ann. Appl. Probab. 5 (1994) 779-790
- [2] K.B. Athreya and P.E Ney, Branching Process. Springer, Berlin (1972).
- [3] K. Azuma, Weighted sums of certain dependent random variables. Tôhoku Math. J. 19 (1967) 357-367.
- [4] V. Bansaye, J.F. Delmas, L. Marsalle and V.C. Tran, Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. Ann. Appl. Probab. 21 (2011) 2263-2314
- [5] G. Bennett, Probability inequalities for sum of independant random variables. J. Am. Stat. Assoc. 57 (1962) 33-45.
- B. Bercu, and V. Blandin, A Rademacher-Menchov approach for random coefficient bifurcating autoregressive processes. Stochastic Processes Appl. 125 (2015) 1218-1243.
- [7] B. Bercu, B. De Saporta and A. Gégout-Petit, Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic. J. Probab.* 14 (2009) 2492–2526.
- [8] S.V. Bitseki Penda and H. Djellout, Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models. Ann. Inst. Henri Poincaré 50 (2014) 806–844.
- [9] S.V. Bitseki Penda, H. Djellout and A. Guillin, Deviation inequalities, Moderate deviations and some limit theorems for bifurcating Markov chains with application. Ann. Appl. Probab. 24 (2014) 235–291.
- [10] V. Blandin, Asymptotic results for random coefficient bifurcating autoregressive processes. Statistics 48 (2013) 1202–1232.
- [11] R. Cowan and R.G. Staudte, The bifurcating autoregressive model in cell lineage studies. Biometrics 42 (1986) 769–783.
- [12] B. De Saporta, A. Gégout-Petit and L. Marsalle, Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* 5 (2011) 1313–1353.
- [13] B. De Saporta, A. Gégout-Petit and L. Marsalle, Asymmetry tests for Bifurcating Auto-Regressive Processes with missing data. Stat. Probab. Lett. 82 (2012) 1439–1444.
- [14] B. De Saporta, A. Gégout-Petit and L. Marsalle, Random coefficients bifurcating autoregressive processes ESAIM: PS 18 (2014) 365–399.
- [15] J.F. Delmas and L. Marsalle, Detection of cellular aging in Galton–Watson process. Stochastic Process. Appl. 120 (2010) 2495-2519
- [16] J.P. Dion and N.M. Yanev, Limit theorems and estimation theory for branching processes with increasing random number of ancestors. J. Appl. Probab. 34 (1997) 309–327.
- [17] J. Guyon, Limit theorems for bifurcating markov chains. Application to the detection of cellular aging. Ann. Appl. Probab. 17 (2007) 1538–1569.
- [18] J. Guyon, A. Bize, G. Paul, E.J. Stewart, J.F. Delmas and F. Taddéi, Statistical study of cellular aging. Proc. of CEMRACS 2004. ESAIM: Proc. 14 (2005) 100–114.
- [19] W. Hoeffding, Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. 58 (1963) 13-30.
- [20] R.B. Karp and Y. Zhang, Finite branching processes and AND/OR tree evaluation TR. International Computer Science Institute; 93-043, ICSI. Berkeley, Calif. (1994).
- [21] C. McDiarmid, On the method of bounded differences. Electron. Comm. Probab. 11 (2006) 64–77.
- [22] E.J. Stewart, R. Madden, G. Paul and F. Taddéi, Aging and death in an organism that reproduces by morphologically symmetric division. PLoS Biol. 3 (2005) e45.