Bernoulli **23**(4B), 2017, 3213–3242
DOI: 10.3150/16-BEJ843

Transportation and concentration inequalities for bifurcating Markov chains

S. VALÈRE BITSEKI PENDA[1], MIKAEL ESCOBAR-BACH[2] and ARNAUD GUILLIN[3]

[1]*Institut de Mathématiques de Bourgogne (IMB), CNRS UMR 5584, Université Bourgogne Franche-Comté, 9 avenue Alain Savary-BP 47870, 21078 DIJON Cedex – France.*
E-mail: simeon-valere.bitseki-penda@u-bourgogne.fr
[2]*Department of Mathematics and Computer Science, University of Southern Denmark, Campus 55, 5230 Odense M, Denmark. E-mail: escobar@sdu.dk*
[3]*Institut Universitaire de France et Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, 24 avenue des Landais, BP 80026, 63177 Aubière, France.*
E-mail: guillin@math.univ-bpclermont.fr

We investigate the transportation inequality for bifurcating Markov chains which are a class of processes indexed by a regular binary tree. Fitting well models like cell growth when each individual gives birth to exactly two offsprings, we use transportation inequalities to provide useful concentration inequalities. We also study deviation inequalities for the empirical means under relaxed assumptions on the Wasserstein contraction for the Markov kernels. Applications to bifurcating nonlinear autoregressive processes are considered for point-wise estimates of the non-linear autoregressive function.

Keywords: bifurcating Markov chains; deviation inequalities; geometric ergodicity; transportation inequalities; Wasserstein distance

1. Introduction

Roughly speaking, a bifurcating Markov chain (BMC) is a Markov chain indexed by a regular binary tree. Introduced by Guyon [25] in application of the Escherichia coli cellular aging, this class of processes is well adapted for the study of lineage data where individuals give birth to exactly two descendants. Several models of BMC have been recently studied [13,18,25,27,28] with a great interest in cell division topics. There is now an important literature covering asymptotic results for BMC [4,7,13,14,16,25,26]. These limit theorems are particularly useful when applied to the statistics of bifurcating processes, enabling to provide efficient test to assert if the cell aging is different for the two offspring (see [26] for real case study). Of course, they may be considered only in the "ergodic" case, that is, when the law of the random lineage chain has an unique invariant measure. However, this could be unusable in practice since one is often faced to study limited sized data. Thus, natural questions arise about the control of statistics outside the limits. Such control could be reach with deviation inequalities (or concentration inequalities) and have been recently the subject of many studies. We refer to the books of Ledoux [29] and Massart [33] for nice introductions on the subject, developing both i.i.d. and dependent cases with a wide variety of tools (Laplace controls, functional inequalities, Efron–Stein, . . .). It was

1350-7265

one of the goal of Bitseki *et al.* [7] to investigate deviation inequalities for additive functionals of BMC. In their work, one of the main hypothesis is that the Markov chain associated to a random lineage of the population is uniformly ergodic in a geometric point of view. It is clearly a very strong assumption, nearly reducing interesting models to the compact case. The purpose of this paper is to considerably weaken this hypothesis. More specifically, we obtain deviation inequalities for BMC when the auxiliary Markov chain may satisfy some contraction properties in Wasserstein distance and some (uniform) integrability properties. This will be done with the help of transportation cost inequalities and direct Laplace controls. In order to present our results, we may now define properly the model of BMC and transportation inequalities.

1.1. Bifurcating Markov chains

First, we introduce some useful notations. Let $\mathbb{T}$ be a regular binary tree in which each vertex is seen as a positive integer different from 0. For $n \in \mathbb{N}$, let

$$\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}, \qquad \mathbb{T}_n = \bigcup_{k=0}^{n} \mathbb{G}_k,$$

denote respectively the nth column and the first $(n+1)$ columns of the tree. The whole tree is thus defined by

$$\mathbb{T} = \bigcup_{n=0}^{\infty} \mathbb{G}_n.$$

Reversely, any vertex n belongs to the column $\mathbb{G}_{r_n}$ with $r_n = \lfloor \log_2 n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of the real number x. A vertex n represents an individual as well as the ancestor of the individuals $2n$ and $2n+1$. Ones who belong to $2\mathbb{N}$ (resp. $2\mathbb{N}+1$) will be called of type 0 (resp. type 1). The initial individual will be denoted 1 (see Figure 1).

For each individual n, we associate a random variable X_n, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which takes values in a metric space (E, d) endowed with its Borel σ-algebra $\mathcal{E}$. We assume that each pair of random variables (X_{2n}, X_{2n+1}) depends of the past values $\{X_m, m \in \mathbb{T}_{r_n}\}$ only through X_n. In order to describe this dependance, let us introduce the following notion.

***Definition 1.1** (**$\mathbb{T}$-transition probability, see [25]**). We call $\mathbb{T}$-transition probability any mapping $P : E \times \mathcal{E}^2 \to [0, 1]$ such that*:

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

In particular, for all $x, y, z \in E$, $P(x, dy, dz)$ represents the probability that the quantities associated to a children pair (from same mother) are in the neighbourhood of y and z given that the quantity associated to their ancestor is x.

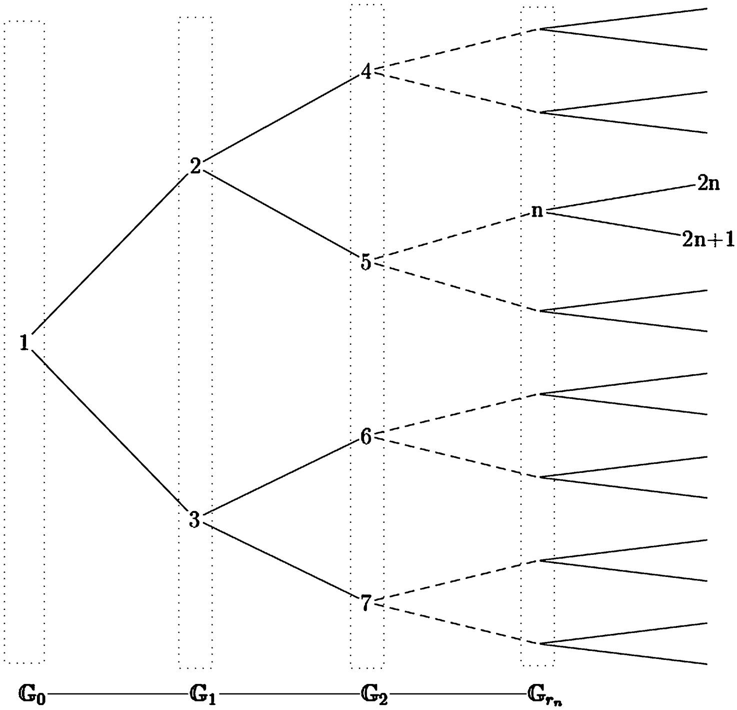

Figure 1. A regular binary tree $\mathbb{T}$.

For a $\mathbb{T}$-transition probability P on $E \times \mathcal{E}^2$, we denote by P_0, P_1 the first and the second marginal of P, that is $P_0(x, A) = P(x, A \times E)$, $P_1(x, A) = P(x, E \times A)$ for all $x \in E$ and $A \in \mathcal{E}$. Then, P_0 (resp. P_1) can be seen as the transition probability associated to individual of type 0 (resp. type 1).

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

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

We are now in position to give a precise definition for a BMC.

Definition 1.2 (BMC, see [25]). *Let $(X_n, n \in \mathbb{T})$ be a family of E-valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$. Let ν be a probability on $(E, \mathcal{E})$ and P be a $\mathbb{T}$-transition probability. We say that $(X_n, n \in \mathbb{T})$ is a $(\mathcal{F}_r)$-bifurcating Markov chain with initial distribution ν and $\mathbb{T}$-transition probability P if:*

- *X_n is $\mathcal{F}_{r_n}$-measurable for all $n \in \mathbb{T}$,*
- *$\mathcal{L}(X_1) = \nu$,*

- *for all* $r \in \mathbb{N}$ *and for all family* $\{f_n, n \in \mathbb{G}_r\} \subseteq \mathcal{B}_b(E^3)$

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

In the following, when unprecise, the filtration implicitly used will be $\mathcal{F}_r = \sigma(X_i, i \in \mathbb{T}_r)$.

Remark 1.1. We may also consider BMC's on a a degree tree (with $a \geq 2$) without any additional technicalities but heavy additional notations. In the same spirit, Markov chains of higher order (such as BAR processes considered in [6]) could be handled by the same techniques. A nontrivial extension would be the case of BMC on a Galton–Watson tree (see, for example, [5] under very strong assumptions) but this will be considered in future works.

1.2. Transportation inequality

We recall that (E, d) is a metric space endowed with its Borel σ-algebra $\mathcal{E}$. Given $p \geq 1$, the L^p-Wasserstein distance between two probability measures μ and ν on E is defined by

$$W_p^d(\nu, \mu) = \inf\left(\iint d(x, y)^p \, d\pi(x, y)\right)^{1/p},$$

where the infimum is taken over all probability measures π on the product space $E \times E$ with marginal distributions μ and ν, called coupling of (μ, ν). This infimum is finite as soon as μ and ν have finite moments of order p. When $d(x, y) = \mathbb{1}_{x\neq y}$ (the trivial measure), $2W_1^d(\mu, \nu) = \|\mu - \nu\|_{TV}$, the total variation of $\mu - \nu$.

The Kullback information (or relative entropy) of ν with respect to μ is defined as

$$H(\nu|\mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} \, d\nu, & \text{if } \nu \ll \mu, \\ +\infty, & \text{else.} \end{cases}$$

Definition 1.3 (L^p-transportation inequality). *We say that the probability measure* μ *satisfies the* L^p*-transportation inequality* T_p *on* (E, d), *denoted* $\mu \in T_p(C)$, *if there is some constant* $C > 0$ *such that for any probability measure* ν,

$$W_p^d(\nu, \mu) \leq \sqrt{2CH(\nu|\mu)}.$$

This transportation inequality has been introduced by Marton [30,31] as a tool for (Gaussian) concentration measure property. Transportation inequality should be called transportation cost-information inequality (as noted in Villani [35]), the transportation cost being the Wasserstein distance and the information Kullback–Leibler information. One may change the information for Fisher information to get other type of concentration result. The following result will be crucial in the sequel. It gives a characterization of L^1-transportation inequality in term of concentration inequality. This is one of the main tool to get deviation inequalities (via Markov inequality).

Theorem 1.4 ([9]). *The measure μ satisfies the L^1-transportation inequality $T_1(C)$ on (E, d) with constant $C > 0$ if and only if for any Lipschitzian function $F : (E, d) \to \mathbb{R}$, F is μ-integrable and*

$$\int_E \exp\bigl(\lambda\bigl(F - \langle F\rangle_\mu\bigr)\bigr)\, d\mu \le \exp\biggl(\frac{\lambda^2}{2} C \|F\|_{\mathrm{Lip}}^2\biggr), \qquad \forall \lambda \in \mathbb{R},$$

where $\langle F\rangle_\mu = \int_E F\, d\mu$ and

$$\|F\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)} < +\infty.$$

In particular, we have the concentration inequality

$$\mu\bigl(F - \langle F\rangle_\mu \le -t\bigr) \vee \mu\bigl(F - \langle F\rangle_\mu \ge t\bigr) \le \exp\biggl(-\frac{t^2}{2C\|F\|_{\mathrm{Lip}}^2}\biggr) \qquad \forall t \in \mathbb{R}.$$

In this work, we will mainly focus on transportation cost inequality T_1 where a considerable literature already exists. As a flavor, let us first cite the characterization of T_1 as a Gaussian integrability property [17] (see also [20]).

Theorem 1.5 ([17]). *μ satisfies the L^1-transportation inequality $T_1(C)$ on (E, d) if and only if there exists $\delta > 0$ and $x_0 \in E$ such that*

$$\mu\bigl(e^{\delta d^2(x, x_0)}\bigr) < \infty,$$

where the constant C can be made explicit.

There is also a large deviations characterization [22]. Recent striking results on transportation inequalities have been obtained for T_2, namely that they are equivalent to dimension free Gaussian concentration [21], or to a restricted class of logarithmic Sobolev inequalities [24]. See also [11] or [12] for practical application based on Lyapunov type criterion and we refer for example to [23] or [35] for surveys on transportation inequality. One of it main aspect is the tensorization property, that is, $\mu^{\otimes n}$ will satisfy some transportation measure if μ does (with dependence on the dimension n). One important development was to consider such a property for dependent sequences such as Markov chains. In [17], Djellout *et al.* have generalized Marton's [32] result providing conditions under which the law of a homogeneous Markov chain $(Y_k)_{1 \le k \le N}$ on E^N satisfies the L^p-transportation inequality T_p with respect to the metric

$$d_{l_p}(x, y) := \Biggl(\sum_{i=1}^N d(x_i, y_i)^p\Biggr)^{1/p}.$$

We will follow similar ideas here to establish the L^p-transportation inequality for the law of a BMC $(X_i)_{1 \le i \le N}$ on E^N. This will allow us to obtain concentration inequalities under hypotheses largely weaker than those of Bitseki *et al.* [7].

Remark 1.2. There are natural generalizations of the $T_1(C)$ inequality often denoted $\alpha - T_1(C)$ inequality, where α is a nonnegative convex lower semi continuous function vanishing at 0. We say that the probability measure μ satisfies $\alpha - T_1(C)$ if for any probability measure ν

$$\alpha\big(W_1(\nu,\mu)\big) \le 2CH(\nu/\mu).$$

The usual $T_1(C)$ inequality is then the case where $\alpha(t) = t^2$. Gozlan [20] has generalized Bobkov–Götze's Laplace transform control [9] and Djellout–Guillin–Wu [17] integrability criterion to this setting enabling to recover sub or super Gaussian concentration. The results of the following section can be generalized to this setting, however adding technical details and heavy notations. Details will thus be left to the reader.

1.3. Objectives and plan of the paper

Our main goal is to furnish easy to verify, and less stringent than uniform ergodicity, conditions to get transportation inequalities and particular concentration inequalities. Let us be more precise.

1. First, we will prove transportation cost inequality for the law $\mathcal{P}$ of the process up to generation n, that is, for every probability measure ν

$$W_p^{d_{l_p}}(\nu,\mathcal{P}) \le \sqrt{2C\big(|\mathbb{T}_n|, p\big)H(\nu|\mathcal{P})},$$

for some function $C(n,p)$ with a "good behaviour" with respect to the dimension n. From this, we may deduce general deviation inequalities: for every 1-Lipschitzian functions F (w.r.t. d_{l_p}) and $t > 0$

$$\mathbb{P}\big(F(X_i)_{i\le|\mathbb{T}_n|} - \mathbb{E}\big(F(X_i)_{i\le|\mathbb{T}_n|}\big) \ge t\big) \le e^{-\frac{t^2}{2C(|\mathbb{T}_n|,p)}}.$$

This will be proved under a contraction assumption in Wasserstein distance of the Markov kernel, and uniform Gaussian integrability property. It is the purpose of Section 2. To get a good dimensional behaviour, the contraction has to be a strict contraction.
2. However transportation inequalities may for some example be too general as it leads to deviation inequalities for all Lipschitzian functions. For statistical purposes, it is more likely to get a functional of interest empirical mean, for which one can hope to get an "averaging" effect between marginal kernels P_0 and P_1. We will prove in Section 3, that it is indeed possible to relax slightly the contraction assumption on each kernel. This allows one kernel to have expanding bounds (rather than strict contraction).
3. We finally apply these deviation inequalities in the nonparametric statistical problem of kernel estimation of the leading function in (nonlinear) bifurcating autoregressive processes. It is the subject of Section 4. It enables us to give nonasymptotic deviation inequalities (sharp wrt the dimension) for these estimators.

2. Transportation cost inequalities for bifurcating Markov chains

Let $(X_i, i \in \mathbb{T})$ be a BMC on E with $\mathbb{T}$-probability transition P and initial measure ν. For $p \geq 1$ and $C > 0$, we consider the following assumption that we shall call $H_p(C)$ in the sequel.

Assumption 2.1 ***($H_p(C)$).*** *There exists $q > 0$ with the following statements*:

(a) $\nu \in T_p(C)$;
(b) $P(x, \cdot, \cdot) \in T_p(C)$ *on* (E^2, d_{l_p}), $\forall x \in E$;
(c) $W_p^{d_{l_p}}(P(x, \cdot, \cdot), P(\tilde{x}, \cdot, \cdot)) \leq q d(x, \tilde{x})$, $\forall x, \tilde{x} \in E$.

Recall that P_0 and P_1 are the margin measures of P. It is important to remark that under $H_p(C)$, condition (b) implies that P_0 and P_1 also satisfies (uniformly) a transportation inequality. Moreover, condition (c) implies that there exist positive constants q_0 and q_1 smaller than q such that P_0 and P_1 fit the same condition. Furthermore, if $P(x, dy, dz) = P_0(x, dy)P_1(x, dz)$, we have $q \leq (q_0^p + q_1^p)^{1/p}$.

As a proof for these comments, let us state the following proposition.

Proposition 2.2. *Assume $C > 0$ and $p \geq 1$ two constants such that $H_p(C)$ is available with an associated constant $q > 0$, then there exists q_0 and q_1 in $(0, q)$ such that for any $b = 0, 1$*:

(i) $P_b(x, \cdot) \in T_p(C)$ *on* (E, d), $\forall x \in E$;
(ii) $W_p^d(P_b(x, \cdot), P_b(\tilde{x}, \cdot)) \leq q_b d(x, \tilde{x})$, $\forall x, \tilde{x} \in E$.

Furthermore, if $P(x, dy, dz) = P_0(x, dy)P_1(x, dz)$, then $q \leq (q_0^p + q_1^p)^{1/p}$.

Proof. Assume π be any coupling of $(P(x, \cdot, \cdot), P(\tilde{x}, \cdot, \cdot))$ for $x, \tilde{x} \in E$ fixed:

$$\begin{aligned}\int_{E^4} d_{l_p}(y, z)^p \, d\pi(dy, dz) &:= \int_{E^4} d(y_0, z_0)^p + d(y_1, z_1)^p \pi(dy_0, dy_1, dz_0, dz_1) \\ &= \int_{E^2} d(y_0, z_0)^p \pi_0(dy_0, dz_0) + \int_{E^2} d(y_1, z_1)^p \pi_1(dy_1, dz_1),\end{aligned}$$

where $\pi_0(dy_0, dz_0) = \pi(dy_0, E, dz_0, E)$ and $\pi_1(dy_1, dz_1) = \pi(E, dy_1, E, dz_1)$. Thus, π_1 and π_2 are both respectively, coupling of $(P_0(x, \cdot), P_0(\tilde{x}, \cdot))$ and $(P_1(x, \cdot), P_1(\tilde{x}, \cdot))$. Firstly for $b = 0, 1$, if π is an optimal coupling, that is, π realizes the minimum overall possible couplings (see [35] for existence):

$$\begin{aligned}W_p^{d_{l_p}}\big(P(x, \cdot, \cdot), P(\tilde{x}, \cdot, \cdot)\big)^p &\geq \int_{E^2} d(y_b, z_b)^p \pi_b(dy_b, dz_b) \\ &\geq W_p^d\big(P_b(x, \cdot), P_b(\tilde{x}, \cdot)\big)^p.\end{aligned} \tag{2.1}$$

Thus, existence of q_b is proven.

Again, if both π_0 and π_1 realize the minimums, then with $\pi = \pi_0 \otimes \pi_1$ coupling for $(P(x,\cdot,\cdot), P(\tilde{x},\cdot,\cdot))$ when $P(x,\cdot,\cdot) = P_0(\cdot,dy)P_1(\cdot,dz)$:

$$\begin{aligned} W_p^d\big(P_0(x,\cdot),P_0(\tilde{x},\cdot)\big)^p + W_p^d\big(P_1(x,\cdot),P_1(\tilde{x},\cdot)\big)^p &= \int_{E^4} d_{l_p}(y,z)^p \pi(dy,dz) \\ &\geq W_p^{d_{l_p}}\big(P(x,\cdot,\cdot),P(\tilde{x},\cdot,\cdot)\big)^p, \end{aligned}$$

and the last assertion of the Proposition 2.2 then follows from this inequality.

For the first statement, we focus on $b = 0$ since the same arguments hold for $b = 1$. Let denote $P^1(x,\cdot|x_0)$ the conditional probability measure according to the first coordinate from its coupling $P(x,\cdot,\cdot)$. Assume ν be any probability measure on E and $\pi(dx_0,dx_1) = \nu(dx_0)P^1(x,dx_1|x_0)$ a coupling of $(\nu, P^1(x,\cdot|x_0))$. Similarly to (2.1)

$$W_p^d\big(P_b(x,\cdot),\nu\big) \leq W_p^{d_{l_p}}\big(P(x,\cdot,\cdot),\pi\big) \leq \sqrt{2CH\big(\pi|P(x,\cdot,\cdot)\big)}.$$

Now according to lemma 2.4 (below)

$$H\big(\pi|P(x,\cdot,\cdot)\big) = H\big(\nu|P_0(x,\cdot)\big) + \int_E H\big(P^1(x,\cdot|x_0)|P^1(x,\cdot|x_0)\big)\nu(dx_0).$$

Result comes naturally since $H(\mu|\mu) = 0$ whatever the probability measure μ is. □

Let us note thanks to the Hölder inequality that $H_p(C)$ implies $H_1(C)$.

We do not suppose here that q, q_0 and q_1 are strictly less than 1, and thus the two marginal chains, as well as the bifurcating one, are not in principle strict contractions. We are thus considering here both "stable" and "unstable" cases, that is, the case where the two marginal Markov chains with transition P_0 and P_1 are ergodic and the case where only one of them is ergodic. We deduce now the following result for the law of the whole trajectory on the binary tree. In the sequel, the symbol "$\asymp$" will denote equality up to a multiplicative constant which does not depend on the size of the sample.

Theorem 2.3. *Let $n \in \mathbb{N}$ and let $\mathcal{P}$ be the law of $(X_i)_{i\in\mathbb{T}_n}$ and denote $N = |\mathbb{T}_n|$. We assume $H_p(C)$ for $1 \leq p \leq 2$ and $C > 0$. Then $\mathcal{P} \in T_p(C_N)$ where*

$$C_N \asymp \begin{cases} C\dfrac{N^{2/p-1}}{(1-q)^2}, & \text{if } q < 1, \\ C\exp\Big(2-\dfrac{2}{p}\Big)N^{2/p+1}, & \text{if } q = 1, \\ C(N+1)\Big(\dfrac{\exp(q-1)r^{pN}}{r^p-1}\Big)^{2/p}, & \text{if } q > 1. \end{cases}$$

Before the proof of this result, let us make the following notations.

Let χ be any Polish space and denote $\mathcal{M}_1(\chi)$ the space of probability measures on χ. We assume now E separable meaning that whatever $N \in \mathbb{N}^*$, E^N is also a Polish space. For any element $x \in E^N$ and $1 \le i \le N$, we denote $x^i := (x_1, \dots, x_i)$. Let $(X_1, \dots, X_N) \in E^N$ be a random vector with joint distribution $\mu \in \mathcal{M}_1(E^N)$.

Similarly we denote:

- for $1 \le i \le N$, μ^i the law of X^i,
- for $1 \le j < i \le N$, $\mu^i_{x^j}$ the conditional law of $(X_i, X_{i-1}, \dots, X_{j+1})$ given $X^j = x^j$ with the convention $\mu^1_{x^0} = \mu^1$, where $x^0 = x_0$ is some fixed point.

In particular, if μ is the law of a BMC with $\mathbb{T}$-probability transition P, then $\mu^{2i+1}_{x^{2i-1}} = P(x_i, \cdot, \cdot)$. For the convenience of the readers, we recall the formula of additivity of entropy (see, for example, [35], Lemma 22.8).

Lemma 2.4. *Let $N \in \mathbb{N}$, let $\chi_1, \dots, \chi_N$ be Polish spaces and $\mathcal{P}, \mathcal{Q} \in \mathcal{M}_1(\chi)$ where $\chi = \prod_{i=1}^N \chi_i$. Then*

$$H(\mathcal{Q}|\mathcal{P}) = \sum_{i=1}^N \int_\chi H\big(\mathcal{Q}^i_{x^{i-1}} | \mathcal{P}^i_{x^{i-1}}\big) \mathcal{Q}(dx).$$

We can now prove the theorem.

Proof of the Theorem 2.3. Let $\varepsilon > 0$ and $\mathcal{Q} \in \mathcal{M}_1(E^N)$. Assume that $H(\mathcal{Q}|\mathcal{P}) < \infty$ (trivial otherwise). The idea is to conditionally study each generation by pairs w.r.t the previous $\mathbb{G}_{n-1}$. Conditionally to their ancestors, every pair of offsprings of an individual is independent of the others from the same generation. Let i be a member of generation $\mathbb{G}_{j-1}$, and define for a realization x on the tree $\mathbb{T}_i(x) := (x_1, \dots, x_{|\mathbb{T}_j|})$. By the definition of the Wasserstein distance, there is a coupling $\pi^{2i+1}_{y^{2i-1}, x^{2i-1}}$ of $(\mathcal{Q}^{2i+1}_{y^{2i-1}}, \mathcal{P}^{2i+1}_{x^{2i-1}})$ such that

$$\begin{aligned}
\mathcal{A}_i &:= \int \big(d(y_{2i}, x_{2i})^p + d(y_{2i+1}, x_{2i+1})^p\big)\, d\pi^{2i+1}_{y^{2i-1}, x^{2i-1}} \\
&\le (1+\varepsilon) W_p^{d_{l_p}} \big(\mathcal{Q}^{2i+1}_{y^{2i-1}}, \mathcal{P}^{2i+1}_{x^{2i-1}}\big)^p \\
&\le (1+\varepsilon)\big[W_p^{d_{l_p}} \big(\mathcal{Q}^{2i+1}_{y^{2i-1}}, \mathcal{P}^{2i+1}_{y^{2i-1}}\big) + W_p^{d_{l_p}} \big(\mathcal{P}^{2i+1}_{y^{2i-1}}, \mathcal{P}^{2i+1}_{x^{2i-1}}\big)\big]^p \\
&= (1+\varepsilon)\big[W_p^{d_{l_p}} \big(\mathcal{Q}^{2i+1}_{y^{2i-1}}, P(y_i, \cdot, \cdot)\big) + W_p^{d_{l_p}} \big(P(y_i, \cdot, \cdot), P(x_i, \cdot, \cdot)\big)\big]^p,
\end{aligned}$$

where the second inequality is obtained thanks to the triangle inequality for the W_p^d distance and the equality is a consequence of the Markov property. By $H_p(C)$ and the convexity of the function $x \mapsto x^p$, we obtain, for $a, b > 1$ such that $1/a + 1/b = 1$,

$$\begin{aligned}
\mathcal{A}_i &\le (1+\varepsilon)\Big(\sqrt{2CH_i\big(y^{2i-1}\big)} + q d(y_i, x_i)\Big)^p \\
&\le (1+\varepsilon)\Big(a^{p-1}\Big(\sqrt{2CH_i\big(y^{2i-1}\big)}\Big)^p + b^{p-1} q^p d^p(y_i, x_i)\Big),
\end{aligned}$$

where $H_i(x) = H(\mathcal{Q}^{2i+1}_x | \mathcal{P}^{2i+1}_x)$.

By recurrence on i, this entails that $\mathbb{E}[d(Y_i, X_i)^p] < +\infty$ for all $i \in \{1, \dots, N\}$ where $(Y_i)_{1\leq i\leq N}$ is distributed as $\mathcal{Q}$. Taking the expected value under $\mathcal{Q}$ and summing on i, we obtain

$$\sum_{i=0}^{|\mathbb{T}_{n-1}|} \mathbb{E}[\mathcal{A}_i] \leq (1+\varepsilon)\left(a^{p-1}(2C)^{p/2} \sum_{i=1}^{|\mathbb{T}_{n-1}|} \mathbb{E}\big[H_i\big(Y^{2i-1}\big)^{p/2}\big]\right) + \left(b^{p-1}q^p \sum_{i=0}^{|\mathbb{T}_{n-2}|} \mathbb{E}[\mathcal{A}_i]\right).$$

Letting ε goes to 0^+, we are led to

$$\sum_{i=0}^{|\mathbb{T}_{n-1}|} \mathbb{E}[\mathcal{A}_i] \leq \sum_{i=1}^{N}\big(a^{p-1}(2C)^{p/2}\mathbb{E}\big[H_i\big(Y^{2i-1}\big)^{p/2}\big]\big) + \left(b^{p-1}q^p \sum_{i=0}^{|\mathbb{T}_{n-2}|} \mathbb{E}[\mathcal{A}_i]\right).$$

Now, in the same way as before, one can show that

$$\sum_{i=0}^{|\mathbb{T}_{n-2}|} \mathbb{E}[\mathcal{A}_i] \leq \sum_{i=1}^{N-1}\big(a^{p-1}(2C)^{p/2}\mathbb{E}\big[H_i\big(Y^{2i-1}\big)^{p/2}\big]\big) + \left(b^{p-1}q^p \sum_{i=0}^{|\mathbb{T}_{n-3}|} \mathbb{E}[\mathcal{A}_i]\right),$$

and more generally, for all $k \in \{1, \dots, n\}$, we have

$$\sum_{i=0}^{|\mathbb{T}_{n-k}|} \mathbb{E}[\mathcal{A}_i] \leq \sum_{i=1}^{N-k+1}\big(a^{p-1}(2C)^{p/2}\mathbb{E}\big[H_i\big(Y^{2i-1}\big)^{p/2}\big]\big) + \left(b^{p-1}q^p \sum_{i=0}^{|\mathbb{T}_{n-k-1}|} \mathbb{E}[\mathcal{A}_i]\right).$$

We set $h_i = a^{p-1}(2C)^{p/2}\mathbb{E}[H_i(Y^{2i-1})^{p/2}]$. Then, using the previous inequalities, we obtain

$$\begin{aligned}
\sum_{i=0}^{|\mathbb{T}_{n-1}|} \mathbb{E}[\mathcal{A}_i] &\leq \sum_{i=N-n}^{N}\left(\sum_{j=1}^{i} h_j\right)\big(b^{p-1}q^p\big)^{N-i} \\
&\leq \sum_{i=1}^{N}\left(\sum_{j=1}^{i} h_j\right)\big(b^{p-1}q^p\big)^{N-i} \\
&= \sum_{i=1}^{N} h_i \sum_{j=0}^{N-i}\big(b^{p-1}q^p\big)^j \\
&\leq \left(\sum_{i=1}^{N} h_i^{2/p}\right)^{p/2} \left(\sum_{i=1}^{N}\left(\sum_{j=0}^{N-i}\big(b^{p-1}q^p\big)^j\right)^{\frac{2}{2-p}}\right)^{\frac{2-p}{2}},
\end{aligned}$$

where the second inequality was obtained using the fact that the quantities h_i are non-negative and the last inequality was obtained thanks to Hölder inequality. By the definition of the Wasserstein distance, the additivity of entropy and using the concavity of the function $x \mapsto x^{p/2}$ for

$p \in [1,2]$, we obtain

$$W_p^{d_{l_p}}(\mathcal{Q},\mathcal{P})^p \le a^{p-1}\big(2CH(\mathcal{Q}|\mathcal{P})\big)^{p/2}\left(\sum_{i=1}^{N}\left(\sum_{j=0}^{N-i}\big(b^{p-1}q^p\big)^j\right)^{\frac{2}{2-p}}\right)^{\frac{2-p}{2}}$$
$$\le a^{p-1}\big(2CH(\mathcal{Q}|\mathcal{P})\big)^{p/2}N^{1-\frac{p}{2}}\sum_{j=0}^{N-1}\big(b^{p-1}q^p\big)^j.$$

When $q<1$, we take $b=q^{-1}$, so that $b^{p-1}q^p=r<1$ and the desired result follows easily. When $q\ge 1$, we take $b=1+1/N$ and the results follow from simple analysis and this ends the proof. □

Remark 2.1. For $q<1$ and $p=1$, the constant C_N increases linearly on the dimension N. However, for $p=2$ this constant is independent of the dimension as in the i.i.d. case.

Remark 2.2. As we will see in the next section, still when $q<1$, Theorems 2.3 and 1.4 applied to $F(X_1,\dots,X_N)=(1/N)\sum_{i=1}^N f(X_i)$, where f is a Lipschitzian function defined on E, gives us deviation inequalities with a good order of N. But, without any averaging effect as for example for $F(X_1,\dots,X_N)=f(X_N)$, deviation inequalities are not anymore efficient as a function of N. The same remark holds when $F(X_1,\dots,X_N)=g(X_n,X_{2n},X_{2n+1})$ with $n\in\{1,\dots,(N-N[2])\}$ and g a Lipschitzian function defined on E^3. Since this last question is of great interest for L^1-transportation cost inequality of the BMC invariant measure, we give the following results.

Proposition 2.5. *Under $H_1(C)$, for any $n\in\mathbb{T}$ and $x\in E$*

$$\mathcal{L}(X_n|X_1=x)\in T_1(c_n),$$

where

$$c_n=C\sum_{k=0}^{r_n-1}q_0^{2(k-a_k)}q_1^{2a_k};\qquad a_0=0,$$

and for all $k\in\{1,\dots,r_n-1\}$, a_k is the number of ancestor of type 1 of X_n which are between the r_n-k+1th generation and the r_nth generation.

Before the proof, we introduce some additional notations. Let $n\in\mathbb{T}$, we denote $(b_1,\dots,b_{r_n})\in\{0,1\}^{r_n}$ the type representation of the unique path from 1 to n. More precisely, recall that each individual belongs to type 0 or 1. Since $\mathbb{T}$ is a regular binary tree, one can find only one path between two individuals where each individual within has a type b. The type representation for a path is thus the types of the individuals within it. Note that we keep the chronological order, that is, the ith type represents the ith individual in the path.

Then, for all $i \in \{1, \dots, r_n\}$, b_i is the ancestor's type of n which is in the ith generation and the quantities a_k defined in the Proposition 2.5 are given by

$$a_k = \sum_{i=r_n-k+1}^{r_n} b_i.$$

For all $k \in \{1, \dots, r_n\}$, we denote by P^k and P^{-k} the iterated of the transition probabilities redP_0 and redP_1 defined $\forall f \in \mathcal{B}(E^p)$:

$$P^k f := P_{b_1} \circ \cdots \circ P_{b_k} f \quad \text{and} \quad P^{-k} f := P_{b_{r_n-k}} \circ \cdots \circ P_{b_{r_n}} f.$$

Proof of the Proposition 2.5. First, note that since

$$W_1^d(\nu, \mu) = \sup_{f:\|f\|_{\text{Lip}} \le 1} \left| \int_S f \, d\mu - \int_S f \, d\nu \right|,$$

condition (c) of $H_1(C)$ implies that

$$\|P_b f\|_{\text{Lip}} \le q_b \|f\|_{\text{Lip}} \qquad \forall b \in \{0, 1\}.$$

Now let f be a Lipschitzian function defined on E. By (b)–(c) of $H_1(C)$ and Theorem 1.4 applied to $P_{b_{r_n}}(e^f)$, we have

$$P^{r_n}(e^f) \le P^{r_n-1}\left(\exp\left(P_{b_{r_n}} f + \frac{C\|f\|_{\text{Lip}}^2}{2}\right)\right).$$

Once again, applying Theorem 1.4 on $P_{b_{r_n-1}} \circ P_{b_{r_n}}(e^f)$, we obtain

$$P^{r_n}(e^f) \le P^{r_n-2}\left(\exp\left(P^{-1} f + \frac{C\|f\|_{\text{Lip}}^2}{2} + \frac{C\|P_{b_{r_n}} f\|_{\text{Lip}}^2}{2}\right)\right).$$

By iterating this method, we are led to

$$P^{r_n}(e^f) \le \exp\left(P^{-r_n+1} f + \left(1 + q_{b_{r_n}}^2 + q_{b_{r_n}}^2 q_{b_{r_n-1}}^2 + \cdots + \prod_{i=2}^{r_n} q_{b_i}^2\right) \frac{C\|f\|_{\text{Lip}}^2}{2}\right).$$

Since

$$1 + q_{b_{r_n}}^2 + q_{b_{r_n}}^2 q_{b_{r_n-1}}^2 + \cdots + \prod_{i=2}^{q_n} q_{b_i}^2 = \sum_{k=0}^{r_n-1} q_0^{2(k-a_k)} q_1^{2a_k} \quad \text{and} \quad P^{-r_n+1} f = P^{r_n} f,$$

we conclude the proof thanks to Theorem 1.4. □

The next result is a consequence of the previous proposition.

Corollary 2.6. *Assume $H_1(C)$ and $m := \max\{q_0, q_1\} < 1$. Then,* $\forall x_1 \in E$

$$\mathcal{L}(X_n|X_1 = x_1) \in T_1(c_\infty) \quad \textit{and} \quad \mathcal{L}\big((X_n, X_{2n}, X_{2n+1})|X_1 = x_1\big) \in T_1\big(c'_\infty\big),$$

where

$$c_\infty = \frac{C}{1-m^2} \quad \textit{and} \quad c'_\infty = C\left(1 + \frac{(1+q)^2}{1-m^2}\right).$$

Proof. That $\mathcal{L}(X_n|X_1 = x) \in T_1(c_\infty)$ is a direct consequence of Proposition 2.5 bounding q_0 and q_1 by m.

In order to deal with the ancestor-offspring case (X_n, X_{2n}, X_{2n+1}), we first focus on generations r_{2n} and r_n. Let $f : (E^3, d_{l_1}) \to \mathbb{R}$ be a Lipschitzian function. We have

$$\|Pf\|_{\mathrm{Lip}} = \sup_{x,\tilde{x}\in E} \frac{|\int f(x,y,z)P(x,dy,dz) - \int f(\tilde{x},y,z)P(\tilde{x},dy,dz)|}{d(x,\tilde{x})}.$$

Thanks to condition (c) of $H_1(C)$, we have the following inequalities

$$\begin{aligned}
&\left|\int f(x,y,z)P(x,dy,dz) - \int f(\tilde{x},y,z)P(\tilde{x},dy,dz)\right| \\
&\quad \le \|f\|_{\mathrm{Lip}}\big(d(x,\tilde{x}) + W_1^{d_{l_1}}\big(P(x,\cdot), P(\tilde{x},\cdot)\big)\big) \\
&\quad \le (q+1)\|f\|_{\mathrm{Lip}} d(x,\tilde{x}),
\end{aligned}$$

and then

$$\|Pf\|_{\mathrm{Lip}} \le (q+1)\|f\|_{\mathrm{Lip}}.$$

By definition, we have

$$\mathbb{E}\big[\exp\big(f(X_n, X_{2n}, X_{2n+1})\big)\big] = P^{r_n}\big(Pe^f\big).$$

Now, associating the latter remark and $H_1(C)$, we use the same strategy as proof of Proposition 2.5 for the other generations, we are led to

$$\begin{aligned}
&\mathbb{E}\big[\exp\big(f(X_n, X_{2n}, X_{2n+1})\big)\big] \\
&\quad \le \exp\left(P_{b_1}\cdots P_{b_{r_n}} Pf + \frac{C\|f\|_{\mathrm{Lip}}^2}{2} + \frac{C(1+q)^2\|f\|_{\mathrm{Lip}}^2}{2}\sum_{i=0}^{r_n-1} m^{2i}\right).
\end{aligned}$$

Since $P_{b_1}\cdots P_{b_{r_n}} Pf = \mathbb{E}[f(X_n, X_{2n}, X_{2n+1})]$ and $\sum_{i=0}^{r_n-1} m^{2i} \le 1/(1-m^2)$, we obtain

$$\mathbb{E}\big[\exp\big(f(X_n, X_{2n}, X_{2n+1})\big)\big] \le \exp\big(\mathbb{E}\big[f(X_n, X_{2n}, X_{2n+1})\big] + c'_\infty\big),$$

with c'_∞ given in the corollary. We then conclude the proof thanks to Theorem 1.4. □

3. Concentration inequalities for bifurcating Markov chains

3.1. Direct applications of the Theorem 2.3

We will now focus our attention on concentration inequalities for additive functional applied to BMC. Specifically, let $N \in \mathbb{N}^*$ and I be a subset of $\{1, \dots, N\}$ with cardinality $|I|$. Let f be a real function on E or E^3 and set

$$\overline{M}_I(f) = \frac{1}{|I|} \sum_{i \in I} f(\Delta_i),$$

where $\Delta_i = X_i$ if f is defined on E or $\Delta_i = (X_i, X_{2i}, X_{2i+1})$ if f is defined on E^3.

In statistical applications, the cases $N = |\mathbb{T}_n|$ and $I = \mathbb{G}_m$ (for $m \in \{0, \dots, n\}$) or $I = \mathbb{T}_n$ are the more relevant ones (see, for example, [7]). First, we will establish concentration inequalities when f is a real Lipschitzian function defined on E. Then, for $p \geq 1$, $\overline{M}_I(f)$ is also a Lipschitzian function on (E^N, d_{l_p}) and we have

$$\big\|\overline{M}_I(f)\big\|_{\mathrm{Lip}} \leq |I|^{-1/p} \|f\|_{\mathrm{Lip}}.$$

The following result is a direct consequence of Theorem 2.3.

Proposition 3.1. *Let $n \in \mathbb{N}$ and let $\mathcal{P}$ be the law of $(X_i)_{1 \leq i \leq N}$ where we denote $N = |\mathbb{T}_n|$. Let f be a real Lipschitzian function on (E, d). Then, under $H_p(C)$ for $1 \leq p \leq 2$,*

$$\mathcal{P} \circ \overline{M}_I(f)^{-1} \in T_p\big(C_N |I|^{-2/p} \|f\|^2_{\mathrm{Lip}}\big),$$

where C_N is given in the Theorem 2.3 and $\mathcal{P} \circ \overline{M}_I(f)^{-1}$ is the push forward measure of $\mathcal{P}$ by $\overline{M}_I(f)$. In particular, for all $t > 0$ we have

$$\begin{aligned}&\mathbb{P}\big(\overline{M}_I(f)\big(X^N\big) \leq -t + \mathbb{E}\big[\overline{M}_I(f)\big(X^N\big)\big]\big) \vee \mathbb{P}\big(\overline{M}_I(f)\big(X^N\big) \geq t + \mathbb{E}\big[\overline{M}_I(f)\big(X^N\big)\big]\big)\\ &\quad \leq \exp\left(-\frac{t^2 |I|^{2/p}}{2C_N \|f\|^2_{\mathrm{Lip}}}\right).\end{aligned}$$

Proof. The first part is a direct consequence of Theorem 2.3 and Lemma 2.1 of [17]. The second part is an application of Theorem 1.4. □

For the next concentration inequality, we assume that f is a real Lipschitzian function defined on (E^3, d_{l_1}), which means that

$$\big|f(x) - f(y)\big| \leq \|f\|_{\mathrm{Lip}} \sum_{i=1}^{3} d(x_i, y_i) \qquad \forall x, y \in E^3.$$

We assume that N is an odd number. Let I be a subset of $\{1, \dots, (N-1)/2\}$. Now, we denote by $\overline{M}_I(f)$ the real function defined on E^N

$$\overline{M}_I(f)\left(x^N\right) = \frac{1}{|I|}\sum_{i\in I} f(x_i, x_{2i}, x_{2i+1}).$$

For all $x^N, y^N \in E^N$ we have

$$\begin{aligned}\left|\overline{M}_I(f)\left(x^N\right) - \overline{M}_I(f)\left(y^N\right)\right| &\le \frac{\|f\|_{\mathrm{Lip}}}{|I|}\sum_{i\in I}\left(d(x_i, y_i) + d(x_{2i}, y_{2i}) + d(x_{2i+1}, y_{2i+1})\right) \\ &\le \frac{2\|f\|_{\mathrm{Lip}}}{|I|^{1/p}} d_{l_p}\left(x^N, y^N\right).\end{aligned}$$

The constant 2 comes from the fact that one can count at most 2 times the distance between a pair (x_i, y_i) since we sum along triplets. Distance d_{l_p} comes from the convexity of $x \to x^p$.

This implies that $\overline{M}_I(f)$ is a Lipschitzian function on (E^N, d_{l_p}) and $\|\overline{M}_I(f)\|_{\mathrm{Lip}} \le 2\|f\|_{\mathrm{Lip}}/|I|^{1/p}$. We then have the following result.

Proposition 3.2. *Let $n \in \mathbb{N}$ and let $\mathcal{P}$ be the law of $(X_i)_{1\le i\le N}$ where we denote $N = |\mathbb{T}_n|$. Let f be a real Lipschitzian function on (E^3, d_{l_1}). Then, under $H_p(C)$ for $1 \le p \le 2$,*

$$\mathcal{P} \circ \overline{M}_I(f)^{-1} \in T_p\left(2C_N|I|^{-2/p}\|f\|^2_{\mathrm{Lip}}\right),$$

where C_N is given in the Theorem 2.3 and $\mathcal{P} \circ \overline{M}_I(f)^{-1}$ is the push forward measure of $\mathcal{P}$ by $\overline{M}_I(f)$. In particular, for all $t > 0$ we have

$$\begin{aligned}&\mathbb{P}\left(\overline{M}_I(f)\left(X^N\right) \le -t + \mathbb{E}\left[\overline{M}_I(f)\left(X^N\right)\right]\right) \vee \mathbb{P}\left(\overline{M}_I(f)\left(X^N\right) \ge t + \mathbb{E}\left[\overline{M}_I(f)\left(X^N\right)\right]\right) \\ &\quad\le \exp\left(-\frac{t^2|I|^{2/p}}{4C_N\|f\|^2_{\mathrm{Lip}}}\right).\end{aligned}$$

Proof. The proof is a direct consequence of Theorem 2.3, Lemma 2.1 of [17] and Theorem 1.4. □

In the case $p = 1$, $q < 1$ and f Lipschitzian, the previous results applied to the empirical means $\overline{M}_{\mathbb{G}_n}(f)$ and $\overline{M}_{\mathbb{T}_n}(f)$ give us relevant concentration inequalities, that is with good order size w.r.t cardinality subset. However, when $q \ge 1$, only a global effect is considered and no compensation between the contraction of P_0 and P_1 comes into play. The goal of the next subsection is to obtain relevant concentration inequalities for the empirical means $\overline{M}_{\mathbb{G}_n}(f)$ and $\overline{M}_{\mathbb{T}_n}(f)$ when $q \ge 1$ when one of the marginal Markov chain may be "unstable" but however compensated by a strict contraction of the other one.

3.2. Gaussian concentration inequalities for the empirical means $\overline{M}_{\mathbb{G}_n}(f)$ and $\overline{M}_{\mathbb{T}_n}(f)$

Throughout this section, we will focus only in the case $p = 1$, and will assume $H_1(C)$. We set $s = q_0 + q_1$.

The main goal of this subsection is to broaden the range of application of deviation inequalities of $\overline{M}_{\mathbb{G}_n}(f)$ and $\overline{M}_{\mathbb{T}_n}(f)$ to cases where $s > 1$, namely when it is possible that one of the two marginal Markov chains is not a strict contraction, that is, q_0 or $q_1 < 1$. The transportation inequality of Theorem 2.3 is a very powerful tool to get deviation inequalities for all Lipschitzian functions of the whole trajectory (up to generation n), and may thus concern for example Lipschitzian function *exclusively applied to* offspring generated by P_0 or P_1. Consequently, to get "consistent" deviation inequalities, both marginal Markov chains have to be contractions in Wasserstein distance.

However when dealing with $\overline{M}_{\mathbb{G}_n}(f)$ or $\overline{M}_{\mathbb{T}_n}(f)$, we may hope for an averaging effect, that is, if one is not a contraction and the other one a strong contraction it may in a sense compensate. Such averaging effect have been observed at the level of the LLN and CLT in [16,25] but only asymptotically. Our purpose here will be then to show that such averaging effect will also affect deviation inequalities.

We will use, directly inspired by Bobkov-Götze's Laplace transform control, what we call Gaussian Concentration property: for $\kappa > 0$, we will say that a random variable X satisfies $GC(\kappa)$ if

$$\mathbb{E}\big[\exp\big(t\big(X - \mathbb{E}[X]\big)\big)\big] \le \exp\big(\kappa t^2/2\big) \qquad \forall t \in \mathbb{R}.$$

Using Markov's inequality and optimization, this Gaussian concentration property immediately implies that

$$\mathbb{P}\big(X - \mathbb{E}(X) \ge r\big) \le e^{-\frac{r^2}{2\kappa}}.$$

We may thus focus here only on the Gaussian concentration property (GC).

Proposition 3.3. *Let f be a real Lipschitzian function on E and $n \in \mathbb{N}$. Assume that $H_1(C)$ holds. Then $\overline{M}_{\mathbb{G}_n}(f)$ satisfies $GC(\gamma_n)$ where*

$$\gamma_n \asymp \begin{cases} \dfrac{2C\|f\|_{\mathrm{Lip}}^2}{|\mathbb{G}_n|}\left(\dfrac{1-(s^2/2)^{n+1}}{1-s^2/2}\right), & \text{if } s \neq \sqrt{2}, \\ \dfrac{2C\|f\|_{\mathrm{Lip}}^2(n+1)}{|\mathbb{G}_n|}, & \text{if } s = \sqrt{2}. \end{cases}$$

We recall that here $s = q_0 + q_1$.

Remark 3.1. One can observe that for $s < \sqrt{2}$, the previous inequalities are on the same order of magnitude that the inequalities obtained thanks to Proposition 3.1 with $q < 1$. For $\sqrt{2} \le s < 2$ the above inequalities remain relevant since we just have a negligible loss with respect to $|\mathbb{G}_n|$. But for $s \ge 2$, these inequalities are not significant (see the same type of limitations at the CLT level in [16]) and reflects that the non-contracting Markov kernel entails a quite different and explosive behaviour.

Proof. Let f be a real Lipschitzian function on E, $n \in \mathbb{N}$ and $t \in \mathbb{R}$. We have

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i)\Big)\Big] \\
&\quad = \mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_{n-1}} (P_0+P_1)f(X_i)\Big) \\
&\qquad \times \mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_{n-1}} \big(f(X_{2i})+f(X_{2i+1})-(P_0+P_1)f(X_i)\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\Big].
\end{aligned}$$

Thanks to the Markov property, we have

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_{n-1}} \big(f(X_{2i})+f(X_{2i+1})-(P_0+P_1)f(X_i)\big)\Big)\Big|\mathcal{F}_{n-1}\Big] \\
&\quad = \prod_{i\in\mathbb{G}_{n-1}} P\big(\exp\big(t2^{-n}\big(f\oplus f-(P_0+P_1)f\big)\big)\big)(X_i),
\end{aligned}$$

where $f\oplus f$ is the function on E^2 defined by $(f\oplus f)(x,y)=f(x)+f(y)$. We recall that from $H_1(C)$ we have $P(x,\cdot,\cdot)\in T_1(C)$ for all $x\in E$. Now, thanks to Theorem 1.4, we have

$$\prod_{i\in\mathbb{G}_{n-1}} P\big(\exp\big(t2^{-n}\big(f\oplus f-(P_0+P_1)f\big)\big)\big)(X_i) \le \prod_{i\in\mathbb{G}_{n-1}} \exp\Big(\frac{t^2C\|f\oplus f\|_{\mathrm{Lip}}^2}{2\times 2^{2n}}\Big).$$

Since $\|f\oplus f\|_{\mathrm{Lip}} \le 2\|f\|_{\mathrm{Lip}}$, we are led to

$$\mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i)\Big)\Big] \le \exp\Big(\frac{2^2t^22^{n-1}C\|f\|_{\mathrm{Lip}}^2}{2\times 2^{2n}}\Big)\mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_{n-1}} (P_0+P_1)f(X_i)\Big)\Big].$$

Doing the same for $\mathbb{E}[\exp(t2^{-n}\sum_{i\in\mathbb{G}_{n-1}}(P_0+P_1)f(X_i))]$ with $(P_0+P_1)f$ replacing f and using the inequality

$$\big\|(P_0+P_1)f\oplus(P_0+P_1)f\big\|_{\mathrm{Lip}} \le 2s\|f\|_{\mathrm{Lip}},$$

we are led to

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i)\Big)\Big] &\le \mathbb{E}\Big[\exp\Big(t2^{-n}\sum_{i\in\mathbb{G}_{n-2}} (P_0+P_1)^2f(X_i)\Big)\Big] \\
&\qquad \times \exp\Big(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2 2^{n-1}}{2\times 2^{2n}}\Big)\exp\Big(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2 s^2 2^{n-2}}{2\times 2^{2n}}\Big).
\end{aligned}$$

Iterating this method and using the inequalities

$$\left\|(P_0+P_1)^k f \oplus (P_0+P_1)^k f\right\|_{\mathrm{Lip}} \le 2s^k \|f\|_{\mathrm{Lip}} \qquad \forall k \in \{1,\dots,n-1\},$$

we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i)\right)\right] \\
&\quad \le \exp\left(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2}{2\times 2^{2n}}\sum_{k=0}^{n-1} s^{2k}2^{n-k-1}\right)\mathbb{E}\bigl[\exp\bigl(t2^{-n}(P_0+P_1)^n f(X_1)\bigr)\bigr].
\end{aligned}$$

Since $\mathbb{E}[t2^{-n}(P_0+P_1)^n f(X_1)] = \mathbb{E}[t2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i)] = t2^{-n}\nu(P_0+P_1)^n f$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t2^{-n}\left(\sum_{i\in\mathbb{G}_n} f(X_i) - \nu(P_0+P_1)^n f\right)\right)\right] \\
&\quad \le \exp\left(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2}{2\times 2^{2n}}\sum_{k=0}^{n-1} s^{2k}2^{n-k-1}\right)\mathbb{E}\bigl[\exp\bigl(t2^{-n}\bigl((P_0+P_1)^n f(X_1)\bigr) - \nu(P_0+P_1)^n f\bigr)\bigr].
\end{aligned}$$

Thanks to $H_1(C)$, we conclude that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t2^{-n}\left(\sum_{i\in\mathbb{G}_n} f(X_i) - \nu(P_0+P_1)^n f\right)\right)\right] \\
&\quad \le \exp\left(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2}{2\times 2^{2n}}\sum_{k=0}^{n} s^{2k}2^{n-k-1}\right)
\end{aligned}$$

and the results of the proposition follows from this last inequality. □

For the ancestor-offspring triangle (X_i, X_{2i}, X_{2i+1}), we have the following result which can be seen as a consequence of the Proposition 3.3.

Corollary 3.4. *Let f be a real Lipschitzian function on E^3 and $n\in\mathbb{N}$. Assume that $H_1(C)$ holds. Then $\overline{M}_{\mathbb{G}_n}(f)$ satisfies $GC(\gamma_n')$ where*

$$\gamma_n' \asymp \begin{cases} \dfrac{2C(1+q)^2\|f\|_{\mathrm{Lip}}^2}{s^2|\mathbb{G}_n|}\left(\dfrac{1-(s^2/2)^{n+2}}{1-s^2/2}\right), & \text{if } s\neq\sqrt{2}, \\[2ex] \dfrac{2C(1+q)^2\|f\|_{\mathrm{Lip}}^2(n+2)}{|\mathbb{G}_n|}, & \text{if } s=\sqrt{2}. \end{cases}$$

Proof. Let f be a real Lipschitzian function on E^3, $n \in \mathbb{N}$ and $t \in \mathbb{R}$. We have

$$\mathbb{E}\left[\exp\left(t2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i, X_{2i}, X_{2i+1})\right)\right]$$
$$= \mathbb{E}\left[\exp\left(t2^{-n}\sum_{i\in\mathbb{G}_n} Pf(X_i)\right)\mathbb{E}\left[\exp\left(t2^{-n}\sum_{i\in\mathbb{G}_n}\big(f(X_i, X_{2i}, X_{2i+1}) - Pf(X_i)\big)\right)\Big|\mathcal{F}_n\right]\right].$$

By the Markov property and thanks to the Proposition 2.3 and the Theorem 1.4, we have

$$\mathbb{E}\left[\exp\left(t2^{-n}\sum_{i\in\mathbb{G}_n}\big(f(X_i, X_{2i}, X_{2i+1}) - Pf(X_i)\big)\right)\Big|\mathcal{F}_n\right] \le \exp\left(\frac{t^2C\|f\|^2_{\mathrm{Lip}}2^n}{2\times 2^{2n}}\right).$$

Now, using Pf instead of f in the proof of the Proposition 3.3 and using the fact that $\|Pf\|_{\mathrm{Lip}} \le (1+q)\|f\|_{\mathrm{Lip}}$ and

$$\mathbb{E}\left[2^{-n}\sum_{i\in\mathbb{G}_n} f(X_i, X_{2i}, X_{2i+1})\right] = \mathbb{E}\left[2^{-n}\sum_{i\in\mathbb{G}_n} Pf(X_i)\right] = 2^{-n}\nu(P_0+P_1)^nPf,$$

we are led to

$$\mathbb{E}\left[\exp\left(t2^{-n}\left(\sum_{i\in\mathbb{G}_n} f(X_i, X_{2i}, X_{2i+1}) - \nu(P_0+P_1)^nPf\right)\right)\right]$$
$$\le \exp\left(\frac{4t^2C(1+q)^2\|f\|^2_{\mathrm{Lip}}}{2^2\times 2^n}\sum_{k=-1}^{n}\left(\frac{s^2}{2}\right)^k\right).$$

The results then follows by easy calculations. □

For the subtree $\mathbb{T}_n$, we have the following result.

Proposition 3.5. *Let f be a real Lipschitzian function on E and $n \in \mathbb{N}$. Assume that $H_1(C)$ holds. Then $\overline{M}_{\mathbb{T}_n}(f)$ satisfies $GC(\tau_n)$ where*

$$\tau_n \asymp \begin{cases} \dfrac{2C\|f\|^2_{\mathrm{Lip}}}{(s-1)^2|\mathbb{T}_n|}\left(1+\dfrac{1-(s^2/2)^{n+1}}{1-s^2/2}\right), & \text{if } s\ne\sqrt{2}, s\ne 1,\\ \dfrac{2C\|f\|^2_{\mathrm{Lip}}}{(s-1)^2|\mathbb{T}_n|}\big(s^2(n+1)+1\big), & \text{if } s=\sqrt{2},\\ \dfrac{2C\|f\|^2_{\mathrm{Lip}}}{|\mathbb{T}_n|^2}\left(|\mathbb{T}_n|-\dfrac{n+1}{2}\right), & \text{if } s=1.\end{cases}$$

Proof. Let f be a real Lipschitzian function on E and $n \in \mathbb{N}$. Note that

$$\mathbb{E}\Big[\sum_{i\in\mathbb{T}_n} f(X_i)\Big] = \nu\Big(\sum_{m=0}^{n}(P_0+P_1)^m f\Big).$$

We have

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{T}_n} f(X_i)\Big)\Big] \\
&\quad = \mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{T}_{n-2}} f(X_i)\Big)\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{G}_{n-1}}\big(f+(P_0+P_1)f\big)(X_i)\Big) \\
&\qquad \times \mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{G}_{n-1}}\big(f(X_{2i})+f(X_{2i+1})-(P_0+P_1)f(X_i)\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\Big].
\end{aligned}$$

As in the proof of Proposition 3.3, we have

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{G}_{n-1}}\big(f(X_{2i})+f(X_{2i+1})-(P_0+P_1)f(X_i)\big)\Big)\Big|\mathcal{F}_{n-1}\Big] \\
&\quad \le \exp\Big(\frac{2^2Ct^2\|f\|_{\mathrm{Lip}}^2 2^{n-1}}{2|\mathbb{T}_n|^2}\Big).
\end{aligned}$$

This leads us to

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{T}_n} f(X_i)\Big)\Big] &\le \exp\Big(\frac{2^2Ct^2\|f\|_{\mathrm{Lip}}^2 2^{n-1}}{2|\mathbb{T}_n|^2}\Big) \\
&\quad \times \mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{T}_{n-2}} f(X_i)\Big) \\
&\quad \times \exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{G}_{n-1}}\big(f+(P_0+P_1)f\big)(X_i)\Big)\Big].
\end{aligned}$$

Iterating this method, we are led to

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{i\in\mathbb{T}_n} f(X_i)\Big)\Big] &\le \exp\Big(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2}{2|\mathbb{T}_n|^2}\sum_{k=0}^{n-1}\Big(\sum_{l=0}^{k}s^l\Big)^2 2^{n-k-1}\Big) \\
&\quad \times \mathbb{E}\Big[\exp\Big(\frac{t}{|\mathbb{T}_n|}\sum_{m=0}^{n}(P_0+P_1)^m f(X_1)\Big)\Big],
\end{aligned}$$

and we then obtain thanks to (a) of $H_1(C)$ and Theorem 1.4

$$\mathbb{E}\left[\exp\left(\frac{t}{|\mathbb{T}_n|}\left(\sum_{i\in\mathbb{T}_n} f(X_i) - \nu\left(\sum_{m=0}^{n}(P_0+P_1)^m f\right)\right)\right)\right] \leq \exp\left(\frac{2^2t^2C\|f\|_{\mathrm{Lip}}^2}{2|\mathbb{T}_n|^2}\sum_{k=0}^{n}\left(\sum_{l=0}^{k}s^l\right)^2 2^{n-k-1}\right).$$

In the last inequality, we have used

$$\left\|\sum_{m=0}^{n}(P_0+P_1)^m f\right\|_{\mathrm{Lip}} \leq \left(\sum_{k=0}^{n}s^k\right)\|f\|_{\mathrm{Lip}}.$$

The results then easily follows. □

For the ancestor-offspring triangle we have the following results which can be seen as a consequence of the Proposition 3.5.

Corollary 3.6. *Let f be a real Lipschitzian function on E^3 and $n \in \mathbb{N}$. Assume that $H_1(C)$ holds. Then $\overline{M}_{\mathbb{T}_n}(f)$ satisfies $GC(\tau_n')$ where*

$$\tau_n' \asymp \begin{cases} \dfrac{2^3C(1+q)^2\|f\|_{\mathrm{Lip}}^2}{|\mathbb{T}_n|}\left(1+\dfrac{1}{(s-1)^2}\left(1+\dfrac{s^2(1-(s^2/2)^{n+1})}{1-s^2/2}\right)\right), & \text{if } s\neq\sqrt{2},\ s\neq 1,\\ \dfrac{2^3C(1+q)^2\|f\|_{\mathrm{Lip}}^2}{|\mathbb{T}_n|}\left(1+\dfrac{1+s^2(n+1)}{(s-1)^2}\right), & \text{if } s=\sqrt{2},\\ \dfrac{2^3C(1+q)^2\|f\|_{\mathrm{Lip}}^2}{|\mathbb{T}_n|^2}\left(2|\mathbb{T}_n|-\dfrac{n+1}{2}\right), & \text{if } s=1. \end{cases}$$

Proof. Let f be a real Lipschitzian function on E^3 and $n \in \mathbb{N}$. By Hölder inequality and using the fact that

$$\mathbb{E}\left[\sum_{i\in\mathbb{T}_n} f(\Delta_i)\right] = \mathbb{E}\left[\sum_{i\in\mathbb{T}_n} Pf(X_i)\right],$$

we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{t}{|\mathbb{T}_n|}\left(\sum_{i\in\mathbb{T}_n} f(\Delta_i) - \mathbb{E}\left[\sum_{i\in\mathbb{T}_n} f(\Delta_i)\right]\right)\right)\right] &\leq \left(\mathbb{E}\left[\exp\left(\frac{2t}{|\mathbb{T}_n|}\left(\sum_{i\in\mathbb{T}_n}\big(f(\Delta_i)-Pf(X_i)\big)\right)\right)\right]\right)^{1/2}\\ &\quad\times\left(\mathbb{E}\left[\exp\left(\frac{2t}{|\mathbb{T}_n|}\left(\sum_{i\in\mathbb{T}_n} Pf(X_i) - \mathbb{E}\left[\sum_{i\in\mathbb{T}_n} Pf(X_i)\right]\right)\right)\right]\right)^{1/2}.\end{aligned}$$

We bound the first term of the right-hand side of the previous inequality by using the same calculations as in the first iteration of the proof of Corollary 3.4. We then have

$$\left(\mathbb{E}\left[\exp\left(\frac{2t}{|\mathbb{T}_n|}\left(\sum_{i\in\mathbb{T}_n}\big(f(\Delta_i)-Pf(X_i)\big)\right)\right)\right]\right)^{1/2}\le\exp\left(\frac{2t^2C\|f\|_{\mathrm{Lip}}^2|\mathbb{T}_n|}{2|\mathbb{T}_n|^2}\right).$$

For the second term, we use the proof of the Proposition 3.5 with Pf instead of f. We then have

$$\left(\mathbb{E}\left[\exp\left(\frac{2t}{|\mathbb{T}_n|}\left(\sum_{i\in\mathbb{T}_n}Pf(X_i)-\mathbb{E}\left[\sum_{i\in\mathbb{T}_n}Pf(X_i)\right]\right)\right)\right]\right)^{1/2}$$
$$\le\exp\left(\frac{2^3t^2C(1+q)^2\|f\|_{\mathrm{Lip}}^2}{2|\mathbb{T}_n|^2}\sum_{k=0}^{n}\left(\sum_{l=0}^{k}s^l\right)^2 2^{n-k-1}\right).$$

The results follows by easy analysis and this ends the proof. □

3.3. Deviation inequalities towards the invariant measure of the randomly drawn chain

All the previous results do not assume any "stability" of the Markov chain on the binary tree, whereas for usual asymptotic theorem the convergence is towards mean of the function with respect to the invariant probability measure of the random lineage chain. To reinforce this asymptotic result by nonasymptotic deviation inequality, it is thus fundamental to be able to replace for example $\mathbb{E}(\overline{M}_{\mathbb{T}_n}(f))$ by some asymptotic quantity. This random lineage chain is a Markov chain with transition kernel $Q=(P_0+P_1)/2$. We shall now suppose the existence of a probability measure π such that $\pi Q=\pi$. We will consider a slight modification of our main assumption and as we are mainly interested in concentration inequalities, let us focus in the $p=1$ case:

***Assumption 3.7** ($H_1'(C)$).* *There exist $q, q_1, q_0>0$, $q_0+q_1<2$ such that*:

(a) $\nu\in T_1(C)$;
(b) $P_b(x,\cdot)\in T_1(C)$, $\forall x\in E$, $b=0,1$;
(c) $W_1^{d_{l_1}}(P(x,\cdot,\cdot),P(\tilde{x},\cdot,\cdot))\le qd(x,\tilde{x})$, $\forall x,\tilde{x}\in E$.
(c′) $W_1^d(P_b(x,\cdot),P_b(\tilde{x},\cdot))\le q_bd(x,\tilde{x})$, $\forall x,\tilde{x}\in E$, $b=0,1$.

Remark 3.2. We have already remarked (see Proposition 2.2) that (c) implies (c′) but we require moreover that $q_0+q_1<2$.

Under this assumption, using the convexity of W_1 (see [35]), we easily see that

$$W_1^d\big(Q(x,\cdot),Q(\tilde{x},\cdot)\big)\le\frac{q_0+q_1}{2}d(x,\tilde{x}),\qquad\forall x,\tilde{x}$$

ensuring the strict contraction of Q, and then the exponential convergence towards π in Wasserstein distance, namely (assuming that π has a first moment)

$$W_1^d\big(Q^n(x,\cdot),\pi\big) \le \left(\frac{q_0+q_1}{2}\right)^n \int d(x,y)\pi(dy).$$

Let us show that we may now control easily the distance between $\mathbb{E}(\overline{M}_{\mathbb{T}_n}(f))$ and $\pi(f)$. Indeed, we may first remark that

$$\mathbb{E}\bigg(\sum_{k\in\mathbb{G}_n} f(X_k)\bigg) = \nu(P_0+P_1)^n f$$

so that assuming that f is 1-lipschitzian, and by the dual version of the Wasserstein distance

$$\begin{aligned}
\big|\mathbb{E}\big(\overline{M}_{\mathbb{T}_n}(f)\big)-\pi(f)\big| &= \frac{1}{|\mathbb{T}_n|}\left|\sum_{j=1}^n\bigg(\sum_{k\in\mathbb{G}_j}\big(f(X_k)-\pi(f)\big)\bigg)\right| \\
&= \frac{1}{|\mathbb{T}_n|}\left|\sum_{j=1}^n 2^j\nu\left(\frac{P_0+P_1}{2}\right)^j\big(f-\pi(f)\big)\right| \\
&\le \frac{1}{|\mathbb{T}_n|}\sum_{j=1}^n 2^j W_1^d\big(\nu Q^j,\pi\big) \\
&\le \frac{1}{|\mathbb{T}_n|}\sum_{j=1}^n (q_0+q_1)^j \\
&\le c_n := \begin{cases} c\left(\dfrac{q_0+q_1}{2}\right)^{n+1}, & \text{if } q_0+q_1\neq 1, \\ c\dfrac{n}{2^{n+1}}, & \text{if } q_0+q_1=1 \end{cases}
\end{aligned}$$

for some universal constant c. The sequence c_n goes to 0 exponentially fast as soon as $q_0+q_1<2$ which was assumed in $H_1'(C)$. We may then see that for $r>c_n$

$$\mathbb{P}\big(\overline{M}_{\mathbb{T}_n}(f)-\pi(f)>r\big) \le \mathbb{P}\big(\overline{M}_{\mathbb{T}_n}(f)-\mathbb{E}\big(\overline{M}_{\mathbb{T}_n}(f)\big)>r-c_n\big)$$

and one then applies the result of the previous subsection.

4. Application to nonlinear bifurcating autoregressive models

The setting will be here the case of the nonlinear bifurcating autoregressive models. It has been considered as a particular realistic model to study cell aging [34], and the asymptotic behavior of parametric estimators as well as nonparametric estimators has been considered in an important series of work, see, for example, [1–4,6,14–16,25,27,28] (and for example, in the random coefficient setting in [13]).

We will then consider the following simplified model where $E = \mathbb{R}$ and $\mathcal{L}(X_1) = \mu_0$ satisfies T_1. We recursively define on the binary tree as before

$$\begin{cases} X_{2k} = f_0(X_k) + \varepsilon_{2k}; \\ X_{2k+1} = f_1(X_k) + \varepsilon_{2k+1}, \end{cases} \tag{4.1}$$

with the following assumptions:

Assumption 4.1 (NL). *f_0 and f_1 are Lipschitz continuous functions.*

Assumption 4.2 (No). *$(\varepsilon_k)_{k\geq 1}$ are centered i.i.d.r.v. and for all $k \geq 0$, ε_k have law μ_ε and satisfy for some positive constant δ_ε, $\mu_\varepsilon(e^{\delta_\varepsilon x^2}) < \infty$. Equivalently, μ_ε satisfies $T_1(C_\varepsilon)$.*

It is then easy to deduce that under these two assumptions, we perfectly match the previous framework. Denoting P_0 and P_1 as previously, we see that H_1' is verified, with the additional fact that $P = P_0 \otimes P_1$. We will do the proof for P_0, being the same for P_1. The conclusion follows for P by conditional independence of X_{2k} and X_{2k+1}. Let us first prove that $P_0(x, \cdot)$ satisfies T_1. Indeed $P_0(x, \cdot)$ is the law of $f_0(x) + \varepsilon_{2k}$, and we have thus to verify the Gaussian integrability property of Theorem 1.5. To this end, consider $x_0 = f(x)$, and choose δ_ε of condition (No) to verify the Gaussian integrability property. We have thus that P_0 satisfies $T_1(C_P)$.

We prove now the Wasserstein contraction property. $P_0(x, \cdot)$ is of course the law of $f_0(x) + \varepsilon_k$. Here ε_k denotes a generic random variable and thus the law of $P_0(y, \cdot)$ is the law of $f_0(y) + \varepsilon_k$. An upper bound of the Wasserstein distance between $P_0(x, \cdot)$ and $P_0(y, \cdot)$ can thus be obtained using a proper coupling where we choose the same noise ε_k for the realization of the two marginal laws. Let f be any Lipschitz function such that $\|f\|_{\mathrm{Lip}} \leq 1$

$$\begin{aligned} \left| \int_S f(z) P_0(x, dz) - \int_S f(z) P_0(y, dz) \right| &= \mathbb{E}\big[f\big(f_0(x) + \varepsilon_1\big) - f\big(f_0(y) + \varepsilon_1\big)\big] \\ &\leq \|f\|_{\mathrm{Lip}} \big| f_0(x) - f_0(y) \big|. \end{aligned}$$

By the Monge–Kantorovitch duality expression of the Wasserstein distance (see, for example, [35], Chapter 6), one has then

$$W_1\big(P_0(x, \cdot), P_0(y, \cdot)\big) \leq \big| f_0(x) - f_0(y) \big| \leq \|f_0\|_{\mathrm{Lip}} |x - y|.$$

Thus under (NL) and (No), our model fits the framework of the previous section with $q = \|f_0\|_{\mathrm{Lip}} + \|f_1\|_{\mathrm{Lip}}$, $q_0 = \|f_0\|_{\mathrm{Lip}}$ and $q_1 = \|f_1\|_{\mathrm{Lip}}$. We stress that Assumption 4.2 is valid in particular for Gaussian and bounded random variables.

We will be interested here in the nonparametric estimation of the autoregression functions f_0 and f_1, and we will use Nadaraya–Watson kernel type estimator, as considered in [8]. Let K be a kernel satisfying the following assumption.

Assumption 4.3 (Ker). *The function K is nonnegative, has compact support $[-R, R]$, is Lipschitz continuous with constant $\|K\|_{\mathrm{Lip}}$ and such that $\int K(z)\,dz = 1$.*

Let us also introduce as usual a bandwidth h_n which will be taken to simplify as $h_n := |\mathbb{T}_n|^{-\alpha}$ for some $0 < \alpha < 1$. The Nadaraya–Watson estimators are then defined for $x \in \mathbb{R}$ as

$$\widehat{f}_{0,n}(x) := \frac{\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K(\frac{X_k-x}{h_n})X_{2k}}{\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K(\frac{X_k-x}{h_n})},$$

$$\widehat{f}_{1,n}(x) := \frac{\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K(\frac{X_k-x}{h_n})X_{2k+1}}{\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K(\frac{X_k-x}{h_n})}.$$

Let us focus on f_0, as it will be exactly the same for f_1 and fix $x \in \mathbb{R}$. We will be interested here in deviation inequalities of $\widehat{f}_{0,n}(x)$ with respect to $f_0(x)$. One has to face two problems. First, it is an autonormalized estimator. It will be dealt with considering deviation inequalities for the numerator and denominator separately before reunite them. Second, $(x, y) \to K(x)y$ is in fact not Lipschitzian in general state space, so that the result of the previous section for deviation inequalities for Lipschitzian function of ancestor-offspring may not be applied directly. Let us tackle this problem. By definition

$$\widehat{f}_{0,n}(x) - f_0(x) = \frac{\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K(\frac{X_k-x}{h_n})[f_0(X_k) - f_0(x) + \varepsilon_{2k}]}{\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K(\frac{X_k-x}{h_n})}$$
$$:= \frac{N_n + M_n}{D_n},$$

where

$$N_n := \sum_{k\in\mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right)\left[f_0(X_k) - f_0(x)\right];$$

$$M_n := \sum_{k\in\mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right)\varepsilon_{2k};$$

$$D_n := \sum_{k\in\mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right).$$

Denote also $\tilde{N}_n = N_n/(|\mathbb{T}_n|h_n)$, $\tilde{M}_n = M_n/(|\mathbb{T}_n|h_n)$, $\tilde{D}_n = D_n/(|\mathbb{T}_n|h_n)$. Let us remark that D_n and M_n completely enter the framework of Proposition 3.5. We may thus prove

Proposition 4.4. *Let us assume that (NL), (No) and (Ker) holds, and $q = \|f_0\|_{\mathrm{Lip}} + \|f_1\|_{\mathrm{Lip}} < \sqrt{2}$. Let us also suppose that $\alpha < 1/4$. Then for all $r > 0$ such that $r > \mathbb{E}(\tilde{N}_n)/\mathbb{E}(\tilde{D}_n)$, there exists* *constants $C, C', C'' > 0$ such that*

$$\mathbb{P}\big(\big|\widehat{f}_{0,n}(x) - f_0(x)\big| > r\big) \leq 2\exp\big(-C\big(r\mathbb{E}(\tilde{D}_n) - \mathbb{E}(\tilde{N}_n)\big)^2|\mathbb{T}_n|h_n^2\big) + 2\exp\left(-C'\frac{(r\mathbb{E}(\tilde{D}_n) - \mathbb{E}(\tilde{N}_n))^2|\mathbb{T}_n|h_n^2}{1 + C''\frac{r^2}{h_n^2}}\right).$$

Proof. Remark first that, by (Ker), K is Lipschitz continuous so that $y \to K(\frac{y-x}{h_n})$ is also Lipschitzian with constant $\|K\|_{\text{Lip}}/h_n$. The mapping

$$H : y \to K\left(\frac{y-x}{h_n}\right)\big(f_0(y) - f_0(x)\big),$$

as K has a compact support and f_0 is Lipschitzian, is also Lipschitzian with constant

$$R\|K\|_{\text{Lip}}\|f_0\|_{\text{Lip}} + \|f_0\|_{\text{Lip}}\|K\|_\infty.$$

Indeed, let y and z be two reals numbers. Since the support of K is $[-R, R]$, we can assume that $|y-x| \leq Rh_n$ and $|z-x| \leq Rh_n$. Now we have

$$\left|K\left(\frac{y-x}{h_n}\right)\big(f_0(y) - f_0(x)\big) - K\left(\frac{z-x}{h_n}\right)\big(f_0(z) - f_0(x)\big)\right| \leq \left|K\left(\frac{y-x}{h_n}\right)\right|\big|f_0(y) - f_0(z)\big| + \big|f_0(z) - f_0(x)\big|\left|K\left(\frac{y-x}{h_n}\right) - K\left(\frac{z-x}{h_n}\right)\right|.$$

By the Lipschitzianity of f_0, we have

$$\left|K\left(\frac{y-x}{h_n}\right)\right|\big|f_0(y) - f_0(z)\big| \leq \|f_0\|_{\text{Lip}}\|K\|_\infty|y-z|.$$

Next, by the Lipschitzianity of f_0 and K, and since $|z-x| \leq Rh_n$, we have

$$\big|f_0(z) - f_0(x)\big|\left|K\left(\frac{y-x}{h_n}\right) - K\left(\frac{z-x}{h_n}\right)\right| \leq R\|K\|_{\text{Lip}}\|f_0\|_{\text{Lip}}.$$

It therefore follows that H is a Lipschitzian function. We can then use Proposition 3.5 to get deviation inequalities for D_n. For all positive r there exists a constant L (explicitly given through Proposition 3.5), such that

$$\mathbb{P}\big(\big|D_n - \mathbb{E}(D_n)\big| > r|\mathbb{T}_n|h_n\big) \leq 2\exp\big(-Lr^2|\mathbb{T}_n|h_n^4/\|K\|_{\text{Lip}}^2\big).$$

For $N_n + M_n$ we cannot directly apply Proposition 3.5 due to the successive dependence of X_k at generation n and ε_{2k} of generation $n-1$. But as we are interested in deviation inequalities, we may split the deviation coming from each term. For N_n, it is once again a simple application of

Proposition 3.5,

$$\mathbb{P}\big(\big|N_n - \mathbb{E}(N_n)\big| > r|\mathbb{T}_n|h_n\big) \le 2\exp\biggl(\frac{-Lr^2|\mathbb{T}_n|h_n^2}{(R\|K\|_{\mathrm{Lip}}\|f_0\|_{\mathrm{Lip}} + \|f_0\|_{\mathrm{Lip}}\|K\|_\infty)^2}\biggr).$$

Note that ε_{2k} is independent of X_k, and centered so that $\mathbb{E}(M_n) = 0$, and satisfies a transportation inequality. Note also that K is bounded. By simple conditioning argument (as in the proof of Proposition 3.3), we may control the Laplace transform of M_n quite simply. We then have for all positive r

$$\mathbb{P}\big(|M_n| > r|\mathbb{T}_n|h_n\big) \le 2\exp\biggl(-r^2\frac{|\mathbb{T}_n|h_n^2\|K\|_\infty^2}{2C\varepsilon}\biggr).$$

However, we cannot use directly these estimations as the estimator is autonormalized. Instead

$$\begin{aligned}
&\mathbb{P}\big(\widehat{f}_{0,n}(x) - f_0(x) > r\big)\\
&\quad\le \mathbb{P}(\tilde{N}_n + \tilde{M}_n > r\tilde{D}_n)\\
&\quad\le \mathbb{P}\big(\tilde{N}_n - \mathbb{E}(\tilde{N}_n) - r\big(\tilde{D}_n - \mathbb{E}(\tilde{D}_n)\big) + \tilde{M}_n > r\mathbb{E}(\tilde{D}_n) - \mathbb{E}(\tilde{N}_n)\big)\\
&\quad\le \mathbb{P}\big(\tilde{N}_n - \mathbb{E}(\tilde{N}_n) - r\big(\tilde{D}_n - \mathbb{E}(\tilde{D}_n)\big) > \big(r\mathbb{E}(\tilde{D}_n) - \mathbb{E}(\tilde{N}_n)\big)/2\big)\\
&\qquad + \mathbb{P}\big(\tilde{M}_n > \big(r\mathbb{E}(\tilde{D}_n) - \mathbb{E}(\tilde{N}_n)\big)/2\big).
\end{aligned}$$

Remark now to conclude that $K((y-x)/h_n)(f(y) - f(x)) + K((y-x)/h_n)$ is

$$\big(R\|K\|_{\mathrm{Lip}}\|f_0\|_{\mathrm{Lip}} + \|f_0\|_{\mathrm{Lip}}\|K\|_\infty + \|K\|_{\mathrm{Lip}}/h_n\big)\text{-Lipschitzian},$$

and we may then proceed as before. □

Remark 4.1. In order to get fully practical deviation inequalities, let us remark that

$$\mathbb{E}[\tilde{D}_n] = \frac{1}{|\mathbb{T}_n|h_n}\sum_{m=0}^{n} 2^m \mu_0 Q^m H \underset{n\to+\infty}{\longrightarrow} \nu(x),$$

where $H(y) = K((y-x)/h_n)$, $\nu(\cdot)$ is the invariant density of the Markov chain associated to a random lineage and

$$\mathbb{E}[\tilde{N}_n] = \frac{1}{|\mathbb{T}_n|h_n}\sum_{m=0}^{n} 2^m \big(\mu_0 Q^m(Hf_0) - f_0(x)\mu_0 Q^m H\big) \underset{n\to+\infty}{\longrightarrow} 0.$$

We refer to [8] for quantitative versions of these limits.

Remark 4.2. Of course this nonparametric estimation is in some sense incomplete, as we would have liked to consider a deviation inequality for $\sup_x |\hat{f}_{0,n}(x) - f_0(x)|$. The problem is somewhat much more complicated here, as the estimator is self normalized. However, it is a crucial problem that we will consider in the near future. For some ideas which could be useful here, let us cite the results of [10] for (uniform) deviation inequalities for estimators of density in the i.i.d. case, and to [19] for control of the Wasserstein distance of the empirical measure of i.i.d.r.v. or of Markov chains.

Remark 4.3 (Estimation of the $\mathbb{T}$-transition probability). We assume that the process has as initial law, the invariant probability ν. We denote by f the density of (X_1, X_2, X_3). For the estimation of f, we propose the estimator $\widehat{f}_n$ defined by

$$\widehat{f}_n(x, y, z) = \frac{1}{|\mathbb{T}_n| h_n} \sum_{k \in \mathbb{T}_n} K\left(\frac{x - X_k}{h_n}\right) K\left(\frac{y - X_{2k}}{h_n}\right) K\left(\frac{z - X_{2k+1}}{h_n}\right).$$

An estimator of the $\mathbb{T}$-probability transition is then given by

$$\widehat{P}_n(x, y, z) = \frac{\widehat{f}_n(x, y, z)}{\tilde{D}_n}.$$

For $x, y, z \in \mathbb{R}$, one can observe that the function G defined on $\mathbb{R}^3$ by

$$G(u, v, w) = K\left(\frac{x - u}{h_n}\right) K\left(\frac{y - v}{h_n}\right) K\left(\frac{z - w}{h_n}\right),$$

is Lipschitzian with $\|G\|_{\mathrm{Lip}} \leq (\|K\|_\infty^2 \|K\|_{\mathrm{Lip}})/h_n$. We have

$$\widehat{P}_n(x, y, z) - P(x, y, z) = \frac{\widehat{f}_n(x, y, z) - f(x, y, z)}{\tilde{D}_n} + \frac{f(x, y, z)(\nu(x) - \tilde{D}_n)}{\nu(x) \tilde{D}_n}.$$

Now using the decomposition

$$\widehat{f}_n(x, y, z) - f(x, y, z) = \left(\widehat{f}_n(x, y, z) - \mathbb{E}\left[\widehat{f}_n(x, y, z)\right]\right) + \left(\mathbb{E}\left[\widehat{f}_n(x, y, z)\right] - f(x, y, z)\right),$$

and the convergence of $\mathbb{E}[\widehat{f}_n(x, y, z)]$ to $f(x, y, z)$, we obtain a deviation inequality for $|\widehat{P}_n(x, y, z) - P(x, y, z)|$ similar to that obtained at the Proposition 4.4.

When the density g_ε of $(\varepsilon_2, \varepsilon_3)$ is known, another strategy for the estimation of the $\mathbb{T}$-transition probability is to observe that $P(x, y, z) = g_\varepsilon(y - f_0(x), z - f_1(x))$. An estimator of $P(x, y, z)$ is then given by $\widehat{P}_n(x, y, z) = g_\varepsilon(y - \widehat{f}_{0,n}(x), z - \widehat{f}_{1,n}(x))$ where $\widehat{f}_{0,n}$ and $\widehat{f}_{1,n}$ are estimators defined above. If g_ε is Lipschitzian, we have

$$\left|\widehat{P}_n(x, y, z) - P(x, y, z)\right| \leq \|g_\varepsilon\|_{\mathrm{Lip}} \left(\left|\widehat{f}_{0,n}(x) - f_0(x)\right| + \left|\widehat{f}_{1,n}(x) - f_1(x)\right|\right),$$

and the deviation inequalities for $|\widehat{P}_n(x, y, z) - P(x, y, z)|$ are thus of the same order that those given by the Proposition 4.4.

Acknowledgements

S.V. Bitseki Penda was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH. A. Guillin was supported by the French ANR STAB project. M. Escobar-Bach was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416).

References

[1] Basawa, I.V. and Zhou, J. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* **41A** 55–64. MR2057565

[2] Basawa, I.V. Zhou, J. and (2005). Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **1** 77–88.

[3] Basawa, I.V. Zhou, J. and (2005). Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Series Anal.* **26** 825–842.

[4] Bercu, B., de Saporta, B. and Gégout-Petit, A. (2009). Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* **14** 2492–2526. MR2563249

[5] Bitseki Penda, S.V. (2015). Deviation inequalities for bifurcating Markov chains on Galton–Watson tree. *ESAIM Probab. Stat.* **19** 689–724. MR3433433

[6] Bitseki Penda, S.V. and Djellout, H. (2014). Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 806–844. MR3224290

[7] Bitseki Penda, S.V., Djellout, H. and Guillin, A. (2014). Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Ann. Appl. Probab.* **24** 235–291. MR3161647

[8] Bitseki Penda, S.V. and Olivier, A. (2016). Autoregressive functions estimation in nonlinear bifurcating autoregressive models. Preprint. Available at arXiv:1506.01842.

[9] Bobkov, S.G. and Götze, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772

[10] Bolley, F., Guillin, A. and Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137** 541–593. MR2280433

[11] Cattiaux, P. and Guillin, A. (2006). On quadratic transportation cost inequalities. *J. Math. Pures Appl.* (9) **86** 341–361. MR2257848

[12] Cattiaux, P., Guillin, A. and Wu, L.-M. (2010). A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields* **148** 285–304. MR2653230

[13] de Saporta, B., Gégout-Petit, A. and Marsalle, L. (2014). Random coefficients bifurcating autoregressive processes. *ESAIM Probab. Stat.* **18** 365–399. MR3333995

[14] Delmas, J.-F. and Marsalle, L. (2010). Detection of cellular aging in a Galton–Watson process. *Stochastic Process. Appl.* **120** 2495–2519. MR2728175

[15] De Saporta, B., Gégout-Petit, A. and Marsalle, L. (2011). Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* **5** 1313–1353. MR2842907

[16] De Saporta, B., Gégout-Petit, A. and Marsalle, L. (2012). Asymmetry tests for bifurcating autoregressive processes with missing data. *Statist. Probab. Lett.* **82** 1439–1444. MR2929798

[17] Djellout, H., Guillin, A. and Wu, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** 2702–2732. MR2078555

[18] Doumic, M., Hoffmann, M., Krell, N. and Robert, L. (2015). Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli* **21** 1760–1799. MR3352060

[19] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341
[20] Gozlan, N. (2006). Integral criteria for transportation-cost inequalities. *Electron. Commun. Probab.* **11** 64–77 (electronic). MR2231734
[21] Gozlan, N. (2009). A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.* **37** 2480–2498. MR2573565
[22] Gozlan, N. and Léonard, C. (2007). A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** 235–283. MR2322697
[23] Gozlan, N. and Léonard, C. (2010). Transport inequalities. A survey. *Markov Process. Related Fields* **16** 635–736. MR2895086
[24] Gozlan, N., Roberto, C. and Samson, P.-M. (2011). A new characterization of Talagrand's transport-entropy inequalities and applications. *Ann. Probab.* **39** 857–880. MR2789577
[25] Guyon, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** 1538–1569. MR2358633
[26] Guyon, J., Bize, A., Paul, G., Stewart, E., Delmas, J.-F. and Taddéi, F. (2005). Statistical study of cellular aging. In *CEMRACS* 2004 – *Mathematics and Applications to Biology and Medicine*. *ESAIM Proc.* **14** 100–114 (electronic). Les Ulis: EDP Sci. MR2226805
[27] Huggins, R.M. and Basawa, I.V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36** 1225–1233. MR1746406
[28] Huggins, R.M. and Basawa, I.V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42** 423–432. MR1802966
[29] Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347
[30] Marton, K. (1986). A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory* **32** 445–446. MR0838213
[31] Marton, K. (1996). Bounding $\overline{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** 857–866. MR1404531
[32] Marton, K. (1996). A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** 556–571. MR1392329
[33] Massart, P. (2007). *Concentration Inequalities and Model Selection*. *Lecture Notes in Math.* **1896**. Berlin: Springer. MR2319879
[34] Stewart, E.J., Madden, R., Paul, G. and Taddéi, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol* **3** 2:e45.
[35] Villani, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Berlin: Springer. MR2459454

Received September 2015