



**HAL**  
open science

## Transportation cost-information and concentration inequalities for bifurcating Markov chains

Siméon Valère Bitseki Penda, Mikael Escobar-Bach, Arnaud Guillin

► **To cite this version:**

Siméon Valère Bitseki Penda, Mikael Escobar-Bach, Arnaud Guillin. Transportation cost-information and concentration inequalities for bifurcating Markov chains. *Bernoulli*, 2017, 23 (4B), pp.3213-3242. 10.3150/16-BEJ843 . hal-01109760

**HAL Id: hal-01109760**

**<https://hal.science/hal-01109760>**

Submitted on 26 Jan 2015

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

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

# Transportation cost-information and concentration inequalities for bifurcating Markov chains

S. Valère Bitseki Penda\*      Mikael Escobar-Bach †

Arnaud Guillin ‡

January 27, 2015

## Abstract

We investigate the transportation cost-information inequalities for bifurcating Markov chains which are a class of processes indexed by binary tree. These processes provide models for cell growth when each individual in one generation gives birth to two offsprings in the next one. Transportation cost inequalities provide useful concentration inequalities. We also study deviation inequalities for the empirical means under relaxed assumptions on the Wasserstein contraction of the Markov kernels. Applications to bifurcating non linear autoregressive processes are considered: deviation inequalities for pointwise estimates of the non linear leading functions.

*Keywords:* Transportation cost-information inequalities, Wasserstein distance, bifurcating Markov chains, deviation inequalities, geometric ergodicity.

## 1 Introduction

Roughly speaking, a bifurcating Markov chain is a Markov chain indexed by a binary regular tree. This class of processes are well adapted for the study of populations where each individual in one generation gives birth

---

\*CMAP, Ecole polytechnique route de Saclay, 91128, Palaiseau, France. **email:** bitseki@map.polytechnique.fr

†Department of Mathematics and Computer Science, University of Southern Denmark, Campus 55, 5230 Odense M, Denmark **email:**escobar@sdu.dk

‡Institut Universitaire de France et Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, 24 avenue des Landais, BP 80026, 63177 Aubière, France. **email:** guillin@math.univ-bpclermont.fr

to two offsprings in the next one. They were introduced by Guyon [29] in order to study the *Escherichia coli* aging process. Namely, when a cell divides into two offsprings, are the genetical traits identical for the two daughter cells? Recently, several models of bifurcating Markov chains, or models using the theory of bifurcating Markov chains, for example under the form of bifurcating autoregressive processes, have been studied [1, 2, 29, 21, 18], showing that these processes are of great importance to analysis of cell division. There is now an important literature covering asymptotic theorems for bifurcating Markov chains such as Law of Large Numbers, Central Limit Theorems, Moderate Deviation Principle, Law of Iterated Logarithm, see for example [29, 30, 6, 19, 15, 18, 8] for recent references. These limit theorems are particularly useful when applied to the statistics of the bifurcating processes, enabling to provide efficient tests to assert if the aging of the cell is different for the two offsprings (see [30] for real case study). Of course, these limit theorems may be considered only in the "ergodic" case, i.e. when the law of the random lineage chain has an unique invariant measure.

However, limit theorems are only asymptotical results and one is often faced to study only datas with a size limited population. It is thus very natural to control the statistics non asymptotically. Such deviation inequalities (or concentration inequalities) have been recently the subject of many studies and we refer to the books of Ledoux [31] and Massart [35] for nice introductions on the subject, developing both i.i.d. case and dependent case with a wide variety of tools (Laplace controls, functional inequalities, Efron-Stein,...). It was one of the goal of Bitseki et al. [8] to investigate deviation inequalities for additive functionals of bifurcating Markov chain. In their work, one of the main hypothesis is that the Markov chain associated to a random lineage of the population is uniformly geometrically ergodic. 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, our aim is to obtain deviation inequalities for bifurcating Markov chain when the auxiliary Markov chains may satisfy some contraction properties in Wasserstein distance, and some (uniform) integrability property. This will be done with the help of transportation cost-information inequalities and direct Laplace controls. In order to present our result, we may now define properly the model of bifurcating Markov chains.

## 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  $r \in \mathbb{N}$ , let

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

which denote respectively the  $r$ -th column and the first  $(r + 1)$  columns of the tree. The whole tree is thus defined by

$$\mathbb{T} = \bigcup_{r=0}^{\infty} \mathbb{G}_r.$$

A column of a given vertex  $n$  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  $x$ .

In the sequel, we will see  $\mathbb{T}$  as a given population in which each individual in one generation gives birth to two offsprings in the next one. This will make easier the introduction of different notions. The vertex  $n$  will denote the individual  $n$  and the ancestor of individuals  $2n$  and  $2n + 1$ . The individuals who belong to  $2\mathbb{N}$  (resp.  $2\mathbb{N}+1$ ) will be called individual of type 0 (resp. type 1). The column  $\mathbb{G}_r$  and the first  $(r + 1)$  columns  $\mathbb{T}_r$  will denote respectively the  $r$ -th generation and the first  $(r + 1)$  generations. The initial individual will be denoted 1.

For each individual  $n$ , we look into a random variable  $X_n$ , defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and which takes its values in a metric space  $(E, d)$  endowed with its Borel  $\sigma$ -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 ([29])). *We call  $\mathbb{T}$ -transition probability any mapping  $P : E \times \mathcal{E}^2 \rightarrow [0, 1]$  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  $x, y, z \in E$ ,  $P(x, dy, dz)$  denotes the probability that the couple of the quantities associated with the children are in the neighbourhood of  $y$  and  $z$  given that the quantity associated with their mother is  $x$ .

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

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

We are now in position to give a precise definition of bifurcating Markov chain.

**Definition 1.2** (Bifurcating Markov Chains, see ([29])). *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  $\nu$  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  $\nu$  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 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 unprecised, the filtration implicitly used will be  $\mathcal{F}_r = \sigma(X_i, i \in \mathbb{T}_r)$ .

*Remark 1.3.* We may of course also consider in this work bifurcating Markov chains on a  $a$ -ary tree (with  $a \geq 2$ ) with no additional technicalities, but heavy additional notations. In the same spirit, Markov chains of higher order (such as BAR processes considered in [7]) could be handled by the same techniques. A non trivial extension would be the case of bifurcating Markov chains on a Galton-Watson tree (see for example [10] under very strong assumptions), that we will consider elsewhere.

## 1.2 Transportation cost-information inequality

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

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

where the infimum is taken over all probability measures  $\pi$  on the product space  $E \times E$  with marginal distributions  $\mu$  and  $\nu$  (say, coupling of  $(\mu, \nu)$ ). This infimum is finite as soon as  $\mu$  and  $\nu$  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  $\nu$  with respect to  $\mu$  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 1.4** ( $L^p$ -transportation cost-inequality). *We say that the probability measure  $\mu$  satisfies the  $L^p$ -transportation cost-information inequality on  $(E, d)$  (and we write  $\mu \in T_p(C)$ ) if there is some constant  $C > 0$  such that for any probability measure  $\nu$ ,*

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

This transportation inequality have been introduced by Marton [32, 33] as a tool for (Gaussian) concentration of measure property. The following result will be crucial in the sequel. It gives a characterization of  $L^1$ -transportation cost-inequality in term of concentration inequality. It is of course one of the main tool to get deviation inequalities (via Markov inequality).

**Theorem 1.5** ([11]).  *$\mu$  satisfies the  $L^1$ -transportation cost-information inequality (say  $T_1$ ) on  $(E, d)$  with constant  $C > 0$ , that is,  $\mu \in T_1(C)$ , if and only if for any Lipschitzian function  $F : (E, d) \rightarrow \mathbb{R}$ ,  $F$  is  $\mu$ -integrable and*

$$\int_E \exp(\lambda(F - \langle F \rangle_\mu)) d\mu \leq \exp\left(\frac{\lambda^2}{2} C \|F\|_{Lip}^2\right) \quad \forall \lambda \in \mathbb{R},$$

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

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

In particular, we have the concentration inequality

$$\mu(F - \langle F \rangle_\mu \leq -t) \vee \mu(F - \langle F \rangle_\mu \geq t) \leq \exp\left(-\frac{t^2}{2C\|F\|_{Lip}^2}\right) \quad \forall t \in \mathbb{R}.$$

In this work we will focus on transportation inequality  $T_1$  mainly. There is now a considerable literature around these transportation inequalities. As a flavor, let us cite first the characterization of  $T_1$  as a Gaussian integrability property [20] (see also [23]).

**Theorem 1.6** ([20]).  *$\mu$  satisfies the  $L^1$ -transportation cost-information inequality (say  $T_1$ ) on  $(E, d)$  if and only if there exists  $\delta > 0$  and  $x_0 \in E$  such that*

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

*and the constant of the Transportation inequality can be made explicit.*

There is also a large deviations characterization [25]. Recent striking results on transportation inequalities have been obtained for  $T_2$ , namely that they are equivalent to dimension free Gaussian concentration [24], or to a restricted class of logarithmic Sobolev inequalities [27]. See also [13] or [14] for practical criterion based on Lyapunov type criterion and we refer for example to [26] or [37] for surveys on transportation inequality. One of the main aspect of transportation inequality is their tensorization property, i.e.  $\mu^{\otimes n}$  will satisfy some transportation measure if  $\mu$  does (with dependence on the dimension  $n^{2/p-1}$ ). One important development was to consider such a property for dependent sequences such as Markov chains. In [20], Djellout et al., generalizing result of Marton [34], have provided conditions under which the law of a homogeneous Markov chain  $(Y_k)_{1 \leq k \leq n}$  on  $E^n$  satisfies the  $L^p$ -transportation cost-information inequality  $T_p$  with respect to the metric

$$d_{l_p}(x, y) := \left(\sum_{i=1}^n d(x_i, y_i)^p\right)^{1/p}.$$

We will follow similar ideas here to establish the  $L^p$ -transportation cost-information inequality for the law of a bifurcating Markov chain  $(X_i)_{1 \leq i \leq N}$  on  $E^N$ . This will allow us to obtain concentration inequalities for bifurcating Markov chains under hypotheses largely weaker than those of Bitseki et al. [8]. It would also be tempting to generalize the approach of [28] to Markov chains and bifurcating Markov chains to get directly deviation inequalities for Markov chains, w.r.t. the invariant measure. However it would need to restrict to reversible Markov chains and thus not directly suited to bifurcating Markov chains and would thus require new ideas.

*Remark 1.7.* There are natural generalizations of the  $T_1$  inequality often denoted  $\alpha - T_1$  inequality, where  $\alpha$  is a non negative convex lower semi continuous function vanishing at 0. We say that the probability measure  $\mu$  satisfies  $\alpha - T_p(C)$  if for any probability measure  $\nu$

$$\alpha(W_1(\nu, \mu)) \leq 2C H(\nu/\mu).$$

The usual  $T_1$  inequality is then the case where  $\alpha(t) = t^2$ . Gozlan [23] has generalized Bobkov-Götze's Laplace transform control [11] and Djellout-Guillin-Wu [20] integrability criterion to this setting enabling to recover sub or super Gaussian concentration. The result 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.

## 2 Transportation cost-information inequalities for bifurcating Markov chains

Let  $(X_i, i \in \mathbb{T})$  be a bifurcating Markov chain on  $E$  with  $\mathbb{T}$ -probability transition  $P$  and initial measure  $\nu$ . 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**  $(H_p(C))$ .

- (a)  $\nu \in T_p(C)$ ;
- (b)  $P(x, \cdot, \cdot) \in T_p(C), \forall x \in E$  ;
- (c)  $W_p^d(P(x, \cdot, \cdot), P(\tilde{x}, \cdot, \cdot)) \leq q d(x, \tilde{x}), \forall x, \tilde{x} \in E$  and some  $q > 0$ .

It is important to remark that under  $(H_p(C)), (c)$  we have that there exists  $r_0$  and  $r_1$  smaller than  $q$  such that for  $b = 0, 1$

$$W_p^d(P_b(x, \cdot), P_b(\tilde{x}, \cdot)) \leq r_b d(x, \tilde{x}), \quad \forall x, \tilde{x} \in E.$$

Note also that when  $P(x, dy, dz) = P_0(x, dy)P_1(x, dz)$ , then these last two stability results in Wasserstein contraction implies  $(H_p(C)), (c)$  with  $q \leq (r_0^p + r_1^p)^{1/p}$  (using trivial coupling). We may remark also that by  $(H_p(C)), (b)$ ,  $P_0$  and  $P_1$  also satisfies (uniformly) a transportation inequality. Let us note that thanks to the Hölder inequality,  $(H_p(C))$  implies  $(H_1(C))$ .

We do not suppose here that  $q, r_0$  and  $r_1$  are strictly less than 1, and thus the two marginal chains, as well as the bifurcating one, are not a priori



contractions. We are thus considering here both "stable" and "unstable" cases.

We then have the following result for the law of the whole trajectory on the binary tree.

**Theorem 2.2.** *Let  $n \in \mathbb{N}$  and let  $\mathcal{P}$  be the law of  $(X_i)_{1 \leq i \leq \mathbb{T}_n}$  and denote  $N = |\mathbb{T}_n|$ . We assume Assumption 2.1 for  $1 \leq p \leq 2$ . Then  $\mathcal{P} \in T_p(C_N)$  where*

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

Before the proof of this result, let us make the following notations. For a Polish space  $\chi$ , we denote by  $\mathcal{M}_1(\chi)$  the space of probability measures on  $\chi$ . For  $x \in E^N$ ,  $x^i := (x_1, \dots, x_i)$ . For  $\mu \in \mathcal{M}_1(E^N)$ , let  $(x_1, \dots, x_N) \in E^N$  be distributed randomly according to  $\mu$ . We denote by  $\mu^i$  the law of  $x^{2i+1}$ , and by  $\mu_{x^{2i-1}}^i$  the conditional law of  $(x_{2i}, x_{2i+1})$  given  $x^{2i-1}$  with the convention  $\mu_{x^0}^1 = \mu^1$ , where  $x^0 = x_0$  is some fixed point. In particular, if  $\mu$  is the law of a bifurcating Markov chain with  $\mathbb{T}$ -probability transition  $P$ , then  $\mu_{x^{2i-1}}^i = P(x_i, \cdot, \cdot)$ .

For the convenience of the readers, we recall the formula of additivity of entropy (see for e.g. [37], Lemma 22.8).

**Lemma 2.3.** *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*

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

where  $\mathcal{P}_{x^{i-1}}^i$  and  $\mathcal{Q}_{x^{i-1}}^i$  are defined in the same way as above.

We can now prove the Theorem.

*Proof of the Theorem 2.2.* Let  $\mathcal{Q} \in \mathcal{M}_1(E^N)$ . Assume that  $H(\mathcal{Q}|\mathcal{P}) < \infty$  (trivial otherwise). Let  $\varepsilon > 0$ . The idea is of course to do a conditionnement with respect to the previous generation, i.e. to  $\mathbb{G}_{n-1}$  but we will do it sequentially by pairs. Conditionally to their ancestors, every pair of offspring of an individual is independent of the offspring of the other individuals for 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  $\pi_{y^{2i-1}, x^{2i-1}}^i$  of  $(\mathcal{Q}_{y^{2i-1}}^i, \mathcal{P}_{x^{2i-1}}^i)$  such that

$$\begin{aligned}
\mathcal{A}_i &:= \int (d(y_{2i}, x_{2i})^p + d(y_{2i+1}, x_{2i+1})^p) d\pi_{y^{2i-1}, x^{2i-1}}^i \\
&\leq (1 + \epsilon) W_p^d \left( \mathcal{Q}_{y^{2i-1}}^i, \mathcal{P}_{x^{2i-1}}^i \right)^p \\
&\leq (1 + \epsilon) \left[ W_p^d \left( \mathcal{Q}_{y^{2i-1}}^i, \mathcal{P}_{y^{2i-1}}^i \right) + W_p^d \left( \mathcal{P}_{y^{2i-1}}^i, \mathcal{P}_{x^{2i-1}}^i \right) \right]^p \\
&\leq (1 + \epsilon) \left[ W_p^d \left( \mathcal{Q}_{y^{2i-1}}^i, P(y_i, \cdot, \cdot) \right) + W_p^d \left( P(y_i, \cdot, \cdot), P(x_i, \cdot, \cdot) \right) \right]^p,
\end{aligned}$$

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 Assumption 2.1, and the convexity of the function  $x \mapsto x^p$ , we obtain, for  $a, b > 1$  such that  $1/a + 1/b = 1$ ,

$$\begin{aligned}
\mathcal{A}_i &\leq (1 + \epsilon) \left( \sqrt{2CH_i(y^{2i-1})} + qd(y_i, x_i) \right)^p \\
&\leq (1 + \epsilon) \left( a^{p-1} \left( \sqrt{2CH_i(y^{2i-1})} \right)^p + b^{p-1} q^p d^p(y_i, x_i) \right)
\end{aligned}$$

where  $H_i(y^{2i-1}) = H(\mathcal{Q}_{y^{2i-1}}^i | \mathcal{P}_{y^{2i-1}}^i)$ . By recurrence, it leads to the finiteness of  $p$ -moments. Taking the average with respect to the whole law and summing on  $i$ , we obtain

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

Letting  $\epsilon$  goes to  $0^+$ , we are led to

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

Iterating the latter inequality, increasing some terms and thanks to Hölder inequality, we obtain

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

where  $h_i = a^{p-1}(2C)^{p/2}\mathbb{E}[H_i(Y^{i-1})^{p/2}]$ . 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

$$\begin{aligned} W_p^{d_{lp}}(\mathcal{Q}, \mathcal{P})^p &\leq a^{p-1} (2CH(\mathcal{Q}|\mathcal{P}))^{p/2} \left( \sum_{i=1}^N \left( \sum_{j=0}^{N-i} (b^{p-1}q^p)^j \right)^{\frac{2}{2-p}} \right)^{\frac{2-p}{2}} \\ &\leq a^{p-1} (2CH(\mathcal{Q}|\mathcal{P}))^{p/2} N^{1-\frac{p}{2}} \sum_{j=0}^{N-1} (b^{p-1}q^p)^j. \end{aligned}$$

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

*Remark 2.4.* For  $q < 1$ , we then have that the constant  $C_N$  of  $T_1$  inequality for  $\mathcal{P}$  increases linearly on the dimension  $N$ . However, for  $T_2$  this constant is independent of the dimension as in the i.i.d. case.

*Remark 2.5.* As we will see in the next section, still when  $q < 1$ , Theorem 2.2 and Theorem 1.5 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, when they are applied to  $F(X_1, \dots, X_N) = f(X_N)$ , deviation inequalities that we obtain does not furnish the good order of  $N$  when  $N$  is large. 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  $E^3$ . As this last question is important for the  $L^1$ -transportation cost-information inequality of the invariant measure of a bifurcating Markov chain, we give the following results.

**Proposition 2.6.** *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

$$c_n = C \sum_{k=0}^{r_n-1} r_0^{2(k-a_k)} r_1^{2a_k}; \quad a_0 = 0$$

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  $r_n - k + 1$ -th generation and the  $r_n$ -th generation.

Before the proof, we introduce some more notations. Let  $n \in \mathbb{T}$ . We denote by  $(z_1, \dots, z_{r_n}) \in \{0, 1\}^{r_n}$  the unique path from the root 1 to  $n$ . Then, for all  $i \in \{1, \dots, r_n\}$ ,  $z_i$  is the type of the ancestor of  $n$  which is in the  $i$ -th generation and the quantities  $a_k$  defined in the Proposition 2.6 are given by

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

For all  $k \in \{1, \dots, r_n\}$ , we denote by  $P^k$  and  $P^{-k}$  the iterated of the transition probabilities  $P_0$  and  $P_1$  defined by

$$P^k := P_{z_1} \circ \dots \circ P_{z_k} \quad \text{and} \quad P^{-k} := P_{z_{r_n-k}} \circ \dots \circ P_{z_{r_n}}.$$

*Proof of the Proposition 2.6.* First note that since

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

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

$$\|P_b f\|_{Lip} \leq r_b \|f\|_{Lip} \quad \forall b \in \{0, 1\}.$$

Now let  $f$  be a Lipschitzian function defined on  $E$ . By (b)-(c) of  $(H_1(C))$  and Theorem 1.5, we have

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

Once again, applying Theorem 1.5, we obtain

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

By iterating this method, we are led to

$$P^{r_n}(e^f) \leq \exp \left( P^{-r_n+1}f + (1 + r_{z_{r_n}}^2 + r_{z_{r_n}}^2 r_{z_{r_n-1}}^2 + \cdots + \prod_{i=2}^{r_n} r_{z_i}^2) \frac{C \|f\|_{Lip}^2}{2} \right).$$

Since

$$1 + r_{z_{r_n}}^2 + r_{z_{r_n}}^2 r_{z_{r_n-1}}^2 + \cdots + \prod_{i=2}^{r_n} r_{z_i}^2 = \sum_{k=0}^{r_n-1} r_0^{2(k-a_k)} r_1^{2a_k} \quad \text{and} \quad P^{-r_n+1}f = P^{r_n}f,$$

we conclude the proof thanks to Theorem 1.5.  $\square$

The next result is a consequence of the previous Proposition.

**Corollary 2.7.** *Assume  $(H_1(C))$  and  $r := \max\{r_0, r_1\} < 1$ . Then*

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

where

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

*Proof.* That  $\mathcal{L}(X_n | X_1 = x) \in T_1(c_\infty)$  is a direct consequence of Proposition 2.6. It suffices to bound  $r_0$  and  $r_1$  by  $r$ .

In order to deal with the ancestor-offspring case  $(X_n, X_{2n}, X_{2n+1})$ , we do the following remarks.

Let  $f : (E^3, d_{l_1}) \rightarrow \mathbb{R}$  be a Lipschitzian function. We have

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

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

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

and then,

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

We recall that  $X_1 = x$ . We have

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

Now, from  $(H_1(C))$ , the previous remarks and using the same strategy as in the proof of Proposition 2.6, we are led to

$$\begin{aligned} & \mathbb{E}[\exp(f(X_n, X_{2n}, X_{2n+1}))] \\ & \leq \exp\left(P_{z_1} \cdots P_{z_{r_n}} P f(x) + \frac{C\|f\|_{Lip}^2}{2} + \frac{C(1+q)^2\|f\|_{Lip}^2}{2} \sum_{i=0}^{r_n-1} r^{2i}\right). \end{aligned}$$

Since  $P_{z_1} \cdots P_{z_{r_n}} P f(x) = \mathbb{E}[f(X_n, X_{2n}, X_{2n+1})]$  and  $\sum_{i=0}^{r_n-1} r^{2i} \leq 1/(1-r^2)$ , we obtain

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

with  $c'_\infty$  given in the Corollary. We then conclude the proof thanks to Theorem 1.5.  $\square$

### 3 Concentration inequalities for bifurcating Markov chains

#### 3.1 Direct applications of the Theorem 2.2

We are now interested in the concentration inequalities for the additive functionals of bifurcating Markov chains. Specifically, let  $N \in \mathbb{N}^*$  and  $I$  be a subset of  $\{1, \dots, N\}$ . Let  $f$  be a real function on  $E$  or  $E^3$ . We set

$$M_I(f) = \sum_{i \in I} f(\Delta_i)$$

where  $\Delta_i = X_i$  if  $f$  is defined on  $E$  and  $\Delta_i = (X_i, X_{2i}, X_{2i+1})$  if  $f$  is defined on  $E^3$ . We also consider the empirical mean  $\overline{M}_I(f)$  over  $I$  defined by  $\overline{M}_I(f) = (1/|I|)M_I(f)$  where  $|I|$  denotes the cardinality of  $I$ . In the 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 relevant (see for e.g. [8]).

First, we will establish concentration inequalities when  $f$  is a real Lipschitzian function defined on  $E$ . For a subset  $I$  of  $\{1, \dots, N\}$ , let  $F_I$  be the function defined on  $(E^N, d_{l_p})$ ,  $p \geq 1$  by  $F_I(x^N) = 1/(|I|) \sum_{i \in I} f(x_i)$  for all  $x^N \in E^N$ . Then  $F_I$  is also a Lipschitzian function on  $(E^N, d_{l_p})$  and we have  $\|F_I\|_{Lip} \leq |I|^{-1/p} \|f\|_{Lip}$ . The following result is a direct consequence of Theorem 2.2.

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

$$\mathcal{P} \circ F_I^{-1} \in T_p(C_N |I|^{-2/p} \|f\|_{Lip}^2)$$

where  $C_N$  is given in the Theorem 2.2 and  $\mathcal{P} \circ F_I^{-1}$  is the image law of  $\mathcal{P}$  under  $F_I$ . In particular, for all  $t > 0$  we have

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

*Proof.* The first part is a direct consequence of Theorem 2.2 and Lemma 2.1 of [20]. The second part is an application of Theorem 1.5.  $\square$

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

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

We assume that  $N$  is a odd number. Let  $I$  be a subset of  $\{1, \dots, (N-1)/2\}$ . Now, we denote by  $F_I$  the real function defined on  $(E^N, d_{l_p})$  by  $F_I(x^N) = (1/|I|) \sum_{i \in I} f(x_i, x_{2i}, x_{2i+1})$ . For all  $x^N, y^N \in E^N$  we have for some universal constant  $c$

$$\begin{aligned} |F_I(x^N) - F_I(y^N)| &\leq \frac{\|f\|_{Lip}}{|I|} \sum_{i \in I} (d(x_i, y_i) + d(x_{2i}, y_{2i}) + d(x_{2i+1}, y_{2i+1})) \\ &\leq \frac{c\|f\|_{Lip}}{|I|^{1/p}} d_{l_p}(x^N, y^N). \end{aligned}$$

$F_I$  is then a Lipschitzian function on  $(E^N, d_{l_p})$  and  $\|F_I\|_{Lip} \leq c\|f\|_{Lip}/|I|^{1/p}$ . We then have the following result.

**Proposition 3.2.** *Let  $N \in \mathbb{N}^*$  be a odd number and let  $\mathcal{P}$  be the law of  $(X_i)_{1 \leq i \leq N}$ . Let  $f$  be a real Lipschitzian function on  $(E^3, d_{l_1})$ . Then, under  $(H_p(C))$  for  $1 \leq p \leq 2$ ,*

$$\mathcal{P} \circ F_I^{-1} \in T_p(cC_N |I|^{-2/p} \|f\|_{Lip}^2)$$

where  $C_N$  is given in the Theorem 2.2 and  $\mathcal{P} \circ F_I^{-1}$  is the image law of  $\mathcal{P}$  under  $F_I$ . In particular, for all  $t > 0$  we have

$$\begin{aligned} \mathbb{P}(F_I(X^N) \leq -t + \mathbb{E}[F_I(X^N)]) \vee \mathbb{P}(F_I(X^N) \geq t + \mathbb{E}[F_I(X^N)]) \\ \leq \exp\left(-\frac{t^2|I|^{2/p}}{2cC_N\|f\|_{Lip}^2}\right). \end{aligned}$$

*Proof.* The proof is a direct consequence of Theorem 2.2, Lemma 2.1 of [20] and Theorem 1.5.  $\square$

*Remark 3.3.* The previous results applied with  $p = 1$  to the empirical means  $\overline{M}_{\mathbb{G}_n}(f)$  and  $\overline{M}_{\mathbb{T}_n}(f)$  ( $f$  being a real Lipschitzian function) give us relevant concentration inequalities, that is with the good order size of the index set, when  $q < 1$ . For example, for  $\overline{M}_{\mathbb{G}_n}(f)$ , it suffices to take  $N = |\mathbb{T}_n|$  and  $I = \mathbb{G}_n$  in the Propositions 3.1 and 3.2. But for  $q \geq 1$ , the concentration inequalities obtained thanks to these results are not satisfactory. In the sequel, we will be interested in obtaining relevant concentration inequalities for the empirical means  $\overline{M}_{\mathbb{G}_n}(f)$  and  $\overline{M}_{\mathbb{T}_n}(f)$  when  $q \geq 1$ .

### 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  $r = r_0 + r_1$ .

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  $r > 1$ , namely when it is possible that one of the two marginal Markov chains is not a strict contraction. The transportation inequality of Theorem 2.2 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 of only 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, i.e. 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 [29, 16] 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

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

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

$$\mathbb{P}(X - \mathbb{E}(X) \geq r) \leq e^{-\frac{r^2}{2\kappa}}.$$

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

**Proposition 3.4.** *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*

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

We recall that here  $r = r_0 + r_1$ .

*Remark 3.5.* One can observe that for  $r < \sqrt{2}$ , the previous inequalities are on the same order of magnitude that the inequalities obtained thanks to Proposition 3.1 with  $q < 1$ . For  $r < 2$  the above inequalities remain relevant since we just have a negligible loss with respect to  $|\mathbb{G}_n|$ . But for  $r \geq \sqrt{2}$ , these inequalities are not significant (see the same type of limitations at the CLT level in [16]).

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

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

Thanks to the Markov property, we have

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

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 1.5, we have

$$\begin{aligned} \prod_{i \in \mathbb{G}_{n-1}} P \left( \exp (t2^{-n} (f \oplus f - (P_0 + P_1)f)) \right) (X_i) \\ \leq \prod_{i \in \mathbb{G}_{n-1}} \exp \left( \frac{t^2 C \|f \oplus f\|_{Lip}^2}{2 \times 2^{2n}} \right). \end{aligned}$$

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

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

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

$$\|(P_0 + P_1)f \oplus (P_0 + P_1)f\|_{Lip} \leq 2r\|f\|_{Lip},$$

we are led to

$$\begin{aligned} \mathbb{E} \left[ \exp \left( t2^{-n} \sum_{i \in \mathbb{G}_n} f(X_i) \right) \right] \leq \mathbb{E} \left[ \exp \left( t2^{-n} \sum_{i \in \mathbb{G}_{n-2}} (P_0 + P_1)^2 f(X_i) \right) \right] \\ \times \exp \left( \frac{2^2 t^2 C \|f\|_{Lip}^2 2^{n-1}}{2 \times 2^{2n}} \right) \exp \left( \frac{2^2 t^2 C \|f\|_{Lip}^2 r^2 2^{n-2}}{2 \times 2^{2n}} \right). \end{aligned}$$

Iterating this method and using the inequalities

$$\|(P_0 + P_1)^k f \oplus (P_0 + P_1)^k f\|_{Lip} \leq 2r^k \|f\|_{Lip} \quad \forall k \in \{1, \dots, n-1\},$$

we obtain

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

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

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

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

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

and the results of the Proposition then follow from this last inequality.  $\square$

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 3.4.

**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  $\overline{M}_{\mathbb{G}_n}(f)$  satisfies  $GC(\gamma'_n)$  where*

$$\gamma'_n = \begin{cases} \frac{2C(1+q)^2 \|f\|_{Lip}^2}{r^{2|\mathbb{G}_n|}} \left( \frac{1-(r^2/2)^{n+2}}{1-r^2/2} \right) & \text{if } r \neq \sqrt{2} \\ \frac{2C(1+q)^2 \|f\|_{Lip}^2 (n+2)}{|\mathbb{G}_n|} & \text{if } r = \sqrt{2}. \end{cases}$$

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

have

$$\begin{aligned} \mathbb{E} \left[ \exp \left( t2^{-n} \sum_{i \in \mathbb{G}_n} f(X_i, X_{2i}, X_{2i+1}) \right) \right] &= \mathbb{E} \left[ \exp \left( t2^{-n} \sum_{i \in \mathbb{G}_n} Pf(X_i) \right) \right. \\ &\quad \times \left. \mathbb{E} \left[ \exp \left( t2^{-n} \sum_{i \in \mathbb{G}_n} (f(X_i, X_{2i}, X_{2i+1}) - Pf(X_i)) \right) \middle| \mathcal{F}_n \right] \right]. \end{aligned}$$

By the Markov property and thanks to the Proposition 2.2 and the Theorem 1.5, we have

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

Now, using  $Pf$  instead of  $f$  in the proof of the Proposition 3.4 and using the fact that  $\|Pf\|_{Lip} \leq (1+q)\|f\|_{Lip}$  and

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

we are led to

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

The results then follow by easy calculations.  $\square$

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

**Proposition 3.7.** *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*

$$\tau_n = \begin{cases} \frac{2C\|f\|_{Lip}^2}{(r-1)^2|\mathbb{T}_n|} \left( 1 + \frac{1-(r^2/2)^{n+1}}{1-r^2/2} \right) & \text{if } r \neq \sqrt{2}, r \neq 1 \\ \frac{2C\|f\|_{Lip}^2}{(r-1)^2|\mathbb{T}_n|} (r^2(n+1) + 1) & \text{if } r = \sqrt{2} \\ \frac{2C\|f\|_{Lip}^2}{|\mathbb{T}_n|^2} (|\mathbb{T}_n| - \frac{n+1}{2}) & \text{if } r = 1. \end{cases}$$

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

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

We have

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

As in the proof of Proposition 3.4, we have

$$\begin{aligned} \mathbb{E} \left[ \exp \left( \frac{t}{|\mathbb{T}_n|} \sum_{i \in \mathbb{G}_{n-1}} (f(X_{2i}) + f(X_{2i+1}) - (P_0 + P_1)f(X_i)) \right) \middle| \mathcal{F}_{n-1} \right] \\ \leq \exp \left( \frac{2^2 C t^2 \|f\|_{Lip}^2 2^{n-1}}{2|\mathbb{T}_n|^2} \right). \end{aligned}$$

This leads us to

$$\begin{aligned} \mathbb{E} \left[ \exp \left( \frac{t}{|\mathbb{T}_n|} \sum_{i \in \mathbb{T}_n} f(X_i) \right) \right] &\leq \exp \left( \frac{2^2 C t^2 \|f\|_{Lip}^2 2^{n-1}}{2|\mathbb{T}_n|^2} \right) \\ &\times \mathbb{E} \left[ \exp \left( \frac{t}{|\mathbb{T}_n|} \sum_{i \in \mathbb{T}_{n-2}} f(X_i) \right) \exp \left( \frac{t}{|\mathbb{T}_n|} \sum_{i \in \mathbb{G}_{n-1}} (f + (P_0 + P_1)f)(X_i) \right) \right]. \end{aligned}$$

Iterating this method, we are led to

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

and we then obtain thanks to (a) of  $(H_1(C))$  and Theorem 1.5

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

In the last inequality we have used

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

The results then easily follows.  $\square$

For the ancestor-offspring triangle we have the following results which can be seen as a consequence of the Proposition 3.7.

**Corollary 3.8.** *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*

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

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

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

we have

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

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 3.6. We then have

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

For the second term, we use the proof of the Proposition 3.7 with  $Pf$  instead of  $f$ . We then have

$$\begin{aligned} & \left( \mathbb{E} \left[ \exp \left( \frac{2t}{|\mathbb{T}_n|} \left( \sum_{i \in \mathbb{T}_n} Pf(X_i) - \mathbb{E} \left[ \sum_{i \in \mathbb{T}_n} Pf(X_i) \right] \right) \right) \right] \right)^{1/2} \\ & \leq \exp \left( \frac{2^3 t^2 R (1+q)^2 \|f\|_{Lip}^2}{2|\mathbb{T}_n|^2} \sum_{k=0}^n \left( \sum_{l=0}^k r^l \right)^2 2^{n-k-1} \right). \end{aligned}$$

The results then follow by easy analysis and this ends the proof.  $\square$

### 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 non asymptotic 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  $\pi$  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 3.9** ( $H'_1(C)$ ).

- (a)  $\nu \in T_1(C)$ ;
- (b)  $P_b(x, \cdot) \in T_1(C), \forall x \in E, b = 0, 1$  ;

(c)  $W_1^d(P(x, \cdot, \cdot), P(\tilde{x}, \cdot, \cdot)) \leq q d(x, \tilde{x})$ ,  $\forall x, \tilde{x} \in E$  and some  $q > 0$ . And for  $r_0, r_1 > 0$  such that  $r_0 + r_1 < 2$ , for  $b = 0, 1$ ,  $W_1^d(P_b(x, \cdot), P_b(\tilde{x}, \cdot)) \leq r_b d(x, \tilde{x})$ ,  $\forall x, \tilde{x} \in E$ .

Under this assumption, using the convexity of  $W_1$  (see [37]), we easily see that

$$W_1(Q(x, \cdot), Q(\tilde{x}, \cdot)) \leq \frac{r_0 + r_1}{2} d(x, \tilde{x}), \quad \forall x, \tilde{x}$$

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

$$W_1(Q^n(x, \cdot), \pi) \leq \left( \frac{r_0 + r_1}{2} \right)^n \int d(x, y) \pi(dy).$$

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

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

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

$$\begin{aligned} |\mathbb{E}(\overline{M}_{\mathbb{T}_n}(f)) - \pi(f)| &= \frac{1}{|\mathbb{T}_n|} \left| \sum_{j=1}^n \mathbb{E} \left( \sum_{k \in \mathbb{G}_j} (f(X_k) - \pi(f)) \right) \right| \\ &= \frac{1}{|\mathbb{T}_n|} \left| \sum_{j=1}^n 2^j \nu \left( \frac{P_0 + P_1}{2} \right)^j (f - \pi(f)) \right| \\ &\leq \frac{1}{|\mathbb{T}_n|} \sum_{j=1}^n 2^j W_1(\nu Q^j, \pi) \\ &\leq \frac{1}{|\mathbb{T}_n|} \sum_{j=1}^n (r_0 + r_1)^j \\ &\leq c_n := \begin{cases} c \left( \frac{r_0 + r_1}{2} \right)^{n+1} & \text{if } r_0 + r_1 \neq 1 \\ c \frac{n}{2^{n+1}} & \text{if } r_0 + r_1 = 1 \end{cases} \end{aligned}$$

for some universal  $c$ , which goes to 0 exponentially fast as soon as  $r_0 + r_1 < 2$  which was assumed in  $(H'_1(C))$ . We may then see that for  $r > c_n$

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

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 [36], and the asymptotic behavior of parametric estimators as well as non parametric estimators has been considered in an important series of work, see e.g. [1, 2, 3, 4, 5, 29, 6, 19, 15, 17, 7] (and for example in the random coefficient setting in [16]).

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

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

with the following assumptions:

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

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

It is then easy to deduce that under these two assumptions, we perfectly match with 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 1.6. To this end, consider  $x_0 = f(x)$ , and choose  $\delta_\varepsilon$  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  $\varepsilon_k$  denotes a generic random variable and thus the law of  $P_0(y, \cdot)$  is the law of  $f_0(y) + \varepsilon_k$  and an upper bound of the Wasserstein distance between  $P_0(x, \cdot)$  and  $P_0(y, \cdot)$  can then be obtained by the coupling where we really choose the same noise  $\varepsilon_k$  for the realization of the two marginal laws so that

Let  $f$  be any Lipschitz function such that  $\|f\|_{Lip} \leq 1$

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

By the Monge-Kantorovitch duality expression of the Wasserstein distance, one has then

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

Thus under (NL) and (No), our model fits in the framework of the previous section with  $q = \|f_0\|_{Lip} + \|f_1\|_{Lip}$ ,  $r_0 = \|f_0\|_{Lip}$  and  $r_1 = \|f_1\|_{Lip}$ . We will be interested here in the non parametric estimation of the autoregression functions  $f_0$  and  $f_1$ , and we will use Nadaraya-Watson kernel type estimator, as considered in [9]. Let  $K$  be a kernel satisfying the following assumption.

**Assumption 4.3 (Ker).** *The function  $K$  is non negative, has compact support  $[-R, R]$ , is Lipschitz continuous with constant  $\|K\|_{Lip}$  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 as for  $x \in \mathbb{R}$

$$\begin{aligned} \widehat{f}_{0,n}(x) &:= \frac{\frac{1}{|\mathbb{T}_n| h_n} \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right) X_{2k}}{\frac{1}{|\mathbb{T}_n| h_n} \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right)} \\ \widehat{f}_{1,n}(x) &:= \frac{\frac{1}{|\mathbb{T}_n| h_n} \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right) X_{2k+1}}{\frac{1}{|\mathbb{T}_n| h_n} \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right)}. \end{aligned}$$

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  $f(x)$ . 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 and reunite them. Secondly  $(x, y) \rightarrow K(x)y$  is in fact

not Lipschitzian in general state space, so that the result of the previous section for deviation inequalities of Lipschitzian function of ancestor-offspring may not be applied directly. Let us tackle this problem. By definition

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

where

$$\begin{aligned}N_n &:= \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right) [f_0(X_k) - f_0(x)], \\ M_n &:= \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right) \varepsilon_{2k}, \\ D_n &= \sum_{k \in \mathbb{T}_n} K\left(\frac{X_k - x}{h_n}\right).\end{aligned}$$

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 3.7. We may thus prove

**Proposition 4.4.** *Let us assume that (NL), (No) and (Ker) holds, and  $q = \|f_0\|_{Lip} + \|f_1\|_{Lip} < \sqrt{2}$ . 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*

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

*Proof.* Remark first that, by (Ker),  $K$  is Lipschitz continuous so that  $y \rightarrow K\left(\frac{y-x}{h_n}\right)$  is also Lipschitzian with constant  $\|K\|_{Lip}/h_n$ . The mapping  $y \rightarrow K\left(\frac{y-x}{h_n}\right)(f_0(y) - f_0(x))$ , as  $K$  has a compact support and  $f_0$  is Lipschitzian, is also Lipschitzian with constant  $R\|K\|_{Lip}\|f_0\|_{Lip} + \|f_0\|_{Lip}\|K\|_{\infty}$ . We can then use Proposition 3.7 to get deviation inequalities for  $D_n$ . For all positive

$r$  there exists a constant  $L$  (explicitly given through Proposition 3.7), such that

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

For  $N_n + M_n$  we cannot directly apply Proposition 3.7 due to the successive dependence of  $X_k$  at generation  $n$  and  $\varepsilon_{2k}$  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 3.7,

$$\mathbb{P}(|N_n - \mathbb{E}(N_n)| > r|\mathbb{T}_n|h_n) \leq 2 \exp\left(\frac{-Lr^2|\mathbb{T}_n|h_n^2}{(R\|K\|_{Lip}\|f_0\|_{Lip} + \|f_0\|_{Lip}\|K\|_\infty)^2}\right).$$

Note that  $\varepsilon_{2k}$  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, we may control the Laplace transform of  $M_n$  quite simply. We then have for all positive  $r$

$$\mathbb{P}(|M_n| > r|\mathbb{T}_n|h_n) \leq 2 \exp\left(-r^2 \frac{|\mathbb{T}_n|h_n^2\|K\|_\infty^2}{2C\varepsilon}\right).$$

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

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

Remark now to conclude that  $K((y-x)/h_n)(f(y) - f(x)) + K((y-x)/h_n)$  is  $(R\|K\|_{Lip}\|f_0\|_{Lip} + \|f_0\|_{Lip}\|K\|_\infty + r\|K\|_{Lip}/h_n)$ -Lipschitzian, and we may then proceed as before.  $\square$

*Remark 4.5.* In order to get fully practical deviation inequalities, let us remark that

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

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

$$\mathbb{E} \left[ \tilde{N}_n \right] = \frac{1}{|\mathbb{T}_n| h_n} \sum_{m=0}^n 2^m (\mu_0 Q^m (H f_0) - f_0(x) \mu_0 Q^m H) \xrightarrow{n \rightarrow +\infty} 0.$$

We refer to [9] for quantitative versions of these limits.

*Remark 4.6.* Of course this non parametric 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 [12] for (uniform) deviation inequalities for estimators of density in the i.i.d. case, and to [22] for control of the Wasserstein distance of the empirical measure of i.i.d.r.v. or of Markov chains.

*Remark 4.7* (Estimation of the  $\mathbb{T}$ -transition probability). We assume that the process has as initial law, the invariant probability  $\nu$ . We denote by  $f$  the density of  $(X_1, X_2, X_3)$ . For the estimation of  $f$ , we propose the estimator  $\hat{f}_n$  defined by

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

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

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

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

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

Now using the decomposition

$$\begin{aligned} \hat{f}_n(x, y, z) - f(x, y, z) &= \left( \hat{f}_n(x, y, z) - \mathbb{E} \left[ \hat{f}_n(x, y, z) \right] \right) \\ &\quad + \left( \mathbb{E} \left[ \hat{f}_n(x, y, z) \right] - f(x, y, z) \right) \end{aligned}$$

and the convergence of  $\mathbb{E} \left[ \widehat{f}_n(x, y, z) \right]$  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  $g_\varepsilon$  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  $g_\varepsilon$  is Lipschitzian, we have

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

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 sincerely thanks Labex Mathématiques Hadamard. A. Guillin was supported by the French ANR STAB project.

### References

- [1] I. V. Basawa and R. M. Huggins. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.*, 36, 4:1225–1233, 1999.
- [2] I. V. Basawa and R. M. Huggins. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.*, 42, 4:423–432, 2000.
- [3] I. V. Basawa and J. Zhou. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.*, 41A:55–64, 2004.
- [4] I. V. Basawa and J. Zhou. Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab.Lett.*, 1:77–88, 2005.
- [5] I. V. Basawa and J. Zhou. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Ser. Anal.*, 26, 6:825–842, 2005.
- [6] B. Bercu, B. De Saporta, and A. Gégout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic. J. Probab.*, 14:2492–2526, 2009.

- [7] S. V. Bitseki Penda and H. Djellout. Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models. *Annales de l'IHP-PS*, 50(3):806–844, 2014.
- [8] S. V. Bitseki Penda, H. Djellout, and A. Guillin. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *The Annals of Applied Probability*, 24(1):235–291, 2014.
- [9] S. V. Bitseki Penda and A. Olivier. Asymptotic behavior of nonparametric estimators in nonlinear bifurcating autoregressive models. In preparation, 2015.
- [10] S.V. Bitseki Penda. Deviations inequalities for bifurcating markov chains on galton- watson tree. In revision for ESAIM P/S, 2014.
- [11] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic sobolev inequalities. *J. Funct. Anal.*, 163:1–28, 1999.
- [12] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields*, 137(3-4):541–593, 2007.
- [13] P. Cattiaux and A. Guillin. On quadratic transportation cost inequalities. *J. Math. Pures Appl. (9)*, 86(4):341–361, 2006.
- [14] P. Cattiaux, A. Guillin, and L. Wu. A note on Talagrand’s transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields*, 148(1-2):285–304, 2010.
- [15] B. De Saporta, A Gégout-Petit, and L. Marsalle. Asymmetry tests for bifurcating auto-regressive processes with missing data. *preprint*, page 2011.
- [16] B. De Saporta, A Gégout-Petit, and L. Marsalle. Random coefficients bifurcating autoregressive processes. *arXiv:1205.3658v1 [math.PR]*, page 2012.
- [17] B. De Saporta, A. Gégout-Petit, and L. Marsalle. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electronic Journal of Statistics*, 5:1313–1353, 2011.

- [18] B. De Saporta, A. Gégout-Petit, and L. Marsalle. Random coefficients bifurcating autoregressive processes. *ESAIM PS*, 18:365–399, 2014.
- [19] J. F. Delmas and L. Marsalle. Detection of cellular aging in Galton-Watson process. *Stochastic Processes and their Applications*, 120:2495–2519, 2010.
- [20] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Annals of Probability*, 32, No. 3B:2702–2732, 2004.
- [21] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, to appear, 2014.
- [22] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, To appear, 2015.
- [23] N. Gozlan. Integral criteria for transportation-cost inequalities. *Electron. Comm. Probab.*, 11:64–77 (electronic), 2006.
- [24] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.
- [25] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields*, 139(1-2):235–283, 2007.
- [26] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.
- [27] N. Gozlan, C. Roberto, and P-M. Samson. A new characterization of Talagrand’s transport-entropy inequalities and applications. *Ann. Probab.*, 39(3):857–880, 2011.
- [28] A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for Markov processes. *Probab. Theory Related Fields*, 144(3-4):669–695, 2009.
- [29] J. Guyon. Limit theorems for bifurcating Markov chains. application to the detection of cellular aging. *Ann. Appl. Probab.*, Vol. 17, No. 5-6:1538–1569, 2007.



- [30] J. Guyon, A. Bize, G. Paul, E.J. Stewart, J.F. Delmas, and F. Taddéi. Statistical study of cellular aging. *CEMRACS 2004 Proceedings, ESAIM Proceedings*, 14:100–114, 2005.
- [31] M. Ledoux. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.
- [32] K. Marton. A simple proof of the blowing-up lemma. *IEEE Transactions on Information Theory*, 32:445–446, 1986.
- [33] K. Marton. Bounding  $\bar{d}$ -distance by informational divergence: a way to prove measure concentration. *Annals of Probability*, 24:857–866, 1996.
- [34] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6:556–571, 1997.
- [35] P. Massart. *Concentration inequalities and model selection*. Lecture Notes in Mathematics, 1896. Springer, Berlin, 2007.
- [36] 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(2):e45, 2005.
- [37] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.