

# Parameters estimation for asymmetric bifurcating autoregressive processes with missing data

Benoîte de Saporta

*Université de Bordeaux, GREThA CNRS UMR 5113, IMB CNRS UMR 5251  
and INRIA Bordeaux Sud Ouest team CQFD, France  
e-mail: [saporta@math.u-bordeaux1.fr](mailto:saporta@math.u-bordeaux1.fr)*

Anne Gégout-Petit

*Université de Bordeaux, IMB, CNRS UMR 525  
and INRIA Bordeaux Sud Ouest team CQFD, France  
e-mail: [anne.petit@u-bordeaux2.fr](mailto:anne.petit@u-bordeaux2.fr)*

and

Laurence Marsalle

*Université de Lille 1, Laboratoire Paul Painlevé, CNRS UMR 8524, France  
e-mail: [laurence.marsalle@univ-lille1.fr](mailto:laurence.marsalle@univ-lille1.fr)*

**Abstract:** We estimate the unknown parameters of an asymmetric bifurcating autoregressive process (BAR) when some of the data are missing. In this aim, we model the observed data by a two-type Galton-Watson process consistent with the binary tree structure of the data. Under independence between the process leading to the missing data and the BAR process and suitable assumptions on the driven noise, we establish the strong consistency of our estimators on the set of non-extinction of the Galton-Watson process, via a martingale approach. We also prove a quadratic strong law and the asymptotic normality.

**AMS 2000 subject classifications:** Primary 62F12, 62M09; secondary 60J80, 92D25, 60G42.

**Keywords and phrases:** Least squares estimation, bifurcating autoregressive process, missing data, Galton-Watson process, joint model, martingales, limit theorems.

Received June 2011.

## 1. Introduction

Bifurcating autoregressive processes (BAR) generalize autoregressive (AR) processes, when the data have a binary tree structure. Typically, they are involved in modeling cell lineage data, since each cell in one generation gives birth to two offspring in the next one. Cell lineage data usually consist of observations of some quantitative characteristic of the cells, over several generations descended

from an initial cell. BAR processes take into account both inherited and environmental effects to explain the evolution of the quantitative characteristic under study. They were first introduced by Cowan and Staudte [4]. In their paper, the original BAR process was defined as follows. The initial cell is labelled 1, and the two offspring of cell  $k$  are labelled  $2k$  and  $2k + 1$ . If  $X_k$  denotes the quantitative characteristic of individual  $k$ , then the first-order BAR process is given, for all  $k \geq 1$ , by

$$\begin{cases} X_{2k} &= a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} &= a + bX_k + \varepsilon_{2k+1}. \end{cases}$$

The noise sequence  $(\varepsilon_{2k}, \varepsilon_{2k+1})$  represents environmental effects, while  $a, b$  are unknown real parameters, with  $|b| < 1$ , related to the inherited effects. The driven noise  $(\varepsilon_{2k}, \varepsilon_{2k+1})$  was originally supposed to be independent and identically distributed with normal distribution. But since two sister cells are in the same environment at their birth,  $\varepsilon_{2k}$  and  $\varepsilon_{2k+1}$  are allowed to be correlated, inducing a correlation between sister cells, distinct from the correlation inherited from their mother.

Recently, experiments made by biologists on aging of *Escherichia coli* [15], motivated mathematical and statistical studies of the asymmetric BAR process, that is when the quantitative characteristics of the even and odd sisters are allowed to depend on their mother's through different sets of parameters  $(a, b)$ , see Equation (2.1) below. In [9, 8], Guyon proposes an interpretation of the asymmetric BAR process as a bifurcating Markov chain, which allows him to derive laws of large numbers and central limit theorems for the least squares estimators of the unknown parameters of the process. This Markov chain approach was further developed by Bansaye [2] in the context of cell division with parasite infection, and by Delmas and Marsalle [5], where the cells are allowed to die. Another approach based on martingales theory was proposed by Bercu, de Saporta and Gégout-Petit [3], to sharpen the asymptotic analysis of Guyon under weaker assumptions.

The originality of this paper is that we take into account possibly missing data in the estimation procedure of the parameters of the asymmetric BAR process, see Figure 1 for an example. This is a problem of practical interest, as experimental data are often incomplete, either because some cells died, or because the measurement of the characteristic under study was impossible or faulty. For instance, among the 94 colonies dividing up to 9 times studied in [15], in average, there are about 47% of missing data. It is important to take this phenomenon into account in the model for a rigorous statistical study.

Missing data in bifurcating processes were first modeled by Delmas and Marsalle [5]. They defined the genealogy of the cells through a Galton-Watson process, but they took into account the possible asymmetry problem only by differentiating the reproduction laws according to the daughter's type (even or odd). The bifurcating process was thus still a Markov chain. However, considering the biological issue of aging in *E. coli* naturally leads to introduce the possibility that two cells of different types may not have the same reproduction law. In this paper, we thus introduce a two-type Galton-Watson process to

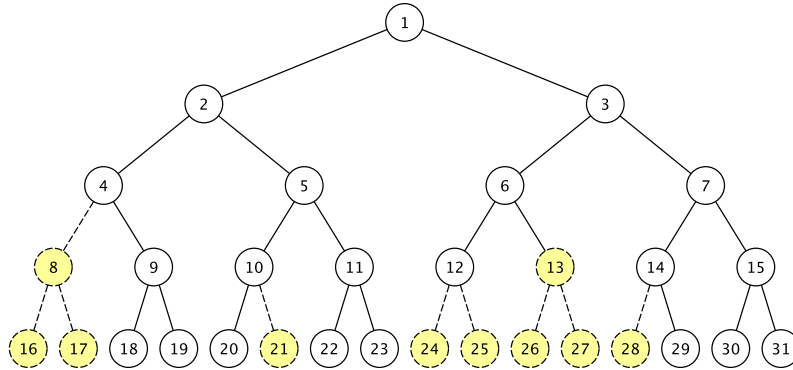


FIG 1. A tree associated with the bifurcating auto-regressive process up to the 4th generation. The dashed cells are not observed.

model the genealogy, and lose the Markovian structure of the bifurcating chain, so that we cannot use the same approach as [5]. Instead, we use the martingale approach introduced in [3]. It must be pointed out that missing data are not dealt with in [3], so that we cannot directly use their results either. In particular, the observation process is another source of randomness that requires stronger moment assumptions on the driven noise of the BAR process and careful choice between various filtrations. In addition, the normalizing terms are now random and the convergences are only available on the random non-extinction set of the observed process.

The *naive* approach to handle missing data would be to replace the sums over all data in the estimators by sums over the observed data only. Our approach is slightly more subtle, as we distinguish whether a cell has even or odd daughters. We propose a joint model where the structure for the observed data is based on a two-type Galton-Watson process consistent with the possibly asymmetric structure of the BAR process. See e.g. [12, 1, 10] for a presentation of multi-type Galton-Watson processes and general branching processes. Note also that our estimation procedure does not require the previous knowledge of the parameters of the two-type Galton-Watson process.

This paper is organized as follows. In Section 2, we first introduce our BAR model as well as related notation, then we define and recall results on the two-type Galton-Watson process used to model the observation process. In Section 3, we give the least square estimator for the parameters of observed BAR process and we state our main results on the convergence and asymptotic normality of our estimators as well as estimation results on data. The proofs are detailed in the following sections.

## 2. Joint model

We now introduce our joint model, starting with the asymmetric BAR process for the variables of interest.

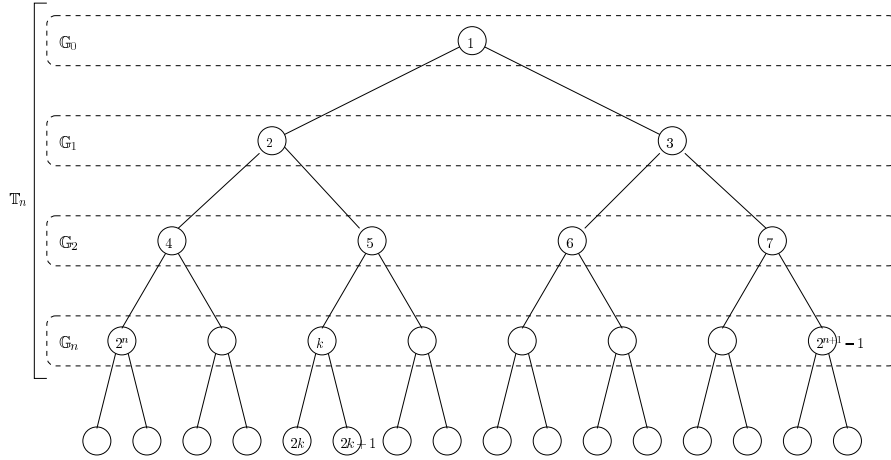


FIG 2. The tree associated with the bifurcating auto-regressive process.

**2.1. Bifurcating autoregressive processes**

On the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ , we consider the first-order asymmetric BAR process given, for all  $k \geq 1$ , by

$$\begin{cases} X_{2k} &= a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} &= c + dX_k + \varepsilon_{2k+1}. \end{cases} \tag{2.1}$$

The initial state  $X_1$  is the characteristic of the ancestor, while  $(\varepsilon_{2k}, \varepsilon_{2k+1})$  is the driven noise of the process. In all the sequel, we shall assume that  $\mathbb{E}[X_1^8] < \infty$ . Moreover, as in the previous literature, the parameters  $(a, b, c, d)$  belong to  $\mathbb{R}^4$  with

$$0 < \max(|b|, |d|) < 1.$$

This assumption ensures the stability (non explosion) of the BAR process. As explained in the introduction, one can see this BAR process as a first-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor, see Figure 2 for an illustration. We use the same notation as in [3]. For all  $n \geq 1$ , denote the  $n$ -th generation by  $\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}$ . In particular,  $\mathbb{G}_0 = \{1\}$  is the initial generation, and  $\mathbb{G}_1 = \{2, 3\}$  is the first generation of offspring from the first ancestor. Let  $\mathbb{G}_{r_k}$  be the generation of individual  $k$ , which means that  $r_k = \lceil \log_2(k) \rceil$ , where  $\lceil x \rceil$  denotes the largest integer less than or equal to  $x$ . Recall that the two offspring of individual  $k$  are labelled  $2k$  and  $2k + 1$ , or conversely, the mother of individual  $k$  is  $\lfloor k/2 \rfloor$ . More generally, the ancestors of individual  $k$  are  $\lfloor k/2 \rfloor, \lfloor k/2^2 \rfloor, \dots, \lfloor k/2^{r_k} \rfloor$ . Denote by  $\mathbb{T}_n = \bigcup_{\ell=0}^n \mathbb{G}_\ell$  the sub-tree of all individuals from the original individual up to the  $n$ -th generation. Note that the cardinality  $|\mathbb{G}_n|$  of  $\mathbb{G}_n$  is  $2^n$ , while that of  $\mathbb{T}_n$  is  $|\mathbb{T}_n| = 2^{n+1} - 1$ . Next,  $\mathbb{T}$  denotes the complete tree, so to speak  $\mathbb{T} = \bigcup_{n \geq 0} \mathbb{G}_n = \bigcup_{n \geq 0} \mathbb{T}_n = \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ .

Finally, we need to distinguish the individuals in  $\mathbb{G}_n$  and  $\mathbb{T}_n$  according to their type. Since we are dealing with the types even and odd, that we will also label 0 and 1, we set

$$\begin{aligned} \mathbb{G}_n^0 &= \mathbb{G}_n \cap (2\mathbb{N}), & \mathbb{G}_n^1 &= \mathbb{G}_n \cap (2\mathbb{N} + 1), & \mathbb{T}_n^0 &= \mathbb{T}_n \cap (2\mathbb{N}), \\ \mathbb{T}_n^1 &= \mathbb{T}_n \cap (2\mathbb{N} + 1), & \mathbb{T}^0 &= \mathbb{T} \cap (2\mathbb{N}) & \text{and} & \mathbb{T}^1 = \mathbb{T} \cap (2\mathbb{N} + 1). \end{aligned} \quad (2.2)$$

We now state our assumptions on the noise sequence. Denote by  $\mathbb{F} = (\mathcal{F}_n)$  the natural filtration associated with the first-order BAR process, which means that  $\mathcal{F}_n$  is the  $\sigma$ -algebra generated by all individuals up to the  $n$ -th generation,  $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$ . In all the sequel, we shall make use of the following moment and independence hypotheses.

**(HN.1)** For all  $n \geq 0$  and for all  $k \in \mathbb{G}_{n+1}$ ,  $\varepsilon_k$  belongs to  $L^8$ . Moreover, there exist  $(\sigma^2, \tau^4, \kappa^8) \in (0, +\infty)^3$ ,  $(|\rho'|, \nu^2, \lambda^4) \in [0, 1]^3$  such that:

- $\forall n \geq 0$  and  $k \in \mathbb{G}_{n+1}$ ,

$$\mathbb{E}[\varepsilon_k | \mathcal{F}_n] = 0, \quad \mathbb{E}[\varepsilon_k^2 | \mathcal{F}_n] = \sigma^2, \quad \mathbb{E}[\varepsilon_k^4 | \mathcal{F}_n] = \tau^4, \quad \mathbb{E}[\varepsilon_k^8 | \mathcal{F}_n] = \kappa^8 \quad \text{a.s.}$$

- $\forall n \geq 0 \quad \forall k \neq l \in \mathbb{G}_{n+1}$  with  $[k/2] = [l/2]$ ,

$$\mathbb{E}[\varepsilon_k \varepsilon_l | \mathcal{F}_n] = \rho = \rho' \sigma^2, \quad \mathbb{E}[\varepsilon_{2k}^2 \varepsilon_{2k+1}^2 | \mathcal{F}_n] = \nu^2 \tau^4, \quad \mathbb{E}[\varepsilon_{2k}^4 \varepsilon_{2k+1}^4 | \mathcal{F}_n] = \lambda^4 \kappa^8 \quad \text{a.s.}$$

**(HN.2)** For all  $n \geq 0$  the random vectors  $\{(\varepsilon_{2k}, \varepsilon_{2k+1}), k \in \mathbb{G}_n\}$  are conditionally independent given  $\mathcal{F}_n$ .

## 2.2. Observation process

We now turn to the modeling of the observation process. The observation process is intended to encode if a datum is missing or not. The natural property it has thus to satisfy is the following: if the datum is missing for some individual, it is also missing for all its descendants. Indeed, the datum may be missing because of the death of the individual, or because the individual is the last of its lineage at the end of the data's gathering, see Figure 3 for an example of partially observed tree.

### 2.2.1. Definition of the observation process

Mathematically, we define the observation process,  $(\delta_k)_{k \in \mathbb{T}}$ , as follows. We set  $\delta_1 = 1$  and define recursively the sequence through the following equalities:

$$\delta_{2k} = \delta_k \zeta_k^0 \quad \text{and} \quad \delta_{2k+1} = \delta_k \zeta_k^1, \quad (2.3)$$

where  $(\zeta_k = (\zeta_k^0, \zeta_k^1))$  is a sequence of independent random vectors of  $\{0, 1\}^2$ ,  $\zeta_k^i$  standing for the number (0 or 1) of descendants of type  $i$  of individual  $k$ . The sequences  $(\zeta_k, k \in 2\mathbb{N}^*)$  and  $(\zeta_k, k \in 2\mathbb{N} + 1)$  are sequences of identically

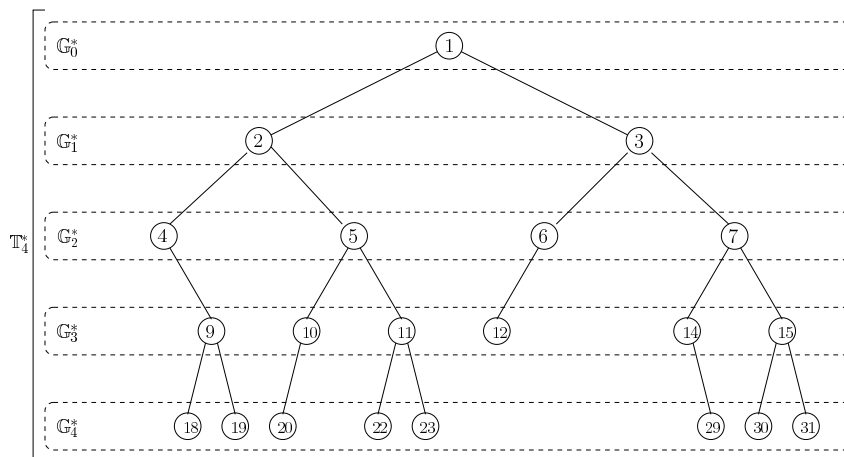


FIG 3. The tree associated with the observed data of the tree in Figure 1.

distributed random vectors. We specify the common laws of these two sequences using their generating functions,  $f^{(0)}$  and  $f^{(1)}$  respectively:

$$\begin{aligned} f^{(0)}(s_0, s_1) &= p^{(0)}(0, 0) + p^{(0)}(1, 0)s_0 + p^{(0)}(0, 1)s_1 + p^{(0)}(1, 1)s_0s_1, \\ f^{(1)}(s_0, s_1) &= p^{(1)}(0, 0) + p^{(1)}(1, 0)s_0 + p^{(1)}(0, 1)s_1 + p^{(1)}(1, 1)s_0s_1, \end{aligned}$$

where  $p^{(i)}(j_0, j_1)$  is the probability that an individual of type  $i$  gives birth to  $j_0$  descendants of type 0, and  $j_1$  of type 1. The sequence  $(\delta_k)$  is thus completely defined. We also assume that the observation process is independent from the BAR process.

**(HI)** The sequences  $(\delta_k)$  and  $(\zeta_k)$  are independent from the sequences  $(X_k)$  and  $(\varepsilon_k)$ .

Remark that, since both  $\zeta_k^0$  and  $\zeta_k^1$  take values in  $\{0, 1\}$  for all  $k$ , the observation process  $(\delta_k)$  is itself taking values in  $\{0, 1\}$ . Finally, Equation (2.3) ensures that if  $\delta_k = 0$  for some  $k \geq 2$ , then for all its descendants  $j$ ,  $\delta_j = 0$ . In relation with the observation process  $(\delta_k)$ , we introduce two filtrations:  $\mathcal{Z}_n = \sigma\{\zeta_k, k \in \mathbb{T}_n\}$ ,  $\mathcal{O}_n = \sigma\{\delta_k, k \in \mathbb{T}_n\}$ , and the sigma field  $\mathcal{O} = \sigma\{\delta_k, k \in \mathbb{T}\}$ . Notice that  $\mathcal{O}_{n+1} \subset \mathcal{Z}_n$ . We also define the sets of observed individuals as follows:

$$\mathbb{G}_n^* = \{k \in \mathbb{G}_n : \delta_k = 1\} \quad \text{and} \quad \mathbb{T}_n^* = \{k \in \mathbb{T}_n : \delta_k = 1\}.$$

Finally, let  $\mathcal{E}$  be the event corresponding to the cases when there are no individual left to observe. More precisely,

$$\mathcal{E} = \bigcup_{n \geq 1} \{|\mathbb{G}_n^*| = 0\}. \tag{2.4}$$

We will denote  $\bar{\mathcal{E}}$  the complementary set of  $\mathcal{E}$ .

2.2.2. Results on the observation process

Let us introduce some additional notation. For  $n \geq 1$ , we define the number of observed individuals among the  $n$ -th generation, distinguishing according to their types:

$$Z_n^0 = |\mathbb{G}_n^* \cap 2\mathbb{N}| \quad \text{and} \quad Z_n^1 = |\mathbb{G}_n^* \cap (2\mathbb{N} + 1)|, \tag{2.5}$$

and we set, for all  $n \geq 1$ ,  $\mathbf{Z}_n = (Z_n^0, Z_n^1)$ . Note that for  $i \in \{0, 1\}$  and  $n \geq 1$  one has

$$Z_n^i = \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i}.$$

One has  $\mathbb{G}_0^* = \mathbb{G}_0 = \{1\}$ , but, even if 1 is odd, the individual whose lineage we study may as well be of type 0 as of type 1. Consequently, we will work with possibly two different initial laws:  $\mathbb{P}(\cdot | \mathbf{Z}_0 = \mathbf{e}_i)$ , for  $i \in \{0, 1\}$ , where  $\mathbf{e}_0 = (1, 0)$  and  $\mathbf{e}_1 = (0, 1)$ . The process  $(\mathbf{Z}_n, n \geq 0)$  is thus a two-type Galton-Watson process, and all the results we are giving in this section mainly come from [12]. Notice that the law of  $\zeta_k$ , for even  $k$ , is the law of reproduction of an individual of type 0, the first component of  $\zeta_k$  giving the number of children of type 0, the second the number of children of type 1. The same holds for  $\zeta_k$ , with odd  $k$ , *mutatis mutandis*. This ensures the existence of moments of all order for these reproduction laws, and we can thus define the descendants matrix  $\mathbf{P}$

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},$$

where  $p_{i0} = p^{(i)}(1, 0) + p^{(i)}(1, 1)$  and  $p_{i1} = p^{(i)}(0, 1) + p^{(i)}(1, 1)$ , for  $i \in \{0, 1\}$ . The quantity  $p_{ij} = \mathbb{E}[\zeta_{2+i}^j]$  is thus the expected number of descendants of type  $j$  of an individual of type  $i$ . We also introduce the variance of the laws of reproduction:  $\sigma_{ij}^2 = \mathbb{E}[(\zeta_{2+i}^j - p_{ij})^2]$ , for  $(i, j) \in \{0, 1\}^2$ . Note that  $\sigma_{ij}^2 = p_{ij}(1 - p_{ij})$ . It is well-known (see e.g. Theorem 5.1 of [12]) that when all the entries of the matrix  $\mathbf{P}$  are positive,  $\mathbf{P}$  has a positive strictly dominant eigenvalue, denoted  $\pi$ , which is also simple. We make the following main assumptions on the matrix  $\mathbf{P}$ .

**(HO)** All entries of the matrix  $\mathbf{P}$  are positive: for all  $(i, j) \in \{0, 1\}^2$ ,  $p_{ij} > 0$ , and the dominant eigenvalue is greater than one:  $\pi > 1$ .

Hence, still following Theorem 5.1 of [12], we know that there exist left and right eigenvectors for  $\pi$  which are positive, in the sense that each component of the vector is positive. We call  $\mathbf{y} = (y^0, y^1)^t$  such a right eigenvector, and  $\mathbf{z} = (z^0, z^1)$  such a left one; without loss of generality, we choose  $\mathbf{z}$  such that  $z^0 + z^1 = 1$ . Regarding the two-type Galton-Watson process  $(\mathbf{Z}_n)$ ,  $\pi$  plays the same role as the expected number of offspring, in the case of standard Galton-Watson processes. In particular,  $\pi$  is related to the extinction of the process, where the set of extinction of  $(\mathbf{Z}_n)$  is defined as  $\cup_{n \geq 1} \{\mathbf{Z}_n = (0, 0)\}$ . Notice that  $\{\mathbf{Z}_n = (0, 0)\} = \{Z_n^0 + Z_n^1 = 0\} = \{|\mathbb{G}_n^*| = 0\}$ , so that this set coincides with  $\mathcal{E}$ , defined by Eq. (2.4). Now let  $\mathbf{q} = (q^0, q^1)$ , where, for  $i \in \{0, 1\}$ ,

$$q^i = \mathbb{P}(\mathcal{E} | \mathbf{Z}_0 = \mathbf{e}_i).$$

The probability  $q^i$  is thus the extinction probability if initially there is one individual of type  $i$ . These two probabilities allow to compute the extinction probability under any initial distribution, since  $\mathbb{P}(\mathcal{E}) = \mathbb{E}[(q^0)^{Z_0^0}(q^1)^{Z_0^1}]$ , thanks to the branching property. Hypothesis **(HO)** means that the Galton-Watson process  $(Z_n)$  is super-critical, and ensures that  $0 \leq q^i < 1$ , for both  $i = 0$  and  $i = 1$ . This immediately yields

$$\mathbb{P}(\mathcal{E}) < 1. \tag{2.6}$$

Under that condition, we also have the existence of a non-negative random variable  $W$  such that for any initial distribution of  $Z_0$

$$\lim_{n \rightarrow +\infty} \frac{Z_n}{\pi^n} = \lim_{n \rightarrow +\infty} \frac{\pi - 1}{\pi^{n+1} - 1} \sum_{\ell=0}^n Z_\ell = Wz \quad \text{a.s.} \tag{2.7}$$

It is well-known that  $\{W = 0\} = \mathcal{E}$  a.s., so that the set  $\{W > 0\}$  can be viewed as the set of non-extinction  $\bar{\mathcal{E}}$  of  $(Z_n)$ , up to a negligible set. These results give the asymptotic behavior of the number of observed individuals, since  $|\mathbb{G}_n^*| = Z_n^0 + Z_n^1$ , and  $|\mathbb{T}_n^*| = \sum_{\ell=0}^n (Z_\ell^0 + Z_\ell^1)$ :

$$\lim_{n \rightarrow +\infty} \frac{|\mathbb{G}_n^*|}{\pi^n} = \lim_{n \rightarrow +\infty} \frac{\pi - 1}{\pi^{n+1} - 1} |\mathbb{T}_n^*| = W \quad \text{a.s.}$$

Roughly speaking, this means that  $\pi^n$  is a deterministic equivalent of  $|\mathbb{T}_n^*|$  and Eq. (2.7) implies that  $z^i$  is the asymptotic proportion of cells of type  $i$  in a given generation. We will thus very often replace  $|\mathbb{T}_n^*|$  by  $\pi^n$  for computations, and the next lemma will be used frequently to replace  $\pi^n$  by  $|\mathbb{T}_n^*|$ .

**Lemma 2.1.** *Under assumption (HO), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\pi^n}{|\mathbb{T}_n^*|} = \frac{\pi - 1}{\pi} \frac{1}{W} \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

### 2.3. Joint model

The model under study in this paper is therefore the observed BAR process defined by

$$\begin{cases} \delta_{2k} X_{2k} &= \delta_{2k} (a + bX_k + \varepsilon_{2k}), \\ \delta_{2k+1} X_{2k+1} &= \delta_{2k+1} (c + dX_k + \varepsilon_{2k+1}). \end{cases}$$

The aim of this paper is to study the sharp asymptotic properties of the least-squares estimators of the parameters  $(a, b, c, d)$  and the variance matrix of the noise process.

### 3. Least-squares estimation

Our goal is to estimate  $\theta = (a, b, c, d)^t$  from the observed individuals up to the  $n$ -th generation, that is the observed sub-tree  $\mathbb{T}_n^*$ .



3.1. Definition of the estimators

We propose to make use of the standard least-squares (LS) estimator  $\widehat{\boldsymbol{\theta}}_n$  which minimizes

$$\Delta_n(\boldsymbol{\theta}) = \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k}(X_{2k} - a - bX_k)^2 + \delta_{2k+1}(X_{2k+1} - c - dX_k)^2.$$

Consequently, we obviously have for all  $n \geq 1$

$$\widehat{\boldsymbol{\theta}}_n = \begin{pmatrix} \widehat{a}_n \\ \widehat{b}_n \\ \widehat{c}_n \\ \widehat{d}_n \end{pmatrix} = \boldsymbol{\Sigma}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} X_{2k} \\ \delta_{2k} X_k X_{2k} \\ \delta_{2k+1} X_{2k+1} \\ \delta_{2k+1} X_k X_{2k+1} \end{pmatrix}, \tag{3.1}$$

where, for all  $n \geq 0$ ,

$$\boldsymbol{\Sigma}_n = \begin{pmatrix} \mathbf{S}_n^0 & 0 \\ 0 & \mathbf{S}_n^1 \end{pmatrix}, \quad \text{and} \quad \mathbf{S}_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix},$$

for  $i \in \{0, 1\}$ . In order to avoid intricate invertibility assumption, we shall assume, without loss of generality, that for all  $n \geq 0$ ,  $\boldsymbol{\Sigma}_n$  is invertible. Otherwise, we only have to add the identity matrix  $\mathbf{I}_4$  to  $\boldsymbol{\Sigma}_n$ , as Proposition 4.2 states that the normalized limit of  $\boldsymbol{\Sigma}_n$  is positive definite.

**Remark 3.1.** Note that when all data are observed, that is when all  $\delta_k$  equal 1, this is simply the least squares estimator described in the previous literature. However, one must be careful here with the indices in the normalizing matrix, as there are now two different matrices  $\mathbf{S}_n^0$  and  $\mathbf{S}_n^1$ , while there was only one in the fully observed problem. The intuitive way to deal with missing data would be to restrict the sums to the observed data only. Note that our estimator is more complex as it involves sums depending on the absence or presence of even- or odd-type *daughters* of the available data.

We now turn to the estimation of the parameters  $\sigma^2$  and  $\rho$ . We propose to estimate the conditional variance  $\sigma^2$  and the conditional covariance  $\rho$  by

$$\widehat{\sigma}_n^2 = \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}^*} (\widehat{\varepsilon}_{2k}^2 + \widehat{\varepsilon}_{2k+1}^2), \quad \widehat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}} \widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1},$$

where for all  $k \in \mathbb{G}_n$ ,

$$\begin{cases} \widehat{\varepsilon}_{2k} &= \delta_{2k}(X_{2k} - \widehat{a}_n - \widehat{b}_n X_k), \\ \widehat{\varepsilon}_{2k+1} &= \delta_{2k+1}(X_{2k+1} - \widehat{c}_n - \widehat{d}_n X_k). \end{cases},$$

and

$$\mathbb{T}_n^{*01} = \{k \in \mathbb{T}_n : \delta_{2k} \delta_{2k+1} = 1\},$$

so to speak  $\mathbb{T}_{n-1}^{*01}$  is the set of the cells of the tree  $\mathbb{T}_{n-1}$  which have exactly two offspring.

### 3.2. Main results

We can now state the sharp convergence results we obtain for the estimators above. We introduce additional notation. For  $i \in \{0, 1\}$ , let us denote:

$$\mathbf{L}^i = \begin{pmatrix} \pi z^i & h^i \\ h^i & k^i \end{pmatrix} \quad \mathbf{L}^{0,1} = \begin{pmatrix} \bar{p}(1, 1) & h^{0,1} \\ h^{0,1} & k^{0,1} \end{pmatrix}$$

where  $\mathbf{z} = (z^0, z^1)$  is the left eigenvector for the dominant eigenvalue  $\pi$  of the descendants matrix  $\mathbf{P}$  introduced in section 2.2.2,  $h^i, k^i$  are defined in Propositions 6.3 and 6.5 and the four terms of  $L^{0,1}$  defined in Proposition 6.6. We also define the  $4 \times 4$  matrices

$$\Sigma = \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix}, \quad \text{and} \quad \Gamma = \begin{pmatrix} \sigma^2 \mathbf{L}^0 & \rho \mathbf{L}^{0,1} \\ \rho \mathbf{L}^{0,1} & \sigma^2 \mathbf{L}^1 \end{pmatrix}. \tag{3.2}$$

Our first result deals with the strong consistency of the LS estimator  $\widehat{\boldsymbol{\theta}}_n$ .

**Theorem 3.2.** *Under assumptions (HN.1), (HN.2), (HO) and (HI),  $\widehat{\boldsymbol{\theta}}_n$  converges to  $\boldsymbol{\theta}$  almost surely on  $\bar{\mathcal{E}}$  with the rate of convergence*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.3}$$

In addition, we also have the quadratic strong law

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \Sigma (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) = 4 \frac{\pi - 1}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.4}$$

Our second result is devoted to the almost sure asymptotic properties of the variance and covariance estimators  $\widehat{\sigma}_n^2$  and  $\widehat{\rho}_n$ . Let

$$\sigma_n^2 = \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}^*} (\delta_{2k} \varepsilon_{2k}^2 + \delta_{2k+1} \varepsilon_{2k+1}^2), \quad \rho_n = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k} \varepsilon_{2k} \delta_{2k+1} \varepsilon_{2k+1}.$$

**Theorem 3.3.** *Under assumptions (HN.1), (HN.2), (HO) and (HI),  $\widehat{\sigma}_n^2$  converges almost surely to  $\sigma^2$  on  $\bar{\mathcal{E}}$ . More precisely, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \sum_{i=0}^1 \delta_{2k+i} (\widehat{\varepsilon}_{2k+i} - \varepsilon_{2k+i})^2 = 4(\pi - 1) \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.5}$$

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\widehat{\sigma}_n^2 - \sigma_n^2) = 4(\pi - 1) \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.6}$$

In addition,  $\widehat{\rho}_n$  converges almost surely to  $\rho$  on  $\bar{\mathcal{E}}$  and one has

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \delta_{2k+1} (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \\ = \rho \frac{\pi - 1}{\pi} \text{tr}((\mathbf{L}^1)^{-1} (\mathbf{L}^{0,1})^2 (\mathbf{L}^0)^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned} \tag{3.7}$$

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|G_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\hat{\rho}_n - \rho_n) = \rho \frac{\pi - 1}{\bar{p}(1, 1)} \text{tr}((\mathbf{L}^1)^{-1}(\mathbf{L}^{0,1})^2(\mathbf{L}^0)^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.8}$$

Our third result concerns the asymptotic normality for all our estimators  $\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n^2$  and  $\hat{\rho}_n$  given the non-extinction of the underlying Galton-Watson process. For this, using the fact that  $\mathbb{P}(\bar{\mathcal{E}}) \neq 0$  thanks to Eq. (2.6), we define the probability  $\mathbb{P}_{\bar{\mathcal{E}}}$  by

$$\mathbb{P}_{\bar{\mathcal{E}}}(A) = \frac{\mathbb{P}(A \cap \bar{\mathcal{E}})}{\mathbb{P}(\bar{\mathcal{E}})} \quad \text{for all } A \in \mathcal{A}.$$

**Theorem 3.4.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have the central limit theorem*

$$\sqrt{|\mathbb{T}_{n-1}^*|} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Sigma}^{-1}) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}). \tag{3.9}$$

In addition, we also have

$$\sqrt{|\mathbb{T}_n^*|} (\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\pi(\tau^4 - \sigma^4) + 2\bar{p}(1, 1)(\nu^2\tau^4 - \sigma^4)}{\pi}\right) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}), \tag{3.10}$$

where  $\bar{p}(1, 1)$  is defined in Eq. (6.6) and

$$\sqrt{|\mathbb{T}_{n-1}^{*01}|} (\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2\tau^4 - \rho^2) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}). \tag{3.11}$$

The proof of our main results is going to be detailed in the next sections. It is based on martingale properties, and we will exhibit our main martingale  $(\mathbf{M}_n)$  in Section 4. Sections 5 to 7 are devoted proving to the sharp asymptotic properties of  $(\mathbf{M}_n)$ . Finally, in Section 8 we prove our main results. Before turning to the definition of the martingale  $(\mathbf{M}_n)$ , we present a short application of our estimation procedure on data.

### 3.3. Results on real data

The biological issue addressed by Stewart et al. in [15] is aging in the single cell organism *Escherichia coli*, see also [7] for further biological details. *E. coli* is a rod-shaped bacterium that reproduces by dividing in the middle. Each cell has thus a new end (or *pole*), and an older one. The cell that inherits the old pole of its mother is called the old pole cell, the cell that inherits the new pole of its mother is called the new pole cell. Therefore, each cell has a *type*: old pole (even) or new pole (odd) cell, inducing asymmetry in the cell division.

Stewart et al. filmed colonies of dividing cells, determining the complete lineage and the growth rate of each cell. Their statistical study of the averaged genealogy and pair-wise comparison of sister cells showed that the old pole cells exhibit cumulatively slowed growth, less offspring biomass production and an increased probability of death. Note that their test assumes independence between the averaged pairs of sister cells which is not verified in the lineage.

TABLE 1  
*Estimation on the data set penna-2002-10-04-4*

parameter	$a$	$c$
estimation	0.03627	0.03058
C.I.	[0.03276; 0.03979]	[0.02696; 0.03420]
parameter	$b$	$d$
estimation	0.02662	0.17055
C.I.	[-0.06866; 0.12191]	[0.07247; 0.26863]

Another analysis was proposed in [9]. They model the growth rate by a Markovian bifurcating process, allowing single-experiment statistical analysis instead of averaging all the genealogical trees. Asymptotic properties of a more general asymmetric Markovian bifurcating autoregressive process are then investigated in [8], where a Wald's type test is rigorously constructed to study the asymmetry of the process. These results cannot be compared to ours because this model does not take into account the possibly missing data from the genealogies, and it is not clear how the author manages them, as not a single tree from the data of [15] is complete. In [5], the authors take missing data into account but, contrary to our approach, they allow different sets of parameters for cells with two, one or no offspring, making the direct comparison with our estimator again impossible.

We have applied our methodology on the set of data `penna-2002-10-04-4` from the experiments of [15]. It is the largest data set of the experiment. It contains 663 cells up to generation 9 (note that there would be 1023 cells in a full tree up to generation 9). In particular, we have performed

- point estimation of the vector  $\theta$ ,
- interval estimation for the coefficients  $(a, b, c, d)$ ,
- Wald's type symmetry tests for the entries of  $\hat{\theta}_n$ .

Table 1 gives the estimation  $\hat{\theta}_9$  of  $\theta$  with the 95% Confidence Interval (C.I.) of each coefficient. The variance given by the CLT for  $\theta$  in Eq. (3.9), is approximated by  $\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}$  thanks to the convergence given in Corollary 4.3. The confidence intervals of  $b$  and  $d$  show that the *non explosion* assumption ( $|b| < 1$  and  $|d| < 1$ ) is satisfied. Some empirical computation on the process  $(\delta_k)$  gives the following estimation for the highest eigenvalue of the Galton-Watson process:  $\hat{\pi} = 1.35669$  (with confidence interval [1.27979, 1.43361], see [14]), also satisfying the super-criticality assumption. Wald tests of comparison between the coefficients of  $\theta$  have been deduced of the CLT. The null hypotheses  $(a, b) = (c, d)$  (resp.  $a = c$ ,  $b = d$ ) are rejected with p-values  $p = 0.0211$  (resp.  $p = 0.0158$  and  $p = 0.0244$ ). Hence on this data set the cell division is indeed statistically asymmetric.

#### 4. Martingale approach

To establish all the asymptotic properties of our estimators, we shall make use of a martingale approach, similar to [3]. However, their results cannot be

used in our framework, since the randomness comes now not only from the state process, but also from the time space (genealogy). These two mixed randomness sources require careful choice between various filtrations, and stronger moment assumptions on the driven noise of the BAR process. For all  $n \geq 1$ , denote

$$M_n = \sum_{k \in \mathbb{T}_{n-1}} (\delta_{2k} \varepsilon_{2k}, \delta_{2k} X_k \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1}, \delta_{2k+1} X_k \varepsilon_{2k+1})^t.$$

Thus, for all  $n \geq 2$ , we readily deduce from Equations (3.1) and (2.1) that

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} \varepsilon_{2k} \\ \delta_{2k} X_k \varepsilon_{2k} \\ \delta_{2k+1} \varepsilon_{2k+1} \\ \delta_{2k+1} X_k \varepsilon_{2k+1} \end{pmatrix} = \Sigma_{n-1}^{-1} M_n. \tag{4.1}$$

The key point of our approach is that  $(M_n)$  is a martingale for a well chosen filtration.

**4.1. Martingale property**

Recall that  $\mathcal{O} = \sigma\{\delta_k, k \in \mathbb{T}\}$  is the  $\sigma$ -field generated by the observation process. We shall assume that all the history of the process  $(\delta_k)$  is known at time 0 and use the filtration  $\mathbb{F}^{\mathcal{O}} = (\mathcal{F}_n^{\mathcal{O}})$  defined for all  $n$  by

$$\mathcal{F}_n^{\mathcal{O}} = \mathcal{O} \vee \sigma\{\delta_k X_k, k \in \mathbb{T}_n\} = \mathcal{O} \vee \sigma\{X_k, k \in \mathbb{T}_n^*\},$$

where  $\mathcal{F} \vee \mathcal{G}$  denotes the  $\sigma$ -field generated by both  $\mathcal{F}$  and  $\mathcal{G}$ . Note that for all  $n$ ,  $\mathcal{F}_n^{\mathcal{O}}$  is a sub  $\sigma$ -field of  $\mathcal{O} \vee \mathcal{F}_n$ .

**Proposition 4.1.** *Under assumptions (HN.1), (HN.2) and (HI), the process  $(M_n)$  is a square integrable  $\mathbb{F}^{\mathcal{O}}$ -martingale with increasing process given, for all  $n \geq 1$ , by*

$$\langle M \rangle_n = \Gamma_{n-1} = \begin{pmatrix} \sigma^2 \mathbf{S}_{n-1}^0 & \rho \mathbf{S}_{n-1}^{0,1} \\ \rho \mathbf{S}_{n-1}^{0,1} & \sigma^2 \mathbf{S}_{n-1}^1 \end{pmatrix},$$

where  $\mathbf{S}_n^0$  and  $\mathbf{S}_n^1$  are defined in section 3.1 and

$$\mathbf{S}_n^{0,1} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}.$$

*Proof.* First, notice that for all  $n \geq 1$ , one has

$$\Delta M_n = M_n - M_{n-1} = \sum_{k \in \mathbb{G}_{n-1}} \begin{pmatrix} \delta_{2k} \varepsilon_{2k} \\ \delta_{2k} X_k \varepsilon_{2k} \\ \delta_{2k+1} \varepsilon_{2k+1} \\ \delta_{2k+1} X_k \varepsilon_{2k+1} \end{pmatrix}.$$

Now, we use the fact that for all  $n$ ,  $\mathcal{F}_n^\mathcal{O}$  is a sub- $\sigma$  field of  $\mathcal{O} \vee \mathcal{F}_n$ , the independence between  $\mathcal{O}$  and  $\mathcal{F}_n$  under assumption **(HI)** and the moment hypothesis **(HN.1)** to obtain

$$\begin{aligned} \mathbb{E}[\delta_{2k}\varepsilon_{2k} \mid \mathcal{F}_{n-1}^\mathcal{O}] &= \delta_{2k}\mathbb{E}[\mathbb{E}[\varepsilon_{2k} \mid \mathcal{O} \vee \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}^\mathcal{O}] \\ &= \delta_{2k}\mathbb{E}[\mathbb{E}[\varepsilon_{2k} \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}^\mathcal{O}] = 0. \end{aligned}$$

We obtain similar results for the other entries of  $\Delta M_n$  as  $\delta_{2k+1}$  and  $X_k$  are also  $\mathcal{F}_{n-1}^\mathcal{O}$ -measurable. Hence,  $(M_n)$  is a  $\mathbb{F}^\mathcal{O}$ -martingale. It is clearly square integrable from assumption **(HN.1)**. The same measurability arguments together with assumption **(HN.2)** yield

$$\begin{aligned} &\mathbb{E}[\Delta M_n(\Delta M_n)^t \mid \mathcal{F}_{n-1}^\mathcal{O}] \\ &= \sum_{k \in \mathbb{G}_{n-1}} \begin{pmatrix} \sigma^2\delta_{2k} & \sigma^2\delta_{2k}X_k & \rho\delta_{2k}\delta_{2k+1} & \rho\delta_{2k}\delta_{2k+1}X_k \\ \sigma^2\delta_{2k}X_k & \sigma^2\delta_{2k}X_k^2 & \rho\delta_{2k}\delta_{2k+1}X_k & \rho\delta_{2k}\delta_{2k+1}X_k^2 \\ \rho\delta_{2k}\delta_{2k+1} & \rho\delta_{2k}\delta_{2k+1}X_k & \sigma^2\delta_{2k+1} & \sigma^2\delta_{2k+1}X_k \\ \rho\delta_{2k}\delta_{2k+1}X_k & \rho\delta_{2k}\delta_{2k+1}X_k^2 & \sigma^2\delta_{2k+1}X_k & \sigma^2\delta_{2k+1}X_k^2 \end{pmatrix}. \end{aligned}$$

Hence the result as  $\langle M \rangle_n = \sum_{\ell=1}^n \mathbb{E}[\Delta M_\ell(\Delta M_\ell)^t \mid \mathcal{F}_{\ell-1}^\mathcal{O}]$ . □

Our main results are direct consequences of the sharp asymptotic properties of the martingale  $(M_n)$ . In particular, we will extensively use the strong law of large numbers for locally square integrable real martingales given in Theorem 1.3.15 of [6]. Throughout this paper, we shall also use other auxiliary martingales, either with respect to the same filtration  $\mathbb{F}^\mathcal{O}$ , or with respect to other filtrations naturally embedded in our process, see Lemma 5.1.

### 4.2. Asymptotic results

We first give the asymptotic behavior of the matrices  $S_n^0$ ,  $S_n^1$  and  $S_n^{0,1}$ . This is the first step of our asymptotic results.

**Proposition 4.2.** *Suppose that assumptions **(HN.1)**, **(HN.2)**, **(HO)** and **(HI)** are satisfied. Then, for  $i \in \{0, 1\}$ , we have*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{S_n^i}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\varepsilon}} L^i \quad a.s. \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{S_n^{0,1}}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\varepsilon}} L^{0,1} \quad a.s.$$

In addition,  $L^0$  and  $L^1$ , hence  $\Sigma$  are definite positive.

A consequence of this proposition is the asymptotic behavior of the increasing process of the martingale  $(M_n)$ .

**Corollary 4.3.** *Suppose that assumptions **(HN.1)**, **(HN.2)**, **(HO)** and **(HI)** are satisfied. Then, we have*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Sigma_n}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\varepsilon}} \Sigma, \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Gamma_n}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\varepsilon}} \Gamma.$$

This result is the keystone of our asymptotic analysis. It enables us to prove sharp asymptotic properties for the martingale  $(M_n)$ .

**Theorem 4.4.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} M_n^t \Sigma_{n-1}^{-1} M_n = \mathcal{O}(n) \quad a.s. \tag{4.2}$$

In addition, we also have

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \frac{1}{n} \sum_{\ell=1}^n M_\ell^t \Sigma_{\ell-1}^{-1} M_\ell = 4\sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{4.3}$$

Moreover, we have the central limit theorem on  $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}).$$

As seen in Eq. (4.1),  $(\hat{\theta}_n - \theta)$  is closely linked to  $M_n$  and this last theorem is then the major step to establish the asymptotic properties of our estimators. The proof of this Theorem is given in Section 7. As explained before, it is a consequence of Proposition 4.2 which proof is detailed in Section 6. In between, Section 5 presents preliminary results in the form of laws of large number for the observation, noise and BAR processes.

### 5. Laws of large numbers

We now state some laws of large numbers involving the observation, noise and BAR processes. They are based on martingale convergence results, and we start with giving a general result of convergence for martingales adapted to our framework.

#### 5.1. Martingale convergence results

The following result is nothing but the strong law of large numbers for square integrable martingales, written in our peculiar setting, and will be repeatedly used.

**Lemma 5.1.** *Let  $\mathcal{G} = (\mathcal{G}_n)$  be some filtration,  $(H_n)$  and  $(G_n)$  be two sequences of random variables satisfying the following hypotheses:*

- (i) *for all  $n \geq 1$ , for all  $k \in \mathbb{G}_n$ ,  $H_k$  is  $\mathcal{G}_{n-1}$ -measurable,  $G_k$  is  $\mathcal{G}_n$ -measurable, and  $\mathbb{E}[(H_k G_k)^2] < +\infty$ ,*
- (ii) *there exist  $c^2 > 0$ ,  $r \in [-1, 1]$ , such that for all  $n \geq 1$ , for all  $k, p \in \mathbb{G}_n$ ,*

$$\mathbb{E}[G_k | \mathcal{G}_{n-1}] = 0, \quad \mathbb{E}[G_k G_p | \mathcal{G}_{n-1}] = \begin{cases} c^2 & \text{if } k = p, \\ rc^2 & \text{if } k \neq p \text{ and } [k/2] = [p/2], \\ 0 & \text{otherwise,} \end{cases}$$

(iii) there exists a sequence of real numbers  $(a_n)$  that tends to  $\infty$  such that  $\sum_{k \in \mathbb{T}_n} H_k^2 = \mathcal{O}(a_n)$ .

Then  $\sum_{k \in \mathbb{T}_n} H_k G_k$  is a  $\mathcal{G}$ -martingale and one has

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{k \in \mathbb{T}_n} H_k G_k = 0 \quad a.s.$$

*Proof.* Define  $D_n = \sum_{k \in \mathbb{T}_n} H_k G_k$ . Assumptions (i) and (ii) clearly yield that  $(D_n)$  is a square integrable martingale with respect to the filtration  $(\mathcal{G}_n)$ . Thanks to (ii), its increasing process satisfies

$$\begin{aligned} \langle D \rangle_n &= c^2 \left( \sum_{k \in \mathbb{T}_n} H_k^2 + 2r \sum_{k \in \mathbb{T}_{n-1}} H_{2k} H_{2k+1} \right) \\ &\leq c^2 \left( \sum_{k \in \mathbb{T}_n} H_k^2 + r \sum_{k \in \mathbb{T}_{n-1}} (H_{2k}^2 + H_{2k+1}^2) \right) \\ &\leq c^2 (r + 1) \sum_{k \in \mathbb{T}_n} H_k^2, \end{aligned}$$

and now, (iii) implies that  $\langle D \rangle_n = \mathcal{O}(a_n)$ . Finally, since the sequence  $(a_n)$  tends to  $\infty$ , Theorem 1.3.15 of [6] ensures that  $D_n = o(a_n)$  a.s.  $\square$

We also recall Lemma A.3 of [3] that will be useful in the sequel.

**Lemma 5.2.** *Let  $(A_n)$  be a sequence of real-valued matrices such that*

$$\sum_{n=0}^{\infty} \|A_n\| < \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \sum_{k=0}^n A_k = A.$$

*In addition, let  $(X_n)$  be a sequence of real-valued vectors which converges to a limiting value  $X$ . Then,*

$$\lim_{n \rightarrow \infty} \sum_{\ell=0}^n A_{n-\ell} X_\ell = AX.$$

### 5.2. Laws of large numbers for the observation process

We now give more specific results on the asymptotic behavior of the observation process  $(\delta_k)_{k \geq 1}$ . Recall the notation  $\mathbb{T}_n^i$  defined in (2.2).

**Lemma 5.3.** *Under the assumption (HO), we have the following convergences, for  $(i, j)$  in  $\{0, 1\}^2$*

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i} \delta_{2k+j} = p_{ij} \frac{\pi}{\pi - 1} W z^i \quad a.s.$$

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i} \delta_{2k} \delta_{2k+1} = p^{(i)}(1, 1) \frac{\pi}{\pi - 1} W z^i \quad a.s.$$



*Proof.* Recall that  $\delta_{2k+j} = \delta_k \zeta_k^j$ , so that

$$\sum_{k \in \mathbb{T}_n^i} \delta_{2k+j} = p_{ij} \sum_{k \in \mathbb{T}_n^i} \delta_k + \sum_{k \in \mathbb{T}_n^i} \delta_k (\zeta_k^j - p_{ij}) = p_{ij} \left( i + \sum_{\ell=1}^n Z_\ell^i \right) + D_n,$$

since  $\mathbb{G}_0 = \{1\}$ , so that  $\mathbb{T}_n^i$  contains 1 or not, according to  $i = 1$  or not, and where  $D_n = \sum_{k \in \mathbb{T}_n^i} \delta_k (\zeta_k^j - p_{ij})$ . To deal with  $D_n$ , we use Lemma 5.1, with  $\mathcal{G} = (\mathcal{Z}_n)$  (recall that  $\mathcal{Z}_n = \sigma\{\zeta_k, k \in \mathbb{T}_n\}$ ),  $H_k = \delta_k \mathbb{1}_{\{k \in \mathbb{T}^i\}}$ , and  $G_k = (\zeta_k^j - p_{ij}) \mathbb{1}_{\{k \in \mathbb{T}^i\}}$ . Assumption (i) of Lemma 5.1 is obviously satisfied, since  $\delta_k$ , for  $k \in \mathbb{G}_n$ , is  $\mathcal{Z}_{n-1}$ -measurable. Regarding (ii), since the sequence  $(\zeta_k^j)$  is a sequence of i.i.d. random variables with expectation  $p_{ij}$  and variance  $\sigma_{ij}^2$ , we have  $\mathbb{E}[G_k | \mathcal{Z}_{n-1}] = 0$  and  $\mathbb{E}[G_k^2 | \mathcal{Z}_{n-1}] = \sigma_{ij}^2$ , for  $k \in \mathbb{G}_n$ , and  $\mathbb{E}[G_k G_p | \mathcal{Z}_{n-1}] = 0$ , for  $k \neq p \in \mathbb{G}_n$ . Finally, we turn to assumption (iii):

$$\sum_{k \in \mathbb{T}_n} H_k^2 = \sum_{k \in \mathbb{T}_n^i} \delta_k = i + \sum_{\ell=1}^n Z_\ell^i = \mathcal{O}(\pi^n),$$

thanks to (HO) and Eq. (2.7). Finally,  $D_n = o(\pi^n)$ , and again using Eq. (2.7), we obtain the first limit. The proof of the second one is similar using the  $\mathcal{Z}$ -martingale:

$$\sum_{k \in \mathbb{T}_n^i} \delta_k (\delta_{2k} \delta_{2k+1} - p^{(i)}(1, 1)) = \sum_{k \in \mathbb{T}_n} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} \delta_k}_{H_k} \underbrace{(\zeta_k^0 \zeta_k^1 - p^{(i)}(1, 1))}_{G_k},$$

and Lemma 5.1 again. □

### 5.3. Laws of large numbers for the noise process

We need to establish strong laws of large numbers for the noise sequence  $(\varepsilon_n)$  restricted to the observed indices.

**Lemma 5.4.** *Under assumptions (HN.1), (HN.2), (HO), (HI) and for  $i \in \{0, 1\}$ , one has*

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k+i} \varepsilon_{2k+i} = 0 \quad a.s.$$

*Proof.* Set

$$P_n^i = \sum_{k \in \mathbb{T}_n} \underbrace{\delta_{2k+i}}_{H_k} \underbrace{\varepsilon_{2k+i}}_{G_k}.$$

We use Lemma 5.1, with  $\mathcal{G} = (\mathbb{F}_{n+1}^\mathcal{O})$ . Assumption (i) is obvious. For  $k \in \mathbb{G}_{n+1}^i$ , we have  $\mathbb{E}[G_k | \mathbb{F}_{n+1}^\mathcal{O}] = 0$  and  $\mathbb{E}[G_k^2 | \mathbb{F}_{n+1}^\mathcal{O}] = \sigma^2$ , and  $\mathbb{E}[G_k G_p | \mathbb{F}_{n+1}^\mathcal{O}] = 0$ , for  $k \neq p \in \mathbb{G}_{n+1}^i$ . Finally, we turn to assumption (iii):

$$\sum_{k \in \mathbb{T}_n} H_k^2 = \sum_{k \in \mathbb{T}_n} \delta_{2k+i}^2 = \sum_{\ell=1}^{n+1} Z_\ell^i = \mathcal{O}(\pi^n),$$

as  $n$  tends to infinity, thanks to Eq. (2.7). We obtain the result.  $\square$

**Lemma 5.5.** *Under assumptions (HN.1), (HN.2), (HO), (HI) and for  $i \in \{0, 1\}$ , one has*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_k &= \sigma^2 \frac{\pi}{\pi - 1} W z^i \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \varepsilon_{2k} \varepsilon_{2k+1} &= \rho p^{(i)}(1, 1) \frac{\pi}{\pi - 1} W z^i \quad a.s. \end{aligned}$$

*Proof.* In order to prove the first convergence, we apply again Lemma 5.1 to the  $\mathbb{F}^{\mathcal{O}}$ -martingale:

$$Q_n = \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} (\varepsilon_k^2 - \sigma^2) \delta_k = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} \delta_k}_{H_k} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} (\varepsilon_k^2 - \sigma^2)}_{G_k},$$

Under (HN.1), (HN.2), we have  $\mathbb{E}[G_k | \mathbb{F}_n^{\mathcal{O}}] = 0$  and  $\mathbb{E}[G_k^2 | \mathbb{F}_n^{\mathcal{O}}] = \tau^4 - \sigma^4$ , and  $\mathbb{E}[G_k G_p | \mathbb{F}_n^{\mathcal{O}}] = 0$ , for  $k \neq p \in \mathbb{G}_n$ . Thanks to Eq. (2.7), we have:

$$\frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i} \delta_k = \frac{1}{\pi^n} \sum_{\ell=1}^n Z_\ell^i \xrightarrow{n \rightarrow \infty} \frac{\pi}{\pi - 1} W z^i \quad a.s.$$

which both implies assumption (iii) and the first convergence. To prove the second convergence, we write

$$\begin{aligned} &\frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \varepsilon_{2k} \varepsilon_{2k+1} \\ &= \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} \delta_{2k} \delta_{2k+1}}_{H_k} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} (\varepsilon_{2k} \varepsilon_{2k+1} - \rho)}_{G_k} + \frac{1}{\pi^n} \rho \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \end{aligned}$$

We use Lemma 5.1 to prove that the first term converges to 0; Lemma 5.3 gives the limit of the second term.  $\square$

**Corollary 5.6.** *Under assumptions (HN.1), (HN.2), (HO), (HI) and for  $i \in \{0, 1\}$ , one has*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_{2k+j} &= \sigma^2 p_{ij} \frac{\pi}{\pi - 1} W z^i \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \varepsilon_{2k} \varepsilon_{2k+1} &= \rho \bar{p}(1, 1) \frac{\pi}{\pi - 1} W \quad a.s. \end{aligned}$$

*Proof.* The proof of the first limit is similar to the preceding ones, using the decomposition  $\delta_{2k+j} = \delta_k \zeta_k^j$  and the properties of the sequence  $(\zeta_n^j)$ . Using Lemma 5.5 the second one is straightforward.  $\square$

**Lemma 5.7.** *Under assumptions (HN.1), (HN.2), (HO), (HI) and for  $i \in \{0, 1\}$ , we have*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_k \varepsilon_k^4 &= \tau^4 \frac{\pi}{\pi - 1} W z^i \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_{n-1}^i} \delta_{2k} \delta_{2k+1} \varepsilon_{2k}^2 \varepsilon_{2k+1}^2 &= \nu^2 \tau^4 p^{(i)}(1, 1) \frac{\pi}{\pi - 1} W z^i \quad a.s. \end{aligned}$$

*Proof.* The proof follows essentially the same lines as the proof of Lemma 5.5 using the square integrable real martingales

$$Q_n = \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_k (\varepsilon_k^4 - \tau^4), \quad \text{and} \quad R_n = \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2j} \delta_{2j+1} (\varepsilon_{2j}^2 \varepsilon_{2j+1}^2 - \nu^2 \tau^4).$$

It is therefore left to the reader. □

## 6. Convergence of the increasing process

We can now turn to the proof of our keystone result, the convergence of the increasing process of the main martingale  $(M_n)$ .

### 6.1. Preliminary results

We first need an upper bound of the normalized sums of the  $\delta_{2n+i} X_n^2$ , and  $\delta_{2n} \delta_{2n+1} X_n^2$  before being able to deduce their limits.

**Lemma 6.1.** *Under assumptions (HN.1), (HN.2), (HI) and (HO), and for  $i \in \{0, 1\}$ , we have*

$$\sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k^2 = \mathcal{O}(\pi^n) \quad \text{and} \quad \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^2 = \mathcal{O}(\pi^n) \quad a.s.$$

*Proof.* In all the sequel, for all  $k \geq 1$ , define  $a_{2k} = a$ ,  $b_{2k} = b$ ,  $a_{2k+1} = c$ ,  $b_{2k+1} = d$  and  $\eta_k = a_k + \varepsilon_k$  with the convention that  $\eta_1 = 0$ . It follows from a recursive application of relation (2.1) that, for all  $k \geq 1$ ,

$$X_k = \left( \prod_{\ell=0}^{r_k-1} b_{\lfloor \frac{k}{2^\ell} \rfloor} \right) X_1 + \sum_{\ell=0}^{r_k-1} \left( \prod_{p=0}^{\ell-1} b_{\lfloor \frac{k}{2^p} \rfloor} \right) \eta_{\lfloor \frac{k}{2^\ell} \rfloor}, \quad (6.1)$$

with the convention that an empty product equals 1. Set  $\alpha = \max(|a|, |c|)$ ,  $\beta = \max(|b|, |d|)$  and notice that  $0 < \beta < 1$ . The proof of Lemma A.5 in [3] yields

$$\begin{aligned} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} X_k^2 &\leq \frac{4}{1-\beta} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \varepsilon_{\lfloor \frac{k}{2^\ell} \rfloor}^2 + \frac{4\alpha^2}{1-\beta} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \\ &\quad + 2X_1^2 \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \beta^{2r_k}, \\ &\leq \frac{4}{1-\beta} A_n^i + \frac{4\alpha^2}{1-\beta} B_n^i + 2X_1^2 C_n^i, \end{aligned} \tag{6.2}$$

where, for  $i \in \{0, 1\}$ ,

$$A_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \varepsilon_{\lfloor \frac{k}{2^\ell} \rfloor}^2, \quad B_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell, \quad C_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \beta^{2r_k}.$$

The last two terms above are readily evaluated by splitting the sums generation-wise. Indeed, the last term can be rewritten as

$$C_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \beta^{2\ell} = \sum_{\ell=1}^n \beta^{2\ell} Z_{\ell+1}^i = \pi^n \sum_{\ell=1}^n (\pi^{-1})^{n-\ell} \left( \beta^{2\ell} \frac{Z_{\ell+1}^i}{\pi^\ell} \right).$$

We now use Lemma 5.2 with  $\mathbf{A}_n = \pi^{-n}$  and  $\mathbf{X}_n = \beta^{2n} Z_{n+1}^i \pi^{-n}$ . On the one hand, the series of  $(\pi^{-n})$  converges to  $\pi/(\pi-1)$  as  $\pi > 1$  by assumption; on the other hand,  $\beta^{2n}$  tends to 0 as  $n$  tends to infinity as  $\beta < 1$ , and  $Z_n^i \pi^{-n}$  converges a.s. to  $Wz^i$  according to Eq. (2.7), hence  $\beta^{2n} Z_{n+1}^i \pi^{-n}$  tends to 0 as  $n$  tends to infinity. Lemma 5.2 thus yields

$$\lim_{n \rightarrow \infty} \sum_{\ell=1}^n (\pi^{-1})^{n-\ell} \left( \beta^{2\ell} \frac{Z_{\ell+1}^i}{\pi^\ell} \right) = 0 \quad \text{and} \quad C_n^i = o(\pi^n) \quad \text{a.s.}$$

We now turn to the term  $B_n^i$ :

$$B_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \frac{1-\beta^\ell}{1-\beta} \leq \frac{1}{(1-\beta)} \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \leq \frac{|\mathbb{T}_{n+1}^*|}{(1-\beta)} = \mathcal{O}(\pi^n),$$

due to Lemma 2.1. It remains to control the first term  $A_n^i$ . Note that  $\varepsilon_k$  appears in  $A_n^i$  as many times as it has descendants up to the  $n$ -th generation, and its multiplicative factor for its  $p$ -th generation descendant  $k$  is  $\beta^p \delta_{2k}$ . This leads to

$$A_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \varepsilon_k^2 \sum_{p=0}^{n-\ell} \beta^p \sum_{m=0}^{2^p-1} \delta_{2(2^p k+m)+i}.$$

Now, note that  $\sum_{m=0}^{2^p-1} \delta_{2(2^p k+m)+i} = \delta_k \sum_{m=0}^{2^p-1} \delta_{2(2^p k+m)+i}$  is the number of descendants of type  $i$  of individual  $k$  after  $p+1$  generations. We denote it  $Z_{p+1}^i(k)$ , and split  $A_n^i$  the following way:

$$A_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \sigma^2 \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k) + \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k). \tag{6.3}$$

We first deal with the second term of the above sum.

$$\begin{aligned} \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k) &= \sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \delta_k Z_{p+1}^i(k) \\ &= \sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} Y_{\ell,p}^i, \end{aligned}$$

where  $Y_{\ell,p}^i = \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \delta_k Z_{p+1}^i(k)$ . Tedious but straightforward computations lead to the following expression for the second order moment of  $Y_{\ell,p}^i$ , relying on assumptions **(HI)**, **(HN.1)** and **(HN.2)**. We also use the fact that, for  $k \in \mathbb{G}_\ell$ , conditionally to  $\{\delta_k = 1\}$ ,  $Z_{p+1}^i(k)$  follows the same law as  $Z_{p+1}^i$ , and is independent of any  $Z_{p+1}^i(k')$ , for  $k' \neq k \in \mathbb{G}_\ell$ .

$$\begin{aligned} \mathbb{E}[(Y_{\ell,p}^i)^2] &= (\tau^4 - \sigma^4) \mathbb{E}[Z_\ell^0 + Z_\ell^1] \mathbb{E}[(Z_{p+1}^i)^2] \\ &\quad + (\nu^2 \tau^4 - \sigma^4) \mathbb{E}[Z_{p+1}^i]^2 \mathbb{E}\left[\sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k} \delta_{2k+1}\right] \\ &\leq (\tau^4 - \sigma^4) \mathbb{E}[Z_\ell^0 + Z_\ell^1] \left(\mathbb{E}[(Z_{p+1}^i)^2] + \mathbb{E}[Z_{p+1}^i]^2\right), \end{aligned}$$

since  $\sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k} \delta_{2k+1} \leq \sum_{k \in \mathbb{G}_{\ell-1}} (\delta_{2k} + \delta_{2k+1}) = Z_\ell^0 + Z_\ell^1$ . Now, using results on the moments of a two-type Galton-Watson process (see e.g. [12]), we know that  $\mathbb{E}[(Z_{p+1}^i)^2] = \mathcal{O}(\pi^{2p})$ . Recall Eq. (2.7) to obtain that  $\mathbb{E}[(Y_{\ell,p}^i)^2] = \mathcal{O}(\pi^\ell \pi^{2p})$ , which immediately entails that  $|Y_{\ell,p}^i| = o(\pi^{\alpha\ell} \pi^{\gamma p})$  a.s., for any  $\alpha > 1/2$  and  $\gamma > 1$ . We thus one gets

$$\sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} Y_{\ell,p}^i = \mathcal{O}((\beta\pi^\gamma)^n) = \mathcal{O}(\pi^n) \quad \text{a.s.},$$

since we can choose  $\gamma$  close enough to 1 to get  $\beta\pi^\gamma \leq \pi$ , as  $\beta < 1$ . We have thus proved that the second term in the sum in (6.3) is  $\mathcal{O}(\pi^n)$ , we now turn to the first one

$$\begin{aligned} &\sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \sigma^2 \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k) \\ &= \sigma^2 \sum_{\ell=1}^n \sum_{p=0}^{n-\ell} \beta^p \sum_{k \in \mathbb{G}_\ell} \delta_k Z_{p+1}^i(k) = \sigma^2 \sum_{\ell=1}^n \sum_{p=0}^{n-\ell} \beta^p Z_{\ell+p+1}^i \\ &= \sigma^2 \sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} Z_{\ell+p+1}^i \leq \sigma^2 \sum_{p=0}^{n-1} \beta^p |\mathbb{T}_{n+1}^*| = \mathcal{O}(\pi^n) \quad \text{a.s.} \end{aligned}$$

Finally,  $A_n^i = \mathcal{O}(\pi^n)$ , and the first result of the Lemma is proved. The second result follows immediately from the remark that the second sum in Lemma 6.1 is clearly smaller than the first one.  $\square$

**Lemma 6.2.** *Under assumptions (HN.1), (HN.2), (HI) and (HO), and for  $i \in \{0, 1\}$ , we have*

$$\sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k^4 = \mathcal{O}(\pi^n) \quad \text{and} \quad \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^4 = \mathcal{O}(\pi^n) \quad \text{a.s.}$$

*Proof.* The proof mimics that of Lemma 6.1. Instead of Equation (6.2), we have

$$\sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} X_k^4 \leq \frac{64}{(1-\beta)^3} A_n^i + \frac{64\alpha^4}{(1-\beta)^3} B_n^i + 8X_1^4 C_n^i$$

with, for  $i$  in  $\{0, 1\}$

$$A_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \varepsilon_{[\frac{k}{2^\ell}]}, \quad B_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell, \quad C_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \beta^{4r_k}.$$

We can easily prove that  $(B_n^i + C_n^i) = \mathcal{O}(\pi^n)$ . Therefore, we only need a sharper estimate for  $A_n^i$ . Via the same lines as in the proof of Lemma 6.1, but dealing with  $\varepsilon_k^4$  instead of  $\varepsilon_k^2$ , we can show that  $A_n^i = \mathcal{O}(\pi^n)$  a.s. which immediately yields the first result. The second one is obtained by remarking that the second sum is less than the first one. □

**6.2. Asymptotic behavior of the sum of observed data**

We now turn to the asymptotic behavior of the sums of the observed data. More precisely, set  $H_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k$ , for  $i$  in  $\{0, 1\}$ , and  $\mathbf{H}_n = (H_n^0, H_n^1)^t$ . The following result gives the asymptotic behavior of  $(\mathbf{H}_n)$ .

**Proposition 6.3.** *Under assumptions (HN.1), (HN.2), (HI) and (HO), we have the convergence:*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{H}_n}{\pi^n} = \frac{\pi}{\pi-1} \mathbf{W} \mathbf{h} \quad \text{a.s.,}$$

where

$$\mathbf{h} = \begin{pmatrix} h^0 \\ h^1 \end{pmatrix} = (\mathbf{I}_2 - \tilde{\mathbf{P}}_1)^{-1} \mathbf{P}^t \begin{pmatrix} az^0 \\ cz^1 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{P}}_1 = \frac{1}{\pi} \mathbf{P}^t \begin{pmatrix} b & 0 \\ 0 & d \end{pmatrix}.$$

*Proof.* We first prove that the sequence  $(\mathbf{H}_n)$  satisfies a recursive property using Equation (2.1).

$$\begin{aligned} H_n^0 &= X_1 \delta_2 + \sum_{k \in \mathbb{T}_n^0} \left( a + bX_{[\frac{k}{2}]} + \varepsilon_k \right) \delta_{2k} + \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \left( c + dX_{[\frac{k}{2}]} + \varepsilon_k \right) \delta_{2k} \\ &= X_1 \delta_2 + a \sum_{k \in \mathbb{T}_n^0} \delta_{2k} + b \sum_{k \in \mathbb{T}_n^0} X_{[\frac{k}{2}]} \delta_{2k} + c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k} + d \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} X_{[\frac{k}{2}]} \delta_{2k} \\ &\quad + \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k} \\ &= bp_{00} H_{n-1}^0 + dp_{10} H_{n-1}^1 + B_n^0, \end{aligned}$$

with

$$B_n^0 = X_1\delta_2 + a \sum_{k \in \mathbb{T}_n^0} \delta_{2k} + c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k} + \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k} \\ + b \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k} (\delta_{4k} - p_{00}) + d \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+1} (\delta_{4k+2} - p_{10}).$$

Similarly, we have

$$H_n^1 = bp_{01}H_{n-1}^0 + dp_{11}H_{n-1}^1 + B_n^1,$$

with

$$B_n^1 = X_1\delta_3 + a \sum_{k \in \mathbb{T}_n^0} \delta_{2k+1} + c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k+1} + \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k+1} \\ + b \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k} (\delta_{4k+1} - p_{01}) + d \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+1} (\delta_{4k+3} - p_{11}).$$

Let us denote  $\mathbf{B}_n = (B_n^0, B_n^1)^t$ . The last equations yield in the matrix form:

$$\frac{\mathbf{H}_n}{\pi^n} = \tilde{\mathbf{P}}_1 \frac{\mathbf{H}_{n-1}}{\pi^{n-1}} + \frac{\mathbf{B}_n}{\pi^n} = \tilde{\mathbf{P}}_1^n \mathbf{H}_0 + \sum_{k=1}^n \tilde{\mathbf{P}}_1^{n-k} \frac{\mathbf{B}_k}{\pi^k},$$

with

$$\tilde{\mathbf{P}}_1 = \frac{1}{\pi} \begin{pmatrix} bp_{00} & dp_{10} \\ bp_{01} & dp_{11} \end{pmatrix} = \frac{1}{\pi} \mathbf{P}^t \begin{pmatrix} b & 0 \\ 0 & d \end{pmatrix}.$$

One has  $\|\tilde{\mathbf{P}}_1^n\| \leq \pi^{-n} \beta^n \|\mathbf{P}^n\|$ . It is well known that  $\pi^{-n} \mathbf{P}^n$  converges to a fixed matrix (see e.g. [13]) as  $\mathbf{P}$  is a positive matrix with dominant eigenvalue  $\pi$ . Since  $\beta < 1$ , the sequence  $\tilde{\mathbf{P}}_1^n$  thus converges to 0 as  $n$  tends to infinity. In addition,  $\sum \|\tilde{\mathbf{P}}_1^n\|$  is bounded,  $\mathbf{I}_2 - \tilde{\mathbf{P}}_1$  is invertible and  $\sum_{n \geq 0} \tilde{\mathbf{P}}_1^n$  converges to  $(\mathbf{I}_2 - \tilde{\mathbf{P}}_1)^{-1}$ . In order to use Lemma 5.2, we need to compute the limit of  $\mathbf{B}_n/\pi^n$ . First, we prove that

$$\sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k+i} = o(\pi^n), \tag{6.4}$$

for  $i \in \{0, 1\}$ , thanks to Lemma 5.1. Indeed, set  $\mathcal{G} = \mathbb{F}^{\mathcal{O}}$ ,  $H_k = \delta_{2k+i}$ ,  $G_k = \varepsilon_k$ . Thus hypothesis (i) of Lemma 5.1 is obvious, (ii) comes from (HN.1) and (HN.2). Finally, the last assumption (iii) holds, since

$$\sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i}^2 = \sum_{\ell=1}^{n+1} Z_\ell^i = \mathcal{O}(\pi^n),$$

the last equality coming from (2.7), which holds thanks to (HO). Now, we turn to the terms

$$\sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij}) = \sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} (\zeta_{2k+i}^j - p_{ij}),$$

for  $(i, j) \in \{0, 1\}^2$ . We use again Lemma 5.1, with the following setting:  $(\mathcal{G}_n) = (\mathcal{Z}_{n+1} \vee \mathcal{F}_{n+1})$ ,  $H_k = X_k \delta_{2k+i}$ ,  $G_k = \zeta_{2k+i}^j - p_{ij}$ . For  $k \in \mathbb{T}_n$ , we check that  $X_k \delta_{2k+i}$  is  $\mathcal{G}_{n-1}$ -measurable, since  $X_k$  is  $\mathcal{F}_n$ -measurable and  $\delta_{2k+i}$  is  $\mathcal{Z}_n$ -measurable. Next, because of (HI) and of the independence of the sequence  $(\zeta_k^j)$ ,  $\mathbb{E}[\zeta_{2k+i}^j - p_{ij} | \mathcal{Z}_n \vee \mathcal{F}_n] = 0$ . The same independence hypothesis yields that  $\mathbb{E}[G_k G_p | \mathcal{Z}_n \vee \mathcal{F}_n] \neq 0$  only if  $k = p$ , and then equals  $\sigma_{ij}^2$ . Finally,

$$\sum_{k \in \mathbb{T}_n} (X_k \delta_{2k+i})^2 = \sum_{k \in \mathbb{T}_n} X_k^2 \delta_{2k+i} = \mathcal{O}(\pi^n),$$

thanks to Lemma 6.1. Now, Lemma 5.1 allows to conclude that

$$\sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij}) = o(\pi^n), \tag{6.5}$$

for  $(i, j) \in \{0, 1\}^2$ . Next, Lemma 5.3 gives the limit of the term  $\sum_{k \in \mathbb{T}_n^i} \delta_{2k+j}$ , for  $(i, j) \in \{0, 1\}^2$ , so that we finally obtain:

$$\lim_{n \rightarrow \infty} \frac{\mathbf{B}_n}{\pi^n} = W \frac{\pi}{\pi - 1} \begin{pmatrix} az^0 p_{00} + cz^1 p_{10} \\ az^0 p_{01} + cz^1 p_{11} \end{pmatrix} = W \frac{\pi}{\pi - 1} \mathbf{P}^t \begin{pmatrix} az^0 \\ cz^1 \end{pmatrix} \quad \text{a.s.}$$

and we use Lemma 5.2 to conclude. □

**Remark 6.4.** Putting together Proposition 6.3 and Eq. (6.5) above, we immediately get that under the same assumptions as that of Proposition 6.3,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} \delta_{2(2k+i)+j} = \frac{\pi}{\pi - 1} h^i p_{ij} W \quad \text{a.s.}$$

for all  $(i, j) \in \{0, 1\}^2$ , result we will use for the study of the limit of  $\sum X_k^2 \delta_{2k+i}$ .

### 6.3. Asymptotic behavior of the sum of squared observed data

We now turn to the asymptotic behavior of the sums of the squared observed data. Set  $K_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k^2$ , for  $i$  in  $\{0, 1\}$ , and  $\mathbf{K}_n = (K_n^0, K_n^1)^t$ . The following result gives the asymptotic behavior of  $(\mathbf{K}_n)$ .

**Proposition 6.5.** *Under assumptions (HN.1), (HN.2), (HI) and (HO), we have the convergence:*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{K}_n}{\pi^n} = \frac{\pi}{\pi - 1} W \mathbf{k} \quad \text{a.s.},$$

where

$$\mathbf{k} = \begin{pmatrix} k^0 \\ k^1 \end{pmatrix} = (\mathbf{I}_2 - \tilde{\mathbf{P}}_2)^{-1} \mathbf{P}^t \begin{pmatrix} (a^2 + \sigma^2)z^0 + \frac{2}{\pi}abh^0 \\ (c^2 + \sigma^2)z^1 + \frac{2}{\pi}cdh^1 \end{pmatrix},$$

and

$$\tilde{\mathbf{P}}_2 = \frac{1}{\pi} \mathbf{P}^t \begin{pmatrix} b^2 & 0 \\ 0 & d^2 \end{pmatrix}.$$



*Proof.* We use again Equation (2.1) to prove a recursive property for the sequence  $(\mathbf{K}_n)$ . Following the same lines as in the proof of Proposition 6.3, we obtain:

$$\frac{\mathbf{K}_n}{\pi^n} = \tilde{\mathbf{P}}_2 \frac{\mathbf{K}_{n-1}}{\pi^{n-1}} + \frac{\mathbf{C}_n}{\pi^n} = \tilde{\mathbf{P}}_2^n \mathbf{K}_0 + \sum_{\ell=1}^n \tilde{\mathbf{P}}_2^{n-\ell} \frac{\mathbf{C}_\ell}{\pi^\ell},$$

where  $\mathbf{C}_n = (C_n^0, C_n^1)^t$  is defined by

$$\begin{aligned} C_n^i &= X_1^2 \delta_{2+i} + a^2 \sum_{k \in \mathbb{T}_n^0} \delta_{2k+i} + b^2 \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k} (\delta_{4k+i} - p_{0i}) \\ &+ 2ab \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k} \delta_{4k+i} + 2a \sum_{k \in \mathbb{T}_n^0} \varepsilon_k \delta_{2k+i} + 2b \sum_{k \in \mathbb{T}_n^0} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+i} \\ &+ \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_{2k+i} + c^2 \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k+i} + d^2 \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+1} (\delta_{4k+2+i} - p_{1i}) \\ &+ 2cd \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+1} \delta_{4k+2+i} + 2c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k+i} + 2d \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+i}, \end{aligned}$$

for  $i \in \{0, 1\}$ . Note that  $\|\tilde{\mathbf{P}}_2^n\| \leq \pi^{-n} \beta^{2n} \|\mathbf{P}^n\|$ , so that  $\tilde{\mathbf{P}}_2^n$  converges to 0. In addition,  $\sum \|\tilde{\mathbf{P}}_2^n\|$  is bounded,  $\mathbf{I}_2 - \tilde{\mathbf{P}}_2$  is invertible and  $\sum_{n \geq 0} \tilde{\mathbf{P}}_2^n$  converges to  $(\mathbf{I}_2 - \tilde{\mathbf{P}}_2)^{-1}$ . In order to use Lemma 5.2, we have to compute the limit of  $\mathbf{C}_n/\pi^n$ . Following the proof of (6.4), we already have, for  $(i, j) \in \{0, 1\}^2$ ,

$$\sum_{k \in \mathbb{T}_n^j} \varepsilon_k \delta_{2k+i} = o(\pi^n) \quad \text{a.s.}$$

We now turn to the terms  $\sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij})$ , for  $(i, j) \in \{0, 1\}^2$ . To deal with these terms, we use Lemma 5.1 with the same setting we used to prove Eq. (6.5), except that we replace  $X_k$  with  $X_k^2$ . Assumptions (i) and (ii) of Lemma 5.1 have thus already been checked, and regarding (iii), we have  $\sum_{k \in \mathbb{T}_{n-1}} X_k^4 \delta_{2k+i} = \mathcal{O}(\pi^n)$  a.s. thanks to Lemma 6.2. We conclude that

$$\sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij}) = o(\pi^n) \quad \text{a.s.}$$

Next, we study  $\sum_{k \in \mathbb{T}_n^i} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+j}$ , for  $(i, j) \in \{0, 1\}^2$ . We use the same martingale tool, so to speak Lemma 5.1, with  $\mathcal{G} = \mathbb{F}^\mathcal{O}$ ,  $H_k = X_{[\frac{k}{2}]} \delta_{2k+j} \mathbb{1}_{\{k \in \mathbb{T}^i\}}$  and  $G_k = \varepsilon_k$ . Assumptions (i) and (ii) are easily checked, and since

$$\sum_{k \in \mathbb{T}_n^i} X_{[\frac{k}{2}]}^2 \delta_{2k+j} = \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2(2k+i)+j} \leq \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+i} = \mathcal{O}(\pi^n),$$

the last equality coming from Lemma 6.1, assumption (iii) is satisfied and

$$\sum_{k \in \mathbb{T}_n^i} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+j} = o(\pi^n) \quad \text{a.s.}$$

Now, Corollary 5.6 yields that for  $i \in \{0, 1\}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_{2k+i} = \sigma^2(p_{0i}z^0 + p_{1i}z^1) \frac{\pi}{\pi - 1} W \quad \text{a.s.}$$

Finally, Remark 6.4 gives the limit of  $\pi^{-n} \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+i} \delta_{2(2k+i)+j}$ , and Lemma 5.3 that of  $\pi^{-n} \sum_{k \in \mathbb{T}_n^j} \delta_{2k+i}$ , so that we finally obtain

$$\lim_{n \rightarrow \infty} \frac{C_n}{\pi^n} = \frac{W\pi}{\pi - 1} \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} \times \begin{pmatrix} (a^2 + \sigma^2)z^0 + \frac{2}{\pi}abh^0 \\ (c^2 + \sigma^2)z^1 + \frac{2}{\pi}cdh^1 \end{pmatrix} \text{a.s.}$$

And we conclude using Lemma 5.2 again. □

Propositions 6.3 and 6.5 together with Equation (2.7) give the asymptotic behavior of the matrices  $S_n^0$  and  $S_n^1$ . The next result gives the behavior of matrix  $S_n^{0,1}$  given through the quantities  $\sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k$  and  $\sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^2$ . It is an easy consequence of Propositions 6.3 and 6.5, together with Lemma 5.3 for the first limit.

#### 6.4. Asymptotic behavior of covariance terms

Finally, we turn to the asymptotic behavior of the covariance terms, which are involved in matrix  $S_n^{0,1}$ . We thus define  $H_n^{01} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k$  and  $K_n^{01} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^2$ .

**Proposition 6.6.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have the almost sure convergences:*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} &= \frac{\pi}{\pi - 1} W \bar{p}(1, 1), \\ \lim_{n \rightarrow \infty} \frac{H_n^{01}}{\pi^n} &= \frac{\pi}{\pi - 1} W h^{0,1} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{K_n^{01}}{\pi^n} = \frac{\pi}{\pi - 1} W k^{0,1}, \end{aligned}$$

where

$$\begin{aligned} \bar{p}(1, 1) &= p^{(0)}(1, 1)z^0 + p^{(1)}(1, 1)z^1, \\ h^{0,1} &= p^{(0)}(1, 1) \left( az^0 + b \frac{h^0}{\pi} \right) + p^{(1)}(1, 1) \left( cz^1 + d \frac{h^1}{\pi} \right), \\ k^{0,1} &= p^{(0)}(1, 1) \left( a^2z^0 + b^2 \frac{k^0}{\pi} + 2ab \frac{h^0}{\pi} \right) \\ &\quad + p^{(1)}(1, 1) \left( c^2z^1 + d^2 \frac{k^1}{\pi} + 2cd \frac{h^1}{\pi} \right) + \sigma^2 \bar{p}(1, 1). \end{aligned} \tag{6.6}$$

*Proof.* The first limit is a consequence of Lemma 5.3. Next, using Eq. (2.1) we obtain  $H_n^{01} \pi^{-n}$  and  $K_n^{01} \pi^{-n}$  in terms of  $\pi^{-n} \sum_{k \in \mathbb{T}_{n-1}^i} \delta_k$ ,  $H_{n-1}^i \pi^{-n}$  and  $K_{n-1}^i \pi^{-n}$  and the result follows from Propositions 6.3 and 6.5. □

*Proof of Proposition 4.2.* We are now in a position to complete the proof of Proposition 4.2. Simply notice that we have proved in Propositions 6.3, 6.5 and 6.6 all the wished convergences, except that we normalized the sums with  $\pi^n$ . Thanks to Lemma 2.1, we end the proof.  $\square$

**Remark 6.7.** In the case of fully observed data, the matrix  $\mathbf{P}$  is a  $2 \times 2$  matrix with all entries equal to 1,  $\pi$  equals 2 and the normalized eigenvector  $\mathbf{z}$  equals  $(1/2, 1/2)$ . One can check that in that case, our limits correspond to those of [3].

### 7. Asymptotic behavior of the main martingale

Theorem 4.4 is a strong law of large numbers for the martingale  $(M_n)$ . The standard strong law for martingales is unhelpful here. Indeed, it is valid for martingales that can be decomposed in a sum of the form  $\sum_{\ell=1}^n \Psi_{\ell-1} \xi_{\ell}$  where  $(\Psi_{\ell})$  is predictable and  $(\xi_{\ell})$  is a martingale difference sequence. In addition,  $(\Psi_{\ell})$  and  $(\xi_{\ell})$  are required to be sequences of *fixed-size* vectors. Such a decomposition with fixed-sized vectors is impossible in our context (see Lemma A.2), essentially because the number of observed data in each generation asymptotically grows exponentially fast as  $\pi^n$ . Consequently, we are led to propose a new strong law of large numbers for  $(M_n)$ , adapted to our framework.

For all  $n \geq 1$ , let  $\mathcal{V}_n = M_n^t \Sigma_{n-1}^{-1} M_n$  where  $\Sigma_n$  is defined in Section 3.1. First of all, we have

$$\begin{aligned} \mathcal{V}_{n+1} &= (M_n + \Delta M_{n+1})^t \Sigma_n^{-1} (M_n + \Delta M_{n+1}), \\ &= \mathcal{V}_n - M_n^t (\Sigma_{n-1}^{-1} - \Sigma_n^{-1}) M_n + 2M_n^t \Sigma_n^{-1} \Delta M_{n+1} + \Delta M_{n+1}^t \Sigma_n^{-1} \Delta M_{n+1}. \end{aligned}$$

Note that  $M_n^t \Sigma_n^{-1} \Delta M_{n+1}$  and  $\Delta M_{n+1}^t \Sigma_n^{-1} M_{n+1}$  are scalars, hence they are equal to their own transpose and as a result, one has  $M_n^t \Sigma_n^{-1} \Delta M_{n+1} = \Delta M_{n+1}^t \Sigma_n^{-1} M_{n+1}$ . By summing over the identity above, we obtain the main decomposition

$$\mathcal{V}_{n+1} + \mathcal{A}_n = \mathcal{V}_1 + \mathcal{B}_{n+1} + \mathcal{W}_{n+1}, \tag{7.1}$$

where

$$\begin{aligned} \mathcal{A}_n &= \sum_{\ell=1}^n M_{\ell}^t (\Sigma_{\ell-1}^{-1} - \Sigma_{\ell}^{-1}) M_{\ell}, \\ \mathcal{B}_{n+1} &= 2 \sum_{\ell=1}^n M_{\ell}^t \Sigma_{\ell}^{-1} \Delta M_{\ell+1}, \quad \mathcal{W}_{n+1} = \sum_{\ell=1}^n \Delta M_{\ell+1}^t \Sigma_{\ell}^{-1} \Delta M_{\ell+1}. \end{aligned}$$

The asymptotic behavior of the left-hand side of (7.1) is as follows.

**Proposition 7.1.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|G_n^*| > 0\}} \frac{\mathcal{V}_{n+1} + \mathcal{A}_n}{n} = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

*Proof.* Thanks to the laws of large numbers derived in Sections 5 and 6, the proof of Proposition 7.1 follows essentially the same lines as [3] and is given in Appendix A.  $\square$

Since  $(\mathcal{V}_n)$  and  $(\mathcal{A}_n)$  are two sequences of non negative real numbers, Proposition 7.1 yields that  $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}\mathcal{V}_n = \mathcal{O}(n)$  a.s. which proves Equation (4.2). We now turn to the proof of Equation (4.3). We start with a sharp rate of convergence for  $(\mathbf{M}_n)$ .

**Proposition 7.2.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have, for all  $\eta > 1/2$ ,*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \|\mathbf{M}_n\|^2 = o(|\mathbb{T}_{n-1}^*|n^\eta) \quad a.s.$$

*Proof.* The result is obvious on  $\mathcal{E}$ . On  $\bar{\mathcal{E}}$ , the proof follows again the same lines as [3] thanks to the laws of large numbers derived in Sections 5 and 6. It is given in Appendix B.  $\square$

A direct application of Proposition 7.2 ensures that  $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}\mathcal{V}_n = o(n^\eta)$  a.s. for all  $\eta > 1/2$ . Hence, Proposition 7.1 immediately leads to the following result.

**Corollary 7.3.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \frac{\mathcal{A}_n}{n} = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

*Proof of Result (4.3) of Theorem 4.4.* First of all,  $\mathcal{A}_n$  may be rewritten as

$$\mathcal{A}_n = \sum_{\ell=1}^n \mathbf{M}_\ell^t (\boldsymbol{\Sigma}_{\ell-1}^{-1} - \boldsymbol{\Sigma}_\ell^{-1}) \mathbf{M}_\ell = \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \boldsymbol{\Delta}_\ell \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \mathbf{M}_\ell,$$

where  $\boldsymbol{\Delta}_n = \mathbf{I}_4 - \boldsymbol{\Sigma}_{n-1}^{1/2} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_{n-1}^{1/2}$ . Thanks to Corollary 4.3, we know that

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \boldsymbol{\Delta}_n = \frac{\pi - 1}{\pi} \mathbf{I}_4 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Besides, Corollary 7.3 yields that  $\mathcal{A}_n \sim n \frac{\pi-1}{\pi} 4\sigma^2$  a.s. on  $\bar{\mathcal{E}}$ . Plugging these two results into the equality

$$\mathcal{A}_n = \frac{\pi - 1}{\pi} \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell + \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \left( \boldsymbol{\Delta}_\ell - \frac{\pi - 1}{\pi} \mathbf{I}_4 \right) \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \mathbf{M}_\ell$$

gives that  $\sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell \sim \mathcal{A}_n \frac{\pi}{\pi-1}$  a.s. on  $\bar{\mathcal{E}}$  and convergence (4.3) directly follows.  $\square$

### 8. Proof of the main results

We can now proceed to proving our main results.

**8.1. Strong consistency for  $\widehat{\boldsymbol{\theta}}_n$**

Theorem 3.2 is a direct consequence of Theorem 4.4.

*Proof of result (3.3) of Theorem 3.2.* Recall that  $\mathcal{V}_n = \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n$ . It clearly follows from Equation (4.1) that

$$\mathcal{V}_n = (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^t \boldsymbol{\Sigma}_{n-1} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}).$$

Consequently, the asymptotic behavior of  $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$  is related to the one of  $\mathcal{V}_n$ . More precisely, we can deduce from Corollary 4.3 and the fact that the eigenvalues of a matrix are continuous functions of its coefficients the following result

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\lambda_{\min}(\boldsymbol{\Sigma}_n)}{|\mathbb{T}_n^*|} = \lambda_{\min}(\boldsymbol{\Sigma}) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

where  $\lambda_{\min}(\mathbf{A})$  denotes the smallest eigenvalue of matrix  $\mathbf{A}$ . Since  $\mathbf{L}$  as well as  $\boldsymbol{\Sigma}$  is definite positive, one has  $\lambda_{\min}(\boldsymbol{\Sigma}) > 0$ . Therefore, as

$$\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 \leq \frac{\mathcal{V}_n}{\lambda_{\min}(\boldsymbol{\Sigma}_{n-1})},$$

we use Result (4.2) of Theorem 4.4 to conclude that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = \mathcal{O}\left(\frac{n}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\overline{\mathcal{E}}} = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

which completes the proof of results (3.3). □

We now prove the quadratic strong law (QSL).

*Proof of result (3.4) of Theorem 3.2.* The QSL is a direct consequence of result (4.3) of Theorem 4.4 together with the fact that  $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n$ . Indeed, we have

$$\begin{aligned} & \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell \\ &= \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \boldsymbol{\Sigma}_{\ell-1} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) \\ &= \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \mathbb{1}_{\{|\mathbb{G}_{\ell-1}^*| > 0\}} \frac{\boldsymbol{\Sigma}_{\ell-1}}{|\mathbb{T}_{\ell-1}^*|} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) \\ &= \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \boldsymbol{\Sigma} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) + o(1) \quad \text{a.s.} \end{aligned}$$

which completes the proof. □

### 8.2. Strong consistency for the variance estimators

For  $n \geq 1$ , set

$$\mathbf{V}_k = (\delta_{2k}\varepsilon_{2k}, \delta_{2k+1}\varepsilon_{2k+1})^t, \quad \widehat{\mathbf{V}}_k = (\delta_{2k}\widehat{\varepsilon}_{2k}, \delta_{2k+1}\widehat{\varepsilon}_{2k+1})^t.$$

The almost sure convergence of  $\widehat{\sigma}_n^2$  and  $\widehat{\rho}_n$  is strongly related to that of  $\widehat{\mathbf{V}}_k - \mathbf{V}_k$ .

*Proof of result (3.5) of Theorem 3.3.* Equation(3.5) can be rewritten as

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 = 4(\pi - 1)\sigma^2 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

Once again, we are searching for a link between the sum of  $\|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|$  and the processes  $(\mathcal{A}_n)$  and  $(\mathcal{V}_n)$  whose convergence properties were previously investigated. For  $i \in \{0, 1\}$  and  $n \geq 0$ , let

$$\Phi_n^i = \begin{pmatrix} \delta_{2(2^n)+i} & \delta_{2(2^{n+1})+i} & \cdots & \delta_{2(2^{n+1}-1)+i} \\ \delta_{2(2^n)+i}X_{2^n} & \delta_{2(2^{n+1})+i}X_{2^{n+1}} & \cdots & \delta_{2(2^{n+1}-1)+i}X_{2^{n+1}-1} \end{pmatrix}$$

be the collection of  $(\delta_{2k+i}, \delta_{2k+i}X_k)^t$ ,  $k \in \mathbb{G}_n$ , and set

$$\Psi_n = \begin{pmatrix} \Phi_n^0 & 0 \\ 0 & \Phi_n^1 \end{pmatrix}.$$

Note that  $\Psi_n$  is a  $4 \times 2^{n+1}$  matrix. For all  $n \geq 1$ , we thus have, in the matrix form

$$\begin{aligned} \sum_{k \in \mathbb{G}_n} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 &= \sum_{k \in \mathbb{G}_n} \delta_{2k}(\widehat{\varepsilon}_{2k} - \varepsilon_{2k})^2 + \delta_{2k+1}(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1})^2, \\ &= (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^t \Psi_n \Psi_n^t (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}), \\ &= \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \Psi_n \Psi_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n, \\ &= \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1/2} \boldsymbol{\Delta}_n \boldsymbol{\Sigma}_{n-1}^{-1/2} \mathbf{M}_n, \end{aligned}$$

where

$$\boldsymbol{\Delta}_n = \boldsymbol{\Sigma}_{n-1}^{-1/2} \Psi_n \Psi_n^t \boldsymbol{\Sigma}_{n-1}^{-1/2} = \boldsymbol{\Sigma}_{n-1}^{-1/2} (\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_{n-1}) \boldsymbol{\Sigma}_{n-1}^{-1/2}.$$

Now, we can deduce from Corollary (4.3) that

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \boldsymbol{\Delta}_n = (\pi - 1) \mathbf{I}_4 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

which implies that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \sum_{k \in \mathbb{G}_n} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 = \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n (\pi - 1 + o(1)) \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \quad \text{a.s.}$$

Therefore, we can conclude via convergence (4.3) that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 \\ &= \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n(\pi - 1)} \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell = 4(\pi - 1)\sigma^2 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

which completes the proof. □

*Proof of result (3.6) of Theorem 3.3.* First of all, one has

$$\begin{aligned} \widehat{\sigma}_n^2 - \sigma_n^2 &= \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}} (\|\widehat{\mathbf{V}}_k\|^2 - \|\mathbf{V}_k\|^2), \\ &= \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}} (\|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 + 2(\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k). \end{aligned}$$

Set

$$P_n = \sum_{k \in \mathbb{T}_{n-1}} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k.$$

We clearly have

$$\Delta P_{n+1} = P_{n+1} - P_n = \sum_{k \in \mathbb{G}_n} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k.$$

One can observe that for all  $k \in \mathbb{G}_n$ ,  $\widehat{\mathbf{V}}_k - \mathbf{V}_k$  is  $\mathcal{F}_n^{\mathcal{O}}$ -measurable. Consequently,  $(P_n)$  is a real martingale transform for the filtration  $\mathbb{F}^{\mathcal{O}}$ . Hence, we can deduce from the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [6] together with (3.5) that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} P_n = o\left(\sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2\right) = o(n) \quad \text{a.s.}$$

It ensures once again via convergence (3.5) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\widehat{\sigma}_n^2 - \sigma_n^2) &= \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 \\ &= 4(\pi - 1)\sigma^2 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

which completes the proof of result (3.6). □

*Proof of results (3.7) and (3.8) of Theorem 3.3.* We now turn to the study of the covariance estimator  $\widehat{\rho}_n$ . We have

$$\begin{aligned} \widehat{\rho}_n - \rho_n &= \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1} - \varepsilon_{2k} \varepsilon_{2k+1}), \\ &= \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \delta_{2k+1} (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) + \frac{1}{|\mathbb{T}_{n-1}^{*01}|} Q_n, \end{aligned}$$

where

$$\begin{aligned} Q_n &= \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \varepsilon_{2k+1} + \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \varepsilon_{2k} \\ &= \sum_{k \in \mathbb{T}_{n-1}} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{J}_2 \mathbf{V}_k, \end{aligned}$$

with

$$\mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The process  $(Q_n)$  is a real martingale transform for the filtration  $\mathbb{F}^\mathcal{O}$  satisfying

$$Q_n = o\left(\sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2\right) = o(n) \quad \text{a.s.}$$

It now remains to prove that

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \\ &= \lim_{n \rightarrow \infty} \frac{R_n}{n} = \rho(\pi - 1) \text{tr}((\mathbf{L}^1)^{-1} (\mathbf{L}^{0,1})^2 (\mathbf{L}^0)^{-1}) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \quad (8.1) \end{aligned}$$

where

$$R_n = \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} (\mathbf{J}_2 \otimes \boldsymbol{\Phi}_\ell^{01} (\boldsymbol{\Phi}_\ell^{01})^t) \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell,$$

where  $\otimes$  denotes the Kronecker product of matrices, i.e.

$$\mathbf{J}_2 \otimes \boldsymbol{\Phi}_\ell^{01} (\boldsymbol{\Phi}_\ell^{01})^t = \begin{pmatrix} 0 & \boldsymbol{\Phi}_\ell^{01} (\boldsymbol{\Phi}_\ell^{01})^t \\ \boldsymbol{\Phi}_\ell^{01} (\boldsymbol{\Phi}_\ell^{01})^t & 0 \end{pmatrix},$$

and  $\boldsymbol{\Phi}_\ell^{01}$  is defined similarly as  $\boldsymbol{\Phi}_\ell^0$  and  $\boldsymbol{\Phi}_\ell^1$  by the collection of  $(\delta_{2k} \delta_{2k+1}, \delta_{2k} \delta_{2k+1} X_k)^t$ ,  $k \in \mathbb{G}_\ell$ . As  $\boldsymbol{\Phi}_n^{01} (\boldsymbol{\Phi}_n^{01})^t = \mathbf{S}_n^{01} - \mathbf{S}_{n-1}^{01}$ , proposition 4.2 implies that

$$\lim_{n \rightarrow \infty} \boldsymbol{\Sigma}_{n-1}^{-1/2} (\mathbf{J}_2 \otimes \boldsymbol{\Phi}_n^{01} (\boldsymbol{\Phi}_n^{01})^t) \boldsymbol{\Sigma}_{n-1}^{-1/2} = (\pi - 1) \boldsymbol{\Sigma}^{-1/2} \mathbf{J}_2 \otimes \mathbf{L}^{01} \boldsymbol{\Sigma}^{-1/2} \quad \text{a.s.}$$

so that the asymptotic behavior of  $R_n/n$  boils down to that of

$$\sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1/2} (\mathbf{J}_2 \otimes \mathbf{L}^{01}) \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \mathbf{M}_\ell.$$

A proof along the same lines as in Section 7 finally yields the expected results, i.e.

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{R_n}{n} = \rho \frac{\pi - 1}{\pi} \text{tr}((\mathbf{L}^1)^{-1} (\mathbf{L}^{0,1})^2 (\mathbf{L}^0)^{-1}) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$



which completes the proof of convergence (8.1). We then obtain

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{C}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\widehat{\rho}_n - \rho_n) = \rho \frac{\pi - 1}{\bar{p}(1, 1)} \text{tr}((\mathbf{L}^1)^{-1}(\mathbf{L}^{0,1})^2(\mathbf{L}^0)^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

which completes the proof of Theorem 3.3. □

### 8.3. Asymptotic normality

Contrary to the previous literature on BAR processes, we cannot use the central limit theorem given by Propositions 7.8 and 7.9 of [11] as in [8, 3] because the normalizing term is now the number of observations and is therefore random. The approach used in [5] strongly relies on the gaussian assumption for the noise sequence that does not hold here. Instead, we use the central limit theorem for martingales given in Theorem 3.II.10 of Duflo [6]. However, unlike the previous sections, this theorem can not be directly applied to the martingale  $(\mathbf{M}_n)$  because the number of observed data in a given generation grows exponentially fast and the Lindeberg condition does not hold. The solution is to use a new filtration. Namely, instead of using the observed generation-wise filtration, we will use the sister pair-wise one. Let

$$\mathcal{G}_p^{\mathcal{O}} = \mathcal{O} \vee \sigma\{\delta_1 X_1, (\delta_{2k} X_{2k}, \delta_{2k+1} X_{2k+1}), 1 \leq k \leq p\}$$

be the  $\sigma$ -algebra generated by the whole history  $\mathcal{O}$  of the Galton-Watson process and all observed individuals up to the offspring of individual  $p$ . Hence  $(\delta_{2k} \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1})$  is  $\mathcal{G}_k^{\mathcal{O}}$ -measurable. In addition, assumptions **(HN.1)** and **(HI)** imply that the processes  $(\delta_{2k} \varepsilon_{2k}, X_k \delta_{2k} \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1}, X_k \delta_{2k+1} \varepsilon_{2k+1})^t$ ,  $(\delta_{2k} \varepsilon_{2k}^2 + \delta_{2k+1} \varepsilon_{2k+1}^2 - (\delta_{2k} + \delta_{2k+1}) \sigma^2)$  and  $(\delta_{2k} \delta_{2k+1} (\varepsilon_{2k} \varepsilon_{2k+1} - \rho))$  are  $\mathcal{G}_k^{\mathcal{O}}$ -martingale difference sequences. In all the sequel, we will work under the probability  $\mathbb{P}_{\bar{\mathcal{E}}}$  and we denote by  $\mathbb{E}_{\bar{\mathcal{E}}}$  the corresponding expectation.

*Proof of Theorem 3.4, first step.* We apply Theorem 3.II.10 of [6] to the  $\mathcal{G}_k^{\mathcal{O}}$ -martingale  $\mathbf{M}^{(n)} = (\mathbf{M}_p^{(n)})_{\{p \geq 1\}}$  defined by

$$\mathbf{M}_p^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k=1}^p \mathbf{D}_k \quad \text{with} \quad \mathbf{D}_k = \begin{pmatrix} \delta_{2k} \varepsilon_{2k} \\ X_k \delta_{2k} \varepsilon_{2k} \\ \delta_{2k+1} \varepsilon_{2k+1} \\ X_k \delta_{2k+1} \varepsilon_{2k+1} \end{pmatrix}.$$

Set  $\nu_n = |\mathbb{T}_n| = 2^{n+1} - 1$ . Note that if  $k \notin \mathbb{T}_n^*$ , then  $\mathbf{D}_k = 0$  which implies that

$$\mathbf{M}_{\nu_n}^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k=1}^{|\mathbb{T}_n|} \mathbf{D}_k = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k \in \mathbb{T}_n^*} \mathbf{D}_k.$$

As the non-extinction set  $\bar{\mathcal{E}}$  is in  $\mathcal{G}_k^{\mathcal{O}}$  for every  $k \geq 1$ , it is easy to prove that

$$\begin{aligned} \mathbb{E}_{\bar{\mathcal{E}}}[D_k D_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] &= \mathbb{E}[D_k D_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] \\ &= \begin{pmatrix} \sigma^2 \delta_{2k} & \sigma^2 \delta_{2k} X_k & \rho \delta_{2k} \delta_{2k+1} & \rho \delta_{2k} \delta_{2k+1} X_k \\ \sigma^2 \delta_{2k} X_k & \sigma^2 \delta_{2k} X_k^2 & \rho \delta_{2k} \delta_{2k+1} X_k & \rho \delta_{2k} \delta_{2k+1} X_k^2 \\ \rho \delta_{2k} \delta_{2k+1} & \rho \delta_{2k} \delta_{2k+1} X_k & \sigma^2 \delta_{2k+1} & \sigma^2 \delta_{2k+1} X_k \\ \rho \delta_{2k} \delta_{2k+1} X_k & \rho \delta_{2k} \delta_{2k+1} X_k^2 & \sigma^2 \delta_{2k+1} X_k & \sigma^2 \delta_{2k+1} X_k^2 \end{pmatrix}, \end{aligned}$$

and Corollary 4.3 gives the  $\mathbb{P}_{\bar{\mathcal{E}}}$  almost sure limit of the increasing process

$$\langle M^{(n)} \rangle_{\nu_n} = \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[D_k D_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] = \frac{\mathbf{\Gamma}_n}{|\mathbb{T}_n^*|} \xrightarrow{n \rightarrow \infty} \mathbf{\Gamma}. \tag{8.2}$$

Therefore, the first assumption of Theorem 3.II.10 of [6] holds under  $\mathbb{P}_{\bar{\mathcal{E}}}$ . We now want to prove the Lindeberg condition that is the convergence in probability to 0 of the following expression  $L_n$  for all  $\epsilon > 0$ :

$$\begin{aligned} L_n &= \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\|D_k\|^2 \mathbb{1}_{\{\|D_k\| > \epsilon \sqrt{|\mathbb{T}_n^*|}\}} | \mathcal{G}_{k-1}^{\mathcal{O}}] \\ &\leq \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\|D_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] \mathbb{P}_{\bar{\mathcal{E}}}(\|D_k\| > \epsilon \sqrt{|\mathbb{T}_n^*|} | \mathcal{G}_{k-1}^{\mathcal{O}}) \\ &\leq \frac{\sup_{k \geq 0} \mathbb{E}[\|D_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}]}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \frac{\mathbb{E}_{\bar{\mathcal{E}}}[\|D_k\|^2 | \mathcal{G}_{k-1}^{\mathcal{O}}]}{\epsilon^2 |\mathbb{T}_n^*|}, \end{aligned}$$

for some  $r > 2$  and thanks to Hölder and Chebyshev inequalities. Besides, using Eq. (6.1) and similar calculations as in Lemma 6.1, one readily obtains

$$X_n^8 \leq 2^7 (1 - \beta)^{-7} \sum_{k=0}^{r_n-1} \beta^k |\eta_{[\frac{r_n}{2^k}]}^8| + 2^7 \beta^{8r_n} X_1^8.$$

Now, assumption **(HN.1)** together with  $\beta < 1$  yield the existence of a constant  $C$  such that

$$\sup_{k \geq 0} \mathbb{E}[X_k^8] \leq C(1 + \mathbb{E}[X_1^8]),$$

and recall that  $\mathbb{E}[X_1^8] < \infty$ . Finally, since the entries of  $D_k$  are combinations of  $\varepsilon_{2k+i}$  and  $X_k$ , using again **(HN.1)** and **(HI)**, one obtains that

$$\sup_{k \geq 0} \mathbb{E}[\|D_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty \quad \text{a.s.}$$

with  $r = 8$ . The Lindeberg condition is thus proved, plugging the convergence (8.2) into the following equality:

$$\frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\|D_k\|^2 | \mathcal{G}_{k-1}^{\mathcal{O}}] = \text{tr} \left( \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[D_k D_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] \right) \xrightarrow{n \rightarrow \infty} \text{tr}(\mathbf{\Gamma}).$$

We can now conclude that under  $\mathbb{P}_{\bar{\mathcal{E}}}$

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} \sum_{k \in \mathbb{T}_{n-1}^*} \mathbf{D}_k = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} \mathbf{M}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}).$$

Finally, result (3.9) follows from Eq. (4.1) and Corollary 4.3 together with Slutsky’s Lemma.  $\square$

*Proof of Theorem 3.4, second step.* On the one hand, we apply Theorem 3.II.10 of [6] to the  $\mathcal{G}_p^{\mathcal{O}}$ -martingale  $M^{(n)} = (M_p^{(n)})_{\{p \geq 1\}}$  defined by

$$M_p^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k=1}^p v_k \quad \text{and} \quad v_k = \delta_{2k} \varepsilon_{2k}^2 + \delta_{2k+1} \varepsilon_{k+1}^2 - (\delta_{2k} + \delta_{2k+1}) \sigma^2.$$

As above, one clearly has

$$M_{\nu_n}^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k \in \mathbb{T}_{n-1}^*} v_k = \sqrt{|\mathbb{T}_n^*|} (\sigma_n^2 - \sigma^2).$$

Using assumptions (HN.1), (HI) and Lemma 5.3 we compute the limit of the increasing process under  $\mathbb{P}_{\bar{\mathcal{E}}}$

$$\lim_{n \rightarrow \infty} \langle M^{(n)} \rangle_{\nu_n} = (\tau^4 - \sigma^4) + \frac{2\bar{p}(1, 1)}{\pi} (\nu^2 \tau^4 - \sigma^4) \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

Therefore, the first assumption of Theorem 3.II.10 of [6] holds under  $\mathbb{P}_{\bar{\mathcal{E}}}$ . Thanks to assumptions (HN.2) and (HI) we can prove that for some  $r > 2$ ,

$$\sup_{k \geq 0} \mathbb{E}_{\bar{\mathcal{E}}} [\|v_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty \quad \text{a.s.}$$

which implies the Lindeberg condition. Therefore, we obtain that under  $\mathbb{P}_{\bar{\mathcal{E}}}$

$$\sqrt{|\mathbb{T}_n^*|} (\sigma_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\tau^4 - \sigma^4) + \frac{2\bar{p}(1, 1)}{\pi} (\nu^2 \tau^4 - \sigma^4)).$$

Furthermore, we infer from Eq. (3.6) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_n^*|} (\hat{\sigma}_n^2 - \sigma_n^2) = 0 \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

which yields result (3.10).

We turn now to the proof of result (3.11) with another  $\mathcal{G}_p^{\mathcal{O}}$ -martingale  $(M^{(n)})$  defined by

$$M_p^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^{*01}|}} \sum_{k=1}^p \delta_{2k} \delta_{2k+1} (\varepsilon_{2k} \varepsilon_{2k+1} - \rho).$$

As above, one easily shows that

$$M_{\nu_n}^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^{*01}|}} \sum_{i \in \mathbb{T}_{n-1}^*} \delta_{2i} \delta_{2i+1} (\varepsilon_{2i} \varepsilon_{2i+1} - \rho) = \sqrt{|\mathbb{T}_{n-1}^{*01}|} (\rho_n - \rho).$$

Using assumptions **(HN.1)** and **(H.I)**, we compute the limit of the increasing process

$$\lim_{n \rightarrow \infty} \langle M^{(n)} \rangle_{\nu_n} = \nu^2 \tau^4 - \rho^2 \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

We also derive the Lindeberg condition. Consequently, we obtain that under  $\mathbb{P}_{\bar{\mathcal{E}}}$ , one has

$$\sqrt{|\mathbb{T}_{n-1}^{*01}|}(\rho_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 \tau^4 - \rho^2).$$

Furthermore, we infer from **(3.8)** that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_{n-1}^{*01}|}(\hat{\rho}_n - \rho_n) = 0 \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

Finally, result **(3.11)** follows, which completes the proof of Theorem **3.4**.  $\square$

### Appendix A: Quadratic strong law

In order to establish the quadratic strong law for the main martingale  $(M_n)$ , we are going to study separately the asymptotic behavior of  $(W_n)$  and  $(B_n)$  which appear in the main decomposition given by Equation **(7.1)**.

**Lemma A.1.** *Under assumptions **(HN.1)**, **(HN.2)**, **(HO)** and **(HI)**, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} W_n = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

*Proof.* First of all, we have the decomposition  $W_{n+1} = \mathcal{T}_{n+1} + \mathcal{R}_{n+1}$  where

$$\begin{aligned} \mathcal{T}_{n+1} &= \sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t \Sigma^{-1} \Delta M_{\ell+1}}{|\mathbb{T}_{\ell}^*|}, \\ \mathcal{R}_{n+1} &= \sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t (|\mathbb{T}_{\ell}^*| \Sigma_{\ell}^{-1} - \Sigma^{-1}) \Delta M_{\ell+1}}{|\mathbb{T}_{\ell}^*|}. \end{aligned}$$

We first prove that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathcal{T}_n = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \tag{A.1}$$

As  $\mathcal{T}_n$  is a scalar and the trace is commutative, one can rewrite  $\mathcal{T}_{n+1}$  as  $\mathcal{T}_{n+1} = \text{tr}(\Sigma^{-1/2} \mathcal{H}_{n+1} \Sigma^{-1/2})$  where

$$\mathcal{H}_{n+1} = \sum_{\ell=1}^n \frac{\Delta M_{\ell+1} \Delta M_{\ell+1}^t}{|\mathbb{T}_{\ell}^*|}.$$

Our goal is to make use of the strong law of large numbers for martingale transforms, so we start by adding and subtracting a term involving the conditional expectation of  $\Delta \mathcal{H}_{n+1}$  given  $\mathcal{F}_n^O$ . We have already seen in Section **4.1** that for

all  $n$ ,  $\mathbb{E}[\Delta M_{n+1} \Delta M_{n+1}^t | \mathcal{F}_n^{\mathcal{O}}] = \Gamma_n - \Gamma_{n-1}$ . Consequently, we can split  $\mathcal{H}_{n+1}$  into two terms

$$\mathcal{H}_{n+1} = \sum_{\ell=1}^n \frac{\Gamma_{\ell} - \Gamma_{\ell-1}}{|\mathbb{T}_{\ell}^*|} + \mathcal{K}_{n+1},$$

where

$$\mathcal{K}_{n+1} = \sum_{\ell=1}^n \frac{\Delta M_{\ell+1} \Delta M_{\ell+1}^t - (\Gamma_{\ell} - \Gamma_{\ell-1})}{|\mathbb{T}_{\ell}^*|}.$$

On the one hand, it follows from Corollary 4.3 and Lemma 2.1 that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Gamma_n - \Gamma_{n-1}}{|\mathbb{T}_n^*|} = \frac{\pi - 1}{\pi} \Gamma \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

Thus, Cesaro convergence and the remark that  $\{|\mathbb{G}_{\ell}^*| = 0\} \subset \{|\mathbb{G}_n^*| = 0\}$  for all  $\ell \leq n$  yield

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \frac{\Gamma_{\ell} - \Gamma_{\ell-1}}{|\mathbb{T}_{\ell}^*|} &= \lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbb{1}_{\{|\mathbb{G}_{\ell}^*| > 0\}} \frac{\Gamma_{\ell} - \Gamma_{\ell-1}}{|\mathbb{T}_{\ell}^*|} \\ &= \frac{\pi - 1}{\pi} \Gamma \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

On the other hand, the sequence  $(\mathcal{K}_n)$  is obviously a matrix martingale transform and tedious but straightforward calculations, together with Lemmas 6.1 and 6.2 and the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [6] imply that  $\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \mathcal{K}_n = o(n)$  a.s. Hence, we infer from the equation above that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathcal{H}_n = \frac{\pi - 1}{\pi} \Gamma \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \quad (\text{A.2})$$

Finally, we obtain

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathcal{T}_n = \frac{\pi - 1}{\pi} \text{tr}(\Sigma^{-1/2} \Gamma \Sigma^{-1/2}) \mathbb{1}_{\bar{\mathcal{E}}} = \frac{\pi - 1}{\pi} 4\sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

which proves (A.1). We now turn to the asymptotic behavior of  $\mathcal{R}_{n+1}$ . We know from Proposition 4.2 that  $\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} (|\mathbb{T}_n^*| \Sigma_n^{-1} - \Sigma^{-1})$  goes to 0 as  $n$  goes to infinity. Hence, for all positive  $\epsilon$  and for large enough  $n$ , one has

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\Delta M_{n+1}^t (|\mathbb{T}_n^*| \Sigma_n^{-1} - \Sigma^{-1}) \Delta M_{n+1}| \leq \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} 4\epsilon \Delta M_{n+1}^t \Delta M_{n+1}.$$

Using again that  $\{|\mathbb{G}_{\ell}^*| = 0\} \subset \{|\mathbb{G}_{n+1}^*| = 0\}$  for all  $\ell \leq n + 1$ , we rewrite

$$\mathbb{1}_{\{|\mathbb{G}_{n+1}^*| > 0\}} \mathcal{R}_{n+1} = \mathbb{1}_{\{|\mathbb{G}_{n+1}^*| > 0\}} \sum_{\ell=1}^n \mathbb{1}_{\{|\mathbb{G}_{\ell}^*| > 0\}} \frac{\Delta M_{\ell+1}^t (|\mathbb{T}_{\ell}^*| \Sigma_{\ell}^{-1} - \Sigma^{-1}) \Delta M_{\ell+1}}{|\mathbb{T}_{\ell}^*|}.$$

Hence,

$$\begin{aligned} \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}}|\mathcal{R}_{n+1}| &\leq 4\epsilon \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}} \sum_{\ell=1}^n \mathbb{1}_{\{|\mathbb{G}_\ell^*|>0\}} \frac{\Delta M_{\ell+1}^t \Delta M_{\ell+1}}{|\mathbb{T}_\ell^*|} \\ &\leq 4\epsilon \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}} \sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t \Delta M_{\ell+1}}{|\mathbb{T}_\ell^*|} \\ &\leq 4\epsilon \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}} \text{tr}(\mathcal{H}_{n+1}). \end{aligned}$$

This last inequality holding for any positive  $\epsilon$  and large enough  $n$ , the limit given by Equation (A.2) entails that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \frac{1}{n} \mathcal{R}_n = 0 \quad \text{a.s.}$$

which completes the proof of Lemma A.1. □

**Lemma A.2.** *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \frac{1}{n} \mathcal{B}_n = 0 \quad \text{a.s.}$$

*Proof.* The result is obvious on the extinction set  $\mathcal{E}$ . Now let us work on  $\bar{\mathcal{E}}$ . Now for  $i \in \{0, 1\}$  and  $n \geq 1$ , let  $\boldsymbol{\xi}_n^i = (\varepsilon_{2^n+i}, \varepsilon_{2^n+2+i}, \dots, \varepsilon_{2^{n+1}-2+i})^t$ , be the collection of  $\varepsilon_k$ ,  $k \in \mathbb{G}_n^i$ , and set  $\boldsymbol{\xi}_n = (\boldsymbol{\xi}_n^0, \boldsymbol{\xi}_n^1)^t$ . Note that  $\boldsymbol{\xi}_n$  is a column vector of size  $2^{n+1}$ . With these notation, one has

$$\mathcal{B}_{n+1} = 2 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Delta M_{\ell+1} = 2 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Psi_\ell \boldsymbol{\xi}_{\ell+1}.$$

The sequence  $(\mathcal{B}_n)$  is a real martingale transform satisfying

$$\Delta \mathcal{B}_{n+1} = \mathcal{B}_{n+1} - \mathcal{B}_n = 2 M_n^t \Sigma_n^{-1} \Psi_n \boldsymbol{\xi}_{n+1}.$$

Consequently, via the strong law of large numbers for martingale transforms, we find that either  $(\mathcal{B}_n)$  converges a.s. or  $\mathcal{B}_n = o(\langle \mathcal{B} \rangle_n)$  a.s. where

$$\langle \mathcal{B} \rangle_{n+1} = 4 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Psi_\ell \mathbf{C} \Psi_\ell^t \Sigma_\ell^{-1} M_\ell,$$

with

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix} \otimes \mathbf{I}_{2^n}.$$

As  $\mathbf{C}$  is definite positive under assumption (HN.1), one has  $\mathbf{C} \leq 2\sigma^2 \mathbf{I}_{2^{n+1}}$  in the sense that  $2\sigma^2 \mathbf{I}_{2^{n+1}} - \mathbf{C}$  is semi definite positive. Hence, one has

$$\langle \mathcal{B} \rangle_{n+1} \leq 8\sigma^2 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Psi_\ell \Psi_\ell^t \Sigma_\ell^{-1} M_\ell.$$

Now, by definition, one has

$$\Sigma_\ell^{-1} \Psi_\ell \Psi_\ell^t \Sigma_\ell^{-1} = \begin{pmatrix} (\mathbf{S}_\ell^0)^{-1} \Phi_\ell^0 (\Phi_\ell^0)^t (\mathbf{S}_\ell^0)^{-1} & 0 \\ 0 & (\mathbf{S}_\ell^1)^{-1} \Phi_\ell^1 (\Phi_\ell^1)^t (\mathbf{S}_\ell^1)^{-1} \end{pmatrix}.$$

We now use Lemma B.1 of [3] on each entry to obtain

$$\Sigma_\ell^{-1} \Psi_\ell \Psi_\ell^t \Sigma_\ell^{-1} \leq \Sigma_{\ell-1}^{-1} - \Sigma_\ell^{-1},$$

as the matrix  $l_k$  in that lemma is definite positive. Therefore, we obtain that

$$\langle \mathcal{B} \rangle_{n+1} \leq 8\sigma^2 \sum_{\ell=1}^n M_\ell^t (\Sigma_{\ell-1}^{-1} - \Sigma_\ell^{-1}) M_\ell = 8\sigma^2 \mathcal{A}_n.$$

Finally, we deduce from the main decomposition given by Equation (7.1) and Lemma A.1 that

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} (\mathcal{V}_{n+1} + \mathcal{A}_n) = o(\mathcal{A}_n) + \mathcal{O}(n) \quad \text{a.s.}$$

leading to  $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathcal{V}_{n+1} = \mathcal{O}(n)$  and  $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathcal{A}_n = \mathcal{O}(n)$  a.s. as  $\mathcal{V}_{n+1}$  and  $\mathcal{A}_n$  are non-negative. This implies in turn that  $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathcal{B}_n = o(n)$  a.s. completing the proof of Lemma A.2.  $\square$

### Appendix B: Wei’s Lemma

In order to prove Proposition 7.2, we shall apply Wei’s Lemma given in [16] page 1672, to each entry of the vector-valued main martingale

$$M_n = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} (\delta_{2k} \varepsilon_{2k}, \delta_{2k} X_k \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1}, \delta_{2k+1} X_k \varepsilon_{2k+1})^t.$$

For  $i \in \{0, 1\}$ , denote

$$P_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k+i} \varepsilon_{2k+i} \quad \text{and} \quad Q_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k+i} X_k \varepsilon_{2k+i}.$$

On the set  $\bar{\mathcal{E}}$ , these processes can be rewritten as

$$P_n^i = \sum_{\ell=1}^n \sqrt{|\mathbb{G}_{\ell-1}^*|} v_\ell^i, \quad Q_n^i = \sum_{\ell=1}^n \sqrt{|\mathbb{G}_{\ell-1}|} w_\ell^i,$$

where

$$\begin{aligned} v_n^i &= \mathbb{1}_{\{|\mathbb{G}_{n-1}^*|>0\}} \frac{1}{\sqrt{|\mathbb{G}_{n-1}^*|}} \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i} \varepsilon_{2k+i}, \\ w_n^i &= \mathbb{1}_{\{|\mathbb{G}_{n-1}^*|>0\}} \frac{1}{\sqrt{|\mathbb{G}_{n-1}^*|}} \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i} X_k \varepsilon_{2k+i}. \end{aligned}$$

On the one hand, we clearly have  $\mathbb{E}[v_{n+1}^i | \mathcal{F}_n^{\mathcal{O}}] = 0$  and  $\mathbb{E}[(v_{n+1}^i)^2 | \mathcal{F}_n^{\mathcal{O}}] = \sigma^2 \frac{Z_{n+1}^i}{|\mathbb{G}_n^*|}$  a.s. on  $\bar{\mathcal{E}}$ . Moreover, it follows from Cauchy-Schwarz inequality that

$$\begin{aligned} \mathbb{E}[(v_{n+1}^i)^4 | \mathcal{F}_n^{\mathcal{O}}] &= \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{G}_n^*|^2} \sum_{k \in \mathbb{G}_n} \delta_{2l+i} \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] \\ &\quad + \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{G}_n^*|^2} \sum_{p \in \mathbb{G}_n} \sum_{k \neq p} \delta_{2p+i} \delta_{2k+i} \mathbb{E}[\varepsilon_{2p+i}^2 | \mathcal{F}_n^{\mathcal{O}}] \mathbb{E}[\varepsilon_{2k+i}^2 | \mathcal{F}_n^{\mathcal{O}}] \\ &\leq 3C \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \sup_{k \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] \quad \text{a.s.} \end{aligned}$$

as  $Z_{n+1}^i |\mathbb{G}_n^*|^{-1}$  is bounded. This implies that  $\sup \mathbb{E}[(v_{n+1}^i)^4 | \mathcal{F}_n^{\mathcal{O}}] < +\infty$  a.s. Consequently, we deduce from Wei's Lemma that for all  $\eta > 1/2$ ,

$$\mathbb{1}_{\{|\mathbb{G}_{n-1}^*| > 0\}} (P_n^i)^2 = o(|\mathbb{T}_{n-1}^*| n^\eta) \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

On the other hand, it is not hard to see that  $\mathbb{E}[w_{n+1}^i | \mathcal{F}_n^{\mathcal{O}}] = 0$  a.s. Moreover, it follows from Cauchy-Schwarz inequality that,

$$\begin{aligned} \mathbb{E}[(w_{n+1}^i)^4 | \mathcal{F}_n^{\mathcal{O}}] &\leq \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{G}_n^*|^2} \left( \sum_{k \in \mathbb{G}_n} \delta_{2k+i} X_k^4 \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] + \sigma^4 \sum_{p \in \mathbb{G}_n} \sum_{k \neq p} \delta_{2p+i} \delta_{2k+i} X_p^2 X_k^2 \right) \\ &\leq 3 \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \left( \sup_{k \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] \right) \left( \frac{1}{|\mathbb{G}_n^*|} \sum_{k \in \mathbb{G}_n} \delta_{2k+i} X_k^2 \right)^2 \quad \text{a.s.} \end{aligned}$$

which is finite from Proposition 6.5. We deduce from Wei's Lemma applied to  $Q_n^i$  that for all  $\eta > 1/2$ ,  $\mathbb{1}_{\{|\mathbb{G}_{n-1}^*| > 0\}} \|Q_n^i\|^2 = o(|\mathbb{T}_{n-1}^*| n^\eta)$  a.s. which completes the proof of Proposition 7.2.  $\square$

## References

- [1] KRISHNA B. ATHREYA AND PETER E. NEY. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196. [MR0373040](#)
- [2] VINCENT BANSAYE. Proliferating parasites in dividing cells: Kimmel's branching model revisited. *Ann. Appl. Probab.*, 18(3):967–996, 2008. [MR2418235](#)
- [3] BERNARD BERCU, BENOÎTE DE SAPORTA, AND ANNE GÉGOUT-PETIT. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.*, 14:no. 87, 2492–2526, 2009. [MR2563249](#)
- [4] R. COWAN AND R. G. STAUDTE. The bifurcating autoregressive model in cell lineage studies. *Biometrics*, 42:769–783, 1986.
- [5] JEAN-FRANÇOIS. DELMAS AND LAURENCE MARSALLE. Detection of cellular aging in a Galton-Watson process. *Stoch. Process. and Appl.*, 120:2495–2519, 2010. [MR2728175](#)



- [6] MARIE DUFLO. *Random iterative models*, volume 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997. [MR1485774](#)
- [7] STEVEN N. EVANS AND DAVID STEINSALTZ. Damage segregation at fissioning may increase growth rates: A superprocess model. *Theoretical Population Biology*, 71(4):473–490, 2007.
- [8] JULIEN GUYON. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.*, 17(5-6):1538–1569, 2007. [MR2358633](#)
- [9] JULIEN GUYON, ARIANE BIZE, GRÉGORIE PAUL, ERIC STEWART, JEAN-FRANÇOIS DELMAS, AND FRANÇOIS TADDÉI. Statistical study of cellular aging. In *CEMRACS 2004—mathematics and applications to biology and medicine*, volume 14 of *ESAIM Proc.*, pages 100–114 (electronic). EDP Sci., Les Ulis, 2005. [MR2226805](#)
- [10] PATSY HACCOU, PETER JAGERS, AND VLADIMIR A. VATUTIN. *Branching processes: variation, growth, and extinction of populations*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, Cambridge, 2007. [MR2429372](#)
- [11] JAMES D. HAMILTON. *Time series analysis*. Princeton University Press, Princeton, NJ, 1994. [MR1278033](#)
- [12] THEODORE E. HARRIS. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963. [MR0163361](#)
- [13] ROGER A. HORN AND CHARLES R. JOHNSON. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original. [MR1084815](#)
- [14] F. MAAOUIA AND A. TOUATI. Identification of multitype branching processes. *Ann. Statist.*, 33(6):2655–2694, 2005. [MR2253098](#)
- [15] E.J. STEWART, R. MADDEN, G. PAUL, AND F. TADDEI. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, 3(2):e45, 2005.
- [16] C. Z. WEI. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.*, 15(4):1667–1682, 1987. [MR0913581](#)