

# Statistical study of asymmetry in cell lineage data

Benoîte de Saporta

Univ. Bordeaux, Gretha, UMR 5113, IMB, UMR 5251, F-33400 Talence, France  
CNRS, Gretha, UMR 5113, IMB, UMR 5251, F-33400 Talence, France  
INRIA Bordeaux Sud Ouest, team CQFD, F-33400 Talence, France

Anne Gégout-Petit

Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France  
CNRS, IMB, UMR 5251, F-33400 Talence, France  
INRIA Bordeaux Sud Ouest, team CQFD, F-33400 Talence, France

Laurence Marsalle

Univ. Lille 1, Laboratoire Paul Painlevé, UMR 8524, F-59 655 Villeneuve d'Ascq, France  
CNRS, Laboratoire Paul Painlevé, UMR 8524, F-59 655 Villeneuve d'Ascq, France

July 3, 2018

## Abstract

A rigorous methodology is proposed to study cell division data consisting in several observed genealogical trees of possibly different shapes. The procedure takes into account missing observations, data from different trees, as well as the dependence structure within genealogical trees. Its main new feature is the joint use of all available information from several data sets instead of single data set estimation, to avoid the drawbacks of low accuracy for estimators or low power for tests on small single-trees. The data is modeled by an asymmetric bifurcating autoregressive process and possibly missing observations are taken into account by modeling the genealogies with a two-type Galton-Watson process. Least-squares estimators of the unknown parameters of the processes are given and symmetry tests are derived. Results are applied on real data of *Escherichia coli* division and an empirical study of the convergence rates of the estimators and power of the tests is conducted on simulated data.

## 1 Introduction

Cell lineage data consist of observations of some quantitative characteristic of the cells (e.g. their length, growth rate, time until division, ...) over several generations descended from an initial cell. Track is kept of the genealogy to study the inherited effects on the evolution of the characteristic. As a cell usually gives birth to two offspring by division, such genealogies are structured as binary trees. Cowan and Staudte (1986) first adapted autoregressive processes to this binary tree structure by introducing bifurcating autoregressive processes (BAR). This parametric model takes into account both the environmental and inherited effects. Inference on this model has been proposed based on either a single-tree growing to infinity, see e.g. Cowan and Staudte (1986), Huggins (1996), Huggins and Basawa (2000), Zhou and Basawa (2005) or for an asymptotically infinite number of small replicated trees, see e.g. Huggins and Staudte (1994), Huggins and Basawa (1999).

More recently, studies of aging in single cell organisms by Stewart et al (2005) suggested that cell division may not be symmetric. An asymmetric BAR model was therefore proposed by Guyon (2007), where the two sets of parameters corresponding to sister cells are allowed to be different. Inference for this model was only investigated for single-trees growing to infinity, see Guyon (2007), Bercu et al (2009) for the fully observed model or Delmas and Marsalle (2010), de Saporta et al (2011), de Saporta et al (2012) for missing data models.

Cell division data often consist in recordings over several genealogies of cells evolving in similar experimental conditions. For instance, Stewart et al (2005) filmed 94 colonies of *Escherichia coli*

cells dividing between four and nine times. We therefore propose a new rigorous approach to take into account all the available information. Indeed, we propose an inference based on a finite fixed number of replicated trees when the total number of observed cells tends to infinity. We use the missing data asymmetric BAR model introduced by de Saporta et al (2011). In this approach, the observed genealogies are modeled with a two-type Galton-Watson (GW) process. However, we propose a different least-squares estimator for the parameters of the BAR process that does not correspond to the single-tree estimators averaged on the replicated trees. We also propose an estimator of the parameters of the GW process specific to our binary tree structure and not based simply on the observation of the number of cells of each type in each generation as in Guttorp (1991), Maaouia and Touati (2005). We study the consistency and asymptotic normality of our estimators and derive asymptotic confidence intervals as well as Wald's type tests to investigate the asymmetry of the data for both the BAR and GW processes. Our results are applied to the *Escherichia coli* data of Stewart et al (2005). We also provide an empirical study of the convergence rate of our estimators and of the power of the symmetry tests on simulated data.

The paper is organized as follows. In Section 2, we describe a methodology for least-squares estimation based on multiple data sets in a general framework. In Section 3, we present the BAR and observations models. In Section 4 we give our estimators and state their asymptotic properties. In Section 5, we propose a new investigation of Stewart et al (2005) data. In Section 6 we give simulation results. The precise statement of the convergence results, the explicit form of the asymptotic variance of the estimators and the convergence proofs are postponed to the appendix.

## 2 Methodology

We work with the following general framework. Consider that several data sets are available, obtained in similar experimental conditions and then assumed to come from the same parametric model. Suppose that there exists a consistent least-squares estimator for the parametric model. This estimator can be computed on each individual data set, but we would like to take into account all the data at disposal, which should improve the accuracy of the estimation.

To this aim, we assume that the different data sets are independent realizations of the parametric model. A natural idea is to average the single-set estimators. It may be a good approach if the single-set estimators have roughly the same variance, which is usually the case when the data sets have the same size. However, if the data sets have very different sizes, the single-set estimators may have variances of different orders and this direct approach becomes dubious.

Instead, we propose to use a global least-squares estimator. Suppose that we have  $m$  data sets. Let  $\theta$  be the (possibly multivariate) parameter to be estimated, and  $\hat{\theta}_{j,n}$  the least-squares estimator build with the  $j$ -th data set for  $1 \leq j \leq m$ . The global least-squares estimator  $\hat{\theta}_n$  decomposes as

$$\hat{\theta}_n = \left( \sum_{j=1}^m \Sigma_{j,n} \right)^{-1} \sum_{j=1}^m \mathbf{V}_{j,n},$$

where  $\Sigma_{j,n}$  is a normalizing matrix and  $\mathbf{V}_{j,n}$  a vector of the same size as  $\theta$ , involved in the decomposition of the single-set least-squares estimator  $\hat{\theta}_{j,n}$  as follows

$$\hat{\theta}_{j,n} = \Sigma_{j,n}^{-1} \mathbf{V}_{j,n}.$$

Note that the estimator  $\hat{\theta}_n$  thus constructed is neither an average nor a function of the  $\hat{\theta}_{j,n}$ . Hence, the asymptotic behavior of the global estimator  $\hat{\theta}_n$  cannot be deduced from that of the single-set estimators  $\hat{\theta}_{j,n}$ . Nevertheless, the asymptotic behavior of  $\hat{\theta}_{j,n}$  is often obtained through the convergence of the normalizing matrices  $\Sigma_{j,n}$  and of the vectors  $\mathbf{V}_{j,n}$  separately, which gives the convergence of the global estimator  $\hat{\theta}_n$  as the number  $m$  of data sets is fixed. Note that the asymptotic is not the number  $m$  of data sets.

The aim of this paper is to apply this methodology to cell division data with missing data. In this special case, the convergence of the global estimator  $\hat{\theta}_n$  is not straightforward, because we have to prove it on a set where the convergence of each  $\Sigma_{j,n}$  and  $\mathbf{V}_{j,n}$  is not ensured.

### 3 Model

Our aim is to estimate the parameters of coupled BAR and GW processes through  $m$  i.i.d. realizations of the processes. We first define our parametric model and introduce our notations. The BAR and GW processes have the same dynamics as in de Saporta et al (2011), the main difference is that our inference is here based on several i.i.d. realizations of the processes, instead of a single one. Additional notations together with the precise technical assumptions are specified in A.

#### 3.1 Bifurcating autoregressive model

Consider  $m$  i.i.d. replications of the asymmetric BAR process with coefficient  $\theta = (a_0, b_0, a_1, b_1) \in \mathbb{R}^4$ . More precisely, for  $1 \leq j \leq m$ , the first cell in genealogy  $j$  is labelled  $(j, 1)$  and for  $k \geq 1$ , the two offspring of cell  $(j, k)$  are labelled  $(j, 2k)$  and  $(j, 2k + 1)$ . As we consider an asymmetric model, each cell has a *type* defined by its label:  $(j, 2k)$  has type *even* and  $(j, 2k + 1)$  has type *odd*. The characteristic of cell  $k$  in genealogy  $j$  is denoted by  $X_{(j,k)}$ . The BAR processes are defined recursively as follows: for all  $1 \leq j \leq m$  and  $k \geq 1$ , one has

$$\begin{cases} X_{(j,2k)} &= a_0 + b_0 X_{(j,k)} + \varepsilon_{(j,2k)}, \\ X_{(j,2k+1)} &= a_1 + b_1 X_{(j,k)} + \varepsilon_{(j,2k+1)}. \end{cases} \quad (1)$$

Let us also define the variance and covariance of the noise sequence

$$\sigma_0^2 = \mathbb{E}[\varepsilon_{(j,2k)}^2], \quad \sigma_1^2 = \mathbb{E}[\varepsilon_{(j,2k+1)}^2], \quad \rho = \mathbb{E}[\varepsilon_{(j,2k)}\varepsilon_{(j,2k+1)}].$$

Our goal is to estimate the parameters  $\theta = (a_0, b_0, a_1, b_1)$  and  $(\sigma_0^2, \sigma_1^2, \rho)$ , and then test if  $(a_0, b_0) = (a_1, b_1)$  or not.

#### 3.2 Observation process

We now turn to the observation process  $(\delta_{(j,k)})$  that encodes for the presence or absence of cell measurements in the available data

$$\delta_{(j,k)} = \begin{cases} 1 & \text{if cell } k \text{ in genealogy } j \text{ is observed,} \\ 0 & \text{if cell } k \text{ in genealogy } j \text{ is not observed.} \end{cases}$$

To take into account possible asymmetry in the observation process, we use a two-type Galton-Watson model. The relevance of this model to E. coli data is discussed in section 5. Again, we suppose all the  $m$  observation processes to be drawn independently from the same two-type GW process. More precisely, for all  $1 \leq j \leq m$ , we model the observation process  $(\delta_{(j,k)})_{k \geq 1}$  for the  $j$ -th genealogy as follows. We set  $\delta_{(j,1)} = 1$  and draw  $(\delta_{(j,2k)}, \delta_{(j,2k+1)})$  independently from one another with a law depending on the type of cell  $k$ . More precisely, for  $i \in \{0, 1\}$ , if  $k$  is of type  $i$  we set

$$\begin{aligned} \mathbb{P}\left((\delta_{(j,2k)}, \delta_{(j,2k+1)}) = (l_0, l_1) \mid \delta_{(j,k)} = 1\right) &= p^{(i)}(l_0, l_1), \\ \mathbb{P}\left((\delta_{(j,2k)}, \delta_{(j,2k+1)}) = (0, 0) \mid \delta_{(j,k)} = 0\right) &= 1, \end{aligned}$$

for all  $(l_0, l_1) \in \{0, 1\}^2$ . Thus,  $p^{(i)}(l_0, l_1)$  is the probability that a cell of type  $i$  has  $l_0$  daughter of type 0 and  $l_1$  daughter of type 1. And if a cell is missing, its descendants are missing as well. Figure 1 gives an example of realization of an observation process. We also assume that the observation processes are independent from the BAR processes.

### 4 Inference

Our first goal is to estimate the reproduction probabilities  $p^{(i)}(l_0, l_1)$  of the GW process from the  $m$  genealogies of observed cells up to the  $n$ -th generation to be able to test the symmetry of the GW model itself. Our second goal is to estimate  $\theta = (a_0, b_0, a_1, b_1)^t$  from all the observed individuals

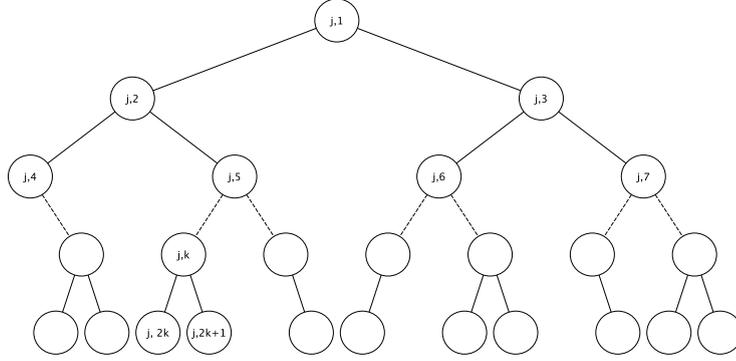


Figure 1: A tree of observed cells.

of the  $m$  trees up to the  $n$ -th generation. We then give the asymptotic properties of our estimator to be able to build confidence intervals and symmetry tests for  $\theta$ .

Denote by  $|\mathbb{T}_n^*|$  the total number of observed cells in the  $m$  trees up to the  $n$ -th generation of offspring from the original ancestors, and let

$$\bar{\mathcal{E}} = \left\{ \lim_{n \rightarrow \infty} |\mathbb{T}_n^*| = \infty \right\}$$

be the non-extinction set, on which the global cell population grows to infinity. Thus, our asymptotic results only hold on the set  $\bar{\mathcal{E}}$ . This global non-extinction set is the union and not the intersection of the non-extinction sets of each single-tree. It means that some trees may extinct, which allows us to take into account trees with a different number of observed generations. We are thus in a case where averaging single-tree estimators is not recommended. The possibility of extinction for some trees is also the reason why the convergence of the multiple-trees estimator  $\hat{\theta}_n$  is not straightforward from existing results in the literature. Conditions for the probability of non-extinction to be positive are given in A.

#### 4.1 Estimation of the reproduction law of the GW process

There are many references on the inference of a multi-type GW process, see for instance (Guttorp, 1991) and (Maaouia and Touati, 2005). Our context of estimation is very specific because the information given by  $(\delta_{(j,k)})$  is more precise than that given by the number of cells of each type in a given generation that is usually used in the literature. Indeed, not only do we know the number of cells of each type in each generation, but we also know their precise positions on the binary tree of cell division. The empiric estimators of the reproduction probabilities using data up to the  $n$ -th generation are then, for  $i, l_0, l_1$  in  $\{0, 1\}$

$$\hat{p}_n^{(i)}(l_0, l_1) = \frac{\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k+i)} \phi_{l_0}(\delta_{(j,2(2k+i))}) \phi_{l_1}(\delta_{(j,2(2k+i)+1})}{\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k+i)}}$$

where  $\phi_0(x) = 1 - x$ ,  $\phi_1(x) = x$ , and if the denominator is non zero, the estimator equalling zero otherwise. Note that the numerator is just the number of cells of type  $i$  in all the trees up to generation  $n - 1$  that have exactly  $l_0$  daughter of type 0 and  $l_1$  daughter of type 1 in the  $n$ -th generation. The denominator is the total number of cells of type  $i$  in all the trees up to generation  $n - 1$ . Set also

$$\mathbf{p}^{(i)} = (p^{(i)}(1, 1), p^{(i)}(1, 0), p^{(i)}(0, 1), p^{(i)}(0, 0))^t,$$

the vector of the 4 reproduction probabilities for a mother of type  $i$ ,  $\mathbf{p} = ((\mathbf{p}^{(0)})^t, (\mathbf{p}^{(1)})^t)^t$  the vector of all 8 reproduction probabilities and  $\hat{\mathbf{p}}_n$  its empirical estimator.

## 4.2 Least-squares estimation for the BAR parameters

For the parameters of the BAR process, we use the standard least-squares (LS) estimator  $\hat{\theta}_n$  with all the available data from the  $m$  trees up to generation  $n$ . It minimizes

$$\Delta_n(\theta) = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j,2k)}(X_{(j,2k)} - a_0 - b_0 X_{(j,k)})^2 + \delta_{(j,2k+1)}(X_{(j,2k+1)} - a_1 - b_1 X_{(j,k)})^2.$$

Consequently, for all  $n \geq 1$  we have  $\hat{\theta}_n = (\hat{a}_{0,n}, \hat{b}_{0,n}, \hat{a}_{1,n}, \hat{b}_{1,n})^t$  with

$$\hat{\theta}_n = \Sigma_{n-1}^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} (\delta_{(j,2k)} X_{(j,2k)}, \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)}, \delta_{(j,2k+1)} X_{(j,2k+1)}, \delta_{(j,2k+1)} X_{(j,k)} X_{(j,2k+1)})^t \quad (2)$$

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

$$\Sigma_n = \begin{pmatrix} \mathbf{S}_n^0 & 0 \\ 0 & \mathbf{S}_n^1 \end{pmatrix}, \quad \mathbf{S}_n^i = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j,2k+i)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix}.$$

Note that in the normalizing matrices  $\mathbf{S}_n^i$  the sum is over all observed cells for which a daughter of type  $i$  is observed, and not merely over all observed cells. To estimate the variance parameters  $\sigma_i^2$  and  $\rho$ , we define the empiric residuals. For all  $2^\ell \leq k \leq 2^{\ell+1} - 1$  and  $1 \leq j \leq m$  set

$$\begin{cases} \hat{\varepsilon}_{(j,2k)} & = \delta_{(j,2k)}(X_{(j,2k)} - \hat{a}_{0,\ell} - \hat{b}_{0,\ell} X_{(j,k)}), \\ \hat{\varepsilon}_{(j,2k+1)} & = \delta_{(j,2k+1)}(X_{(j,2k+1)} - \hat{a}_{1,\ell} - \hat{b}_{1,\ell} X_{(j,k)}). \end{cases} \quad (3)$$

We propose the following empirical estimators

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

where  $|\mathbb{T}_n^{*i}|$  is the set of all cells which have at least one offspring of type  $i$ , for  $i \in \{0, 1\}$  and  $|\mathbb{T}_n^{*01}|$  is the set of all the cells which have exactly two offspring, in the  $m$  trees up to generation  $n$ .

## 4.3 Consistency and normality

We now state the convergence results we obtain for the estimators above. The assumptions (H.1) to (H.6) are given in A.2. These results hold on the non-extinction set  $\bar{\mathcal{E}}$ .

**Theorem 4.1** *Under assumptions (H.5-6) and for all  $i, l_0$  and  $l_1$  in  $\{0, 1\}$ ,  $\hat{p}_n^{(i)}(l_0, l_1)$  converges to  $p^{(i)}(l_0, l_1)$  almost surely on  $\bar{\mathcal{E}}$ . Under assumptions (H.0-6),  $\hat{\theta}_n$ ,  $\hat{\sigma}_{0,n}^2$ ,  $\hat{\sigma}_{1,n}^2$  and  $\hat{\rho}_n$  converge to  $\theta$ ,  $\sigma_0^2$ ,  $\sigma_1^2$  and  $\rho$  respectively, almost surely on  $\bar{\mathcal{E}}$ .*

The asymptotic normality results are only valid conditionally to the non-extinction of the global cell population.

**Theorem 4.2** *Under assumptions (H.5-6) we have*

$$\sqrt{|\mathbb{T}_{n-1}^{*0}|}(\hat{\mathbf{p}}_n - \mathbf{p}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{V}),$$

and under assumptions (H.0-6), we have

$$\begin{aligned} \sqrt{|\mathbb{T}_{n-1}^{*0}|}(\hat{\theta}_n - \theta) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}_\theta), & \sqrt{|\mathbb{T}_n^{*0}|}(\hat{\sigma}_{0,n}^2 - \sigma_0^2, \hat{\sigma}_{1,n}^2 - \sigma_1^2)^t &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}_\sigma), \\ \sqrt{|\mathbb{T}_{n-1}^{*01}|}(\hat{\rho}_n - \rho) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma_\rho), \end{aligned}$$

conditionally to  $\bar{\mathcal{E}}$ . The explicit form of the variance matrices  $\mathbf{V}$ ,  $\mathbf{\Gamma}_\theta$ ,  $\mathbf{\Gamma}_\sigma$  and of  $\gamma_\rho$  is given in Eq. (6), (7), (9) and (10) respectively.

The proofs of these results are given in B.2 and B.3 for the GW process and in C.2 and C.3 for the BAR process. From the asymptotic normality, one can naturally construct confidence intervals and tests. Their explicit formulas are given in B.4 and C.4.

## 5 Data analysis

We applied our procedure to the *Escherichia coli* data of Stewart et al (2005). The biological issue addressed is aging in single cell organisms. *E. coli* is a rod-shaped bacterium that reproduces by dividing in the middle. Each cell has thus a new *pole* (due to the division of its mother) and an old one (one of the two poles of its mother). The cell that inherits the old pole of its mother is called the *old pole* cell, the other one is called the *new pole* cell. Therefore, each cell has a *type*: old pole or new pole cell, inducing asymmetry in the cell division. On a binary tree, the new pole cells are labelled by an even number and the old pole cells by an odd number.

Stewart et al (2005) filmed 94 colonies of dividing *E. coli* cells, determining the complete lineage and the growth rate of each cell. The number of divisions goes from four to nine. The 94 data sets gather  $|\mathbb{T}_9^*| = 22394$  data (11189 of type even and 11205 of type odd). Not a single data tree is complete. Missing data mainly do not come from cell death (only 16 cells are recorded to die) but from measurement difficulties due mostly to overlapping cells or cells wandering away from the field of view. Note also that for a growth rate to be recorded, the cell needs to be observed through its whole life cycle. If this is not the case, there is no record at all, so that a censored data model is not relevant. The observed average growth rate of even (resp. odd) cells is 0.0371 (resp. 0.0369). These data were investigated in (Stewart et al, 2005; Guyon et al, 2005; Guyon, 2007; de Saporta et al, 2011; de Saporta et al, 2012).

Stewart et al (2005) proposed a statistical study of the averaged genealogy and pair-wise comparison of sister cells. They concluded that the old pole cells exhibit cumulatively slowed growth, less offspring biomass production and an increased probability of death whereas single-experiment analyses did not. However they assumed independence between the averaged couples of sister cells, which does not hold in such genealogies.

The other studies are based on single-tree analyses instead of averaging all the genealogical trees. Guyon et al (2005) model the growth rate by a Markovian bifurcating process, but their procedure does not take into account the dependence between pairs of sister cells either. The asymmetry was rejected ( $p\text{-value} < 0.1$ ) in half of the experiments so that a global conclusion was difficult. Guyon (2007) has then investigated the asymptotical properties of a more general asymmetric Markovian bifurcating autoregressive process, and he rigorously constructed a Wald's type test to study the asymmetry of the process. However, his model does not take into account the possibly missing data from the genealogies. The author investigates the method on the 94 data sets but it is not clear how he manages missing data. More recently, de Saporta et al (2011) proposed a single-tree analysis with a rigorous method to deal with the missing data and carried out their analysis on the largest data set, concluding to asymmetry on this single set. Further single-tree studies of the 51 data sets issued from the 94 colonies containing at least 8 generations were conducted in de Saporta et al (2012). The symmetry hypothesis is rejected in one set out of four for  $(a_0, b_0) = (a_1, b_1)$  and one out of eight for  $a_0/(1 - b_0) = a_1/(1 - b_1)$  forbidding a global conclusion. Simulation studies tend to prove that the power of the tests on single-trees is quite low for only eight or nine generations. This is what motivated the present study and urged us to use all the data available in one global estimation, rather than single-tree analyses.

In this section, we propose a new investigation of *E. coli* data of (Stewart et al, 2005) where for the first time the dependence structure between cells within a genealogy is fully taken into account, missing data are taken care of rigorously, all the available data, i.e. the 94 sets, are analyzed at once and both the growth rate and the number/type of descendants are investigated. It is sensible to consider that all the data sets correspond to BAR processes with the same coefficients as the experiments where conducted in similar conditions. Moreover, a direct comparison of single-tree estimations would be meaningless as the data trees do not all have the same number of generations, and it would be impossible to determine whether variations in the computed single-tree estimators come from an intrinsic variability between trees or just the low accuracy of the estimators for small trees. The original estimation procedure described in Section 2 enables us to use all the information available without the drawbacks of low accuracy for estimators or low power for tests on small single-trees.

## 5.1 Symmetry of the BAR process

We now give the results of our new investigation of the E. coli growth rate data of (Stewart et al, 2005). We suppose that the growth rate of cells in each lineage is modeled by the BAR process defined in Eq. (1) and observed through the two-type GW process defined in section 3.2. The experiments were independent and lead in the same conditions corresponding to independence and identical distribution of the processes  $(X_{(j,\cdot)}, \delta_{(j,\cdot)})$ ,  $1 \leq j \leq m$ .

We first give the point and interval estimation for the various parameters of the BAR process. Table 1 gives the estimation  $\hat{\theta}_9$  of  $\theta$  with the 95% confidence interval (CI) of each coefficient together with an estimation of  $a_i/(1-b_i)$ . This value is interesting in itself as  $a_i/(1-b_i)$  is the fixed point of the equation  $\mathbb{E}[X_{2k+i}] = a_i + b_i\mathbb{E}[X_i]$ . Thus it corresponds to the asymptotic mean growth rate of the cells in the lineage always inheriting the new pole from the mother ( $i = 0$ ) or always inheriting the old pole ( $i = 1$ ). The confidence intervals of  $b_0$  and  $b_1$  show that the *non explosion* assumption  $|b_0| < 1$  and  $|b_1| < 1$  is satisfied. Note that although the number of observed generations  $n = 9$  may seem too small to obtain the consistency of our estimators, Theorem 4.2 shows that their variance is of order  $|\mathbb{T}_n^*|^{-1/2}$ . Here the total number of observed cells is high enough as  $|\mathbb{T}_9^*| = 22394$ . In addition, an empirical study of the convergence rate on simulated data is conducted in the next section to validate that 9 observed generations is enough.

parameter	estimation	CI	parameter	estimation	CI
$a_0$	0.0203	[0.0202; 0.0204]	$a_1$	0.0195	[0.0194; 0.0196]
$b_0$	0.4615	[0.4417; 0.4812]	$b_1$	0.4782	[0.4631; 0.4933]
$a_0/(1-b_0)$	0.03773	[0.03756; 0.03790]	$a_1/(1-b_1)$	0.03734	[0.03717; 0.03752]

Table 1: Estimation and 95 % CI of  $\theta$  and  $a_i/(1-b_i)$ .

Table 2 gives the estimations  $\hat{\sigma}_{i,9}^2$  of  $\sigma_i^2$  and  $\hat{\rho}_9$  of  $\rho$  with the 95% CI of each coefficient. The hypothesis of equality of variances  $\sigma_0^2 = \sigma_1^2$  is not rejected (p-value= 0.19). From the biological point of view, this result is not surprising as the noise sequence represents the influence of the environment and both sister cells are born and grow in the same local environment.

We now turn to the results of symmetry tests. The hypothesis of equality of the couples  $(a_0, b_0) = (a_1, b_1)$  is strongly rejected (p-value =  $10^{-5}$ ). The hypothesis of the equality of the two fixed points  $a_0/(1-b_0)$  and  $a_1/(1-b_1)$  of the BAR process is also rejected (p-value =  $2 \cdot 10^{-3}$ ). We can therefore rigorously confirm that there is a statistically significant asymmetry in the division of E. coli. Biologically we can thus conclude that the growth rates of the old pole and new pole cells do have different dynamics. This is interpreted as aging for the single cell organism E. coli, see Stewart et al (2005); Wang et al (2010).

## 5.2 Symmetry of the GW process

Let us now turn to the asymmetry of the GW process itself. Note that to our best knowledge, it is the first time this question is investigated for the E. coli data of (Stewart et al, 2005). We estimated the parameters  $p^{(i)}(l_0, l_1)$  of the reproduction laws of the underlying GW process. Table 3 gives the estimations  $\hat{p}_9^{(i)}(l_0, l_1)$  of the  $p^{(i)}(l_0, l_1)$ . The estimation of the dominant eigenvalue  $\pi$  of the descendants matrix of the GW processes (characterizing extinction, see A.1) is  $\hat{\pi}_9 = 1.204$  with CI [1.191; 1.217]. The non-extinction hypothesis ( $\pi > 1$ ) is thus satisfied.

The means of the two reproduction laws  $\mathbf{p}^{(0)}$  and  $\mathbf{p}^{(1)}$  are estimated at  $\hat{m}_9^0 = 1.2048$  and  $\hat{m}_9^1 = 1.2032$  respectively. The hypothesis of the equality of the mean numbers of offspring is not rejected (p-value = 0.9). However, Table 3 shows that there is a statistically significant difference between vectors  $\mathbf{p}^{(0)}$  and  $\mathbf{p}^{(1)}$  as none of the confidence intervals intersect. Indeed, the symmetry hypothesis  $\mathbf{p}^{(0)} = \mathbf{p}^{(1)}$  is rejected with p-value =  $2 \cdot 10^{-5}$ . However, it is not possible to interpret this asymmetry in terms of the division of E. coli, since the cause of missing data is mostly due to observation difficulties rather than some intrinsic behavior of the cells.

parameter	estimation	CI
$\sigma_0^2$	$2.28 \cdot 10^{-5}$	$[0.88 \cdot 10^{-5}; 3.67 \cdot 10^{-5}]$
$\sigma_1^2$	$1.34 \cdot 10^{-5}$	$[1.29 \cdot 10^{-5}; 1.40 \cdot 10^{-5}]$
$\rho$	$0.48 \cdot 10^{-5}$	$[0.44 \cdot 10^{-5}; 0.52 \cdot 10^{-5}]$

Table 2: Estimation and 95 % CI of  $\sigma_i^2$  and  $\rho$

Table 3: Estimation and 95 % CI of  $\mathbf{p}$ .

parameter	estimation	CI	parameter	estimation	CI
$p^{(0)}(1, 1)$	0.56060	[0.56055; 0.56065]	$p^{(1)}(1, 1)$	0.55928	[0.55923; 0.55933]
$p^{(0)}(1, 0)$	0.03621	[0.03620; 0.03622]	$p^{(1)}(1, 0)$	0.04707	[0.04706; 0.04708]
$p^{(0)}(0, 1)$	0.04740	[0.04739; 0.04741]	$p^{(1)}(0, 1)$	0.03755	[0.03754; 0.03756]
$p^{(0)}(0, 0)$	0.35579	[0.35574; 0.35583]	$p^{(1)}(0, 0)$	0.35611	[0.35606; 0.35616]

## 6 Simulation study

To investigate the empirical rate of convergence of our estimators as well as the power of the symmetry tests we have performed simulations of our coupled BAR-GW model. In particular, we study how they depend both on the ratio of missing data and on the number of observed generations.

In a complete binary tree, the number of descendants of each individual is exactly 2. In our model of GW tree, the number of descendants is random and its average is asymptotically of the order of the dominant eigenvalue  $\pi$  of the descendants matrix of the GW processes, see A.1. Therefore  $\pi$  characterizes the scarcity of data: if  $\pi = 2$ , the whole tree is observed and there are no missing data; as  $\pi$  decreases, the average number of missing data increases (we choose  $\pi > 1$  to avoid almost sure extinction). In addition, for a single GW tree, the number of observed individuals up to generation  $n$  is asymptotically of order  $\pi^n$ .

We have simulated the BAR-GW process for 19 distinct parameters sets, see Tables 4 and 5. Sets 1 to 10 are symmetric with decreasing  $\pi$  (from 2 to 1.08), sets 11 to 19 are asymmetric with decreasing  $\pi$  (from 1.9 to 1.1). The parameters of the BAR process are chosen close to the estimated values on E. coli data whereas the GW parameters are chosen to obtain different values of  $\pi$ . Notice that set 18 is close to the estimated values for E. coli data. For each set, we simulated the BAR-GW process up to generation 15 and ran our estimation procedure on  $m = 100$  replicated trees ( $m = 94$  for E. coli data). Each estimation was repeated 1000 times.

set	$a_0$	$b_0$	$a_1$	$b_1$	$\sigma_0$	$\sigma_1$	$\rho$
1 to 10	0.02	0.47	0.02	0.47	$1.8 \cdot 10^{-5}$	$1.8 \cdot 10^{-5}$	$0.5 \cdot 10^{-5}$
11 to 19	0.0203	0.4615	0.0195	0.4782	$2.28 \cdot 10^{-5}$	$1.34 \cdot 10^{-5}$	$0.48 \cdot 10^{-5}$

Table 4: Parameters sets for the simulated BAR processes.

We first investigate the significant level and power of our symmetry tests on the simulated data. The asymptotic properties of the tests are given in C.4. Table 6 (resp. Table 7) gives the proportion of reject (significant level 5%) under H0 (symmetric sets 1 to 10) and under H1 (asymmetric sets 11 to 19) for the test of symmetry of fixed points H0:  $a_0/(1 - b_0) = a_1/(1 - b_1)$  (resp. the test of equality of vectors H0:  $(a_0, b_0) = (a_1, b_1)$ ). In both cases, the proportion of reject under H0 is close to the significant level regardless of the number of observed generations (from 5 generations on) and of the value of  $\pi$ . We thus can conclude that from  $n = 5$  on the asymptotic  $\chi^2$  law is valid. Under H1, the proportion of reject increases when the number of observed generations increases and decreases when  $\pi$  decreases. Recall that the number of observed individuals up to generation  $n$  is asymptotically of order  $m\pi^n$  ( $m = 100$ ) and the power is strongly linked to the number of

set	$\mathbf{p}^{(0)}$	$\mathbf{p}^{(1)}$	$\pi$
1	(1,0,0,0)	(1,0,0,0)	2
2	(0.90,0.04,0.04,0.02)	(0.90,0.04,0.04,0.02)	1.88
3	(0.85,0.04,0.04,0.07)	(0.85,0.04,0.04,0.07)	1.78
4	(0.80,0.04,0.04,0.12)	(0.80,0.04,0.04,0.12)	1.68
5	(0.75,0.04,0.04,0.17)	(0.75,0.04,0.04,0.17)	1.58
6	(0.70,0.04,0.04,0.22)	(0.70,0.04,0.04,0.22)	1.48
7	(0.65,0.04,0.04,0.27)	(0.65,0.04,0.04,0.27)	1.38
8	(0.60,0.04,0.04,0.32)	(0.60,0.04,0.04,0.32)	1.28
9	(0.55,0.04,0.04,0.37)	(0.55,0.04,0.04,0.37)	1.18
10	(0.50,0.04,0.04,0.42)	(0.50,0.04,0.04,0.42)	1.08
11	(0.901,0.045,0.055,0.019)	(0.899,0.055,0.045,0.021)	1.9
12	(0.851,0.045,0.055,0.069)	(0.849,0.055,0.045,0.071)	1.8
13	(0.801,0.045,0.055,0.119)	(0.799,0.055,0.045,0.121)	1.7
14	(0.751,0.045,0.055,0.169)	(0.749,0.055,0.045,0.171)	1.6
15	(0.701,0.045,0.055,0.219)	(0.699,0.055,0.045,0.221)	1.5
16	(0.651,0.045,0.055,0.269)	(0.649,0.055,0.045,0.271)	1.4
17	(0.601,0.045,0.055,0.319)	(0.659,0.055,0.045,0.321)	1.3
18	(0.551,0.045,0.055,0.369)	(0.549,0.055,0.045,0.371)	1.2
19	(0.501,0.045,0.055,0.419)	(0.499,0.055,0.045,0.421)	1.1

Table 5: Parameters sets for the simulated GW processes.

observed data. For instance, it is perfect for high numbers of observed generations and high  $\pi$  when the expected number of observed data is huge and it is low for low  $\pi$  even for high numbers of observed generations.

Next, we investigate the empirical convergence rate of the estimation error  $\|\hat{\theta}_n - \theta\|_2$  both as a function of the number of observed generations  $n$  and of  $\pi$ . Figure 2 (resp. Figure 3) shows the distribution of  $\|\hat{\theta}_n - \theta\|_2 / \|\theta\|_2$  for  $n = 9$  (reps.  $n = 15$ ) observed generations for the asymmetric parameters sets 11 to 19. It illustrates how the error deteriorates as  $\pi$  decreases, i.e. as the ratio of missing data increases. The two figures have the same scale to illustrate how the relative error decreases when the number of observed generations is higher.

We know from Theorem 4.2 that the variance of  $\hat{\theta}_n$  is of order  $|\mathbb{T}_n^*|^{-1/2}$  which asymptotically has the same order of magnitude as  $\pi^{-n/2}$ . In order to check how soon (in terms of the number  $n$  of observed generations) this asymptotic rate is reached, we fitted the logarithm of the errors  $\|\hat{\theta}_n - \theta\|_2$  (averaged over the 1000 simulations) to a linear function of  $n$  for each parameters set (using the errors from generation 8 to generation 15). The results are shown on Figure 4.

We also compare the computed slopes of the linear functions to the theoretical value  $-\log(\pi)/2$  for the various parameters sets. The results are given in Table 8 and show that the asymptotic rate is reached from generation 8 on. It thus validates the accuracy of the study of *E. coli* data conducted in the previous section.

## 7 Conclusion

In this paper, we first propose a statistical model to estimate and test asymmetry of a quantitative characteristic associated to each node of a family of incomplete binary trees, without aggregating single-tree estimators. An immediate application is the investigation of asymmetry in cell lineage data. This model of coupled GW-BAR process generalizes all the previous methods on this subject in the literature because it rigorously takes into account:

- the dependence of the characteristic of a cell to that of its mother and the correlation between two sisters through the BAR model,
- the possibly missing data through the GW model,

generation	5	6	7	8	9	10	11	12	13	14	15
set 1	0.037	0.050	0.047	0.048	0.046	0.056	0.046	0.047	0.053	0.041	0.042
set 2	0.045	0.047	0.047	0.052	0.048	0.053	0.050	0.042	0.040	0.050	0.049
set 3	0.051	0.048	0.043	0.048	0.057	0.064	0.046	0.045	0.048	0.049	0.052
set 4	0.051	0.055	0.052	0.056	0.049	0.047	0.052	0.050	0.059	0.058	0.051
set 5	0.052	0.052	0.049	0.053	0.061	0.065	0.052	0.054	0.040	0.045	0.042
set 6	0.045	0.036	0.039	0.035	0.051	0.062	0.054	0.061	0.055	0.043	0.046
set 7	0.045	0.048	0.045	0.044	0.048	0.037	0.041	0.044	0.050	0.049	0.049
set 8	0.046	0.044	0.044	0.049	0.047	0.048	0.042	0.038	0.043	0.043	0.054
set 9	0.053	0.052	0.058	0.061	0.060	0.055	0.052	0.052	0.045	0.053	0.051
set 10	0.039	0.038	0.051	0.046	0.054	0.049	0.054	0.046	0.047	0.046	0.039
set 11	0.448	0.697	0.926	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000
set 12	0.356	0.568	0.832	0.975	0.999	1.000	1.000	1.000	1.000	1.000	1.000
set 13	0.305	0.497	0.711	0.894	0.991	1.000	1.000	1.000	1.000	1.000	1.000
set 14	0.252	0.399	0.586	0.777	0.926	0.994	0.999	1.000	1.000	1.000	1.000
set 15	0.208	0.293	0.417	0.608	0.808	0.930	0.990	1.000	1.000	1.000	1.000
set 16	0.200	0.279	0.390	0.502	0.668	0.790	0.905	0.977	0.997	1.000	1.000
set 17	0.174	0.234	0.287	0.364	0.458	0.566	0.696	0.829	0.912	0.967	0.990
set 18	0.130	0.165	0.209	0.255	0.335	0.382	0.451	0.548	0.650	0.725	0.811
set 19	0.118	0.142	0.174	0.190	0.207	0.245	0.300	0.330	0.371	0.416	0.459

Table 6: Proportion of p-values  $\leq 5\%$  for the equality of fixed points test (1000 replications).

generation	5	6	7	8	9	10	11	12	13	14	15
set 1	0.045	0.062	0.038	0.051	0.051	0.051	0.040	0.033	0.060	0.036	0.049
set 2	0.036	0.055	0.049	0.054	0.044	0.048	0.032	0.037	0.039	0.047	0.041
set 3	0.040	0.044	0.045	0.053	0.057	0.042	0.050	0.039	0.053	0.045	0.039
set 4	0.053	0.058	0.055	0.047	0.053	0.056	0.061	0.049	0.052	0.048	0.043
set 5	0.050	0.050	0.049	0.052	0.056	0.049	0.047	0.052	0.044	0.048	0.044
set 6	0.058	0.043	0.040	0.043	0.052	0.053	0.057	0.056	0.048	0.043	0.051
set 7	0.032	0.048	0.042	0.032	0.044	0.040	0.046	0.035	0.041	0.052	0.047
set 8	0.059	0.052	0.058	0.055	0.052	0.050	0.053	0.044	0.050	0.052	0.050
set 9	0.054	0.049	0.046	0.042	0.048	0.042	0.044	0.050	0.042	0.047	0.045
set 10	0.042	0.049	0.045	0.044	0.045	0.053	0.051	0.046	0.043	0.044	0.037
set 11	0.414	0.678	0.920	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
set 12	0.310	0.557	0.833	0.980	0.999	1.000	1.000	1.000	1.000	1.000	1.000
set 13	0.286	0.454	0.703	0.902	0.996	1.000	1.000	1.000	1.000	1.000	1.000
set 14	0.218	0.367	0.555	0.775	0.938	0.995	1.000	1.000	1.000	1.000	1.000
set 15	0.193	0.276	0.391	0.596	0.789	0.934	0.990	1.000	1.000	1.000	1.000
set 16	0.175	0.237	0.354	0.479	0.641	0.800	0.925	0.980	0.997	1.000	1.000
set 17	0.156	0.188	0.246	0.362	0.437	0.540	0.683	0.806	0.919	0.968	0.989
set 18	0.126	0.152	0.193	0.247	0.285	0.359	0.410	0.525	0.633	0.726	0.819
set 19	0.110	0.116	0.140	0.161	0.192	0.229	0.271	0.320	0.365	0.395	0.452

Table 7: Proportion of p-values  $\leq 5\%$  for the test  $(a_0, b_0) = (a_1, b_1)$  (1000 replications).

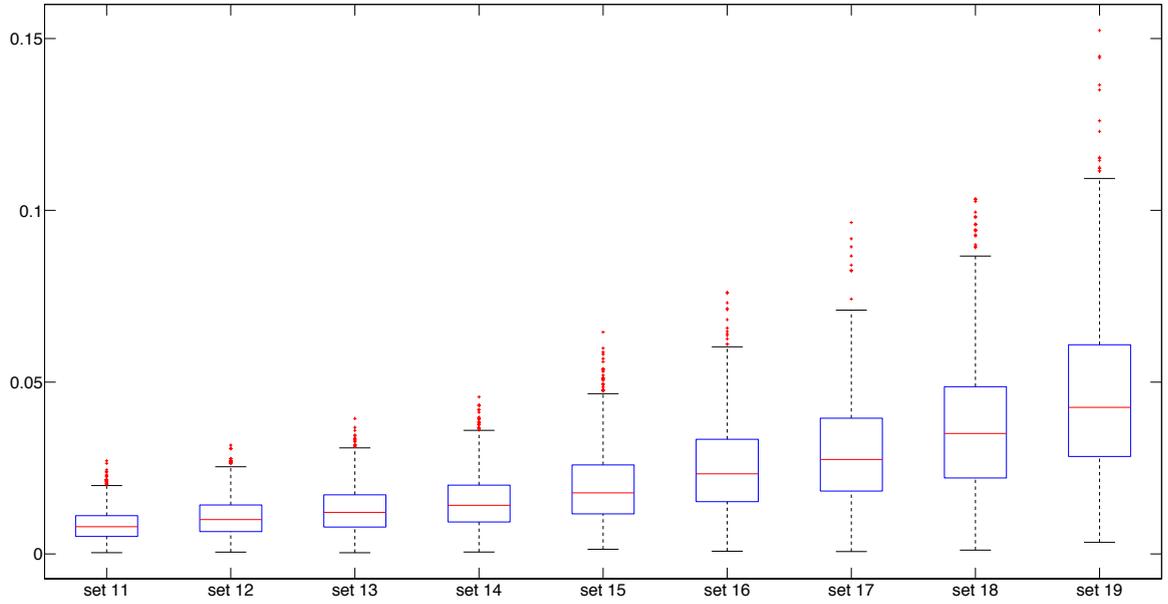


Figure 2: Boxplot of the estimation of the relative error  $\|\hat{\theta}_9 - \theta\|_2 / \|\theta\|_2$  for the data sets 11 to 19 (decreasing  $\pi$ )

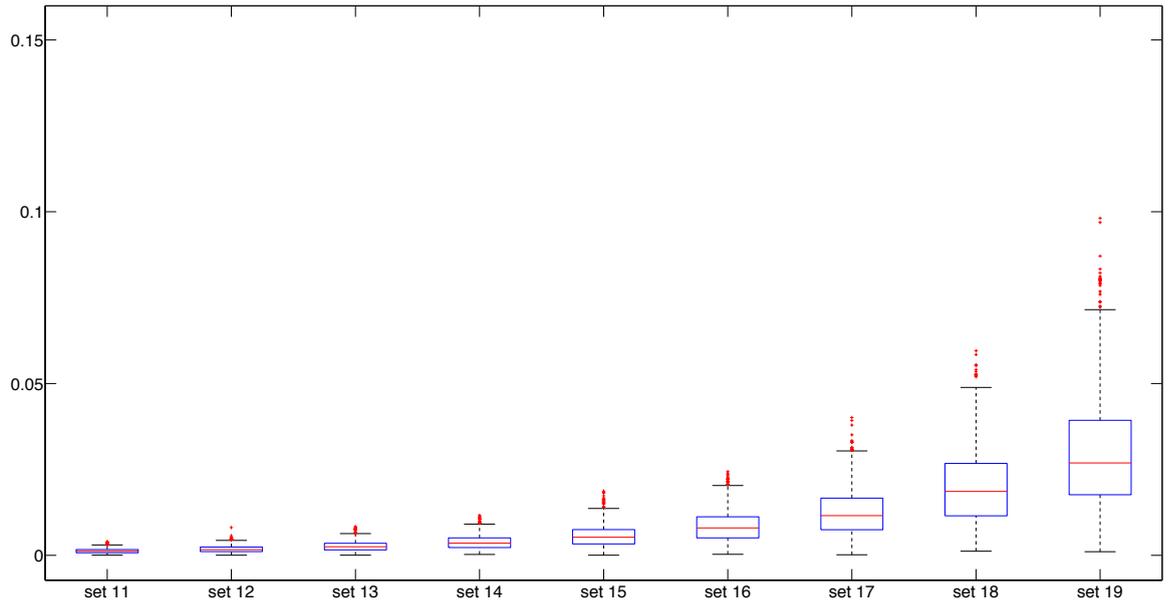


Figure 3: Boxplot of the estimation of the relative error  $\|\hat{\theta}_9 - \theta\|_2 / \|\theta\|_2$  for the data sets 11 to 19 (decreasing  $\pi$ )

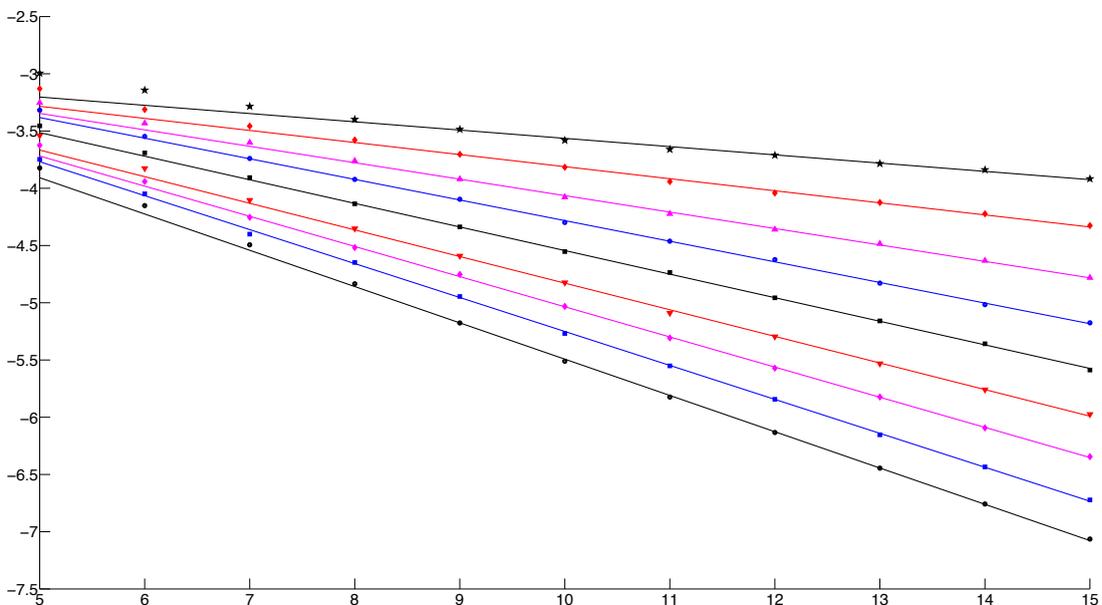


Figure 4: Logarithm of the error  $\log \|\widehat{\theta}_n - \theta\|_2$  as a function of the number  $n$  of observed generations for the asymmetric parameters set 11 to 19 (from bottom to top: set 11-black circles, set 12-blue squares, set 13-magenta diamonds, set 14-red triangles, set 15-black squares, set 16-blue circles, set 17-magenta triangles, set 18-red diamonds, set 19-black stars).

set	11	12	13	14	15	16	17	18	19
empirical slope	-0.3170	-0.2966	-0.2634	-0.2325	-0.2060	-0.1801	-0.1413	-0.0953	-0.0672
$-\log(\pi)/2$	-0.3209	-0.2939	-0.2653	-0.2350	-0.2027	-0.1682	-0.1312	-0.0912	-0.0477

Table 8: Logarithm of empirical convergences rates vs theoretical rate

- the information from several sets of data obtained in similar experimental conditions without the drawbacks of poor accuracy or power for small single-trees.

Furthermore, we propose the estimation of parameters of a two-type GW process in the specific context of a binary tree with a fine observation, namely the presence or absence of each cell of the complete binary tree is known. In the context where missing offspring really come from the intrinsic reproduction, and not from faulty measures, the asymmetry of the parameters of the GW process can be applied to cell lineage data and be interpreted as a difference in the reproduction laws between the two different types of cell.

We applied our procedure to the *E. coli* data of Stewart et al (2005) and concluded there exists a statistically significant asymmetry in this cell division. Results were validated by simulation studies of the empirical rate of convergence of the estimators and power of the tests.

## A Technical assumptions and notation

Our convergence results rely on martingale theory and the use of several carefully chosen filtrations regarding the BAR and/or GW process. The approach is similar to that of de Saporta et al (2011); de Saporta et al (2012), but their results cannot be directly applied here. This is mainly due to our choice of the global non-extinction set as the union and not the intersection of the non-extinction sets of each replicated process preventing us from directly using convergence results on single-tree estimators. We now give some additional notation and the precise assumptions of our convergence theorems.

## A.1 Generations and extinction

We first introduce some notation about the complete and observed genealogy trees that will be used in the sequel. For all  $n \geq 1$ , denote the  $n$ -th generation of any given tree by  $\mathbb{G}_n = \{k, 2^n \leq k \leq 2^{n+1} - 1\}$ . 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. 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  $|\mathbb{T}_n| = 2^{n+1} - 1$ . Finally, we define the sets of observed individuals in each tree  $\mathbb{G}_{j,n}^* = \{k \in \mathbb{G}_n : \delta_{(j,k)} = 1\}$  and  $\mathbb{T}_{j,n}^* = \{k \in \mathbb{T}_n : \delta_{(j,k)} = 1\}$ , and set

$$|\mathbb{G}_n^*| = \sum_{j=1}^m |\mathbb{G}_{j,n}^*| \quad \text{and} \quad |\mathbb{T}_n^*| = \sum_{j=1}^m |\mathbb{T}_{j,n}^*|,$$

the total number of observed cells in all  $m$  trees in generation  $n$  and up to generation  $n$  respectively. We next need to characterize the possible extinction of the GW processes, that is where  $|\mathbb{T}_n^*|$  does not tend to infinity with  $n$ . For  $1 \leq j \leq m$  and  $n \geq 1$ , we define the number of observed cells among the  $n$ -th generation of the  $j$ -th tree, distinguishing according to their type, by

$$Z_{j,n}^0 = \sum_{k \in \mathbb{G}_{n-1}} \delta_{(j,2k)} \quad \text{and} \quad Z_{j,n}^1 = \sum_{k \in \mathbb{G}_{n-1}} \delta_{(j,2k+1)},$$

and we set  $\mathbf{Z}_{j,n} = (Z_{j,n}^0, Z_{j,n}^1)$ . For all  $j$ , the process  $(\mathbf{Z}_{j,n})$  thus defined is a two-type GW process, see Harris (1963). We define the descendants matrix  $\mathbf{P}$  of the GW process by

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

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  $p_{il}$  is thus the expected number of descendants of type  $l$  of an individual of type  $i$ . It is well-known 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 and admits a positive left eigenvector, see e.g. (Harris, 1963, Theorem 5.1). In that case, we denote by  $\mathbf{z} = (z^0, z^1)$  the left eigenvector of  $\mathbf{P}$  associated with the dominant eigenvalue  $\pi$  and satisfying  $z^0 + z^1 = 1$ . Let  $\mathcal{E}_j = \bigcup_{n \geq 1} \{\mathbf{Z}_{j,n} = (0, 0)\}$  be the event corresponding to the case when there are no cells left to observe in the  $j$ -th tree. We will denote  $\bar{\mathcal{E}}_j$  the complementary set of  $\mathcal{E}_j$ . We are interested in asymptotic results on the set where there is an infinity of  $X_{(j,k)}$  to be observed that is on the union of the non-extinction sets  $\bar{\mathcal{E}}_j$  denoted by

$$\bar{\mathcal{E}} = \bigcup_{j=1}^m \bar{\mathcal{E}}_j = \left\{ \lim_{n \rightarrow \infty} |\mathbb{T}_n^*| = \infty \right\}.$$

Note that we allow some trees to extinct, as long as there is at least one tree still growing. This assumption is natural in view of the E. coli data as the collected genealogies do have a significantly different numbers of observed generations (from 4 up to 9).

## A.2 Assumptions

Our inference is based on the  $m$  i.i.d. replicas of the observed BAR process, i.e. the available information is given by the sequence  $(\delta_{(j,k)}, \delta_{(j,k)} X_{(j,k)})_{1 \leq j \leq m, k \geq 1}$ . We first introduce the natural generation-wise filtrations of the BAR processes. For all  $1 \leq j \leq m$ , denote by  $\mathbb{F}_j = (\mathcal{F}_{j,n})_{n \geq 1}$  the natural filtration associated with the  $j$ -th copy of the BAR process, which means that  $\mathcal{F}_{j,n}$  is the  $\sigma$ -algebra generated by all individuals of the  $j$ -th tree up to the  $n$ -th generation,  $\mathcal{F}_{j,n} = \sigma\{X_{(j,k)}, k \in \mathbb{T}_n\}$ . For all  $1 \leq j \leq m$ , we also define the observation filtrations as  $\mathcal{O}_{j,n} = \sigma\{\delta_{(j,k)}, k \in \mathbb{T}_n\}$ , and the sigma fields  $\mathcal{O}_j = \sigma\{\delta_{(j,k)}, k \geq 1\}$ .

We make the following main assumptions on the BAR and GW processes.

**(H.0)** The parameters  $(a_0, b_0, a_1, b_1)$  satisfy the usual stability assumption  $0 < \max\{|b_0|, |b_1|\} < 1$ .

**(H.1)** For all  $1 \leq j \leq m$ ,  $n \geq 0$ ,  $k \in \mathbb{G}_{n+1}$ ,  $\mathbb{E}[\varepsilon_{(j,k)}^{16}] < \infty$  and  $\mathbb{E}[X_{(j,1)}^{16}] < \infty$ .

For all  $1 \leq j \leq m$ ,  $n \geq 0$ ,  $k \in \mathbb{G}_n$  and  $i \in \{0, 1\}$ , one a.s. has

$$\begin{aligned}\mathbb{E}[\varepsilon_{(j,2k+i)}|\mathcal{F}_{j,n}] &= 0, \quad \mathbb{E}[\varepsilon_{(j,2k+i)}^2|\mathcal{F}_{j,n}] = \sigma_i^2, \quad \mathbb{E}[\varepsilon_{(j,2k+i)}^3|\mathcal{F}_{j,n}] = \lambda_i, \\ \mathbb{E}[\varepsilon_{(j,2k+i)}^4|\mathcal{F}_{j,n}] &= \tau_i^4, \quad \mathbb{E}[\varepsilon_{(j,2k+i)}^8|\mathcal{F}_{j,n}] = \gamma_i^8, \quad \mathbb{E}[\varepsilon_{(j,2k+i)}^{16}|\mathcal{F}_{j,n}] = \mu_i^{16}.\end{aligned}$$

For all  $1 \leq j \leq m$ ,  $n \geq 0$ ,  $k \in \mathbb{G}_n$ , one a.s. has

$$\begin{aligned}\mathbb{E}[\varepsilon_{(j,2k)}\varepsilon_{(j,2k+1)}|\mathcal{F}_{j,n}] &= \rho, \quad \mathbb{E}[\varepsilon_{(j,2k)}^2\varepsilon_{(j,2k+1)}^2|\mathcal{F}_{j,n}] = \nu^2, \quad \mathbb{E}[\varepsilon_{(j,2k)}^8\varepsilon_{(j,2k+1)}^8|\mathcal{F}_{j,n}] = \eta^8, \\ \mathbb{E}[\varepsilon_{(j,2k)}^2\varepsilon_{(j,2k+1)}|\mathcal{F}_{j,n}] &= \alpha, \quad \mathbb{E}[\varepsilon_{(j,2k)}\varepsilon_{(j,2k+1)}^2|\mathcal{F}_{j,n}] = \beta.\end{aligned}$$

**(H.2)** For all  $1 \leq j \leq m$  and  $n \geq 0$  the vectors  $\{(\varepsilon_{(j,2k)}, \varepsilon_{(j,2k+1)}), k \in \mathbb{G}_n\}$  are conditionally independent given  $\mathcal{F}_{j,n}$ .

**(H.3)** The sequences  $(\varepsilon_{(1,k)})_{k \geq 2}, (\varepsilon_{(2,k)})_{k \geq 2}, \dots, (\varepsilon_{(m,k)})_{k \geq 2}$  are independent. The random variables  $(X_{(j,1)})_{1 \leq j \leq m}$  are independent and independent from the noise sequences.

**(H.4)** For all  $1 \leq j \leq m$ , the sequence  $(\delta_{(j,k)})_{k \geq 1}$  is independent from the sequences  $(X_{(j,k)})_{k \geq 1}$  and  $(\varepsilon_{(j,k)})_{k \geq 2}$ .

**(H.5)** The sequences  $(\delta_{(1,k)})_{k \geq 2}, (\delta_{(2,k)})_{k \geq 2}, \dots, (\delta_{(m,k)})_{k \geq 2}$  are independent.

We also make the following super criticality assumption on the matrix  $\mathbf{P}$ .

**(H.6)** All entries of the matrix  $\mathbf{P}$  are positive: for all  $(i, l) \in \{0, 1\}^2$ ,  $p_{il} > 0$ , and the dominant eigenvalue is greater than one:  $\pi > 1$ .

If  $\pi > 1$ , it is well known, see e.g. Harris (1963), that the extinction probability of the GW processes is less than one: for all  $1 \leq j \leq m$ ,  $\mathbb{P}(\mathcal{E}_j) = p < 1$ . Under assumptions (H.5-6), one thus clearly has  $\mathbb{P}(\bar{\mathcal{E}}) = 1 - p^m > 0$ .

Note that under these assumptions, it is proved in de Saporta et al (2011) that the single-tree estimators  $\hat{\theta}_{j,n}$  are consistent on the single-tree non-extinction sets  $\bar{\mathcal{E}}_j$ . This result is based on the separate convergence of  $\mathbf{\Sigma}_{j,n}$  and  $\mathbf{\Sigma}_{j,n}\hat{\theta}_{j,n}$ . Therefore, the convergence of our global estimator  $\hat{\theta}_n$  is readily obtained on the intersection of the single-tree non-extinction sets  $\cap_{j=1}^m \bar{\mathcal{E}}_j$ , see Section 2. However, we are interested in the convergence of the global estimator on the larger set  $\bar{\mathcal{E}} = \cup_{j=1}^m \bar{\mathcal{E}}_j$ . This is why we cannot directly use the results of de Saporta et al (2011). We explain in the following sections how the ideas therein have to be adapted to this new framework.

### A.3 Additional estimators

From the estimators of the reproductions probabilities of the GW process, one can easily construct an estimator of the spectral radius  $\pi$  of the descendants matrix  $\mathbf{P}$  of the GW process. Indeed,  $\mathbf{P}$  is a  $2 \times 2$  matrix so that its spectral radius can be computed explicitly as a function of its coefficients, namely

$$\pi = \frac{1}{2} \left( \text{tr}(\mathbf{P}) + (\text{tr}(\mathbf{P})^2 - 4 \det(\mathbf{P}))^{1/2} \right).$$

Replacing the coefficients of  $\mathbf{P}$  by their empirical estimators, one obtains

$$\hat{\pi}_n = \frac{1}{2} (\hat{T}_n + (\hat{T}_n^2 - 4\hat{D}_n)^{1/2}).$$

where

$$\begin{aligned}\hat{T}_n &= \hat{p}_n^{(0)}(1, 0) + \hat{p}_n^{(0)}(1, 1) + \hat{p}_n^{(1)}(0, 1) + \hat{p}_n^{(1)}(1, 1), \\ \hat{D}_n &= (\hat{p}_n^{(0)}(1, 0) + \hat{p}_n^{(0)}(1, 1))(\hat{p}_n^{(1)}(0, 1) + \hat{p}_n^{(1)}(1, 1)) - (\hat{p}_n^{(0)}(0, 1) + \hat{p}_n^{(0)}(1, 1))(\hat{p}_n^{(1)}(1, 0) + \hat{p}_n^{(1)}(1, 1))\end{aligned}$$

are the empirical estimator of the trace  $\text{tr}(\mathbf{P})$  and the determinant  $\det(\mathbf{P})$  respectively. Finally, to compute confidence intervals for  $\sigma_i^2$  and  $\rho$ , we need an estimation of higher moments. We use again empirical estimators

$$\hat{\tau}_{i,n}^4 = \frac{1}{|\mathbb{T}_{n-1}^{*i}|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \hat{\varepsilon}_{(j,2k+i)}^4, \quad \hat{\nu}_n^2 = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \hat{\varepsilon}_{(j,2k)}^2 \hat{\varepsilon}_{(j,2k+1)}^2.$$

## B Convergence of estimators for the GW process

We now prove the convergence of the estimators for the GW process, that is the first parts of Theorems 4.1 and 4.2, together with additional technical results.

### B.1 Preliminary results: from single-trees to multiple trees

Our objective is to show that we can adapt the results in de Saporta et al (2011) to the multiple tree framework despite our choice of considering the union and not the intersection of the single-tree non-extinction sets. To this aim, we first need to recall Lemma A.3 of Bercu et al (2009).

**Lemma B.1** *Let  $(\mathbf{A}_n)$  be a sequence of real-valued matrices such that*

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

*In addition, let  $(\mathbf{X}_n)$  be a sequence of real-valued vectors which converges to a limiting value  $\mathbf{X}$ . Then, one has*

$$\lim_{n \rightarrow \infty} \sum_{\ell=0}^n \mathbf{A}_{n-\ell} \mathbf{X}_\ell = \mathbf{A} \mathbf{X}.$$

The next result is an adaptation of Lemma A.2 in Bercu et al (2009) to the GW tree framework. It gives a correspondence between sums on one generation and sums on the whole tree.

**Lemma B.2** *Let  $(x_n)$  be a sequence of real numbers and  $\pi > 1$ . One has*

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_k = x \iff \lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{G}_n} x_k = \frac{\pi - 1}{\pi} x.$$

*Proof:* Suppose that  $\pi^{-n} \sum_{k \in \mathbb{T}_n} x_k$  converges to  $x$ . Then one has

$$\frac{1}{\pi^n} \sum_{k \in \mathbb{G}_n} x_k = \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_k - \frac{1}{\pi} \frac{1}{\pi^{n-1}} \sum_{k \in \mathbb{T}_{n-1}} x_k \xrightarrow{n \rightarrow \infty} x - \frac{1}{\pi} x = \frac{\pi - 1}{\pi} x.$$

Conversely, if  $\pi^{-n} \sum_{k \in \mathbb{G}_n} x_k$  converges to  $y$ , as  $\mathbb{T}_n = \cup_{\ell=0}^n \mathbb{G}_\ell$ , one has

$$\frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_k = \sum_{\ell=0}^n \frac{1}{\pi^{n-\ell}} \frac{1}{\pi^\ell} \sum_{k \in \mathbb{G}_\ell} x_k \xrightarrow{n \rightarrow \infty} \frac{\pi}{\pi - 1} y,$$

using Lemma B.1 with  $\mathbf{A}_n = \pi^{-n}$  and  $\mathbf{X}_n = \pi^{-n} \sum_{k \in \mathbb{G}_n} x_k$ . □

We now adapt Lemma 2.1 of de Saporta et al (2011) to our multiple tree framework.

**Lemma B.3** *Under assumption (H.5-6), there exist a nonnegative random variable  $W$  such that for all sequences  $(x_{(1,n)}), \dots, (x_{(m,n)})$  of real numbers one has a.s.*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = x \mathbb{1}_{\bar{\mathcal{E}}} \iff \lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = x \frac{\pi}{\pi - 1} W.$$

*Proof:* We use a well known property of super-critical GW processes, see e.g. Harris (1963): for all  $j$ , there exists a non negative random variable  $W_j$  such that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_{j,n}^*|}{\pi^n} = \frac{\pi}{\pi - 1} W_j \quad \text{a.s.} \quad (4)$$

and in addition  $\{W_j > 0\} = \bar{\mathcal{E}}_j = \lim\{\{|\mathbb{G}_{j,n}^*| > 0\}\}$ . Therefore, one has

$$\lim_{n \rightarrow \infty} \sum_{j=1}^m \frac{|\mathbb{T}_{j,n}^*|}{\pi^n} = \lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n^*|}{\pi^n} = \frac{\pi}{\pi - 1} \sum_{j=1}^m W_j \quad \text{a.s.}$$

The result is obtained by setting  $W = \sum_{j=1}^m W_j$  and noticing that  $\bar{\mathcal{E}} = \cup_{j=1}^m \bar{\mathcal{E}}_j = \{\sum_{j=1}^m W_j > 0\} = \lim\{\mathbb{1}_{\{\mathbb{G}_n^* > 0\}}\}$ .  $\square$

Finally, the main result of this section is new and explains how convergence results on multiple trees can be obtained from convergence results on a single-tree. This will allow us to directly use results from de Saporta et al (2011) in all the sequel.

**Lemma B.4** *Let  $(x_{(1,n)}), \dots, (x_{(m,n)})$  be  $m$  sequences of real numbers such that for all  $1 \leq j \leq m$  one has the a.s. limit*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_{j,n}^*| > 0\}}}{|\mathbb{T}_{j,n}^*|} \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \mathbb{1}_{\bar{\mathcal{E}}_j}, \quad (5)$$

then under assumptions (H.5-6) one also has

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

*Proof:* Equations (5) and (4) yield, for all  $j$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \frac{\pi - 1}{\pi} W_j.$$

Summing over  $j$ , one obtains

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \frac{\pi - 1}{\pi} \sum_{j=1}^m W_j = \ell \frac{\pi - 1}{\pi} W.$$

Finally, we use Lemma B.3 to conclude.  $\square$

## B.2 Strong consistency for the estimators of the GW process

To prove the convergence of the  $\hat{p}_n^{(i)}(l_0, l_1)$  we first need to derive a convergence result for a sum of independent GW processes.

**Lemma B.5** *Suppose that assumptions (H.5-6) are satisfied. Then for  $i \in \{0, 1\}$  one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+i)} = z^i \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

*Proof* Remarking that  $\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+i)} = \sum_{j=1}^m \sum_{l=1}^n Z_{j,l}^i$ , the lemma is a direct consequence of Lemma B.4 and the well-known property of super-critical GW processes  $\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{l=1}^n \mathbf{Z}_{j,l} \rightarrow z \mathbb{1}_{\bar{\mathcal{E}}}$ , for all  $0 \leq j \leq m$ .  $\square$

*Proof of Theorem 4.1, first part* We give the details of the convergence of  $\hat{p}_n^{(1)}(1, 1)$  to  $p^{(1)}(1, 1)$ , the other convergences are derived similarly. The proof relies on the convergence of square integrable scalar martingales. Set

$$M_n = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j, 2k+1)} (\delta_{(j, 4k+2)} \delta_{(j, 4k+3)} - p^{(1)}(1, 1)).$$

We are going to prove that  $(M_n)$  is a martingale for a well chosen filtration. Recall that  $\mathcal{O}_{j,n} = \sigma\{\delta_{(j,k)}, k \in \mathbb{T}_n\}$ , and set  $\mathcal{O}_n = \vee_{j=1}^m \mathcal{O}_{j,n}$ . Then  $(M_n)$  is clearly a square integrable real  $(\mathcal{O}_n)$ -martingale. Using the independence assumption (H.5), its increasing process is

$$\langle M \rangle_n = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j, 2k+1)} p^{(1)}(1, 1) (1 - p^{(1)}(1, 1)) = p^{(1)}(1, 1) (1 - p^{(1)}(1, 1)) \sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^1.$$

Hence, Lemma B.5 implies that  $|\mathbb{T}_{n-1}^*|^{-1} \langle M \rangle_n$  converges almost surely on the non-extinction set  $\bar{\mathcal{E}}$ . The law of large numbers for scalar martingales thus yields that  $|\mathbb{T}_{n-1}^*|^{-1} M_n$  tends to 0 as  $n$  tends to infinity on  $\bar{\mathcal{E}}$ . Finally, notice that

$$\widehat{p}_n^{(1)}(1, 1) - p^{(1)}(1, 1) = \frac{M_n}{\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j, 2k+1)}} = \frac{M_n}{\sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^1},$$

so that Lemma B.5 again implies the almost sure convergence of  $\widehat{p}_n^{(1)}(1, 1)$  to  $p^{(1)}(1, 1)$  on the non-extinction set  $\bar{\mathcal{E}}$ .  $\square$

As a direct consequence, one obtains the a.s. convergence of  $\widehat{\pi}_n$  to  $\pi$  on  $\bar{\mathcal{E}}$ .

### B.3 Asymptotic normality for the estimators of the GW process

As  $\mathbb{P}(\bar{\mathcal{E}}) \neq 0$ , we can define a new probability  $\mathbb{P}_{\bar{\mathcal{E}}}$  by  $\mathbb{P}_{\bar{\mathcal{E}}}(A) = \mathbb{P}(A \cap \bar{\mathcal{E}})/\mathbb{P}(\bar{\mathcal{E}})$  for all event  $A$ . In all the sequel of this section, we will work on the space  $\bar{\mathcal{E}}$  under the probability  $\mathbb{P}_{\bar{\mathcal{E}}}$  and we denote by  $\mathbb{E}_{\bar{\mathcal{E}}}$  the corresponding expectation. We can now turn to the proof of the asymptotic normality of  $\widehat{\mathbf{p}}_n$ . The proof also relies on martingale theory. As the normalizing term in our central limit theorem is random, we use the central limit theorem for martingales given in Theorem 2.1.9 of Duflo (1997) that we first recall as Theorem B.6 for self-completeness.

**Theorem B.6** *Suppose that  $(\Omega, \mathcal{A}, P)$  is a probability space and that for each  $n$  we have a filtration  $\mathbb{F}_n = (\mathcal{F}_k^{(n)})$ , a stopping time  $\nu_n$  relative to  $\mathbb{F}_n$  and a real, square-integrable vector martingale  $M^{(n)} = (M_k^{(n)})_{k \geq 0}$  which is adapted to  $\mathbb{F}_n$  and has hook denoted by  $\langle M \rangle^{(n)}$ . We make the following two assumptions.*

**A.1** *For a deterministic symmetric positive semi-definite matrix  $\Gamma$*

$$\langle M \rangle_{\nu_n}^{(n)} \xrightarrow{P} \Gamma.$$

**A.2** *Lindeberg's condition holds; in other words, for all  $\epsilon > 0$ ,*

$$\sum_{k=1}^{\nu_n} \mathbb{E} \left( \|M_k^{(n)} - M_{k-1}^{(n)}\|^2 \mathbb{1}_{\{\|M_k^{(n)} - M_{k-1}^{(n)}\| > \epsilon\}} \mid \mathcal{F}_{k-1}^{(n)} \right) \xrightarrow{P} 0.$$

*Then:*

$$M_{\nu_n}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma).$$

*Proof of Theorem B.6, first part* First, set

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}^0/z^0 & 0 \\ 0 & \mathbf{V}^1/z^1 \end{pmatrix} \quad (6)$$

where for all  $i$  in  $\{0, 1\}$ ,  $\mathbf{V}^i = \mathbf{W}^i - \mathbf{p}^{(i)}(\mathbf{p}^{(i)})^t$ ,  $\mathbf{W}^i$  is a  $4 \times 4$  matrix with the entries of  $\mathbf{p}^{(i)}$  on the diagonal and 0 elsewhere. We are going to prove that  $\mathbf{V}$  is the asymptotic variance of  $\widehat{\mathbf{p}}_n - \mathbf{p}$  suitably normalized. We use Theorem B.6. We first need to define a suitable filtration. Here, we use the first cousins filtration defined as follows. Let

$$\mathcal{H}_{j,p} = \sigma\{\delta_{(j,1)}, \dots, \delta_{(j,3)}, (\delta_{(j,4k)}, \dots, \delta_{(j,4k+3)}), 1 \leq k \leq p\}$$

be the  $\sigma$ -field generated by all the 4-tuples of observed cousin cells up the granddaughters of cell  $(j, p)$  in the  $j$ -th tree and  $\mathcal{H}_p = \bigvee_{j=1}^m \mathcal{H}_{j,p}$ . Hence, the 4-tuple  $(\delta_{(j,4k)}, \dots, \delta_{(j,4k+3)})$  is  $\mathcal{H}_k$ -measurable for all  $j$ . By definition of the reproduction probabilities  $p^{(i)}(l_0, l_1)$ , the processes

$$(\delta_{(j,2k+i)}(\phi_{l_0}(\delta_{(j,2(2k+i))})\phi_{l_1}(\delta_{(j,2(2k+i)+1})) - p^{(i)}(l_0, l_1))_{k \geq 1}$$

are  $(\mathcal{H}_k)$ -martingale difference sequences. We thus introduce a sequence of  $(\mathcal{H}_k)$ -martingales  $(\mathbf{M}_p^{(n)})_{\{p \geq 1\}}$  defined for all  $n \geq 1$  and  $p \geq 1$  by

$$\mathbf{M}_p^{(n)} = |\mathbb{T}_{n-1}^*|^{-1/2} \sum_{k=1}^p \sum_{j=1}^m \mathbf{D}_{(j,k)},$$

with  $\mathbf{D}_{(j,k)} = ((\mathbf{D}_{(j,k)}^0)^t, (\mathbf{D}_{(j,k)}^1)^t)^t$  and

$$\mathbf{D}_{(j,k)}^i = \delta_{(j,2k+i)} \begin{pmatrix} \delta_{(j,2(2k+i))} \delta_{(j,2(2k+i)+1)} - p^{(i)}(1,1) \\ \delta_{(j,2(2k+i))} (1 - \delta_{(j,2(2k+i)+1)}) - p^{(i)}(1,0) \\ (1 - \delta_{(j,2(2k+i))}) \delta_{(j,2(2k+i)+1)} - p^{(i)}(0,1) \\ (1 - \delta_{(j,2(2k+i))}) (1 - \delta_{(j,2(2k+i)+1)}) - p^{(i)}(0,0) \end{pmatrix}.$$

We also introduce the sequence of stopping times  $\nu_n = |\mathbb{T}_{n-2}| = 2^{n-1} - 1$ . One has

$$\mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_{(j,k)} \mathbf{D}_{(j,k)}^t | \mathcal{H}_{k-1}] = \begin{pmatrix} \delta_{(j,2k)} \mathbf{V}^0 & 0 \\ 0 & \delta_{(j,2k+1)} \mathbf{V}^1 \end{pmatrix}.$$

Therefore the one has  $\langle \mathbf{M}^{(n)} \rangle_{\nu_n} = |\mathbb{T}_{n-1}^*|^{-1} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \begin{pmatrix} Z_{j,\ell}^0 \mathbf{V}^0 & 0 \\ 0 & Z_{j,\ell}^1 \mathbf{V}^1 \end{pmatrix}$ , so that its  $\mathbb{P}_{\bar{\mathcal{E}}}$  almost sure limit is

$$\mathbf{\Gamma}' = \begin{pmatrix} z^0 \mathbf{V}^0 & 0 \\ 0 & z^1 \mathbf{V}^1 \end{pmatrix},$$

thanks to Lemma B.5. Therefore, assumption A.1 of Theorem B.6 holds under  $\mathbb{P}_{\bar{\mathcal{E}}}$ . The Lindeberg condition A.2 is obviously satisfied as we deal with finite support distributions. We then conclude that under  $\mathbb{P}_{\bar{\mathcal{E}}}$  one has

$$|\mathbb{T}_{n-1}^*|^{-1/2} \mathbf{M}_{\nu_n}^{(n)} = |\mathbb{T}_{n-1}^*|^{-1/2} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}^*} \mathbf{D}_{(j,k)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}').$$

Using the relation

$$\hat{\mathbf{p}}_n - \mathbf{p} = \begin{pmatrix} (\sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^0) \mathbf{I}_4 & 0 \\ 0 & (\sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^1) \mathbf{I}_4 \end{pmatrix}^{-1} \mathbf{M}_{\nu_n}^{(n)},$$

Lemma B.5 and Slutsky's Lemma give the first part of Theorem 4.2.  $\square$

## B.4 Interval estimation and tests for the GW process

From the central limit theorem 4.2 one can easily build asymptotic confidence intervals for our estimators. In our context,  $Y_n$  and  $Y'_n$  being two random variables, we will say that  $[Y_n; Y'_n]$  is an asymptotic confidence interval with confidence level  $1 - \epsilon$  for the parameter  $Y$  if  $\mathbb{P}_{\bar{\mathcal{E}}}(Y_n \leq Y \leq Y'_n) \xrightarrow[n \rightarrow \infty]{} (1 - \epsilon)$ . For any  $0 \leq \epsilon \leq 1$ , let  $q_{1-\epsilon/2}$  be the  $1 - \epsilon/2$  quantile of the standard normal law.

For all  $n \geq 2$ , define the  $8 \times 8$  matrix

$$\hat{\mathbf{V}}_n = \begin{pmatrix} \hat{\mathbf{V}}_n^0 (\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k)})^{-1} & 0 \\ 0 & \hat{\mathbf{V}}_n^1 (\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k+1)})^{-1} \end{pmatrix},$$

where for all  $i$  in  $\{0, 1\}$ ,  $\hat{\mathbf{V}}_n^i = \hat{\mathbf{W}}_n^i - \hat{\mathbf{p}}_n^{(i)} (\hat{\mathbf{p}}_n^{(i)})^t$ ,  $\hat{\mathbf{W}}_n^i$  is a  $4 \times 4$  matrix with the entries of  $\hat{\mathbf{p}}_n^{(i)}$  on the diagonal and 0 elsewhere. Thus,  $|\mathbb{T}_{n-1}^*| \hat{\mathbf{V}}_n$  is an empirical estimator of the covariance matrix  $\mathbf{V}$ .

**Theorem B.7** Under assumptions (H.5-6), for  $i, l_0, l_1$  in  $\{0, 1\}$  and for any  $0 < \epsilon < 1$ , the random interval defined by

$$[\widehat{p}_n^{(i)}(l_0, l_1) - q_{1-\epsilon/2}(\widehat{\mathbf{V}}_n^{1/2})_{\ell, \ell} ; \widehat{p}_n^{(i)}(l_0, l_1) + q_{1-\epsilon/2}(\widehat{\mathbf{V}}_n^{1/2})_{\ell, \ell}]$$

is an asymptotic confidence interval with level  $1 - \epsilon$  for  $p^{(i)}(l_0, l_1)$ ; where  $(\ell, \ell)$  is the coordinate of  $\mathbf{V}_n$  corresponding to  $p^{(i)}(l_0, l_1)$ , namely  $\ell = 4(i + 1) - (2l_0 + l_1)$ .

*Proof* This is a straightforward consequence of the central limit Theorem 4.2 together with Slutsky's lemma as  $\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^*| \widehat{\mathbf{V}}_n = \mathbf{V}$   $\mathbb{P}_{\bar{\mathcal{E}}}$  a.s. thanks to Lemma B.5 and Theorem 4.1.  $\square$

Set  $\widehat{G}_n = \widehat{\mathbf{F}}_n^t \widehat{\mathbf{V}}_n \widehat{\mathbf{F}}_n$ , where  $\widehat{\mathbf{F}}_n$  is the  $8 \times 1$  vector defined by

$$\widehat{\mathbf{F}}_n = \frac{1}{2}(1, 1, 0, 0, 1, 0, 1, 0)^t + \frac{1}{2}(\widehat{T}_n^2 - 4\widehat{D}_n)^{-1/2} \widehat{\mathbf{H}}_n$$

and

$$\widehat{\mathbf{H}}_n = \begin{pmatrix} \widehat{p}_n^{(0)}(1, 1) + \widehat{p}_n^{(0)}(1, 0) + \widehat{p}_n^{(1)}(1, 1) + 2\widehat{p}_n^{(1)}(1, 0) - \widehat{p}_n^{(1)}(0, 1) \\ \widehat{p}_n^{(0)}(1, 1) + \widehat{p}_n^{(0)}(1, 0) - \widehat{p}_n^{(1)}(1, 1) - \widehat{p}_n^{(1)}(0, 1) \\ 2\widehat{p}_n^{(1)}(1, 1) + 2\widehat{p}_n^{(1)}(1, 0) \\ 0 \\ \widehat{p}_n^{(0)}(1, 1) - \widehat{p}_n^{(0)}(1, 0) + 2\widehat{p}_n^{(0)}(0, 1) + \widehat{p}_n^{(1)}(1, 1) + \widehat{p}_n^{(1)}(0, 1) \\ 2\widehat{p}_n^{(0)}(1, 1) + 2\widehat{p}_n^{(0)}(0, 1) \\ -\widehat{p}_n^{(0)}(1, 1) - \widehat{p}_n^{(0)}(1, 0) + \widehat{p}_n^{(1)}(1, 1) + \widehat{p}_n^{(1)}(0, 1) \\ 0 \end{pmatrix}.$$

**Theorem B.8** Under assumptions (H.5-6), for any  $0 < \epsilon < 1$  one has that

$$[\widehat{\pi}_n - q_{1-\epsilon/2} \widehat{G}_n^{1/2} ; \widehat{\pi}_n + q_{1-\epsilon/2} \widehat{G}_n^{1/2}]$$

is an asymptotic confidence interval with level  $1 - \epsilon$  for  $\pi$ .

*Proof* This is again a straightforward consequence of the central limit Theorem 4.2 together with Slutsky's lemma as  $\widehat{\mathbf{F}}_n$  is the gradient of the function that maps the vector  $\widehat{\mathbf{p}}$  onto the estimator  $\widehat{\pi}_n$ .  $\square$

We propose two symmetry tests for the GW process. The first one compares the average number of offspring  $m_0$  of a cell of type 0:  $m_0 = p^{(0)}(1, 0) + p^{(0)}(0, 1) + 2p^{(0)}(1, 1)$  to that of a cell of type 1:  $m_1 = p^{(1)}(1, 0) + p^{(1)}(0, 1) + 2p^{(1)}(1, 1)$ . Denote by  $\widehat{m}_n^0$  and  $\widehat{m}_n^1$  their empirical estimators. Set

- $\mathbf{H}_0^m$ :  $m_0 = m_1$  the symmetry hypothesis,
- $\mathbf{H}_1^m$ :  $m_0 \neq m_1$  the alternative hypothesis.

Let  $Y_n^m$  be the test statistic defined by

$$Y_n^m = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^m)^{-1/2} (\widehat{m}_n^0 - \widehat{m}_n^1),$$

where  $\widehat{\Delta}_n^m = |\mathbb{T}_{n-1}^*| \mathbf{d}\mathbf{g}_m^t \widehat{\mathbf{V}}_n \mathbf{d}\mathbf{g}_m$  and  $\mathbf{d}\mathbf{g}_m = (2, 1, 1, 0, -2 - 1, -1, 0)^t$ . This test statistic has the following asymptotic properties.

**Theorem B.9** Under assumptions (H.5-6) and the null hypothesis  $\mathbf{H}_0^m$ , one has

$$(Y_n^m)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

on  $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$ ; and under the alternative hypothesis  $\mathbf{H}_1^m$ , almost surely on  $\bar{\mathcal{E}}$  one has

$$\lim_{n \rightarrow \infty} (Y_n^m)^2 = +\infty.$$

*Proof* Let  $g_m$  be the function defined from  $\mathbb{R}^8$  onto  $\mathbb{R}$  by  $g_m(x_1, \dots, x_8) = 2x_1 + x_2 + x_3 + 2x_5 - x_6 - x_7$  so that  $\mathbf{d}g_m$  is the gradient of  $g_m$ . Thus, the central limit Theorem 4.2 yields

$$\sqrt{|\mathbb{T}_{n-1}^*|} (g_m(\widehat{\mathbf{p}}_n) - g_m(\mathbf{p})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{d}g_m^t \mathbf{V} \mathbf{d}g_m) = \mathcal{N}(0, \Delta^m)$$

on  $(\overline{\mathcal{E}}, \mathbb{P}_{\overline{\mathcal{E}}})$ . Under the null hypothesis  $\mathbf{H}_0^m$ ,  $g_m(\mathbf{p}) = 0$ , so that one has

$$|\mathbb{T}_{n-1}^*| (\Delta^m)^{-1} g_m(\widehat{\mathbf{p}}_n)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

on  $(\overline{\mathcal{E}}, \mathbb{P}_{\overline{\mathcal{E}}})$ . Lemma B.5 and Theorem 4.1 give the almost sure convergence of  $\widehat{\Delta}_n^m$  to  $\Delta^m$ . Hence Slutsky's Lemma yields the expected result. Under the alternative hypothesis  $\mathbf{H}_1^m$ , one has

$$Y_n^m = (\widehat{\Delta}_n^m)^{-1/2} (\sqrt{|\mathbb{T}_{n-1}^*|} (g_m(\widehat{\mathbf{p}}_n) - g_m(\mathbf{p})) + \sqrt{|\mathbb{T}_{n-1}^*|} g_m(\mathbf{p})).$$

The first term converges to a centered normal law and the second term tends to infinity as  $|\mathbb{T}_{n-1}^*|$  tends to infinity a.s. on  $(\overline{\mathcal{E}}, \mathbb{P}_{\overline{\mathcal{E}}})$ .  $\square$

Our next test compares the reproduction probability vectors of mother cells of type 0 and 1.

- $\mathbf{H}_0^p$ :  $\mathbf{p}^{(0)} = \mathbf{p}^{(1)}$  the symmetry hypothesis,
- $\mathbf{H}_1^p$ :  $\mathbf{p}^{(0)} \neq \mathbf{p}^{(1)}$  the alternative hypothesis.

Let  $(\mathbf{Y}_n^p)^t \mathbf{Y}_n^p$  be the test statistic defined by

$$\mathbf{Y}_n^p = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^p)^{-1/2} (\widehat{\mathbf{p}}^{(0)} - \widehat{\mathbf{p}}^{(1)}),$$

where  $\widehat{\Delta}_n^p = |\mathbb{T}_{n-1}^*| \mathbf{d}g_p^t \widehat{\mathbf{V}}_n \mathbf{d}g_p$  and  $\mathbf{d}g_p = \begin{pmatrix} \mathbf{I}_4 \\ -\mathbf{I}_4 \end{pmatrix}$ . This test statistic has the following asymptotic properties.

**Theorem B.10** *Under assumptions (H.5-6) and the null hypothesis  $\mathbf{H}_0^p$ , one has*

$$(\mathbf{Y}_n^p)^t \mathbf{Y}_n^p \xrightarrow{\mathcal{L}} \chi^2(4)$$

on  $(\overline{\mathcal{E}}, \mathbb{P}_{\overline{\mathcal{E}}})$ ; and under the alternative hypothesis  $\mathbf{H}_1^p$ , almost surely on  $\overline{\mathcal{E}}$  one has

$$\lim_{n \rightarrow \infty} \|\mathbf{Y}_n^p\|^2 = +\infty.$$

*Proof* We mimic the proof of Theorem B.9 with  $g_p$  the function defined from  $\mathbb{R}^8$  onto  $\mathbb{R}^4$  by  $g_p(x_1, \dots, x_8) = (x_1 - x_5, x_2 - x_6, x_3 - x_7, x_4 - x_8)^t$ , so that  $\mathbf{d}g_p$  is the gradient of  $g_p$ .  $\square$

## C Convergence of estimators for the BAR process

We now prove the convergence of the estimators for the BAR process, that is the parts of Theorems 4.1 and 4.2 concerning  $\widehat{\theta}_n$ ,  $\widehat{\sigma}_{n,i}^2$  and  $\widehat{\rho}_n$ , together with additional technical results, especially the convergence of higher moment estimators required to estimate the asymptotic variances.

### C.1 Preliminary results: laws of large numbers

In this section, we want to study the asymptotic behavior of various sums of observed data. Most of the results are directly taken from de Saporta et al (2011). All external references in this section refer to that paper that will not be cited each time. However, we need additional results concerning higher moments of the BAR process in order to obtain the consistency of  $\widehat{\tau}_{i,n}^4$  and  $\widehat{\nu}_n^2$ , as there is no such result in de Saporta et al (2011). We also give all the explicit formulas so that the interested reader can actually compute the various asymptotic variances.

Again, our work relies on the strong law of large numbers for square integrable martingales. To ensure that the increasing processes of our martingales are at most  $\mathcal{O}(\pi^n)$  we first need the following lemma.

**Lemma C.1** Under assumptions (H.0-6), for all  $i \in \{0, 1\}$  one has

$$\sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^8 = \mathcal{O}(\pi^n) \quad a.s.$$

*Proof* The proof follows the same lines as that of Lemma 6.1. The constants before the terms  $A_n^i$ ,  $B_n^i$  and  $C_n^i$  therein are replaced respectively by  $(4/(1-\beta))^7$ ,  $\alpha^8(4/(1-\beta))^7$  and  $2^8$ ; in the term  $A_n^i$ ,  $\varepsilon^2$  is replaced by  $\varepsilon^8$ ; in the term  $C_n^i$ ,  $\beta^{2rk}$  is replaced by  $\beta^{8rk}$ ; the term  $B_n^i$  is unchanged. In the expression of  $\mathbb{E}[(Y_{\ell,p}^i)^2]$ , one just needs to replace  $\tau^4$  by  $\mu_i^{16}$ ,  $\sigma^4$  by  $\gamma_i^{16}$  and  $\nu^2\tau^4$  by  $\eta^8$ . Note that the various moments of the noise sequence are defined in assumption (H.1). The rest of the proof is unchanged.  $\square$

We also state some laws of large numbers for the noise processes.

**Lemma C.2** Under assumptions (H.0-6), for all  $i \in \{0, 1\}$  and for all integers  $0 \leq q \leq 4$ , one has

$$\frac{1}{\pi^n} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+i)} \varepsilon_{(j, 2k+i)}^q = \frac{\pi}{\pi-1} W z^i \mathbb{E}[\varepsilon_{(1, 2+i)}^q] \quad a.s.$$

*Proof* This is also a direct consequence of de Saporta et al (2011) thanks to Lemmas B.3 and B.4. Lemma 5.3 provides the result for  $q = 0$ , Lemma 5.5 for  $q = 1$ , Corollary 5.6 for  $q = 2$  and Lemma 5.7 for  $q = 4$ . The result for  $q = 3$  is obtained similarly.  $\square$

In view of these new stronger results, we can now state our first laws of large numbers for the observed BAR process. For  $i \in \{0, 1\}$  and all integers  $1 \leq q \leq 4$  let us now define

$$\begin{aligned} H_n^i(q) &= \sum_{j=1}^m H_{j,n}^i(q) = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^q, \\ H_n^{01}(q) &= \sum_{j=1}^m H_{j,n}^{01}(q) = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k)} \delta_{(j, 2k+1)} X_{(j,k)}^q, \end{aligned}$$

and  $\mathbf{H}_n(q) = (H_n^0(q), H_n^1(q))^t$ .

**Lemma C.3** Under assumptions (H.0-6) and for all integers  $1 \leq q \leq 4$ , one has the following a.s. limits on the non-extinction set  $\bar{\mathcal{E}}$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \mathbf{H}_n(q) &= \mathbf{h}(q) = (\mathbf{I}_2 - \tilde{\mathbf{P}}_q)^{-1} \mathbf{P}^t \tilde{\mathbf{h}}(q), \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} H_n^{01}(q) &= h^{01}(q) = p^{(0)}(1, 1) \left( \tilde{h}^0(q) + b_0^q \frac{h^0(q)}{\pi} \right) \\ &\quad + p^{(1)}(1, 1) \left( \tilde{h}^1(q) + b_1^q \frac{h^1(q)}{\pi} \right), \end{aligned}$$

where

$$\tilde{\mathbf{P}}_q = \pi^{-1} \mathbf{P}^t \begin{pmatrix} b_0^q & 0 \\ 0 & b_1^q \end{pmatrix}, \quad \mathbf{h}(q) = \begin{pmatrix} h^0(q) \\ h^1(q) \end{pmatrix}, \quad \tilde{\mathbf{h}}(q) = \begin{pmatrix} \tilde{h}^0(q) \\ \tilde{h}^1(q) \end{pmatrix},$$

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

$$\begin{aligned} \tilde{h}^i(1) &= a_i z^i, \\ \tilde{h}^i(2) &= (a_i^2 + \sigma_i^2) z^i + 2a_i b_i h^i(1) \pi^{-1}, \\ \tilde{h}^i(3) &= (a_i^3 + 3a_i \sigma_i^2 + \lambda_i) z^i + 3b_i (a_i^2 + \sigma_i^2) h^i(1) \pi^{-1} + 3a_i b_i^2 h^i(2) \pi^{-1}, \\ \tilde{h}^i(4) &= (a_i^4 + 6a_i^2 \sigma_i^2 + 4a_i \lambda_i + \tau_i^4) z^i + 4b_i (a_i^3 + 3a_i \sigma_i^2 + \lambda_i) h^i(1) \pi^{-1} \\ &\quad + 6b_i^2 (a_i^2 + \sigma_i^2) h^i(2) \pi^{-1} + 4a_i b_i^3 h^i(3) \pi^{-1}. \end{aligned}$$

*Proof* The results for  $q = 1$  and  $q = 2$  come from Propositions 6.3, 6.5 and 6.6 together with Lemma B.4. The proofs for  $q \geq 3$  follow the same lines, using Lemma C.2 when required and Lemma C.1 to bound the increasing processes of the various martingales at stake.  $\square$

To prove the consistency of our estimators, we also need some additional families of laws of large numbers.

**Lemma C.4** *Under assumptions (H.0-6), for  $i \in \{0, 1\}$  and for all integers  $1 \leq p + q \leq 4$ , one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p \varepsilon_{(j, 2k+i)}^q = \mathbb{E}[\varepsilon_{2+i}^q] h^i(p) \mathbb{1}_{\bar{\mathcal{E}}}.$$

*Proof* The proof is similar to that of Theorem 4.1. For all  $1 \leq j \leq m$ , one has

$$\begin{aligned} & \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p \varepsilon_{(j, 2k+i)}^q \\ &= \sum_{\ell=0}^n \sum_{k \in \mathbb{G}_\ell} \delta_{(j, 2k+i)} X_{(j,k)}^p (\varepsilon_{(j, 2k+i)}^q - \mathbb{E}[\varepsilon_{(j, 2k+i)}^q \mid \mathcal{F}_{j,\ell}^{\mathcal{O}}]) + \mathbb{E}[\varepsilon_{2+i}^q] \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p, \end{aligned}$$

as the conditional moment of  $\varepsilon_{2k+i}$  are constants by assumption (H.1). The first term is a square integrable  $(\mathcal{F}_{j,n}^{\mathcal{O}})$ -martingale and its increasing process is  $\mathcal{O}(\pi^n)$  thanks to Lemma C.1, thus the first term is  $o(\pi^n)$ . The limit of the second term is given by Lemma C.3.  $\square$

**Lemma C.5** *Under assumptions (H.0-6), for  $i \in \{0, 1\}$  and for all integers  $1 \leq q \leq 4$ , one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j, 2k+i)}^q = (\pi \tilde{h}^i(q) + b_i^q h^i(q)) \mathbb{1}_{\bar{\mathcal{E}}}.$$

*Proof* The proof is obtained by replacing  $X_{(j, 2k+i)}$  by  $a_i + b_i X_k + \varepsilon_{2k+i}$ . One then develops the exponent and uses Lemmas B.5, C.2, C.3 and C.4 to conclude.  $\square$

**Lemma C.6** *Under assumptions (H.0-6), for  $i \in \{0, 1\}$  and for all integers  $1 \leq p + q \leq 4$ , one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p X_{(j, 2k+i)}^q = h^i(p, q) \mathbb{1}_{\bar{\mathcal{E}}},$$

with

$$\begin{aligned} h^i(p, 1) &= a_i h^i(p) + b_i h^i(p+1), \\ h^i(p, 2) &= (a_i^2 + \sigma_i^2) h^i(p) + 2a_i b_i h^i(p+1) + b_i^2 h^i(p+2), \\ h^i(p, 3) &= (a_i^3 + 3a_i \sigma_i^2 + \lambda_i) h^i(p) + 3b_i (a_i^2 + \sigma_i^2) h^i(p+1) + 3a_i b_i^2 h^i(p+2) + b_i^3 h^i(p+3), \end{aligned}$$

where we used the convention  $h^i(0) = z^i \pi$ .

*Proof* As above, the proof is obtained by replacing  $X_{(j, 2k+i)}$  and developing the exponents. Then one uses Lemmas B.5, C.2, C.3 and C.4 to compute the limits.  $\square$

**Lemma C.7** *Under assumptions (H.5-6), one has the following a.s. limit*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2(2k+i))} \delta_{(j, 2(2k+i)+1)} = p^{(i)}(1, 1) z^i \pi \mathbb{1}_{\bar{\mathcal{E}}}.$$

*Proof* First note that  $\delta_{(j,2(2k+i))}\delta_{(j,2(2k+i)+1)} = \delta_{(j,2k+i)}\delta_{(j,2(2k+i))}\delta_{(j,2(2k+i)+1)}$ . The proof is then similar to that of Theorem 4.1. One adds and subtract  $p^{(i)}(1,1)$  so that a martingale similar to  $(M_n)$  naturally appears. The limit of the remaining term is given by Lemma B.5.  $\square$

**Lemma C.8** *Under assumptions (H.0-6), for all integers  $0 \leq p+q+r \leq 4$ , one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j,2k)}\delta_{(j,2k+1)}X_{(j,k)}^p \varepsilon_{(j,2k)}^q \varepsilon_{(j,2k+1)}^r = \mathbb{E}[\varepsilon_2^q \varepsilon_3^r] h^{01}(p) \mathbb{1}_{\bar{\mathcal{E}}},$$

where we used the convention  $h^{01}(0) = p^{(0)}(1,1)z^0 + p^{(1)}(1,1)z^1$ .

*Proof* The proof is similar to Lemma C.4, one adds and subtracts the constant  $\mathbb{E}[\varepsilon_{(j,2k)}^q \varepsilon_{(j,2k+1)}^r \mid \mathcal{F}_{j,\ell}^{\mathcal{O}}]$ .  $\square$

**Lemma C.9** *Under assumptions (H.0-6), for all integers  $1 \leq p+q+r \leq 4$ , one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j,2k)}\delta_{(j,2k+1)}X_{(j,k)}^p X_{(j,2k)}^q X_{(j,2k+1)}^r = h^{01}(p, q, r) \mathbb{1}_{\bar{\mathcal{E}}},$$

with

$$\begin{aligned} h^{01}(p, 1, 0) &= a_0 h^{01}(p) + b_0 h^{01}(p+1), & h^{01}(p, 0, 1) &= a_1 h^{01}(p) + b_1 h^{01}(p+1), \\ h^{01}(p, 2, 0) &= (a_0^2 + \sigma_0^2) h^{01}(p) + 2a_0 b_0 h^{01}(p+1) + b_0^2 h^{01}(p+2), \\ h^{01}(p, 0, 2) &= (a_1^2 + \sigma_1^2) h^{01}(p) + 2a_1 b_1 h^{01}(p+1) + b_1^2 h^{01}(p+2), \\ h^{01}(p, 3, 0) &= (a_0^3 + 3a_0 \sigma_0^2 + \lambda_0) h^{01}(p) + 3b_0 (a_0^2 + \sigma_0^2) h^{01}(p+1) + 3a_0 b_0^2 h^{01}(p+2) + b_0^3 h^{01}(p+3), \\ h^{01}(p, 0, 3) &= (a_1^3 + 3a_1 \sigma_1^2 + \lambda_1) h^{01}(p) + 3b_1 (a_1^2 + \sigma_1^2) h^{01}(p+1) + 3a_1 b_1^2 h^{01}(p+2) + b_1^3 h^{01}(p+3), \\ h^{01}(p, 1, 1) &= (a_0 a_1 + \rho) h^{01}(p) + (a_0 b_1 + b_0 a_1) h^{01}(p+1) + b_0 b_1 h^{01}(p+2), \\ h^{01}(p, 2, 1) &= ((a_0^2 + \sigma_0^2) a_1 + 2a_0 \rho + \alpha) h^{01}(p) + ((a_0^2 + \sigma_0^2) b_1 + 2(a_0 a_1 + \rho) b_0) h^{01}(p+1) \\ &\quad + b_0 (2a_0 b_1 + b_0 a_1) h^{01}(p+2) + b_0^2 b_1 h^{01}(p+3), \\ h^{01}(p, 1, 2) &= ((a_1^2 + \sigma_1^2) a_0 + 2a_1 \rho + \beta) h^{01}(p) + ((a_1^2 + \sigma_1^2) b_0 + 2(a_0 a_1 + \rho) b_1) h^{01}(p+1) \\ &\quad + b_1 (a_0 b_1 + 2b_0 a_1) h^{01}(p+2) + b_0 b_1^2 h^{01}(p+3), \\ h^{01}(0, 2, 2) &= (a_0^2 a_1^2 + a_0^2 \sigma_1^2 + a_1^2 \sigma_0^2 + \nu^2 + 2a_0 \beta + 2a_1 \alpha + 4a_0 a_1 \rho) h^{01}(0) \\ &\quad + 2(b_0 (a_0 (a_1^2 + \sigma_1^2) + \beta + 2a_1 \rho) + b_1 (a_1 (a_0^2 + \sigma_0^2) + \alpha + 2a_0 \rho)) h^{01}(1) \\ &\quad + (b_0^2 (a_1^2 + \sigma_1^2) + b_1^2 (a_0^2 + \sigma_0^2) + 4b_0 b_1 (a_0 a_1 + \rho)) h^{01}(2) \\ &\quad + 2b_0 b_1 (a_0 b_1 + b_0 a_1) h^{01}(3) + b_0^2 b_1^2 h^{01}(4). \end{aligned}$$

*Proof* The proof is obtained by replacing  $X_{(j,2k+i)}$  by  $a_i + b_i X_k + \varepsilon_{2k+i}$  and developing the exponents. One uses Lemmas C.3 and C.8 to compute the limits.  $\square$

To conclude this section, we prove the convergence of the normalizing matrices  $\mathbf{S}_n^0$ ,  $\mathbf{S}_n^1$  and  $\mathbf{S}_n^{01}$  where

$$\mathbf{S}_n^{01} = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j,2k)}\delta_{(j,2k+1)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix},$$

with the sum taken over all observed cells that have observed daughters of both types.

**Lemma C.10** *Suppose that assumptions (H.0-6) are satisfied. Then, there exist definite positive matrices  $\mathbf{L}^0$ ,  $\mathbf{L}^1$  and  $\mathbf{L}^{01}$  such that for  $i \in \{0, 1\}$  one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1} \mathbf{S}_n^i = \mathbb{1}_{\bar{\mathcal{E}}} \mathbf{L}^i, \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1} \mathbf{S}_n^{01} = \mathbb{1}_{\bar{\mathcal{E}}} \mathbf{L}^{01} \quad a.s.$$

where

$$\mathbf{L}^i = \begin{pmatrix} h^i(0) & h^i(1) \\ h^i(1) & h^i(2) \end{pmatrix}, \quad \mathbf{L}^{01} = \begin{pmatrix} h^{01}(0) & h^{01}(1) \\ h^{01}(1) & h^{01}(2) \end{pmatrix}.$$

*Proof* This is a direct consequence of Lemmas B.5 and C.3.  $\square$

## C.2 Strong consistency for the estimators of the BAR process

We could obtain the convergences of our estimators by sharp martingales results as in de Saporta et al (2011), see also B.2. However, we chose the direct approach here. Indeed, our convergences are now direct consequences of the laws of large numbers given in C.1.

*Proof of Theorem 4.1, convergence of  $\hat{\theta}_n$*  This is a direct consequence of Lemmas C.10 and C.6. Indeed, by Lemma C.6 one has

$$\frac{\mathbb{1}_{\{|\mathbb{G}_{n-1}^*|>0\}}}{|\mathbb{T}_{n-1}^*|} \Sigma_{n-1} \hat{\theta}_n = \frac{\mathbb{1}_{\{|\mathbb{G}_{n-1}^*|>0\}}}{|\mathbb{T}_{n-1}^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{(j,2k)} X_{(j,2k)} \\ \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)} \\ \delta_{(j,2k+1)} X_{(j,2k+1)} \\ \delta_{(j,2k+1)} X_{(j,k)} X_{(j,2k+1)} \end{pmatrix} \xrightarrow{n \rightarrow \infty} \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix} \theta \mathbb{1}_{\bar{\mathcal{E}}}.$$

And one concludes using Lemma C.10.  $\square$

*Proof of Theorem 4.1, convergence of  $\hat{\sigma}_{i,n}^2$  and  $\hat{\rho}_n$*  This result is not as direct as the preceding one because of the presence of the  $\hat{\varepsilon}_k$  in the various estimators. Take for instance the estimator  $\hat{\sigma}_{i,n}^2$ . For all  $1 \leq j \leq m$ , one has

$$\begin{aligned} \sum_{k \in \mathbb{T}_{n-1}} \hat{\varepsilon}_{(j,2k+i)}^2 &= \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} (X_{(j,2k+i)} - \hat{a}_{i,\ell} - \hat{b}_{i,\ell} X_{(j,k)})^2 \\ &= \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j,2k+i)} X_{(j,2k+i)}^2 + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \hat{a}_{i,\ell}^2 \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} \\ &\quad + 2 \sum_{\ell=0}^{n-1} \hat{a}_{i,\ell} \hat{b}_{i,\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,k)} + \sum_{\ell=0}^{n-1} \hat{b}_{i,\ell}^2 \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,k)}^2 \\ &\quad - 2 \sum_{\ell=0}^{n-1} \hat{a}_{i,\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,2k+i)} - 2 \sum_{\ell=0}^{n-1} \hat{b}_{i,\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,k)} X_{(j,2k+i)}. \end{aligned}$$

Let us study the limit of the last term. One has

$$\frac{1}{\pi^n} \sum_{\ell=0}^{n-1} \hat{b}_{i,\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,k)} X_{(j,2k+i)} = \frac{1}{\pi} \sum_{\ell=0}^{n-1} \frac{1}{\pi^{n-1-\ell}} \left( \hat{b}_{i,\ell} \frac{1}{\pi^\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,k)} X_{(j,2k+i)} \right).$$

We now use Lemma B.1 with  $\mathbf{A}_n = \pi^{-n}$  and  $\mathbf{X}_n = \hat{b}_{i,n} \pi^{-n} \sum_{k \in \mathbb{G}_n} \delta_{(j,2k+i)} X_{(j,k)} X_{(j,2k+i)}$ . We know from Lemma C.6 together with Lemma B.2 that  $\pi^{-n} \sum_{k \in \mathbb{G}_n} \delta_{(j,2k+i)} X_{(j,k)} X_{(j,2k+i)}$  converges to  $h^0(1,1)W_j$ , and the previous proof gives the convergence of  $\hat{b}_{i,n}$ . Thus, one obtains

$$\frac{1}{\pi^n} \sum_{\ell=0}^{n-1} \hat{b}_{i,\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k+i)} X_{(j,k)} X_{(j,2k+i)} \xrightarrow{n \rightarrow \infty} \frac{\pi^2}{\pi-1} W_j b h^0(1,1).$$

We deal with the other terms in the decomposition of the sum of  $\hat{\varepsilon}_{2k}^2$  in a similar way, using either Lemma C.3, C.5 or C.6. Finally, one obtains the almost sure limit on  $\bar{\mathcal{E}}$

$$\begin{aligned} \hat{\sigma}_{i,n}^2 &\xrightarrow{n \rightarrow \infty} (\tilde{h}^i(2) + b_i^2 h^i(2) \pi^{-1} + a_i^2 z^i + 2a_i b_i h^i(1) \pi^{-1} - 2a_i (\tilde{h}^i(1) + b_i^2 h^i(1) \pi^{-1})) (z^i)^{-1} \\ &= \sigma_i^2. \end{aligned}$$

To obtain the convergence of  $\hat{\rho}_n$  the approach is similar, using the convergence results given in Lemmas C.3, C.7, C.8 and C.9.  $\square$

**Theorem C.11** *Under assumptions (H.0-6),  $\hat{\tau}_{i,n}^4$  and  $\hat{\nu}_n^2$  converge almost surely to  $\tau_i^4$  and  $\nu^2$  respectively on  $\bar{\mathcal{E}}$ .*

*Proof* We work exactly along the same lines as the previous proof with higher powers.  $\square$

### C.3 Asymptotic normality for the estimators of the BAR process

We first give the asymptotic normality for  $\widehat{\theta}_n$ .

*Proof of Theorem 4.2 for  $\widehat{\theta}_n$*  Define the  $4 \times 4$  matrices

$$\Sigma = \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \sigma_0^2 \mathbf{L}^0 & \rho \mathbf{L}^{01} \\ \rho \mathbf{L}^{01} & \sigma_1^2 \mathbf{L}^1 \end{pmatrix}, \quad \Gamma_\theta = \Sigma^{-1} \Gamma \Sigma^{-1}. \quad (7)$$

We now follow the same lines as the proof of the first part of Theorem 4.2 with a different filtration. This time we use the observed sister pair-wise filtration defined as follows. For  $0 \leq j \leq m$  and  $p \geq 0$ , let

$$\mathcal{G}_{j,p}^\mathcal{O} = \mathcal{O}_j \vee \sigma\{\delta_{(j,1)} X_{(j,1)}, (\delta_{(j,2k)} X_{(j,2k)}, \delta_{(j,2k+1)} X_{(j,2k+1)}), 1 \leq k \leq p\} \quad (8)$$

be the  $\sigma$ -field generated by the  $j$ -th GW tree and all the pairs of observed sister cells in genealogy  $j$  up to the daughters of cell  $(j, p)$ , and let  $\mathcal{G}_p^\mathcal{O} = \bigvee_{j=1}^m \mathcal{G}_{j,p}^\mathcal{O}$  be the  $\sigma$ -field generated by the union of all  $\mathcal{G}_{j,p}^\mathcal{O}$  for  $1 \leq j \leq m$ . Hence, for instance,  $(\delta_{(j,2k)} \varepsilon_{(j,2k)}, \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)})$  is  $\mathcal{G}_k^\mathcal{O}$ -measurable for all  $j$ . In addition, assumptions (H.1) and (H.4-5) imply that the process

$$(\delta_{(j,2k)} \varepsilon_{(j,2k)}, X_{(j,k)} \delta_{(j,2k)} \varepsilon_{(j,2k)}, \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)}, X_{(j,k)} \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)})^t$$

is a  $(\mathcal{G}_k^\mathcal{O})$ -martingale difference sequence. Indeed, as the non-extinction set  $\bar{\mathcal{E}}$  is in  $\mathcal{G}_k^\mathcal{O}$  for every  $k \geq 1$ , it is first easy to prove that  $\mathbb{E}_{\bar{\mathcal{E}}}[\delta_{(j,2k)} \varepsilon_{(j,2k)} | \mathcal{G}_{k-1}^\mathcal{O}] = \mathbb{E}[\delta_{(j,2k)} \varepsilon_{(j,2k)} | \mathcal{G}_{k-1}^\mathcal{O}]$ . Then, for  $k \in \mathbb{G}_n$ , using repeatedly the independence properties, one has

$$\begin{aligned} & \mathbb{E}[\delta_{(j,2k)} \varepsilon_{(j,2k)} | \mathcal{G}_{k-1}^\mathcal{O}] \\ &= \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{O} \vee \mathcal{F}_n \vee \sigma(\varepsilon_{j,p}, 1 \leq j \leq m, p \in \mathbb{G}_{n+1}, p \leq 2k-1)] | \mathcal{G}_{k-1}^\mathcal{O}] \\ &= \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{F}_n \vee \sigma(\varepsilon_{j,p}, 1 \leq j \leq m, p \in \mathbb{G}_{n+1}, p \leq 2k-1)] | \mathcal{G}_{k-1}^\mathcal{O}] \\ &= \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{F}_n] | \mathcal{G}_{k-1}^\mathcal{O}] = \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{F}_{j,n}] | \mathcal{G}_{k-1}^\mathcal{O}] = 0. \end{aligned}$$

We introduce a sequence of  $(\mathcal{G}_k^\mathcal{O})$ -martingales  $(\mathbf{M}_p^{(n)})_{\{p \geq 1\}}$  defined for all  $n, p \geq 1$  by  $\mathbf{M}_p^{(n)} = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^p \mathbf{D}_k$ , with

$$\mathbf{D}_k = \sum_{j=1}^m \mathbf{D}_{(j,k)} = \sum_{j=1}^m \begin{pmatrix} \delta_{(j,2k)} \varepsilon_{(j,2k)} \\ X_{(j,k)} \delta_{(j,2k)} \varepsilon_{(j,2k)} \\ \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)} \\ X_{(j,k)} \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)} \end{pmatrix}.$$

We also introduce the sequence of stopping times  $\nu_n = |\mathbb{T}_n| = 2^{n+1} - 1$ . We are interested in the convergence of the process  $\mathbf{M}_{\nu_n}^{(n)} = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^{|\mathbb{T}_n|} \mathbf{D}_k$ . Again, it is easy to prove that

$$\mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^\mathcal{O}] = \mathbb{E}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^\mathcal{O}] = \sum_{j=1}^m \begin{pmatrix} \sigma_0^2 \varphi_{(j,k)}^0 & \rho \varphi_{(j,k)}^{01} \\ \rho \varphi_{(j,k)}^{01} & \sigma_1^2 \varphi_{(j,k)}^1 \end{pmatrix},$$

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

$$\varphi_{(j,k)}^i = \delta_{(j,2k+i)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix}, \quad \varphi_{(j,k)}^{01} = \delta_{(j,2k)} \delta_{(j,2k+1)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix}.$$

Lemma C.10 yields that the  $\mathbb{P}_{\bar{\mathcal{E}}}$  almost sure limit of the process  $\langle \mathbf{M}^{(n)} \rangle_{\nu_n} = |\mathbb{T}_n^*|^{-1} \sum_{k \in \mathbb{T}_n} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^\mathcal{O}]$  is  $\Gamma$ , as

$$\sum_{k \in \mathbb{T}_n} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^\mathcal{O}] = \begin{pmatrix} \sigma_0^2 \mathbf{S}_n^0 & \rho \mathbf{S}_n^{01} \\ \rho \mathbf{S}_n^{01} & \sigma_1^2 \mathbf{S}_n^1 \end{pmatrix}.$$

Therefore, the assumption A.1 of Theorem B.6 holds under  $\mathbb{P}_{\bar{\mathcal{E}}}$ . Thanks to assumptions (H.1) and (H.4-5) we can easily prove that for some  $r > 2$ , one has  $\sup_{k \geq 0} \mathbb{E}[\|\mathbf{D}_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty$  a.s. which in turn implies the Lindeberg condition A.2. We can now conclude that under  $\mathbb{P}_{\bar{\mathcal{E}}}$  one has

$$|\mathbb{T}_{n-1}^*|^{-1/2} \sum_{k \in \mathbb{T}_{n-1}^*} \mathbf{D}_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma).$$

Finally Eq. (2) implies that  $\sum_{k \in \mathbb{T}_{n-1}^*} \mathbf{D}_k = \Sigma_{n-1}(\hat{\theta}_n - \theta)$ . Therefore, the result is a direct consequence of Lemma C.10 together with Slutsky's Lemma.  $\square$

We now turn to the asymptotic normality of  $\hat{\sigma}_{i,n}^2$  and  $\hat{\rho}_n$ . The direct application of the central limit theorem for martingales to  $\hat{\sigma}_{i,n}^2$  and  $\hat{\rho}_n$  is not obvious because of the  $\hat{\varepsilon}_{(j,2k+i)}$ . We proceed along the same lines as in the proof of the convergence of  $\hat{\sigma}_{i,n}^2$ , using the decomposition along the generations. However, this time we need a convergence rate for  $\hat{\theta}_n$  in order to apply Lemma B.1.

**Theorem C.12** *Under assumptions (H.0-6), one has*

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

*Proof* : This result is based on the asymptotic behavior of the martingale  $(\mathbf{M}_n)$  defined as follows

$$\mathbf{M}_n = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{(j,2k)} \varepsilon_{2j,k}, \\ \delta_{(j,2k)} X_{(j,k)} \varepsilon_{(j,2k)}, \\ \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)}, \\ \delta_{(j,2k+1)} X_{(j,k)} \varepsilon_{(j,2k+1)} \end{pmatrix}.$$

For all  $n \geq 2$ , we readily deduce from the definitions of the BAR process and of our estimator  $\hat{\theta}_n$  that

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{(j,2k)} \varepsilon_{(j,2k)} \\ \delta_{(j,2k)} X_{(j,k)} \varepsilon_{(j,2k)} \\ \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)} \\ \delta_{(j,2k+1)} X_{(j,k)} \varepsilon_{(j,2k+1)} \end{pmatrix} = \Sigma_{n-1}^{-1} \mathbf{M}_n.$$

The sharp asymptotic behavior of  $(\mathbf{M}_n)$  relies on properties of vector martingales. Thanks to Lemma B.4, the proof follows exactly the same lines as that of the first part of Theorem 3.2 of de Saporta et al (2011) and is not repeated here.  $\square$

We can now turn to the end of the proof of Theorem 4.2 concerning the asymptotic normality of  $\hat{\sigma}_{i,n}^2$  and  $\hat{\rho}_n$ .

*Proof of Theorem 4.2, asymptotic normality of  $\hat{\sigma}_{i,n}^2$*  Thanks to Eq. (1) and (3), we decompose  $\hat{\sigma}_{i,n}^2 - \sigma_i^2$  into two parts  $U_n^i$  and  $V_n^i$

$$\begin{aligned} |\mathbb{T}_{n-1}^{*i}|(\hat{\sigma}_{i,n}^2 - \sigma_i^2) &= \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \hat{\varepsilon}_{(j,2k+i)}^2 - \varepsilon_{(j,2k+i)}^2 + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k+i)} (\varepsilon_{(j,2k+i)}^2 - \sigma_i^2) \\ &= \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} u_{(j,k)}^i + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} v_{(j,k)}^i = U_n^i + V_n^i, \end{aligned}$$

with

$$\begin{aligned} u_{(j,k)}^i &= \delta_{(j,2k+i)} ((a_i - \hat{a}_{i,\ell})^2 + (b_i - \hat{b}_{i,\ell})^2 X_{(j,k)}^2) + 2(a_i - \hat{a}_{i,\ell})(b_i - \hat{b}_{i,\ell}) X_{(j,k)} \\ v_{(j,k)}^i &= \delta_{(j,2k+i)} (2((a_i - \hat{a}_{i,\ell}) + (b_i - \hat{b}_{i,\ell}) X_{(j,k)}) \varepsilon_{(j,2k+i)} + \varepsilon_{(j,2k+i)}^2 - \sigma_i^2). \end{aligned}$$

We first deal with  $U_n^i$  and study the limit of  $\pi^{-n/2}U_n^i$ . Let us just detail the first term

$$\begin{aligned} \frac{1}{\pi^{n/2}} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k+i)} (a_i - \widehat{a}_{i,\ell})^2 &= \sum_{\ell=0}^{n-1} \pi^{(\ell-n)/2} \frac{\ell}{\pi^{\ell/2}} \frac{(a_i - \widehat{a}_{i,\ell})^2}{\ell \pi^{-\ell}} \left( \frac{1}{\pi^\ell} \sum_{j=1}^m \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k+i)} \right) \\ &= \sum_{\ell=0}^{n-1} \pi^{(\ell-n)/2} x_{i,\ell}. \end{aligned}$$

On the one hand, Lemmas B.5, B.3 and B.2 imply that  $\pi^{-\ell} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k+i)}$  converges a.s. to a finite limit. On the other hand, thanks to Theorem C.12, one has  $(a_i - \widehat{a}_{i,\ell})^2 (\ell \pi^{-\ell})^{-1} = \mathcal{O}(1)$  a.s. As a result, one obtains  $\lim_{\ell \rightarrow \infty} x_{i,\ell} = 0$  a.s. as  $\pi > 1$  by assumption. Therefore, Lemma B.1 yields

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^{n/2}} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{j,2k+i} (a_i - \widehat{a}_{i,\ell})^2 = 0 \quad \text{a.s..}$$

The other terms in  $U_n^i$  are dealt with similarly, using Lemma C.3 instead of Lemma B.5. One obtains  $\lim_{n \rightarrow \infty} \pi^{-n/2}U_n^i = 0$  a.s. and as a result Lemma B.3 yields  $\lim_{n \rightarrow \infty} |\mathbb{T}_n^*|^{-1/2}U_n^i = 0$ . Let us now deal with the martingale terms  $V_n^i$ . Set  $\mathbf{V}_n = (V_n^0, V_n^1)^t$ . Let us remark that  $|\mathbb{T}_n^*|^{-1/2}\mathbf{V}_n = \mathbf{M}_{\nu_n}^{(n)}$  with  $\mathbf{M}^{(n)} = (\mathbf{M}_p^{(n)})_{\{p \geq 1\}}$  the sequence of  $\mathcal{G}_p^{\mathcal{O}}$ -vector martingales defined by

$$\mathbf{M}_p^{(n)} = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^p (v_k^0, v_k^1)^t = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^p \sum_{j=1}^m (v_{(j,k)}^0, v_{(j,k)}^1)^t$$

and  $\nu_n = 2^n - 1$  ( $\mathcal{G}_p^{\mathcal{O}}$  defined by (8)). We want now to apply Theorem B.6 to  $\mathbf{M}^{(n)}$ . Using Lemmas C.3-C.9 together with Lemma B.1 and Theorem C.12 along the same lines as above, we obtain the following limit conditionally to  $\bar{\mathcal{E}}$

$$\lim_{n \rightarrow \infty} \langle \mathbf{M} \rangle_{\nu_n} = \begin{pmatrix} (\tau_0^4 - \sigma_0^4)z^0 & (\nu^2 - \sigma_0^2\sigma_1^2)h^{01}(0)\pi^{-1} \\ (\nu^2 - \sigma_0^2\sigma_1^2)h^{01}(0)\pi^{-1} & (\tau_1^4 - \sigma_1^4)z^1 \end{pmatrix} = \Gamma_{\mathbf{V}}.$$

Therefore, assumption A.1 of Theorem B.6 holds under  $\mathbb{P}_{\bar{\mathcal{E}}}$ . Thanks to assumptions (H.1) and (H.4-5) we can prove that for some  $r > 2$ ,  $\sup_{k \geq 0} \mathbb{E}_{\bar{\mathcal{E}}}[\|v_k^i\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty$  a.s. which implies the Lindeberg condition. Therefore, we obtain that under  $\mathbb{P}_{\bar{\mathcal{E}}}$

$$|\mathbb{T}_n^*|^{-1/2}\mathbf{V}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma_{\mathbf{V}}).$$

If one sets

$$\Gamma_{\sigma} = \begin{pmatrix} (\tau_0^4 - \sigma_0^4)(z^0)^{-1} & (\nu^2 - \sigma_0^2\sigma_1^2)h^{01}(0)(\pi z^0 z^1)^{-1} \\ (\nu^2 - \sigma_0^2\sigma_1^2)h^{01}(0)(\pi z^0 z^1)^{-1} & (\tau_1^4 - \sigma_1^4)(z^1)^{-1} \end{pmatrix}, \quad (9)$$

one obtains the expected result using Slutsky's lemma.  $\square$

*Proof of Theorem 4.2, Asymptotic normality of  $\widehat{\rho}_n$ .* Along the same lines, we show the central limit theorem for  $\widehat{\rho}_n$ . One has

$$\begin{aligned} |\mathbb{T}_{n-1}^{*01}|(\widehat{\rho}_n - \rho) &= \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} (\widehat{\varepsilon}_{(j,2k)} \widehat{\varepsilon}_{(j,2k+1)} - \varepsilon_{(j,2k)} \varepsilon_{(j,2k+1)}) \\ &= \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} u'_{(j,k)} + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} v'_{(j,k)} = U'_n + V'_n, \end{aligned}$$

with

$$\begin{aligned} u'_{(j,k)} &= \delta_{(j,2k)} \delta_{(j,2k+1)} ((a_0 - \widehat{a}_{0,\ell})(a_1 - \widehat{a}_{1,\ell}) + (b_0 - \widehat{b}_{0,\ell})(b_1 - \widehat{b}_{1,\ell}) X_{(j,k)}^2) \\ &\quad + ((a_0 - \widehat{a}_{0,\ell})(b_1 - \widehat{b}_{1,\ell}) + (b_0 - \widehat{b}_{0,\ell})(a_1 - \widehat{a}_{1,\ell})) X_{(j,k)}, \\ v'_{(j,k)} &= \delta_{(j,2k)} \delta_{(j,2k+1)} (((a_0 - \widehat{a}_{0,\ell}) + (b_0 - \widehat{b}_{0,\ell}) X_{(j,k)}) \varepsilon_{(j,2k+1)} \\ &\quad + ((a_1 - \widehat{a}_{1,\ell}) + (b_1 - \widehat{b}_{1,\ell}) X_{(j,k)}) \varepsilon_{(j,2k)} + \varepsilon_{(j,2k)} \varepsilon_{(j,2k+1)} - \rho). \end{aligned}$$

Thanks to Theorem C.12, it is easy to check that  $\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^{*01}|^{1/2} U'_n = 0$  a.s. Let us define a new sequence of  $\mathcal{G}_p^{\mathcal{O}}$ -martingales  $(M^{(n)})$  by

$$M_p^{(n)} = |\mathbb{T}_{n-1}^{*01}|^{-1/2} \sum_{k=1}^p v'_k = |\mathbb{T}_{n-1}^{*01}|^{-1/2} \sum_{k=1}^p \sum_{j=1}^m v'(j, k).$$

We clearly have  $M_{\nu_n}^{(n)} = |\mathbb{T}_{n-1}^{*01}|^{1/2} V'_n$ . We obtain the  $\mathbb{P}_{\bar{\mathcal{E}}}$ - a.s. limit

$$\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^{*01}|^{-1} \sum_{k \in \mathbb{T}_n} \mathbb{E}_{\bar{\mathcal{E}}} [v_k^2 | \mathcal{G}_{k-1}^{\mathcal{O}}] = \nu^2 - \rho^2.$$

So we have assumption A.1 of Theorem B.6. We also derive the Lindeberg condition A.2. Consequently, we obtain that under  $\mathbb{P}_{\bar{\mathcal{E}}}$ , one has

$$\sqrt{|\mathbb{T}_{n-1}^{*01}|} V'_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 - \rho^2).$$

Setting

$$\gamma_\rho = \nu^2 - \rho^2, \tag{10}$$

completes the proof of Theorem 4.2.  $\square$

#### C.4 Interval estimation and tests for the BAR process

For all  $n \geq 1$ , define the  $4 \times 4$  matrices  $\hat{\mathbf{\Gamma}}_n$  and  $\hat{\mathbf{\Omega}}_n$  by

$$\hat{\mathbf{\Gamma}}_n = |\mathbb{T}_n^*|^{-1} \begin{pmatrix} \hat{\sigma}_{0,n}^2 \mathbf{S}_n^0 & \hat{\rho}_n \mathbf{S}_n^{01} \\ \hat{\rho}_n \mathbf{S}_n^{01} & \hat{\sigma}_{1,n}^2 \mathbf{S}_n^1 \end{pmatrix}, \quad \text{and} \quad \hat{\mathbf{\Omega}}_n = \mathbf{\Sigma}_n^{-1} \hat{\mathbf{\Gamma}}_n \mathbf{\Sigma}_n^{-1}.$$

Note that the matrix  $\hat{\mathbf{\Gamma}}_n$  is the empirical estimator of matrix  $\mathbf{\Gamma}$  while  $\hat{\mathbf{\Omega}}_n$  is the empirical estimator of the asymptotic variance of  $\hat{\theta}_n - \theta$ .

**Theorem C.13** *Under assumptions (H.0-6), for any  $0 < \epsilon < 1$ , the intervals*

$$\begin{aligned} & [\hat{a}_{0,n} - q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{1,1}; \hat{a}_{0,n} + q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{1,1}], & [\hat{b}_{0,n} - q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{2,2}; \hat{b}_{0,n} + q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{2,2}], \\ & [\hat{a}_{1,n} - q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{3,3}; \hat{a}_{1,n} + q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{3,3}], & [\hat{b}_{1,n} - q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{4,4}; \hat{b}_{1,n} + q_{1-\epsilon/2}(\hat{\mathbf{\Omega}}_{n-1}^{1/2})_{4,4}] \end{aligned}$$

are asymptotic confidence intervals with level  $1 - \epsilon$  of the parameters  $a_0$ ,  $b_0$ ,  $a_1$  and  $b_1$  respectively.

*Proof* This is a straightforward consequence of the central limit Theorem 4.2 together with Slutsky's lemma as  $\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^*| \hat{\mathbf{\Omega}}_{n-1} = \mathbf{\Sigma}^{-1} \mathbf{\Gamma} \mathbf{\Sigma}^{-1}$   $\mathbb{P}_{\bar{\mathcal{E}}}$  a.s. thanks to Lemma C.10 and Theorem 4.1.  $\square$

Let

$$\hat{h}_n^{01}(0) = \hat{p}_n^{(0)}(1, 1) |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j,2k)} + \hat{p}_n^{(1)}(1, 1) |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j,2k+1)}$$

be an empirical estimator of  $h^{01}(0)$  and

$$\hat{\mathbf{\Gamma}}_{\sigma,n} = \begin{pmatrix} (\hat{\tau}_{0,n}^4 - \hat{\sigma}_{0,n}^4) \frac{|\mathbb{T}_n^*|}{|\mathbb{T}_{n-1}^{*0}|} & (\hat{\nu}_n^2 - \hat{\sigma}_{0,n}^2 \hat{\sigma}_{1,n}^2) h^{01}(0) \hat{\pi}_n^{-1} \frac{|\mathbb{T}_n^*|^2}{|\mathbb{T}_{n-1}^{*0}| |\mathbb{T}_{n-1}^{*1}|} \\ (\hat{\nu}_n^2 - \hat{\sigma}_{0,n}^2 \hat{\sigma}_{1,n}^2) h^{01}(0) \hat{\pi}_n^{-1} \frac{|\mathbb{T}_n^*|^2}{|\mathbb{T}_{n-1}^{*0}| |\mathbb{T}_{n-1}^{*1}|} & (\hat{\tau}_{1,n}^4 - \hat{\sigma}_{1,n}^4) \frac{|\mathbb{T}_n^*|}{|\mathbb{T}_{n-1}^{*1}|} \end{pmatrix},$$

be an empirical estimator of the variance term in the central limit theorem regarding  $\sigma_i^2$ .

**Theorem C.14** Under assumptions (H.0-6), for any  $0 < \epsilon < 1$ , the intervals

$$\left[ \widehat{\sigma}_{i,n}^2 - q_{1-\epsilon/2} \left( \frac{\widehat{\mathbf{\Gamma}}_{\sigma,n}}{|\mathbb{T}_n^*|} \right)_{i,i}^{1/2} ; \widehat{\sigma}_{i,n}^2 + q_{1-\epsilon/2} \left( \frac{\widehat{\mathbf{\Gamma}}_{\sigma,n}}{|\mathbb{T}_n^*|} \right)_{i,i}^{1/2} \right],$$

$$\left[ \widehat{\rho}_n - q_{1-\epsilon/2} \left( \frac{\widehat{\nu}_n^2 - \widehat{\rho}_n^2}{|\mathbb{T}_{n-1}^{*01}|} \right)^{1/2} ; \widehat{\rho}_n + q_{1-\epsilon/2} \left( \frac{\widehat{\nu}_n^2 - \widehat{\rho}_n^2}{|\mathbb{T}_{n-1}^{*01}|} \right)^{1/2} \right]$$

are asymptotic confidence intervals with level  $1 - \epsilon$  of the parameters  $\sigma_i^2$  and  $\rho$  respectively.

*Proof* This is a again straightforward consequence of the central limit Theorem 4.2 together with Slutsky's lemma as

$$\lim_{n \rightarrow \infty} \widehat{\mathbf{\Gamma}}_{\sigma,n} = \mathbf{\Gamma}_{\sigma}, \quad \lim_{n \rightarrow \infty} \widehat{\nu}_n^2 - \widehat{\rho}_n^2 = \nu^2 - \rho^2,$$

$\mathbb{P}_{\bar{\mathcal{E}}}$  almost surely thanks to Lemma B.5 and Theorems 4.1 and C.11.  $\square$

We now propose two different symmetry tests for the BAR process based on the central limit Theorem 4.2. The first one compares the couples  $(a_0, b_0)$  and  $(a_1, b_1)$ . Set

- $\mathbf{H}_0^c$ :  $(a_0, b_0) = (a_1, b_1)$  the symmetry hypothesis,
- $\mathbf{H}_1^c$ :  $(a_0, b_0) \neq (a_1, b_1)$  the alternative hypothesis.

Let  $(\mathbf{Y}_n^c)^t \mathbf{Y}_n^c$  be the test statistic defined by

$$\mathbf{Y}_n^c = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\mathbf{\Delta}}_n^c)^{-1/2} (\widehat{a}_{0,n} - \widehat{a}_{1,n}, \widehat{b}_{0,n} - \widehat{b}_{1,n})^t,$$

where

$$\widehat{\mathbf{\Delta}}_n^c = |\mathbb{T}_{n-1}^*| \mathbf{d}\mathbf{g}_c^t \widehat{\mathbf{\Omega}}_{n-1} \mathbf{d}\mathbf{g}_c, \quad \mathbf{d}\mathbf{g}_c = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}^t.$$

**Theorem C.15** Under assumptions (H.0-6) and the null hypothesis  $\mathbf{H}_0^c$  one has

$$(\mathbf{Y}_n^c)^t \mathbf{Y}_n^c \xrightarrow{\mathcal{L}} \chi^2(2)$$

on  $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$ ; and under the alternative hypothesis  $\mathbf{H}_1^c$ , almost surely on  $\bar{\mathcal{E}}$  one has

$$\lim_{n \rightarrow \infty} \|\mathbf{Y}_n^c\|^2 = +\infty.$$

*Proof* We mimic again the proof of Theorem B.9 with  $g_c$  the function defined from  $\mathbb{R}^4$  onto  $\mathbb{R}^2$  by  $g_c(x_1, x_2, x_3, x_4) = (x_1 - x_3, x_2 - x_4)^t$ , so that  $\mathbf{d}\mathbf{g}_c$  is the gradient of  $g_c$ .  $\square$

Our next test compares the fixed points  $a_0/(1 - b_0)$  and  $a_1/(1 - b_1)$ , which are the asymptotic means of  $X_{(j,2k)}$  and  $X_{(j,2k+1)}$  respectively. Set

- $\mathbf{H}_0^f$ :  $a_0/(1 - b_0) = a_1/(1 - b_1)$  the symmetry hypothesis,
- $\mathbf{H}_1^f$ :  $a_0/(1 - b_0) \neq a_1/(1 - b_1)$  the alternative hypothesis.

Let  $(Y_n^f)^2$  be the test statistic defined by

$$Y_n^f = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\mathbf{\Delta}}_n^f)^{-1/2} (\widehat{a}_{0,n}/(1 - \widehat{b}_{0,n}) - \widehat{a}_{1,n}/(1 - \widehat{b}_{1,n})),$$

where  $\widehat{\mathbf{\Delta}}_n^f = |\mathbb{T}_{n-1}^*| \mathbf{d}\mathbf{g}_f^t \widehat{\mathbf{\Omega}}_{n-1} \mathbf{d}\mathbf{g}_f$ , and  $\mathbf{d}\mathbf{g}_f = (1/(1 - \widehat{b}_{0,n}), \widehat{a}_{0,n}/(1 - \widehat{b}_{0,n})^2, -1/(1 - \widehat{b}_{1,n}), -\widehat{a}_{1,n}/(1 - \widehat{b}_{1,n})^2)^t$ . This test statistic has the following asymptotic properties.

**Theorem C.16** Under assumptions (H.0-6) and the null hypothesis  $\mathbf{H}_0^f$ , one has

$$(Y_n^f)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

on  $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$ ; and under the alternative hypothesis  $\mathbf{H}_1^f$ , almost surely on  $\bar{\mathcal{E}}$  one has

$$\lim_{n \rightarrow \infty} (Y_n^f)^2 = +\infty.$$

*Proof* We mimic again proof of Theorem B.9 with  $g_f$  the function defined from  $\mathbb{R}^4$  onto  $\mathbb{R}$  by  $g_f(x_1, x_2, x_3, x_4) = (x_1/(1-x_2) - x_3/(1-x_4))$ , so that  $\mathbf{d}g_f$  is the gradient of  $g_f$ .  $\square$

Finally, our last test compares the even and odd variances  $\sigma_0^2$  and  $\sigma_1^2$  of the noise sequence. Set

- $\mathbf{H}_0^\sigma$ :  $\sigma_0^2 = \sigma_1^2$  the symmetry hypothesis,
- $\mathbf{H}_1^\sigma$ :  $\sigma_0^2 \neq \sigma_1^2$  the alternative hypothesis.

Let  $(Y_n^\sigma)^2$  be the test statistic defined by

$$Y_n^\sigma = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^\sigma)^{-1/2} (\widehat{\sigma}_{0,n}^2 - \widehat{\sigma}_{1,n}^2),$$

where  $\widehat{\Delta}_n^\sigma = |\mathbb{T}_{n-1}^*| \mathbf{d}g_\sigma^t \widehat{\Gamma}_{\sigma, n-1} \mathbf{d}g_\sigma$ , and  $\mathbf{d}g_\sigma = (1, -1)^t$ . This test statistic has the following asymptotic properties.

**Theorem C.17** *Under assumptions (H.0-6) and the null hypothesis  $\mathbf{H}_0^\sigma$ , one has*

$$(Y_n^\sigma)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

*on  $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$ ; and under the alternative hypothesis  $\mathbf{H}_1^\sigma$ , almost surely on  $\bar{\mathcal{E}}$  one has*

$$\lim_{n \rightarrow \infty} (Y_n^\sigma)^2 = +\infty.$$

*Proof* We mimic one last time the proof of Theorem B.9 with  $g_\sigma$  the function defined from  $\mathbb{R}^2$  onto  $\mathbb{R}$  by  $g_\sigma(x_1, x_2) = (x_1 - x_2)$ , so that  $\mathbf{d}g_\sigma$  is the gradient of  $g_\sigma$ .  $\square$

## References

- Bercu B, de Saporta B, Gégout-Petit A (2009) Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron J Probab* 14:no. 87, 2492–2526
- Cowan R, Staudte RG (1986) The bifurcating autoregressive model in cell lineage studies. *Biometrics* 42:769–783
- Delmas JF, Marsalle L (2010) Detection of cellular aging in a Galton-Watson process. *Stoch Process and Appl* 120:2495–2519
- Duflo M (1997) *Random iterative models, Applications of Mathematics*, vol 34. Springer-Verlag, Berlin
- Guttorp P (1991) *Statistical inference for branching processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, John Wiley & Sons Inc., New York, a Wiley-Interscience Publication
- Guyon J (2007) Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann Appl Probab* 17(5-6):1538–1569
- Guyon J, Bize A, Paul G, Stewart E, Delmas JF, Taddéi F (2005) Statistical study of cellular aging. In: *CEMRACS 2004—mathematics and applications to biology and medicine*, ESAIM Proc., vol 14, EDP Sci., Les Ulis, pp 100–114 (electronic)
- Harris TE (1963) *The theory of branching processes. Die Grundlehren der Mathematischen Wissenschaften*, Bd. 119, Springer-Verlag, Berlin
- Huggins RM (1996) Robust inference for variance components models for single trees of cell lineage data. *Ann Statist* 24(3):1145–1160
- Huggins RM, Basawa IV (1999) Extensions of the bifurcating autoregressive model for cell lineage studies. *J Appl Probab* 36(4):1225–1233

- Huggins RM, Basawa IV (2000) Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust N Z J Stat* 42(4):423–432
- Huggins RM, Staudte RG (1994) Variance components models for dependent cell populations. *J AMS* 89(425):19–29
- Maaouia F, Touati A (2005) Identification of multitype branching processes. *Ann Statist* 33(6):2655–2694
- de Saporta B, Gégout-Petit A, Marsalle L (2011) Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron J Statist* 5:1313–1353
- de Saporta B, Gégout Petit A, Marsalle L (2012) Asymmetry tests for bifurcating autoregressive processes with missing data. *Statistics & Probability Letters* 82(7):1439–1444
- Stewart E, Madden R, Paul G, Taddei F (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol* 3(2):e45
- Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, Jun S (2012) Robust growth of *escherichia coli*. *Current Biology*, 20(12):1099 – 1103
- Zhou J, Basawa IV (2005) Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J Time Ser Anal* 26(6):825–842