

Maximum likelihood estimation and expectation-maximization algorithm for controlled branching processes

M. González, C. Minuesa, I. del Puerto

Department of Mathematics, University of Extremadura, 06006 - Badajoz, Spain.

Abstract

The controlled branching process is a generalization of the classical Bienaymé-Galton-Watson branching process. It is a useful model for describing the evolution of populations in which the population size at each generation needs to be controlled. The maximum likelihood estimation of the parameters of interest for this process is addressed under various sample schemes. Firstly, assuming that the entire family tree can be observed, the corresponding estimators are obtained and their asymptotic properties investigated. Secondly, since in practice it is not usual to observe such a sample, the maximum likelihood estimation is initially considered using the sample given by the total number of individuals and progenitors of each generation, and then using the sample given by only the generation sizes. Expectation-maximization algorithms are developed to address these problems as incomplete data estimation problems. The accuracy of the procedures is illustrated by means of a simulated example.

Keywords: Maximum likelihood estimation, expectation-maximization algorithm, branching process, controlled process.

1. Introduction

Controlled branching processes are a class of discrete-time stochastic growth population models characterized by the existence of a random control mechanism to determine in each generation (non-overlapping generations) how many progenitors

*Corresponding Author: Phone: +34 924289300 ext. 86820. Fax: +34 924272911. E-mail: cminuesa@unex.es

This is the plain accepted version of the following paper published in the journal *Computational Statistics and Data Analysis* (see the official journal website at <https://doi.org/10.1016/j.csda.2015.01.015>):

González, M., Minuesa, C., del Puerto, I., 2016. Maximum likelihood estimation and expectation-maximization algorithm for controlled branching processes. *Computational Statistics & Data Analysis* **93**, 209–227. DOI: 10.1016/j.csda.2015.01.015

participate in the subsequent reproduction process. Once the number of progenitors is known, each one reproduces independently of the others according to the same probability law, called the offspring distribution, as usual in the framework of branching processes.

In general, the notion of branching has had relevance in the development of theoretical approaches to problems in such applied fields as the growth and extinction of populations, biology (gene amplification, clonal resistance theory of cancer cells, polymerase chain reactions, etc.), epidemiology (the evolution of infectious diseases), cell proliferation kinetics (stem cells, etc.), genetics (sex-linked genes, mitochondrial DNA, etc.) and algorithm and data structures (see, for example, the monographs Kimmel and Axelrod (2002) and Haccou et al. (2005)). In particular, the novelty of adding to the branching notion a mechanism that fixes the number of progenitors in each generation can allow a great variety of random migratory movements to be modeled. The control mechanism can be defined either by a degenerate distribution giving rise to deterministic control or in a random way (through control probability distributions), in both cases with dependence on the number of individuals in each generation. For example, a practical situation that can be modeled by this kind of process is the evolution of an animal population that is threatened by the existence of predators. In each generation, the survival of each animal (and therefore the possibility of giving new births) will be strongly affected by this factor, making the introduction of a random mechanism (a binomial control process would be reasonable) necessary to model the evolution of this kind of population. One can also model phenomena concerning the introduction or re-introduction of animal species to inhabit environments in which they are in potential danger of disappearance or have previously become extinct. This re-population can be achieved by the controlled introduction of new animals until the species has become firmly established in that habitat.

The family of controlled branching processes includes as particular cases various models previously introduced in the branching process literature, such as branching processes with immigration (see Sriram (1994)), with immigration at state zero (see Bruss and Slavtchova-Bojkova (1999)), with random migration (see Yanev and Yanev (1996)), with bounded emigration (see del Puerto and Yanev (2008)), with adaptive control (see Bercu (1999)), and with continuous state space (see Rahimov and Al-Sabah (2007)).

The probability theory of this model has been extensively studied from the pioneering work of Yanev (1976) until the recent paper of González and del Puerto (2012) (see also the references therein). In the last few years, interest in these processes has mainly focused on the development of their inference theory in order to guarantee the applicability of these models. Results in this line from a frequentist standpoint may be found in González et al. (2004, 2005a) for deterministic control models, using maximum likelihood estimation, and in Dion and Essebbar (1995) and Sriram et al. (2007) for models with random control distributions, using martingale theory (for a multiplicative control function) and weighted conditional least squares

estimation, respectively.

The objective of this paper is to consider the maximum likelihood estimation of the parameters of interest for a controlled branching process with random control distributions under various sample schemes. Firstly, we consider the entire family tree until some fixed generation can be observed. The results obtained under the observation of this sample generalize those in González et al. (2004, 2005a). Secondly, since, in practice, it is not usual to observe the entire family tree, we consider the maximum likelihood estimation using initially the sample given by the total number of individuals and progenitors of each generation, and then the sample given by only the generation sizes. We deal with these problems as incomplete data estimation problems, and develop expectation-maximization (EM) algorithms to this end (see McLachlan and Krishnan (2008), for details of this methodological approach or for recent applications of this methodology in Bernhardt et al. (2015) and Wang et al. (2015)). EM algorithms have been successfully used to approximate maximum likelihood estimators when there are missing or incomplete data, although there are only a few articles on their use in the context of branching processes (see Veen and Schoenberg (2008), González et al. (2012), Daskalova (2014) and Hautphenne and Fackrell (2014)), and in no case for models which consider random control mechanisms.

After this Introduction, the paper is organized as follows. We begin by describing the probability model in Section 2, in which we introduce some notation and the working assumptions for the subsequent study. Section 3 is devoted to the maximum likelihood estimation based on the complete family tree and to studying the asymptotic properties of the estimators obtained. In Section 4, we address the problem of obtaining maximum likelihood estimates under incomplete sampling schemes, developing the EM algorithms. The accuracy of these algorithms is illustrated by means of a simulated example in Section 5 (see the supplementary material for data sets and a further discussion of some aspects of the example). Some concluding remarks are provided in Section 6. Finally, in order to allow a more readily comprehensible reading, Appendices A, B, and C are devoted to giving the proofs of the theoretical results set out in the paper.

2. The Probability Model

We shall focus our attention on the class of the controlled branching process with random control function (CBP). Mathematically, this process is a discrete-time stochastic growth population model $\{Z_n\}_{n \geq 0}$ defined recursively as follows:

$$Z_0 = N, \quad Z_{n+1} = \sum_{j=1}^{\phi_n(Z_n)} X_{nj}, \quad n = 0, 1, \dots, \quad (1)$$

where N is a non-negative integer, $\{X_{nj} : n = 0, 1, \dots; j = 1, 2, \dots\}$ and $\{\phi_n(k) : n, k = 0, 1, \dots\}$ are two independent families of non-negative integer valued random

variables. Also, X_{nj} , $n = 0, 1, \dots$, $j = 1, 2, \dots$, are independent and identically distributed (i.i.d.) random variables, and, for each $n = 0, 1, \dots$, $\{\phi_n(k)\}_{k \geq 0}$, are independent stochastic processes with equal one-dimensional probability distributions. The empty sum in (1) is considered to be 0. Let $p = \{p_k\}_{k \geq 0}$ denote the common probability distribution of the random variables X_{nj} , i.e., $p_k = P[X_{nj} = k]$, $k \geq 0$, and m and σ^2 its mean and variance (assumed finite), respectively. We also denote by $\varepsilon(k) = E[\phi_0(k)]$ and $\sigma^2(k) = Var[\phi_0(k)]$ the mean and the variance of the control variables (assumed finite too).

Intuitively, Z_n denotes the number of individuals (particles) in the n -th generation and X_{nj} the number of offspring of the j -th individual in the n -th generation. The probability law p is called the offspring distribution, and m and σ^2 are the offspring mean and variance, respectively. The variable $\phi_n(Z_n)$ represents a control on the number of progenitors in each generation, in such a way that when $\phi_n(Z_n) = k$ then k will be the number of individuals who will take part in the reproduction process that will determine Z_{n+1} . Thus, if $\phi_n(Z_n) < Z_n$ then $Z_n - \phi_n(Z_n)$ individuals are removed from the population (emigration, presence of predators, etc.), and therefore do not participate in the future evolution of the process. If $\phi_n(Z_n) > Z_n$ then $\phi_n(Z_n) - Z_n$ new individuals of the same type are added to the population (immigration, re-population, etc.). No control is applied to the population when $\phi_n(Z_n) = Z_n$. Obviously, if $\phi_n(k) = k$ for all k , one obtains the standard Bienaymé-Galton-Watson process.

It is easy to verify that $\{Z_n\}_{n \geq 0}$ is a Markov chain with stationary transition probabilities. Moreover, assuming

- (a) $p_0 > 0$ or $P[\phi_n(k) = 0] > 0$, $k > 0$,
- (b) $\phi_n(0) = 0$ almost surely (a.s.),

then 0 is an absorbing state and the states $k = 1, 2, \dots$ are transient. Whence it is verified that $P[Z_n \rightarrow 0] + P[Z_n \rightarrow \infty] = 1$.

Let us fix the main parameters of interest and the working assumptions for the development of their maximum likelihood estimation. Consider a CBP with an offspring distribution p , whose mean and variance are m and σ^2 , respectively. Given that one has different control laws for different population sizes, the problem of estimating the control parameters would seem intractable based on samples with a finite dimension unless the control process is assumed to have a structure that is stable over time. In this sense, formally we consider CBPs given by (1) with control distributions belonging to the power series family of distributions, i.e., for each $k \geq 0$,

$$P[\phi_n(k) = j] = a_k(j)\theta^j A_k(\theta)^{-1}, \quad j \geq 0; \theta \in \Theta_k, \quad (2)$$

with $a_k(j)$ taking known non-negative values, $A_k(\theta) = \sum_{j=0}^{\infty} a_k(j)\theta^j$, and $\Theta_k = \{\theta > 0 : 0 < A_k(\theta) < \infty\}$ being an open subset of \mathbb{R} . We also assume that the sets Θ_k are independent of k , so that we shall henceforth drop the index k from Θ_k , the control

parameter space. Moreover, we assume the following regularity condition:

$$\prod_{k \in C} A_k(\theta) = A_{\sum_{k \in C} k}(\theta), \quad \text{for every } C \subseteq \mathbb{N}; \theta \in \Theta. \quad (3)$$

Remark 2.1. *The distribution given in (2) is an exponential family which includes many important discrete distributions (e.g., Poisson, binomial, negative binomial, etc.). The condition (3) is a technical hypothesis, satisfied by a wide set of probability distributions belonging to the exponential family. Hence, the control distributions in the model depend on a single parameter θ , termed the control parameter, and on the size of the population, say k .*

It is well known that:

$$\begin{aligned} \varepsilon(k) &= \varepsilon(k, \theta) = E[\phi_0(k)] = \theta \frac{d}{d\theta} \log A_k(\theta), \\ \sigma^2(k) &= \sigma^2(k, \theta) = \text{Var}[\phi_0(k)] = \theta \frac{d}{d\theta} \varepsilon(k, \theta). \end{aligned}$$

Under condition (3), it can be deduced that $\varepsilon(k, \theta) = k\mu(\theta)$, $k \geq 0$, $\theta \in \Theta$, where $\mu(\cdot)$ is a continuous and invertible function. From (3), $A_k(\theta) = A_1(\theta)^k$, $k \geq 1$, so that

$$\varepsilon(k, \theta) = \frac{\theta \frac{d}{d\theta} A_k(\theta)}{A_1(\theta)^k} = k \frac{\theta \frac{d}{d\theta} A_1(\theta)}{A_1(\theta)} = k\theta \frac{d}{d\theta} \log(A_1(\theta)) = k\varepsilon(1, \theta).$$

Therefore, a family of distributions which verifies (3) can be re-parametrized making use of the parameter $\mu = \mu(\theta) = \varepsilon(1, \theta)$. This parameter can be termed the migration parameter because of its intuitive interpretation: if $\mu < 1$, the control law allows one to model processes with expected emigration; if $\mu > 1$, one can model processes with expected immigration; and if $\mu = 1$, no migration is expected. One also notes that, under assumption (3), $\sigma^2(k, \theta) = k\theta\mu'(\theta)$, with $\mu'(\cdot)$ denoting the first derivative of $\mu(\cdot)$.

Remark 2.2. *Three interesting particular cases of distributions which verify (2) and (3) are the following:*

- (i) *For each $k \geq 0$, take $\phi_n(k)$ to follow a Poisson distribution of parameter $k\theta$. Consequently, $\mu(\theta) = \theta$. Hence, depending on the value of θ , a CBP with this control function can model different migratory processes. It is easy to verify that conditions (2) and (3) hold by setting $a_k(j) = k^j/j!$ and $A_k(\theta) = e^{k\theta}$.*
- (ii) *For each $k \geq 0$, take $\phi_n(k)$ to follow a binomial distribution of parameters k and q . Taking $\theta = q(1-q)^{-1}$, $a_k(j) = \binom{k}{j}$, and $A_k(\theta) = (1+\theta)^k$, conditions (2) and (3) can be checked straightforwardly, and $\mu(\theta) = \theta(1+\theta)^{-1} = q$. From a practical viewpoint, this could be a reasonable control mechanism with which*

to model situations in which, in each generation, each individual can give birth to offspring in the next generation with probability q , and is removed from the population with probability $1 - q$, not participating in its subsequent evolution. As $\mu(\theta) < 1$, a CBP with this control distribution always models a case of expected emigration, and, for example, could be useful to model the presence of predators in an animal population.

- (iii) For each $k \geq 0$, take $\phi_n(k)$ to follow a negative binomial distribution of parameters k and q . In this case, conditions (2) and (3) can be checked by setting $\theta = 1 - q$, $a_k(j) = \binom{j+k-1}{j}$, and $A_k(\theta) = (1 - \theta)^{-k}$. Moreover, $\mu(\theta) = \theta(1 - \theta)^{-1}$. As also was the case for the model considered in (i), this process can model either expected immigration or expected emigration.

Finally, another parameter of great interest for this family of processes is what is termed the asymptotic mean growth rate. This is denoted by τ_m , and is defined in general as $\lim_{k \rightarrow \infty} k^{-1} E[Z_{n+1} | Z_n = k] = \lim_{k \rightarrow \infty} k^{-1} m\varepsilon(k)$ (whenever it exists). Under condition (3), $\tau_m = m\mu(\theta)$. This is the threshold parameter that determines the behaviour of a CBP in relation to its extinction. Following the classification of CBPs set out in González et al. (2005b), we shall term a CBP as subcritical, critical, or supercritical depending on whether τ_m is less than, equal to, or greater than unity (emulating the Bienaymé–Galton–Watson process classification).

In summary, we deal with the problem of estimating p , m , σ^2 , θ , $\mu(\theta)$, and τ_m by making use of the maximum likelihood estimation based on different samples.

3. Maximum Likelihood Estimators with Complete Data

In this section, we shall consider the maximum likelihood estimation of the aforementioned parameters of interest by assuming that one can observe the entire family tree up to generation n (complete data), i.e., the random variables $\{X_{li} : 1 \leq i \leq \phi_l(Z_l); 0 \leq l \leq n-1\}$, or at least $\mathcal{Z}_n^* = \{Z_l(k) : 0 \leq l \leq n-1; k \geq 0\}$, where $Z_l(k) = \sum_{i=1}^{\phi_l(Z_l)} I_{\{X_{li}=k\}}$, $0 \leq l \leq n-1$, $k \geq 0$, with I_A standing for the indicator function of the set A . Intuitively, $Z_l(k)$ represents the number of individuals in generation l who have exactly k offspring. It is easily deduced that $\phi_l(Z_l) = \sum_{k=0}^{\infty} Z_l(k)$ and $Z_{l+1} = \sum_{k=0}^{\infty} kZ_l(k)$, $l = 0, \dots, n-1$.

Let $Y_l = \sum_{j=0}^l Z_j$, $\Delta_l = \sum_{j=0}^l \phi_j(Z_j)$, and $Y_l(k) = \sum_{j=0}^l Z_j(k)$, $l \geq 0$, $k \geq 0$. Intuitively, Y_l and Δ_l denote the total number of individuals and the total number of parents until the l -th generation, respectively, and $Y_l(k)$ represents the accumulated number up to generation l of individuals who have exactly k offspring. The results presented in this section generalize those given in González et al. (2004, 2005a) for CBPs with a deterministic control function.

Theorem 3.1. Let $\{Z_n\}_{n \geq 0}$ be a CBP verifying (2) and (3). The maximum likelihood estimators (MLEs) of p_k , $k \geq 0$, and θ , based on \mathcal{Z}_n^* , are, respectively:

$$\hat{p}_{k,n} = \frac{Y_{n-1}(k)}{\Delta_{n-1}}, \quad k \geq 0, \quad \text{and} \quad \hat{\theta}_n = \mu^{-1} \left(\frac{\Delta_{n-1}}{Y_{n-1}} \right),$$

where $\mu^{-1}(\cdot)$ denotes the inverse of the function $\mu(\cdot)$.

The proof is given in Appendix A.

Using this theorem and the invariance of the MLEs under re-parametrization, the following result is immediate:

Corollary 3.2. Let $\{Z_n\}_{n \geq 0}$ be a CBP verifying (2) and (3). The MLEs of m , σ^2 , $\mu(\theta)$, and τ_m based on \mathcal{Z}_n^* , are, respectively:

$$\hat{m}_n = \frac{Y_n - Z_0}{\Delta_{n-1}}, \quad \hat{\sigma}_n^2 = \sum_{k=0}^{\infty} (k - \hat{m}_n)^2 \hat{p}_{k,n}, \quad \hat{\mu}_n = \frac{\Delta_{n-1}}{Y_{n-1}}, \quad \text{and} \quad \hat{\tau}_{m,n} = \frac{Y_n - Z_0}{Y_{n-1}}.$$

For simplicity, when the meaning is clear, we shall drop the index n from $\hat{p}_{k,n}$ and $\hat{\tau}_{m,n}$ and write simply \hat{p}_k and $\hat{\tau}_m$.

Remark 3.3. (i) It is worth noting that to obtain the MLE of the offspring distribution, p , and its associated parameters, m and σ^2 , it is not necessary to impose the requirement of any knowledge about the control distribution. One can thus address this problem in a nonparametric framework, obtaining the same estimators for these three parameters.

(ii) The MLEs of p_k and m are intuitively very reasonable because we estimate the probability that an individual gives rise to k offspring by the relative proportion of parents with k offspring, and the offspring mean is estimated by the total number of offspring up to a certain generation divided by the number of progenitors who have generated those offspring.

(iii) It can be proved that \hat{m}_n , $\hat{\theta}_n$, $\hat{\mu}_n$, and $\hat{\tau}_m$ are also the MLEs of m , θ , $\mu(\theta)$, and τ_m , respectively, based on the sample $\{Z_0, \dots, Z_n, \phi_0(Z_0), \dots, \phi_{n-1}(Z_{n-1})\}$ (see Jagers (1975), Lemma 2.13.2). Moreover, $\hat{\tau}_m$ is also the MLE of τ_m based on $\{Z_0, \dots, Z_n\}$, following similar arguments.

3.1. Asymptotic behaviour

In order to investigate the asymptotic properties of the proposed estimators, it will be necessary to make some working assumptions. To parameters associated with the offspring distribution, one does not need to assume that the control variables belong to a power series family of distributions. Instead, one only needs to assume

that the CBP $\{Z_n\}_{n \geq 0}$ verifies the following conditions:

- (a) There exists $\tau = \lim_{k \rightarrow \infty} \varepsilon(k)k^{-1} < \infty$, and the sequence $\{\sigma^2(k)k^{-1}\}_{k \geq 1}$ is bounded.
- (b) $\tau_m = \tau m > 1$, and Z_0 large enough such that $P[Z_n \rightarrow \infty] > 0$.
- (c) $\{Z_n \tau_m^{-n}\}_{n \geq 0}$ converges *a.s.* to a finite random variable W such that $P[W > 0] > 0$.
- (d) $\{W > 0\} = \{Z_n \rightarrow \infty\}$ *a.s.*

Remark 3.4. (i) In González et al. (2002), conditions are provided that guarantee (b) in (4). Also, in González et al. (2006), conditions are established under which $\{W > 0\} = \{Z_n \rightarrow \infty\}$ *a.s.* is verified.

(ii) It can be proved (see González et al. (2002), Theorem 4) that, under condition (4), on the set $\{Z_n \rightarrow \infty\}$ one has that

$$\lim_{n \rightarrow \infty} Z_n^{-1} Z_{n+1} = \tau_m \quad \text{a.s.}$$

We shall now establish a preliminary result that will be used in the study of the estimators' asymptotic properties. The proof is omitted because it is a consequence of Remark 3.4(ii) and the Stolz-Cesàro Lemma.

Proposition 3.5. Let $\{Z_n\}_{n \geq 0}$ be a CBP verifying the conditions given in (4). Then, on the set $\{Z_n \rightarrow \infty\}$, it is verified that:

- (i) $\lim_{n \rightarrow \infty} Z_n^{-1} \phi_n(Z_n) = \tau$ *a.s.*
- (ii) $\sum_{n=0}^{\infty} \phi_n(Z_n)^{-1} < \infty$ *a.s.*
- (iii) $\lim_{n \rightarrow \infty} Y_n^{-1} Y_{n+1} = \tau_m$ *a.s.*
- (iv) $\lim_{n \rightarrow \infty} Y_n^{-1} \Delta_n = \tau$ *a.s.*
- (v) $\lim_{n \rightarrow \infty} \Delta_n^{-1} \phi_n(Z_n) = \tau_m^{-1}(\tau_m - 1)$ *a.s.*
- (vi) $\lim_{n \rightarrow \infty} \varepsilon(Z_n)^{-1} \phi_n(Z_n) = 1$ *a.s.*

In the following result, we study asymptotic properties of the estimators related to the offspring distribution, i.e., \hat{p}_k , $k \geq 0$, \hat{m}_n , and $\hat{\sigma}_n^2$. For simplicity, we shall use the notation $\mathcal{D} = \{Z_n \rightarrow \infty\}$ and $P_{\mathcal{D}}[\cdot] = P[\cdot | \mathcal{D}]$. The result holds whether or not conditions (2) and (3) on the control are satisfied.

Theorem 3.6. Let $\{Z_n\}_{n \geq 0}$ be a CBP verifying (4). Then it holds that:

- (i) \hat{p}_k , \hat{m}_n , and $\hat{\sigma}_n^2$ are strongly consistent for p_k , m , and σ^2 , respectively, on $\{Z_n \rightarrow \infty\}$.
- (ii) If P' is a probability measure dominated by $P_{\mathcal{D}}$, then for any $x \in \mathbb{R}$:
 - (a) $\lim_{n \rightarrow \infty} P'[(p_k(1 - p_k))^{-1/2} \Delta_{n-1}^{1/2} (\hat{p}_k - p_k) \leq x] = \Phi(x)$,

- (b) $\lim_{n \rightarrow \infty} P'[\sigma^{-1} \Delta_{n-1}^{1/2} (\widehat{m}_n - m) \leq x] = \Phi(x)$,
- (c) If $E[X_{01}^4] < \infty$, then $\lim_{n \rightarrow \infty} P'[Var[(X_{01} - m)^2]^{-1/2} \Delta_{n-1}^{1/2} (\widehat{\sigma}_n^2 - \sigma^2) \leq x] = \Phi(x)$,
- with $\Phi(\cdot)$ denoting the standard normal distribution function.

The proof is given in Appendix B.

Remark 3.7. Using the previous theorem and Lemma 2.3 in Guttorp (1991), it is immediate to prove that (ii) also holds for $P[\cdot | Z_n > 0]$. Then, taking into account Theorem 3.6 and Slutsky's Theorem, and assuming $Z_n > 0$, one can obtain asymptotic confidence intervals for the parameters p , m , and σ^2 . Thus, for example, the asymptotic confidence interval for m at the $1 - \alpha$ level, $0 < \alpha < 1$, is given by

$$\left[\widehat{m}_n - z_\alpha (\widehat{\sigma}_n^2 \Delta_{n-1}^{-1})^{1/2}, \widehat{m}_n + z_\alpha (\widehat{\sigma}_n^2 \Delta_{n-1}^{-1})^{1/2} \right],$$

with z_α being such that $1 - \Phi(z_\alpha) = \alpha/2$.

Considering now the parameters of the control law, let us recall that if the latter belongs to the power series family of distributions then (4)(a) holds trivially, and $\tau = \mu(\theta)$. Denoting equal in distribution by $\stackrel{d}{=}$, one has the following result:

Theorem 3.8. Let $\{Z_n\}_{n \geq 0}$ be a CBP verifying (2), (3), and (4). Then it holds that:

- (i) $\widehat{\theta}_n$, $\widehat{\mu}_n$ and $\widehat{\tau}_m$ are strongly consistent for θ , $\mu(\theta)$ and τ_m , respectively, on $\{Z_n \rightarrow \infty\}$.
- (ii) If, for each $l \geq 0$ and $z \geq 0$, $\phi_l(z) \stackrel{d}{=} \sum_{s=1}^z X_s(l, z)$, with $\{X_s(l, z) : 1 \leq s \leq z; z \geq 0; l \geq 0\}$ being i.i.d. random variables with mean $\mu(\theta)$ and variance $\theta \mu'(\theta)$ then, for any $x \in \mathbb{R}$,
- (a) $\lim_{n \rightarrow \infty} P_{\mathcal{D}} \left[(\theta \mu'(\theta))^{-1/2} Y_{n-1}^{1/2} (\widehat{\mu}_n - \mu(\theta)) \leq x \right] = \Phi(x)$,
- (b) $\lim_{n \rightarrow \infty} P_{\mathcal{D}} \left[(\sigma^2 \mu(\theta) + m^2 \theta \mu'(\theta))^{-1/2} Y_{n-1}^{1/2} (\widehat{\tau}_m - \tau_m) \leq x \right] = \Phi(x)$,
- with $\Phi(\cdot)$ denoting the standard normal distribution function.

The proof is given in Appendix C.

Remark 3.9. (i) It is worthy of note that the condition set out in Theorem 3.8(ii) is satisfied by the control distributions introduced in Remark 2.2.

(ii) Theorem 3.8 (ii) also holds for $P[\cdot | Z_n > 0]$. Again, assuming $Z_n > 0$, from this theorem and Slutsky's Theorem, and replacing the values m , σ^2 , θ , and $\mu'(\theta)$ by \widehat{m}_n , $\widehat{\sigma}_n^2$, $\widehat{\theta}_n$, and $\mu'(\widehat{\theta}_n)$, respectively, one can obtain asymptotic confidence intervals for the parameters $\mu(\theta)$ and τ_m at the $1 - \alpha$ level, $0 < \alpha < 1$:

$$\left[\widehat{\mu}_n - z_\alpha \left(\widehat{\theta}_n \mu'(\widehat{\theta}_n) Y_{n-1}^{-1} \right)^{1/2}, \widehat{\mu}_n + z_\alpha \left(\widehat{\theta}_n \mu'(\widehat{\theta}_n) Y_{n-1}^{-1} \right)^{1/2} \right],$$

$$\left[\widehat{\tau}_m - z_\alpha \left((\widehat{\sigma}_n^2 \mu(\widehat{\theta}_n) + \widehat{m}_n^2 \widehat{\theta}_n \mu'(\widehat{\theta}_n)) Y_{n-1}^{-1} \right)^{1/2}, \widehat{\tau}_m + z_\alpha \left((\widehat{\sigma}_n^2 \mu(\widehat{\theta}_n) + \widehat{m}_n^2 \widehat{\theta}_n \mu'(\widehat{\theta}_n)) Y_{n-1}^{-1} \right)^{1/2} \right],$$

where z_α is such that $1 - \Phi(z_\alpha) = \alpha/2$.

(iii) Notice that $\hat{\tau}_m$ is also a strongly consistent estimator for τ_m on $\{Z_n \rightarrow \infty\}$ for CBPs only verifying (4).

4. Maximum Likelihood Estimators with Incomplete Data

In the previous section, we obtained the MLE of the parameters of interest (p , m , σ^2 , θ , $\mu(\theta)$, and τ_m) based on the sample \mathcal{Z}_n^* . However, in practice, it might be difficult to observe the entire family tree or the variables in \mathcal{Z}_n^* . More realistic would be to suppose that only the total number of individuals and of progenitors of each generation are known, or even only the generation sizes. Notice that, with these two samples, $\hat{\tau}_m$ is the MLE of τ_m (see Remark 3.3(iii)). Hence, we shall focus attention on the rest of the parameters. We shall address the problem of the maximum likelihood estimation under the aforementioned samples as an incomplete data estimation procedure, making use of the EM algorithm and considering \mathcal{Z}_n^* as *hidden* variables. Starting with an initial probability distribution, $p^{(0)}$, and an initial value of the control parameter, $\theta^{(0)}$, we will construct a sequence $\{(p^{(i)}, \theta^{(i)})\}_{i \geq 0}$ that will converge to the MLE of (p, θ) . This iterative method consists of two alternating steps which are iterated until convergence: the E and the M steps. In the E step, the expectation of the complete log-likelihood is calculated using the distribution of the unobserved data. The values of the parameters which maximize this expectation are calculated in the following M step.

4.1. Maximum likelihood estimators based on the sample $\{Z_0, \dots, Z_n, \phi_0(Z_0), \dots, \phi_{n-1}(Z_{n-1})\}$

We shall determine the MLE of the main parameters of the model assuming that only the set of random variables $\bar{\mathcal{Z}}_n = \{Z_0, \dots, Z_n, \phi_0(Z_0), \dots, \phi_{n-1}(Z_{n-1})\}$ can be observed.

Notice that, in accordance with Remark 3.3(iii), the MLEs of m , θ , and $\mu(\theta)$ based on the sample $\bar{\mathcal{Z}}_n$ are \hat{m}_n , $\hat{\theta}_n$, and $\hat{\mu}_n$, respectively. Hence, we shall focus on finding the MLEs of p and σ^2 based on this sample, although we present the method in a general way, considering all the parameters.

4.1.1. The E step

We shall present the E step of the EM algorithm in the $(i + 1)$ -st iteration. For each i , let $p^{(i)} = \{p_k^{(i)}\}_{k \geq 0}$ and $\theta^{(i)}$ be the probability distribution and the control parameter, respectively, obtained in the i -th iteration, and $\mathcal{Z}_n^* | (\bar{\mathcal{Z}}_n, \{p^{(i)}, \theta^{(i)}\})$ the probability distribution of the random vector \mathcal{Z}_n^* given the sample $\bar{\mathcal{Z}}_n$ and the parameters $p^{(i)}$ and $\theta^{(i)}$. For simplicity, in the following, we shall use the notation $E_i^*[\cdot] = E_{\mathcal{Z}_n^* | (\bar{\mathcal{Z}}_n, \{p^{(i)}, \theta^{(i)}\})}[\cdot]$.

In the proof of Theorem 3.1 (see Appendix A), Equation (10) gives the log-likelihood function $\ell(p, \theta | \mathcal{Z}_n^*, \bar{\mathcal{Z}}_n) = \ell(p, \theta | \mathcal{Z}_n^*)$, which depends on the unobserved variables $Z_l(k)$, $0 \leq l \leq n - 1$, $k \geq 0$. The expectation of the log-likelihood with

respect to the distribution $\mathcal{Z}_n^*|(\bar{\mathcal{Z}}_n, \{p^{(i)}, \theta^{(i)}\})$ is:

$$E_i^*[\ell(p, \theta | \mathcal{Z}_n^*, \bar{\mathcal{Z}}_n)] = \Delta_{n-1} \log \theta - \log(A_{Y_{n-1}}(\theta)) + \sum_{l=0}^{n-1} \sum_{k=0}^{\infty} E_i^*[Z_l(k)] \log p_k + E_i^*[K]. \quad (5)$$

Thus, to obtain the value of the above expectation, one has to determine the distribution of \mathcal{Z}_n^* given $\bar{\mathcal{Z}}_n$ when the parameters of the models are $p^{(i)}$ and $\theta^{(i)}$. Since the individuals reproduce independently, and the control distributions are independent of the offspring distribution, one has that, for $z_0, z_{l+1}, \phi_l^*, z_l(k) \in \mathbb{N} \cup \{0\}$, $k \geq 0$, $0 \leq l \leq n-1$ satisfying the constraints $z_{l+1} = \sum_{k=0}^{\infty} k z_l(k)$ and $\phi_l^* = \sum_{k=0}^{\infty} z_l(k)$,

$$\begin{aligned} P[Z_l(k) = z_l(k), 0 \leq l \leq n-1, k \geq 0 | Z_0 = z_0, Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^*, 0 \leq l \leq n-1] &= \\ &= \frac{P[\{Z_0 = z_0\} \cap \bigcap_{l=0}^{n-1} \{Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^*, Z_l(k) = z_l(k), k \geq 0\}]}{P[\{Z_0 = z_0\} \cap \bigcap_{l=0}^{n-1} \{Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^*\}]} \\ &= \prod_{l=0}^{n-1} \frac{P[Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^*, Z_l(k) = z_l(k), k \geq 0 | Z_l = z_l]}{P[Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^* | Z_l = z_l]} \\ &= \prod_{l=0}^{n-1} \frac{P[\sum_{k=0}^{\infty} k Z_l(k) = z_{l+1}, \phi_l(Z_l) = \phi_l^*, Z_l(k) = z_l(k), k \geq 0 | Z_l = z_l]}{P[Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^* | Z_l = z_l]} \\ &= \prod_{l=0}^{n-1} \frac{P[\phi_l(Z_l) = \phi_l^*, Z_l(k) = z_l(k), k \geq 0 | Z_l = z_l]}{P[Z_{l+1} = z_{l+1}, \phi_l(Z_l) = \phi_l^* | Z_l = z_l]} \\ &= \prod_{l=0}^{n-1} \frac{P[\phi_l(z_l) = \phi_l^*, \sum_{i=1}^{\phi_l(z_l)} I_{\{X_{li}=k\}} = z_l(k), k \geq 0]}{P[\sum_{i=1}^{\phi_l^*} X_{li} = z_{l+1}, \phi_l(z_l) = \phi_l^*]} \\ &= \prod_{l=0}^{n-1} \frac{P[\sum_{i=1}^{\phi_l^*} I_{\{X_{li}=k\}} = z_l(k), k \geq 0]}{P[\sum_{i=1}^{\phi_l^*} X_{li} = z_{l+1}]} \\ &= \prod_{l=0}^{n-1} \frac{1}{P[\sum_{i=1}^{\phi_l^*} X_{li} = z_{l+1}]} \cdot \frac{\phi_l^*!}{\prod_{k=0}^{\infty} z_l(k)!} \prod_{k=0}^{\infty} p_k^{(i) z_l(k)}. \end{aligned} \quad (6)$$

Notice that, although the cardinality of the support of the reproduction law may be infinite, for each $0 \leq l \leq n-1$, once z_{l+1} and ϕ_l^* are known, since $z_{l+1} = \sum_{k=0}^{\infty} k z_l(k)$ and $\phi_l^* = \sum_{k=0}^{\infty} z_l(k)$, only a finite number of coordinates of the sequence $\{z_l(k) : k \geq 0\}$ are non-null. From (6), it is clear that to obtain the distribution $\mathcal{Z}_n^*|(\bar{\mathcal{Z}}_n, \{p^{(i)}, \theta^{(i)}\})$, first it is enough to know the distributions $(Z_l(k), k \geq 0) | (Z_l, \phi_l(Z_l), Z_{l+1}, \{p^{(i)}, \theta^{(i)}\})$, for each $l = 0, \dots, n-1$. Now, given a fixed generation, say l , assuming that $Z_l = z_l$, $Z_{l+1} = z_{l+1}$ and $\phi_l(z_l) = \phi_l^*$, it is needed to determine the sample space of the vector $(Z_l(k), k \geq 0)$ taking into account that its possible values $(z_l(k), k \geq 0)$ must verify the constraints $z_{l+1} = \sum_{k=0}^{\infty} k z_l(k)$ and $\phi_l^* = \sum_{k=0}^{\infty} z_l(k)$. After that, their corresponding probabilities must be obtained

following the equation

$$\frac{1}{P[\sum_{i=1}^{\phi_i^*} X_{li} = z_{l+1}]} \cdot \frac{\phi_l^*!}{\prod_{k=0}^{\infty} z_l(k)!} \prod_{k=0}^{\infty} p_k^{(i)z_l(k)}.$$

To this end, it is enough to calculate them from a multinomial distribution of parameters ϕ_l^* and $p^{(i)}$ and normalize the obtained probabilities. From this, it is straightforward to obtain the expected values $E_i^*[Z_l(k)]$, $k \geq 0$. Notice that this distribution does not depend on $\theta^{(i)}$ and hence it has no influence on obtaining $E_i^*[Z_l(k)]$.

4.1.2. The M step

In the M step, one calculates the values of the parameters p and θ which maximize the expectation of the complete log-likelihood, determined in the previous step. In other words, one has to find the values $p^{(i+1)} = \{p_k^{(i+1)}\}_{k \geq 0}$ and $\theta^{(i+1)}$ which maximize the expression (5), subject to the constraints $\sum_{k=0}^{\infty} p_k^{(i+1)} = 1$, $p_k^{(i+1)} \geq 0$, $k \geq 0$.

With a procedure similar to that in the proof of Theorem 3.1 (see Appendix A) to obtain the MLEs based on the entire family tree, one obtains that the values for p and θ in the $(i+1)$ -st iteration are given by

$$p_k^{(i+1)} = \frac{\sum_{l=0}^{n-1} E_i^*[Z_l(k)]}{\sum_{k=0}^{\infty} \sum_{l=0}^{n-1} E_i^*[Z_l(k)]} = \frac{\sum_{l=0}^{n-1} E_i^*[Z_l(k)]}{\sum_{l=0}^{n-1} E_i^*[\sum_{k=0}^{\infty} Z_l(k)]} = \frac{\sum_{l=0}^{n-1} E_i^*[Z_l(k)]}{\Delta_{n-1}}, \quad k \geq 0,$$

and

$$\theta^{(i+1)} = \mu^{-1} \left(\frac{\Delta_{n-1}}{Y_{n-1}} \right).$$

Intuitively, $p_k^{(i+1)}$ represents the ratio of the average number (with respect to the probability distribution determined in the E step) of parents with k offspring to the total number of progenitors. Notice that $\theta^{(i+1)}$ does not depend on the iteration i because it is only based on \bar{Z}_n , which is observed, so that the algorithm reaches the value $\hat{\theta}_n$ at the first iteration and then never leaves it. Hence, as $\theta^{(i)}$ plays no role in calculating $E_i^*[Z_l(k)]$, at each iteration of the algorithm based on \bar{Z}_n only $p_k^{(i)}$ is updated. Nonetheless, we include $\theta^{(i)}$ in the description of the procedure in order for it to be essentially valid in both cases considered: when \bar{Z}_n is observed and when the sample is only $\{Z_0, \dots, Z_n\}$ (we shall deal with the latter case in Subsection 4.2).

Indeed, in general, the values $p^{(i+1)} = \{p_k^{(i+1)}\}_{k \geq 0}$ and $\theta^{(i+1)}$ obtained in the M step are used to begin another E step and the process is repeated until the convergence criterion is satisfied, in which case the process stops, and the final values are obtained, which we shall denote by $\hat{p}_n^{(EM)} = \{\hat{p}_{k,n}^{(EM)}\}_{k \geq 0}$ and $\hat{\theta}_n^{(EM)}$, respectively. When \bar{Z}_n is observed, $\theta^{(i+1)}$ is not needed to begin another E step, and obviously $\hat{\theta}_n^{(EM)} = \hat{\theta}_n$.

It is straightforward to verify the convergence of the algorithm by checking the

conditions given in McLachlan and Krishnan (2008) on the continuity and differentiability of the expectation of the complete log-likelihood function. Consequently, the sequence $\{(p^{(i)}, \theta^{(i)})\}_{i \geq 0}$ converges to the MLE of (p, θ) based on the sample $\bar{\mathcal{Z}}_n$ provided that the likelihood function $\mathcal{L}(p, \theta | \bar{\mathcal{Z}}_n)$ is unimodal.

The EM algorithm also provides the MLE of m , σ^2 , and $\mu(\theta)$ based on $\bar{\mathcal{Z}}_n$ from the estimates obtained for the parameter p :

$$\hat{m}_n^{(EM)} = \sum_{k=0}^{\infty} k \hat{p}_{k,n}^{(EM)}, \quad \hat{\sigma}_n^{2(EM)} = \sum_{k=0}^{\infty} (k - \hat{m}_n^{(EM)})^2 \hat{p}_{k,n}^{(EM)}, \quad \hat{\mu}_n^{(EM)} = \mu(\hat{\theta}_n^{(EM)}).$$

Obviously, $\hat{m}_n^{(EM)} = \hat{m}_n$ and $\hat{\mu}_n^{(EM)} = \hat{\mu}_n$. Indeed, for each $i \geq 0$,

$$m_n^{(i+1)} = \sum_{k=0}^{\infty} k p_k^{(i+1)} = \frac{\sum_{k=0}^{\infty} k \sum_{l=0}^{n-1} E_i^*[Z_l(k)]}{\sum_{k=0}^{\infty} \sum_{l=0}^{n-1} E_i^*[Z_l(k)]} = \frac{Y_n - Z_0}{\Delta_{n-1}} = \hat{m}_n.$$

In summary, and presented in a general way, the method to estimate the parameters p and θ , and consequently m , σ^2 , and $\mu(\theta)$, consists of:

- Step 0 $i = 0$. Choose values $\theta^{(0)}$, $0 \leq p_k^{(0)} \leq 1$, with $\sum_{k=0}^{\infty} p_k^{(0)} = 1$.
- Step 1 *E step*. Based on $p^{(i)}$ and $\theta^{(i)}$
- (a) determine $\mathcal{Z}_n^* | (\bar{\mathcal{Z}}_n, \{p^{(i)}, \theta^{(i)}\})$,
 - (b) calculate $E_i[\ell(p, \theta | \mathcal{Z}_n^*, \bar{\mathcal{Z}}_n)]$.
- Step 2 *M step*. Calculate the values

$$(p^{(i+1)}, \theta^{(i+1)}) = \arg \max_{p, \theta} E_i[\ell(p, \theta | \mathcal{Z}_n^*, \bar{\mathcal{Z}}_n)].$$

- Step 3 If $\max\{|p_k^{(i+1)} - p_k^{(i)}|, k \geq 0, |\theta^{(i+1)} - \theta^{(i)}|\}$ is less than some convergence criterion, the algorithm halts, and the final values are denoted by $\hat{p}_n^{(EM)}$ and $\hat{\theta}_n^{(EM)}$. Otherwise, i is incremented by one unit, and Steps 1-3 are repeated.

4.2. Maximum likelihood estimators based on the sample $\{Z_0, \dots, Z_n\}$

Now, we shall estimate the parameters with reduced sample information, assuming that only the total number of individuals at each generation can be observed. Let us write $\mathcal{Z}_n = \{Z_0, \dots, Z_n\}$. Although we do not know exactly what the control function is like or the values $\phi_0(Z_0), \dots, \phi_{n-1}(Z_{n-1})$, some information on the kind of control we are dealing with is necessary, as will be seen below.

The procedure to obtain the MLE of the model parameters is almost identical to that of the previous case: making use of the EM algorithm, one constructs a

sequence $\{p^{(i)}, \theta^{(i)}\}_{i \geq 0}$ which will converge to the MLE of (p, θ) based on the sample \mathcal{Z}_n .

In this case, to determine the expectation of the log-likelihood in the E step, which is

$$E_i[\ell(p, \theta | \mathcal{Z}_n^*, \mathcal{Z}_n)] = E_i[\Delta_{n-1}] \log \theta - \log(A_{Y_{n-1}}(\theta)) + \sum_{l=0}^{n-1} \sum_{k=0}^{\infty} E_i[Z_l(k)] \log p_k + E_i[K], \quad (7)$$

where now $E_i[\cdot] = E_{\mathcal{Z}_n^* | (\mathcal{Z}_n, \{p^{(i)}, \theta^{(i)}\})}[\cdot]$, one has to know the distribution of \mathcal{Z}_n^* given \mathcal{Z}_n when the parameters are $p^{(i)}$ and $\theta^{(i)}$. In this case, it can be proved that

$$\begin{aligned} P[Z_l(k) = z_l(k), k \geq 0, l = 0, \dots, n-1 | Z_0 = z_0, \dots, Z_n = z_n] &= \\ &= \prod_{l=0}^{n-1} \frac{a_{z_l}(\phi_l^*) \theta^{\phi_l^*} A_{z_l}(\theta)^{-1}}{P[Z_{l+1} = z_{l+1} | Z_l = z_l]} \cdot \frac{\phi_l^*!}{\prod_{k=0}^{\infty} z_l(k)!} \prod_{k=0}^{\infty} p_k^{(i)z_l(k)}, \end{aligned} \quad (8)$$

where $z_0, z_{l+1}, z_l(k) \in \mathbb{N} \cup \{0\}$, $k \geq 0$, $0 \leq l \leq n-1$, satisfying $\sum_{k=0}^{\infty} k z_l(k) = z_{l+1}$, and with $\phi_l^* = \sum_{k=0}^{\infty} z_l(k)$, $0 \leq l \leq n-1$. Equation (8) means that to determine the distribution $\mathcal{Z}_n^* | (\mathcal{Z}_n, \{p^{(i)}, \theta^{(i)}\})$ is enough to know the distributions $(Z_l(k), k \geq 0) | (Z_l, Z_{l+1}, \{p^{(i)}, \theta^{(i)}\})$, $0 \leq l \leq n-1$. Now, for each fixed l , to obtain $(Z_l(k), k \geq 0) | (Z_l, Z_{l+1}, \{p^{(i)}, \theta^{(i)}\})$, and given $Z_l = z_l$ and $Z_{l+1} = z_{l+1}$, first one must consider the possible values for ϕ_l^* , determined from the control distribution with parameters $\theta^{(i)}$ and z_l (notice that, for this purpose, the kind of control distribution of the process has to be known). Then, for each fixed value ϕ_l^* , it is needed to obtain the sample space of the vector $(Z_l(k), k \geq 0)$ taking into account that its possible values $(z_l(k), k \geq 0)$ must verify the constrains $z_{l+1} = \sum_{k=0}^{\infty} k z_l(k)$ and $\phi_l^* = \sum_{k=0}^{\infty} z_l(k)$. Finally their corresponding probabilities are obtained as the product of probabilities from a multinomial distribution with parameters ϕ_l^* and $p^{(i)}$ by the probability that the control distribution takes the value ϕ_l^* (suitably normalized).

The values of the parameters p and θ which maximize the expectation of the complete log-likelihood (7), subject to the constraints $\sum_{k=0}^{\infty} p_k^{(i+1)} = 1$, $p_k^{(i+1)} \geq 0$, $k \geq 0$, are:

$$p_k^{(i+1)} = \frac{\sum_{l=0}^{n-1} E_i[Z_l(k)]}{\sum_{k=0}^{\infty} \sum_{l=0}^{n-1} E_i[Z_l(k)]} = \frac{\sum_{l=0}^{n-1} E_i[Z_l(k)]}{\sum_{l=0}^{n-1} E_i[\sum_{k=0}^{\infty} Z_l(k)]} = \frac{\sum_{l=0}^{n-1} E_i[Z_l(k)]}{E_i[\Delta_{n-1}]}, \quad k \geq 0,$$

and

$$\theta^{(i+1)} = \mu^{-1} \left(\frac{E_i[\Delta_{n-1}]}{Y_{n-1}} \right).$$

We shall denote the final values after applying the algorithm to convergence by $\tilde{p}_n^{(EM)} = \{\tilde{p}_{k,n}^{(EM)}\}_{k \geq 0}$ and $\tilde{\theta}_n^{(EM)}$, respectively.

Again, it can be checked that the conditions given in McLachlan and Krishnan (2008) on the continuity and differentiability of the expectation of the complete

log-likelihood function are satisfied by (7). In this case, the method also provides estimators for m , σ^2 , and $\mu(\theta)$ based on \mathcal{Z}_n :

$$\tilde{m}_n^{(EM)} = \sum_{k=0}^{\infty} k \tilde{p}_{k,n}^{(EM)}, \quad \tilde{\sigma}_n^{2(EM)} = \sum_{k=0}^{\infty} (k - \tilde{m}_n^{(EM)})^2 \tilde{p}_{k,n}^{(EM)}, \quad \tilde{\mu}_n^{(EM)} = \mu(\tilde{\theta}_n^{(EM)}).$$

5. Simulated Example

We shall illustrate the foregoing results with a simulated example. We consider a CBP whose offspring distribution is given by $p_0 = 0.1084$, $p_1 = 0.2709$, $p_2 = 0.3386$, and $p_3 = 0.2822$, and the control variables $\phi_n(k)$ follow binomial distributions with parameters k and $q = 0.6$. Thus, the offspring mean and variance are $m = 1.7946$ and $\sigma^2 = 0.9443$, respectively; $\theta = 1.5$, $\mu(\theta) = 0.6$, and the mean growth rate is $\tau_m = 1.0767$.

In practice, a CBP with control functions of this kind may be useful to model the evolution of a population with the presence of predators. Recall that this binomial control mechanism models situations in which each individual can give birth to offspring in the next generation with probability q , or can be removed from the population, not participating in its subsequent evolution, with probability $1 - q$.

Notice that both θ and $\mu(\theta)$ determine the control distribution when the population size is known. Consequently, we shall focus on the migration parameter $\mu(\theta)$, which in this case is easily interpretable. According to the classification of these models (see Section 2), the process considered in this example is a supercritical CBP with an offspring mean that is also supercritical, i.e., greater than unity. Notice that 40% of the individuals do not participate in the reproduction process for the next generation (i.e., they are taken by predators).

We simulate the first 30 generations of such a process starting with $Z_0 = 1$ individual. The different samples will be denoted by z_{30}^* , \bar{z}_{30} , and z_{30} for that based on the entire family tree, on the individuals and progenitors, and on the population size only, respectively – see the supplementary material. Figure 1 shows the evolution of the number of individuals and progenitors. One sees that the reproduction process makes up for the control process, and, despite the emigration/predators, the process grows. Thus, this path seems to belong to the set $\{Z_n \rightarrow \infty\}$. Under the conditions of the example, in González et al. (2002) and González et al. (2006) it is proved that, on the set $\{Z_n \rightarrow \infty\}$, the process grows exponentially with rate τ_m (hence, the assumption set out in (4) holds).

First, we determined the MLEs and their approximate 95% confidence intervals based on the entire family tree, z_{30}^* , for p , m , σ^2 , $\mu(\theta)$, and τ_m . The estimates are given in Table 1. Figures 2–4 show their behaviours over the course of generations. In these figures we plot the estimates obtained based on the samples restricted to the first n generations, for n going from 0 to 30. One observes that they approach the true values of the parameters, in accordance with Theorems 3.6 and 3.8 and Remark 3.9(ii).

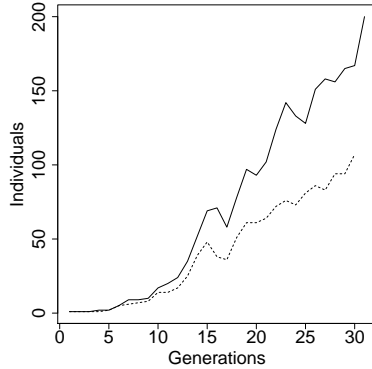


Figure 1: Evolution of the number of individuals (solid line) and the number of progenitors (dashed line).

PARAMETERS								
SAMPLE	p_0	p_1	p_2	p_3	m	σ^2	$\mu(\theta)$	τ_m
z_{30}^*	.1027	.2765	.3389	.2820	1.8002	.9293	.6087	1.0957
\bar{z}_{30}	.1211	.2528	.3308	.2953	1.8002	.9927	.6087	1.0957
z_{30}	.1299	.3083	.3283	.2335	1.6653	.9496	.6579	1.0957
TRUE VALUE	.1084	.2709	.3386	.2822	1.7946	.9443	.6000	1.0767

Table 1: Estimates of the parameters of interest based on the samples considered z_{30}^* , \bar{z}_{30} , and z_{30} .

We shall now illustrate the performance of the EM algorithm in the two situations studied above: using the sample given by the total number of individuals and progenitors in each generation, and the sample given by only the generation sizes. In both cases, assuming that there is no information available about the offspring distribution, only the maximum number of offspring per progenitor, we start the algorithm with a uniform distribution on $\{0, 1, 2, 3\}$ and $\theta^{(0)} = 1/2$. The maximum number of offspring per progenitor in an animal population is a datum that is likely to be known once the reproductive cycle of the corresponding species is understood. Even if this information is unavailable, one can try various (but reasonable) values for this maximum number of offspring per progenitor, and compare the results using the Akaike Information Criterion (AIC) in order to choose the optimal value (we shall illustrate this procedure below).

Using the first sample, individuals plus progenitors, we ran the algorithm until attaining a difference between two consecutive iterations smaller than 10^{-6} , with this convergence occurring from iteration 733 onwards. The resulting estimates are given in Table 1. We repeated this procedure considering samples \bar{z}_j , $j = 1, \dots, 30$, to assess the consistency of the estimates. Figures 5 and 6 (right) show the evolution of these estimates obtained after convergence of the EM algorithm, and based on the samples \bar{z}_j , $j = 1, \dots, 30$ (dashed lines), together with MLEs based on the entire family tree, for the parameters p_k , $k = 0, 1, 2, 3$, and σ^2 . As was mentioned above,

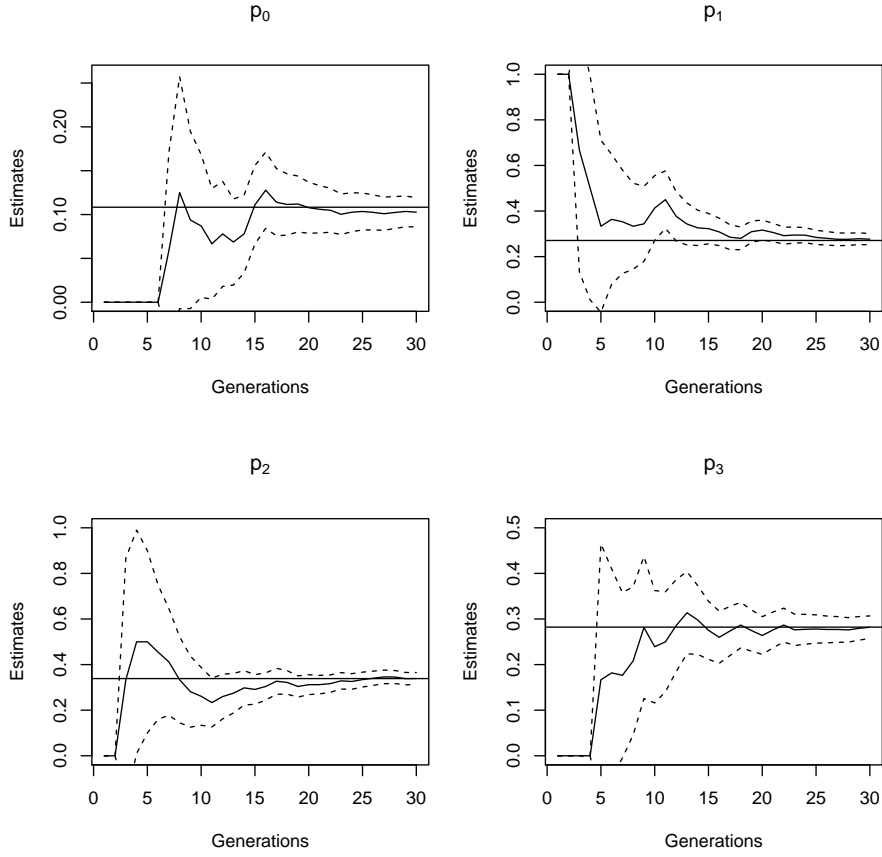


Figure 2: Evolution of the estimates given by \hat{p}_0 , \hat{p}_1 , \hat{p}_2 , and \hat{p}_3 (solid line), and their approximate 95% confidence intervals (dashed line), together with the true value of the parameters (horizontal line).

the EM algorithm is not needed to approximate the MLEs of m , θ , and $\mu(\theta)$ based on the total number of individuals and progenitors in each generation.

We also applied the EM algorithm using the sample defined by only the population sizes, z_{30} . Recall that it is necessary in this case to know the kind of control distribution with which one is working. In practice, this information can come from knowledge of how the population has developed. For instance, if there are predators in the environment, a binomial control distribution would be clearly justified. In the simulation, we observed convergence from iteration 1164 onwards (again for a precision of 10^{-6}). The estimates of the parameters are listed in Table 1 and their temporal evolution is plotted in Figures 5, 6, and 7 (left). One observes in the figures that all the parameters approach their respective true values over the course of the generations.

We studied the influence of the values of $(p^{(0)}, \theta^{(0)})$ on the convergence of the algorithms using discrete sensitivity analysis. The methods were started with 300

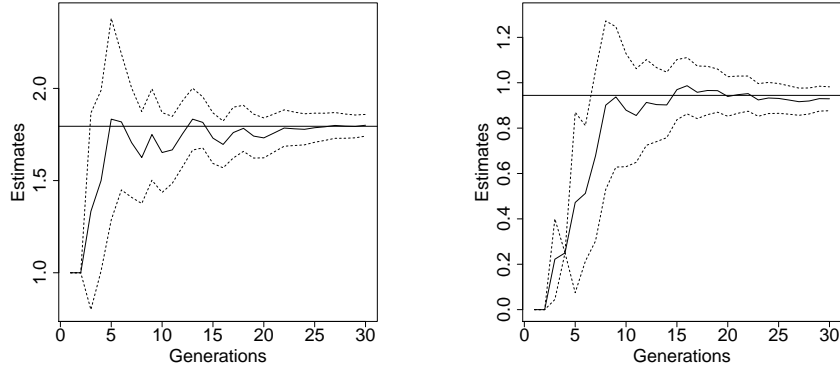


Figure 3: Evolution of the estimates given by \hat{m}_n (left) and $\hat{\sigma}_n^2$ (right) over the course of the generations (solid line) and their approximate 95% confidence intervals (dashed line). The horizontal line represents the true value of each parameter.

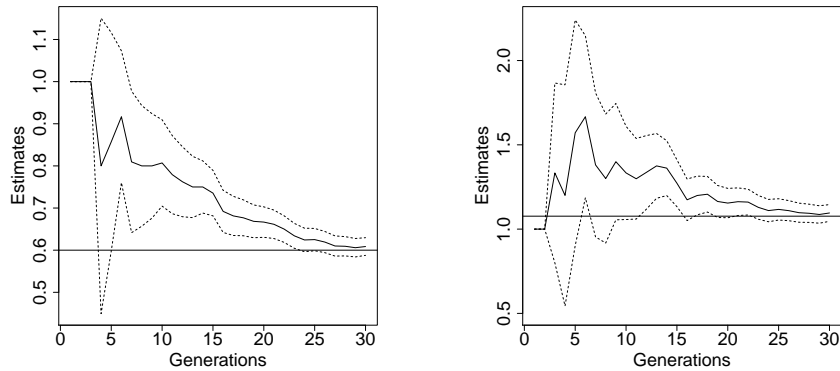


Figure 4: Evolution of the estimates given by $\hat{\mu}_n$ (left) and $\hat{\tau}_m$ (right) over the course of the generations (solid line) and their approximate 95% confidence intervals (dashed line). The horizontal line represents the true value of each parameter.

different random initial values. The distributions $p^{(0)}$ with support $\{0, 1, 2, 3\}$ were randomly chosen from a Dirichlet distribution with all the parameters equal to unity (i.e., by sampling uniformly from the unit simplex), and the values of $\theta^{(0)}$ through the equation $\theta^{(0)} = q^{(0)}(1 - q^{(0)})^{-1}$, with $q^{(0)}$ sampled from a uniform distribution on the open interval $(0, 1)$. Clearly, the EM algorithm based on the sample \bar{z}_{30} is insensitive to such choices. But the EM algorithm based on z_{30} was observed to not be at all robust to the choice of initial values, with convergence to different estimates that could have been local maxima or saddle points. In order to choose the best approximation to the MLE based on z_{30} (which we will call the EM estimate), we propose the following methodological approach.

The log-likelihood function based on the sample \mathcal{Z}_n , denoted by $\ell(p, \theta \mid \mathcal{Z}_n)$, is

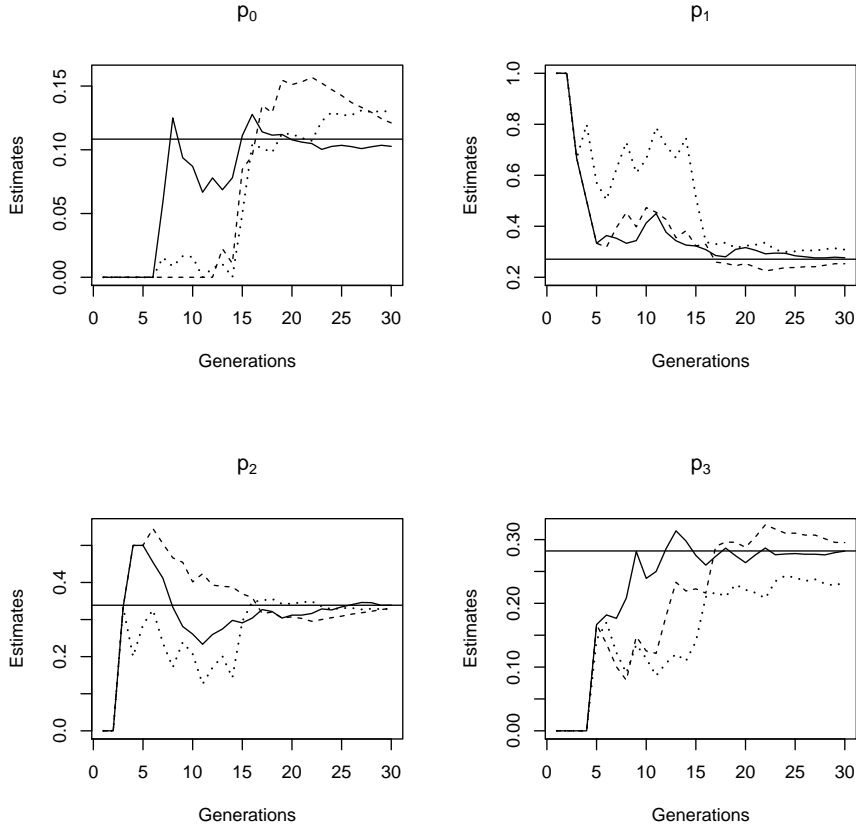


Figure 5: Evolution of the estimates given by \hat{p} (solid line), $\hat{p}^{(EM)}$ (dashed line), and $\tilde{p}^{(EM)}$ (dotted line).

given by the expression

$$\ell(p, \theta \mid Z_l = z_l, l = 0, \dots, n) = \sum_{j=0}^{n-1} \log \left(\sum_{l=0}^{z_j} P_{z_{j+1}}^{*l} \binom{z_j}{l} \frac{\theta^l}{(1+\theta)^{z_j}} \right) \quad (9)$$

with P^{*l} denoting the l -fold convolution of the offspring law p . While maximization of $\ell(p, \theta \mid \mathcal{Z}_n)$ would seem to be intractable using standard methods, (9) can be evaluated for each particular (p, θ) . Our proposal is, therefore, to take as EM estimates of the parameters those associated with the greatest log-likelihood when it is evaluated at the convergence points of the EM algorithm started with different randomly chosen values of the parameters. In our example, the maximum is obtained on the estimates given in Table 1 (see the supplementary material for a further discussion). This methodological strategy can be also followed when the sample is $\bar{\mathcal{Z}}_n$ (if necessary -not for our sample observed, \bar{z}_{30}), taking into account

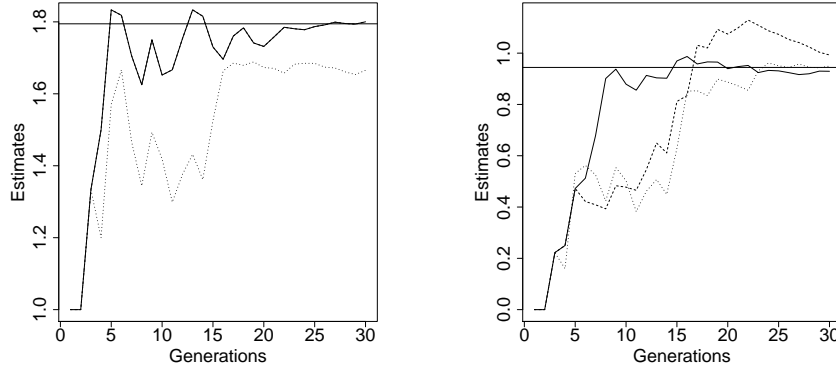


Figure 6: Evolution of the estimates of m (left) and σ^2 (right) based on the entire family tree (solid line), on the total number of individuals and progenitors (dashed line) –for estimates of m , this coincides with the solid line– and on the total number of individuals (dotted line), together with the true value of each parameter (horizontal line).

that

$$\ell(p, \theta \mid Z_l = z_l, \phi_l(Z_l) = \phi_l^*, l = 0, \dots, n-1; Z_n = z_n) = \sum_{j=0}^{n-1} \log \left(P_{z_{j+1}}^{*\phi_l^*} \left(z_j \right) \frac{\theta \phi_l^*}{(1 + \theta)^{z_j}} \right).$$

Moreover, it can be extended to processes with whatever type of control function by only assuming (as has been assumed in our example) knowledge of the kind of control distribution and of the maximum number of offspring per progenitor (denoted by s_{max}). Besides, the possibility of calculating the log-likelihood functions under the samples \mathcal{Z}_n and $\bar{\mathcal{Z}}_n$ allows us to evaluate the influence of the control distribution and of the value of s_{max} , applying the above method with different control distributions and/or values of s_{max} , and using the AIC to compare the resulting models. We have made this study considering the sample \bar{z}_{30} . The results obtained are given in Table 2, in which one observes that for any value of s_{max} , the minimum AIC corresponds to the binomial control distributions. With respect to the influence of s_{max} , the cases $s_{max} = 3$ and 4 led to values that differed little from each other. Considering therefore *parsimonious parametrization*, it would be reasonable to choose $s_{max} = 3$ as optimal. In summary, for problems in which there is no precise knowledge of s_{max} or of the kind of control, a satisfactory procedure would be one like the foregoing, based on comparing in terms of the AIC several fitted models (allowing both expected emigration and expected immigration).

Finally, to approximate the sampling distributions of $\hat{p}_{30}^{(EM)}$, $\tilde{p}_{30}^{(EM)}$, and $\tilde{\theta}_{30}^{(EM)}$ and of their associated parameters, $\hat{\sigma}_{30}^{2(EM)}$, $\tilde{\sigma}_{30}^{2(EM)}$, $\tilde{m}_{30}^{(EM)}$, and $\tilde{\mu}_{30}^{(EM)}$, we applied a bootstrap procedure. We use $\hat{p}_{30}^{(EM)}$ and $\tilde{\theta}_{30}^{(EM)}$, based on \bar{z}_{30} , as parameters to perform a Monte Carlo simulation of 1000 processes up to generation 30. We applied the EM algorithm for each of these bootstrapped samples, obtaining bootstrap approximations to the sampling distributions of $\hat{p}_{30}^{(EM)}$, and consequently of $\hat{\sigma}_{30}^{2(EM)}$.

s_{max}	Iterations	Control distribution					
		Binomial		N Binomial		Poisson	
		Log	AIC	Log	AIC	Log	AIC
3	733	-166.2663	341.2469	-176.1572	361.0288	-170.9058	350.5259
4	4143	-164.8032	340.6973	-174.6942	360.4792	-169.4427	349.9763
5	4244	-164.8032	343.1620	-174.6942	362.9439	-169.4427	352.4410
6	4690	-164.8032	345.7196	-174.6942	365.5015	-169.4427	354.9986

Table 2: Summary of the results for the influence of the control distributions and s_{max} values. Log denotes $\ell(\hat{p}^{(EM)}, \hat{\theta}^{(EM)} | \bar{z}_{30})$. N Binomial denotes the negative binomial distribution. The Iterations column corresponds to the number of iterations needed to attain a precision of 10^{-6} in the EM procedure for the different s_{max} values.

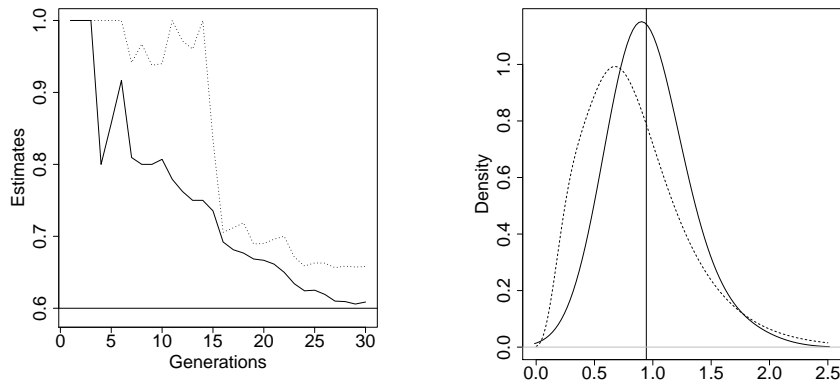


Figure 7: Evolution of the estimates of $\mu(\theta)$ (left) based on the entire family tree (solid line) and on the total number of individuals per generation (dotted line), together with the true value of the parameter (horizontal line). Bootstrap sampling densities of $\hat{\sigma}_{30}^{2(EM)}$ (solid line) and $\tilde{\sigma}_{30}^{2(EM)}$ (dotted line).

Analogously, using the estimates based on the sample z_{30} , we obtained the bootstrap approximations of the sampling distributions of the corresponding estimators. To illustrate these results without excessive repetition, we shall focus on the offspring mean and variance and on the migration parameter. Figure 7 (right) shows the bootstrap sampling distributions of $\hat{\sigma}_{30}^{2(EM)}$ and $\tilde{\sigma}_{30}^{2(EM)}$. One observes that the distribution of $\tilde{\sigma}_{30}^{2(EM)}$ is more variable than that of $\hat{\sigma}_{30}^{2(EM)}$. This is a consequence of the lack of information represented by the control variables not being observed. Figure 8 shows the joint distribution of $(\tilde{m}_{30}^{(EM)}, \tilde{\mu}_{30}^{(EM)})$ and its marginal distributions. One observes how strongly these estimators are related, with their being distributed around the curve given by $\tau_m = m\mu(\theta) = 1.0767$.

Based on the foregoing bootstrap distributions, one can calculate the bootstrap estimates of the mean squared error (MSE) of the respective estimators based on the samples \bar{z}_{30} and z_{30} , and compare the accuracy of the different estimators through their relative efficiency (eff) (Table 3). One observes from the table that the es-

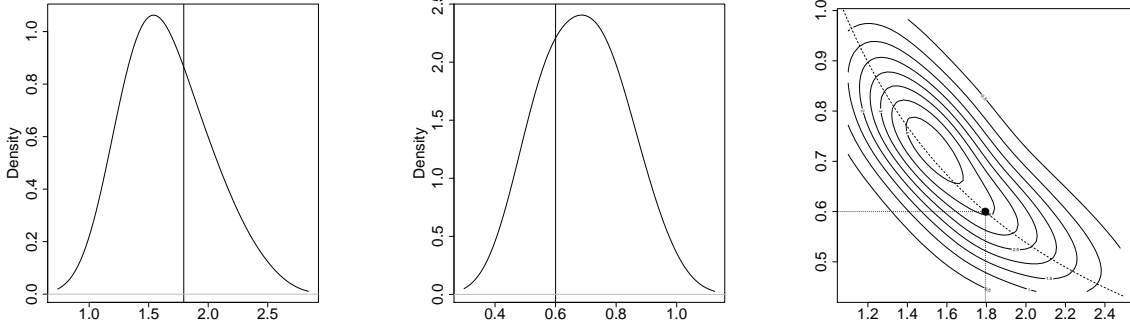


Figure 8: Bootstrap sampling densities of $\tilde{m}_{30}^{(EM)}$ (left) and $\tilde{\mu}_{30}^{(EM)}$ (center) and joint distribution of $(\tilde{m}_{30}^{(EM)}, \tilde{\mu}_{30}^{(EM)})$ with the curve $m\mu(\theta) = 1.0767$ (right), together with the true values of the parameters.

	p_0	p_1	p_2	p_3	m	σ^2	$\mu(\theta)$
MSE based on \bar{z}_{30}	0.0081	0.0324	0.0332	0.0085	0.0022	0.0808	0.0004
MSE based on z_{30}	0.0136	0.0613	0.0637	0.0260	0.1243	0.1438	0.0247
eff	1.6779	1.8906	1.9208	3.0594	56.5083	1.7790	62.6719

Table 3: Efficiency of the estimators based on \bar{z}_{30} relative to the estimators based on z_{30} for the parameters of interest.

imates provided by assuming observation of \bar{z}_{30} are preferable to those assuming observation of z_{30} . This can be understood as a reflection of the greater information content of the former of the two samples.

Computational complexity

With the aim of determining the order of the computational complexity of each iteration of the two EM algorithms proposed, we evaluate the number of operations needed to obtain $E_i[Z_l(k)]$ and $E_i^*[Z_l(k)]$, $l = 0, 1, \dots, n-1$; $k = 0, 1, \dots, s_{max}$, respectively (recall s_{max} is the maximum number of offspring per progenitor).

Let $E^{(i)} = (E_i[Z_l(k)])_{0 \leq l \leq n-1; 0 \leq k \leq s_{max}}$ and $E^{*(i)} = (E_i^*[Z_l(k)])_{0 \leq l \leq n-1; 0 \leq k \leq s_{max}}$. Considering the sample \bar{Z}_n , let B_l be the matrix storing the tree associated to the transition from $\phi_l(Z_l)$ to Z_{l+1} , that is, it stores by rows the possible values of the vector $(Z_l(0), \dots, Z_l(s_{max}))$ such that $\sum_{k=0}^{s_{max}} Z_l(k) = \phi_l(Z_l)$ and $\sum_{k=0}^{s_{max}} k Z_l(k) = Z_{l+1}$, $l = 0, 1, \dots, n-1$. Let us denote b_l the number of rows of B_l , $l = 0, 1, \dots, n-1$. Finally, for $l = 0, 1, \dots, n-1$, let P_l be a row vector whose elements are the probabilities of each row of B_l , obtained by equation (6), that is, if $\phi_l(Z_l) = \phi_l^*$ and $Z_{l+1} = z_{l+1}$, the corresponding element of P_l for the row of B_l given by $(z_l(0), \dots, z_l(s_{max}))$ is equal to

$$\frac{1}{P[\sum_{i=1}^{\phi_l^*} x_{li} = z_{l+1}]} \frac{\phi_l^*!}{\prod_{k=0}^{s_{max}} z_l(k)!} \prod_{k=0}^{s_{max}} p_k^{(i)z_l(k)}.$$

Then, the l -th row of $E^{(i)}$ is equal to the product $P_l \cdot B_l$, $l = 0, 1, \dots, n - 1$.

Analogously, assuming the sample \mathcal{Z}_n , for $l = 0, 1, \dots, n - 1$, let B_l^* be the matrix storing the tree associated to the transition from Z_l to Z_{l+1} , that is, its rows store all the possible values of the vector $(Z_l(0), \dots, Z_l(s_{max}))$ that allow reaching Z_{l+1} from Z_l . To obtain such a matrix, if $Z_l = z_l$ and $Z_{l+1} = z_{l+1}$, we consider for each possible value of $\phi_l(z_l)$, say ϕ_l^* , every vector $(z_l(0), \dots, z_l(s_{max}))$ such that $\sum_{k=0}^{s_{max}} z_l(k) = \phi_l^*$ and $\sum_{k=0}^{s_{max}} k z_l(k) = z_{l+1}$, $l = 0, 1, \dots, n - 1$. Now, for each one of these vectors we obtain the probabilities (see equation (8))

$$\frac{a_{z_l}(\phi_l^*) \theta^{\phi_l^*} A_{z_l}(\theta)^{-1}}{P[Z_{l+1} = z_{l+1} | Z_l = z_l]} \frac{\phi_l^*!}{\prod_{k=0}^{s_{max}} z_l(k)!} \prod_{k=0}^{s_{max}} p_k^{(i)z_l(k)},$$

which are ordered in the row vector P_l^* . Then, the l -th row of $E^{*(i)}$ is equal to the product $P_l^* \cdot B_l^*$, $l = 0, 1, \dots, n - 1$. Let us denote by b_l^* the number of the rows of B_l^* , $l = 0, 1, \dots, n - 1$.

Hence, for each iteration of both methods we can determine the order of the computational complexity as $s_{max} \sum_{l=0}^{n-1} b_l$ and $s_{max} \sum_{l=0}^{n-1} b_l^*$, respectively. Now, for each $l = 0, 1, \dots, n - 1$, b_l depends on the values of s_{max} , $\phi_l(Z_l)$ and Z_{l+1} , and b_l^* on s_{max} , Z_l and Z_{l+1} , but it is not possible to obtain closed forms of them. To obtain an upper bound of b_l one can obtain the dimension of the biggest transition tree. In the case of binomial control functions, this tree can be generated by considering $\phi_l(Z_l) = Z_l$ (the maximum number of progenitors). Clearly, the dimension of this tree is greater than or equal to that of the one that leads to Z_{l+1} . By an empirical study (see supplementary material for details) we have determined that $b_l = O(Z_l^{s_{max}-1})$. In a similar way, an upper bound of b_l^* is given by the dimension of the biggest tree that can be generated by Z_l individuals under the lack of awareness of the exact number of progenitors $\phi_l(Z_l)$. Again, we have determined empirically (see supplementary material for details) that, in the case of binomial control functions, $b_l^* = O(Z_l^{s_{max}})$. This fact allows us to compare, at least roughly, the computational complexity of both methods, indicating that for a generation of size z , one needs to generate trees of dimension z times greater when only the population size is observed than when the number of progenitors, $\phi_l(z)$, is also available. Figure 9 reveals this fact in our numerical example. This implies that the EM procedure requires much more time in each iteration when storing only \mathcal{Z}_n , compared to when storing $\bar{\mathcal{Z}}_n$. In particular, in our example, the same number of iterations of the procedure required a factor of 128 less time under the sample with observation of offspring and progenitors than under the sample based only on generation sizes. Also, the second of these two procedures needed more iterations to reach convergence. Hence, as was to be expected due to the relative loss of information, the second method is far more costly computationally than the first (by a factor of roughly 170 for a precision of 10^{-6} , in terms of computational time). Moreover, this second procedure involves post-processing which involves running it several times for different seeds, and evaluating the exact likelihood at the convergence points.

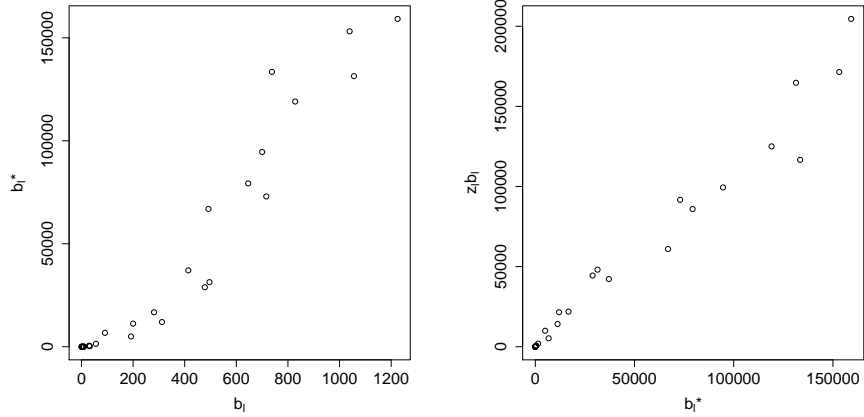


Figure 9: Evolution of (b_l, b_l^*) -left- and $(b_l^*, z_l b_l)$ -right-, $l = 0, \dots, 29$, given the samples \bar{z}_{30} and z_{30} , and by considering $s_{max} = 3$ and binomial control distributions.

Remark 5.1. *The example simulations were performed by parallel computing using the **R** statistical software environment (see R Core Team (2014)). For the estimator density and the exact log-likelihood function calculations, we used the `sm` and `polynom` packages (see Bowman and Azzalini (2014) and Venables et al. (2014)), respectively.*

6. Concluding Remarks

We have studied the maximum likelihood estimation of the main parameters of the CBP with random control function considering a nonparametric framework for the offspring distribution and a parametric scheme for the control process. First, assuming the entire family tree is observable, we determined the MLEs of the parameters associated with the offspring distribution and with the control law, and established their consistency and limiting normality. These results generalized those that had been obtained for the parameters associated with the offspring law for CBPs with deterministic control function. We also provided new results on the estimation of the control and migration parameters, with particular note made of their asymptotic properties.

Since in practice it is difficult to observe the entire family tree, we considered two more realistic situations, one assuming that the only observable data are the total number of individuals and progenitors in each generation, \bar{Z}_n , and the other that even only the generation sizes are observable, Z_n . In both cases, we addressed the problem of obtaining the MLEs of the main parameters of the model by an incomplete data estimation procedure. To this end, we made use of the EM algorithm. A simulated example showed that this seems to work appropriately based on the sample \bar{Z}_n . Based on the sample Z_n , we encountered the problem that the algorithm may converge to local maxima or saddle points. In such a case, we proposed running the algorithm with a large number of different starting values, and choosing the ones

associated with the highest value of the log-likelihood function (this function can be evaluated although it can not be maximized by standard methods). The simulated example showed this methodological strategy to also work adequately. The procedure based on knowledge of the total numbers of individuals and progenitors converges rapidly, providing adequate accuracy with reasonably short computation times. Storing only \mathcal{Z}_n however, we found the EM algorithm to require not only much more time for each iteration but also more iterations to reach convergence (with the same precision).

In the simulated example, we also illustrated the consistency of the estimates based on the three samples. (The only case established theoretically in the paper was that corresponding to observing the entire family tree.) We then used a bootstrapping approach to get approximations to the sampling distributions of the estimators obtained by the EM algorithm, finding that the more information that the samples contained, the smaller was the variability of the estimator.

Acknowledgements

The authors thank the referee for her/his careful reading of our paper and for her/his constructive comments which have improved its presentation. Also, the authors would like to thank Horacio González-Velasco and Carlos García-Orellana for providing them with computational support. Research supported by the Ministerio de Educación, Cultura y Deporte (grant FPU13/03213), the Ministerio de Economía y Competitividad and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2012-31235.

This version of the article has been accepted for publication, after peer review and is subject to Elsevier's terms of use for *Computational Statistics & Data Analysis*, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <https://doi.org/10.1016/j.csda.2015.01.015>.

Appendix A. Proof of Theorem 3.1

It is immediate to verify that the likelihood function based on \mathcal{Z}_n^* is:

$$\begin{aligned} \mathcal{L}(p, \theta | Z_l(k) = z_l(k), 0 \leq l \leq n-1; k \geq 0) &= \\ &= \theta^{\sum_{l=0}^{n-1} \phi_l^*} A_{\sum_{l=0}^{n-1} z_l}(\theta)^{-1} \prod_{l=0}^{n-1} \frac{\phi_l^*! a_{z_l}(\phi_l^*)}{\prod_{k=0}^{\infty} z_l(k)!} \prod_{k=0}^{\infty} p_k^{z_l(k)}, \end{aligned}$$

where $\phi_l^* = \sum_{k=0}^{\infty} z_l(k)$. Consequently, the log-likelihood function based on \mathcal{Z}_n^* is:

$$\ell(p, \theta | \mathcal{Z}_n^*) = f(p) + g(\theta) + K, \quad (10)$$

with $f(p) = \sum_{l=0}^{n-1} \sum_{k=0}^{\infty} Z_l(k) \log p_k$, $g(\theta) = \Delta_{n-1} \log \theta - \log(A_{Y_{n-1}}(\theta))$ and K some positive random variable whose expression does not depend on p or θ .

From (10), one has to maximize $f(p) + g(\theta)$ subject to the constraints $\sum_{k=0}^{\infty} p_k = 1$, $p_k \geq 0$, $k \geq 0$. Using the non-negativity of the Kullback-Leibler divergence, it can be verified that the value of p which maximizes the function $f(p)\Delta_{n-1}$, and hence, $f(p)$, is

$$\hat{p}_{k,n} = \frac{\sum_{l=0}^{n-1} Z_l(k)}{\sum_{l=0}^{n-1} \sum_{k=0}^{\infty} Z_l(k)} = \frac{Y_{n-1}(k)}{\Delta_{n-1}}, \quad k \geq 0.$$

Moreover, it can easily be shown that

$$\hat{\theta}_n = \mu^{-1} \left(\frac{\Delta_{n-1}}{Y_{n-1}} \right)$$

is an extremum of the function $g(\theta)$. Taking into account that

$$\left. \frac{d^2 g(\theta)}{d\theta^2} \right|_{\theta=\hat{\theta}_n} = - \left. \frac{\Delta_{n-1} - \varepsilon(Y_{n-1}, \theta) + \sigma^2(Y_{n-1}, \theta)}{\theta^2} \right|_{\theta=\hat{\theta}_n} < 0,$$

one has that $\hat{\theta}_n$ is the maximum of $g(\theta)$ and then $(\hat{p}_n, \hat{\theta}_n)$ maximizes $f(p) + g(\theta)$.

Appendix B. Proof of Theorem 3.6

(i) We shall prove that \hat{p}_k is strongly consistent for p_k , making use of a strong law of large numbers for martingales. We shall fix $k \geq 0$, and prove that, as $n \rightarrow \infty$,

$$\hat{p}_k = \frac{1}{\sum_{j=0}^{n-1} \phi_j(Z_j)} \sum_{i=1}^n \sum_{j=1}^{\phi_{i-1}(Z_{i-1})} I_{\{X_{i-1j}=k\}} \rightarrow p_k \quad a.s. \text{ on } \{Z_n \rightarrow \infty\}. \quad (11)$$

For simplicity, we will consider $P[Z_n \rightarrow \infty] = 1$. For each $i = 1, 2, \dots$, let

$$\begin{aligned} V_i(k) &= \sum_{j=1}^{\phi_{i-1}(Z_{i-1})} (I_{\{X_{i-1j}=k\}} - p_k), \\ \mathcal{H}_i &= \sigma(X_{l-1j}, \phi_{l-1}(k) : 1 \leq l \leq i, j \geq 1, k \geq 0). \end{aligned}$$

It is verified that $\{V_i(k), \mathcal{H}_i\}_{i \geq 0}$ is a martingale difference. In these terms, $\hat{p}_k - p_k = \Delta_{n-1}^{-1} \sum_{i=1}^n V_i(k)$.

For each $n \geq 0$, let $U_n = Y_{n-1}$. Then, taking into account Proposition 3.5(iv), to prove (11) one only needs to obtain that, as $n \rightarrow \infty$,

$$U_n^{-1} \sum_{i=1}^n V_i(k) \rightarrow 0 \quad a.s. \quad (12)$$

Since $U_n \rightarrow \infty$, to prove (12), using Theorem 2.18 in Hall and Heyde (1980), it is enough to verify that $\sum_{i=1}^{\infty} U_i^{-2} E[|V_i(k)|^2 | \mathcal{H}_{i-1}] < \infty$ *a.s.* Now, let $M =$

$\sup_{n \geq 0} \Delta_n Y_n^{-1} < \infty$ *a.s.* and $N = \sup_{n \geq 0} \varepsilon(Z_n) \phi_n(Z_n)^{-1} < \infty$ *a.s.* (guaranteed by Proposition 3.5(iv) and (vi), respectively). Then, one has

$$\begin{aligned}
\sum_{i=1}^{\infty} U_i^{-2} E[|V_i(k)|^2 | \mathcal{H}_{i-1}] &= \sum_{i=1}^{\infty} \frac{E \left[\text{Var} \left[\sum_{j=1}^{\phi_{i-1}(Z_{i-1})} I_{\{X_{i-1j}=k\}} \middle| \phi_{i-1}(Z_{i-1}) \right] \right]}{Y_{i-1}^2} \\
&= \sum_{i=1}^{\infty} \frac{\varepsilon(Z_{i-1}) p_k (1 - p_k)}{Y_{i-1}^2} \\
&= p_k (1 - p_k) \sum_{i=1}^{\infty} \frac{\varepsilon(Z_{i-1})}{\phi_{i-1}(Z_{i-1})} \cdot \frac{\phi_{i-1}(Z_{i-1})}{\Delta_{i-1}^2} \cdot \left(\frac{\Delta_{i-1}}{Y_{i-1}} \right)^2 \\
&\leq p_k (1 - p_k) N M^2 \sum_{i=1}^{\infty} \frac{1}{\phi_{i-1}(Z_{i-1})} < \infty \quad \textit{a.s.},
\end{aligned}$$

where the last inequality is true due to $\phi_{i-1}(Z_{i-1}) \leq \Delta_{i-1}$, $i \geq 1$, and Proposition 3.5(ii).

The strong consistency of \widehat{m}_n is a direct consequence of Proposition 3.5(ii)-(v).

Taking into account that both \widehat{m}_n and \widehat{p}_k are strongly consistent for m and p_k , respectively, on $\{Z_n \rightarrow \infty\}$, it is deduced that $\widehat{\sigma}_n^2$ is strongly consistent for σ^2 .

(ii) The key to proving (ii) (a) and (b) is to rewrite

$$\widehat{p}_k - p_k \stackrel{d}{=} \frac{1}{\Delta_{n-1}} \sum_{l=1}^{\Delta_{n-1}} (I_{\{X_l=k\}} - p_k), \quad \widehat{m}_n - m \stackrel{d}{=} \frac{1}{\Delta_{n-1}} \sum_{l=1}^{\Delta_{n-1}} (X_l - m),$$

with $\stackrel{d}{=}$, as one recalls, denoting equal in distribution, and $\{X_l\}_{l \geq 1}$ being a sequence of i.i.d. random variables with common distribution being the offspring distribution. The results are derived by applying a central limit theorem for random sums as was done, *mutatis mutandis*, in the proofs of Theorems 3.2 and 4.2 in González et al. (2004) for CBPs with deterministic control function.

Finally, to prove (ii) (c), we adapt the proof established in Theorem 3.1 in González et al. (2005a) for CBPs with deterministic control function. We here provide just a brief scheme. The result is firstly proved for $\sum_{k=0}^{\infty} (k - m)^2 \widehat{p}_k$, i.e., when m is considered known. Due to the fact that one can write

$$\sum_{k=0}^{\infty} (k - m)^2 \widehat{p}_k \stackrel{d}{=} \frac{1}{\Delta_{n-1}} \sum_{l=1}^{\Delta_{n-1}} (X_l - m)^2,$$

the result holds by using the central limit theorem cited above, following similar steps to those in the proof of Theorem 3.1 in González et al. (2005a). Now, notice that

$\sum_{k=0}^{\infty} (k-m)^2 \widehat{p}_k - \widehat{\sigma}_n^2 = (\widehat{m}_n - m)^2$, so that, by considering (ii) (b), $\sigma^2 \Delta_{n-1}^{-1/2} \xrightarrow{P'} 0$, and Slutsky's Theorem, one has

$$\left(\sum_{k=0}^{\infty} (k-m)^2 \widehat{p}_k - \widehat{\sigma}_n^2 \right) \Delta_{n-1}^{1/2} \xrightarrow{P'} 0, \text{ as } n \rightarrow \infty.$$

Hence, together with the fact that the result holds when m is known and Slutsky's Theorem, (ii) (c) follows.

Appendix C. Proof of Theorem 3.8

(i) This is immediate from Proposition 3.5(iv).

(ii) (a) For simplicity, we shall suppose $P[Z_n \rightarrow \infty] = 1$. Let $D_i = \phi_{i-1}(Z_{i-1}) - \mu(\theta)Z_{i-1}$ and $\mathcal{F}_i = \sigma(Z_0, \dots, Z_i, \phi_0(Z_0), \dots, \phi_{i-1}(Z_{i-1}))$, $i = 1, \dots, n$, $n = 1, 2, \dots$. We have

$$\begin{aligned} Y_{n-1}^{1/2} (\widehat{\mu} - \mu(\theta)) &= Y_{n-1}^{-1/2} \left[\sum_{i=1}^n \left((Z_{i-1} + 1)^{1/2} - (\tau_m^{i-1} W)^{1/2} \right) \frac{D_i}{(Z_{i-1} + 1)^{1/2}} \right. \\ &\quad \left. + W^{1/2} \sum_{i=1}^n \tau_m^{(i-1)/2} \frac{D_i}{(Z_{i-1} + 1)^{1/2}} \right] \end{aligned}$$

with W being the limit variable introduced in (4)(c). Taking into account $\tau_m^{-n} Y_{n-1} \rightarrow (\tau_m - 1)^{-1} W$ a.s. as $n \rightarrow \infty$, it follows that it is enough to prove

$$(I) = \sum_{i=1}^n \left((Z_{i-1} + 1)^{1/2} - (\tau_m^{i-1} W)^{1/2} \right) \frac{D_i}{(Z_{i-1} + 1)^{1/2}} = o_P(\tau_m^{n/2}) \quad (13)$$

and

$$(II) = (\tau_m - 1)^{1/2} \sum_{i=1}^n \tau_m^{-(n-i+1)/2} \frac{D_i}{(Z_{i-1} + 1)^{1/2}} \xrightarrow{d} N(0, \theta \mu'(\theta)), \quad (14)$$

as $n \rightarrow \infty$, with $o_P(\cdot)$ denoting the stochastic order analogue of $o(\cdot)$ (i.e., write $X_n = o_P(Y_n)$ to mean $P(|X_n| \geq \epsilon |Y_n|) \rightarrow 0$, as $n \rightarrow \infty$, for each $\epsilon > 0$). The proof follows similar steps to those given in Sriram et al. (2007), Theorem 2. For each $n \geq 0$, let

$$A_n = \sum_{i=1}^n \tau_m^{(i-1)/2} \left(\left(\frac{Z_{i-1} + 1}{\tau_m^{i-1}} \right)^{1/2} - W^{1/2} \right)^2 \quad \text{and} \quad B_n = \sum_{i=1}^n \tau_m^{(i-1)/2} \frac{D_i^2}{Z_{i-1} + 1}.$$

Then, applying the Cauchy-Schwarz inequality, $|(I)| \leq A_n^{1/2} B_n^{1/2}$. By (4)(c), one obtains $(\tau_m^{-(i-1)} (Z_{i-1} + 1))^{1/2} - W^{1/2} \rightarrow 0$ a.s., and consequently, using the Stolz-Cèsaro Lemma, $A_n = o_P \left(\sum_{i=1}^n \tau_m^{(i-1)/2} \right) = o_P \left(\tau_m^{n/2} \right)$. Now, using

$$E[D_i^2 | \mathcal{F}_{i-1}] = \theta \mu'(\theta) Z_{i-1}, \quad i \geq 1, \quad (15)$$

one has that $E[B_n] = O\left(\sum_{i=1}^n \tau_m^{(i-1)/2}\right) = O\left(\tau_m^{n/2}\right)$, as $n \rightarrow \infty$, which implies that $|B_n| = O_P\left(\sum_{i=1}^n \tau_m^{(i-1)/2}\right) = O_P\left(\tau_m^{n/2}\right)$ as $n \rightarrow \infty$, with $O_P(\cdot)$ denoting the stochastic order analogue of $O(\cdot)$ (i.e., write $X_n = O_P(Y_n)$ to mean: for each $\epsilon > 0$ there exists a real number M such that $P(|X_n| \geq M|Y_n|) < \epsilon$ if n is large enough). Hence (13) follows.

To establish (14), let $\gamma_{nj} = D_{n-j+1}(Z_{n-j} + 1)^{-1/2}$, $j = 1, \dots, n$. Then

$$\begin{aligned} (\tau_m - 1)^{1/2} \sum_{i=1}^n \tau_m^{-(n-i+1)/2} \frac{D_i}{(Z_{i-1} + 1)^{1/2}} &= (\tau_m - 1)^{1/2} \sum_{j=1}^n \tau_m^{-j/2} \frac{D_{n-j+1}}{(Z_{n-j} + 1)^{1/2}} \\ &= (\tau_m - 1)^{1/2} \left(\sum_{j=1}^J \tau_m^{-j/2} \gamma_{nj} + \sum_{j=J+1}^n \tau_m^{-j/2} \gamma_{nj} \right) \\ &= U_{Jn} + (\tau_m - 1)^{1/2} \sum_{j=J+1}^n \tau_m^{-j/2} \gamma_{nj} = U_{nn}, \end{aligned} \quad (16)$$

with $U_{Jn} = (\tau_m - 1)^{1/2} \sum_{j=1}^J \tau_m^{-j/2} \gamma_{nj}$, $J = 1, \dots, n$. For $J \geq 1$ and given $(t_1, \dots, t_J) \in \mathbb{R}^J$, it can be proved, using analogous arguments to those given in the proof of Theorem 1 in Heyde and Brown (1971), jointly with the condition assumed in (ii), that

$$E \left[e^{i \sum_{j=1}^J t_j \tau_m^{-j/2} \gamma_{nj}} \right] \rightarrow e^{-\frac{1}{2} \theta \mu'(\theta) \sum_{j=1}^J t_j^2 \tau_m^{-j}}, \text{ as } n \rightarrow \infty.$$

Consequently, for each $J = 1, \dots, n$, the vector $(\tau_m^{-1/2} \gamma_{n1}, \dots, \tau_m^{-J/2} \gamma_{nJ})$ is asymptotically multivariate normal as $n \rightarrow \infty$, and therefore $U_{Jn} \xrightarrow{d} U_J$, with U_J following a $N(0, \theta \mu'(\theta) (\tau_m - 1) \sum_{j=1}^J \tau_m^{-j})$. Now, from Chebyshev's inequality, (15), and (16), for every $n \geq 0$ and $\epsilon > 0$, one has $P[|U_{Jn} - U_{nn}| > \epsilon] \leq \epsilon^{-2} (\tau_m - 1) \theta \mu'(\theta) \sum_{j=J+1}^{\infty} \tau_m^{-j}$. In consequence, for some constant k_0 ,

$$\limsup_{n \rightarrow \infty} P[|U_{Jn} - U_{nn}| > \epsilon] \leq k_0 \sum_{j=J+1}^{\infty} \tau_m^{-j} \rightarrow 0, \text{ as } J \rightarrow \infty.$$

Therefore, from the fact that $U_J \xrightarrow{d} N(0, \theta \mu'(\theta))$ as $J \rightarrow \infty$, and Theorem 25.5 in Billingsley (1979), one obtains

$$U_{nn} \xrightarrow{d} N(0, \theta \mu'(\theta)),$$

as $n \rightarrow \infty$, and hence (14) is proved.

(ii) (b) This is proved with identical arguments to those of (ii) (a), setting in this case $D_i = Z_i - \tau_m Z_{i-1}$ and $\mathcal{F}_i = \sigma(Z_0, \dots, Z_i)$, $i = 1, \dots, n$, $n = 1, 2, \dots$. Consequently, $E[D_i^2 | \mathcal{F}_{i-1}] = (\sigma^2 \mu(\theta) + m^2 \theta \mu'(\theta)) Z_{i-1}$, $i \geq 1$, and now it is verified that

$$E \left[e^{i \sum_{j=1}^J t_j \tau_m^{-j/2} \gamma_{nj}} \right] \rightarrow e^{-\frac{1}{2} (\sigma^2 \mu(\theta) + m^2 \theta \mu'(\theta)) \sum_{j=1}^J t_j^2 \tau_m^{-j}}, \text{ as } n \rightarrow \infty.$$

References

- Bercu, B., 1999. Weighted estimation and tracking for Bienaymè–Galton–Watson processes with adaptive control. *Statistics & Probability Letters* 42, 415–421.
- Bernhardt, P.H., Zhang, D., Wang, H.J. 2015. A fast EM algorithm for fitting joint models of a binary response and multiple longitudinal covariates subject to detection limits. *Computational Statistics & Data Analysis*, 85, 37–53.
- Billingsley, P., 1979. *Probability and Measure*. Wiley and Sons, Inc.
- Bowman, A., Azzalini, A., 2014. R package *sm*: nonparametric smoothing methods (version 2.2-5.4). University of Glasgow, UK and Università di Padova, Italia.
- Bruss, F., Slavtchova-Bojkova, M., 1999. On waiting times to populate an environment and a question of statistical inference. *Journal of Applied Probability* 36, 261–267.
- Daskalova, N., 2014. *Nonlinear Dynamics of Electronic Systems*. Vol. 438 of *Communications in Computer and Information Science*. Springer, Ch. EM Algorithm for Estimation of the Offspring Probabilities in Some Branching Models, pp. 181–188.
- del Puerto, I., Yanev, N., 2008. *Leading-Edge Applied Mathematical Modeling Research*. Nova Science Publishers, Inc, Ch. 11: Branching Processes with Multitype Random Control Functions: subcritical case, pp. 363–374.
- Dion, J. P., Essebbar, B., 1995. On the statistics of controlled branching processes. *Lecture Notes in Statistics* 99, 14–21.
- González, M., del Puerto, I., 2012. Diffusion approximation of an array of controlled branching processes. *Methodology and Computing in Applied Probability* 14, 843–861.
- González, M., Gutiérrez, C., Martínez, R., 2012. Expectation-Maximization algorithm for determining natural selection of Y-linked genes through two-sex branching processes. *Journal of Computational Biology* 19, 1015–1026.
- González, M., Martínez, R., del Puerto, I., 2004. Nonparametric estimation of the offspring distribution and mean for a controlled branching process. *Test* 13, 465–479.
- González, M., Martínez, R., del Puerto, I., 2005a. Estimation of the variance for a controlled branching process. *Test* 14, 199–213.
- González, M., Molina, M., del Puerto, I., 2002. On the class of controlled branching process with random control functions. *Journal of Applied Probability* 39, 804–815.

- González, M., Molina, M., del Puerto, I., 2005b. Asymptotic behaviour of the critical controlled branching processes with random control functions. *Journal of Applied Probability* 42, 463–477.
- González, M., Molina, M., del Puerto, I., 2006. Geometric growth for stochastic difference equations with application to branching populations. *Bernoulli* 12 (5), 931–942.
- Guttorp, P., 1991. *Statistical inference for branching processes*. John Wiley and Sons, Inc.
- Haccou, P., Jagers, P., Vatutin, V., 2005. *Branching processes: variation, growth and extinction of populations*. Cambridge University Press.
- Hall, C., Heyde, C. C., 1980. *Martingale limit theory and its applications*. Academic Press, Inc.
- Hautphenne, S., Fackrell, M., 2014. An EM algorithm for the model fitting of Markovian binary trees. *Computational Statistics & Data Analysis* 70, 19–34.
- Heyde, C. C., Brown, B. M., 1971. An invariance principle and some convergence rate results for branching processes. *Z. Wahrscheinlichkeitsth. Verw. Geb.* 20, 271–278.
- Jagers, P., 1975. *Branching processes with biological applications*. John Wiley and Sons, Inc.
- Kimmel, M., Axelrod, D. E., 2002. *Branching processes in biology*. Springer-Verlag.
- McLachlan, G., Krishnan, T., 2008. *The EM algorithm and extensions*. John Wiley and Sons, Inc.
- R Core Team, 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>
- Rahimov, I., Al-Sabah, W., 2007. Controlled branching processes with continuous states. *Journal of Applied Probability and Statistics* 2, 123–137.
- Sriram, T., 1994. Invalidation of bootstrap for critical branching processes with immigration. *The Annals of Statistics* 22, 565–1114.
- Sriram, T. N., Bhattacharya, A., González, M., Martínez, R., del Puerto, I., 2007. Estimation of the offspring mean in a controlled branching process with a random control function. *Stochastic Processes and their Applications* 117, 928–946.

- Veen, A., Schoenberg, F., 2008. Estimation of space–time branching process models in seismology using an EM–type algorithm. *Journal of the American Statistical Association* 103, 614–624.
- Venables, B., Hornik, K., Maechler, M., 2014. `polynom`: A Collection of Functions to Implement a Class for Univariate Polynomial Manipulations. R package version 1.3-8. S original by Bill Venables, packages for R by Kurt Hornik and Martin Maechler.
- Wang, N., Wang, L., McMahan, C.S. 2015. Regression analysis of bivariate current status data under the Gamma-frailty proportional hazards model using the EM algorithm. *Computational Statistics & Data Analysis*, 83, 140-150.
- Yanev, G. P., Yanev, N. M., 1996. Branching processes with two types of emigration and state-dependent immigration. *Lecture Notes in Statistics* 114, 216–228.
- Yanev, N. M., 1976. Conditions for degeneracy of ϕ -branching processes with random ϕ . *Theory of Probability and its Applications* 20, 421–428.