

# Estimation of the infection parameter of an epidemic modeled by a branching process

Sophie Péni<sup>s</sup>son

*Université Paris-Est, LAMA (UMR 8050), UPEMLV, UPEC, CNRS, 94010 Créteil, France*

*e-mail: [sophie.penisson@u-pec.fr](mailto:sophie.penisson@u-pec.fr)*

**Abstract:** The aim of this paper is to build estimators of the infection parameter in the different phases of an epidemic (growth and decay phases). The epidemic is modeled by a Markovian process of order  $d \geq 1$ , and can be written as a multitype branching process. We propose three estimators suitable for the different classes of criticality of the process, and consequently for different phases of the epidemic. We prove their consistency and asymptotic normality when the number of ancestors (resp. number of generations) tends to infinity.

**AMS 2000 subject classifications:** Primary 60J80, 60J85, 62P10; secondary 62M05, 62F12, 62J02.

**Keywords and phrases:** Multitype branching process, conditioned branching process, CLSE, consistency and asymptotic normality, epidemiology, SEI.

Received January 2013.

## Contents

1	Introduction . . . . .	2159
2	The epidemic model . . . . .	2160
3	Estimation of the infection parameter . . . . .	2162
3.1	A closed-form estimator with asymptotic properties, as $n \rightarrow \infty$ . . . . .	2165
3.2	A closed-form CLSE with asymptotic properties, as $n \rightarrow \infty$ and as $ \mathbf{x}_0  \rightarrow \infty$ . . . . .	2167
3.3	A CLSE with asymptotic properties, as $n \rightarrow \infty$ . . . . .	2169
4	Comparison and choice of the estimators . . . . .	2172
4.1	Comparison of the estimators . . . . .	2172
4.2	Choice of the estimator in practice and illustration with a simulated multi-phase epidemic . . . . .	2173
A	Proofs related to Section 3.1 . . . . .	2177
B	Proofs related to Section 3.2 . . . . .	2178
C	Proofs related to Section 3.3 . . . . .	2179
	References . . . . .	2186

## 1. Introduction

The purpose of this paper is to quantify the infection of an epidemic in its different phases (growth and decay) by providing appropriate estimators of the infection parameter for each of these phases. The epidemic is modeled by a Markovian process of order  $d \geq 1$  with Poissonian transitions, which can be seen as a multitype Bienaymé-Galton-Watson (BGW) branching process with  $d$  types. This model is suitable for any rare transmissible *SEIR* disease in a large branching population following a Reed-Frost model for the infection ([11]). The process corresponds to the incidence of clinical cases (individuals or animals with clinical symptoms), which are assumed to be the only available observations. An epidemic phase is here defined as a phase for which the transmission mechanism of the disease remains unchanged, in particular for which the infection parameter remains constant. In our setting, this results in an increase or a decrease of the clinical cases incidence, and we refer to the corresponding phases as growth or decay phases.

Branching processes are useful models to describe the extinction and growth of populations, and as such have been applied to many biological problems (see *e.g.* [7]). The estimation of a key parameter such as the mean number of offspring or the Perron's root (largest eigenvalue of the mean matrix), which determines the asymptotic average growth rate of the population and whether or not extinction is certain, is consequently of a very large interest. Since the parameter quantifying the infection of the epidemic is, in our model, an explicit function of the Perron's root (see (3.5)), it is very natural either to build a new estimator specifically designed for the model, or to investigate the existing results in the estimation theory of the Perron's root. As detailed in Section 3, estimators of the Perron's root for general multitype branching processes usually require the knowledge of the whole or partial genealogy of the process (for example individual offspring sizes, or parent-offspring type combination counts), which are data that are mostly non available in the epidemiological context. Asmussen and Keiding, however, introduced in [1] an explicit estimator based only on the total generation sizes, corresponding in our model to the incidence of cases at each time, assumed to be the only available observations. We deduce from this estimator a first estimator of the infection parameter. Despite the potentially large order of the Markovian process that we consider, its Poissonian character ensures many properties which make it easy to derive estimators with interesting characteristics. We thus build two conditional least squares estimators (CLSE) based either on the chosen process or on the process conditioned on non-extinction at each time step, and finally compare these three estimators.

After presenting the epidemic model as well as the underlying general mathematical process in Section 2, we provide in Section 3 the three estimators of the infection parameter, which are all only based on the available observations. We aim at asymptotic results, either as the size of the initial population tends to infinity, or as the number of observations tends to infinity. It could be of a great mathematical interest to study the asymptotic behavior when both the initial size and the number of observations simultaneously tend to infinity, as it is done

in [6] for the single-type case, but we choose to focus on asymptotic properties of an immediate practical interest. We first present in Section 3.1 an explicit estimator  $\tilde{\theta}$  derived from the estimator of the Perron's root introduced in [1], and in Section 3.2 a CLSE  $\hat{\theta}$ , which are both consistent and asymptotically normal in the supercritical case, on the set of non-extinction, as the number of observations tends to infinity. These two estimators are thus especially suitable for a long growth phase of the epidemic. We prove moreover in Section 3.2 that the CLSE  $\hat{\theta}$  is consistent and asymptotically normal, as the initial population size grows to infinity, no matter the class of criticality. This estimator is thus also appropriate for any phase of the epidemic, provided that the number of clinical cases at the beginning of the phase is large. We then focus in Section 3.3 on a CLSE  $\check{\theta}$  based on the process conditioned on its non-extinction at each time step, which is consistent and asymptotically normal in the subcritical case, as the number of observations tends to infinity. This estimator is thus particularly designed for a long decay phase of the epidemic. We finally proceed in Section 4 to a numerical comparison of these estimators, focusing especially on the comparison of  $\tilde{\theta}$  against  $\hat{\theta}$  in a growth phase, and of  $\check{\theta}$  against  $\hat{\theta}$  in a decay phase. We describe the practical interest of each of these estimators, and illustrate on a simulated multi-phase epidemic which one should be chosen on which type of data.

## 2. The epidemic model

Throughout this paper we consider the following Markovian process of order  $d \geq 1$ , with deterministic initial values  $X_{-d+1} = x_{-d+1}, \dots, X_0 = x_0$ :

$$X_n = \sum_{k=1}^d \sum_{i=1}^{X_{n-k}} Y_{n-k,n,i}, \quad n \geq 1, \quad (2.1)$$

where the  $\{Y_{n-k,n,i}\}_{i,k}$  are independent given  $\mathcal{F}_{n-1} := \sigma(\{X_{n-k}\}_{k \geq 1})$ , and the  $\{Y_{n-k,n,i}\}_i$  are independent and identically distributed (i.i.d.) given  $\mathcal{F}_{n-1}$ , following a Poisson distribution with some parameter  $\Psi_k > 0$  independent of  $n$ .

Denoting by  $X_n$  the incidence of infectives at time  $n$ , and assuming that an infective can transmit the disease during one time unit at most, then  $X_n$  also corresponds to the incidence of cases at time  $n$ , which are usually the only available observations. In accordance to the usual terminology for branching process, we refer to  $X_n$  as the size of the generation  $n$ . The Markovian process (2.1) describes in a recursive way how one single infective indirectly generates new infectives (so-called ‘‘secondary cases’’)  $k$  time units later, for  $1 \leq k \leq d$ . In (2.1), the variable  $Y_{n-k,n,i}$  then corresponds to the incidence of secondary cases produced at time  $n$  with a delay  $k$  (latent period) by individual  $i$  infectious at time  $n - k$ . We refer to [11] for a justification of this modelling in the context of a rare infectious disease of the *SEIR* kind.

Let us first point out that in the simple case  $d = 1$ , the process is a single-type BGW branching process with a Poisson offspring distribution. For any

information on single-type or multitype branching processes, we refer to [2]. Let  $d \geq 1$ . We easily derive from (2.1) the conditional law of  $X_n$ ,

$$X_n | (X_{n-1}, \dots, X_{n-d}) \stackrel{\mathcal{D}}{=} \text{Poisson} \left( \sum_{k=1}^d X_{n-k} \Psi_k \right),$$

where  $\stackrel{\mathcal{D}}{=}$  denotes the equality in distribution. Moreover,  $X_n$  may be written as a multitype BGW process. For this purpose, we define the  $d$ -dimensional process  $\mathbf{X}_n := (X_{n,1}, \dots, X_{n,d})$  such that, for all  $i = 1 \dots d$ ,  $X_{n,i} := X_{n-i+1}$ , namely  $\mathbf{X}_n = (X_n, X_{n-1}, \dots, X_{n-d+1})$ . In particular, the first coordinate of  $\mathbf{X}_n$  corresponds to the value of the one-dimensional process  $(X_n)_n$  at time  $n$ . Then  $(\mathbf{X}_n)_n$  is a  $d$ -type BGW process, where the types correspond to the memory of the one-dimensional process.

Let us denote by  $\mathbf{f} = (f_1, \dots, f_d)$  its offspring generating function defined on  $[0, 1]^d$  by  $f_i(\mathbf{r}) := \mathbb{E}[\mathbf{r}^{\mathbf{X}_1} | \mathbf{X}_0 = \mathbf{e}_i]$ , where  $\mathbf{e}_i := (0, \dots, 1, \dots, 0)$  is the  $i$ th basis vector of  $\mathbb{R}^d$ . Then, for each  $i = 1 \dots d - 1$ ,  $f_i(\mathbf{r}) = e^{-\Psi_i(1-r_1)} r_{i+1}$ , and  $f_d(\mathbf{r}) = e^{-\Psi_d(1-r_1)}$ . Denoting by  $\mathbf{M}$  its mean matrix defined by its entries  $m_{ij} := \mathbb{E}[X_{1,j} | \mathbf{X}_0 = \mathbf{e}_i]$ , we obtain

$$\mathbf{M} := \begin{pmatrix} \Psi_1 & 1 & 0 & \dots & 0 \\ \Psi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \Psi_{d-1} & 0 & \dots & \dots & 1 \\ \Psi_d & 0 & \dots & \dots & 0 \end{pmatrix}. \tag{2.2}$$

The branching process  $(\mathbf{X}_n)_n$  is nonsingular and positive regular. Moreover, it can be easily shown that it satisfies the  $X \log X$  assumption, that is for each  $i, j = 1 \dots d$ ,

$$\mathbb{E}[X_{1,j} \ln X_{1,j} | \mathbf{X}_0 = \mathbf{e}_i] < \infty. \tag{2.3}$$

Let us denote by  $\rho$  the Perron's root of the mean matrix  $\mathbf{M}$ , and let  $\boldsymbol{\xi}$  and  $\boldsymbol{\eta}$  be the right and left eigenvectors of  $\mathbf{M}$  for the eigenvalue  $\rho$ , with normalization  $\boldsymbol{\xi} \cdot \mathbf{1} = \boldsymbol{\xi} \cdot \boldsymbol{\eta} = 1$ , where  $\mathbf{u} \cdot \mathbf{v}$  denotes the scalar product  $\sum_{k=1}^d u_k v_k$ . We compute that for all  $i = 1 \dots d$ ,

$$\xi_i = \frac{\rho^i \sum_{j=i}^d \Psi_j \rho^{-j}}{\sum_{k=1}^d \rho^k \sum_{j=k}^d \Psi_j \rho^{-j}}, \quad \eta_i = \frac{\sum_{k=1}^d \rho^k \sum_{j=k}^d \Psi_j \rho^{-j}}{\rho^i \sum_{k=1}^d \sum_{j=k}^d \Psi_j \rho^{-j}}.$$

By the Perron-Frobenius theorem, the Perron's root  $\rho$  provides information about the asymptotic behavior of the powers of  $\mathbf{M}$ , and thus about the expected incidence of cases at time  $n$ , for  $n$  large. Given the initial values  $X_{-d+1}, \dots, X_0$ , the expected number of new cases satisfies

$$\mathbb{E}(X_n) \stackrel{n \rightarrow \infty}{\sim} \rho^n \eta_1 \sum_{i=1}^d X_{1-i} \xi_i.$$

The asymptotic growth or decay of the average number of cases thus depends on the sign of  $\rho - 1$ . We call the process  $(\mathbf{X}_n)_n$  critical (resp. subcritical, supercritical) if  $\rho = 1$  (resp.  $< 1$ ,  $> 1$ ). By the theory of multitype positive regular and nonsingular BGW processes (see *e.g.* [2], Theorem 5.3.2), the extinction of the process  $(\mathbf{X}_n)_n$  occurs almost surely if and only if  $\rho \leq 1$ . In the supercritical case  $\rho > 1$ , there exists a real-valued random variable  $W$  such that

$$\lim_{n \rightarrow \infty} \rho^{-n} \mathbf{X}_n \stackrel{a.s.}{=} W \boldsymbol{\eta}. \quad (2.4)$$

Moreover, (2.3) implies that  $\mathbb{P}(W > 0) > 0$  ([2], Theorem 5.6.1).

As we will see in the next section, the class of criticality of the process can play a role in the choice of the estimation method. This can be problematic if the range of  $\rho$ , which has no explicit form, is unknown. However, a computation of  $\det(\mathbf{M} - \rho \mathbf{I})$  shows that  $\rho$  is solution of the equation

$$\sum_{k=1}^d \Psi_k \rho^{-k} = 1, \quad (2.5)$$

which implies that  $(\mathbf{X}_n)_n$  is critical (resp. subcritical, supercritical) if  $\sum_{k=1}^d \Psi_k = 1$  (resp.  $< 1$ ,  $> 1$ ). If the order of magnitude of the parameters  $\Psi_k$  is known, we thus obtain an alternative and more handy criticality criteria.

### 3. Estimation of the infection parameter

From now on we assume that  $d$  is fixed, and that the  $\Psi_k$ 's affinely depend on some unknown parameter  $\theta_0$ , that is for all  $k = 1 \dots d$ ,

$$\Psi_k(\theta_0) = a_k \theta_0 + b_k, \quad (3.1)$$

where  $a_k > 0$  and  $b_k \geq 0$  are known. In what follows we will refer to  $\theta_0$  as the infection parameter (see Remark 3.1). We shall denote by  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\boldsymbol{\Psi}(\theta_0)$  the  $d$ -dimensional vectors with coordinates  $a_k$ ,  $b_k$  and  $\Psi_k(\theta_0)$  respectively. From Section 2 it immediately follows that  $(\mathbf{X}_n)_n$  is critical (resp. subcritical, supercritical) if

$$\theta_0 = \frac{1 - \mathbf{b} \cdot \mathbf{1}}{\mathbf{a} \cdot \mathbf{1}} \quad (\text{resp. } <, >). \quad (3.2)$$

**Remark 3.1.** The assumption (3.1) is for instance relevant in the epidemiological context of rare *SEIR* diseases in large populations. Indeed, in this case (see [11]), the process of cases incidences  $(X_n)_n$  is of the form (2.1), where  $d + 1$  is the largest survival age in the whole population, and for each  $k = 1 \dots d$ ,

$$\Psi_k(\theta_0) = \sum_{a=k+1}^{d+1} L_{a-k,k} S_a \theta_0 + L_{1,k} S_{k+1} p.$$

The unknown parameter in  $\Psi_k$  would commonly be the horizontal infection parameter  $\theta_0$ , which is the mean number of individuals infected by an infective

per time unit by horizontal route. The other parameters involved, that is  $L_{a,k}$  (probability for an individual aged  $a$  at infection to have a latent period equal to  $k$ , given his survival),  $S_a$  (probability to be aged  $a$  at the end of the latent period) and  $p$  (vertical infection parameter, representing the maternal transmission probability), can be reasonably assumed to be known from previous experiences or from the study of a previous phase of the epidemic. However, the infection parameter  $\theta_0$  is often unknown and is likely to vary. Indeed, it strongly depends on the environment which might have been modified by control or sanitary measures during the course of the epidemic. Estimating this parameter then not only provides information about the efficiency of the new measures, but also enables predictions of the future spread of the disease under the same measures.

Assuming (3.1), our aim is to provide estimators of  $\theta_0$  based on the observations  $(\mathbf{X}_0, \dots, \mathbf{X}_n)$ , with asymptotic properties corresponding to interesting characteristics in the epidemiological context. We are thus looking for estimators suitable in the subcritical and/or supercritical cases, with asymptotic properties, as the initial population size grows to infinity, or as the number of observations  $n$  tends to infinity. We would thus entirely cover the problem of estimating the infection parameter in the growth and extinction phases of the epidemic, offering moreover several alternatives depending on which asymptotic assumption is suitable regarding the available data.

There are numerous results in the literature dedicated to the estimation theory for general branching processes. In its early paper [8] in 1948, Harris provided an estimator for the mean value  $m_0$  of a single-type BGW process  $X_0, \dots, X_n$ . It is a maximum likelihood estimator (MLE) based on observed values of the individual offspring size for each individual at each time:

$$\hat{m}^{MLE} = \frac{X_1 + \dots + X_n}{X_0 + \dots + X_{n-1}}, \tag{3.3}$$

which is consistent as  $n \rightarrow \infty$  in the supercritical case, on the set of non-extinction. Note that  $\hat{m}^{MLE}$  only involves  $X_0, \dots, X_n$ . It is actually proved in [5] that it is also the MLE of  $m_0$  based on the observed values of  $X_0, \dots, X_n$  only. It is straightforward to show that  $\hat{m}^{MLE}$  is also the weighted conditional least squares estimator (CLSE) based on the process  $X_k/\sqrt{X_{k-1}}$ , defined as follows

$$\hat{m}^{CLSE} = \arg \min_{m \geq 0} \sum_{k=1}^n \frac{(X_k - mX_{k-1})^2}{X_{k-1}}.$$

Similar estimation problems are considered in the multitype case. In [1], Asmussen and Keiding proposed a maximum likelihood estimator of the Perron's root  $\rho_0$  based on the observations of the whole genealogy of the population (*i.e.* each offspring vector produced by every individual). It is proved in [10] that this estimator is also the maximum likelihood estimator solely based on the observations at each time of the total number of individuals of type  $j$  whose parents were of type  $i$ , for every  $i, j = 1 \dots d$ . However in epidemiology this kind of

variables are generally not observable. For our model this would imply indeed that, considering the number of cases at a given time, we could say how many of them were infected exactly  $k$  time-units earlier. We are thus more interested in estimations based on the number of cases at each time, such as the other estimator presented in [1],

$$\tilde{\rho} = \frac{|\mathbf{X}_1| + \cdots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \cdots + |\mathbf{X}_{n-1}|}, \quad (3.4)$$

where  $|\cdot|$  denotes the  $\mathcal{L}^1$ -norm  $|\mathbf{X}_k| := X_{k,1} + \cdots + X_{k,d}$ . For  $d = 1$ ,  $\tilde{\rho}$  clearly reduces to the Harris estimator defined in (3.3). Note that relation (2.5) implies that

$$\theta_0 = \frac{1 - \sum_{k=1}^d b_k \rho_0^{-k}}{\sum_{k=1}^d a_k \rho_0^{-k}}. \quad (3.5)$$

Hence an estimation of  $\rho_0$  would provide an estimation of  $\theta_0$  (the opposite is not true since  $\rho_0$  cannot in general be expressed as an explicit function of  $\theta_0$ ). In the supercritical case, the estimator  $\tilde{\rho}$  was shown to be consistent, as  $n \rightarrow \infty$ , on the set of non-extinction, with an explicit asymptotic distribution ([1]).

Due to the Poissonian character of the transitions of the process  $(\mathbf{X}_n)_{n \geq 0}$ , it is possible, in our setting, to express the joint probability function of the observations  $X_0, \dots, X_n$ , without involving the whole or partial genealogy of the process. The likelihood function is indeed given by two factors, one of which is independent of  $\theta_0$ , the logarithm of the other being  $L(\theta_0) := -\theta_0 \sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1} + \sum_{k=1}^n X_k \ln(\Psi(\theta_0) \cdot \mathbf{X}_{k-1})$ . The MLE of  $\theta_0$  based on the observations  $X_0, \dots, X_n$  is thus a solution of  $L'(\theta) = 0$ , where  $L'(\theta) = -\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1} + \sum_{k=1}^n X_k (\mathbf{a} \cdot \mathbf{X}_{k-1}) (\Psi(\theta) \cdot \mathbf{X}_{k-1})^{-1}$ . This equation has in general no explicit solution, except for simple cases such as the one-dimensional case  $d = 1$ , for which the MLE is

$$\hat{\theta}^{MLE} = \frac{\sum_{k=1}^n (X_k - bX_{k-1})}{\sum_{k=1}^n aX_{k-1}},$$

or the linear case  $\mathbf{b} = \mathbf{0}$ , for which the MLE is

$$\hat{\theta}^{MLE} = \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}. \quad (3.6)$$

As shown later (see (3.11)), it corresponds in these cases to the CLSE of  $\theta_0$ . It is however in general not the case, and we choose to focus on the CLSE.

In Section 3.1 we derive from the estimator (3.4) a first estimator of  $\theta_0$ ,

$$\tilde{\theta} := \frac{1 - \sum_{k=1}^d b_k \tilde{\rho}^{-k}}{\sum_{k=1}^d a_k \tilde{\rho}^{-k}},$$

and deduce from [1] asymptotic properties of  $\tilde{\theta}$ , as  $n \rightarrow \infty$ , in the supercritical case, on the set of non-extinction. In a second instance, we study in Section 3.2 the weighted CLSE

$$\hat{\theta} := \arg \min_{\theta \in \Theta} \sum_{k=1}^n \frac{[X_k - \mathbb{E}_\theta(X_k | \mathbf{X}_{k-1})]^2}{\mathbf{a} \cdot \mathbf{X}_{k-1}}$$

which has a closed-form expression given by (3.11). We show its asymptotic properties, on the one hand as  $n \rightarrow \infty$  in the supercritical case on the set of non-extinction, and on the other hand as  $|\mathbf{x}_0| \rightarrow \infty$  for any class of criticality, where  $|\mathbf{x}_0| = x_{-d+1} + \dots + x_0$  is the deterministic initial population size. Finally, since we also aim at finding an estimator with asymptotic properties, as  $n \rightarrow \infty$ , holding in the subcritical case, we consider in Section 3.3 the CLSE associated with a Markov chain  $(\mathbf{Z}_k)_{k \geq 0}$  satisfying  $\mathbb{P}(\mathbf{Z}_k = \mathbf{j} | \mathbf{Z}_{k-1} = \mathbf{i}) = \mathbb{P}(\mathbf{X}_k = \mathbf{j} | \mathbf{X}_{k-1} = \mathbf{i}, \mathbf{X}_k \neq \mathbf{0})$  (see the detailed definition in Section 3.3):

$$\check{\theta} := \arg \min_{\theta \in \Theta} \sum_{k=1}^n \frac{[Z_k - \mathbb{E}_\theta(Z_k | \mathbf{Z}_{k-1})]^2}{\mathbf{a} \cdot \mathbf{Z}_{k-1}}.$$

The practical interest and efficiency of these three estimators will be discussed and compared in Section 4.

### 3.1. A closed-form estimator with asymptotic properties, as $n \rightarrow \infty$

The aim of this section is to provide an estimator with asymptotic properties, as the number of observations tends to infinity, in the supercritical case, which in the epidemiological context would correspond to the growth phase of the epidemic. For this purpose, we deduce from relation (3.5) the following explicit estimator of  $\theta_0$ ,

$$\tilde{\theta} := \frac{1 - \sum_{k=1}^d b_k \tilde{\rho}^{-k}}{\sum_{k=1}^d a_k \tilde{\rho}^{-k}}. \tag{3.7}$$

where  $\tilde{\rho}$  is the estimator of  $\rho_0$  introduced in [1] and given by (3.4). Note that what follows can be applied to any process of the form (2.1), where the  $\{Y_{n-k,n,i}\}_i$  do not necessarily follow a Poisson distribution, but satisfy the relation  $\mathbb{E}_{\theta_0}(Y_{n-k,n,i} | \mathcal{F}_{n-1}) = \Psi_k(\theta_0)$ .

Let us denote by  $W_0$  the limiting random variable introduced in (2.4). First, as pointed out by Becker in [3], the estimator  $\tilde{\rho}$  is strongly consistent on the set of non-extinction  $\{W_0 > 0\}$ . Second, as proven by Asmussen and Keiding in [1], once adequately normalized,  $\tilde{\rho} - \rho_0$  is asymptotically normal. However, the asymptotic behavior of  $\tilde{\rho} - \rho_0$  depends qualitatively on the relative sizes of  $\rho_0$  and  $\lambda^2$ , where  $\lambda$  is the modulus of a certain eigenvalue of  $\mathbf{M}(\theta_0)$ . For the sake of clarity, we need to recall the constants introduced in [1] to study the asymptotic properties of  $\tilde{\rho}$ , which play a role in the study of  $\tilde{\theta}$  as well. Let  $\{\lambda_i\}_{i=1 \dots s}$  be the spectrum of  $\mathbf{M}(\theta_0)$ , and for each  $i = 1 \dots s$ , let  $r_i$  be the algebraic multiplicity of  $\lambda_i$ . We denote by  $\mathcal{B} = \{\mathbf{u}_{i,j}, i = 1 \dots s, j = 1 \dots r_i\}$  the base of the Jordan canonical decomposition of  $\mathbf{M}(\theta_0)$ . Let us define the vector  $\zeta := \mathbf{1} - |\boldsymbol{\eta}_0| \boldsymbol{\xi}_0$ , where  $\boldsymbol{\xi}_0$  is the eigenvector of  $\mathbf{M}(\theta_0)$  defined in Section 2. We denote  $(\zeta_{i,j})_{i=1 \dots s, j=1 \dots r_i}$  its coordinates in  $\mathcal{B}$ . Then  $\lambda$  is defined as follows,

$$\lambda = \lambda(\zeta) := \max_{i=1 \dots s} \{|\lambda_i| : \exists j = 1 \dots r_i \text{ such that } \zeta_{i,j} \neq 0\},$$



and

$$\gamma = \gamma(\zeta) := \max_{\substack{i=1,\dots,s: \\ |\lambda_i|=\lambda}} \{j = 1 \dots r_i : \zeta_{i,j} \neq 0\}.$$

We similarly define  $\lambda(\mathbf{x})$  and  $\gamma(\mathbf{x})$  for any complex vector  $\mathbf{x} \in \mathbb{C}^d$ . As detailed in [1],  $\tilde{\rho} - \rho_0 = (S_n + T_n) (|\mathbf{X}_0| + \dots + |\mathbf{X}_{n-1}|)^{-1}$ , where (to avoid heavy notation, when no confusion is possible, we do not write differently column and row vectors when multiplied by a matrix)

$$S_n := \sum_{k=0}^{n-1} (\mathbf{X}_{k+1} - \mathbf{X}_k \mathbf{M}(\theta_0)) \cdot \mathbf{1}, \quad T_n := \sum_{k=0}^{n-1} \mathbf{X}_k \cdot \boldsymbol{\kappa}, \quad \boldsymbol{\kappa} := (\mathbf{M}(\theta_0) - \rho_0 \mathbf{I}) \mathbf{1}.$$

It appears that  $S_n$  and  $T_n$  are of the same order of magnitude when  $\lambda^2 < \rho_0$ , while  $T_n$  dominates  $S_n$  if  $\lambda^2 \geq \rho_0$ . For each  $i = 1 \dots d$  and  $n \in \mathbb{N}$ , we define the covariance matrix  $\mathbf{V}_n^i(\theta_0)$  with entries

$$[\mathbf{V}_n^i(\theta_0)]_{jk} := \mathbb{E}_{\theta_0} (X_{n,j} X_{n,k} | \mathbf{X}_0 = \mathbf{e}_i) - \mathbb{E}_{\theta_0} (X_{n,j} | \mathbf{X}_0 = \mathbf{e}_i) \mathbb{E}_{\theta_0} (X_{n,k} | \mathbf{X}_0 = \mathbf{e}_i).$$

In particular,  $[\mathbf{V}_1^i]_{jk} = \Psi_i(\theta_0)$  if  $j = k = 1$ , and is null otherwise. In order to deal with the case  $\lambda^2 < \rho_0$ , we define for all  $n \in \mathbb{N}$ ,  $\boldsymbol{\nu}_n := \mathbf{1} + \sum_{k=0}^{n-1} \mathbf{M}(\theta_0)^k \boldsymbol{\kappa}$ , and

$$C_1 := (\rho_0 - 1) \sum_{n=1}^{\infty} \rho_0^{-n} \sum_{i=1}^d \eta_{i,0} \Psi_i(\theta_0) \boldsymbol{\nu}_{n,1}^2.$$

If  $\lambda^2 \geq \rho_0$ , then there exist vectors  $\boldsymbol{\zeta}^1$  and  $\boldsymbol{\zeta}^2$  such that  $(\mathbf{M}(\theta_0) - \rho_0 \mathbf{I}) \boldsymbol{\zeta} = (\mathbf{M}(\theta_0) - \mathbf{I}) \boldsymbol{\zeta}^1 + \boldsymbol{\zeta}^2$ , with  $\lambda(\boldsymbol{\zeta}^1) = \lambda$ ,  $\gamma(\boldsymbol{\zeta}^1) = \gamma$  and  $\lambda(\boldsymbol{\zeta}^2) \leq 1$ . If  $\lambda^2 = \rho_0$ , we set moreover

$$C_2 := \left(1 - \frac{1}{\rho_0}\right) \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^d \eta_{i,0} \boldsymbol{\zeta}^1 \mathbf{V}_n^i(\theta_0) \boldsymbol{\zeta}^1}{\rho_0^n n^{2\gamma-1}}.$$

We finally define the constant

$$C_0 := \left( \frac{\sum_{k=1}^d \rho_0^{-k} \left( k a_k + a_k \sum_{i=1}^d i b_i \rho_0^{-i} - b_k \sum_{i=1}^d i a_i \rho_0^{-i} \right)}{\rho_0 \left( \sum_{k=1}^d a_k \rho_0^{-k} \right)^2} \right)^2.$$

We recall that (2.3) implies  $\mathbb{P}_{\theta_0}(W_0 > 0) > 0$ . For notational convenience, it is assumed in this section just as in [1] that  $\mathbb{P}_{\theta_0}(W_0 = 0) = 0$ . The following results are thus valid on the set of non-extinction. The proof of Theorem 3.2 is postponed in Appendix A.

**Theorem 3.2.** *Let us assume that the process  $(\mathbf{X}_k)_{k \geq 0}$  is supercritical. Then, on the set of non-extinction, the estimator  $\tilde{\theta}$  is strongly consistent:*

$$\lim_{n \rightarrow \infty} \tilde{\theta} \stackrel{a.s.}{=} \theta_0,$$

and has the following asymptotic distribution.

If  $\lambda^2 < \rho_0$ ,

$$\lim_{n \rightarrow \infty} \sqrt{\frac{W_0 (1 + \dots + \rho_0^{n-1})}{C_1 C_0}} (\tilde{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1). \tag{3.8}$$

If  $\lambda^2 = \rho_0$  and  $C_2 > 0$ ,

$$\lim_{n \rightarrow \infty} \sqrt{\frac{W_0 (1 + \dots + \rho_0^{n-1})}{n^{2\gamma-1} C_2 C_0}} (\tilde{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1). \tag{3.9}$$

If  $\lambda^2 > \rho_0$ , there exist random variables  $H_n$  with  $\overline{\lim} |H_n| < \infty$ , such that

$$\lim_{n \rightarrow \infty} \left[ \frac{W_0 (1 + \dots + \rho_0^{n-1})}{\lambda^{n-1} (n-1)^{\gamma-1}} (\tilde{\theta} - \theta_0) - H_{n-1} \right] \stackrel{a.s.}{=} 0. \tag{3.10}$$

**Remark 3.3.** Unlike the strong consistency, the second part of the theorem is seldom of direct practical applicability, in particular because of the differentiation between the three cases  $\lambda^2 < \rho_0$ ,  $\lambda^2 = \rho_0$  and  $\lambda^2 > \rho_0$ . We cannot provide here an asymptotic confidence interval solely based on the observations, as we can do for the estimators  $\hat{\theta}$  and  $\tilde{\theta}$  (see Theorem 3.6 and Theorem 3.8).

**Remark 3.4.** If  $d > 1$ , the estimator  $\tilde{\theta}$  seems to have no interesting asymptotic properties as  $|\mathbf{x}_0| \rightarrow \infty$ , for  $n$  fixed, unless we assume that for all  $i = 1 \dots d$ ,  $\lim_{|\mathbf{x}_0| \rightarrow \infty} x_{0,i} |\mathbf{x}_0|^{-1} = \eta_{0,i} |\boldsymbol{\eta}|^{-1}$ . If this holds, then for any class of criticality,  $\lim_{|\mathbf{x}_0| \rightarrow \infty} \tilde{\theta} = \theta_0$  almost surely. This assumption is however very restrictive. Nevertheless, if  $d = 1$  then  $\tilde{\theta}$  reduces to the CLSE  $\hat{\theta}$  studied in Section 3.2 with asymptotic properties as  $|\mathbf{x}_0| \rightarrow \infty$ .

**3.2. A closed-form CLSE with asymptotic properties, as  $n \rightarrow \infty$  and as  $|\mathbf{x}_0| \rightarrow \infty$**

In this section, we provide an estimator with asymptotic properties as the number of observations  $n$  or as the initial (deterministic) population size  $|\mathbf{x}_0|$  tends to infinity. Note that this estimator was already introduced in [11], but only the limit in  $|\mathbf{x}_0|$  was investigated. We recall these known results in Theorem 3.6, and present the additional limit results as  $n$  tends to infinity in Theorem 3.5, whose proof is postponed in Appendix B. Please observe that the latter theorem is only valid in the supercritical case, while Theorem 3.6 holds for any class of criticality. We thus present here an estimator which is on the one hand suitable for a long growth phase of an epidemic (Theorem 3.5), just as  $\tilde{\theta}$ , and is on the other hand suitable for any phase with a large initial population size (Theorem 3.6).

We consider the weighted CLSE based on the process  $Y_k := X_k / \sqrt{\mathbf{a} \cdot \mathbf{X}_{k-1}}$ ,

$$\hat{\theta} := \arg \min_{\theta \in \Theta} \sum_{k=1}^n (Y_k - \mathbb{E}_\theta(Y_k | \mathbf{X}_{k-1}))^2 = \arg \min_{\theta \in \Theta} \sum_{k=1}^n \frac{(X_k - \boldsymbol{\Psi}(\theta) \cdot \mathbf{X}_{k-1})^2}{\mathbf{a} \cdot \mathbf{X}_{k-1}},$$

where  $\Theta := ]\theta_{\min}, \theta_{\max}[$  with  $\theta_{\max} > \theta_{\min} > 0$ . We easily derive the following explicit form

$$\hat{\theta} = \frac{\sum_{k=1}^n (X_k - \mathbf{b} \cdot \mathbf{X}_{k-1})}{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}. \tag{3.11}$$

Note that the normalization of the process  $X_k$  by  $\sqrt{\mathbf{a} \cdot \mathbf{X}_{k-1}}$  generalizes the normalization  $X_k/\sqrt{aX_{k-1}}$  in the monotype case, leading to the Harris estimator of  $m_0 = a\theta_0 + b$ . Moreover,  $\hat{\theta}$  also corresponds to the MLE of  $\theta_0$  if  $\mathbf{b} = \mathbf{0}$  (see (3.6)). In addition, the conditional variance of the error term in the stochastic regression equation for the normalized process  $Y_k = \mathbb{E}_{\theta_0}(Y_k|\mathbf{X}_{k-1}) + Y_k - \mathbb{E}_{\theta_0}(Y_k|\mathbf{X}_{k-1})$  is invariant under multiplication of the whole process, and bounded respectively to  $(\mathbf{X}_k)_{k \geq 0}$ .

**Theorem 3.5.** *Let us assume that the process  $(\mathbf{X}_k)_{k \geq 0}$  is supercritical. Then, on the set of non-extinction, the estimator  $\hat{\theta}$  is strongly consistent:*

$$\lim_{n \rightarrow \infty} \hat{\theta} \stackrel{a.s.}{=} \theta_0,$$

and is asymptotically normally distributed:

$$\lim_{n \rightarrow \infty} \sqrt{\frac{(\mathbf{a} \cdot \boldsymbol{\eta}_0)^2 W_0 (1 + \dots + \rho_0^{n-1})}{\rho_0 \eta_{0,1}}} (\hat{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

**Theorem 3.6.** *Let us assume that there exist some  $\alpha_i \in [0, 1]$ ,  $i = 1 \dots d$ , such that, for all  $i = 1 \dots d$ ,  $\lim_{|\mathbf{x}_0| \rightarrow \infty} x_{0,i} |\mathbf{x}_0|^{-1} = \alpha_i$ . Then the estimator  $\hat{\theta}$  is strongly consistent:*

$$\lim_{|\mathbf{x}_0| \rightarrow \infty} \hat{\theta} \stackrel{a.s.}{=} \theta_0,$$

and is asymptotically normally distributed:

$$\lim_{|\mathbf{x}_0| \rightarrow \infty} \sqrt{\frac{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}{\sigma^2(\hat{\theta})}} (\hat{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1),$$

where

$$\sigma^2(\theta) := \theta + \frac{\sum_{k=1}^n \sum_{i,j=1}^d \alpha_j b_i m_{ji}^{(k-1)}(\theta)}{\sum_{k=1}^n \sum_{i,j=1}^d \alpha_j a_i m_{ji}^{(k-1)}(\theta)}.$$

and for every  $i, j = 1 \dots d$ ,  $k \geq 1$ ,  $m_{ij}^{(k)}(\theta)$  denotes the  $(i, j)$ -th entry in the  $k$ -th power of the matrix  $\mathbf{M}(\theta)$  given by (2.2).

**Remark 3.7.** We point out that in the proofs of Theorem 3.5 and Theorem 3.6 we do not use the Poissonian character of the transitions of the process (2.1) to derive the properties of  $\hat{\theta}$ , but we simply need its first and second order moments. The results of these theorems can thus be applied to any process of the form (2.1), where the  $\{Y_{n-k,n,i}\}_i$  do not necessarily follow a Poisson distribution, but satisfy  $\mathbb{E}_{\theta_0}(Y_{n-k,n,i}|\mathcal{F}_{n-1}) = \Psi_k(\theta_0)$ . For Theorem 3.6, the variance should be either known, or previously estimated, and the process should be normalized accordingly such that the error term in the stochastic regression equation remains bounded.

**3.3. A CLSE with asymptotic properties, as  $n \rightarrow \infty$**

In this section we study a CLSE associated with a conditioned process, and obtain asymptotic properties as the number of observations  $n$  tends to infinity in the subcritical case, despite the almost sure extinction of the original process  $(\mathbf{X}_k)_{k \geq 0}$ . We point out that this estimator is particularly adapted for the study of the extinction phase of an epidemic, and does not require a large initial number of cases.

For this purpose, let us consider instead of  $(\mathbf{X}_k)_{k \geq 0}$  the associated process conditioned on non-extinction at each time-step. By this we mean the homogeneous Markov chain  $(\mathbf{Z}_k)_{k \geq 0}$  with deterministic initial value  $\mathbf{Z}_0 = \mathbf{x}_0$  and with the following transition probabilities

$$Q(\mathbf{i}, \mathbf{j}) := \mathbb{P}_{\theta_0}(\mathbf{Z}_k = \mathbf{j} | \mathbf{Z}_{k-1} = \mathbf{i}) = \mathbb{P}_{\theta_0}(\mathbf{X}_k = \mathbf{j} | \mathbf{X}_{k-1} = \mathbf{i}, \mathbf{X}_k \neq \mathbf{0}), \quad \mathbf{i}, \mathbf{j} \neq \mathbf{0}.$$

Denoting by  $P(\mathbf{i}, \mathbf{j})$  the transition probabilities of the process  $(\mathbf{X}_k)_{k \geq 0}$ , we thus have

$$Q(\mathbf{i}, \mathbf{j}) = \frac{P(\mathbf{i}, \mathbf{j})}{1 - P(\mathbf{i}, \mathbf{0})}, \quad \mathbf{i}, \mathbf{j} \neq \mathbf{0}.$$

By definition of  $(\mathbf{X}_k)_{k \geq 0}$ ,  $P(\mathbf{i}, \mathbf{j}) = 0$  as soon as  $(j_2, \dots, j_d) \neq (i_1, \dots, i_{d-1})$ , hence the process  $(\mathbf{Z}_k)_{k \geq 0}$  satisfies for all  $i = 2 \dots d$  and  $k \geq 0$ ,  $Z_{k,i} = Z_{k-1,i-1}$ . We also define the one-dimensional  $d$ -Markovian process  $(Z_k)_{k \geq 0}$  corresponding to the first coordinate: for all  $n \geq 0$ ,  $Z_k := Z_{k,1}$ .

For all  $d$ -dimensional vector  $\mathbf{u}$ , we define the truncated sum  $\lceil \mathbf{u} \rceil := u_1 + \dots + u_{d-1}$ . By definition,

$$\mathbb{P}_{\theta_0}(Z_k = j | \mathbf{Z}_{k-1}) = \frac{(\Psi(\theta_0) \cdot \mathbf{Z}_{k-1})^j e^{-\Psi(\theta_0) \cdot \mathbf{Z}_{k-1}}}{j! (1 - \mathbf{1}_{\{\lceil \mathbf{z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{k-d}})}.$$

We consider the CLSE corresponding to the normalized process  $Z_k / \sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}}$ ,

$$\check{\theta} := \arg \min_{\theta \in \Theta} S_n(\theta), \quad S_n(\theta) := \sum_{k=1}^n \left( \frac{Z_k}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}}} - f(\theta, \mathbf{Z}_{k-1}) \right)^2, \quad (3.12)$$

where  $\Theta := ]\theta_{\min}, \theta_{\max}[$  with  $\theta_{\max} > \theta_{\min} > 0$ , and

$$\begin{aligned} f(\theta, \mathbf{Z}_{k-1}) &:= \mathbb{E}_{\theta} \left( \frac{Z_k}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}}} \middle| \mathbf{Z}_{k-1} \right) \\ &= \frac{\Psi(\theta) \cdot \mathbf{Z}_{k-1}}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}} (1 - \mathbf{1}_{\{\lceil \mathbf{z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta) Z_{k-d}})}. \end{aligned} \quad (3.13)$$

This estimator has no closed expression, but the continuity of  $S_n$  on  $\overline{\Theta}$  ensures the existence of a least-squares estimate, which can be numerically approximated by minimization of  $S_n$  on a suitable lattice embedded into  $\overline{\Theta}$  as described in Remark 3.11.

Denoting by  $f'$  the derivative of  $f$  with respect to  $\theta$ , we thus have, for all  $\theta \in \Theta$  and all  $\mathbf{j} \in \mathbb{N}^d, \mathbf{j} \neq \mathbf{0}$ ,

$$f'(\theta, \mathbf{j}) = \begin{cases} \sqrt{a_d j_d} \frac{1 - (1 + \Psi_d(\theta) j_d) e^{-\Psi_d(\theta) j_d}}{(1 - e^{-\Psi_d(\theta) j_d})^2} & \text{if } [\mathbf{j}] = 0, \\ \sqrt{\mathbf{a} \cdot \mathbf{j}} & \text{otherwise.} \end{cases} \tag{3.14}$$

We finally define  $\varepsilon_k := Z_k / \sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}} - f(\theta_0, \mathbf{Z}_{k-1})$ , which implies that

$$\mathbb{E}_{\theta_0} (\varepsilon_k^2 | \mathbf{Z}_{k-1}) = \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{k-1}}{\mathbf{a} \cdot \mathbf{Z}_{k-1} (1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{k-d}})},$$

and the conditional variance of the error term  $\varepsilon_k$  in the stochastic regression equation is consequently bounded. Indeed, defining for any vector  $\mathbf{u}$ ,  $\underline{\mathbf{u}} := \min_i u_i$  and  $\bar{\mathbf{u}} := \max_i u_i$ , we have

$$\theta_0 + \frac{\underline{\mathbf{b}}}{\bar{\mathbf{a}}} \leq \mathbb{E}_{\theta_0} (\varepsilon_k^2 | \mathbf{Z}_{k-1}) \leq (1 - e^{-\Psi_d(\theta_0)})^{-1} \left( \theta_0 + \frac{\bar{\mathbf{b}}}{\underline{\mathbf{a}}} \right). \tag{3.15}$$

**Theorem 3.8.** *The estimator  $\check{\theta}$  is strongly consistent:*

$$\lim_{n \rightarrow \infty} \check{\theta} \stackrel{a.s.}{=} \theta_0.$$

Moreover, if the process  $(\mathbf{X}_k)_{k \geq 0}$  is subcritical, then the estimator  $\check{\theta}$  is asymptotically normally distributed:

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^n (f'(\check{\theta}, \mathbf{Z}_k))^2}{\sqrt{\sum_{k=0}^n (f'(\check{\theta}, \mathbf{Z}_k))^2 f(\check{\theta}, \mathbf{Z}_k) (\mathbf{a} \cdot \mathbf{Z}_k)^{-1/2}}} (\check{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

The proof of this theorem can be found in Appendix C.

**Remark 3.9.** For many observations, the estimations of  $\theta_0$  provided by  $\hat{\theta}$  and  $\check{\theta}$  are by construction equal. More precisely, for any  $\mathbb{N}^d$ -valued sequence  $(\mathbf{x}_0, \dots, \mathbf{x}_n)$  such that for all  $k \leq n$ ,  $x_{k,i} = x_{k-1,i-1}$  and such that  $\mathbf{x}_k \notin \mathcal{S}$ , where  $\mathcal{S}$  is the subset of  $\mathbb{N}^d$  defined by

$$\mathcal{S} := \{(0, \dots, 0, x), x \in \mathbb{N}, x \neq 0\}, \tag{3.16}$$

then  $\hat{\theta}(\mathbf{x}_0, \dots, \mathbf{x}_n) = \check{\theta}(\mathbf{x}_0, \dots, \mathbf{x}_n)$ . In other words, if there is no sequence of  $d - 1$  successive null values in the observation  $(x_{-d+1}, \dots, x_n)$ , then both estimations are equal.

**Remark 3.10.** Defining  $T^X := \inf\{k \geq 1, \mathbf{X}_k \in \mathcal{S}\}$  and  $T^Z := \inf\{k \geq 1, \mathbf{Z}_k \in \mathcal{S}\}$ , we have

$$\lim_{\substack{|\mathbf{x}_0| \rightarrow \infty \\ \mathbf{x}_0 \notin \mathcal{S}}} \mathbb{P}(T^Z \leq n | \mathbf{Z}_0 = \mathbf{x}_0) = \lim_{\substack{|\mathbf{x}_0| \rightarrow \infty \\ \mathbf{x}_0 \notin \mathcal{S}}} \mathbb{P}(T^X \leq n | \mathbf{X}_0 = \mathbf{x}_0) = 0,$$

and we can deduce from Theorem 3.6 (we omit here the detailed proof) that, for any  $n \in \mathbb{N}$  fixed,

$$\lim_{|\mathbf{x}_0| \rightarrow \infty} \check{\theta} \stackrel{a.s.}{=} \theta_0, \quad \text{and} \quad \lim_{|\mathbf{x}_0| \rightarrow \infty} \sqrt{\frac{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{Z}_{k-1}}{\sigma^2(\check{\theta})}} (\check{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1).$$

**Remark 3.11.** Since the estimator  $\check{\theta}$  defined by (3.12) has no closed expression, we provide a simple algorithm for its calculation, whose pseudo-code is written below. The algorithm consists in computing the value of  $S_n(\theta)$  for each  $\theta$  on a finite lattice embedded into  $\Theta$ , and in returning one value of the lattice for which  $S_n$  reaches its minimum. The input arguments of the algorithm are the vectors  $\mathbf{a}$ ,  $\mathbf{b}$  (denoted  $a$  and  $b$  in the pseudo-code), the vector  $Z = (Z_{-d+1}, \dots, Z_n)$ , the bounds  $\theta_{\min}$ ,  $\theta_{\max}$  of the open set  $\Theta$ , and the step length of the lattice chosen by the user.

---

```

function  $S(a, b, Z, \theta)$  ▷ definition of the function  $S_n(\theta)$ 
   $\Psi \leftarrow a\theta + b$ 
   $i \leftarrow 0$ 
   $d \leftarrow \text{length}(a)$ 
   $n \leftarrow \text{length}(Z) - d$ 
   $s \leftarrow 0$ 
  for  $k \leftarrow 1, n$  do
    if  $(Z_{k-d+1}, \dots, Z_{k-1}) = (0, \dots, 0)$  then
       $s \leftarrow s + \frac{(Z_k - \Psi_d Z_{k-d})^2}{a_d Z_{k-d} (1 - e^{-\Psi_d Z_{k-d}})^2}$ 
    else
       $s \leftarrow s + \frac{(Z_k - \Psi \cdot (Z_{k-d}, \dots, Z_{k-1}))^2}{a \cdot (Z_{k-d}, \dots, Z_{k-1})}$ 
    end if
  end for
  return  $s$ 
end function

function  $\text{CLSE}(a, b, Z, \theta_{\min}, \theta_{\max}, \text{step})$  ▷ computation of a least square estimate
   $\text{currentargmin} \leftarrow \theta_{\min} + \text{step}$ 
   $\text{currentmin} \leftarrow S(a, b, Z, \theta_{\min} + \text{step})$ 
  if  $(\theta_{\max} - \theta_{\min}) / \text{step}$  is an integer then
     $m \leftarrow (\theta_{\max} - \theta_{\min}) / \text{step} - 1$ 
  else
     $m \leftarrow \text{floor}((\theta_{\max} - \theta_{\min}) / \text{step})$ 
  end if
  for  $k \leftarrow 2, m$  do
    if  $S(a, b, Z, \theta_{\min} + k * \text{step}) < \text{currentmin}$  then
       $\text{currentargmin} \leftarrow \theta_{\min} + k * \text{step}$ 
       $\text{currentmin} \leftarrow S(a, b, Z, \text{currentargmin})$ 
    end if
  end for
  return  $\text{currentargmin}$ 
end function

```

---

The accuracy of the result provided by this algorithm then obviously depends on the choice of the step. We can see on examples (e.g. Table 1) that this is entirely satisfying. Indeed, for any observation  $(x_{-d+1}, \dots, x_n)$  which does not con-

tain  $d - 1$  successive null values, hence for which  $\tilde{\theta}(\mathbf{x}_0, \dots, \mathbf{x}_n) = \hat{\theta}(\mathbf{x}_0, \dots, \mathbf{x}_n)$ , the previous algorithm provides the rounded value (according to the step length of the lattice) of the estimation given by the closed-form estimator  $\hat{\theta}$ .

#### 4. Comparison and choice of the estimators

The goal of this section is to investigate and compare the performances of the three estimators  $\tilde{\theta}$ ,  $\hat{\theta}$  and  $\check{\theta}$  discussed in this paper. Due to the many unknown variables and constants involved in the asymptotic results of Theorems 3.2, 3.5, 3.6 and 3.8, and because of the numerous cases and different asymptotic limits which are studied, a unified theorem about the asymptotic relative efficiencies, which could be applicable in practice, does not seem conceivable. We consequently choose to compare numerically the efficiencies in Section 4.1. We then deduce in Section 4.2 which estimator should be chosen in practice for a given scenario, which we illustrate on a simulated multi-phase epidemic.

##### 4.1. Comparison of the estimators

In order to compare the performance of the estimators presented above, we perform a large number of simulations and examine the efficiencies of the estimators for finite sample sizes. More precisely, for a given infection parameter  $\theta_0$  and sample size  $(|\mathbf{x}_0|, n)$ , we define the average squared error of an estimator  $T$  of  $\theta_0$  as follows:

$$\frac{1}{1000} \sum_{i=1}^{1000} \left( T(\mathbf{x}^{(i)}) - \theta_0 \right)^2, \quad (4.1)$$

and the approximated relative efficiency of two estimators  $T_1$  and  $T_2$ :

$$e_{\theta_0}(T_1, T_2) := \frac{\sum_{i=1}^{1000} (T_2(\mathbf{x}^{(i)}) - \theta_0)^2}{\sum_{i=1}^{1000} (T_1(\mathbf{x}^{(i)}) - \theta_0)^2}, \quad (4.2)$$

where for each  $i = 1 \dots 1000$ ,  $\mathbf{x}^{(i)}$  is a simulation of length  $n$  of the branching process (2.1) with infection parameter  $\theta_0$ , such that  $|\mathbf{x}_0^{(i)}| = |\mathbf{x}_0|$  and  $\mathbf{x}_n^{(i)} \neq \mathbf{0}$ . For the sake of simplicity we omit in the notation of  $e_{\theta_0}(T_1, T_2)$  the dependence in  $(|\mathbf{x}_0|, n)$ . The large number of replications ensures that the average squared error (4.1) is an accurate approximation of the mean squared error  $\mathbb{E}_{\theta_0}[(T - \theta_0)^2]$ . In what follows, we say that for a given  $\theta_0$  and a given sample size  $(|\mathbf{x}_0|, n)$ , an estimator  $T_1$  is more efficient than  $T_2$  if  $e_{\theta_0}(T_1, T_2) > 1$ .

All simulations are performed using the branching process (2.1) with parameters  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$  and  $\mathbf{b} = (0.05, 0.05)$ . The criticality threshold of such a process is then  $\theta_0 = 10$  (see (3.2)). Note that in this simple 2-dimensional case, we are able to determine the value of  $\lambda^2$ , which plays a major role in the asymptotic limit distribution of  $\tilde{\theta}$  (see Theorem 3.2). We compute here that  $\lambda^2 < \rho_0$ , except for very large values of  $\theta_0$  ( $\theta_0 > 23$ ) which will not be taken account. As a consequence we are in this example only considering the case (3.8)

of Theorem 3.2, for which the convergence rates in  $n$  of  $\tilde{\theta} - \theta_0$  and  $\hat{\theta} - \theta_0$  to the normal distribution are equal up to a multiplicative constant. If  $\lambda^2 \geq \rho_0$ , the convergence rate of  $\tilde{\theta} - \theta_0$  to its limiting distribution is smaller than for  $\hat{\theta} - \theta_0$ , hence we do not investigate this case.

We first plot in Figure 1 the average squared errors of the estimators against the initial population size  $|\mathbf{x}_0|$ , for  $\theta_0 \in \{9, 10, 11\}$  and for a number of observations fixed at  $n = 10$ . The average errors are computed over the same 1000 simulations for each of the three estimators. Keeping in mind Remark 3.9, this explains why the plots for  $\tilde{\theta}$  and  $\hat{\theta}$  are indistinguishable. We observe that the average squared error for  $\tilde{\theta}$  is also extremely close and decreases as rapidly as for  $\hat{\theta}$ . We then plot in Figure 2 the average squared errors against the number of observations  $n$ , for  $\theta_0 \in \{9, 11\}$  and for an initial population size fixed at  $|\mathbf{x}_0| = 100$ . For  $\theta_0 = 10$  and  $\theta_0 = 11$  (critical and supercritical cases), the average squared errors for the three estimators are again indistinguishable at this scale. For  $\theta_0 = 9$ , we observe that the average errors are increasing as  $n$  becomes very large. This is due to the fact that, if  $n$  is large compared to  $|\mathbf{x}_0| = 100$ , the probability that  $\mathbf{X}_n \neq \mathbf{0}$  is low in the subcritical case. Hence the 1000 simulations correspond to rare events and are not typical. As we will see in Figure 4, the estimator  $\check{\theta}$  is in this case the most appropriate, intuitively because it takes into account the additional information that the simulated data have still not become extinct at time  $n$ .

Finally, in order to differentiate more precisely the performances of the estimators, we focus on estimators which have asymptotic properties for the same class of criticality and study their approximated relative efficiency defined by (4.2). We thus compare  $\tilde{\theta}$  against  $\hat{\theta}$  by plotting  $e_{\theta_0}(\tilde{\theta}, \hat{\theta})$  in Figure 3, and  $\check{\theta}$  against  $\hat{\theta}$  by plotting  $e_{\theta_0}(\check{\theta}, \hat{\theta})$  in Figure 4, for  $\theta_0 \in [8, 12]$ . We consider two sample sizes for  $|\mathbf{x}_0|$  ( $|\mathbf{x}_0| = 10$  and  $|\mathbf{x}_0| = 1000$ ) and for  $n$  ( $n = 10$  and  $n = 100$ ). What appears from these results is that  $\hat{\theta}$  is here the most efficient estimator when  $|\mathbf{x}_0| = 10$  and  $n = 10$ , no matter the class of criticality. It appears also that in the supercritical case,  $\tilde{\theta}$  and  $\hat{\theta}$  perform equivalently well and better than  $\check{\theta}$ , no matter the sample sizes. Finally,  $\check{\theta}$  is here the most efficient estimator in the subcritical case, for  $|\mathbf{x}_0| = 10$  and  $n = 100$ .

#### 4.2. Choice of the estimator in practice and illustration with a simulated multi-phase epidemic

Let us deduce from the above simulated results and from the theoretical study of Section 3 how to choose the appropriate estimator, given epidemic data, and illustrate our method on an example plotted in Figure 5.

Our example mimics the evolution of an epidemic during which three control measures are taken, at time 50, 60 and 70. It is obtained by simulation of the branching process (2.1) with parameters  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$ ,  $\mathbf{b} = (0.05, 0.05)$  and initial value  $\mathbf{x}_0 = (1, 1)$ . First we choose a supercritical infection parameter  $\theta_0 = 13$  and simulate over 50 time steps, and then simulate three phases with respective infection parameter  $\theta_0 = 11$ ,  $\theta_0 = 9.5$ ,  $\theta_0 = 7$ , over, respectively, 10,



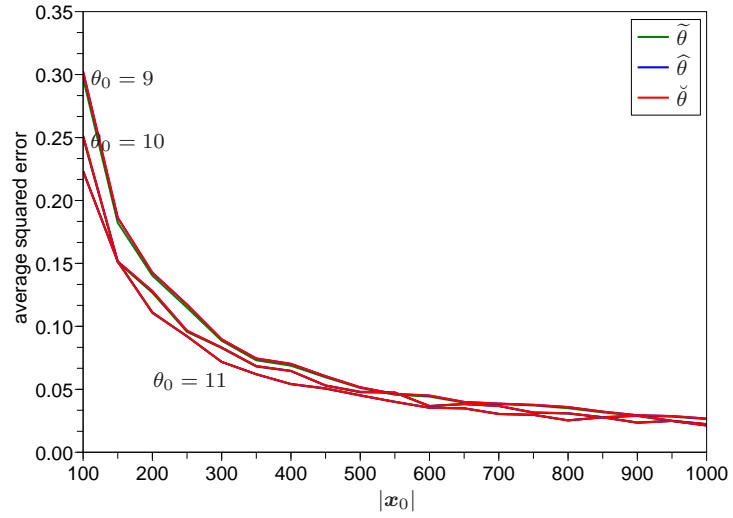


FIG 1. Average squared error (4.1) over 1000 simulations of (2.1) with parameters  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$ ,  $\mathbf{b} = (0.05, 0.05)$  and  $n = 10$  observations, against the initial population size  $|x_0|$ .

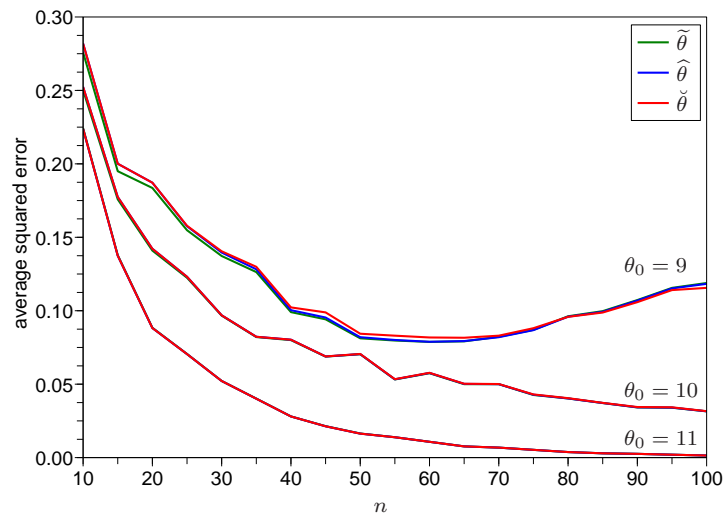


FIG 2. Average squared error (4.1) over 1000 simulations of (2.1) with parameters  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$ ,  $\mathbf{b} = (0.05, 0.05)$  and initial population size  $|x_0| = 100$ , against the number of observations  $n$ .

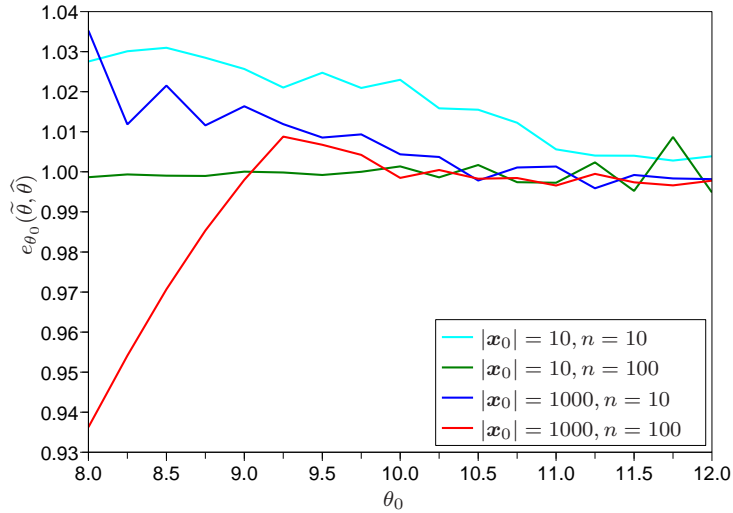


FIG 3. Approximated relative efficiency  $e_{\theta_0}(\tilde{\theta}, \hat{\theta})$  defined by (4.2), based on 1000 simulations of (2.1) with parameters  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$  and  $\mathbf{b} = (0.05, 0.05)$ , against the infection parameter  $\theta_0$  in the supercritical case, for different sample sizes  $(|x_0|, n)$ .

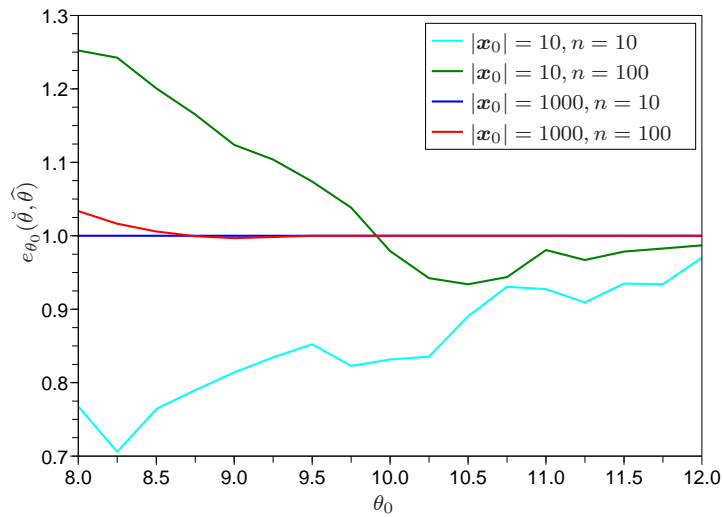


FIG 4. Approximated relative efficiency  $e_{\theta_0}(\check{\theta}, \hat{\theta})$  defined by (4.2), based on 1000 simulations of (2.1) with parameters  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$  and  $\mathbf{b} = (0.05, 0.05)$ , against the infection parameter  $\theta_0$  in the subcritical case, for different sample sizes  $(|x_0|, n)$ .

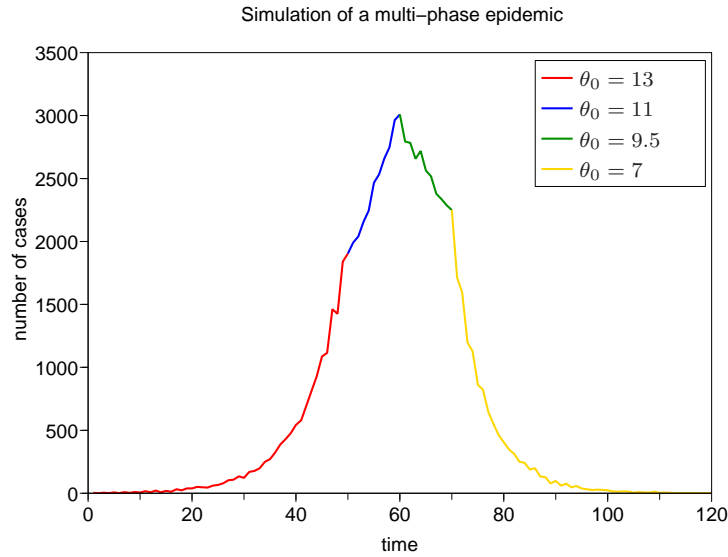


FIG 5. Simulated branching process of the form (2.1) with  $d = 2$ ,  $\mathbf{a} = (0.01, 0.08)$ ,  $\mathbf{b} = (0.05, 0.05)$ , and initial value  $\mathbf{x}_0 = (1, 1)$ . The infection parameter  $\theta_0$  is modified at time 50, 60 and 70.

TABLE 1

Estimations of  $\theta_0$  and 95% confidence intervals based on the data set  $(\mathbf{x}_0, \dots, \mathbf{x}_{120})$  plotted in Figure 5, for different phases corresponding to different data subsets. The first line corresponds to the estimations based on the whole data set

data	$\theta_0$	est. with $\tilde{\theta}_n^X$	est. with $\hat{\theta}$ , 95% conf. interval	est. with $\hat{\theta}_n^Z$ , %95 conf. interval
$(\mathbf{x}_0, \dots, \mathbf{x}_{120})$	–	9.99987	9.99988 [9.92236; 10.0774]	9.99973 [9.92182; 10.0776]
$(\mathbf{x}_0, \dots, \mathbf{x}_{50})$	13	13.0548	13.0514 [12.8334; 13.2694]	13.0512 [12.8464; 13.2561]
$(\mathbf{x}_{50}, \dots, \mathbf{x}_{60})$	11	11.1501	11.0087 [10.8585; 11.1590]	11.0087 [10.8666; 11.1508]
$(\mathbf{x}_{60}, \dots, \mathbf{x}_{70})$	9.5	9.59054	9.44354 [9.3142; 9.57289]	9.44354 [9.31881; 9.56827]
$(\mathbf{x}_{70}, \dots, \mathbf{x}_{120})$	7	7.45274	7.11740 [6.97098; 7.26381]	7.11660 [6.96973; 7.26346]
$(\mathbf{x}_{90}, \dots, \mathbf{x}_{120})$	7	7.39712	7.36783 [6.6480; 8.08765]	7.34980 [6.62794; 8.07166]

10 and 50 time steps. We thus obtain the data  $(\mathbf{x}_0, \dots, \mathbf{x}_{120})$ . The estimations provided by the three estimators on distinct phases are listed in Table 1, as well as the 95% confidence intervals stemming from Theorem 3.6 and Theorem 3.8. On the first line of Table 1, we report the estimations based on the whole data  $(\mathbf{x}_0, \dots, \mathbf{x}_{120})$ , which thus entails the different phases and does not correspond to a process with constant infection parameter.

Since each of the three estimators discussed in this paper is relevant for a branching process of the form (2.1) with a *constant* offspring distribution over time, it is necessary to work with data for which the infection parameter is known or assumed to remain unchanged. If the latter has been potentially modified during the course of the epidemic, for instance because of control or sanitary measures, then the corresponding phases of the epidemic must be isolated (as illustrated with our example in Figure 5). For each phase, the unknown infection parameter might then be estimated with the most suitable estimator.

Due to its numerous properties (Theorem 3.5 and Theorem 3.6),  $\hat{\theta}$  appears to be suitable in most cases, namely for any phase with large initial population size (such as the three last phases in Figure 5) or for a long growth phase (such as the first phase in Figure 5). Thanks to Theorem 3.6,  $\hat{\theta}$  provides in addition an explicit confidence interval, for  $|\mathbf{x}_0|$  large enough, which is of a great practical interest. By contrast, the rate of convergence of  $\tilde{\theta}$  to its limiting distribution is in practice unknown (Theorem 3.2). For these different reasons, we prefer  $\hat{\theta}$  over  $\tilde{\theta}$ , even though  $\tilde{\theta}$  might in some cases perform slightly better than  $\hat{\theta}$  (see Figure 3).

Similarly, we prefer in most cases  $\hat{\theta}$  over  $\check{\theta}$ . We recall that the estimations provided by these two estimators only differ from each other if the data set contains some sequences of  $d-1$  consecutive null values. However, for long decay phases with small initial population size (such as  $(\mathbf{x}_{90}, \dots, \mathbf{x}_{120})$  in Figure 5, for which  $|\mathbf{x}_{90}| = 176$ ), then  $\check{\theta}$  appears to be more relevant (see Figure 4 for  $n = 100$  and  $|\mathbf{x}_0| = 10$  in the subcritical case) and would be chosen over  $\hat{\theta}$ .

### Appendix A: Proofs related to Section 3.1

*Proof of Theorem 3.2.* We deduce Theorem 3.2 from Theorems 6.1–6.3 of [1], which state that if the process  $(\mathbf{X}_k)_{k \geq 0}$  is supercritical, then, on the set of non-extinction, the estimator  $\tilde{\rho}$  is strongly consistent:

$$\lim_{n \rightarrow \infty} \tilde{\rho} \stackrel{a.s.}{=} \rho_0, \tag{A.1}$$

and has the following asymptotic distribution: if  $\lambda^2 < \rho_0$ ,

$$\lim_{n \rightarrow \infty} \sqrt{W_0 (1 + \dots + \rho_0^{n-1})} (\tilde{\rho} - \rho_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, C_1); \tag{A.2}$$

if  $\lambda^2 = \rho_0$  and  $C_2 > 0$ ,

$$\lim_{n \rightarrow \infty} \sqrt{\frac{W_0 (1 + \dots + \rho_0^{n-1})}{n^{2\gamma-1}}} (\tilde{\rho} - \rho_0) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, C_2); \tag{A.3}$$

if  $\lambda^2 > \rho_0$ , there exist random variables  $H_n$  with  $\overline{\lim} |H_n| < \infty$ , such that

$$\lim_{n \rightarrow \infty} \left[ \frac{W_0 (1 + \dots + \rho_0^{n-1})}{\lambda^{n-1} (n-1)^{\gamma-1}} (\tilde{\rho} - \rho_0) - H_{n-1} \right] \stackrel{a.s.}{=} 0. \tag{A.4}$$

The strong consistency in Theorem 3.2 is immediate from (3.5) and (A.1). We then express  $\tilde{\theta} - \theta_0$  as a function of  $\tilde{\rho} - \rho_0$ , in order to deduce its asymptotic distribution from (A.2)–(A.4). We write

$$\begin{aligned} \tilde{\theta} - \theta_0 &= \left( \sum_{k=1}^d a_k \tilde{\rho}^{-k} \sum_{k=1}^d a_k \rho_0^{-k} \right)^{-1} \left[ \sum_{k=1}^d a_k (\rho_0^{-k} - \tilde{\rho}^{-k}) \right. \\ &\quad \left. + \sum_{k=1}^d a_k \rho_0^{-k} \sum_{k=1}^d b_k (\rho_0^{-k} - \tilde{\rho}^{-k}) - \sum_{k=1}^d b_k \rho_0^{-k} \sum_{k=1}^d a_k (\rho_0^{-k} - \tilde{\rho}^{-k}) \right] \end{aligned}$$

and use the fact that, for all  $k = 1 \dots d$ ,

$$\rho_0^{-k} - \tilde{\rho}^{-k} = (\tilde{\rho} - \rho_0) \frac{\sum_{l=1}^k \rho_0^{l-k} \tilde{\rho}^{-l}}{\rho_0},$$

in order to obtain

$$\begin{aligned} \tilde{\theta} - \theta_0 &= (\tilde{\rho} - \rho_0) \left[ \left( \rho_0 \sum_{k=1}^d a_k \tilde{\rho}^{-k} \sum_{k=1}^d a_k \rho_0^{-k} \right)^{-1} \left( \sum_{k=1}^d a_k \sum_{l=1}^k \rho_0^{l-k} \tilde{\rho}^{-l} \right. \right. \\ &\quad \left. \left. + \sum_{k=1}^d a_k \rho_0^{-k} \sum_{k=1}^d b_k \sum_{l=1}^k \rho_0^{l-k} \tilde{\rho}^{-l} - \sum_{k=1}^d b_k \rho_0^{-k} \sum_{k=1}^d a_k \sum_{l=1}^k \rho_0^{l-k} \tilde{\rho}^{-l} \right) \right]. \quad (\text{A.5}) \end{aligned}$$

By (A.1), the square bracket of (A.5) almost surely converges to  $\sqrt{C_0}$ , and (3.8)–(3.10) are immediately deduced from (A.2)–(A.4).  $\square$

**Appendix B: Proofs related to Section 3.2**

*Proof of Theorem 3.5.* We deduce from the basic limit result in the supercritical case that for any vector  $\mathbf{u} \in \mathbb{R}^d$ ,  $\lim_{n \rightarrow \infty} \rho_0^{-n} \mathbf{X}_n \stackrel{a.s.}{=} W_0 \mathbf{u} \cdot \boldsymbol{\eta}_0$ . The Toeplitz Lemma applied to the Toeplitz array  $a_{n,k} = \rho_0^k [\sum_{k=0}^n \rho_0^k]^{-1} \mathbf{1}_{k \leq n}$  thus implies that

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{+\infty} a_{n,k} \rho_0^{-k} \mathbf{u} \cdot \mathbf{X}_k \stackrel{a.s.}{=} W_0 \mathbf{u} \cdot \boldsymbol{\eta}_0,$$

which immediately leads to

$$\lim_{n \rightarrow \infty} \rho_0^{-n} \sum_{k=0}^n \mathbf{u} \cdot \mathbf{X}_k \stackrel{a.s.}{=} W_0 \frac{\rho_0}{\rho_0 - 1} \mathbf{u} \cdot \boldsymbol{\eta}_0. \quad (\text{B.1})$$

As a consequence, on the set of non-extinction  $\{W_0 > 0\}$ ,

$$\lim_{n \rightarrow \infty} \hat{\theta} = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \mathbf{e}_1 \cdot \mathbf{X}_k - \sum_{k=0}^{n-1} \mathbf{b} \cdot \mathbf{X}_k}{\sum_{k=0}^{n-1} \mathbf{a} \cdot \mathbf{X}_k} \stackrel{a.s.}{=} \frac{\rho_0 \mathbf{e}_1 \cdot \boldsymbol{\eta}_0 - \mathbf{b} \cdot \boldsymbol{\eta}_0}{\mathbf{a} \cdot \boldsymbol{\eta}_0} = \theta_0.$$

The last equality comes from the fact that by definition  $\boldsymbol{\eta}_0 \mathbf{M}(\theta_0) = \rho_0 \boldsymbol{\eta}_0$ , which implies in particular  $\boldsymbol{\Psi}(\theta_0) \cdot \boldsymbol{\eta}_0 = \rho_0 \boldsymbol{\eta}_{0,1}$ .

Note that we have the identity

$$\hat{\theta} - \theta_0 = \frac{\sum_{k=1}^n [\mathbf{X}_k - \mathbf{X}_{k-1} \mathbf{M}(\theta_0)] \cdot \mathbf{e}_1}{\sum_{k=0}^{n-1} \mathbf{a} \cdot \mathbf{X}_k}. \tag{B.2}$$

We deduce from the martingale central limit theorem given by Theorem 2.2 in [1] (applied to the vectors denoted in [1] as  $a_n = \mathbf{e}_1$  and to the real constants  $\gamma_n = 1$ ) that

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{W_0 \eta_{0,1} \rho_0 (1 + \dots + \rho_0^{n-1})}} \sum_{k=1}^n [\mathbf{X}_k - \mathbf{X}_{k-1} \mathbf{M}(\theta_0)] \cdot \mathbf{e}_1 \stackrel{\mathcal{D}}{=} \mathcal{N}(0, 1),$$

which combined with (B.1) leads to the desired asymptotic normality. □

The proof of Theorem 3.6 can be found in [11].

**Appendix C: Proofs related to Section 3.3**

*Proof of the strong consistency of  $\check{\theta}$  in Theorem 3.8.* According to Proposition 3.1 in [9], sufficient conditions for the strong consistency of  $\check{\theta}$  are that  $f(\cdot, \mathbf{Z}_{k-1})$  is Lipschitz, in the sense that there exists a nonnegative  $\sigma(\mathbf{Z}_0, \dots, \mathbf{Z}_{k-1})$ -measurable function  $C_k$  satisfying for all  $\theta_1, \theta_2 \in \Theta$ ,

$$|f(\theta_1, \mathbf{Z}_{k-1}) - f(\theta_2, \mathbf{Z}_{k-1})| \stackrel{a.s.}{\leq} C_k |\theta_1 - \theta_2|,$$

that  $\overline{\lim}_{k \rightarrow \infty} \mathbb{E}_{\theta_0} (\varepsilon_k^2 | \mathbf{Z}_{k-1}) \stackrel{a.s.}{<} \infty$ , and that for any  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \inf_{|\theta - \theta_0| \geq \delta} \sum_{k=1}^n (f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1}))^2 \stackrel{a.s.}{=} \infty. \tag{C.1}$$

The Lipschitz condition is satisfied thanks to (3.14), which shows that  $f'(\cdot, \mathbf{Z}_{k-1})$  is bounded on  $\Theta$ . The second condition follows from (3.15). Let  $\delta > 0$  and  $\theta \in \Theta$  such that  $|\theta - \theta_0| \geq \delta$ . We assume for convenience that  $\theta_0 > \theta$ . In order to prove (C.1), we apply the mean value theorem to the function  $f(\cdot, \mathbf{Z}_{k-1})$ , and obtain that there exists some  $t_k \in ]\theta, \theta_0[$  such that  $f'(t_k, \mathbf{Z}_{k-1}) = (f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1})) (\theta_0 - \theta)^{-1}$ . Consequently,

$$\begin{aligned} \sum_{k=1}^n (f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1}))^2 &= (\theta_0 - \theta)^2 \sum_{k=1}^n (f'(t_k, \mathbf{Z}_{k-1}))^2 \\ &= (\theta_0 - \theta)^2 \sum_{k=1}^n \mathbf{a} \cdot \mathbf{Z}_{k-1} \frac{(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} (1 + \Psi_d(t_k) Z_{k-d}) e^{-\Psi_d(t_k) Z_{k-d}})^2}{(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(t_k) Z_{k-d}})^4} \\ &\geq (\theta_0 - \theta)^2 \left(1 - (1 + \Psi_d(\theta_1)) e^{-\Psi_d(\theta_1)}\right)^2 \sum_{k=1}^n \mathbf{a} \cdot \mathbf{Z}_{k-1} \\ &\geq \delta^2 \left(1 - (1 + \Psi_d(\theta_1)) e^{-\Psi_d(\theta_1)}\right)^2 \mathbf{a} n, \end{aligned}$$

which implies (C.1). □

In order to provide the asymptotic distribution of  $\check{\theta} - \theta_0$  in Theorem 3.8, we shall state the positive recurrence of the Markov chain  $(\mathbf{Z}_k)_{k \geq 0}$ , and the finiteness of the first and second-order moments of its stationary distribution.

**Proposition C.1.** *Let us assume that the process  $(\mathbf{X}_k)_{k \geq 0}$  is subcritical. Then the homogeneous Markov chain  $(\mathbf{Z}_k)_{k \geq 0}$  is irreducible positive recurrent, and its stationary distribution  $\nu_{\theta_0}$  satisfies for all  $i, j = 1 \dots d$ ,*

$$\sum_{\mathbf{k} \in \mathbb{N}^d} k_i \nu_{\theta_0}(\mathbf{k}) < \infty, \quad \sum_{\mathbf{k} \in \mathbb{N}^d} k_i k_j \nu_{\theta_0}(\mathbf{k}) < \infty. \tag{C.2}$$

*Proof of Proposition C.1.* Clearly, the chain is irreducible: due to the Poisson random variables coming in play, any nonzero state is attainable from any other nonzero state in a finite time. The positive recurrence then follows from a criterion given e.g. in [14], Theorem 3.1: positive recurrence is implied if there exists a finite set  $A \subset \mathbb{N}^d \setminus \{\mathbf{0}\}$  and a non-negative function  $g$  on  $\mathbb{N}^d \setminus \{\mathbf{0}\}$  such that

$$\sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) g(\mathbf{j}) \leq g(\mathbf{i}) - 1, \quad \mathbf{i} \notin A. \tag{C.3}$$

Let  $g(\mathbf{i}) = \mathbf{i} \cdot \boldsymbol{\xi}$ , where  $\boldsymbol{\xi}$  is the eigenvector defined in Section 2. Then, for all  $\mathbf{i} \in \mathbb{N}^d \setminus \{\mathbf{0}\}$ ,

$$\begin{aligned} g(\mathbf{i}) - \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) g(\mathbf{j}) &= \mathbf{i} \cdot \boldsymbol{\xi} - \frac{1}{1 - P(\mathbf{i}, \mathbf{0})} \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} P(\mathbf{i}, \mathbf{j}) \mathbf{j} \cdot \boldsymbol{\xi} \\ &= \mathbf{i} \cdot \boldsymbol{\xi} - \frac{1}{1 - P(\mathbf{i}, \mathbf{0})} \mathbb{E}_{\theta_0}(\mathbf{X}_1 \cdot \boldsymbol{\xi} | \mathbf{X}_0 = \mathbf{i}) \\ &= \mathbf{i} \cdot \boldsymbol{\xi} \left( 1 - \frac{\rho_0}{1 - P(\mathbf{i}, \mathbf{0})} \right). \end{aligned}$$

Since  $\rho_0 < 1$ , the right term tends to  $+\infty$  as  $|\mathbf{i}| \rightarrow \infty$ . Consequently there exists some finite set  $A$  satisfying (C.3), and the chain is positive recurrent.

Let us now prove that its stationary measure  $\nu_{\theta_0}$  admits finite first-order moments. First, we point out that by definition of  $\mathbf{Z}_n$ , we have for all  $i = 1 \dots d$ ,

$$\sum_{\mathbf{k} \in \mathbb{N}^d} k_i \nu_{\theta_0}(\mathbf{k}) = \mathbb{E}_{\theta_0}(\lim_{n \rightarrow \infty} Z_{n,i}) = \mathbb{E}_{\theta_0}(\lim_{n \rightarrow \infty} Z_{n-i+1,1}) = \sum_{\mathbf{k} \in \mathbb{N}^d} k_1 \nu_{\theta_0}(\mathbf{k}) =: m^{\nu_{\theta_0}}. \tag{C.4}$$

It is thus enough to demonstrate (C.2) for  $i = 1$ . According to [15], Theorem 1, in order that  $\sum_{\mathbf{i} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} i_1 \nu_{\theta_0}(\mathbf{i}) < \infty$ , it suffices that for some non-empty finite set  $B$  and some function  $h$  with  $h(\mathbf{i}) \geq i_1, \mathbf{i} \notin B$ ,

$$\sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) h(\mathbf{j}) \leq h(\mathbf{i}) - i_1, \quad \mathbf{i} \notin B. \tag{C.5}$$

Taking  $h(\mathbf{i}) := \sum_{k=1}^d b_k i_k$  with

$$b_1 = \frac{2}{1 - (\Psi_1(\theta_0) + \dots + \Psi_d(\theta_0))}, \quad b_k = b_1 (\Psi_k(\theta_0) + \dots + \Psi_d(\theta_0)), \quad k = 2 \dots d,$$

we have for all  $\mathbf{i} \neq \mathbf{0}$ ,  $h(\mathbf{i}) \geq i_1$ , and

$$h(\mathbf{i}) - i_1 - \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) h(\mathbf{j}) \sim_{|\mathbf{i}| \rightarrow \infty} \frac{1}{1 - (\Psi_1(\theta_0) + \dots + \Psi_d(\theta_0))} i_1,$$

which ensures the existence of some finite set  $B$  satisfying (C.5).

Consequently,  $\sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} j_1 \nu_{\theta_0}(\mathbf{j}) < \infty$ , and the quantity  $m^{\nu_{\theta_0}}$  defined in (C.4) is finite. Due to the specific properties of the process  $(\mathbf{Z}_k)_{k \geq 0}$ , it is possible to deduce from this that the second-order moments of the stationary measure  $\nu_{\theta_0}$  are finite as well. Indeed, for all  $i = 1 \dots d - 1$ , using the inequality  $x(1 - e^{-x})^{-1} \leq 1 + x$ ,  $x \geq 0$ ,

$$\begin{aligned} \mathbb{E}_{\theta_0}(Z_n Z_{n-i}) &= \mathbb{E}_{\theta_0} \left[ Z_{n-i} \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{n-1}}{1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{n-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{n-d}}} \right] \\ &\leq \mathbb{E}_{\theta_0} \left[ Z_{n-i} \left( 1 + \Psi(\theta_0) \cdot \mathbf{Z}_{n-1} \right) \right], \end{aligned}$$

hence

$$\underline{\lim} \mathbb{E}_{\theta_0}(Z_n Z_{n-i}) \leq m^{\nu_{\theta_0}} + \max_{k=0 \dots d-1} \underline{\lim} \mathbb{E}_{\theta_0}(Z_n Z_{n-k}) \sum_{j=1}^d \Psi_j(\theta_0). \quad (\text{C.6})$$

Similarly,

$$\begin{aligned} \mathbb{E}_{\theta_0}(Z_n^2) &= \mathbb{E}_{\theta_0} \left[ \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{n-1}}{1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{n-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{n-d}}} \left( 1 + \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{n-1}}{1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{n-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{n-d}}} \right) \right] \\ &\leq \mathbb{E}_{\theta_0} \left[ \left( 2 + \Psi(\theta_0) \cdot \mathbf{Z}_{n-1} \right)^2 \right] \\ &= 4 + 4 \sum_{j=1}^d \Psi_j(\theta_0) \mathbb{E}_{\theta_0}(Z_{n-j}) + \sum_{j=1}^d \sum_{l=1}^d \Psi_j(\theta_0) \Psi_l(\theta_0) \mathbb{E}_{\theta_0}(Z_{n-j} Z_{n-l}), \end{aligned}$$

which by Fatou's lemma and (C.4) leads to (using the fact that  $\sum_{j=1}^d \Psi_j(\theta_0) < 1$ )

$$\underline{\lim} \mathbb{E}_{\theta_0}(Z_n^2) \leq 4 + 4m^{\nu_{\theta_0}} + \max_{k=0 \dots d-1} \underline{\lim} \mathbb{E}_{\theta_0}(Z_n Z_{n-k}) \sum_{j=1}^d \Psi_j(\theta_0).$$

Together with (C.6) this implies that

$$\begin{aligned} \max_{k=0 \dots d-1} \underline{\lim} \mathbb{E}_{\theta_0}(Z_n Z_{n-k}) \\ \leq 4 + 4m^{\nu_{\theta_0}} + \max_{k=0 \dots d-1} \underline{\lim} \mathbb{E}_{\theta_0}(Z_n Z_{n-k}) \sum_{j=1}^d \Psi_j(\theta_0), \end{aligned}$$



and thus

$$\max_{k=0\dots d-1} \underline{\lim} \mathbb{E}_{\theta_0} (Z_n Z_{n-k}) \leq \frac{4 + 4m^{\nu_{\theta_0}}}{1 - \sum_{j=1}^d \Psi_j(\theta_0)} < \infty.$$

We then obtain by means of Fatou’s lemma that for every  $i, j = 1 \dots d$ ,

$$\begin{aligned} \sum_{\mathbf{k} \in \mathbb{N}^d} k_i k_j \nu_{\theta_0}(\mathbf{k}) &= \mathbb{E}_{\theta_0} \left( \lim_{n \rightarrow \infty} Z_{n,i} Z_{n,j} \right) = \mathbb{E}_{\theta_0} \left( \lim_{n \rightarrow \infty} Z_n Z_{n-|i-j|} \right) \\ &\leq \underline{\lim} \mathbb{E}_{\theta_0} (Z_n Z_{n-|i-j|}) \leq \max_{k=0\dots d-1} \underline{\lim} \mathbb{E}_{\theta_0} (Z_n Z_{n-k}) < \infty. \end{aligned}$$

□

The asymptotic normality of  $\check{\theta} - \theta_0$  in Theorem 3.8 is a consequence of the following result.

**Proposition C.2.** *Let us assume that the process  $(\mathbf{X}_k)_{k \geq 0}$  is subcritical. Then the estimator  $\check{\theta}$  is asymptotically normally distributed:*

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{\left( \sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 \nu_{\theta_0}(\mathbf{j}) \right)^2}{\sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 f(\theta_0, \mathbf{j}) (\mathbf{a} \cdot \mathbf{j})^{-1/2} \nu_{\theta_0}(\mathbf{j})} (\check{\theta} - \theta_0) \stackrel{D}{=} \mathcal{N}(0, 1),$$

where  $f$  is given by (3.13).

*Proof of Proposition C.2.* Let us first quote from [4], Theorem 1.1 and 1.3, the following strong law of large numbers for homogeneous irreducible positive recurrent Markov chains which can be applied here thanks to Proposition C.1: for every  $\nu_{\theta_0}$ -integrable function  $g : \mathbb{N}^d \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\mathbf{Z}_k) \stackrel{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} g(\mathbf{j}) \nu_{\theta_0}(\mathbf{j}). \tag{C.7}$$

In order to prove Theorem C.2, we follow the steps of the proof of Proposition 6.1 in [9]. Writing the Taylor expansion of  $S'_n(\check{\theta})$  in the neighborhood of  $\theta_0$ , we obtain that for some  $\tilde{\theta}_n = \theta_0 + t_n(\check{\theta} - \theta_0)$ , with  $t_n \in ]0, 1[$ ,

$$\begin{aligned} \sqrt{n} (\check{\theta} - \theta_0) &= -\sqrt{n} \frac{S'_n(\theta_0)}{S''_n(\tilde{\theta}_n)} = \sqrt{n} \frac{2 \sum_{k=1}^n \varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{S''_n(\tilde{\theta}_n)} \\ &= \frac{\sum_{k=1}^n \varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \left( \frac{F_n}{n} \right)^{-1} \left( \frac{1}{2} \frac{S''_n(\tilde{\theta}_n)}{F_n} \right)^{-1}, \end{aligned} \tag{C.8}$$

where  $F_n := \sum_{k=1}^n (f'(\theta_0, \mathbf{Z}_{k-1}))^2$ . By (3.14), for all  $\mathbf{j} \in \mathbb{N}^d, \mathbf{j} \neq \mathbf{0}$ ,  $0 \leq f'(\theta_0, \mathbf{j}) \leq \sqrt{\mathbf{a} \cdot \mathbf{j}} (1 - e^{-\Psi_d(\theta_0)})^{-2}$ , hence we deduce by means of (C.7) and (C.2) that

$$\lim_{n \rightarrow \infty} \frac{F_n}{n} \stackrel{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 \nu_{\theta_0}(\mathbf{j}). \tag{C.9}$$

Let us now prove that

$$\lim_{n \rightarrow \infty} \frac{S_n''(\tilde{\theta}_n)}{F_n} \stackrel{a.s.}{=} 2. \tag{C.10}$$

Since  $S_n(\theta) = \sum_{k=1}^n (\varepsilon_k + f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1}))^2$ , it appears that (C.10) is true, as soon as the following holds:

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \frac{\left| \sum_{k=1}^n \varepsilon_k f''(\theta, \mathbf{Z}_{k-1}) \right|}{F_n} \stackrel{a.s.}{=} 0, \tag{C.11}$$

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \left( f'(\tilde{\theta}_n, \mathbf{Z}_{k-1}) \right)^2}{F_n} \stackrel{a.s.}{=} 1, \tag{C.12}$$

and

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \left( f(\theta_0, \mathbf{Z}_{k-1}) - f(\tilde{\theta}_n, \mathbf{Z}_{k-1}) \right) f''(\tilde{\theta}_n, \mathbf{Z}_{k-1})}{F_n} \stackrel{a.s.}{=} 0. \tag{C.13}$$

Let us prove (C.11)–(C.13). Note that, for every  $\mathbf{j} \neq \mathbf{0}$ ,  $f''(\theta, \mathbf{j}) = 0$  if  $[\mathbf{j}] \neq 0$ , and

$$f''(\theta, \mathbf{j}) = \frac{(a_d j_d)^{3/2} e^{-\Psi_d(\theta) j_d} \left[ e^{-\Psi_d(\theta) j_d} (\Psi_d(\theta) j_d + 2) + \Psi_d(\theta) j_d - 2 \right]}{\left( 1 - e^{-\Psi_d(\theta) j_d} \right)^3}$$

otherwise. First, (C.11) is given by a strong law of large numbers proved in [9], Proposition 5.1. The latter can be indeed applied since  $f''(\cdot, \mathbf{Z}_{k-1})$  fulfils the required Lipschitz condition, and  $\lim_n F_n \stackrel{a.s.}{=} \infty$  (as an immediate consequence of the stronger result (C.9)). In view of (C.12) we consider the function  $(f'(\theta, \mathbf{j}))^2$  and its derivative  $2f'(\theta, \mathbf{j})f''(\theta, \mathbf{j})$ . For all  $\theta \in \Theta$  and all  $\mathbf{j} \neq \mathbf{0}$  with  $[\mathbf{j}] = 0$ ,

$$\begin{aligned} |2f'(\theta, \mathbf{j})f''(\theta, \mathbf{j})| &\leq 4 \frac{(a_d j_d)^2 e^{-\Psi_d(\theta) j_d} (\Psi_d(\theta) j_d + 2)}{\left( 1 - e^{-\Psi_d(\theta) j_d} \right)^5} \\ &\leq \frac{4 \max_{x \geq 0} (x + 2)^3 e^{-x}}{\left( 1 - e^{-\Psi_d(\theta_{\min})} \right)^5} =: c_1. \end{aligned}$$

Consequently,

$$\frac{\left| \sum_{k=1}^n f'(\tilde{\theta}_n, \mathbf{Z}_{k-1})^2 - f'(\theta_0, \mathbf{Z}_{k-1})^2 \right|}{F_n} \leq c_1 \left| \tilde{\theta} - \theta_0 \right| \left( \frac{F_n}{n} \right)^{-1},$$

which by (C.9) and the strong consistency of  $\tilde{\theta}$  almost surely tends to 0. Writing

$$\frac{\sum_{k=1}^n f'(\tilde{\theta}_n, \mathbf{Z}_{k-1})^2}{F_n} = 1 + \frac{\sum_{k=1}^n \left( f'(\tilde{\theta}_n, \mathbf{Z}_{k-1})^2 - f'(\theta_0, \mathbf{Z}_{k-1})^2 \right)}{F_n},$$

this implies (C.12). It now remains to prove (C.13). With similar computations as above, one shows that there exists a deterministic constant  $c_2 > 0$  such that

$$\left| \frac{\sum_{k=1}^n \left( f(\theta_0, \mathbf{Z}_{k-1}) - f(\tilde{\theta}_n, \mathbf{Z}_{k-1}) \right) f''(\tilde{\theta}_n, \mathbf{Z}_{k-1})}{F_n} \right| \leq c_2 |\check{\theta} - \theta_0| \left( \frac{F_n}{n} \right)^{-1},$$

which thanks to (C.9) and the strong consistency of  $\check{\theta}$  implies (C.13).

In view of (C.8), we finally want to prove that  $\sum_{k=1}^n \varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1}) / \sqrt{n}$  converges in distribution. For this purpose, we define for every  $1 \leq k \leq n$ ,  $M_k^{(n)} := \sum_{l=1}^k \varepsilon_l f'(\theta_0, \mathbf{Z}_{l-1}) / \sqrt{n}$ . First, for every  $k \leq n$ ,  $\mathbb{E}_{\theta_0} (\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1}) / \sqrt{n} | \mathbf{Z}_{k-1}) = 0$ . Second,

$$\mathbb{E}_{\theta_0} \left( \left( \frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \right)^2 \middle| \mathbf{Z}_{k-1} \right) = \frac{(f'(\theta_0, \mathbf{Z}_{k-1}))^2 f(\theta_0, \mathbf{Z}_{k-1})}{n \sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}}}.$$

Hence for each  $n \geq 1$ ,  $(M_k^{(n)})_{1 \leq k \leq n}$  is a sequence of square integrable martingales. Let us denote by  $((M)_k^{(n)})_{1 \leq k \leq n}$  the associated Meyer process. By (C.2),

$$\sum_{\mathbf{j} \in \mathbb{N}^d} \frac{(f'(\theta_0, \mathbf{j}))^2 f(\theta_0, \mathbf{j})}{\sqrt{\mathbf{a} \cdot \mathbf{j}}} \nu_{\theta_0}(\mathbf{j}) \leq \frac{1}{(1 - e^{-\Psi_d(\theta_0)})^5} \sum_{\mathbf{j} \in \mathbb{N}^d} \Psi(\theta_0) \cdot \mathbf{j} \nu_{\theta_0}(\mathbf{j}) < \infty, \tag{C.14}$$

so by means of (C.7),

$$\begin{aligned} \lim_{n \rightarrow \infty} \langle M_n \rangle^{(n)} &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}_{\theta_0} \left( \left( \frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \right)^2 \middle| \mathbf{Z}_{k-1} \right) \\ &\stackrel{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} \frac{(f'(\theta_0, \mathbf{j}))^2 f(\theta_0, \mathbf{j})}{\sqrt{\mathbf{a} \cdot \mathbf{j}}} \nu_{\theta_0}(\mathbf{j}) := c^2. \end{aligned} \tag{C.15}$$

Third, using Cauchy-Schwarz and Bienaymé-Chebyshev inequalities,

$$\begin{aligned} &\sum_{k=1}^n \mathbb{E} \left[ \left| M_k^{(n)} - M_{k-1}^{(n)} \right|^2 \mathbf{1}_{\{|M_k^{(n)} - M_{k-1}^{(n)}| \geq \varepsilon\}} \middle| \mathbf{Z}_{k-1} \right] \\ &= \sum_{k=1}^n \mathbb{E}_{\theta_0} \left[ \left| \frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \right|^2 \mathbf{1}_{\{|\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}| \geq \varepsilon\}} \middle| \mathbf{Z}_{k-1} \right] \\ &\leq \sum_{k=1}^n \left( \mathbb{E}_{\theta_0} \left[ \left| \frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \right|^4 \middle| \mathbf{Z}_{k-1} \right] \right)^{\frac{1}{2}} \left( \mathbb{P}_{\theta_0} \left[ \left| \frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \right| \geq \varepsilon \middle| \mathbf{Z}_{k-1} \right] \right)^{\frac{1}{2}} \\ &\leq \frac{1}{n^{\frac{3}{2}} \varepsilon} \sum_{k=1}^n |f'(\theta_0, \mathbf{Z}_{k-1})|^3 (\mathbb{E}_{\theta_0} [\varepsilon_k^4 | \mathbf{Z}_{k-1}])^{\frac{1}{2}} (\mathbb{E}_{\theta_0} [\varepsilon_k^2 | \mathbf{Z}_{k-1}])^{\frac{1}{2}}. \end{aligned} \tag{C.16}$$

We have

$$|f'(\theta_0, \mathbf{Z}_{k-1})| \leq \frac{\sqrt{\mathbf{a} \cdot \mathbf{Z}_{k-1}}}{(1 - e^{-\Psi_d(\theta_0)})^2}, \quad \mathbb{E}_{\theta_0} (\varepsilon_k^2 | \mathbf{Z}_{k-1}) \leq \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{k-1}}{\mathbf{a} \cdot \mathbf{Z}_{k-1} (1 - e^{-\Psi_d(\theta_0)})},$$

and

$$\mathbb{E}_{\theta_0} (\varepsilon_k^4 | \mathbf{Z}_{k-1}) = \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{k-1} (1 + 3\Psi(\theta_0) \cdot \mathbf{Z}_{k-1})}{(\mathbf{a} \cdot \mathbf{Z}_{k-1})^2 (1 - e^{-\Psi_d(\theta_0)})},$$

hence

$$\begin{aligned} |f'(\theta_0, \mathbf{Z}_{k-1})|^3 (\mathbb{E}_{\theta_0} [\varepsilon_k^4 | \mathbf{Z}_{k-1}])^{\frac{1}{2}} (\mathbb{E}_{\theta_0} [\varepsilon_k^2 | \mathbf{Z}_{k-1}])^{\frac{1}{2}} \\ \leq \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{k-1} (1 + 3\Psi(\theta_0) \cdot \mathbf{Z}_{k-1})^{\frac{3}{2}}}{(1 - e^{-\Psi_d(\theta_0)})^7} \\ \leq \frac{\Psi(\theta_0) \cdot \mathbf{Z}_{k-1} + \sqrt{3} (\Psi(\theta_0) \cdot \mathbf{Z}_{k-1})^{\frac{3}{2}}}{(1 - e^{-\Psi_d(\theta_0)})^7}. \end{aligned} \quad (\text{C.17})$$

by Proposition C.1 the stationary distribution  $\nu_{\theta_0}$  has finite second-order moments, we deduce from (C.16) and (C.17) by virtue of (C.7) that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} \left[ \left| M_k^{(n)} - M_{k-1}^{(n)} \right|^2 \mathbf{1}_{\{|M_k^{(n)} - M_{k-1}^{(n)}| \geq \varepsilon\}} \middle| \mathbf{Z}_{k-1} \right] \stackrel{a.s.}{=} 0.$$

This convergence result together with (C.15) enables us to use a central limit theorem for sequences of martingales (see e.g. [13, 12]), leading to the fact that  $\lim_{n \rightarrow \infty} M_n^{(n)} \stackrel{\mathcal{D}}{=} \mathcal{N}(0, c^2)$ , i.e.

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}} \stackrel{\mathcal{D}}{=} \mathcal{N} \left( 0, \sum_{\mathbf{j} \in \mathbb{N}^d} \frac{(f'(\theta_0, \mathbf{j}))^2 f(\theta_0, \mathbf{j})}{\sqrt{\mathbf{a} \cdot \mathbf{j}}} \nu_{\theta_0}(\mathbf{j}) \right). \quad (\text{C.18})$$

Finally, (C.8) together with (C.9), (C.10), (C.18) and Slutsky's theorem imply that

$$\lim_{n \rightarrow \infty} \sqrt{n} (\check{\theta} - \theta_0) \stackrel{\mathcal{D}}{=} \mathcal{N} \left( 0, \frac{\sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 f(\theta_0, \mathbf{j}) (\mathbf{a} \cdot \mathbf{j})^{-1/2} \nu_{\theta_0}(\mathbf{j})}{\left( \sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 \nu_{\theta_0}(\mathbf{j}) \right)^2} \right).$$

□

*Proof of the asymptotic normality of  $\check{\theta} - \theta_0$  in Theorem 3.8.* The result is immediately deduced from Proposition C.2 as soon as we prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \frac{(f'(\check{\theta}, \mathbf{Z}_k))^2 f(\check{\theta}, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} \stackrel{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} \frac{(f'(\theta_0, \mathbf{j}))^2 f(\theta_0, \mathbf{j})}{\sqrt{\mathbf{a} \cdot \mathbf{j}}} \nu_{\theta_0}(\mathbf{j}), \quad (\text{C.19})$$

as well as the equivalent result for the numerator. For this purpose, we write

$$\begin{aligned} \sum_{k=0}^n \frac{\left(f'(\check{\theta}, \mathbf{Z}_k)\right)^2 f(\check{\theta}, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} &= \sum_{k=0}^n \frac{\left(f'(\theta_0, \mathbf{Z}_k)\right)^2 f(\theta_0, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} \\ &+ \sum_{k=0}^n \left[ \frac{\left(f'(\check{\theta}, \mathbf{Z}_k)\right)^2 f(\check{\theta}, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} - \frac{\left(f'(\theta_0, \mathbf{Z}_k)\right)^2 f(\theta_0, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} \right], \end{aligned} \quad (\text{C.20})$$

and show that  $(f'(\cdot, \mathbf{j}))^2 f(\cdot, \mathbf{j}) (\mathbf{a} \cdot \mathbf{j})^{-1/2}$  has a bounded derivative and is thus Lipschitz:

$$\left| \frac{2f''(\theta, \mathbf{j})f'(\theta, \mathbf{j})f(\theta, \mathbf{j}) + (f'(\theta, \mathbf{j}))^3}{\sqrt{\mathbf{a} \cdot \mathbf{j}}} \right| \leq 2c_1 \frac{\Psi(\theta_{\max}) \cdot \mathbf{j}}{(1 - e^{-\Psi_d(\theta_{\min})})^3},$$

which enables to write

$$\begin{aligned} \frac{1}{n+1} \sum_{k=0}^n \left| \frac{\left(f'(\check{\theta}, \mathbf{Z}_k)\right)^2 f(\check{\theta}, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} - \frac{\left(f'(\theta_0, \mathbf{Z}_k)\right)^2 f(\theta_0, \mathbf{Z}_k)}{\sqrt{\mathbf{a} \cdot \mathbf{Z}_k}} \right| \\ \leq \left| \check{\theta} - \theta_0 \right| \frac{2c_1}{(1 - e^{-\Psi_d(\theta_{\min})})^3} \frac{1}{n+1} \sum_{k=0}^n \Psi(\theta_{\max}) \cdot \mathbf{Z}_k. \end{aligned} \quad (\text{C.21})$$

By the strong consistency of  $\check{\theta}$  together with (C.7) and (C.2), (C.21) almost surely tends to zero. Combined with (C.7) and (C.14) in (C.20), this implies (C.19).  $\square$

## References

- [1] ASMUSSEN, S. AND KEIDING, N. (1978). Martingale central limit theorems and asymptotic estimations theory for multitype branching processes. *Adv. Appl. Prob.* **10**, 109–129. [MR0471027](#)
- [2] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*, Springer-Verlag, New-York. [MR0373040](#)
- [3] BECKER, N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics* **33**, 515–522. [MR0654337](#)
- [4] BILLINGSLEY, P. (1961). Statistical inference for Markov processes. *Statistical Research Monographs, Vol. II*. The University of Chicago Press, Chicago, Ill. [MR0123419](#)
- [5] DION, J. P. (1972). Estimation des probabilités initiales et de la moyenne d'un processus de Galton-Watson. Thèse de doctorat, Université de Montréal. [MR2940527](#)
- [6] DION, J. P. AND YANEV, N. M. (1997). Limit theorems and estimation theory for branching processes with an increasing random number of ancestors. *J. Appl. Probab.* **34**, 309–327. [MR1447337](#)

- [7] HACCOU, P., JAGERS, P. AND VATUTIN, V. A. (2005). *Branching Processes: Variation, Growth and Extinction of Populations*, Cambridge Studies in Adaptive Dynamics, Cambridge. [MR2429372](#)
- [8] HARRIS, T. E. (1948). Branching processes. *Ann. Math. Statist.* **19**, 474–494. [MR0027465](#)
- [9] JACOB, C. (2009). Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.* **38**, 566–597. [MR2590051](#)
- [10] KEIDING, N. AND LAURITZEN, S. L. (1978). Marginal maximum likelihood estimates and estimation of the offsprings mean in a branching process. *Scand. J. Statist.* **5**, 106–110. [MR0497064](#)
- [11] PÉNISSON, S. AND JACOB C. (2012). Stochastic methodology for the study of an epidemic decay phase, based on a branching model. *Int. J. Stoch. Anal.*, Art. ID 598701, 32 pp. [MR2999458](#)
- [12] PRAKASA RAO, B. L. S. (1999). *Semimartingales and Their Statistical Inference*, Monographs on Statistics and Applied Probability 83, Chapman & Hall/CRC. [MR1689166](#)
- [13] REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Probab. Theory Relat. Fields* **51**, 269–286. [MR0566321](#)
- [14] TWEEDIE, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stoch. Proc. Appl.* **3**, 385–403. [MR0436324](#)
- [15] TWEEDIE, R. L. (1983). The existence of moments for stationary Markov chains. *J. Appl. Probab.* **20**, 191–196. [MR0688095](#)