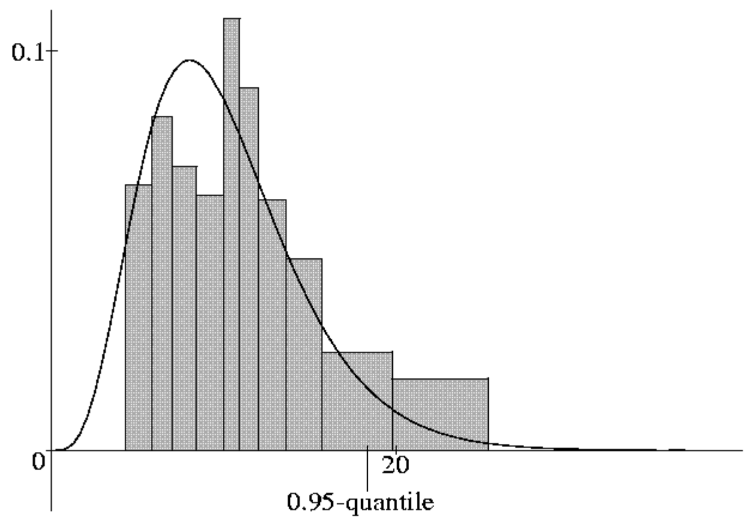
UNIVERSITÄT POTSDAM
Institut für Mathematik

Estimation of the infection parameter in the different phases of an epidemic modeled by a branching process

Sophie Pénisson



Mathematatische Statistik und
Wahrscheinlichkeitstheorie

Universität Potsdam – Institut für Mathematik

Mathematische Statistik und Wahrscheinlichkeitstheorie

Estimation of the infection parameter in the different phases of an epidemic modeled by a branching process

Sophie Pénisson

Institut für Mathematik der Universität Potsdam
e-mail: penisson@math.tu-berlin.de

Preprint 2010/01

März 2010

Impressum

Herausgeber:	Mathematische Statistik und Wahrscheinlichkeitstheorie am Institut für Mathematik
Adresse:	Universität Potsdam PF 60 15 53 14415 Potsdam
Telefon:	
Fax:	+49-331-977 1500
E-mail:	+49-331-977 1578
	neisse@math.uni-potsdam.de

ISSN 1613-3307

Estimation of the infection parameter in the different phases of an epidemic modeled by a branching process*

Sophie Pénisson†

March 2010

Abstract

The aim of this paper is to build and compare estimators of the infection parameter in the different phases of an epidemic (growth and extinction phases). The epidemic is modeled by a Markovian process of order $d \geqslant 1$ (allowing non-Markovian life spans), and can be written as a multitype branching process. We propose three estimators suitable for the different classes of criticality of the process, in particular for the subcritical case corresponding to the extinction phase. We prove their consistency and asymptotic normality for two asymptotics, when the number of ancestors (resp. number of generations) tends to infinity. We illustrate the asymptotic properties with simulated examples, and finally use our estimators to study the infection intensity in the extinction phase of the BSE epidemic in Great-Britain.

Keywords: multitype branching process; conditioned branching process; estimator; CLSE; consistency and asymptotic normality; epidemiology; SEI.

Mathematics Subject Classification. 60J80; 60J85; 62P10; 62M05; 62F12; 62J02.

1 Introduction

The purpose of this paper is to quantify the infection of an epidemic in its different phases (growth and extinction) by providing appropriate estimators of the infection parameter for each of these phases. The epidemic is modeled by a Markovian process of order $d \geqslant 1$ with Poissonian transitions, which can be seen as a multitype Bienaymé-Galton-Watson (BGW) branching process with d types corresponding to the memory of the process. The method proposed in this paper can thus apply to any multitype branching process of the same form, for which one wishes to estimate a parameter acting affinely on the mean matrix. The epidemic model which is used here has been elaborated in [12], and is suitable for any rare transmissible $SEIR$ disease in a large branching population following a Reed-Frost model for the infection. The process corresponds to the incidence of the clinical cases, which are assumed to be the only available observations.

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.* [8]). The estimation of a key parameter such as the mean number of offsprings or of the Perron's root (largest eigenvalue of the mean matrix), which determines 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, 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. S. Asmussen and N. Keiding, however, introduced in [1] an explicit estimator based only on the total generation sizes, which is thus of direct practical applicability for our model. We deduce from this estimator a first estimator of the infection parameter. Despite the potentially large order of the Markovian process that we

*Research supported by the Université franco-allemande and the IRTG Stochastic Models of Complex Processes. The author especially wishes to thank her PhD thesis supervisor, C. Jacob (INRA, France).

†Institut für Mathematik, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany; email: penisson@math.tu-berlin.de

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) with different asymptotic properties.

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 time tends to infinity. It could be of a great mathematical interest to study the asymptotic when both the number of ancestors and the number of generations of the branching process simultaneously tend to infinity, as it is done in [7] for the single-type case, but we choose to focus on asymptotics of an immediate practical interest. We first present in Section 3.1 a CLSE which is consistent and asymptotically normal, as the initial population size grows to infinity. This estimator is thus appropriate either in the growth phase of the epidemic or in its decay phase, provided that the initial time of the model corresponds to a large number of clinical cases. In Section 3.2 we focus on the subcritical case that is particularly designed for the extinction phase, and provide a CLSE based on the process conditioned on its non-extinction at the current time. This estimator is consistent and asymptotically normal, as time tends to infinity. This last result is based on the knowledge of the asymptotic distribution of the conditioned branching process, the so-called Yaglom limit. We finally provide in Section 3.3 an explicit estimator derived from the estimator of the Perron's root introduced in [1], and we deduce the consistency and asymptotic normality of our estimator in the supercritical case, on the set of non-extinction, as time tends to infinity. This last estimator is especially suitable in the growth phase of the epidemic. In Section 4 we compare these three estimators for several values of initial population size and time, and illustrate by means of simulations their asymptotic distributions. We next provide in Section 5 an other illustration, this time on a concrete biological problem, the Bovine Spongiform Encephalopathy (BSE) epidemic in Great-Britain. We finally conclude in Section 6 on the relevance of the estimators presented in this paper.

Whenever it is possible, we point out potential generalizations to wider contexts than $SEIR$ diseases, in particular to non-epidemiological problems.

2 The model

In order to allow the reader to apply this method to other problems than the estimation of the infection parameter in $SEIR$ diseases, we first describe in Section 2.1 the underlying mathematical process, before detailing in Section 2.2 this process in our specific epidemiological context.

2.1 The general mathematical model

Throughout this paper we consider the following Markovian process of order $d \geqslant 1$,

$$X_n = \sum_{k=1}^{d} \sum_{i=1}^{X_{n-k}} \zeta_{n-k,n,i}, \tag{2.1}$$

where the $\{\zeta_{n-k,n,i}\}_i$ are i.i.d. given $\mathcal{F}_{n-1} := \sigma\left(\{X_{n-k}\}_{k\geqslant 1}\right)$, and follow a Poisson distribution with some parameter Ψ_k independent of n (time-homogeneous setting). Moreover, the $\{\zeta_{n-k,n,i}\}_{i,k}$ are assumed to be independent given $\mathcal{F}_{n-1}$. The quantity k represents the maturation period needed to produce the offsprings $\zeta_{n-k,n,i}$. The particles thus have a non-Markovian random life span. We 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 branching processes, we refer to [2]. We easily derive the conditional law of X_n,

$$X_n|(X_{n-1}, \ldots, X_{n-d}) \sim \mathcal{P}oisson\left(\sum_{k=1}^{d} X_{n-k}\Psi_k\right). \tag{2.2}$$

Moreover, as shown in [12], Proposition 3.1, X_n may be written as a multitype BGW process. We define the d-dimensional process $\mathbf{X}_n := (X_{n,1}, \ldots, X_{n,d})$ such that, for all $i = 1 \ldots d$, $X_{n,i} := X_{n-i+1}$ (hence the first coordinate $X_{n,1}$ corresponds to the value of the single-type process at time

n). Then $(\mathbf{X}_n)_{n\geqslant 0}$ is a BGW process with offspring generating function

$$\begin{cases} f_i(\mathbf{r}) := \sum_{k=0}^{\infty} \frac{(\Psi_i)^k}{k!} e^{-\Psi_i} r_1^k r_{i+1} = e^{-\Psi_i(1-r_1)} r_{i+1}, \quad i = 1 \dots d-1, \\ f_d(\mathbf{r}) := \sum_{k=0}^{\infty} \frac{(\Psi_d)^k}{k!} e^{-\Psi_d} r_1^k = e^{-\Psi_d(1-r_1)}, \end{cases} \tag{2.3}$$

and mean matrix

$$\mathbf{M} := \begin{pmatrix} \Psi_1 & 1 & 0 & \dots & 0 \\ \Psi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \Psi_{d-1} & 0 & \dots & \dots & 1 \\ \Psi_d & 0 & \dots & \dots & 0 \end{pmatrix}. \tag{2.4}$$

The process $(\mathbf{X}_n)_{n\geqslant 0}$ is obviously nonsingular. We assume throughout this paper that $\Psi_d > 0$, and that there exists some $i = 1 \dots d-1$ with $\Psi_i > 0$, such that the process is positive regular. Moreover, as shown in [12], the process satisfies the $X \log X$ assumption: for all $i, j = 1 \dots d$, denoting $\mathbf{e_i} := (0, \dots, 1, \dots, 0)$ the basis vector of $\mathbb{N}^d$,

$$\mathbb{E}\left[X_{1,j} \ln X_{1,j} | \mathbf{X}_0 = \mathbf{e_i}\right] < \infty. \tag{2.5}$$

The theory of multitype positive regular and nonsingular BGW processes implies that the extinction of the process $(\mathbf{X}_n)_{n\geqslant 0}$ occurs almost surely if and only if the Perron's root ρ of the mean matrix $\mathbf{M}$ is smaller than or equal to 1. A computation of $\det(\mathbf{M} - \rho \mathbf{I})$ shows that ρ is solution of the equation

$$\sum_{k=1}^{d} \Psi_k \rho^{-k} = 1. \tag{2.6}$$

As mentioned in [12] (Proposition 3.1), this implies that

$$\rho \leqslant 1 \iff R := \sum_{k=1}^{d} \Psi_k \leqslant 1. \tag{2.7}$$

In addition, the distributions of quantities such as the extinction probability, the extinction time, and the tree size until extinction can be easily derived from the model (see [12]).

2.2 The epidemic model

It is shown in [12] that the previous age-dependent process (2.1) can be obtained as the limit of a more complex age and population-dependent process $\mathbf{N}_n := \left(N_n^k\right)_{k\in\mathcal{T}}$, describing the population-size at time n for each type $k \in \mathcal{T}$ (corresponding *e.g.* to health stages, locations, ages *etc.*). We first very briefly recall this result, and then describe in which way this can be applied in order to obtain our epidemic model. The number of k individuals at time n is given by

$$N_n^k = \sum_{l=1}^{a_M} \sum_{h\in\mathcal{T}} \sum_{i=1}^{N_{n-l}^h} Y_{n-l,n,i}^{(h),k}, \tag{2.8}$$

where a_M is the largest survival age, and $Y_{n-l,n,i}^{(h),k}$ is the number of k individuals generated at time n by individual i belonging to the type h at time $n-l$. This number $Y_{n-l,n,i}^{(h),k}$ of "mathematical" offsprings depends on the individual transition of i from h to k, as well as on the number of its "true" offsprings and their respective transition from their initial type to k.

Assuming the existence of a subset $\mathcal{K} \subset \mathcal{T}$ of *rare types*, and eventually of a subset $\mathcal{K}' \subset \mathcal{K}$ corresponding to rare types "of interest", the authors of [12] prove in Proposition 5.1 that, under technical assumptions (which are rather weak in the epidemiological context), the process N_n^k converges in distribution, for all $k \in \mathcal{K}$, as the initial population size $N_0 := \sum_{k\in\mathcal{T}} N_0^k$ tends to infinity. Moreover, the process summed on the rare types of interest, $X_n := \sum_{k\in\mathcal{K}'} X_n^k$, where $X_n^k \overset{\mathcal{D}}{:=} \lim_{N_0} N_n^k$, is a process of the form (2.1).

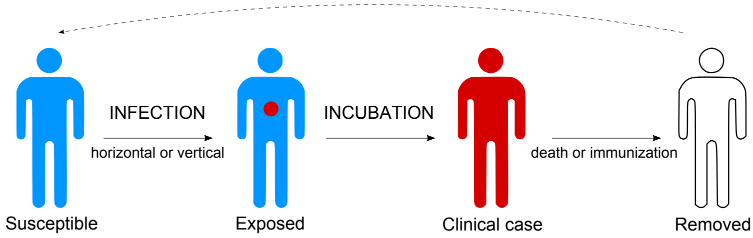


Figure 1: Evolution of the health status of an individual in a *SEIR* disease.

In the epidemiological context of a *SEIR* disease (see Figure 1) with horizontal and vertical infection routes, where the types are the health states, if $\mathcal{K}$ concerns all the infected states and $\mathcal{K}' \subset \mathcal{K}$, the clinical state, then we obtain that the process $X_n := \lim_{N_0} \sum_{k \in \mathcal{K}'} N_n^k$, corresponding to the incidence of the clinical cases at time n, assuming that the initial population size N_0 is very large, is of the form (2.1) ([10], [12]):

$$X_n = \sum_{k=1}^{d} \sum_{i=1}^{X_{n-k}} \zeta_{n-k,n,i}, \tag{2.9}$$

where $d := a_M - 1$, and $\zeta_{n-k,n,i} | \mathcal{F}_{n-1} \sim Poisson\,(\Psi_k)$, with, for all $k = 1 \dots d$, assuming that the infection parameters and the incubation time distribution do not depend on the age at infection,

$$\Psi_k := \sum_{i=k+1}^{a_M} \left(\theta_{hor.} + \mathbf{1}_{\{i=k+1\}}\theta_{vert.}\right) \frac{S_i}{\sum_{j=1}^{a_M} S_j} P_{inc.}(k). \tag{2.10}$$

S_k is the probability that the age of death of S and E individuals is larger than k, $P_{inc.}(k)$ denotes the probability that the intrinsic incubation period equals k, $\theta_{hor.}$ is the horizontal infection parameter (mean number per infective and per time-unit of newly infected individuals *via* horizontal transmission), and $\theta_{vert.}$ is the vertical infection parameter (probability for a newborn with an infectious mother to be infected at birth).

3 Estimation of the infection parameter

Throughout this section we consider the BGW branching process $(\mathbf{X}_n)_{n \geqslant 0}$ introduced in Section 2.1, with generating function (2.3) and mean matrix (2.4). We assume that the Ψ_k's affinely depend on some unknown parameter θ_0: for all $k = 1 \dots d$,

$$\Psi_k(\theta_0) = a_k \theta_0 + b_k, \tag{3.1}$$

where a_k and b_k are known. In the epidemiological context of rare *SEIR* diseases in large populations, θ_0 would correspond either to the horizontal or vertical infection parameter (see (2.10)). In the following, we denote by $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\Psi}(\theta_0)$ the d-dimensional vectors with coordinates a_k, b_k and $\Psi_k(\theta_0)$ respectively.

Our aim is to provide estimators of θ_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 is suitable regarding the available data. As mentioned in the introduction, we first investigate the numerous results in the literature dedicated to the estimation theory for general branching processes, in order to find an appropriate estimator for our model. In its early paper [9] in 1948, T. E. 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, now referred to as the *Harris estimator*, based on observed values of the individual offspring size for each individual in each generation. The estimator is

$$\widehat{m}_n^{MLE} := \frac{X_1 + \ldots + X_n}{X_0 + \ldots + X_{n-1}}, \tag{3.2}$$

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

$$\widehat{m}_n^{CLSE} := \arg\min_{m \geqslant 0} \sum_{k=1}^{n} \frac{(X_k - mX_{k-1})^2}{X_{k-1}}. \tag{3.3}$$

Similar estimation problems are considered in the multitype case. In [1], S. Asmussen and N. Keiding proposed a maximum likelihood estimator of the Perron's root ρ_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 [13] that this estimator is also the maximum likelihood estimator solely based on the observations at each generation of the total number of individuals of type j whose parents were of type i, for every $i, j = 1 \ldots d$. However in epidemiology this kind of variables are generally not observable. For our model this would imply indeed that, considering the number of clinical 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 generations, or on the total size of the generations, such as the other estimator presented in [1],

$$\widetilde{\rho}_n = \frac{|\mathbf{X}_1| + \ldots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \ldots + |\mathbf{X}_{n-1}|}, \tag{3.4}$$

where $|.|$ denotes the $\mathcal{L}^1$-norm $|\mathbf{X}_k| := X_{k,1} + \ldots + X_{k,d}$. For $d = 1$, $\widetilde{\rho}_n$ clearly reduces to the Harris estimator defined in (3.2). Note that the relation (2.6) implies that

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

Hence an estimation of ρ_0 would provide an estimation of θ_0 (the opposite is not true since ρ_0 cannot in general be expressed as an explicit function of θ_0). In the supercritical case, the estimator $\widetilde{\rho}_n$ was shown to be consistent, as $n \to \infty$, on the set of non-extinction, with an explicit asymptotic distribution ([1]).

Due to the Poissonian character of the transitions of the process $(\mathbf{X}_n)_{n\geqslant 0}$, it is possible, in our setting, to express the joint probability function of the observations $X_0, \ldots, X_n$, 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 θ_0, the logarithm of the other being $L(\theta_0) := -\theta_0 \sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1} + \sum_{k=1}^{n} X_k \ln(\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1})$, where $\mathbf{u} \cdot \mathbf{v}$ denotes the scalar product $\sum_{k=1}^{d} u_k v_k$. The MLE of θ_0 based on the observations $X_0, \ldots, X_n$ is thus a solution of $L'(\theta) = 0$, where $L'(\theta) = -\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1} + \sum_{k=1}^{n} X_k (\boldsymbol{a} \cdot \mathbf{X}_{k-1}) (\boldsymbol{\Psi}(\theta) \cdot \mathbf{X}_{k-1})^{-1}$. This equation has in general no explicit solution, except for simple cases such as the one-dimensional case $d = 1$, or the linear case $\boldsymbol{b} = \mathbf{0}$. The MLE is then, respectively,

$$\widehat{\theta}_n^{MLE} \overset{d=1}{=} \frac{\sum_{k=1}^{n} (X_k - bX_{k-1})}{\sum_{k=1}^{n} aX_{k-1}}, \quad \widehat{\theta}_n^{MLE} \overset{\boldsymbol{b}=\mathbf{0}}{=} \frac{\sum_{k=1}^{n} X_k}{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}}. \tag{3.6}$$

As shown later (see (3.10)), it corresponds in these cases to the CLSE of θ_0. It is however in general not the case, and we choose to focus on the CLSE.

In Section 3.1 we first study the weighted CLSE

$$\widehat{\theta}_{|\mathbf{X}_0|}^{X} := \arg\min_{\theta \in \Theta} \sum_{k=1}^{n} \frac{[X_k - \mathbb{E}_\theta (X_k | \mathbf{X}_{k-1})]^2}{\boldsymbol{a} \cdot \mathbf{X}_{k-1}}, \tag{3.7}$$

and its asymptotic properties, as $|\mathbf{X}_0| \to \infty$, for any class of criticality. Since we are only interested in the asymptotic in $|\mathbf{X}_0|$, we omit for the sake of clarity the subscript n in the estimator. In a second instance, since we aim at finding an estimator with asymptotic properties, as $n \to \infty$, holding in the subcritical case, we consider in Section 3.2 the CLSE associated with the conditioned process $\mathbf{Z}_k := (\mathbf{X}_k | \mathbf{X}_k \neq \mathbf{0})$,

$$\widehat{\theta}_n^{Z} := \arg\min_{\theta \in \Theta} \sum_{k=1}^{n} \frac{[Z_k - \mathbb{E}_\theta (Z_k | \mathbf{Z}_{k-1})]^2}{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}, \tag{3.8}$$

where $Z_k := Z_{k,1}$. Finally, thanks to relation (3.5), we derive from the estimator (3.4) a third estimator of θ_0,

$$\widetilde{\theta}_n^X := \frac{1 - \sum_{k=1}^d b_k \widetilde{\rho}_n^{-k}}{\sum_{k=1}^d a_k \widetilde{\rho}_n^{-k}}, \tag{3.9}$$

and deduce in Section 3.3 from [1] asymptotic properties of $\widetilde{\theta}_n^X$, as $n \to \infty$, in the supercritical case, on the set of non-extinction.

In the following, we denote $\Theta :=]\theta_{min}, \theta_{max}[$ with $\theta_{max} > \theta_{min} > 0$.

3.1 A CLSE with asymptotic properties, as $|\mathbf{X}_0| \to \infty$

In this section, we provide an estimator with asymptotic properties, as the initial population size $|\mathbf{X}_0|$ tends to infinity, holding for any class of criticality. We consider the weighted CLSE based on the process $Y_k := X_k / \sqrt{\boldsymbol{a} \cdot \mathbf{X}_{k-1}}$,

$$\widehat{\theta}_{|\mathbf{X}_0|}^X := \arg\min_{\theta \in \Theta} \sum_{k=1}^n \left(Y_k - \mathbb{E}_\theta\left(Y_k | \mathbf{X}_{k-1}\right)\right)^2 = \arg\min_{\theta \in \Theta} \sum_{k=1}^n \frac{\left(X_k - \boldsymbol{\Psi}(\theta) \cdot \mathbf{X}_{k-1}\right)^2}{\boldsymbol{a} \cdot \mathbf{X}_{k-1}}.$$

We easily derive the following explicit form

$$\widehat{\theta}_{|\mathbf{X}_0|}^X = \frac{\sum_{k=1}^n \left(X_k - \boldsymbol{b} \cdot \mathbf{X}_{k-1}\right)}{\sum_{k=1}^n \boldsymbol{a} \cdot \mathbf{X}_{k-1}}. \tag{3.10}$$

The normalization of the process X_k by $\sqrt{\boldsymbol{a} \cdot \mathbf{X}_{k-1}}$ appears to be the most natural and suitable for the following reasons. First, this normalization generalizes the normalization $X_k / \sqrt{a X_{k-1}}$ in the monotype case, which is the one leading to the Harris estimator of $m_0 = a\theta_0 + b$. We have indeed, for $d = 1$, $a\widehat{\theta}_{X_0}^X + b = \widehat{m}_n^{MLE}$. As mentioned in (3.6), it also corresponds, in the linear case, to the MLE of θ_0. In addition, defining for any vector $\mathbf{u}$, $\underline{\mathbf{u}} := \min_i u_i$ and $\overline{\mathbf{u}} := \max_i u_i$, we have

$$\theta_0 + \frac{\underline{\boldsymbol{b}}}{\overline{\boldsymbol{a}}} \leqslant \mathbb{E}_{\theta_0}\left(\left(Y_k - \mathbb{E}_{\theta_0}\left(Y_k | \mathbf{X}_{k-1}\right)\right)^2 | \mathbf{X}_{k-1}\right) = \theta_0 + \frac{\boldsymbol{b} \cdot \mathbf{X}_{k-1}}{\boldsymbol{a} \cdot \mathbf{X}_{k-1}} \leqslant \theta_0 + \frac{\overline{\boldsymbol{b}}}{\underline{\boldsymbol{a}}}, \tag{3.11}$$

hence the conditional variance of the error term $Y_k - \mathbb{E}_{\theta_0}\left(Y_k | \mathbf{X}_{k-1}\right)$ in the stochastic regression equation $Y_k = \mathbb{E}_{\theta_0}\left(Y_k | \mathbf{X}_{k-1}\right) + Y_k - \mathbb{E}_{\theta_0}\left(Y_k | \mathbf{X}_{k-1}\right)$ is invariant under multiplication of the whole process, and bounded respectively to $(\mathbf{X}_k)_{k \geqslant 0}$.

We provide asymptotical results for the estimator $\widehat{\theta}_{|\mathbf{X}_0|}^X$ defined by (3.10), as the initial population size tends to infinity. We introduce the following notation. For every $i, j = 1 \dots d$ and $k \geqslant 1$, $m_{ij}^{(k)}(\theta)$ denotes the (i,j)-th entry in the k-th power of the matrix $\mathbf{M}(\theta)$ given by (2.4). We define

$$\sigma^2(\theta) := \theta + \frac{\sum_{k=1}^n \sum_{j=1}^d \sum_{i=1}^d \alpha_j b_i m_{ji}^{(k-1)}(\theta)}{\sum_{k=1}^n \sum_{j=1}^d \sum_{i=1}^d \alpha_j a_i m_{ji}^{(k-1)}(\theta)}. \tag{3.12}$$

We can now express the main result of this section.

Theorem 3.1. *Let us assume that there exist some $\alpha_i \in [0,1]$, $i = 1 \dots d$, such that, for all $i = 1 \dots d$,*

$$\lim_{|\boldsymbol{X}_0| \to \infty} \frac{X_{0,i}}{|\boldsymbol{X}_0|} \overset{a.s.}{=} \alpha_i. \tag{3.13}$$

Then $\widehat{\theta}_{|\boldsymbol{X}_0|}^X$ is strongly consistent:

$$\lim_{|\boldsymbol{X}_0| \to \infty} \widehat{\theta}_{|\boldsymbol{X}_0|}^X \overset{a.s.}{=} \theta_0, \tag{3.14}$$

and is asymptotically normally distributed:

$$\lim_{|\boldsymbol{X}_0| \to \infty} \sqrt{\frac{\sum_{k=1}^n \boldsymbol{a} \cdot \boldsymbol{X}_{k-1}}{\sigma^2(\widehat{\theta}_{|\boldsymbol{X}_0|}^X)}} \left(\widehat{\theta}_{|\boldsymbol{X}_0|}^X - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0,1). \tag{3.15}$$

In order to prove Theorem 3.1, we first show the following lemma, which takes advantage of the branching property of the process $(\mathbf{X}_k)_{k \geqslant 0}$, and use the strong law of large numbers.

Lemma 3.2. *Assuming* (3.13), *the following holds for all* $k = 1 \dots n$ *and all* $i = 1 \dots d$,

$$\lim_{|X_0| \to \infty} \frac{X_{k,i}}{|\boldsymbol{X}_0|} \overset{a.s.}{=} \sum_{j=1}^{d} \alpha_j m_{ji}^{(k)}(\theta_0). \tag{3.16}$$

Proof of Lemma 3.2. Using the branching property of the process $(\mathbf{X}_k)_{k \geqslant 0}$, we write

$$X_{k,i} = \sum_{j=1}^{X_{0,1}} X_{k,i,j}^{(1)} + \dots + \sum_{j=1}^{X_{0,d}} X_{k,i,j}^{(d)},$$

where for all $l = 1 \dots d$ and $j = 1 \dots X_{0,l}$, $X_{k,i,j}^{(l)}$ is the i-th coordinate of a d-type branching process at time k initiated by a single particle of type l. For k, i and l fixed the random variables $\left\{X_{k,i,j}^{(l)}\right\}_j$ are i.i.d. with mean value $m_{li}^{(k)}(\theta_0)$. According to the strong law of large numbers and under (3.13), we have, for every $l = 1 \dots d$ such that $X_{0,l} \neq 0$,

$$\lim_{|\mathbf{x}_0| \to \infty} \frac{\sum_{j=1}^{X_{0,l}} X_{k,i,j}^{(l)}}{X_{0,l}} \overset{a.s.}{=} m_{li}^{(k)}(\theta_0),$$

which together with (3.13) leads to (3.16). □

Proof of Theorem 3.1. To prove the consistency of $\widehat{\theta}_{|\mathbf{X}_0|}^X$ we apply Lemma 3.2 to (3.10), using the fact that $X_k = X_{k,1}$ and $X_{k-i} = X_{k-1,i}$, and obtain

$$\lim_{|\mathbf{X}_0| \to \infty} \widehat{\theta}_{|\mathbf{X}_0|}^X \overset{a.s.}{=} \frac{\sum_{k=1}^{n} \sum_{j=1}^{d} \alpha_j \left(m_{j1}^{(k)}(\theta_0) - \sum_{i=1}^{d} b_i m_{ji}^{(k-1)}(\theta_0) \right)}{\sum_{k=1}^{n} \sum_{i=1}^{d} \sum_{j=1}^{d} a_i \alpha_j m_{ji}^{(k-1)}(\theta_0)}. \tag{3.17}$$

By definition,

$$m_{j1}^{(k)}(\theta_0) = \sum_{i=1}^{d} m_{ji}^{(k-1)}(\theta_0) m_{i1}(\theta_0) = \sum_{i=1}^{d} m_{ji}^{(k-1)}(\theta_0) \left(a_i \theta_0 + b_i\right),$$

hence (3.17) immediately leads to (3.14).

We are now interested in the asymptotic distribution of $\widehat{\theta}_{|\mathbf{X}_0|}^X - \theta_0$. We derive from (3.10) that

$$\sqrt{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}} \left(\widehat{\theta}_{|\mathbf{X}_0|}^X - \theta_0 \right) = \frac{\sum_{k=1}^{n} \left(X_k - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} \right)}{\sqrt{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}}}. \tag{3.18}$$

By (2.1),

$$X_k - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} = \sum_{i=1}^{d} \sum_{j=1}^{X_{k-i}} \left(\zeta_{k-i,k,j} - \Psi_i\left(\theta_0\right) \right) =: \sum_{i=1}^{d} \sum_{j=1}^{X_{k-i}} \mathring{\zeta}_{k-i,k,j}, \tag{3.19}$$

where the $\left\{\zeta_{k-i,k,j}\right\}_j$ are i.i.d. given $\mathcal{F}_{k-1}$, following a Poisson distribution with parameter $\Psi_i(\theta_0)$, and the $\left\{\zeta_{k-i,k,j}\right\}_{i,j}$ are independent given $\mathcal{F}_{k-1}$. Renumbering the $\mathring{\zeta}_{k-i,k,j}$ we then obtain

$$\sum_{k=1}^{n} \left(X_k - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} \right) = \sum_{i=1}^{d} \sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \mathring{\zeta}_{k-i,k,j}. \tag{3.20}$$

Applying a central limit theorem for the sum of a random number of independent random variables (see *e.g.* [4]), we obtain that for all $i = 1 \dots d$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{\sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \mathring{\zeta}_{k-i,k,j}}{\sqrt{\sum_{k=1}^{n} X_{k-i}}} \overset{\mathcal{D}}{=} \mathcal{N}\left(0, a_i \theta_0 + b_i\right). \tag{3.21}$$

We have used the fact that $|\mathbf{X}_0|$ is a real positive sequence growing to infinity, and $\sum_{k=1}^{n} X_{k-i}$ a sequence of integered-valued random variables such that $\sum_{k=1}^{n} X_{k-i} / |\mathbf{X}_0|$ converges in probability to a finite random variable. In our case the limit is actually deterministic, since we have shown in Lemma 3.2 that

$$\lim_{|\mathbf{X}_0|\to\infty} \frac{\sum_{k=1}^n X_{k-i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \sum_{k=1}^n \sum_{j=1}^d \alpha_j m_{ji}^{(k-1)}(\theta_0).$$

Using (3.20) in (3.18), we write

$$\sqrt{\sum_{k=1}^n \boldsymbol{a}\cdot\mathbf{X}_{k-1}}\left(\widehat{\theta}^X_{|\mathbf{X}_0|}-\theta_0\right) = \sum_{i=1}^d \frac{\sum_{j=1}^{\sum_{k=1}^n X_{k-i}} \mathring{\zeta}_{k-i,k,j}}{\sqrt{\sum_{k=1}^n X_{k-i}}} \frac{\sqrt{\sum_{k=1}^n X_{k-i}}}{\sqrt{\sum_{k=1}^n \boldsymbol{a}\cdot\mathbf{X}_{k-1}}}. \tag{3.22}$$

Using again Lemma 3.2,

$$\lim_{|\mathbf{X}_0|\to\infty} \frac{\sqrt{\sum_{k=1}^n X_{k-i}}}{\sqrt{\sum_{k=1}^n \boldsymbol{a}\cdot\mathbf{X}_{k-1}}} \overset{a.s.}{=} \sqrt{\frac{\sum_{k=1}^n \sum_{j=1}^d \alpha_j m_{ji}^{(k-1)}(\theta_0)}{\sum_{k=1}^n \sum_{j=1}^d \sum_{l=1}^d \alpha_j a_l m_{jl}^{(k-1)}(\theta_0)}},$$

which, combined to (3.21) and (3.22), implies by Slutsky's theorem that

$$\lim_{|\mathbf{X}_0|\to\infty} \sqrt{\sum_{k=1}^n \boldsymbol{a}\cdot\mathbf{X}_{k-1}}\left(\widehat{\theta}^X_{|\mathbf{X}_0|}-\theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0,\sigma^2\left(\theta_0\right)\right). \tag{3.23}$$

By (3.12) and (3.14), $\lim_{|\mathbf{X}_0|}\sqrt{\sigma^2(\theta_0)}/\sqrt{\sigma^2(\widehat{\theta}^X_{|\mathbf{X}_0|})} \overset{a.s.}{=} 1$, from which we deduce (3.15). □

Remark 3.3. We point out that we do not use the Poissonian character of the transitions of the process (2.1) to derive the properties of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, but we simply need its first and second order moments. This estimator can thus be applied to any process of the form (2.1), where the $\{\zeta_{n-k,n,i}\}_i$ do not necessarily follow a Poisson distribution, but satisfy $\mathbb{E}_{\theta_0}(\zeta_{n-k,n,i}|\mathcal{F}_{n-1}) = \Psi_k(\theta_0)$. 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.

3.2 A CLSE with asymptotic properties, as $n\to\infty$

In this section we consider, instead of $(\mathbf{X}_k)_{k\geqslant 0}$, the process conditioned on non-extinction $\mathbf{Z}_k := (\mathbf{X}_k|\mathbf{X}_k\neq\mathbf{0})$, and define for all $k\geqslant 0$ the one-dimensional process corresponding to the first coordinate $Z_k := Z_{k,1}$. We obtain asymptotic properties for the corresponding CLSE, as the number of observations n tends to infinity, even in the subcritical case, despite the almost sure extinction of the process. It is indeed known (see *e.g.* Theorem 5.4.2 in [2]) that in the subcritical case the conditioned process $(\mathbf{Z}_k)_{k\geqslant 0}$ admits a stationary measure ν_{θ_0}, referred to as the Yaglom distribution. We point out that this estimator is particularly adapted for the study of the extinction phase of an epidemic, even if the number of cases at the beginning of the extinction is not very large.

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

$$\mathbb{P}_{\theta_0}\left(Z_k=j|\mathbf{Z}_{k-1}\right) = \frac{\left(\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}\right)^j e^{-\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}}}{j!\left(1-\mathbf{1}_{\{\lceil\mathbf{Z}_{k-1}\rceil=0\}}e^{-\Psi_d(\theta_0)Z_{k-d}}\right)}. \tag{3.24}$$

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

$$\widehat{\theta}^Z_n := \arg\min_{\theta\in\Theta} S_n(\theta), \quad S_n(\theta) := \sum_{k=1}^n \left(\frac{Z_k}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_{k-1}}} - f(\theta,\mathbf{Z}_{k-1})\right)^2, \tag{3.25}$$

where

$$f(\theta_0,\mathbf{Z}_{k-1}) := \mathbb{E}_{\theta_0}\left(\frac{Z_k}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_{k-1}}}\middle|\mathbf{Z}_{k-1}\right) = \frac{\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_{k-1}}\left(1-\mathbf{1}_{\{\lceil\mathbf{Z}_{k-1}\rceil=0\}}e^{-\Psi_d(\theta_0)Z_{k-d}}\right)}. \tag{3.26}$$

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

$$f'(\theta, \mathbf{j}) = \begin{cases} \sqrt{a_d j_d} \frac{1-(1+\Psi_d(\theta) j_d) e^{-\Psi_d(\theta) j_d}}{\left(1-e^{-\Psi_d(\theta) j_d}\right)^2} & \text{if } \lceil \mathbf{j} \rceil = 0, \\ \sqrt{\boldsymbol{a} \cdot \mathbf{j}} & \text{otherwise.} \end{cases} \tag{3.27}$$

We finally define

$$\varepsilon_k := \frac{Z_k}{\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}} - f(\theta_0, \mathbf{Z}_{k-1}), \tag{3.28}$$

which implies that

$$\mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) = \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}}{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}\left(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{k-d}}\right)}, \tag{3.29}$$

and the conditional variance of the error term ε_k in the stochastic regression equation is consequently bounded:

$$\theta_0 + \frac{\underline{\boldsymbol{b}}}{\overline{\boldsymbol{a}}} \leqslant \mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) \leqslant \left(1 - e^{-\Psi_d(\theta_0)}\right)^{-1}\left(\theta_0 + \frac{\overline{\boldsymbol{b}}}{\underline{\boldsymbol{a}}}\right). \tag{3.30}$$

Theorem 3.4. *The estimator $\widehat{\theta}_n^Z$ is strongly consistent:*

$$\lim_{n \to \infty} \widehat{\theta}_n^Z \stackrel{a.s.}{=} \theta_0. \tag{3.31}$$

Proof. According to Proposition 3.1 in [11], sufficient conditions for the strong consistency of $\widehat{\theta}_n^Z$ are that $f(., \mathbf{Z}_{k-1})$ is Lipschitz, in the sense that there exists a nonnegative $\sigma(\mathbf{Z}_0, \ldots, \mathbf{Z}_{k-1})$-measurable function C_k satisfying for all θ_1, $\theta_2 \in \Theta$, $|f(\theta_1, \mathbf{Z}_{k-1}) - f(\theta_2, \mathbf{Z}_{k-1})| \stackrel{a.s.}{\leqslant} C_k\, |\theta_1 - \theta_2|$, that $\overline{\lim}_{k \to \infty} \mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) \stackrel{a.s.}{<} \infty$, and that

$$\lim_{n \to \infty} \inf_{|\theta - \theta_0| \geqslant \delta} \sum_{k=1}^n \left(f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1})\right)^2 \stackrel{a.s.}{=} \infty. \tag{3.32}$$

The Lipschitz condition is satisfied thanks to (3.27), which shows that $f'(., \mathbf{Z}_{k-1})$ is bounded on Θ. The second condition follows from (3.30). Let $\delta > 0$ and $\theta \in \Theta$ such that $|\theta - \theta_0| \geqslant \delta$. We assume for convenience that $\theta_0 > \theta$. In order to prove (3.32), we apply the mean value theorem to the function $f(., \mathbf{Z}_{k-1})$, and obtain that there exists some $\tilde{\theta}_k \in]\theta, \theta_0[$ such that $f'(\tilde{\theta}_k, \mathbf{Z}_{k-1}) = \left(f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1})\right)(\theta_0 - \theta)^{-1}$. Consequently,

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

which implies (3.32). □

To prove the asymptotic distribution of $\widehat{\theta}_n^Z$, we make several times use of the following strong law of large numbers for homogeneous positive recurrent Markov chains: for every ν_{θ_0}-integrable function $g : \mathbb{N}^d \setminus \{\mathbf{0}\} \to \mathbb{R}$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\mathbf{Z}_k) \stackrel{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} g(\mathbf{j}) \nu_{\theta_0}(\mathbf{j}). \tag{3.33}$$

By (2.5) it is ensured (see *e.g.* [2]) that the Yaglom measure ν_{θ_0} satisfies, for all $i = 1 \dots d$,

$$\sum_{\mathbf{k}\in\mathbb{N}^d} k_i \nu_{\theta_0}(\mathbf{k}) < \infty. \tag{3.34}$$

For the following we need to show that ν_{θ_0} admits finite second-order moments as well.

Proposition 3.5. *For all* $i, j = 1 \dots d$,

$$\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_i k_j \nu_{\theta_0}(\boldsymbol{k}) < \infty. \tag{3.35}$$

Proof. First, let us note that by definition of $\mathbf{Z}_n$, we have for all $i = 1 \dots d$,

$$\sum_{\mathbf{k}\in\mathbb{N}^d} k_i \nu_{\theta_0}(\mathbf{k}) = \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} Z_{n,i}\right) = \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} Z_{n-i+1,1}\right) = \sum_{\mathbf{k}\in\mathbb{N}^d} k_1 \nu_{\theta_0}(\mathbf{k}) =: m^{\nu_{\theta_0}}. \tag{3.36}$$

For all $i = 1 \dots d-1$, using the inequality $x\left(1-e^{-x}\right)^{-1} \leqslant 1 + x$, $x \geqslant 0$,

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

hence

$$\underline{\lim}\,\mathbb{E}_{\theta_0}\left(Z_n Z_{n-i}\right) \leqslant m^{\nu_{\theta_0}} + \max_{k=0\dots d-1}\underline{\lim}\,\mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right)\sum_{j=1}^{d}\Psi_j(\theta_0). \tag{3.37}$$

Similarly,

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

which by Fatou's lemma and (3.36) leads to (using the fact that $\sum_{j=1}^{d}\Psi_j(\theta_0) < 1$)

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

Together with (3.37) this implies that

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

and thus

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

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

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

□

We can now prove the following theorem.

Theorem 3.6. *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is subcritical. Then the estimator $\widehat{\theta}_n^Z$ is asymptotically normally distributed:*

$$\lim_{n\to\infty}\sqrt{n\frac{\left(\sum_{\boldsymbol{j}\in\mathbb{N}^d}\left(f'(\theta_0,\boldsymbol{j})\right)^2\nu_{\theta_0}(\boldsymbol{j})\right)^2}{\sum_{\boldsymbol{j}\in\mathbb{N}^d}\left(f'(\theta_0,\boldsymbol{j})\right)^2 f(\theta_0,\boldsymbol{j})\left(\boldsymbol{a}\cdot\boldsymbol{j}\right)^{-1/2}\nu_{\theta_0}(\boldsymbol{j})}}\left(\widehat{\theta}_n^Z-\theta_0\right)\overset{\mathcal{D}}{=}\mathcal{N}\left(0,1\right), \tag{3.38}$$

where f is given by (3.26).

Proof. We follow the steps of the proof of Proposition 6.1 in [11]. Writing the Taylor expansion of $S_n'(\theta)$ in the neighborhood of θ_0, we obtain that

$$\widehat{\theta}_n^Z-\theta_0=-\frac{S_n'(\theta_0)}{S_n''(\tilde{\theta}_n)}, \tag{3.39}$$

for some $\tilde{\theta}_n=\theta_0+t_n\left(\widehat{\theta}_n^Z-\theta_0\right)$, with $t_n\in\,]0,1[$. Since $S_n'(\theta_0)=-2\sum_{k=1}^n\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})$, we can write

$$\sqrt{n}\left(\widehat{\theta}_n^Z-\theta_0\right)=\frac{\sum_{k=1}^n\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\left(\frac{F_n}{n}\right)^{-1}\left(\frac{1}{2}\frac{S_n''(\tilde{\theta}_n)}{F_n}\right)^{-1}, \tag{3.40}$$

where $F_n:=\sum_{k=1}^n\left(f'(\theta_0,\mathbf{Z}_{k-1})\right)^2$. By (3.27), for all $\mathbf{j}\in\mathbb{N}^d$, $\mathbf{j}\neq\mathbf{0}$, $0\leqslant f'(\theta_0,\mathbf{j})\leqslant\sqrt{\boldsymbol{a}\cdot\mathbf{j}}(1-e^{-\Psi_d(\theta_0)})^{-2}$, hence we deduce by means of (3.33) and (3.34) that

$$\lim_{n\to\infty}\frac{F_n}{n}\overset{a.s.}{=}\sum_{\mathbf{j}\in\mathbb{N}^d}\left(f'(\theta_0,\mathbf{j})\right)^2\nu_{\theta_0}(\mathbf{j}). \tag{3.41}$$

In view of (3.40), we now prove that

$$\lim_{n\to\infty}\frac{S_n''(\tilde{\theta}_n)}{F_n}\overset{a.s.}{=}2. \tag{3.42}$$

Computing S_n'' thanks to the formula $S_n(\theta)=\sum_{k=1}^n\left(\varepsilon_k+f(\theta_0,\mathbf{Z}_{k-1})-f(\theta,\mathbf{Z}_{k-1})\right)^2$, it appears that (3.42) is true, as soon as the following holds:

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\frac{\left|\sum_{k=1}^n\varepsilon_k f''(\theta,\mathbf{Z}_{k-1})\right|}{F_n}\overset{a.s.}{=}0, \tag{3.43}$$

$$\lim_{n\to\infty}\frac{\sum_{k=1}^n\left(f'(\tilde{\theta}_n,\mathbf{Z}_{k-1})\right)^2}{F_n}\overset{a.s.}{=}1, \tag{3.44}$$

and

$$\lim_{n\to\infty}\frac{\sum_{k=1}^n\left(f(\theta_0,\mathbf{Z}_{k-1})-f(\tilde{\theta}_n,\mathbf{Z}_{k-1})\right)f''(\tilde{\theta}_n,\mathbf{Z}_{k-1})}{F_n}\overset{a.s.}{=}0. \tag{3.45}$$

Let us prove (3.43)-(3.45). Note that, for every $\mathbf{j}\neq\mathbf{0}$, $f''(\theta,\mathbf{j})=0$ if $\lceil\mathbf{j}\rceil\neq 0$, and

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

otherwise. First, (3.43) is given by a strong law of large numbers proved in [11], Proposition 5.1. The latter can be indeed applied since $f''(.,\mathbf{Z}_{k-1})$ fulfills the required Lipschitz condition, and $\lim_n F_n\overset{a.s.}{=}\infty$ (as an immediate consequence of the stronger result (3.41)). In view of (3.44) we consider the function $\left(f'(\theta,\mathbf{j})\right)^2$ and its derivative $2f'(\theta,\mathbf{j})f''(\theta,\mathbf{j})$. For all $\theta\in\overline{\Theta}$ and all $\mathbf{j}\neq\mathbf{0}$ with $\lceil\mathbf{j}\rceil=0$,

$$\begin{aligned}\left|2f'(\theta,\mathbf{j})f''(\theta,\mathbf{j})\right|&\leqslant 4\frac{(a_dj_d)^2e^{-\Psi_d(\theta)j_d}\left(\Psi_d\left(\theta\right)j_d+2\right)}{\left(1-e^{-\Psi_d(\theta)j_d}\right)^5}\\&\leqslant\frac{4\max_{x\geqslant 0}\left(x+2\right)^3e^{-x}}{\left(1-e^{-\Psi_d(\theta_{min})}\right)^5}=:c_1.\end{aligned} \tag{3.46}$$

Consequently,

$$\frac{\left|\sum_{k=1}^n f'(\tilde{\theta}_n,\mathbf{Z}_{k-1})^2 - f'(\theta_0,\mathbf{Z}_{k-1})^2\right|}{F_n} \leqslant c_1 \left|\widehat{\theta}_n^Z - \theta_0\right| \left(\frac{F_n}{n}\right)^{-1}, \tag{3.47}$$

which by (3.41) and the strong consistency of $\widehat{\theta}_n^Z$ 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 (3.44). It now remains to prove (3.45). With similar computations as above, one shows that there exists a deterministic constant $c_2 > 0$ such that

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

which thanks to (3.41) and the strong consistency of $\widehat{\theta}_n^Z$ implies (3.45).

In view of (3.40), we finally want to prove that $\sum_{k=1}^n \varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})/\sqrt{n}$ converges in distribution, and for this purpose we make use of the following central limit theorem for sequences of martingales (see *e.g.* [15] or [14]).

Proposition 3.7. *Let $\{M_k^{(n)}, \mathcal{F}_k^{(n)}, 1 \leqslant k \leqslant n\}$, $n \geqslant 1$ be a sequence of square integrable martingales. For each $n \geqslant 1$, we denote by $\langle M\rangle^{(n)} = (\langle M\rangle_k^{(n)})_{1\leqslant k\leqslant n}$ the associated Meyer process. We assume that there exists a constant c such that $\lim_{n\to\infty} \langle M_n\rangle^{(n)} \stackrel{\mathcal{P}}{=} c^2$, and assume moreover that for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[\left|M_k^{(n)} - M_{k-1}^{(n)}\right|^2 \mathbf{1}_{\left\{\left|M_k^{(n)} - M_{k-1}^{(n)}\right| \geqslant \varepsilon\right\}} \middle| \mathcal{F}_{k-1}^{(n)}\right] \stackrel{\mathcal{P}}{=} 0.$$

Then $\lim_{n\to\infty} M_n^{(n)} \stackrel{\mathcal{D}}{=} \mathcal{N}(0,c^2)$.

Let us define, for every $k \leqslant n$, $M_k^{(n)} := \sum_{l=1}^k \varepsilon_l f'(\theta_0,\mathbf{Z}_{l-1})/\sqrt{n}$. First, for every $k \leqslant n$, $\mathbb{E}_{\theta_0}\left(\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})/\sqrt{n} | \mathbf{Z}_{k-1}\right) = 0$. Second,

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

and $M_k^{(n)}$ is a sequence of square integrable martingales. Moreover, by (3.34),

$$\sum_{\mathbf{j}\in\mathbb{N}^d} \frac{\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}} \nu_{\theta_0}(\mathbf{j}) \leqslant \frac{1}{\left(1 - e^{-\Psi_d(\theta_0)}\right)^5} \sum_{\mathbf{j}\in\mathbb{N}^d} \mathbf{\Psi}(\theta_0)\cdot\mathbf{j}\nu_{\theta_0}(\mathbf{j}) < \infty, \tag{3.48}$$

so by means of (3.33),

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

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

$$\begin{aligned}
&\sum_{k=1}^n \mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right|^2 \mathbf{1}_{\left\{\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right| \geqslant \varepsilon\right\}} \middle| \mathbf{Z}_{k-1}\right] \\
&\leqslant \sum_{k=1}^n \left(\mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right|^4 \middle| \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \left(\mathbb{P}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right| \geqslant \varepsilon \middle| \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \\
&\leqslant \frac{1}{n^{\frac{3}{2}}\varepsilon} \sum_{k=1}^n |f'(\theta_0,\mathbf{Z}_{k-1})|^3 \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^4 | \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^2 | \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{3.49}$$

We have

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

and

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

hence

$$\begin{aligned}|f'(\theta_0,\mathbf{Z}_{k-1})|^3 &\left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^4|\mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}}\left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^2|\mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \\ &\leqslant \frac{\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}\left(1+3\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}\right)^{\frac{1}{2}}}{\left(1-e^{-\Psi_d(\theta_0)}\right)^7} \\ &\leqslant \frac{\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}+\sqrt{3}\left(\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{k-1}\right)^{\frac{3}{2}}}{\left(1-e^{-\Psi_d(\theta_0)}\right)^7}. \end{aligned} \tag{3.50}$$

Since the Yaglom distribution ν_{θ_0} has finite second-order moments (see Proposition 3.5), we deduce from (3.49) and (3.50) by virtue of (3.33) that

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right|^2 \mathbf{1}_{\left\{\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right|\geqslant\varepsilon\right\}}\middle|\mathbf{Z}_{k-1}\right] \stackrel{a.s.}{=} 0.$$

It then ensues from Proposition 3.7 that

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n}\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}} \stackrel{\mathcal{D}}{=} \mathcal{N}\left(0,\sum_{\mathbf{j}\in\mathbb{N}^d}\frac{\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}}\nu_{\theta_0}(\mathbf{j})\right). \tag{3.51}$$

Finally, (3.40) together with (3.41), (3.42), (3.51) and Slutsky's theorem imply that

$$\lim_{n\to\infty}\sqrt{n}\left(\widehat{\theta}_n^Z-\theta_0\right) \stackrel{\mathcal{D}}{=} \mathcal{N}\left(0,\frac{\sum_{\mathbf{j}\in\mathbb{N}^d}\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)\left(\boldsymbol{a}\cdot\mathbf{j}\right)^{-1/2}\nu_{\theta_0}(\mathbf{j})}{\left(\sum_{\mathbf{j}\in\mathbb{N}^d}\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2\nu_{\theta_0}(\mathbf{j})\right)^2}\right). \tag{3.52}$$

□

Since the Yaglom distribution is in general not explicitly known, Theorem 3.6 is not directly applicable. We can however deduce the following more practical result:

Corollary 3.8. *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is subcritical. Then the estimator $\widehat{\theta}_n^Z$ has the following asymptotic distribution*

$$\lim_{n\to\infty}\frac{\sum_{k=0}^{n}\left(f'\left(\widehat{\theta}_n^Z,\boldsymbol{Z}_k\right)\right)^2}{\sqrt{\sum_{k=0}^{n}\left(f'\left(\widehat{\theta}_n^Z,\boldsymbol{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\boldsymbol{Z}_k\right)\left(\boldsymbol{a}\cdot\boldsymbol{Z}_k\right)^{-1/2}}}\left(\widehat{\theta}_n^Z-\theta_0\right) \stackrel{\mathcal{D}}{=} \mathcal{N}\left(0,1\right). \tag{3.53}$$

Proof. The result is immediate as soon as we prove that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}} \stackrel{a.s.}{=} \sum_{\mathbf{j}\in\mathbb{N}^d}\frac{\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}}\nu_{\theta_0}\left(\mathbf{j}\right), \tag{3.54}$$

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

$$\begin{aligned}\sum_{k=0}^{n}\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}} &= \sum_{k=0}^{n}\frac{\left(f'\left(\theta_0,\mathbf{Z}_k\right)\right)^2 f\left(\theta_0,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}} \\ &+\sum_{k=0}^{n}\left[\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}-\frac{\left(f'\left(\theta_0,\mathbf{Z}_k\right)\right)^2 f\left(\theta_0,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}\right],\end{aligned} \tag{3.55}$$

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

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

which enables to write

$$\frac{1}{n+1}\sum_{k=0}^{n}\left|\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}-\frac{\left(f'\left(\theta_0,\mathbf{Z}_k\right)\right)^2 f\left(\theta_0,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}\right|$$
$$\leqslant\left|\widehat{\theta}_n^Z-\theta_0\right|\frac{2c_1}{\left(1-e^{-\Psi_d(\theta_{min})}\right)^3}\frac{1}{n+1}\sum_{k=0}^{n}\boldsymbol{\Psi}(\theta_{max})\cdot\mathbf{Z}_k. \quad (3.56)$$

By the strong consistency of $\widehat{\theta}_n^Z$ together with (3.33) and (3.34), (3.56) almost surely tends to zero. Combined with (3.33) and (3.48) in (3.55), this implies (3.54).

□

Remark 3.9. One can show that, for a given n, the estimator $\widehat{\theta}_n^Z$ has the same asymptotic distribution, when $|\mathbf{X}_0|$ grows to infinity, as the estimator $\widehat{\theta}_{|\mathbf{X}_0|}^X$. Indeed, if the two estimators differ, this implies that $(\mathbf{X}_k)_{k\leqslant n}\neq(\mathbf{Z}_k)_{k\leqslant n}$. For notational convenience will simply denote this last event by $\{\mathbf{X}\neq\mathbf{Z}\}$. The probability that the estimators are not equal thus satisfies (we omit here the subscript θ_0)

$$\mathbb{P}\left(\widehat{\theta}_{|\mathbf{X}_0|}^X\neq\widehat{\theta}_n^Z\,|\mathbf{X}_0\right)\leqslant\mathbb{P}\left(\mathbf{X}\neq\mathbf{Z}\,|\mathbf{X}_0\right)=\mathbb{P}\left(\exists k\leqslant n:\ X_k=0,\ldots,X_{k-(d-2)}=0\,|\mathbf{X}_0\right)$$
$$\leqslant\mathbb{P}\left(\exists k\leqslant n:\ X_k=0\,|\mathbf{X}_0\right)\leqslant\sum_{k=1}^{n}\mathbb{P}\left(X_k=0\,|\mathbf{X}_0\right).$$

By the branching property of the process $(\mathbf{X}_k)_{k\geqslant 0}$, for all $k\in\mathbb{N}$,

$$\mathbb{P}\left(X_k=0|\mathbf{X}_0\right)=\prod_{i=1}^{d}\mathbb{P}\left(X_k=0|\mathbf{X}_0=\mathbf{e_i}\right)^{X_{0,i}}.$$

Consequently,

$$\lim_{|\mathbf{X}_0|\to\infty}\mathbb{P}\left(\widehat{\theta}_{|\mathbf{X}_0|}^X\neq\widehat{\theta}_n^Z|\mathbf{X}_0\right)=\lim_{|\mathbf{X}_0|\to\infty}\mathbb{P}\left(\mathbf{X}\neq\mathbf{Z}\right)=0.$$

This implies on the one hand the strong consistency of $\widehat{\theta}_n^Z$ as $|\mathbf{X}_0|\to\infty$. On the other hand, for all $u\in\mathbb{R}$,

$$\mathbb{P}\left(\sqrt{\frac{\sum_{k=1}^n\boldsymbol{a}\cdot\mathbf{Z}_{k-1}}{\sigma^2(\widehat{\theta}_n^Z)}}\left(\widehat{\theta}_n^Z-\theta_0\right)\leqslant u\,|\mathbf{X}_0\right)$$
$$=\mathbb{P}\left(\sqrt{\frac{\sum_{k=1}^n\boldsymbol{a}\cdot\mathbf{X}_{k-1}}{\sigma^2(\widehat{\theta}_{|\mathbf{X}_0|}^X)}}\left(\widehat{\theta}_{|\mathbf{X}_0|}^X-\theta_0\right)\leqslant u\,|\mathbf{X}=\mathbf{Z},\,\mathbf{X}_0\right)\mathbb{P}\left(\mathbf{X}=\mathbf{Z}\,|\mathbf{X}_0\right)$$
$$+\mathbb{P}\left(\sqrt{\frac{\sum_{k=1}^n\boldsymbol{a}\cdot\mathbf{Z}_{k-1}}{\sigma^2(\widehat{\theta}_n^Z)}}\left(\widehat{\theta}_n^Z-\theta_0\right)\leqslant u\,|\mathbf{X}\neq\mathbf{Z},\,\mathbf{X}_0\right)\mathbb{P}\left(\mathbf{X}\neq\mathbf{Z}|\,\mathbf{X}_0\right),$$

hence, denoting by Φ the cumulative distribution function of the Gaussian distribution, we deduce from Theorem 3.1 that

$$\lim_{|\mathbf{X}_0|\to\infty}\mathbb{P}\left(\sqrt{\frac{\sum_{k=1}^n\boldsymbol{a}\cdot\mathbf{Z}_{k-1}}{\sigma^2(\widehat{\theta}_n^Z)}}\left(\widehat{\theta}_n^Z-\theta_0\right)\leqslant u\,|\mathbf{X}_0\right)=\Phi(u).$$

3.3 An explicit estimator with asymptotic properties, as $n \to \infty$

The aim of this section is to provide an estimator with asymptotic properties, as time 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 the estimator introduced in [1],

$$\widetilde{\rho}_n := \frac{|\mathbf{X}_1| + \ldots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \ldots + |\mathbf{X}_{n-1}|}, \tag{3.57}$$

using relation (3.5), the following explicit estimator of θ_0,

$$\widetilde{\theta}_n^X := \frac{1 - \sum_{k=1}^d b_k \widetilde{\rho}_n^{-k}}{\sum_{k=1}^d a_k \widetilde{\rho}_n^{-k}}. \tag{3.58}$$

All what follows can be applied to any process of the form (2.1), where the $\{\zeta_{n-k,n,i}\}_i$ do not necessarily follow a Poisson distribution, but satisfy $\mathbb{E}_{\theta_0}(\zeta_{n-k,n,i}|\mathcal{F}_{n-1}) = \Psi_k(\theta_0)$.

For each $i = 1 \ldots d$ and $n \in \mathbb{N}$, we define the covariance matrix $\mathbf{V}_n^i$ with entries

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

In particular, $\left[\mathbf{V}_1^i\right]_{jk} = \Psi_i$ if $j = k = i$, and is null otherwise. Let $\boldsymbol{\eta}$ be the left eigenvector of $\mathbf{M}$ for its Perron's root ρ_0, with normalization $\boldsymbol{\xi} \cdot \boldsymbol{\eta} = 1$. Then, for all $i = 1 \ldots d$,

$$\xi_i = \frac{\rho_0^{i-1} \sum_{j=i}^d \Psi_j \rho_0^{-j}}{\sum_{k=1}^d \rho_0^{k-1} \sum_{j=k}^d \Psi_j \rho_0^{-j}}, \quad \eta_i = \rho_0^{-(i-1)} \frac{\sum_{k=1}^d \rho_0^{k-1} \sum_{j=k}^d \Psi_j \rho_0^{-j}}{\sum_{k=1}^d \sum_{j=k}^d \Psi_j \rho_0^{-j}}. \tag{3.59}$$

The basic limit theorem in the supercritical case states that there exists a random variable W such that (see *e.g.* Theorem 5.6.1 in [2])

$$\lim_{n\to\infty} \rho_0^{-n} \mathbf{X}_n \overset{a.s.}{=} \boldsymbol{\eta} W. \tag{3.60}$$

Let us recall the results obtained by S. Asmussen and N. Keiding in [1], Theorem 6.1. First, as pointed out by N. Becker in [3], the estimator $\widetilde{\rho}_n$ is strongly consistent on the set of non-extinction $\{W > 0\}$. Second, once adequately normalized, $\widetilde{\rho}_n - \rho_0$ is asymptotically normal. However, the asymptotic behavior of $\widetilde{\rho}_n - \rho_0$ depends qualitatively on the relative sizes of ρ and λ^2, where λ is the absolute value of a certain eigenvalue of $\mathbf{M}$. More precisely, let $\{\lambda_i\}_{i=1\ldots s}$ be the spectrum of $\mathbf{M}$, and for each $i = 1 \ldots s$, let r_i be the algebraic multiplicity of λ_i. We denote by $\mathcal{B} = \{\mathbf{u}_{i,j},\ i = 1 \ldots s,\ j = 1 \ldots r_i\}$ the base of the Jordan canonical decomposition of $\mathbf{M}$, *i.e.* such that for all $i = 1 \ldots s$,

$$\mathbf{M}\mathbf{u}_{i,1} = \lambda_i \mathbf{u}_{i,1}, \quad \mathbf{M}\mathbf{u}_{i,j} = \mathbf{u}_{i,j-1} + \lambda_i \mathbf{u}_{i,j},\ j = 2 \ldots r_i.$$

Let us define the vector $\boldsymbol{\zeta} := \mathbf{1} - \boldsymbol{\xi}$ and denote $(\zeta_{i,j})_{\substack{i=1\ldots s,\\ j=1\ldots r_i}}$ its coordinates in $\mathcal{B}$: $\boldsymbol{\zeta} = \sum_{i=1}^s \sum_{j=1}^{r_i} \zeta_{i,j} \mathbf{u}_{i,j}$. Then λ is defined as follows,

$$\lambda = \lambda(\boldsymbol{\zeta}) := \max_{i=1\ldots s} \{|\lambda_i| : \exists j = 1 \ldots r_i \text{ such that } \zeta_{i,j} \neq 0\}, \tag{3.61}$$

and

$$\gamma = \gamma(\boldsymbol{\zeta}) := \max_{\substack{i=1\ldots s:\\ |\lambda_i| = \lambda}} \{j = 1 \ldots r_i : \zeta_{i,j} \neq 0\}. \tag{3.62}$$

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

$$S_n := \sum_{k=0}^{n-1} (\mathbf{X}_{k+1} - \mathbf{X}_k \mathbf{M}) \cdot \mathbf{1}, \quad T_n := \sum_{k=0}^{n-1} \mathbf{X}_k \cdot \boldsymbol{\kappa}, \quad \boldsymbol{\kappa} := (\mathbf{M} - \rho_0 \mathbf{I})\, \mathbf{1}. \tag{3.63}$$

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 $\lambda^2 \geqslant \rho_0$. In order to deal with the case $\lambda^2 < \rho_0$, we define for all $n \in \mathbb{N}$, $\boldsymbol{\nu}_n := \mathbf{1} + \sum_{k=0}^{n-1} \mathbf{M}^k \boldsymbol{\kappa}$, and

$$C_1 := (\rho_0 - 1) \sum_{n=1}^\infty \rho_0^{-n} \sum_{i=1}^d \eta_i \boldsymbol{\nu}_n \mathbf{V}_1^i \boldsymbol{\nu}_n = (\rho_0 - 1) \sum_{n=1}^\infty \rho_0^{-n} \sum_{i=1}^d \eta_i \Psi_i \nu_{n,1}^2.$$

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

$$C_2 := \left(1 - \frac{1}{\rho_0}\right) \lim_{n\to\infty} \frac{\sum_{i=1}^{d} \eta_i \boldsymbol{\zeta}^1 \mathbf{V}_n^i \boldsymbol{\zeta}^1}{\rho_0^n n^{2\gamma-1}}. \tag{3.64}$$

We can now quote Theorem 6.1 of [1]. Note that (2.5) implies that $\mathbb{P}(W>0)>0$ (see *e.g.* [2], Theorem 5.6.1). For notational convenience, it is assumed in this section just as in [1] that $\mathbb{P}(W=0)=0$. The results stated here are thus valid on the set of non-extinction.

Theorem 3.10 (S. Asmussen & N. Keiding, [1] Thm 6.1-6.3)**.** *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is supercritical. Then, on the set of non-extinction, the estimator $\widetilde{\rho}_n$ is consistent:*

$$\lim_{n\to\infty} \widetilde{\rho}_n \overset{a.s.}{=} \rho_0, \tag{3.65}$$

and has the following asymptotic distribution.

If $\lambda^2 < \rho_0$,

$$\lim_{n\to\infty} \sqrt{W\left(1+\ldots+\rho_0^{n-1}\right)}\,(\widetilde{\rho}_n - \rho_0) \overset{\mathcal{D}}{=} \mathcal{N}(0, C_1). \tag{3.66}$$

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

$$\lim_{n\to\infty} \sqrt{\frac{W\left(1+\ldots+\rho_0^{n-1}\right)}{n^{2\gamma-1}}}\,(\widetilde{\rho}_n - \rho_0) \overset{\mathcal{D}}{=} \mathcal{N}(0, C_2). \tag{3.67}$$

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

$$\lim_{n\to\infty} \left[\frac{W\left(1+\ldots+\rho_0^{n-1}\right)}{\lambda^{n-1}(n-1)^{\gamma-1}}\,(\widetilde{\rho}_n - \rho_0) - H_{n-1}\right] \overset{a.s.}{=} 0. \tag{3.68}$$

We now define the constant

$$C_0 := \left(\frac{\sum_{k=1}^{d} k a_k \rho_0^{-k} + \sum_{k=1}^{d} a_k \rho_0^{-k} \sum_{k=1}^{d} k b_k \rho_0^{-k} - \sum_{k=1}^{d} b_k \rho_0^{-k} \sum_{k=1}^{d} k a_k \rho_0^{-k}}{\rho_0 \left(\sum_{k=1}^{d} a_k \rho_0^{-k}\right)^2}\right)^2. \tag{3.69}$$

We immediately deduce from Theorem 3.10 the following result.

Theorem 3.11. *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is supercritical. Then, on the set of non-extinction, the estimator $\widetilde{\theta}_n^X$ is consistent:*

$$\lim_{n\to\infty} \widetilde{\theta}_n^X \overset{a.s.}{=} \theta_0, \tag{3.70}$$

and has the following asymptotic distribution.

If $\lambda^2 < \rho_0$,

$$\lim_{n\to\infty} \sqrt{\frac{W\left(1+\ldots+\rho_0^{n-1}\right)}{C_1 C_0}}\left(\widetilde{\theta}_n^X - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0, 1). \tag{3.71}$$

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

$$\lim_{n\to\infty} \sqrt{\frac{W\left(1+\ldots+\rho_0^{n-1}\right)}{n^{2\gamma-1} C_2 C_0}}\left(\widetilde{\theta}_n^X - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0, 1). \tag{3.72}$$

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

$$\lim_{n\to\infty} \left[\frac{W\left(1+\ldots+\rho_0^{n-1}\right)}{\lambda^{n-1}(n-1)^{\gamma-1}}\left(\widetilde{\theta}_n^X - \theta_0\right) - H_{n-1}\right] \overset{a.s.}{=} 0. \tag{3.73}$$

Proof. The strong consistency is immediate from (3.65). We then express $\widetilde{\theta}_n^X - \theta_0$ as a function of $\widetilde{\rho}_n - \rho_0$, in order to deduce its asymptotic distribution from (3.66)-(3.68). We write

$$\begin{aligned}\widetilde{\theta}_n^X &- \theta_0 \\ &= \frac{\sum_k a_k \left(\rho_0^{-k} - \widetilde{\rho}_n^{-k}\right) + \sum_k a_k \rho_0^{-k} \sum_k b_k \left(\rho_0^{-k} - \widetilde{\rho}_n^{-k}\right) - \sum_k b_k \rho_0^{-k} \sum_k a_k \left(\rho_0^{-k} - \widetilde{\rho}_n^{-k}\right)}{\sum_k a_k \widetilde{\rho}_n^{-k} \sum_k a_k \rho_0^{-k}}, \end{aligned} \tag{3.74}$$

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

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

in order to obtain

$$\begin{aligned}\widetilde{\theta}_n^X - \theta_0 = (\widetilde{\rho}_n - \rho_0) &\left[\frac{\sum_{k=1}^d a_k \sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l}}{\widetilde{\rho}_n \rho_0 \sum_{k=1}^d a_k \widetilde{\rho}_n^{-k} \sum_{k=1}^d a_k \rho_0^{-k}} \right. \\ &\left. + \frac{\sum_{k=1}^d a_k \rho_0^{-k} \sum_{k=1}^d b_k \sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l} - \sum_{k=1}^d b_k \rho_0^{-k} \sum_{k=1}^d a_k \sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l}}{\widetilde{\rho}_n \rho_0 \sum_{k=1}^d a_k \widetilde{\rho}_n^{-k} \sum_{k=1}^d a_k \rho_0^{-k}} \right]. \end{aligned} \tag{3.75}$$

By (3.65), the square bracket of (3.75) almost surely converges to $\sqrt{C_0}$, and (3.71)-(3.73) are immediately deduced from (3.66)-(3.68). □

Unfortunately this 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 can not provide here an asymptotic confidence interval solely based on the observations, as we did for the estimators $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ (see Theorem 3.1 and Corollary 3.8).

Remark 3.12. $\widetilde{\rho}_n$ seems to have no interesting asymptotical properties for n fixed, as $|\mathbf{X}_0| \to \infty$, when $d > 1$ (if $d = 1$ then it reduces to the Harris estimator which is also the CLSE, hence Section 3.1 can be applied), unless we assume that, for all $i = 1 \dots d$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{X_{0,i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \frac{\eta_i}{\boldsymbol{\eta} \cdot \mathbf{1}}. \tag{3.76}$$

If this holds, then for any multitype branching process $(\mathbf{X}_k)_{k \geqslant 0}$ of any class of criticality, $\lim_{|\mathbf{X}_0|} \widetilde{\rho}_n \overset{a.s.}{=} \rho_0$. It is however obvious that assumption (3.76) is much too strong and nearly never applicable.

4 Comparison of the estimators and illustration of the asymptotic

4.1 Comparison of the estimators

In this section we compare the three estimators on a set of simulated trajectories, for several values of $|\mathbf{X}_0|$ and n. As a context of simulation, we choose the BSE epidemic in Great-Britain which, as detailed in Section 5, can be modeled by a multitype branching process $(\mathbf{X}_k)_{k \geqslant 0}$ of the form (2.1), with $d = 9$ types. For all $k = 1 \dots 9$, $\Psi_k(\theta_0) = a_k \theta_0 + b_k$ is of the form (2.10), where $\theta_0 = \theta_{hor.}$ (see Section 5), and the a_k and b_k are given in Table 6. The process is then subcritical (resp. supercritical) for $\theta_0 < \theta_{crit.}$ (resp. $\theta_0 > \theta_{crit.}$), with $\theta_{crit.} \simeq 23$.

We focus on the three following set of trajectories. Fixing the parameter $\theta_0 = 15$, we first simulate trajectories of the unconditioned subcritical process $(\mathbf{X}_k)_{k \geqslant 0}$, and then of the conditioned subcritical process $(\mathbf{X}_k | \mathbf{X}_k \neq \mathbf{0})_{k \geqslant 0}$. We finally simulate, with the parameter $\theta_0 = 35$, trajectories of the unconditioned supercritical process $(\mathbf{X}_k)_{k \geqslant 0}$. We consider different values of $|\mathbf{X}_0|$ and n, namely $|\mathbf{X}_0| = 10, 100, 1000$ and $n = 10, 50, 100$. For every couple $(|\mathbf{X}_0|, n)$, we simulate, in each of the previously mentioned cases, 100 trajectories of length n (*i.e* not extinct at time n), initiated by $\mathbf{X}_0 = (0, \dots, 0, |\mathbf{X}_0|)$, and compute the corresponding empirical means and standard deviations of the estimators. These are reported in Tables 1-3, which allow to compare the three different estimators in each of these situations.

n \ $\lvert\mathbf{X}_0\rvert$		10 mean	10 std. dev.	100 mean	100 std. dev.	1000 mean	1000 std. dev.
10	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	14.7179	4.8811	14.7806	1.5794	14.9930	0.4526
	$\widehat{\theta}^Z_n$	14.7181	4.8805	14.7806	1.5794	14.9930	0.4526
	$\widetilde{\theta}^X_n$	22.2960	3.5440	22.1341	1.1615	22.3036	0.3438
50	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	/	/	15.1834	0.9552	14.9675	0.3370
	$\widehat{\theta}^Z_n$	/	/	15.1834	0.9551	14.9675	0.3371
	$\widetilde{\theta}^X_n$	/	/	19.0956	0.4860	18.9621	0.1803
100	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	/	/	/	/	/	/
	$\widehat{\theta}^Z_n$	/	/	/	/	/	/
	$\widetilde{\theta}^X_n$	/	/	/	/	/	/

Table 1: Empirical means and standard deviations of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, $\widehat{\theta}^Z_n$ and $\widetilde{\theta}^X_n$ corresponding to 100 trajectories of length n of the unconditioned subcritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0,\dots,0,|\mathbf{X}_0|)$ and simulated with the infection parameter $\theta_0 = 15$, for different couples $(|\mathbf{X}_0|, n)$.

n \ $\lvert\mathbf{X}_0\rvert$		10 mean	10 std. dev.	100 mean	100 std. dev.	1000 mean	1000 std. dev.
10	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	14.4306	5.1569	14.8198	1.5400	14.9138	0.5442
	$\widehat{\theta}^Z_n$	14.4306	5.1568	14.8198	1.5400	14.9138	0.5442
	$\widetilde{\theta}^X_n$	22.0041	3.6403	22.1723	1.1272	22.2378	0.4094
50	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	16.0774	2.2719	15.0800	1.0376	15.0420	0.3276
	$\widehat{\theta}^Z_n$	14.6195	3.3079	15.0595	1.0550	15.0428	0.3248
	$\widetilde{\theta}^X_n$	19.7192	1.1291	19.0371	0.5284	18.9985	0.1714
100	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	17.7708	1.3873	15.1534	0.9573	15.0346	0.4027
	$\widehat{\theta}^Z_n$	14.7098	2.6979	14.8563	1.0287	15.0208	0.4047
	$\widetilde{\theta}^X_n$	20.4621	0.7545	19.0074	0.4620	18.9211	0.1943

Table 2: Empirical means and standard deviations of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, $\widehat{\theta}^Z_n$ and $\widetilde{\theta}^X_n$ corresponding to 100 trajectories of length n of the conditioned subcritical process $(\mathbf{X}_k|\mathbf{X}_k \neq \mathbf{0})_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0,\dots,0,|\mathbf{X}_0|)$ and simulated with the infection parameter $\theta_0 = 15$, for different couples $(|\mathbf{X}_0|, n)$.

n \ $\lvert\mathbf{X}_0\rvert$		10 mean	10 std. dev.	100 mean	100 std. dev.	1000 mean	1000 std. dev.
10	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	35.3485	6.2014	35.2777	1.6247	34.9629	0.6258
	$\widehat{\theta}^Z_n$	35.3611	6.1672	35.2777	1.6295	34.9630	0.6271
	$\widetilde{\theta}^X_n$	38.4696	5.3626	38.5015	1.4765	38.2363	0.5670
50	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	34.7898	1.2210	34.9792	0.2760	35.0008	0.0860
	$\widehat{\theta}^Z_n$	34.7898	1.2205	34.9792	0.2764	35.0008	0.0953
	$\widetilde{\theta}^X_n$	34.8578	1.2613	35.0580	0.2816	35.0816	0.0877
100	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	34.9942	0.1014	35.0056	0.0302	35.0021	0.0107
	$\widehat{\theta}^Z_n$	34.9943	0.1042	35.0056	0.0300	35.0000	0.0000
	$\widetilde{\theta}^X_n$	34.9930	0.1025	35.0053	0.0313	35.0032	0.0116

Table 3: Empirical means and standard deviations of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, $\widehat{\theta}^Z_n$ and $\widetilde{\theta}^X_n$ corresponding to 100 trajectories of length n of the unconditioned supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0,\dots,0,|\mathbf{X}_0|)$ and simulated with the infection parameter $\theta_0 = 35$, for different couples $(|\mathbf{X}_0|, n)$.

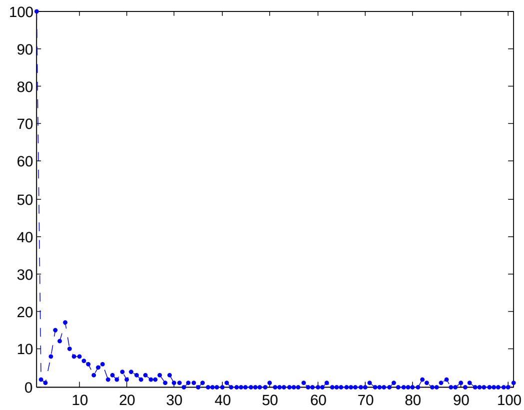


Figure 2: Simulation of a trajectory of length $n = 100$ of the conditioned subcritical process $(\mathbf{X}_k|\mathbf{X}_k \neq \mathbf{0})_{k\geqslant 0}$, initiated by $(0, \dots, 0, 100)$, with the infection parameter $\theta_0 = 15$.

The empty entries in Table 1 are due to the fact that for some given couples $(|\mathbf{X}_0|, n)$, trajectories of the subcritical process initiated by $\mathbf{X}_0$ with an extinction time greater than n occur only with a very small probability. We recall that the estimator $\widehat{\theta}_n^Z$ has no explicit form. Its precision thus depends on the optimization method which is chosen, while the precision for $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widetilde{\theta}_n^X$ solely depends on the computing program. As a consequence, the estimations obtained with $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ might slightly differ from each other, even when they are in theory equal, *i.e.* on trajectories with no sequence of $d - 1 = 8$ zeros. We can see however in Table 1 and Table 3 that, in our example, this approximation error remains very small.

Table 1 enables to compare $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widetilde{\theta}_n^X$. As just mentioned, $\widehat{\theta}_n^Z$ is, in this case, equal to $\widehat{\theta}_{|\mathbf{X}_0|}^X$ since no trajectory contains 8 consecutive zeros. Obviously, $\widehat{\theta}_{|\mathbf{X}_0|}^X$ provides an estimation of the parameter much closer to θ_0 than $\widetilde{\theta}_n^X$, which is of no surprise, since $\widetilde{\theta}_n^X$ is not proved to be consistent in the subcritical case. This table provides moreover an illustration of the consistency of $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and a probable non consistency of $\widetilde{\theta}_n^X$, as $|\mathbf{X}_0|$ tends to infinity, which appears clearly for $n = 10$.

Table 2 illustrates again the fact that $\widetilde{\theta}_n^X$ is not accurate when the process is not supercritical. This table is however very interesting to compare $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ on trajectories which might present one or several sequences of 8 zeros, typically trajectories of the conditioned process $(\mathbf{X}_k|\mathbf{X}_k \neq \mathbf{0})_{k\geqslant 0}$ in the subcritical case, for n large enough (see Figure 2). It appears that for long trajectories (*e.g.* $n = 50$ or $n = 100$), we obtain a better empirical mean with the estimator $\widehat{\theta}_n^Z$, which was expected since it takes into account more information, but a larger standard deviation. This is particularly obvious when the initial size of the clinical population is small ($|\mathbf{X}_0| = 10$). In order to better illustrate this phenomenon, we represent in Figure 3 the estimations obtained with the estimators $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ for the 100 simulated trajectories, used in Table 2, of the conditioned process with the infection parameter $\theta_0 = 15$, with $|\mathbf{X}_0| = 10$, respectively for $n = 50$ and for $n = 100$. It appears that, as n increases, $\widehat{\theta}_{|\mathbf{X}_0|}^X$ tends to overestimate θ_0, while $\widehat{\theta}_n^Z$ remains close to θ_0 but with a larger standard deviation. Moreover, Table 2 illustrates the consistency of $\widehat{\theta}_n^Z$, as n tends to infinity, as well as its consistency, as $|\mathbf{X}_0|$ tends to infinity (see Remark 3.9).

Finally, Table 3 allows to compare $\widehat{\theta}_{|\mathbf{X}_0|}^X$ with $\widetilde{\theta}_n^X$ in the supercritical case (again, $\widehat{\theta}_n^Z$ is here in theory equal to $\widehat{\theta}_{|\mathbf{X}_0|}^X$). On those examples, $\widehat{\theta}_{|\mathbf{X}_0|}^X$ provides a better estimation when the number of observations is small ($n = 10$). However, when n increases, the consistency of $\widetilde{\theta}_n^X$ comes in play, and it appears that $\widetilde{\theta}_n^X$ seems as good as the CLSE $\widehat{\theta}_{|\mathbf{X}_0|}^X$, although $\widetilde{\theta}_n^X$ is originally not built as an estimator with usual characteristics (CLSE, MLE, moments estimator...), but rather as an explicit estimator based upon realistic data.

4.2 Asymptotic normal distribution

The aim of this section is to illustrate the asymptotic normal distribution of each of the three estimators, namely (3.15), (3.53) and (3.71) (occurring in the case $\lambda^2 < \rho_0$).

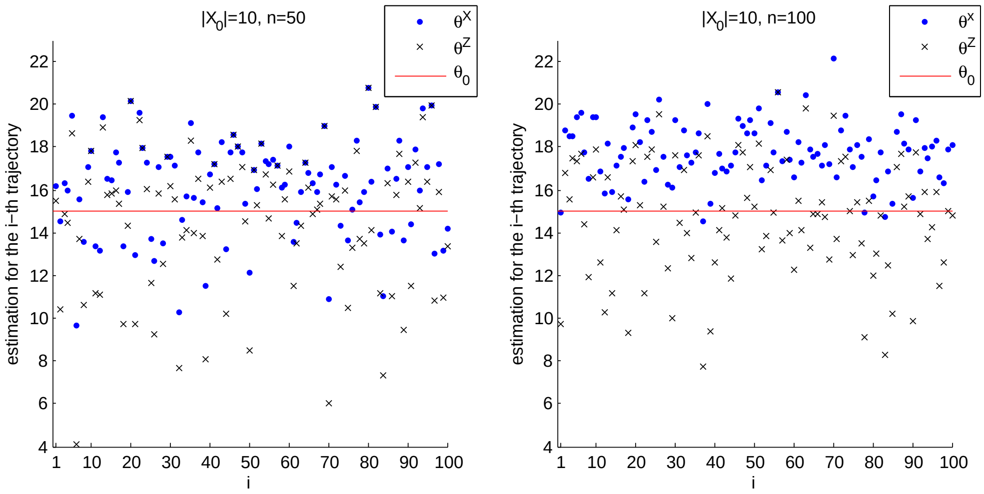


Figure 3: Estimations with $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ for 100 simulated trajectories (used in Table 2) of length n esp.h

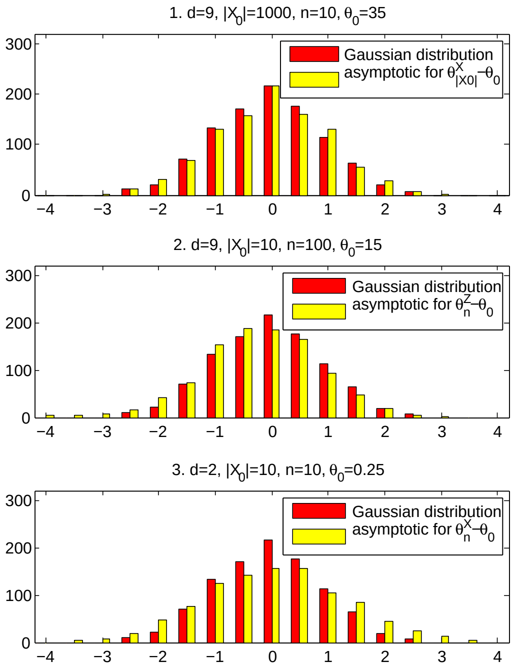


Figure 4: 1) Empirical distribution of (4.1) for 1000 trajectories of length $n = 10$ of the supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0, \ldots, 0, 1000)$, with $\theta_0 = 35$. 2) Empirical distribution of (4.2) for 1000 trajectories of length $n = 100$ of the conditioned subcritical process $(\mathbf{Z}_k)_{k\geqslant 0} = (\mathbf{X}_k|\mathbf{X}_k \neq \mathbf{0})_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0, \ldots, 0, 10)$, with $\theta_0 = 15$. 3) Empirical distribution of (4.3) for 1000 trajectories of length $n = 100$ of the supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0, 10)$, with $\theta_0 = 0.25$. Comparison with the empirical Gaussian distribution.

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Cases	7137	14181	25032	36682	34370	23945	14302	8016	4312	3179
Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Cases	2274	1355	1113	1044	549	309	203	104	53	33

Table 4: Yearly number of cases of BSE reported in Great-Britain from 1989 to 2008 ([17]).

then represent in Figure 4.3 the empirical distribution of

$$\sqrt{\mathbf{X}_n.\boldsymbol{\xi}\left(1+\ldots+\rho_0^{n-1}\right)^2\left(\rho_0^{2n}\beta_n C_0\right)^{-1}}\left(\widetilde{\theta}_n^X-\theta_0\right) \tag{4.3}$$

corresponding to the 1000 trajectories. It appears to be not so close to the Gaussian distribution, which can be due to the several approximations just mentioned.

5 Illustration on the BSE epidemic in Great-Britain

Bovine Spongiform Encephalopathy is a fatal neurodegenerative transmissible disease, the main routes of which are horizontal *via* protein supplements (Meat and Bone Meal, milk replacers), and maternal from a cow to its calf, until the main feed ban regulation introduced by the British government in July 1988. Even 20 years later, cases of BSE are still reported in Great-Britain (33 cases in 2008, see Table 4). Since most of cattle are slaughtered before the age of 10 years, and since a previous statistical study concluded to the full efficiency of the 1988 feed ban (see [10]), this could suggest the existence of an other source of horizontal infection, *e.g. via* the ingestion of excreted prions from other alive infected animals. We aim at quantifying this infection, using the yearly number of cases of BSE reported in Great-Britain from 1989 to 2008 (Table 4).

As detailed in [10] and [12], we can describe the incidence of the clinical cases by model (2.9), with $a_M = 10$. The parameters Ψ_k, $k = 1\ldots 9$, are given by (2.10). The observed survival probabilities in Great-Britain $(S_k)_{1\leqslant k\leqslant 10}$ are given in Table 5 ([16]). Basing ourselves on

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
0.97	0.65	0.36	0.30	0.25	0.18	0.10	0.06	0.02	0.01

Table 5: Observed survival probabilities of cattle in Great-Britain ([16]). S_k is the probability for an (apparently) healthy animal to survive at least until k years.

k	1	2	3	4	5	6	7	8	9
$a_k.10^3$	0.2192	1.9315	5.5275	9.2323	10.4353	8.3260	5.1357	1.8642	0.5569
$b_k.10^4$	0.0738	0.5432	1.8025	3.7227	5.0766	4.3821	3.4238	1.2428	0.5569

Table 6: Values of a_k and b_k defined by (2.10).

[10], we choose for the intrinsic incubation period distribution a discretized Weibull distribution, $P_{inc.}(k) := e^{-\frac{\alpha-1}{\alpha\beta^\alpha}(k-1)^\alpha} - e^{-\frac{\alpha-1}{\alpha\beta^\alpha}k^\alpha}$, where $\alpha = 3.84$ is the shape parameter, and $\beta = 7.46$ is the mode of the probability density of the corresponding Weibull distribution. These values α and β correspond to a maximum *a posteriori* Bayesian estimation, based on 22 observations during the years 1981 and 1987-2007 (see [10]). We set moreover $\theta_{vert.} = 0.1$, which is the maximum likely maternal infection probability usually assumed in BSE. The unknown parameter is the horizontal infection parameter $\theta_0 := \theta_{hor.}$ (corresponding to the mean number per infective and per year of newly infected animals *via* ingestion of excreted prions), and we have, for all $k = 1 \dots 9$, $\Psi_k(\theta_0) = a_k\theta_0 + b_k$, with a_k and b_k given by (2.10). The values of a_k and b_k are given in Table 6. Moreover, applying (2.7) we obtain that

$$\text{the process is subcritical} \iff \sum_{k=1}^{9} \Psi_k(\theta_0) < 1 \iff \theta_0 < \theta_{crit.},$$

where $\theta_{crit.} = (1 - \sum_{k=1}^9 b_k)/\sum_{k=1}^9 a_k \simeq 23$. Considering a previous maximum *a posteriori* Bayesian estimation of θ_0 ($\widehat{\theta}^{MAP} = 2.43$, see [10]), we can reasonably assume that we are in the subcritical case. The initial time for the multitype process $(\mathbf{X}_k)_{k\geqslant 0}$ corresponds here to the year 1997, with the memory over the years 1989-1997. We thus have $\mathbf{X}_0 = (4312, \dots, 7137)$ and $|\mathbf{X}_0| = 167977$. Since, in the data provided in Table 4, there is no sequence of 8 zeros, the estimator $\widehat{\theta}_n^Z$, which concerns the process $\mathbf{X}_k$ conditioned on the event $\mathbf{X}_k \neq \mathbf{0}$, is equal to the estimator $\widehat{\theta}_{|\mathbf{X}_0|}^X$. We obtain the following estimation

$$\widehat{\theta} := \widehat{\theta}_{167977}^X = \widehat{\theta}_{11}^Z = 2.4486. \tag{5.1}$$

Using (3.15) we obtain the following asymptotic confidence interval in $|\mathbf{X}_0|$. We have $|\mathbf{X}_0| = 167977$, hence we choose the following coefficients α_i defined in (3.13), $\alpha_1 = X_{0,1}/|\mathbf{X}_0| = 4312/167977$, …, $\alpha_9 = 7131/167977$. We compute

$$c_1 := \sqrt{\frac{\left(\sum_{k=0}^{10} \boldsymbol{a}\cdot\mathbf{X}_k\right)\left(\sum_{k=0}^{10}\sum_{j=1}^{d}\sum_{i=1}^{d} X_{0,j}a_i m_{ji}^{(k)}(\widehat{\theta})\right)}{\sum_{k=0}^{10}\sum_{j=1}^{d}\sum_{i=1}^{d} X_{0,j}\left(a_i\widehat{\theta}+b_i\right)m_{ji}^{(k)}(\widehat{\theta})}} \simeq 40.3938.$$

A confidence interval with asymptotic probability 95% is then $[\widehat{\theta}-1.96/c_1, \widehat{\theta}+1.96/c_1]$, *i.e.* $\mathbb{P}\Big(\theta_0 \in [2.4000, 2.4971]\Big) \approx 0.95$, which is very narrow.

Using (3.53) we obtain the following asymptotic confidence interval in n. Note that since no sequence of 8 zeros appears in the trajectory, formula (3.53) can be simplified and we have

$$\begin{aligned} c_2 &:= \frac{\sum_{k=0}^{11}\left(f'\left(\widehat{\theta},\mathbf{Z}_k\right)\right)^2}{\sqrt{\sum_{k=0}^{11}\left(f'\left(\widehat{\theta},\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta},\mathbf{Z}_k\right)(\boldsymbol{a}\cdot\mathbf{Z}_k)^{-1/2}}} \\ &= \frac{\sum_{k=0}^{11}\boldsymbol{a}\cdot\mathbf{X}_k}{\sqrt{\sum_{k=0}^{11}\left(\boldsymbol{a}\cdot\mathbf{X}_k+\widehat{\theta}\boldsymbol{b}\cdot\mathbf{X}_k\right)}} \simeq 40.3939. \end{aligned}$$

A confidence interval with asymptotic probability 95% is then $[\widehat{\theta}-1.96/c_2, \widehat{\theta}+1.96/c_2]$, *i.e.* $\mathbb{P}\Big(\theta_0 \in [2.4000, 2.4971]\Big) \approx 0.95$.

We can justify as follows why the two constants c_1 and c_2 are in our example very close. The number of clinical cases for the 9 first years 1989-1997 are very high compared to the following years, so we can roughly approximate c_1 and c_2 by neglecting the values after 1997. The constant c_2 should be thus close to

$$c_2 \approx \frac{\sum_{k=0}^{8}\sum_{j=1}^{9-k} X_{0,j}a_{j+k}}{\sqrt{\sum_{k=1}^{8}\sum_{j=1}^{9-k} X_{0,j}\left(a_{j+k}\widehat{\theta}+b_{j+k}\right)}}.$$

Moreover, since the values of the $\Psi_k(\widehat{\theta})$ are much smaller than 1 ($\sum_{k=1}^{9}\Psi_k(\widehat{\theta}) \approx 0.1$), we can neglect them and approximate the matrix $\mathbf{M}(\widehat{\theta})$ by the null matrix with an upper diagonal of ones. We then have, for all $i, j = 1 \dots d$, $m_{ij}^{(k)}(\widehat{\theta}) \approx \delta_{i,j+k}$ if $0 \leqslant k \leqslant 8$ and $m_{ij}^{(k)} \approx 0$ if $k > 8$. Together with the first approximation, the constant c_1 can thus roughly be approximated by

$$c_1 \approx \sqrt{\frac{\left(\sum_{k=0}^{8}\sum_{j=1}^{9-k} X_{0,j}a_{j+k}\right)\left(\sum_{k=0}^{8}\sum_{j=1}^{9-k} X_{0,j}a_{j+k}\right)}{\sum_{k=0}^{8}\sum_{j=1}^{9-k} X_{0,j}\left(a_{j+k}\widehat{\theta}+b_{j+k}\right)}},$$

which is the same approximation as for c_2 and equals 38.3679.

This estimation of θ_0 suggest that nowadays there could still be a minor infection which is not of maternal type . The average number of newly infected animals *via* this mean of infection is only of the order of 2 to 3 per infective and per year, which is really small compared to the estimations obtained in [10] for the infection *via* Meat and Bone Meal or lactoreplacers (before 1989), which are of the order of 1000. To be more reliable, this estimation should be completed by a sensitivity analysis to the other parameters (incubation period, maternal infection). This will be done in a further publication focusing on the BSE epidemic.

Remark 5.1. Estimating the Perron's root of the process with the estimator (3.58), we obtain $\widetilde{\theta}_{11}^X = 7.5495$, which is far from the expected value given by the Bayesian estimation $\widehat{\theta}^{MAP} = 2.43$.

6 Conclusion

According to the simulations presented in Section 4, the conditional least squares estimators $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ appear to be accurate and equivalent estimators of θ_0 at finite distance $(|\mathbf{X}_0|, n)$ in the epidemic model introduced in Section 2.2, with moreover good asymptotic properties for any class of criticality. In addition, the estimator $\widehat{\theta}_n^Z$, which takes into account more information, provides for long trajectories with sequences of zeros, estimations which are, according to the simulations, better in mean but which have a larger standard deviation. The estimator $\widetilde{\theta}_n^X$ derived from the explicit estimator $\widetilde{\rho}_n$ of the Perron's root introduced in [1], only provides satisfying estimations in the supercritical case, which are, in this case, not as good or are equivalent to the ones obtained with $\widehat{\theta}_{|\mathbf{X}_0|}^X$. Due to the differentiation of three cases and to the presence of unexplicitly known random variables in its asymptotic distribution, it is not possible to build an asymptotic confidence interval of θ_0 based on $\widetilde{\theta}_n^X$. The use of $\widetilde{\theta}_n^X$ is thus in our specific case less appropriate than the use of the CLSE, which is confirmed by the results obtained on the concrete example of Section 5. However, we point out that $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ are of an extremely more limited use that $\widetilde{\rho}_n$, since they do not provide an estimation of the Perron's root and only concern very specific processes, while $\widetilde{\rho}_n$ is suitable for any multitype BGW process.

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.

[2] Athreya, K. B. and Ney, P. E. (1972). *Branching processes*, Springer-Verlag, New-York.

[3] Becker, N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics* **33**, 515–522.

[4] Blum, J. R., Hanson, D. L. and Rosenblatt, J. I. (1963). On the central limit theorem for the sum of a random number of independent random variables. *Prob. Theory Rel. Fields* **1**, 389–393.

[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.

[6] Dion, J. P. (1974). Estimation of the mean and the initial probabilities of a branching process. *J. Appl. Probab.* **11**, 687–694.

[7] 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.

[8] Haccou, P., Jagers, P. and Vatutin, V. A. (2005). *Branching processes: variation, growth and extinction of populations*, Cambridge Studies in Adaptive Dynamics, Cambridge.

[9] Harris, T. E. (1948). Branching processes. *Ann. Math. Statistics* **19**, 474–494.

[10] Jacob, C., Maillard-Teyssier, L., Denis, J. B. and Bidot, C. (2009). A branching process approach for the propagation of the Bovine Spongiform Encephalopathy in Great-Britain. In *Branching processes and their Applications*, Lecture Notes in Statistics, Springer-Verlag, 227–242.

[11] Jacob, C. (2009). Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.* **38**, 566–597.

[12] Jacob, C. and Pénisson, S. (2010). A general multitype branching process with age, memory and population dependance. Submitted to *Adv. Appl. Prob.*

[13] 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.

[14] Prakasa Rao, B. L. S. (1999). *Semimartingales and their statistical inference*, Monographs on Statistics and Applied Probability 83, Chapmann & Hall/CRC.

[15] Rebolledo, R. (1980). Central limit theorems for local martingales. *Probab. Theory Relat. Fields* **51**, 269–286.

[16] Supervie V. and Costagliola, D. (2004). The unrecognised French BSE epidemic. *Vet. Res.* **51**, 349–362.

[17] World organisation for animal health (OIE). `http://www.oie.int/eng/info/en_esbru.htm`.