

ASYMPTOTIC INFERENCE FOR PARTIALLY OBSERVED BRANCHING PROCESSES

ANDREA KVITKOVIČOVÁ* AND

VICTOR M. PANARETOS,** *École Polytechnique Fédérale de Lausanne*

Abstract

We consider the problem of estimation in a partially observed discrete-time Galton–Watson branching process, focusing on the first two moments of the offspring distribution. Our study is motivated by modelling the counts of new cases at the onset of a stochastic epidemic, allowing for the facts that only a part of the cases is detected, and that the detection mechanism may affect the evolution of the epidemic. In this setting, the offspring mean is closely related to the spreading potential of the disease, while the second moment is connected to the variability of the mean estimators. Inference for branching processes is known for its nonstandard characteristics, as compared with classical inference. When, in addition, the true process cannot be directly observed, the problem of inference suffers significant further perturbations. We propose nonparametric estimators related to those used when the underlying process is fully observed, but suitably modified to take into account the intricate dependence structure induced by the partial observation and the interaction scheme. We show consistency, derive the limiting laws of the estimators, and construct asymptotic confidence intervals, all valid conditionally on the explosion set.

Keywords: Epidemic model; Galton–Watson branching process; partial observation; consistency; asymptotic distribution; martingale; stable convergence

2010 Mathematics Subject Classification: Primary 60J80; 62M05
Secondary 60J85; 92D30

1. Introduction

The last few years have witnessed several outbreaks of infectious diseases that spread out rapidly, threatening to become pandemics. The SARS epidemic in 2002, the bird flu epidemic in 2006, and the swine flu pandemic in 2009 are among the most serious recent examples of such infections in humans. When an outbreak of such a disease is recognised, it is important to decide whether measures should be taken to avoid the epidemic escalating into a pandemic. The efficacy of these measures depends on the promptness with which they are adopted, but they typically carry significant costs that would preferably be avoided: massive vaccination programmes (as in the recent swine flu epidemic), quarantining (recall the SARS epidemic), or culling in the case of animal diseases (for example, the foot and mouth outbreak in the UK). Thus, it is useful to have methods available that can guide the authorities in the evaluation of the spreading potential of a disease while it is still in its early stages.

In the initial stages of an epidemic, the number of cases is usually modelled to grow exponentially fast. Within the framework of stochastic models, branching processes provide

Received 25 November 2010; revision received 4 July 2011.

* Postal address: Section de Mathématiques, École Polytechnique Fédérale de Lausanne, EPFL-SB Station 8 - Bâtiment MA, CH-1015 Lausanne, Switzerland.

** Email address: victor.panaretos@epfl.ch

a reasonable approximation to an epidemic process before the depletion of the susceptibles becomes an issue (see Ball and Donnelly (1995)).

Branching processes have been studied extensively in the literature. Their use in practice requires estimation of crucial parameters from the data, which leads to the problem of statistical inference for branching processes (see Guttorp (1991) for example). This involves several nonstandard issues, such as the potential ‘problem’ of extinction. In the epidemic context, however, it is seldom realistic to assume that all cases are reported and recorded, a fact that was raised in Kendall (1956). Instead of the true process, we typically observe a random perturbation, for example a random subset of the cases. This motivates the study of statistical inference for branching processes under specific types of partial observation.

Inference for partially observed epidemic processes has been addressed by several authors, mostly in a continuous-time framework (see Becker and Hasofer (1997) for example). A classical continuous-time model for the number of cases is a linear birth-and-death process. By observing such a process at equally spaced discrete time points, we obtain a discrete-time Galton–Watson branching process, which in turn models the daily counts of cases. A key issue in this context is to maintain an equilibrium between realistic yet ‘solvable’ models. In what follows, we study statistical inference in a partially observed Galton–Watson branching process, a model treated in Meester *et al.* (2002) (who considered a particular instance of the model), Meester and Trapman (2006) (who investigated the effect on statistical inference), and Panaretos (2007) (who gave a coupling construction and investigated the probabilistic structure).

The parameter of central interest, as it crucially influences the behaviour of a Galton–Watson branching process (the extinction/explosion dichotomy), is the offspring mean. When the process is fully observed, the offspring variance plays an important role in the estimation of the mean, as it is related to the variability of the estimators. In this paper we study the estimation of the parameters assuming the roles of the offspring mean and variance in the partially observed branching process.

The estimation of parameters related to the moments of the offspring distribution under partial observation was studied in Meester and Trapman (2006). The authors focused on the consistency, and succeeded in constructing estimators of parameters related to the first three moments of the offspring distribution that are consistent on the explosion set, even though this is not in general possible under full observation.

In this paper we focus on the asymptotic distributions of estimators of parameters related to the first two moments, with the aim of constructing asymptotic confidence intervals. Since the asymptotic distributions of the estimators of Meester and Trapman (2006) are not easy to establish, we propose alternative estimators. We show that these are consistent and asymptotically Gaussian, all conditionally on the explosion set.

In the next section, we formally state the model, including all assumptions, motivate the new estimators, and state their properties. We then formally derive these properties in Sections 3 and 4. In Section 5 we study the performance of the estimators in a simulation study, and, finally, we conclude with a discussion of the results in Section 6.

2. The setting

The basic idea behind the discrete-time branching process epidemic model is as follows. At time k , there are Z_k infectious individuals. Between times k and $k + 1$, the i th individual infects $\xi_{i,k}$ individuals. The number of infectious individuals at time $k + 1$ is $Z_{k+1} = \sum_{i=1}^{Z_k} \xi_{i,k}$. The spreading potential is the same for all individuals, and is formalised by a common distribution on the $\xi_{i,k}$ (offspring distribution); each individual acts independently of the others. Note that this

interpretation implies that ‘a generation’ in the branching process framework is not equivalent to ‘a generation’ in the context of the epidemic. Indeed, one concrete individual can be represented as an offspring of himself or herself in several generations of the branching process.

In the model studied in Meester *et al.* (2002), Meester and Trapman (2006), and Panaretos (2007), this basic idea is extended to accommodate the possibility of partial observation. At time k , there are Z_k infectious individuals able to spread the disease. Out of these individuals, $Y_k \leq Z_k$ are detected between times k and $k+1$. Once an infectious individual is detected, he/she cannot spread the disease further. However, an infectious individual detected between times k and $k+1$ can transmit the disease between time k and the time of detection. Thus, the offspring distribution Υ of a detected individual may be different from the offspring distribution Ξ of an undetected individual. Each individual acts independently of the others (both with respect to spreading the disease and being detected), and the detection probability is the same for all individuals. The process Y_k is observed, while the process Z_k is not.

Meester *et al.* (2002) focused on infections for which it is reasonable to assume that an individual can neither recover nor die from the disease before detection, thus making detection the only means of leaving the infectious state. Meester and Trapman (2006) made the same assumption, but they did not use it to derive their results. Panaretos (2007) did not make this assumption and it is not made here either.

2.1. Notation and assumptions

The model described above can be formally specified as follows. Let Ξ and Υ be distributions of nonnegative integer-valued random variables with finite fourth moments such that $F_\Upsilon \geq F_\Xi$. Assume that Ξ is not a degenerate distribution concentrated at 0. Let ξ and $\{\xi_{i,k}\}_{i \geq 1, k \geq 0}$ be independent random variables with distribution Ξ , and let ζ and $\{\zeta_{i,k}\}_{i \geq 1, k \geq 0}$ be independent random variables with distribution Υ . Furthermore, let $\{B_{i,k}\}_{i \geq 1, k \geq 0}$ be independent Bernoulli random variables with success probability $(1 - \theta)$, where $0 < \theta < 1$. Assume that all these variables are mutually independent. We consider a random process $\{(Z_k, Y_k)^T; k \geq 0\}$ such that

$$\begin{aligned} Z_0 &= 1, \\ Y_k &= \sum_{i=1}^{Z_k} B_{i,k} \quad \text{for every } k \geq 0, \\ Z_{k+1} &= \sum_{i=1}^{Z_k} \xi_{i,k}(1 - B_{i,k}) + \sum_{i=1}^{Z_k} \zeta_{i,k} B_{i,k} \quad \text{for every } k \geq 0. \end{aligned}$$

As discussed in Panaretos (2007), this model has several appealing properties. For example, the process $\{(Z_k, Y_k)^T; k \geq 0\}$ is a Markov chain with time-homogeneous transition probabilities. Moreover, its first coordinate, the process $\{Z_k; k \geq 0\}$, is marginally a Galton–Watson branching process with offspring distribution Ψ —the mixture of the distributions Ξ and Υ in proportions θ and $1 - \theta$, respectively.

We consider only the case where there is a positive probability that the process will explode. That is, we assume that $m = E[Z_{k+1} \mid Z_k = 1] > 1$. Together with establishing asymptotic results conditionally on the explosion set $A = \{Z_n \rightarrow \infty; n \rightarrow \infty\}$, this approach is rather common in the theory of statistical inference for branching processes.

2.2. Simple estimators for fully observed branching processes

If we observe a trajectory of the process $\{Z_k; k \geq 0\}$ up to time n , we can choose from various estimators of the offspring mean m (see Guttorp (1991, Chapter 2) for example). A simple estimator

$$\bar{m}_n = \frac{Z_n}{Z_{n-1}} \mathbf{1}\{Z_{n-1} > 0\} + \mathbf{1}\{Z_{n-1} = 0\}$$

is based only on the last two observations. Alternatively, the estimator

$$\hat{m}_n = \frac{\sum_{k=1}^n Z_k}{\sum_{k=0}^{n-1} Z_k}$$

uses all observations available up to time n . Both estimators are strongly consistent on the explosion set (that is, they converge to m almost surely (a.s.) on A ; see Theorem 2.1 of Guttorp (1991)). Furthermore, they are asymptotically Gaussian conditionally on A . More precisely,

$$\begin{aligned} \sqrt{Z_{n-1}}(\bar{m}_n - m) &\xrightarrow{D} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A, \\ \sqrt{\sum_{k=0}^{n-1} Z_k}(\hat{m}_n - m) &\xrightarrow{D} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A, \end{aligned}$$

where

$$\sigma^2 = \text{var}[Z_{k+1} \mid Z_k = 1]$$

(see the proof of Theorem 2.3 of Guttorp (1991)).

If, in addition, one wishes to construct asymptotic confidence intervals for m , an estimator for σ^2 is needed. Various estimators have been proposed in the literature (see Chapter 3 of Guttorp (1991) for example). Perhaps the simplest one is

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n Z_{k-1} (\bar{m}_k - \hat{m}_n)^2.$$

The estimator is weakly consistent on A (that is, $\hat{\sigma}_n^2$ converges to σ^2 in probability restricted to A) and asymptotically Gaussian conditionally on A . More precisely,

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{D} N(0, 2\sigma^4) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A$$

(see Theorem 3.1 of Guttorp (1991) for both consistency and asymptotic distribution).

2.3. Simple estimators for partially observed branching processes

If instead of the trajectory of the process $\{Z_k; k \geq 0\}$, we observe only the trajectory of the process $\{Y_k; k \geq 0\}$ up to time n , it is not a priori clear how to estimate the offspring mean. Meester and Trapman (2006) proposed the estimators

$$\bar{m}_{Y,n} = \frac{Y_n}{Y_{n-1} + 1} \quad \text{and} \quad \hat{m}_{Y,n} = \frac{\sum_{k=1}^n Y_k}{\sum_{k=0}^{n-1} Y_k}.$$

Both $\bar{m}_{Y,n}$ and $\hat{m}_{Y,n}$ are strongly consistent on A (see Theorem 3.1 of Meester and Trapman (2006)), and $(\hat{m}_{Y,n} - m)$ converges to 0 in probability restricted to A at a rate of order $(\sum_{k=0}^{n-1} Y_k)^{1/2}$ (see Theorem 3.2 of Meester and Trapman (2006)).

As an alternative to σ^2 , Meester and Trapman (2006) considered the parameter

$$\gamma^2 = \theta m + (1 - \theta)\sigma^2 + (2 - \theta)m^2 - 2(1 - \theta)m E \xi.$$

Since we will show later that this is in fact the asymptotic variance of a mean estimator, we prefer to denote the parameter by γ^2 , even though it was introduced as γ in Meester and Trapman (2006). To see that γ^2 is indeed nonnegative, we initially note that the first summand in the definition of γ^2 is positive and the second summand is nonnegative; next, a simple calculation using the fact that $m = \theta E \xi + (1 - \theta) E \zeta$ yields $(2 - \theta)m^2 - 2(1 - \theta)m E \xi = \theta m((1 - \theta) E \xi + \theta E \zeta + E \xi - E \zeta)$. Since we assume that $E \xi \geq E \zeta$, this expression is positive. It follows that γ^2 is positive.

To estimate γ^2 , Meester and Trapman (2006) proposed the estimator

$$\hat{\gamma}_{Y,n}^2 = \frac{1}{n} \sum_{k=1}^n (Y_{k-1} + 1)(\bar{m}_{Y,k} - \hat{m}_{Y,n})^2.$$

The estimator is weakly consistent on A (see Theorem 3.1 of Meester and Trapman (2006)), and $(\hat{\gamma}_{Y,n}^2 - \gamma^2)$ converges to 0 in probability restricted to A with a rate of order at least $n^{-1/2+\delta}$ for every $\delta > 0$ (see Theorem 3.3 of Meester and Trapman (2006)).

The estimators of Meester and Trapman (2006) are analogous to the estimators used when the branching process is fully observed. Some of the properties of their analogues in the full-observation regime are also retained: consistency and (a bound on) the rate of convergence. However, whereas the estimators in the full-observation regime are asymptotically Gaussian, asymptotic distributions of the estimators in the partial-observation regime are not easy to establish. This is because the dependency structure of the process $\{Y_k; k \geq 0\}$ is more complex than that of the process $\{Z_k; k \geq 0\}$.

When studying properties of the estimator $\hat{\gamma}_{Y,n}^2$, Meester and Trapman (2006) resolved the difficulties arising from the complex dependency structure by using a ‘skipping’ idea (explained below). This idea can in fact be further exploited to obtain partial-observation estimators that retain the asymptotic properties of their full-observation analogues, including the asymptotic distribution. To this end, we will exploit the link between the dependency structure of the process $\{Z_k; k \geq 0\}$ and the estimators \hat{m}_n and $\hat{\sigma}_n^2$ established in Duby and Rouault (1982).

The random variables $Z_{n-1}^{-1/2}(Z_k - mZ_{k-1})$ for $k = 1, \dots, n$ are asymptotically independent and Gaussian as $n \rightarrow \infty$ (see the proof of Theorem 2.3 of Guttorp (1991)). Intuitively speaking, $Z_k = mZ_{k-1} + \sigma\sqrt{Z_{k-1}}\varepsilon_k$, where the ε_k are ‘approximately independent and standard Gaussian’. This leads to the ‘log-likelihood’

$$L = -\frac{1}{2}n \log(2\pi\sigma^2) - \frac{1}{2} \sum_{k=1}^n \log Z_{k-1} - \frac{1}{2\sigma^2} \sum_{k=1}^n \frac{(Z_k - mZ_{k-1})^2}{Z_{k-1}},$$

and \hat{m}_n and $\hat{\sigma}_n^2$ are the estimators maximising L .

Judging from the covariance structure of the variables $Y_{n-1}^{-1/2}(Y_k - mY_{k-1})$ for $k = 1, \dots, n$, it does not seem that the variables whose indices differ by 1 are asymptotically independent. However, asymptotic independence can be achieved by ‘skipping one index’ and considering the variables $Y_{2\tilde{n}_o}^{-1/2}(Y_{2k+1} - mY_{2k})$ for $k = 0, \dots, \tilde{n}_o$, where $\tilde{n}_o = \lfloor (n - 1)/2 \rfloor$. In Section 3 we show that these are asymptotically independent and Gaussian as $n \rightarrow \infty$. From $Y_{2k+1} = mY_{2k} + \gamma\sqrt{Y_{2k}}\varepsilon_k$, we obtain the ‘log-likelihood’

$$L_{Y,o} = -\frac{1}{2}\tilde{n}_o \log(2\pi\gamma^2) - \frac{1}{2} \sum_{k=0}^{\tilde{n}_o} \log Y_{2k} - \frac{1}{2\gamma^2} \sum_{k=0}^{\tilde{n}_o} \frac{(Y_{2k+1} - mY_{2k})^2}{Y_{2k}},$$

which is maximised by the estimators

$$\hat{m}_{Y,n,o} = \frac{\sum_{k=0}^{\tilde{n}_o} Y_{2k+1}}{\sum_{k=0}^{\tilde{n}_o} Y_{2k}} \quad \text{and} \quad \tilde{\gamma}^2_{Y,n,o} = \frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} Y_{2k} \left(\frac{Y_{2k+1}}{Y_{2k}} - \hat{m}_{Y,n,o} \right)^2.$$

To avoid problems with $\tilde{\gamma}^2_{Y,n,o}$ in cases where $Y_k = 0$ for some k (which has a positive probability), we consider the modified estimator

$$\hat{\gamma}^2_{Y,n,o} = \frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1) (\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,o})^2.$$

In Section 3 we show that the estimator $\hat{m}_{Y,n,o}$ is strongly consistent on A , and that

$$\sqrt{\sum_{k=0}^{\tilde{n}_o} Y_{2k}} (\hat{m}_{Y,n,o} - m) \xrightarrow{D} N(0, \gamma^2) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A.$$

In Section 4 we will show that the estimator $\hat{\gamma}^2_{Y,n,o}$ is weakly consistent on A , and that

$$\sqrt{\tilde{n}_o} (\hat{\gamma}^2_{Y,n,o} - \gamma^2) \xrightarrow{D} N(0, 2\gamma^4) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A.$$

In this sense, the estimators retain the asymptotic properties of their full-observation analogues.

Clearly, indices can be ‘skipped’ in many ways. For example, a ‘complementary skipping scheme’ gives the variables $Y_{2\tilde{n}_e-1}^{-1/2} (Y_{2k} - mY_{2k-1})$ for $k = 1, \dots, \tilde{n}_e$, where $\tilde{n}_e = \lfloor n/2 \rfloor$. These, following the same reasoning as before, yield the estimators

$$\hat{m}_{Y,n,e} = \frac{\sum_{k=1}^{\tilde{n}_e} Y_{2k}}{\sum_{k=1}^{\tilde{n}_e} Y_{2k-1}} \quad \text{and} \quad \hat{\gamma}^2_{Y,n,e} = \frac{1}{\tilde{n}_e} \sum_{k=1}^{\tilde{n}_e} (Y_{2k-1} + 1) (\bar{m}_{Y,2k} - \hat{m}_{Y,n,e})^2.$$

Both estimators are consistent ($\hat{m}_{Y,n,e}$ strongly and $\hat{\gamma}^2_{Y,n,e}$ weakly) and asymptotically Gaussian conditionally on A . This can be shown in the same way as for the estimators $\hat{m}_{Y,n,o}$ and $\hat{\gamma}^2_{Y,n,o}$ (details can be found in Kvitkovičová and Panaretos (2010)).

Intuitively speaking, to obtain the estimators from the two ‘complementary skipping schemes’, we divide the data into two parts: $Y_0, Y_2, \dots, Y_{2\tilde{n}_e}$ and $Y_1, Y_3, \dots, Y_{2\tilde{n}_o+1}$. We treat one part of the data as a threshold to which we compare the other part, and arrive at estimators of m and γ^2 . We then switch the roles of the two parts of the data and arrive at the second pair of estimators. This may suggest that combining the estimators of the two skipping schemes might improve efficiency. Recall, however, that using both parts of the data ‘on equal terms’ (like the estimators of Meester and Trapman (2006)) complicates the asymptotic distribution. Still, with the question of efficiency in mind, and in order to obtain a single estimator for each parameter, we consider the estimators

$$\begin{aligned} \hat{m}_{Y,n,a} &= \frac{1}{2} (\hat{m}_{Y,n,o} + \hat{m}_{Y,n,e}), \\ \hat{\gamma}^2_{Y,n,a} &= \frac{1}{2} \left(\frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1) (\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,e})^2 + \frac{1}{\tilde{n}_e} \sum_{k=1}^{\tilde{n}_e} (Y_{2k-1} + 1) (\bar{m}_{Y,2k} - \hat{m}_{Y,n,o})^2 \right), \\ \hat{\gamma}^2_{Y,n,aa} &= \frac{1}{2} \left(\frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1) (\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,a})^2 + \frac{1}{\tilde{n}_e} \sum_{k=1}^{\tilde{n}_e} (Y_{2k-1} + 1) (\bar{m}_{Y,2k} - \hat{m}_{Y,n,a})^2 \right). \end{aligned}$$

We do not study asymptotic distributions of these estimators. Note, however, that consistency and the rate of convergence of the estimator $\hat{m}_{Y,n,a}$ follow from the corresponding properties of the estimators $\hat{m}_{Y,n,o}$ and $\hat{m}_{Y,n,e}$. Consistency and the rates of convergence of the estimators $\hat{\gamma}_{Y,n,a}^2$ and $\hat{\gamma}_{Y,n,aa}^2$ follow from results given in Section 4. We comment on the practical performance of all estimators in Section 5.

3. The offspring mean estimator

In this section we prove the properties of the estimator $\hat{m}_{Y,n,o}$. We start with a brief presentation of some background material that provides the basis for proving the properties of interest. We then present some preliminary results, and end by stating and proving the main results.

3.1. The background

To establish the asymptotic distribution of the estimator, we will apply the central limit theorem of Scott (1978). The theorem deals with random sequences with the state space $(\mathbb{R}^{\mathbb{N}}, \mathcal{D})$, where $\mathbb{R}^{\mathbb{N}}$ is the space of real sequences $\mathbf{x} = (x_0, x_1, \dots)$ and \mathcal{D} is the Borel σ -field on $\mathbb{R}^{\mathbb{N}}$ corresponding to the metric $d(\mathbf{x}, \mathbf{y}) = \sup_{n \geq 0} \rho(\sum_{j=0}^n x_j, \sum_{j=0}^n y_j)$ with $\rho(a, b) = |a - b|/(1 + |a - b|)$ for $a, b \in \mathbb{R}$. When dealing with random sequences, it is common to work with the σ -field \mathcal{B} generated by the cylinder sets, so that the distribution of a sequence may be specified by finite-dimensional distributions. It turns out to be possible to switch between the two σ -fields, as $\mathcal{B} \subseteq \mathcal{D}$ and every measure on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B})$ corresponding to a set of consistent finite-dimensional distributions can be uniquely extended to a measure on $(\mathbb{R}^{\mathbb{N}}, \mathcal{D})$ (see Section 2 of Scott (1978)).

Theorem 3.1. (Scott (1978).) *Let $(S_n, \mathcal{F}_n)_{n \geq 0}$ be a martingale with martingale differences $X_0 = S_0$ and $X_n = S_n - S_{n-1}$, $n \geq 1$, conditional variance $V_n^2 = E X_0^2 + \sum_{k=1}^n E[X_k^2 | \mathcal{F}_{k-1}]$, and variance $s_n^2 = E S_n^2$ such that $s_n^2 < \infty$ for every $n \geq 0$. Suppose that*

- (i) $V_n^2/s_n^2 \xrightarrow{P} \eta^2$ as $n \rightarrow \infty$, where η is a random variable such that $\eta \geq 0$ a.s. and $E \eta^2 < \infty$;
- (ii) $s_{n-r}^2/s_n^2 \rightarrow p^{-r}$ as $n \rightarrow \infty$ for every integer $r \geq 0$, where p is a constant such that $p > 1$;
- (iii) $E[\exp\{it X_n/(aV_n)\} | \mathcal{F}_{n-1}] \xrightarrow{P} \exp\{-t^2/2\}$ as $n \rightarrow \infty$ on $\{\eta > 0\}$, where a is a constant such that $a > 0$.

Let $N = (N_0, N_1, \dots)$ be a random sequence with independent components such that $N_j \sim N(0, p^{-j})$, independent of η . For every $n \in \mathbb{N}$, consider a random sequence $\mathbf{M}_n = (X_n/(aV_n), X_{n-1}/(aV_n), \dots, X_1/(aV_n), X_0/(aV_n), 0, 0, \dots)$.

For every $E \in \mathcal{D}$ such that $P(N \in \partial E) = 0$, we have

$$P(\mathbf{M}_n \in E \mid \eta > 0) \rightarrow P(N \in E) \quad \text{as } n \rightarrow \infty.$$

The theorem asserts that the conditional distribution of the sequence \mathbf{M}_n converges weakly to a Gaussian measure. We will in fact only need the asymptotic distribution of a function of that sequence, namely of $\sum_{j=0}^n X_j/(aV_n)$. To establish this, we start by noting that $\sum_{j=0}^n X_j/(aV_n) = \limsup_{k \rightarrow \infty} \sum_{j=0}^k M_{nj}$, where M_{nj} is the j th component of the sequence \mathbf{M}_n . This is because $M_{nj} = 0$ for $j > n$ by definition.

It can be easily shown (directly from the definition of continuity) that the function $\limsup_{k \rightarrow \infty} \sum_{i=0}^k x_i$ is a continuous function from the metric space $(\mathbb{R}^{\mathbb{N}}, d)$ to the metric space $(\mathbb{R} \cup \{\pm\infty\}, \tilde{d})$, where $\tilde{d}(x, y) = |f(x) - f(y)|$, $f(x) = x/(1 + |x|)$ for $x \in \mathbb{R}$, and $f(x) = \pm 1$ for $x = \pm\infty$. By the continuous mapping theorem, it now follows that $\mathcal{L}(\sum_{j=0}^n X_j / (aV_n) \mid \eta > 0) \rightarrow \mathcal{L}(\limsup_{k \rightarrow \infty} \sum_{j=0}^k N_j)$ as $n \rightarrow \infty$.

It remains to determine the distribution on the right-hand side. The assumption that $p > 1$ implies that $\sum_{j=0}^{\infty} \text{var } N_j = \sum_{j=0}^{\infty} p^{-j} = p/(p - 1) < \infty$. Furthermore, the variables $\{N_j\}_{j \geq 0}$ are centred and mutually independent. It follows that $\lim_{k \rightarrow \infty} \sum_{j=0}^k N_j$ a.s. exists. Hence, $\limsup_{k \rightarrow \infty} \sum_{j=0}^k N_j = \lim_{k \rightarrow \infty} \sum_{j=0}^k N_j$ a.s., and they have the same distribution.

Since $\mathcal{L}(\sum_{j=0}^k N_j) = \mathcal{N}(0, \sum_{j=0}^k p^{-j})$ for every $k \in \mathbb{N}$, the characteristic function $\psi_k(t)$ of the random variable $\sum_{j=0}^k N_j$ satisfies

$$\psi_k(t) = \exp\left\{-\left(\sum_{j=0}^k p^{-j}\right)\frac{t^2}{2}\right\} \rightarrow \exp\left\{-\frac{p}{p-1}\frac{t^2}{2}\right\} \text{ as } k \rightarrow \infty.$$

Thus, $\sum_{j=0}^k N_j \xrightarrow{D} \mathcal{N}(0, p/(p - 1))$ as $k \rightarrow \infty$. Since $\lim_{k \rightarrow \infty} \sum_{j=0}^k N_j$ a.s. exists, it follows that its distribution is $\mathcal{N}(0, p/(p - 1))$.

These considerations imply that

$$\mathcal{L}\left(\sum_{j=0}^n \frac{X_j}{aV_n} \mid \eta > 0\right) \rightarrow \mathcal{N}\left(0, \frac{p}{p-1}\right) \text{ as } n \rightarrow \infty.$$

3.2. Preliminary results

We now list several results concerning functionals of the process $\{(Z_k, Y_k)^{\top}; k \geq 0\}$ that will be useful in the proofs of the main results. We give the proof of the crucial lemma here; the remaining lemmas are treated in Appendix A.

Lemma 3.1. *For every $n \geq 0$, we have $E Z_n = m^n$.*

Lemma 3.2. *For every $n \geq 0$, we have*

- (i) $E[Y_{n+1} - mY_n \mid Z_n] = 0$ a.s.;
- (ii) $E[(Y_{n+1} - mY_n)^2 \mid Z_n] = (1 - \theta)\gamma^2 Z_n$ a.s.

Lemma 3.3. *We have $Y_n/Z_n \rightarrow (1 - \theta)$ as $n \rightarrow \infty$ a.s. on A .*

Lemma 3.4. *Let $W = \lim_{n \rightarrow \infty} Z_n/m^n$ a.s. Then*

- (i) $\frac{m^2 - 1}{m^{2n+2} - 1} \sum_{k=0}^n Z_{2k} \rightarrow W$ a.s. as $n \rightarrow \infty$;
- (ii) $\frac{m^2 - 1}{m(m^{2n+2} - 1)} \sum_{k=0}^n Z_{2k+1} \rightarrow W$ a.s. as $n \rightarrow \infty$;
- (iii) $\frac{m^2 - 1}{m^{2n+2} - 1} \sum_{k=0}^n Y_{2k} \rightarrow (1 - \theta)W$ a.s. on A as $n \rightarrow \infty$;
- (iv) $\frac{m^2 - 1}{m(m^{2n+2} - 1)} \sum_{k=0}^n Y_{2k+1} \rightarrow (1 - \theta)W$ a.s. on A as $n \rightarrow \infty$.

Lemma 3.5. *We have*

$$E \left[\exp \left\{ it \sqrt{\frac{1}{(1-\theta)\gamma^2} \frac{Y_{n+1} - mY_n}{\sqrt{Z_n}}} \right\} \middle| Z_n \right] \rightarrow \exp \left\{ -\frac{t^2}{2} \right\} \quad \text{a.s. on } A \text{ as } n \rightarrow \infty.$$

Proof. For every $n \in \mathbb{N}$, let $A_n = Z_n^{-1/2}(Y_{n+1} - mY_n)$, and consider a version of the conditional distribution $P_{A_n | Z_n}$ defined for every $\omega \in \Omega$ and $B \in \mathcal{B}(\mathbb{R})$ as

$$P_{A_n | Z_n}(B, \omega) = \frac{P(A_n \in B, Z_n = Z_n(\omega))}{P(Z_n = Z_n(\omega))} = \frac{P(\{\omega'; A_n(\omega') \in B, Z_n(\omega') = Z_n(\omega)\})}{P(\{\omega'; Z_n(\omega') = Z_n(\omega)\})}.$$

We have

$$\begin{aligned} P(A_n \in B, Z_n = Z_n(\omega)) &= P(Z_n^{-1/2}(Y_{n+1} - mY_n) \in B, Z_n = Z_n(\omega)) \\ &= P\left(Z_n^{-1/2}\left(\sum_{i=1}^{Z_{n+1}} B_{i,n+1} - m \sum_{j=1}^{Z_n} B_{j,n}\right) \in B, Z_n = Z_n(\omega)\right) \\ &= P\left(Z_n^{-1/2}\left(\sum_{i=1}^{Z_n} \xi_{j,n}(1-B_{j,n}) + \zeta_{j,n} B_{j,n} \right) B_{i,n+1} - m \sum_{j=1}^{Z_n} B_{j,n}\right) \in B, Z_n = Z_n(\omega)\Big) \\ &= P\left(Z_n^{-1/2}\left(\sum_{j=1}^{Z_n} \sum_{i=1+\sum_{k=1}^{j-1} \xi_{k,n}(1-B_{k,n}) + \zeta_{k,n} B_{k,n}}^j \xi_{k,n}(1-B_{k,n}) + \zeta_{k,n} B_{k,n} \right) B_{i,n+1} - m \sum_{j=1}^{Z_n} B_{j,n}\right) \in B, \\ &\quad Z_n = Z_n(\omega)\Big) \\ &= P\left(Z_n^{-1/2}\left(\sum_{j=1}^{Z_n} \sum_{i=1}^{\xi_{j,n}(1-B_{j,n}) + \zeta_{j,n} B_{j,n}} \xi_{j,n}(1-B_{j,n}) + \zeta_{j,n} B_{j,n} \right) B_{i,j,n+1} - m \sum_{j=1}^{Z_n} B_{j,n}\right) \in B, Z_n = Z_n(\omega)\Big), \end{aligned}$$

where the $\{B_{i,j,n}\}_{i \in \mathbb{N}, j \in \mathbb{N}, n \in \mathbb{N}}$ are Bernoulli random variables with success probability $1 - \theta$, mutually independent and independent of all the other variables in the model.

Define the random variables

$$C_{j,n} = \sum_{i=1}^{\xi_{j,n}(1-B_{j,n}) + \zeta_{j,n} B_{j,n}} B_{i,j,n+1}$$

for every $n \in \mathbb{N}$ and $j \in \mathbb{N}$. Since $\{\xi_{i,k}\}_{i \in \mathbb{N}, k \geq n}$, $\{\zeta_{i,k}\}_{i \in \mathbb{N}, k \geq n}$, and $\{B_{i,k}\}_{i \in \mathbb{N}, k \geq n}$ are independent of $\{Z_k; k \leq n\}$ by definition, so are the $\{C_{i,k}\}_{i \in \mathbb{N}, k \geq n}$. We can thus write

$$\begin{aligned} P(A_n \in B, Z_n = Z_n(\omega)) &= P\left(Z_n(\omega)^{-1/2} \sum_{j=1}^{Z_n(\omega)} (C_{j,n} - mB_{j,n}) \in B, Z_n = Z_n(\omega)\right) \\ &= P\left(Z_n(\omega)^{-1/2} \sum_{j=1}^{Z_n(\omega)} (C_{j,n} - mB_{j,n}) \in B\right) P(Z_n = Z_n(\omega)). \end{aligned}$$

It follows that

$$P_{A_n | Z_n}(B, \omega) = \frac{P(A_n \in B, Z_n = Z_n(\omega))}{P(Z_n = Z_n(\omega))} = P\left(Z_n(\omega)^{-1/2} \sum_{j=1}^{Z_n(\omega)} (C_{j,n} - mB_{j,n}) \in B\right).$$

Since we are interested in the distribution conditional on $A = \{Z_n \rightarrow \infty\}$, we focus on the distribution of $z_n^{-1/2} \sum_{j=1}^{z_n} (C_{j,n} - mB_{j,n})$ for z_n such that $\lim_{n \rightarrow \infty} z_n = \infty$.

For a fixed n , the random variables $C_{j,n} - mB_{j,n}$ are independent and identically distributed. Furthermore, since $\xi_{j,n}, \zeta_{j,n}, B_{j,n}$, and $B_{i,j,n+1}$ are mutually independent, Wald's identities give $E C_{j,n} = E B_{i,j,n+1} E[\xi_{j,n}(1 - B_{j,n}) + \zeta_{j,n}B_{j,n}] = (1 - \theta)m$. Thus, the random variables $C_{j,n} - mB_{j,n}$ are centred.

We now calculate their variance: $\text{var}[C_{j,n} - mB_{j,n}] = E(C_{j,n} - mB_{j,n})^2 = E C_{j,n}^2 - 2m E C_{j,n}B_{j,n} + m^2 E B_{j,n}^2$. Using Wald's identities, we obtain

$$\begin{aligned} E C_{j,n}^2 &= \text{var } C_{j,n} + (E C_{j,n})^2 \\ &= E[\xi_{j,n}(1 - B_{j,n}) + \zeta_{j,n}B_{j,n}] \text{var } B_{1,j,n+1} \\ &\quad + \text{var}[\xi_{j,n}(1 - B_{j,n}) + \zeta_{j,n}B_{j,n}](E B_{1,j,n+1})^2 + (1 - \theta)^2 m^2 \\ &= \theta(1 - \theta)m + (1 - \theta)^2 \sigma^2 + (1 - \theta)^2 m^2. \end{aligned}$$

Furthermore, by the mutual independence of $\xi_{j,n}, \zeta_{j,n}$, and $B_{j,n}$,

$$\begin{aligned} -2m E C_{j,n}B_{j,n} &= -2m E \left[E \left[B_{j,n} \sum_{i=1}^{\xi_{j,n}(1-B_{j,n})+\zeta_{j,n}B_{j,n}} B_{i,j,n+1} \mid \xi_{j,n}, \zeta_{j,n}, B_{j,n} \right] \right] \\ &= -2(1 - \theta)m E B_{j,n}(\xi_{j,n}(1 - B_{j,n}) + \zeta_{j,n}B_{j,n}) \\ &= -2(1 - \theta)m E[B_{j,n}\xi_{j,n} - B_{j,n}^2\xi_{j,n} + B_{j,n}^2\zeta_{j,n}] \\ &= -2(1 - \theta)^2 m E \zeta. \end{aligned}$$

It follows that $\text{var}[C_{j,n} - mB_{j,n}] = (1 - \theta)\gamma^2$.

The classical central limit theorem now gives

$$((1 - \theta)\gamma^2 z_n)^{-1/2} \sum_{j=1}^{z_n} (C_{j,n} - mB_{j,n}) \xrightarrow{D} N(0, 1) \quad \text{as } n \rightarrow \infty.$$

It follows that

$$\mathcal{L}\left(\frac{Y_{n+1} - mY_n}{\sqrt{(1 - \theta)\gamma^2 Z_n}} \mid Z_n\right) \rightarrow N(0, 1) \quad \text{a.s. on } A \text{ as } n \rightarrow \infty,$$

and the proof is complete.

3.3. Main results

We now derive the properties of the estimator $\hat{m}_{Y,n,0}$. We begin with consistency.

Theorem 3.2. *The estimator $\hat{m}_{Y,n,0}$ is strongly consistent on A .*

Proof. Since $\{Z_k; k \geq 0\}$ is a branching process, we have $W = \lim_{n \rightarrow \infty} Z_n/m^n > 0$ a.s. on A . Using Lemma 3.4, we obtain

$$\begin{aligned} \hat{m}_{Y,n,o} &= \frac{\sum_{k=0}^{\tilde{n}_o} Y_{2k+1}}{\sum_{k=0}^{\tilde{n}_o} Y_{2k}} \\ &= m \left(\frac{m^2 - 1}{m(m^{2\tilde{n}_o+2} - 1)} \sum_{k=0}^{\tilde{n}_o} Y_{2k+1} \right) \left(\frac{m^2 - 1}{m^{2\tilde{n}_o+2} - 1} \sum_{k=0}^{\tilde{n}_o} Y_{2k} \right)^{-1} \\ &\rightarrow m \quad \text{a.s. on } A \text{ as } n \rightarrow \infty. \end{aligned}$$

Next, we proceed to the determination of the asymptotic distribution.

Theorem 3.3. *We have*

$$\sqrt{\sum_{k=0}^{\tilde{n}_o} Y_{2k} (\hat{m}_{Y,n,o} - m)} \xrightarrow{D} N(0, \gamma^2) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A.$$

Proof. We apply Theorem 3.1 in the following setting. For every $k \geq 0$, let

$$\mathcal{F}_k = \sigma\{Y_0, Z_1, Y_1, Z_2, Y_2, \dots, Z_{2k+1}, Y_{2k+1}, Z_{2k+2}\} \quad \text{and} \quad X_k = Y_{2k+1} - mY_{2k}.$$

We first show that $(S_n, \mathcal{F}_n)_{n \geq 0}$ is a martingale. Clearly, X_k is \mathcal{F}_k -measurable for every $k \geq 0$. Furthermore, $E[X_k | \mathcal{F}_{k-1}] = 0$ a.s. for every $k \in \mathbb{N}$ by Lemma 3.2. Finally, using Lemmas 3.2 and 3.1, we obtain

$$\begin{aligned} s_n^2 &= E S_n^2 \\ &= \sum_{k=0}^n E X_k^2 \\ &= E X_0^2 + \sum_{k=1}^n E[E[X_k^2 | \mathcal{F}_{k-1}]] \\ &= (1 - \theta)\gamma^2 \sum_{k=0}^n E Z_{2k} \\ &= (1 - \theta)\gamma^2 \sum_{k=0}^n m^{2k} \\ &= (1 - \theta)\gamma^2 \frac{m^{2n+2} - 1}{m^2 - 1} \end{aligned}$$

for every $n \geq 0$. Thus, $s_n^2 < \infty$ for every $n \geq 0$, and $(S_n, \mathcal{F}_n)_{n \geq 0}$ is a martingale.

For the conditional variance, we have

$$V_n^2 = E X_0^2 + \sum_{k=1}^n E[X_k^2 | \mathcal{F}_{k-1}] = (1 - \theta)\gamma^2 \sum_{k=0}^n Z_{2k} \quad \text{a.s.}$$

Thus, $V_n^2/s_n^2 = (m^2 - 1)/(m^{2n+2} - 1) \sum_{k=0}^n Z_{2k}$ a.s. By Lemma 3.4, it follows that $V_n^2/s_n^2 \rightarrow W$ a.s. as $n \rightarrow \infty$. Since $\{Z_k; k \geq 0\}$ is a branching process, we have $W \geq 0$ a.s.,

$E W = 1$, and $\{W > 0\} = A$ a.s. Hence, Theorem 3.1(i) is satisfied with $\eta^2 = W$ a.s., and $\{\eta > 0\} = A$ a.s.

Furthermore, since $m > 1$, we have $s_{n-r}^2/s_n^2 = (m^{2n-2r+2} - 1)/(m^{2n+2} - 1) \rightarrow m^{-2r}$ as $n \rightarrow \infty$ for every integer $r \geq 0$. Theorem 3.1(ii) is thus satisfied with $p = m^2$.

Finally,

$$\frac{X_n}{V_n} = \frac{Y_{2n+1} - mY_{2n}}{\sqrt{(1-\theta)\gamma^2 \sum_{k=0}^n Z_{2k}}} = \frac{Y_{2n+1} - mY_{2n}}{\sqrt{(1-\theta)\gamma^2 Z_{2n}}} \sqrt{\frac{Z_{2n}}{\sum_{k=0}^n Z_{2k}}}.$$

Since $\{Z_k; k \geq 0\}$ is a branching process, $W > 0$ a.s. on A , and, by Lemma 3.4, we obtain

$$\frac{Z_{2n}}{\sum_{k=0}^n Z_{2k}} = \frac{Z_{2n}}{m^{2n} \left(\frac{m^2 - 1}{m^{2n+2} - 1} \sum_{k=0}^n Z_{2k} \right)^{-1}} \frac{m^{2n}(m^2 - 1)}{m^{2n+2} - 1} \rightarrow 1 - m^{-2} \quad \text{a.s. on } A \text{ as } n \rightarrow \infty.$$

Lemma 3.5 together with Slutsky’s theorem now implies that Theorem 3.1(iii) is satisfied with $a = \sqrt{1 - m^{-2}}$. Recall that we are assuming that $m > 1$, and so $1 - m^{-2} > 0$.

Theorem 3.1 and the discussion following the theorem now gives

$$\begin{aligned} \sum_{j=0}^n \frac{X_j}{aV_n} &= \left((1 - m^{-2})(1 - \theta)\gamma^2 \sum_{k=0}^n Z_{2k} \right)^{-1/2} \sum_{k=0}^n (Y_{2k+1} - mY_{2k}) \\ &\xrightarrow{D} N\left(0, \frac{m^2}{m^2 - 1}\right) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A. \end{aligned} \tag{3.1}$$

By Lemma 3.4 and because $W > 0$ a.s. on A , we have

$$\begin{aligned} \frac{\sum_{k=0}^n Y_{2k}}{\sum_{k=0}^n Z_{2k}} &= \left(\frac{m^2 - 1}{m^{2n+2} - 1} \sum_{k=0}^n Y_{2k} \right) \left(\frac{m^2 - 1}{m^{2n+2} - 1} \sum_{k=0}^n Z_{2k} \right)^{-1} \\ &\rightarrow (1 - \theta) \quad \text{a.s. on } A \text{ as } n \rightarrow \infty. \end{aligned}$$

By Slutsky’s theorem and (3.1), it now follows that

$$\left(\sum_{k=0}^n Y_{2k} \right)^{-1/2} \sum_{k=0}^n (Y_{2k+1} - mY_{2k}) \xrightarrow{D} N(0, \gamma^2) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A,$$

which is equivalent to the statement of the theorem.

In the context of the preceding proof, we note that once the assumptions of Theorem 3.1 are verified, it is not difficult to derive the asymptotic distribution of the simple estimator $\bar{m}_{Y,n}$ of Meester and Trapman (2006). We first apply the continuity theorem to show that

$$\begin{aligned} \frac{X_{\tilde{n}_o}}{aV_{\tilde{n}_o}} &= \left((1 - m^{-2})(1 - \theta)\gamma^2 \sum_{k=0}^{\tilde{n}_o} Z_{2k} \right)^{-1/2} (Y_{2\tilde{n}_o+1} - mY_{2\tilde{n}_o}) \\ &\xrightarrow{D} N(0, 1) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A, \end{aligned}$$

(it is straightforward to verify that the projection on the first coordinate is a continuous function from the metric space $(\mathbb{R}^{\mathbb{N}}, d)$ to the metric space $(\mathbb{R}, |\cdot|)$). Using Lemma 3.4 and Slutsky’s

theorem, we can then show that $(Y_{2\tilde{n}_0} + 1)^{-1/2}(Y_{2\tilde{n}_0+1} - mY_{2\tilde{n}_0}) \xrightarrow{D} N(0, \gamma^2)$ as $n \rightarrow \infty$ conditionally on A , and afterwards that $\sqrt{Y_{2\tilde{n}_0} + 1}(\bar{m}_{Y,2\tilde{n}_0+1} - m) \xrightarrow{D} N(0, \gamma^2)$ as $n \rightarrow \infty$ conditionally on A . If we now denote by $F_n(x)$ the distribution function of the random variable $1/\sqrt{\gamma^2\sqrt{Y_{n-1} + 1}(Y_n/(Y_{n-1} + 1) - m)}$ conditional on A for every $n \in \mathbb{N}$, we have $\lim_{k \rightarrow \infty} F_{2k+1}(x) = \Phi(x)$ for every $x \in \mathbb{R}$. A similar argument for the ‘complementary skipping scheme’ gives $\lim_{k \rightarrow \infty} F_{2k}(x) = \Phi(x)$ for every $x \in \mathbb{R}$. It follows that $\lim_{n \rightarrow \infty} F_n(x) = \Phi(x)$ for every $x \in \mathbb{R}$. This leads to the following result.

Theorem 3.4. *We have*

$$\sqrt{Y_{n-1} + 1}(\bar{m}_{Y,n} - m) \xrightarrow{D} N(0, \gamma^2) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A.$$

4. The estimators of γ^2

Having established the asymptotic properties of the mean estimators, we now proceed to the derivation of the properties of the estimator $\hat{\gamma}_{Y,n,0}^2$ and related estimators. As in Section 3, we first give some background material, then present some preliminary results, and finally prove the main results.

4.1. The background

We will employ the following central limit theorem for martingale arrays.

Theorem 4.1. (Corollary 3.1 of Hall and Heyde (1980).) *Let k_n be constants such that $k_n \nearrow \infty$ as $n \rightarrow \infty$. For every $n \in \mathbb{N}$, let $\{S_{ni}, \mathcal{F}_{ni}; 0 \leq i \leq k_n\}$ be a zero-mean square-integrable martingale with martingale differences $X_{n0} = S_{n0}$ and $X_{nk} = S_{nk} - S_{n,k-1}$ for every $n \in \mathbb{N}$ and $1 \leq k \leq k_n$, and with conditional variance $V_{ni}^2 = \sum_{k=0}^i E[X_{nk}^2 | \mathcal{F}_{n,k-1}]$, where $\mathcal{F}_{n,-1} = \{\emptyset, \Omega\}$. Let η^2 be an a.s. finite random variable. Suppose that*

- (i) *for every $\varepsilon > 0$, $\sum_{k=0}^{k_n} E[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} | \mathcal{F}_{n,k-1}] \xrightarrow{P} 0$ as $n \rightarrow \infty$;*
- (ii) *$V_{nk_n}^2 \xrightarrow{P} \eta^2$ as $n \rightarrow \infty$;*
- (iii) *$\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ for every $n \in \mathbb{N}$ and $0 \leq i \leq k_n$.*

Then

$$S_{nk_n} \xrightarrow{D} Z \quad \text{as } n \rightarrow \infty,$$

where Z is a random variable with the characteristic function $E \exp\{-\frac{1}{2}\eta^2 t^2\}$.

4.2. Preliminary results

We again list several results concerning functionals of the process $\{(Z_k, Y_k)^\top; k \geq 0\}$ that will be useful in the proofs of the main results. We give the proof of the crucial lemma here; the remaining lemmas are treated in Appendix A.

Lemma 4.1. *For every $n \geq 0$ and $Z_n > 0$,*

$$E \left[\frac{1}{Y_n + 1} (Y_{n+1} - m(Y_n + 1))^2 \mid Z_n \right] = \gamma^2 + g(Z_n) \quad \text{a.s.,}$$

where

$$\begin{aligned}
 g(z) &= \frac{1}{(1-\theta)(z+1)}(1-\theta^{z+1}) \\
 &\quad \times (m^2 + \theta^2(E\xi)^2 - 2\theta m E\xi - \theta(1-\theta)E\zeta - (1-\theta)^2 \text{var } \zeta) \\
 &\quad - \theta^{z+1}((1+\theta)(E\xi)^2 + (\theta - 2m)E\xi + (1-\theta) \text{var } \xi) \\
 &\quad - \theta^{z+1}z(1-\theta)(E\xi)^2
 \end{aligned} \tag{4.1}$$

for every $z \in \mathbb{N}$. The function $g(z)$ is bounded. Moreover, there exist $z_g \in \mathbb{N}$ and a constant $B_g > 0$ such that, for every $z \geq z_g$, we have $|g(z)| \leq B_g/(z+1)$. Furthermore, the limit $\sum_{k=0}^\infty |g(Z_k(\omega))|$ exists and is finite a.s. on A .

Lemma 4.2. For every $n \geq 0$ and $Z_n > 0$,

$$\mathbb{E}\left[\frac{1}{(Y_n+1)^2}(Y_{n+1} - m(Y_n+1))^4 \mid Z_n\right] = 3\gamma^4 + h(Z_n) \quad \text{a.s.},$$

where $h(z)$ is a bounded function. Moreover, there exist $z_h \in \mathbb{N}$ and a constant $B_h > 0$ such that, for every $z \geq z_h$, we have $|h(z)| \leq B_h/(z+1)$.

Lemma 4.3. For every $n \geq 0$ and $Z_n > 0$,

$$\text{var}\left[\frac{1}{Y_n+1}(Y_{n+1} - m(Y_n+1))^2 \mid Z_n\right] = 2\gamma^4 + \tilde{h}(Z_n) \quad \text{a.s.},$$

where $\tilde{h}(z)$ is a bounded function. Moreover, there exist $z_{\tilde{h}} \in \mathbb{N}$ and a constant $B_{\tilde{h}} > 0$ such that, for every $z \geq z_{\tilde{h}}$, we have $|\tilde{h}(z)| \leq B_{\tilde{h}}/(z+1)$. In addition, the limit $\sum_{k=0}^\infty |\tilde{h}(Z_k(\omega))|$ exists and is finite a.s. on A .

Lemma 4.4. For every $n \in \mathbb{N}$ and $0 \leq k \leq k_n$, where $k_n = \tilde{n}_\omega$, consider the random variables

$$X_{nk} = \frac{1}{\sqrt{2k_n}\gamma^4} \left(\frac{1}{Y_{2k}+1}(Y_{2k+1} - m(Y_{2k}+1))^2 - \gamma^2 - g(Z_{2k}) \right) \mathbf{1}\{Z_{2k} > 0\}.$$

For every $\varepsilon > 0$, we have $\sum_{k=0}^{k_n} \mathbb{E}[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}] \rightarrow 0$ a.s. as $n \rightarrow \infty$.

Proof. Let us first study the limit behaviour of the conditional expectation on A^c . Since $A^c = \{Z_n \rightarrow 0\}$ a.s., for almost every $\omega \in A^c$, there exists $\tilde{k}(\omega) \geq 0$ such that $Z_{2\tilde{k}(\omega)}(\omega) > 0$ and $Z_{2k}(\omega) = 0$ for every $k > \tilde{k}(\omega)$. This implies that $X_{nk}(\omega) = 0$ for every k such that $k > \tilde{k}(\omega)$. Thus, for almost every $\omega \in A^c$ and every $n > 2\tilde{k}(\omega)$, we have

$$\begin{aligned}
 &\sum_{k=0}^{k_n} \mathbb{E}[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}](\omega) \\
 &\leq \sum_{k=0}^{k_n} \mathbb{E}[X_{nk}^2 \mid Z_{2k}](\omega) \\
 &\quad \tilde{k}(\omega) \\
 &= \sum_{k=0}^{\tilde{k}(\omega)} \mathbb{E}[X_{nk}^2 \mid Z_{2k}](\omega) \\
 &= \frac{1}{k_n} \frac{1}{2\gamma^4} \sum_{k=0}^{\tilde{k}(\omega)} \mathbb{E}\left[\left(\frac{1}{Y_{2k}+1}(Y_{2k+1} - m(Y_{2k}+1))^2 - \gamma^2 - g(Z_{2k})\right)^2 \mid Z_{2k}\right](\omega).
 \end{aligned}$$

Recall that the conditional expectation on the right-hand side is in fact a conditional variance by Lemma 4.1, and, by Lemma 4.3, this conditional variance is bounded. Thus,

$$\sum_{k=0}^{k_n} E[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}] \rightarrow 0 \quad \text{a.s. on } A^c \text{ as } n \rightarrow \infty.$$

To prove the statement on A , we fix an $\varepsilon > 0$ and show that there exists an $\tilde{n} \in \mathbb{N}$ such that, for every $n \geq \tilde{n}$ and $k \leq k_n$, and almost every $\omega \in A$,

$$E[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}](\omega) \leq \frac{1}{k_n} c_n(\omega, \varepsilon), \tag{4.2}$$

where $\lim_{n \rightarrow \infty} c_n(\omega, \varepsilon) = 0$. For every $n \geq \tilde{n}$ and almost every $\omega \in A$, we then have

$$\sum_{k=0}^{k_n} E[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}] \leq \frac{k_n + 1}{k_n} c_n(\omega, \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

To show (4.2), we first observe that $A = \{Z_n \rightarrow \infty\}$ implies that $Z_n(\omega) > 0$ for every $n \in \mathbb{N}$ and every $\omega \in A$. We can thus write

$$\begin{aligned} & E[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}] \\ &= \frac{1}{2k_n \gamma^4} E \left[\left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 - g(Z_{2k}) \right)^2 \right. \\ & \quad \left. \times \mathbf{1} \left\{ \left| \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) \right| > \varepsilon \sqrt{2k_n \gamma^4} \right\} \mid Z_{2k} \right] \end{aligned}$$

a.s. on A . Since $(a + b)^2 \leq 2(a^2 + b^2)$ for every $a, b \in \mathbb{R}$, we obtain

$$\begin{aligned} & E[X_{nk}^2 \mathbf{1}\{|X_{nk}| > \varepsilon\} \mid Z_{2k}] \\ & \leq \frac{1}{k_n \gamma^4} E \left[\frac{(Y_{2k+1} - m(Y_{2k} + 1))^4}{(Y_{2k} + 1)^2} \right. \\ & \quad \left. \times \mathbf{1} \left\{ \left| \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) \right| > \varepsilon \sqrt{2k_n \gamma^4} \right\} \mid Z_{2k} \right] \tag{4.3} \end{aligned}$$

$$\begin{aligned} & + \frac{1}{k_n \gamma^4} E \left[(\gamma^2 + g(Z_{2k}))^2 \right. \\ & \quad \left. \times \mathbf{1} \left\{ \left| \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) \right| > \varepsilon \sqrt{2k_n \gamma^4} \right\} \mid Z_{2k} \right] \tag{4.4} \end{aligned}$$

a.s. on A .

We can write (4.4) as

$$\frac{1}{k_n} \frac{(\gamma^2 + g(Z_{2k}))^2}{\gamma^4} P \left(\left| \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) \right| > \varepsilon \sqrt{2k_n \gamma^4} \mid Z_{2k} \right),$$

which by Chebyshev’s inequality and Lemma 4.3 is a.s. bounded from above on A by

$$\frac{1}{k_n} \frac{(\gamma^2 + g(Z_{2k}))^2}{\gamma^4} \frac{2\gamma^4 + \tilde{h}(Z_{2k})}{\varepsilon^2 2k_n \gamma^4} = \frac{1}{k_n^2} \frac{(\gamma^2 + g(Z_{2k}))^2 (2\gamma^4 + \tilde{h}(Z_{2k}))}{2\gamma^8 \varepsilon^2}.$$

By Lemmas 4.1 and 4.3, both $g(z)$ and $\tilde{h}(z)$ are bounded functions. Thus, the right-hand side of the above equality can be bounded by k_n^{-2} multiplied by a finite constant for every $n \in \mathbb{N}$, every $k \leq k_n$, and almost every fixed $\omega \in A$.

To complete the proof, it suffices to find an appropriate bound for (4.3). We have

$$\begin{aligned} & \mathbf{1} \left\{ \left| \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) \right| > \varepsilon \sqrt{2k_n \gamma^4} \right\} \\ &= \mathbf{1} \left\{ \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) > \varepsilon \sqrt{2k_n \gamma^4} \right\} \\ &+ \mathbf{1} \left\{ \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} - \gamma^2 - g(Z_{2k}) < -\varepsilon \sqrt{2k_n \gamma^4} \right\}. \end{aligned}$$

Thus, (4.3) is a.s. equal to

$$\begin{aligned} & \frac{1}{k_n \gamma^4} \mathbb{E} \left[\frac{(Y_{2k+1} - m(Y_{2k} + 1))^4}{(Y_{2k} + 1)^2} \right. \\ & \quad \times \mathbf{1} \left\{ \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} > \varepsilon \sqrt{2k_n \gamma^4} + \gamma^2 + g(Z_{2k}) \right\} \left. \middle| Z_{2k} \right] \end{aligned} \tag{4.5}$$

$$\begin{aligned} & + \frac{1}{k_n \gamma^4} \mathbb{E} \left[\frac{(Y_{2k+1} - m(Y_{2k} + 1))^4}{(Y_{2k} + 1)^2} \right. \\ & \quad \times \mathbf{1} \left\{ \frac{(Y_{2k+1} - m(Y_{2k} + 1))^2}{Y_{2k} + 1} < -\varepsilon \sqrt{2k_n \gamma^4} + \gamma^2 + g(Z_{2k}) \right\} \left. \middle| Z_{2k} \right]. \end{aligned} \tag{4.6}$$

By Lemma 4.1, the function $g(z)$ is bounded. It follows that there exists an $\tilde{n}_1 \in \mathbb{N}$ such that, for every $n \geq \tilde{n}_1$, every $k \leq k_n$, and every $\omega \in A$, we have $-\varepsilon \sqrt{2k_n \gamma^4} + \gamma^2 + g(Z_{2k}(\omega)) < 0$. Thus, (4.6) is 0 for every $n \geq \tilde{n}_1$, every $k \leq k_n$, and almost every $\omega \in A$. Furthermore, since the function $g(z)$ is bounded, there exists $\tilde{n}_2 \in \mathbb{N}$ such that, for every $n \geq \tilde{n}_2$, every $k \leq k_n$, and every $\omega \in A$, we have $\varepsilon \sqrt{2k_n \gamma^4} + \gamma^2 + g(Z_{2k}(\omega)) > 0$. Thus, (4.5) is equal to

$$\begin{aligned} & \frac{1}{k_n \gamma^4} \mathbb{E} \left[\frac{(Y_{2k+1} - m(Y_{2k} + 1))^4}{(Y_{2k} + 1)^2} \right. \\ & \quad \times \mathbf{1} \left\{ \frac{(Y_{2k+1} - m(Y_{2k} + 1))^4}{(Y_{2k} + 1)^2} > \left(\varepsilon \sqrt{2k_n \gamma^4} + \gamma^2 + g(Z_{2k}) \right)^2 \right\} \left. \middle| Z_{2k} \right] \end{aligned}$$

for every $n \geq \tilde{n}_2$, every $k \leq k_n$, and almost every $\omega \in A$.

Let $A_k = (Y_{2k+1} - m(Y_{2k} + 1))^4 / (Y_{2k} + 1)^2$ for every $k \geq 0$. The random variables $\{A_k\}_{k \geq 0}$ are nonnegative and $\mathbb{E}[A_k \mid Z_{2k}]$ is finite a.s. for every $k \geq 0$ by Lemma 4.2. Furthermore, by Lemma 3.5, together with Lemma 3.3 and Slutsky's theorem, we have

$$\mathcal{L} \left(\frac{Y_{2k+1} - mY_{2k}}{\sqrt{Y_{2k}}} \middle| Z_{2k} \right) \rightarrow \mathcal{N}(0, \gamma^2) \quad \text{a.s. on } A \text{ as } k \rightarrow \infty.$$

It follows that the same result holds for

$$\frac{Y_{2k+1} - m(Y_{2k} + 1)}{\sqrt{Y_{2k} + 1}} = \frac{\sqrt{Y_{2k}}}{\sqrt{Y_{2k} + 1}} \left(\frac{Y_{2k+1} - mY_{2k}}{\sqrt{Y_{2k}}} - \frac{m}{\sqrt{Y_{2k}}} \right).$$

The continuous mapping theorem yields $\mathcal{L}(A_k \mid Z_{2k}) \rightarrow (\mathbf{N}(0, \gamma^2))^4$ a.s. on A as $k \rightarrow \infty$. Thus, the conditional distributions of A_k converge to the distribution of a nonnegative random variable with a finite expectation. Furthermore, by Lemma 4.2 we have $E[A_k \mid Z_{2k}] = 3\gamma^4 + h(Z_{2k}) \rightarrow 3\gamma^4$ a.s. on A as $k \rightarrow \infty$. Since $3\gamma^4 = E(\mathbf{N}(0, \gamma^2))^4$, Theorem 3.6 of Billingsley (1999) on uniform integrability implies that $\lim_{c \rightarrow \infty} \sup_{k \geq 0} E[A_k \mathbf{1}\{A_k \geq c\} \mid Z_{2k}] = 0$ a.s. on A .

Since, by Lemma 4.1, the function $g(z)$ is bounded, for $c_n = (\varepsilon\sqrt{2k_n}\gamma^4 + \gamma^2 + g(Z_{2k}))^2$, we have $\lim_{n \rightarrow \infty} c_n = \infty$. It now suffices to observe that

$$(4.5) = \frac{1}{k_n} \frac{1}{\gamma^4} E[A_k \mathbf{1}\{A_k \geq c_n\} \mid Z_{2k}] \quad \text{a.s.}$$

for every $n \geq \tilde{n}_2$ and $k \leq k_n$. This completes the proof.

4.3. Main results

We now prove the properties of estimators of γ^2 . As the first step, we derive the asymptotic distribution of the random variable $1/\tilde{n}_o \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - m)^2$. Note that the random variable itself could be regarded as a method-of-moments-type estimator of γ^2 in view of Theorem 3.4, if m were assumed to be known.

Lemma 4.5. *We have*

$$\begin{aligned} & \sqrt{\tilde{n}_o} \left(\frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - m)^2 - \gamma^2 \right) \\ & \xrightarrow{D} \mathbf{N}(0, 2\gamma^4) \quad \text{as } n \rightarrow \infty \text{ conditionally on } A. \end{aligned}$$

Proof. We first apply Theorem 4.1 to show that

$$\frac{1}{\sqrt{\tilde{n}_o}} \sum_{k=0}^{\tilde{n}_o} \left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 - g(Z_{2k}) \right) \xrightarrow{D} \mathbf{N}(0, 2\gamma^4) \quad \text{as } n \rightarrow \infty$$

conditionally on A . From this, we will deduce the result for the quantity of interest.

Let $k_n = \tilde{n}_o = \lfloor (n - 1)/2 \rfloor$, and, for every $n \in \mathbb{N}$ and $0 \leq k \leq k_n$, let

$$\mathcal{F}_{nk} = \sigma\{Z_1, Y_1, Z_2, Y_2, \dots, Z_{2k+1}, Y_{2k+1}, Z_{2k+2}\}$$

and

$$X_{nk} = \frac{1}{\sqrt{2k_n}\gamma^4} \left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 - g(Z_{2k}) \right) \mathbf{1}\{Z_{2k} > 0\}$$

with $g(z)$ given by (4.1). Let $S_{ni} = \sum_{k=0}^i X_{nk}$ for every $n \in \mathbb{N}$ and $k \leq k_n$. Set $\mathcal{F}_{n,-1} = \sigma\{Z_0\} = \{\emptyset, \Omega\}$.

Clearly, $k_n \nearrow \infty$ as $n \rightarrow \infty$. Furthermore, X_{nk} is \mathcal{F}_{nk} -measurable for every $n \in \mathbb{N}$ and $k \leq k_n$. Moreover, for every $n \in \mathbb{N}$ and $k \leq k_n$, $E[X_{nk} \mid \mathcal{F}_{n,k-1}] = E[X_{nk} \mid Z_{2k}] = 0$ a.s. by definition on $\{\omega; Z_{2k}(\omega) = 0\}$, and by Lemma 4.1 otherwise. Finally, $E S_{ni}^2 = \sum_{k=0}^i E X_{n,k}^2 \leq \sum_{k=0}^{k_n} E X_{n,k}^2$ for every $n \in \mathbb{N}$ and $0 \leq i \leq k_n$. Using Lemma 4.3, we obtain

$$E S_{ni}^2 \leq \sum_{k=0}^{k_n} E[E[X_{nk}^2 \mid Z_{2k}]] \leq \frac{1}{k_n} \sum_{k=0}^{k_n} \left(1 + \frac{1}{2\gamma^4} E[\tilde{h}(Z_{2k}) \mathbf{1}\{Z_{2k} > 0\}] \right).$$

The expression on the right-hand side is bounded, since, by Lemma 4.3, the function $\tilde{h}(z)$ is bounded. Thus, $\{S_{ni}, \mathcal{F}_{ni}; n \in \mathbb{N}, 0 \leq i \leq k_n\}$ defined above is a zero-mean square-integrable martingale array as described in Theorem 4.1.

We now turn to the calculation of the conditional variance. By Lemma 4.3, on $\{\omega; Z_{2i}(\omega) > 0\}$ we have

$$V_{ni}^2(\omega) = \sum_{k=0}^i E[X_{nk}^2 \mid \mathcal{F}_{n,k-1}](\omega) = \frac{1}{k_n} \left(i + 1 + \frac{1}{2\gamma^4} \sum_{k=0}^i \tilde{h}(Z_{2k}(\omega)) \right) \quad \text{a.s.},$$

and on $\{\omega; \text{there exists } j(\omega) < i; Z_{2j(\omega)}(\omega) > 0, Z_{2j(\omega)+2}(\omega) = 0\}$, we have

$$V_{ni}^2(\omega) = \sum_{k=0}^i E[X_{nk}^2 \mid \mathcal{F}_{n,k-1}](\omega) = \frac{1}{k_n} \left(j(\omega) + 1 + \frac{1}{2\gamma^4} \sum_{k=0}^{j(\omega)} \tilde{h}(Z_{2k}(\omega)) \right) \quad \text{a.s.}$$

Now, $A^c = \{\omega; \text{there exists } j(\omega); Z_{2j(\omega)}(\omega) > 0, Z_{2j(\omega)+2}(\omega) = 0\}$ a.s. Thus, for almost every $\omega \in A^c$ and every $n \in \mathbb{N}$ such that $k_n \geq j(\omega)$,

$$V_{nk_n}^2(\omega) = \frac{1}{k_n} \left(j(\omega) + 1 + \frac{1}{2\gamma^4} \sum_{k=0}^{j(\omega)} \tilde{h}(Z_{2k}(\omega)) \right).$$

It follows that $\lim_{n \rightarrow \infty} V_{nk_n}^2(\omega) = 0$ a.s. on A^c .

On A , we have $Z_n(\omega) > 0$ for every $n \geq 0$. Thus,

$$V_{nk_n}^2(\omega) = \frac{k_n + 1}{k_n} + \frac{1}{2\gamma^4 k_n} \sum_{k=0}^{k_n} \tilde{h}(Z_{2k}(\omega)) \quad \text{a.s. on } A$$

for every $n \in \mathbb{N}$. Concentrating on the limit $\lim_{n \rightarrow \infty} 1/k_n \sum_{k=0}^{k_n} \tilde{h}(Z_{2k})$, we have

$$\left| \sum_{k=0}^n \tilde{h}(Z_{2k}) \right| \leq \sum_{k=0}^n |\tilde{h}(Z_{2k})| \leq \sum_{k=0}^{2n} |\tilde{h}(Z_k)| \quad \text{for every } n \geq 0.$$

The right-hand side of the above inequality has a finite limit for $n \rightarrow \infty$ a.s. on A by Lemma 4.3. It follows that $\lim_{n \rightarrow \infty} 1/k_n \sum_{k=0}^{k_n} \tilde{h}(Z_{2k}) = 0$ a.s. on A . Thus, $\lim_{n \rightarrow \infty} V_{nk_n}^2(\omega) = 1$ a.s. on A .

Combined with the result on A^c , this gives

$$\lim_{n \rightarrow \infty} V_{nk_n}^2(\omega) = \mathbf{1}\{A\} \quad \text{a.s.}$$

Summarising our findings, we have shown that $\{S_{ni}, \mathcal{F}_{ni}; n \in \mathbb{N}, 1 \leq i \leq k_n\}$ is a zero-mean square-integrable martingale array. Since the σ -fields \mathcal{F}_{ni} do not depend on n , Theorem 4.1(iii) is satisfied. We have shown that Theorem 4.1(ii) is satisfied with $\eta^2 = \mathbf{1}\{A\}$. Finally, Lemma 4.4 implies that Theorem 4.1(i) is satisfied as well. Thus, we may invoke Theorem 4.1 to deduce that $S_{nk_n} \xrightarrow{D} Z$ as $n \rightarrow \infty$, where Z is a random variable with the characteristic function $P(A) \exp\{-\frac{1}{2}t^2\} + P(A^c)$.

For the distribution functions F_n of the random variables S_{nk_n} , we have $\lim_{n \rightarrow \infty} F_n(x) = P(A)\Phi(x)$ if $x < 0$ and $\lim_{n \rightarrow \infty} F_n(x) = P(A)\Phi(x) + P(A^c)$ if $x > 0$, where $\Phi(x)$ denotes the

distribution function of a standard Gaussian random variable. Besides, since $A^c = \{Z_n \rightarrow 0\}$ a.s., we have $S_{nk_n} \rightarrow 0$ a.s. on A^c as $n \rightarrow \infty$. It follows that

$$\lim_{n \rightarrow \infty} P(S_{nk_n} \leq x \mid A) = \frac{1}{P(A)} \lim_{n \rightarrow \infty} (P(S_{nk_n} \leq x) - P(S_{nk_n} \leq x, A^c)) = \Phi(x) \tag{4.7}$$

for every $x \in \mathbb{R} \setminus \{0\}$. Now if P_n denotes the conditional distribution of S_{nk_n} given A , (4.7) guarantees that $P_n(B)$ converges to $P(N \in B)$ for every $B \in \mathcal{A}$, where N is a standard Gaussian random variable and

$$\mathcal{A} = \emptyset \cup \{(-\infty, r]; r \in \mathbb{Q} \setminus \{0\}\} \cup \{(r_1, r_2]; r_1, r_2 \in \mathbb{Q} \setminus \{0\}, r_1 < r_2\}.$$

The class \mathcal{A} is a class of convergence-determining sets by Theorem 2.3 of Billingsley (1999). Hence, $S_{nk_n} \xrightarrow{D} N(0, 1)$ as $n \rightarrow \infty$ conditionally on A .

Recall that

$$S_{nk_n} = \frac{1}{\sqrt{2k_n\gamma^4}} \sum_{k=0}^{k_n} \left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 - g(Z_{2k}) \right) \mathbf{1}\{Z_{2k} > 0\},$$

where $k_n = \tilde{n}_0$ and $g(z)$ is given by (4.1). Since $Z_n > 0$ for every $n \in \mathbb{N}$ on A , we have in fact shown that

$$\frac{1}{\sqrt{\tilde{n}_0}} \sum_{k=0}^{\tilde{n}_0} \left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 - g(Z_{2k}) \right) \xrightarrow{D} N(0, 2\gamma^4) \quad \text{as } n \rightarrow \infty$$

conditionally on A . This completes the first part of the proof.

We now use this result to derive the asymptotic distribution of

$$\begin{aligned} & \sqrt{\tilde{n}_0} \left(\frac{1}{\tilde{n}_0} \sum_{k=0}^{\tilde{n}_0} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - m)^2 - \gamma^2 \right) \\ &= \frac{1}{\sqrt{\tilde{n}_0}} \sum_{k=0}^{\tilde{n}_0} \left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 \right). \end{aligned}$$

To this end, it suffices to study the limit $\lim_{n \rightarrow \infty} 1/\sqrt{\tilde{n}_0} \sum_{k=0}^{\tilde{n}_0} g(Z_{2k})$. We have

$$\left| \sum_{k=0}^n g(Z_{2k}) \right| \leq \sum_{k=0}^n |g(Z_{2k})| \leq \sum_{k=0}^{2n} |g(Z_k)| \quad \text{for every } n \in \mathbb{N}.$$

The right-hand side of this inequality has a finite limit for $n \rightarrow \infty$ a.s. on A by Lemma 4.1. It follows that $\lim_{n \rightarrow \infty} 1/\sqrt{\tilde{n}_0} \sum_{k=0}^{\tilde{n}_0} g(Z_{2k}) = 0$ a.s. on A . Slutsky's theorem now gives

$$\frac{1}{\sqrt{\tilde{n}_0}} \sum_{k=0}^{\tilde{n}_0} \left(\frac{1}{Y_{2k} + 1} (Y_{2k+1} - m(Y_{2k} + 1))^2 - \gamma^2 \right) \xrightarrow{D} N(0, 2\gamma^4) \quad \text{as } n \rightarrow \infty$$

conditionally on A , which is equivalent to the statement of the theorem.

We now establish the asymptotic distribution of the estimator $\hat{\gamma}_{Y,n,\circ}^2$.

Theorem 4.2. *We have*

$$\sqrt{\tilde{n}_o}(\hat{\gamma}_{Y,n,o}^2 - \gamma^2) \xrightarrow{D} N(0, 2\gamma^4) \text{ as } n \rightarrow \infty \text{ conditionally on } A.$$

Proof. Recall that $\hat{\gamma}_{Y,n,o}^2 = 1/\tilde{n}_o \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,o})^2$. The statement of the theorem is thus equivalent to

$$\sqrt{\tilde{n}_o} \left(\frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,o})^2 - \gamma^2 \right) \xrightarrow{D} N(0, 2\gamma^4) \text{ as } n \rightarrow \infty$$

conditionally on A . By Lemma 4.5, the convergence holds if we replace $\hat{m}_{Y,n,o}$ in the above expression by m . We now show that the difference

$$\begin{aligned} & \sqrt{\tilde{n}_o} \left(\frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - m)^2 - \gamma^2 \right) \\ & - \sqrt{\tilde{n}_o} \left(\frac{1}{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,o})^2 - \gamma^2 \right) \end{aligned} \tag{4.8}$$

converges to 0 in distribution conditionally on A . This implies convergence to 0 in probability on A . The statement of the theorem then follows by Slutsky’s theorem.

To show that (4.8) converges to 0 in distribution conditionally on A , we write

$$\begin{aligned} (4.8) &= \frac{1}{\sqrt{\tilde{n}_o}} \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1) \left(\left(\frac{Y_{2k+1}}{Y_{2k} + 1} - m \right)^2 - \left(\frac{Y_{2k+1}}{Y_{2k} + 1} - \hat{m}_{Y,n,o} \right)^2 \right) \\ &= \frac{1}{\sqrt{\tilde{n}_o}} \left(2(\hat{m}_{Y,n,o} - m) \sum_{k=0}^{\tilde{n}_o} Y_{2k+1} - (\hat{m}_{Y,n,o}^2 - m^2) \sum_{k=0}^{\tilde{n}_o} (Y_{2k} + 1) \right) \\ &= \frac{1}{\sqrt{\tilde{n}_o}} (\hat{m}_{Y,n,o} - m) \left(2 \sum_{k=0}^{\tilde{n}_o} Y_{2k+1} - (\hat{m}_{Y,n,o} + m) \sum_{k=0}^{\tilde{n}_o} Y_{2k} \right) - \frac{\tilde{n}_o + 1}{\sqrt{\tilde{n}_o}} (\hat{m}_{Y,n,o}^2 - m^2). \end{aligned}$$

Recall that $\hat{m}_{Y,n,o} = (\sum_{k=0}^{\tilde{n}_o} Y_{2k+1}) / (\sum_{k=0}^{\tilde{n}_o} Y_{2k})$. Thus,

$$\begin{aligned} (4.8) &= \frac{1}{\sqrt{\tilde{n}_o}} (\hat{m}_{Y,n,o} - m) \sum_{k=0}^{\tilde{n}_o} Y_{2k} (2\hat{m}_{Y,n,o} - (\hat{m}_{Y,n,o} + m)) - \frac{\tilde{n}_o + 1}{\sqrt{\tilde{n}_o}} (\hat{m}_{Y,n,o}^2 - m^2) \\ &= \frac{1}{\sqrt{\tilde{n}_o}} \sum_{k=0}^{\tilde{n}_o} Y_{2k} (\hat{m}_{Y,n,o} - m)^2 - \frac{\tilde{n}_o + 1}{\sqrt{\tilde{n}_o}} (\hat{m}_{Y,n,o}^2 - m^2). \end{aligned}$$

We now show that the two summands converge to 0 in distribution conditionally on A . Theorem 3.3 together with the continuous mapping theorem implies that $1/\gamma^2 \sum_{k=0}^{\tilde{n}_o} Y_{2k} (\hat{m}_{Y,n,o} - m)^2$ converges in distribution to a χ^2 -distributed random variable conditionally on A . It follows that $1/\sqrt{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} Y_{2k} (\hat{m}_{Y,n,o} - m)^2 \xrightarrow{D} 0$ as $n \rightarrow \infty$ conditionally on A . Strong consistency of $\hat{m}_{Y,n,o}$ on A implies that $(\hat{m}_{Y,n,o} + m) \rightarrow 2m$ a.s. on A as $n \rightarrow \infty$. Furthermore, Lemma 3.4 gives

$$\frac{\tilde{n}_o + 1}{\sqrt{\tilde{n}_o} \sum_{k=0}^{\tilde{n}_o} Y_{2k}} = \frac{\tilde{n}_o + 1}{\sqrt{\tilde{n}_o} (m^{2\tilde{n}_o+2} - 1)} \sqrt{\frac{m^{2\tilde{n}_o+2} - 1}{\sum_{k=0}^{\tilde{n}_o} Y_{2k}}} \rightarrow 0 \text{ a.s. on } A \text{ as } n \rightarrow \infty,$$

since we assume that $m > 1$. Together with Theorem 3.3, this implies that

$$\frac{\tilde{n}_0 + 1}{\sqrt{\tilde{n}_0}} (\hat{m}_{Y,n,o}^2 - m^2) = \frac{\tilde{n}_0 + 1}{\sqrt{\tilde{n}_0} \sum_{k=0}^{\tilde{n}_0} Y_{2k}} \sqrt{\sum_{k=0}^{\tilde{n}_0} Y_{2k}} (\hat{m}_{Y,n,o} - m) (\hat{m}_{Y,n,o} + m) \xrightarrow{D} 0 \quad \text{as } n \rightarrow \infty \text{ conditionally on } A,$$

and the proof is complete.

It now remains to prove weak consistency of the estimator $\hat{\gamma}_{Y,n,o}^2$ on A . In fact, Theorem 4.2 allows us to show even more.

Theorem 4.3. *The estimator $\hat{\gamma}_{Y,n,o}^2$ is weakly consistent on A . Moreover, $\hat{\gamma}_{Y,n,o}^2 = o_P(a_n/n)$ for any sequence a_n such that $1/a_n \xrightarrow{D} 0$ as $n \rightarrow \infty$ on A .*

Proof. By Theorem 4.2 and Slutsky’s theorem, $1/a_n \times n/\sqrt{\tilde{n}_0} \times \sqrt{\tilde{n}_0}(\hat{\gamma}_{Y,n,o}^2 - \gamma^2) \xrightarrow{D} 0$ as $n \rightarrow \infty$ conditionally on A .

Recall that to derive the asymptotic distribution of the estimator $\hat{\gamma}_{Y,n,o}^2$, we first established the asymptotic distribution of the variable $1/\tilde{n}_0 \sum_{k=0}^{\tilde{n}_0} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - m)^2$, and then showed that the distribution does not change when we replace m in the formula by its estimator $\hat{m}_{Y,n,o}$. Alternatively, we can replace m in the formula by $\hat{m}_{Y,n,e}$ to obtain the estimator $1/\tilde{n}_0 \sum_{k=0}^{\tilde{n}_0} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,e})^2$. The asymptotic properties of the estimator can be established in the same way as for $\hat{\gamma}_{Y,n,o}^2$ (and a similar argument applies for an estimator that uses $\hat{m}_{Y,n,a}$ to estimate m).

Indeed, after simplifying the difference

$$\sum_{k=0}^{\tilde{n}_0} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - m)^2 - \sum_{k=0}^{\tilde{n}_0} (Y_{2k} + 1)(\bar{m}_{Y,2k+1} - \hat{m}_{Y,n,e})^2$$

as we did in the proof of Theorem 4.2 with the difference involving $\hat{m}_{Y,n,o}$, and after dividing by $\sqrt{\tilde{n}_0}$, we arrive at

$$\begin{aligned} & \frac{2}{\sqrt{\tilde{n}_0}} \sum_{k=0}^{\tilde{n}_0} Y_{2k} (\hat{m}_{Y,n,e} - m) (\hat{m}_{Y,n,o} - m) \\ & - \frac{1}{\sqrt{\tilde{n}_0}} \sum_{k=0}^{\tilde{n}_0} Y_{2k} (\hat{m}_{Y,n,e} - m)^2 - \frac{\tilde{n}_0 + 1}{\sqrt{\tilde{n}_0}} (\hat{m}_{Y,n,e}^2 - m^2). \end{aligned} \tag{4.9}$$

The first term of (4.9) can be written as

$$\frac{2}{\sqrt{\tilde{n}_0}} \sqrt{\frac{\sum_{k=0}^{\tilde{n}_0} Y_{2k}}{\sum_{k=1}^{\tilde{n}_e} Y_{2k-1}}} \sqrt{\sum_{k=1}^{\tilde{n}_e} Y_{2k-1}} (\hat{m}_{Y,n,e} - m) \sqrt{\sum_{k=0}^{\tilde{n}_0} Y_{2k}} (\hat{m}_{Y,n,o} - m).$$

Its convergence to 0 in probability on A follows by Theorem 3.3 and the corresponding result for the estimator $\hat{m}_{Y,n,e}$, and because $\sum_{k=0}^{\tilde{n}_0} Y_{2k} / \sum_{k=1}^{\tilde{n}_e} Y_{2k-1}$ is bounded in probability by Lemma 3.4. The remaining terms of (4.9) can be shown to converge to 0 in probability on A by using the boundedness of $\sum_{k=0}^{\tilde{n}_0} Y_{2k} / \sum_{k=1}^{\tilde{n}_e} Y_{2k-1}$ and otherwise proceeding as before. Thus, the difference in question is $o_P(\sqrt{\tilde{n}_0})$ on A . As we have already discussed, this implies both asymptotic Gaussianity and weak consistency on A .

5. Simulation results

We have seen that we can choose from several estimators to estimate the parameters m and γ^2 , and can also construct asymptotic confidence intervals based on various combinations of these estimators. In this section we study how they may be expected to perform in practice. A more detailed study can be found in Kvitkovičová and Panaretos (2010). We simulated realisations of the process $\{(Z_k, Y_k)^\top; k \geq 0\}$ under the two scenarios described in Table 1. Geometric and Poisson offspring distributions are chosen; the former is related to a linear birth-and-death process, while the latter was adopted in the applications of Meester *et al.* (2002) and Meester and Trapman (2006). We consider a slowly spreading infection observed for a long period and a quickly spreading infection observed for a short period. For each setting, the number of realisations is chosen so that at the end of the observation period, at least 10 000 of them are nonextinct. The simulations were carried out in the R statistical computing environment, version 2.12.0.

The performance of the mean estimators is summarised in the upper half of Table 2. Between the estimators $\hat{m}_{Y,n,o}$ and $\hat{m}_{Y,n,e}$, we report the results for that which includes the last observation

TABLE 1: Simulation settings. The offspring distributions (Ξ and Υ), the detection probability (θ), the length of the observation period (n), the number of the realisations such that $Z_n > 0$ (N), and the number of all simulated realisations (in parentheses).

Setting	Ξ	Υ	θ	m	γ^2	n	N
1	$\text{Ge}(\frac{2}{5})$	$\text{Ge}(\frac{4}{5})$	$\frac{2}{3}$	1.0833	3.0903	61	10 253 (170 000)
2	$\text{Po}(\frac{3}{2})$	$\text{Po}(\frac{4}{5})$	$\frac{1}{2}$	1.1500	2.2750	30	10 205 (44 000)

TABLE 2: Performance of the estimators. The first three quartiles and the mean of the difference $\hat{\theta} - \theta$, the standard deviation (SD), the mean squared error (MSE), and the mean absolute deviation (MAD) of the estimator.

Setting	Estimator	First quartile	Median	Mean	Third quartile	SD	MSE	MAD
1	$\hat{m}_{Y,n,o}$	-0.0221	-0.0002	-0.0029	0.0187	0.0436	0.0019	0.0204
1	$\hat{m}_{Y,n}$	-0.0111	-0.0007	-0.0039	0.0070	0.0186	0.0004	0.0086
1	$\hat{m}_{Y,n,a}$	-0.0103	-0.0002	-0.0031	0.0074	0.0182	0.0003	0.0086
2	$\hat{m}_{Y,n,e}$	-0.0546	-0.0012	-0.0033	0.0506	0.1032	0.0107	0.0524
2	$\hat{m}_{Y,n}$	-0.0262	-0.0013	-0.0071	0.0184	0.0413	0.0018	0.0216
2	$\hat{m}_{Y,n,a}$	-0.0234	0.0007	-0.0040	0.0206	0.0410	0.0017	0.0218
1	$\hat{\gamma}_{Y,n}^2$	-0.5759	-0.1596	-0.0777	0.3261	0.7159	0.5185	0.4813
1	$\hat{\gamma}_{Y,n,a}^2$	-0.3708	0.1131	0.2300	0.7063	0.8691	0.8081	0.5159
1	$\hat{\gamma}_{Y,n,aa}^2$	-0.5316	-0.1095	-0.0255	0.3850	0.7281	0.5308	0.4761
2	$\hat{\gamma}_{Y,n}^2$	-0.7008	-0.3210	-0.2013	0.1733	0.7172	0.5548	0.5302
2	$\hat{\gamma}_{Y,n,a}^2$	-0.4823	0.0167	0.2434	0.7010	1.0758	1.2165	0.5683
2	$\hat{\gamma}_{Y,n,aa}^2$	-0.6820	-0.2858	-0.1606	0.2249	0.7436	0.5786	0.5273

TABLE 3: Performance of the 95% asymptotic confidence intervals. Coverage, the first three quartiles, the mean, and the standard deviation of their length.

Setting	Based on	Coverage	First quartile	Median	Mean	Third quartile	SD
1	$\hat{m}_{Y,n,o}, \hat{\gamma}_{Y,n}^2$	0.9486	0.0860	0.1195	0.1452	0.1786	0.0860
1	$\hat{m}_{Y,n,o}, \hat{\gamma}_{Y,n,a}^2$	0.9799	0.0896	0.1250	0.1524	0.1876	0.0911
1	$\hat{m}_{Y,n,o}, \hat{\gamma}_{Y,n,aa}^2$	0.9509	0.0868	0.1205	0.1465	0.1804	0.0869
2	$\hat{m}_{Y,n,e}, \hat{\gamma}_{Y,n}^2$	0.9461	0.2125	0.2909	0.3365	0.4166	0.1713
2	$\hat{m}_{Y,n,e}, \hat{\gamma}_{Y,n,a}^2$	0.9940	0.2300	0.3162	0.3699	0.4560	0.2001
2	$\hat{m}_{Y,n,e}, \hat{\gamma}_{Y,n,aa}^2$	0.9495	0.2140	0.2935	0.3399	0.4208	0.1744

(that is, $\hat{m}_{Y,n,o}$ if the sample size is odd, and $\hat{m}_{Y,n,e}$ if the sample size is even). The results are rather promising, showing both low bias and variance. The best performance is exhibited by the estimators $\hat{m}_{Y,n}$ and $\hat{m}_{Y,n,a}$.

The lower half of Table 2 deals with the estimators $\hat{\gamma}_{Y,n}^2$, $\hat{\gamma}_{Y,n,a}^2$, and $\hat{\gamma}_{Y,n,aa}^2$. They are roughly comparable both in terms of bias and variance, and appear to perform better than all the other estimators suggested above ($\hat{\gamma}_{Y,n,o}^2$, $\hat{\gamma}_{Y,n,e}^2$, and their variants). The estimator $\hat{\gamma}_{Y,n,a}^2$ shows a slight tendency to overestimate the true value. This effect may actually follow from the construction. The estimator is based on squared differences between mean estimators, with each estimator in a pair motivated by a different part of the data (recall the discussion in Section 2). Indeed, the estimator $1/2(\hat{\gamma}_{Y,n,o}^2 + \hat{\gamma}_{Y,n,e}^2)$, comparing mean estimators based on the same part of the data, shows a tendency to underestimate the true value (results not shown here). The same tendency is exhibited by the estimator $\hat{\gamma}_{Y,n,aa}^2$ and even more noticeably by $\hat{\gamma}_{Y,n}^2$. Although these estimators often have a larger bias than the estimator $\hat{\gamma}_{Y,n,a}^2$, they have a smaller variance, which gives them an advantage in terms of the median absolute deviation. We prefer this characteristic to the mean squared error here, since the distributions of the estimators seem to have heavy right tails.

Table 3 summarises the performance of 95% asymptotic confidence intervals based on different combinations of point estimators. Overall, they have both reasonable coverage and length, showing the ability to detect even small departures from 1. Note that the latter is an important property, since the construction assumes that the mean is greater than 1 and we are mostly interested in small departures from this threshold. The choice of the estimator of γ^2 is crucial for the performance of the intervals. The intervals based on $\hat{\gamma}_{Y,n,a}^2$ have a higher coverage and are on average longer than those based on the other estimators.

6. Conclusions

We have considered the problem of estimation in a partially observed branching process, previously studied in Meester and Trapman (2006). Similarly to their work, we focused on estimators that can be explicitly defined and mimic those employed when a branching process is fully observed. However, whereas Meester and Trapman (2006) pursued the construction of consistent estimators of parameters related to the first three moments of the offspring distribution, we focused on the construction of asymptotically Gaussian estimators of parameters related to the first two moments.

Motivated by the dependence structure of the observed process, we proposed estimators that are consistent and asymptotically Gaussian, and allow for the construction of asymptotic confidence intervals, all valid conditionally on explosion. We also studied several estimators derived from those proposed originally, which improve properties of the confidence intervals. A simulation study carried out to evaluate the performance of our estimators showed favourable results.

Appendix A. Proofs of the auxiliary lemmas

Proof of Lemma 3.1. Since we assume that $Z_0 = 1$, we have $E Z_0 = 1$. For every $n \in \mathbb{N}$, we then have $E Z_n = E[E[Z_n | Z_{n-1}]] = m E Z_{n-1} = m^n$.

Proof of Lemma 3.3. To prove consistency of $\bar{m}_{Y,n}$, Meester and Trapman (2006) considered a random sequence $\{M_n\}_{n \in \mathbb{N}}$ such that $M_n = \prod_{k=1}^n (Y_k + 1) / ((1 - \theta)Z_k + 1)$. They showed that $M_n \rightarrow M$ a.s. as $n \rightarrow \infty$, where M is a random variable a.s. finite and positive on A . We can use their result to show that $(Y_n + 1) / ((1 - \theta)Z_n + 1) = M_n / M_{n-1} \rightarrow 1$ a.s. on A as $n \rightarrow \infty$. Since $Z_n \rightarrow \infty$ a.s. on A , it follows that $Y_n / ((1 - \theta)Z_n + 1) \rightarrow 1$ a.s. on A as $n \rightarrow \infty$, and finally that $Y_n / ((1 - \theta)Z_n) \rightarrow 1$ a.s. on A as $n \rightarrow \infty$.

Proof of Lemma 3.4. We apply the Toeplitz lemma with $x_n = Z_{2n} / m^{2n}$ and $a_k = m^{2k}$ to show (i), $x_n = Z_{2n+1} / m^{2n+1}$ and $a_k = m^{2k+1}$ to show (ii), $x_n = Y_{2n} / m^{2n}$ and $a_k = m^{2k}$ to show (iii), and with $x_n = Y_{2n+1} / m^{2n+1}$ and $a_k = m^{2k+1}$ to show (iv). Convergence of x_n on the explosion set follows by Lemma 3.3 in the two latter cases.

The proofs of the remaining statements are essentially calculations exploiting conditional distributions in the process $\{(Z_k, Y_k)^T; k \geq 0\}$, properties of the conditional expectation, and properties of the binomial distribution. We now indicate the main steps of the calculations; details can be found in Kvitkovičová and Panaretos (2010).

Proof of Lemma 3.2. (i) We write $Y_{n+1} - mY_n$ as $(Y_{n+1} - (1 - \theta)Z_{n+1}) + (1 - \theta)(Z_{n+1} - mZ_n) + m((1 - \theta)Z_n - Y_n)$. Then

$$\begin{aligned} E[Y_{n+1} - mY_n | Z_n] &= E[Y_{n+1} - (1 - \theta)Z_{n+1} | Z_n] + (1 - \theta)E[Z_{n+1} - mZ_n | Z_n] \\ &\quad + m E[(1 - \theta)Z_n - Y_n | Z_n] \\ &= E[E[Y_{n+1} - (1 - \theta)Z_{n+1} | Z_{n+1}, Z_n] | Z_n] + 0 + 0 \\ &= 0 \quad \text{a.s.} \end{aligned}$$

(ii) We use the same expansion of $Y_{n+1} - mY_n$ as above, and apply the multinomial theorem to obtain a sum of conditional expectations. We then treat each of them separately.

Proof of Lemma 4.1. We write $Y_{n+1} - m(Y_n + 1)$ as $(Y_{n+1} - (1 - \theta)Z_{n+1}) + (1 - \theta)(Z_{n+1} - E[Z_{n+1} | Y_n, Z_n]) + ((1 - \theta)E[Z_{n+1} | Y_n, Z_n] - m(Y_n + 1))$. We again apply the multinomial theorem to write $E[(Y_{n+1} - m(Y_n + 1))^2 / (Y_n + 1) | Z_n]$ as a sum of conditional expectations, which can then be treated separately.

The function $g(z)$ is such that there exist constants c_i such that, for every $z \in \mathbb{N}$, $|g(z)| \leq 1 / (z + 1)(c_1 + \theta^z c_2 + (z + 1)\theta^z c_3 + z(z + 1)\theta^z c_4)$. Since $\lim_{z \rightarrow \infty} z^{k_1} \theta^z = 0$ for every $k_1 \geq 0$, we can choose any $B_g > c_1$ and find z_g so that $c_1 + \theta^z c_2 + (z + 1)\theta^z c_3 + z(z + 1)\theta^z c_4 \leq B_g$ for every $z \geq z_g$. Boundedness now follows.

Finally, since $\{Z_k; k \geq 0\}$ is a branching process, we have $A = \{\omega; Z_n(\omega) / m^n \rightarrow W(\omega) > 0 \text{ as } n \rightarrow \infty\}$ a.s. For every ω in the latter set, fix $\varepsilon(\omega) > 0$ such that, $W(\omega) - \varepsilon(\omega) > 0$.

There exists $\tilde{n}(\omega) \in \mathbb{N}$ such that, for every $n \geq \tilde{n}(\omega)$, we have $Z_n(\omega) \geq z_g$, and, moreover, $|Z_n(\omega)/m^n - W(\omega)| < \varepsilon(\omega)$. It follows that $0 < 1/(Z_n(\omega) + 1) < 1/(m^n(W(\omega) - \varepsilon(\omega)) + 1) < 1/(m^n(W(\omega) - \varepsilon(\omega)))$ for every $n \geq \tilde{n}(\omega)$, and so $|g(Z_n(\omega))| \leq B_g/(Z_n(\omega) + 1) < B_g/(m^n(W(\omega) - \varepsilon(\omega)))$ for every $n \geq \tilde{n}(\omega)$.

Proof of Lemma 4.2. We use the same expansion of $Y_{n+1} - m(Y_n + 1)$ as in the proof of Lemma 4.1 and the multinomial theorem to write $E[(Y_{n+1} - m(Y_n + 1))^4/(Y_n + 1)^2 \mid Z_n]$ as a sum of conditional expectations. We express each of them as a function of $E[1/(Y_n + 1)^2 \mid Z_n]$. This expectation cannot be derived explicitly. However, it corresponds to the expectation of the reciprocal of a positive binomial random variable. To approximate the latter, Stephan (1945) used the fact that

$$\frac{1}{x} = \sum_{i=1}^t \frac{(i-1)! x!}{(x+i)!} + \frac{t!(x-1)!}{(x+t)!}$$

for $x > 0$ and $t \geq 1$. This expansion with $t = 3$ allows us to identify the terms of order at most $1/(Z_n + 1)$ in the expression for $E[(Y_{n+1} - m(Y_n + 1))^4/(Y_n + 1)^2 \mid Z_n]$. By studying these and the remaining terms separately, we eventually obtain the statement of the lemma.

Proof of Lemma 4.3. The statements comprising the lemma follow from Lemmas 4.1 and 4.2; the last one can be shown in the same way as for the function $g(z)$.

Acknowledgements

This research was supported by a Swiss National Science Foundation grant. The authors would like to thank the anonymous referee for a careful reading of the text and for useful suggestions that helped in simplifying the exposition of the results.

References

- BALL, F. AND DONNELLY, P. (1995). Strong approximations for epidemic models. *Stoch. Process. Appl.* **55**, 1–21.
- BECKER, N. G. AND HASOFER, A. M. (1997). Estimation in epidemics with incomplete observations. *J. R. Statist. Soc. B* **59**, 415–429.
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
- DUBY, C. AND ROUAULT, A. (1982). Estimation non paramétrique de l'espérance et de la variance de la loi de reproduction d'un processus de ramification. *Ann. Inst. H. Poincaré B* **18**, 149–163.
- GUTTORP, P. (1991). *Statistical Inference for Branching Processes*. John Wiley, New York.
- HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
- KENDALL, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In *Proc. 3rd Berkeley Symp. on Mathematical Statistics and Probability* Vol. IV, University of California Press, Berkeley, pp. 149–165.
- KVITKOVIČOVÁ, A. AND PANARETOS, V. M. (2010). Asymptotic inference for partially observed branching processes. Technical Report #01/10, Chair of Mathematical Statistics, EPFL.
- MEESTER, R. AND TRAPMAN, P. (2006). Estimation in branching processes with restricted observations. *Adv. Appl. Prob.* **38**, 1098–1115.
- MEESTER, R., DE KONING, J., DE JONG, M. C. M. AND DIEKMANN, O. (2002). Modeling and real-time prediction of classical swine fever epidemics. *Biometrics* **58**, 178–184.
- PANARETOS, V. M. (2007). Partially observed branching processes for stochastic epidemics. *J. Math. Biol.* **54**, 645–668.
- SCOTT, D. J. (1978). A central limit theorem for martingales and an application to branching processes. *Stoch. Process. Appl.* **6**, 241–252.
- STEPHAN, F. F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Statist.* **16**, 50–61.