J. Appl. Prob. **54,** 82–95 (2017)
*doi:*10.1017/jpr.2016.88
© *Applied Probability Trust* 2017

STATISTICAL INFERENCE FOR PARTIALLY OBSERVED BRANCHING PROCESSES WITH IMMIGRATION

IBRAHIM RAHIMOV,* *Zayed University*

Abstract

In the paper we consider the following modification of a discrete-time branching process with stationary immigration. In each generation a binomially distributed subset of the population will be observed. The number of observed individuals constitute a partially observed branching process. After inspection both observed and unobserved individuals may change their offspring distributions. In the subcritical case we investigate the possibility of using the known estimators for the offspring mean and for the mean of the stationary-limiting distribution of the process when the observation of the population sizes is restricted. We prove that, if both the population and the number of immigrants are partially observed, the estimators are still strongly consistent. We also prove that the 'skipped' version of the estimator for the offspring mean is asymptotically normal and the estimator of the stationary distribution's mean is asymptotically normal under additional assumptions.

Keywords: Branching process; subcritical; immigration; restricted observation; offspring mean; stationary distribution; random sum; limit theorem

2010 Mathematics Subject Classification: Primary 60J80
Secondary 62F12; 60G99

1. Introduction

We consider a discrete-time branching stochastic process W_n, $n \geq 0$, $W_0 = 1$, defined by two families of independent, nonnegative integer-valued random variables $\{X_{ni}, (n,i) \in \mathcal{N}_0 \times \mathcal{N}\}$, $\mathcal{N} = \{1, 2, \dots\}$, $\mathcal{N}_0 = \mathcal{N} \cup \{0\}$, and $\{\nu_k,\ k \geq 1\}$ recursively as

$$W_{n+1} = \sum_{i=1}^{W_n} X_{ni} + \nu_{n+1}, \qquad n \geq 0. \tag{1}$$

Assume that X_{ni} have a common distribution for all n and i, and families $\{X_{ni}\}$ and $\{\nu_n\}$ are independent. We also assume that $\{\nu_k,\ k \geq 1\}$ are independent and identically distributed. Variables X_{ni} will be interpreted as the number of offspring of the ith individual in the nth generation and ν_n is the number of immigrating individuals in the nth generation. Then, W_n can be considered as the size of the nth generation of the population.

Suppose that $m = \mathbb{E}X_{ni}$ is the mean number of offspring of a single individual. Process W_n is called *subcritical*, *critical*, or *supercritical* depending on $m < 1$, $m = 1$, or $m > 1$, respectively. The independence assumption for the families $\{X_{ni}\}$ and $\{\nu_n\}$ means that reproduction and immigration processes are independent.

Received 29 September 2015; revision received 16 May 2016.
* Postal address: Department of Mathematics and Statistics, Zayed University, Box 19282, Dubai, UAE.
Email address: ibrahim.rahimov@zu.ac.ae

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

Since the behavior of the branching process is mostly determined by the offspring mean m, the problem of estimation of this parameter is important. Therefore, estimation of the offspring and immigration parameters in the branching process with immigration has been an active area of the research for a long time (see, for example, Yanev (2008)). As a result of this activity, it has been established that a maximum likelihood approach leads to useful results, if the number of immigrating individuals $\{\nu_k,\ k \geq 1\}$ and all offspring sizes $\{X_{ni},\ n, i \geq 1\}$ are observable. Later, it turned out that the offspring mean can successfully be estimated, if population sizes $\{W_i,\ 1 \leq i \leq n\}$ up to some generation are observed. However, in applications it is very often the case that not all individuals existing in the population are observed. As a result, the question about the possibility of estimating the offspring mean based on partial observations is of interest.

The partially observed branching process is defined as follows. Each of W_n individuals existing at time n, independently from other individuals, can be detected (observed) with probability $\theta, 0 < \theta < 1$ or remain undetected with probability $1-\theta$. Assume that the offspring distributions of both detected and undetected individuals may change after inspection. This assumption includes the situation when the inspection 'disturbs' the life of the individuals which leads to the change in the offspring distribution even when they are not detected. The possible change in the offspring distribution of detected individuals is naturally due to isolation of the individuals after detection.

Let $\{\xi_{ni}, (n,i) \in \mathcal{N}_0 \times \mathcal{N}\}$ be a family of independent and identically distributed (i.i.d.) Bernoulli random variables with a probability of success θ and $\{X_{ni}^{(j)},\ (n,i) \in \mathcal{N}_0 \times \mathcal{N}\}$, $j = 1, 2$, two independent families of i.i.d. random variables taking nonnegative integer values. These families may follow different probability distributions for $j = 1$ and 2. Assume also that families $\{\xi_{ni},\ (n,i) \in \mathcal{N}_0 \times \mathcal{N}\}$ and $\{X_{ni}^{(j)},\ (n,i) \in \mathcal{N}_0 \times \mathcal{N}\}$ are independent for all values of n, i, and j. If we take

$$X_{ni} = X_{ni}^{(1)}(1 - \xi_{ni}) + X_{ni}^{(2)}\xi_{ni} \tag{2}$$

in (1), we obtain a new branching process with immigration

$$Z_0 = 1, \qquad Z_{n+1} = \sum_{i=1}^{Z_n} X_{ni}^{(1)}(1 - \xi_{ni}) + \sum_{i=1}^{Z_n} X_{ni}^{(2)}\xi_{ni} + \nu_{n+1}, \qquad n \geq 0.$$

The partially observed branching process with immigration is now defined as

$$Y_{n+1} = \sum_{i=1}^{Z_n} \xi_{ni}, \qquad n \geq 0.$$

The partially observed (or restricted) branching processes have been considered by many authors mostly in the framework of applications in epidemic modeling. It is known that susceptible→infectious→removed (SIR) epidemic models can successfully be approximated by branching processes, when the initial number of susceptible individuals is large. In Andersson and Britton (2000), a systematic study of SIR models based on the branching approximation was provided.

Problems related to statistical inference based on partially observed branching processes without immigration have been considered in Meester *et al.* (2002), Meester and Trapman (2006), Panaretos (2007), and Kvitkovičová and Panaretos (2011). In particular, in Meester and Trapman (2006), it was demonstrated that, if one has a binomially distributed subset of

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

observations of each generation, one will be able to estimate three functions of parameters of the offspring distribution. We also note that the authors of these papers concentrate on the possibility of using traditional Lotka–Nagaev and Harris-type ratio estimators to estimate the offspring mean when the observations are restricted.

In this paper we consider the problem of estimating the parameters of the process with immigration based on partial observations of the branching process. We concentrate on the subcritical case. It is well known that when $m < 1$, the process W_n defined in (1) has a unique stationary probability distribution, if and only if $\mathbb{E}(\log^+ \nu_n)$ is finite (see Pakes (1971)). If, in addition, $\mathbb{E}(\nu_n) < \infty$ then the stationary distribution has a finite mean $\mu = \mathbb{E}(\nu_n)/(1-m)$. Pakes (1971) proposed a strongly consistent and asymptotically normal estimator for μ using his results about the total progeny of the process. Later, it was shown in Nanthi (1979) (see also Venkataraman and Nanthi (1982)) that, if there is an observation of population sizes $\{Z_i,\ i = 1, 2, \ldots, n\}$ and the number of immigrating individuals $\{\nu_i,\ i = 1, 2, \ldots, n\}$, one can construct the maximum likelihood estimator for the offspring mean m, which is asymptotically normal when $n \to \infty$. Sriram (1991) established that the maximum likelihood estimator proposed in Nanthi (1979) is strongly consistent uniformly over a class of offspring distributions with $m \leq 1$ and bounded variances. We use his result in our proofs (see Proposition 2). The question which we address in this paper is simple: will the estimators proposed by Pakes (1971) and Nanthi (1979) remain consistent, if we have only partially observed population sizes and the numbers of immigrating individuals? Another question of interest concerns the rate of convergence of the estimators to the true value of the parameters and asymptotic normality of the estimators. The difficulty in the investigating asymptotic normality is caused by the complexity of the dependency structure of the partially observed process $\{Y_n,\ n \geq 1\}$. We also note that in processes without immigration it was shown that, to obtain asymptotically normal estimators one needs to modify the classic estimators using a so-called 'skipping' method (see Meester and Trapman (2006) and Kvitkovičová and Panaretos (2011)).

It turns out that in the process with immigration, the estimators proposed by Pakes and Nanthi remain strongly consistent estimators for the mean of the stationary distribution (when θ is known) and for the offspring mean. Furthermore, we prove that the difference between the estimators and true values of parameters normed by $n^{1/2}$ in distribution converges to normal random variables with mean zero and finite variances. We note that the asymptotic normality of the Pakes's estimator we establish in the additional assumption $X_{ni}^{(1)} = X_{ni}^{(2)} = X_{ni}$ means that the offspring distributions of the observed and unobserved individuals are the same. To avoid the complexity of the dependency in the estimator of the offspring mean, we also modify the estimator using the 'skipping' method and prove that the modified estimators are asymptotically normal. Unlike the proofs in Meester and Trapman (2006) and Kvitkovičová and Panaretos (2011), in the proof of asymptotic normality of the offspring mean estimator, we use a central limit theorem (CLT) for random sums of independent variables (see Billingsley (1999, p. 152)) and do not rely on the CLT in Scott (1978). The proof of the asymptotic normality of the estimator for $\theta\mu$ is based on a representation of the total progeny of the partially observed process with immigration in terms of processes without immigration generated by the immigrants and on a CLT for a scheme of a series of independent random variables, Petrov (1995).

In Section 2 we introduce the estimators and formulate the main results. In Section 3 we provide two preliminary propositions and the proofs of the main results of the paper.

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

2. Main results and discussion

It is not difficult to see that the new process Z_n is a standard branching process with the same immigration probability generating function as (1) and with offspring generating function

$$f(s) = (1-\theta) f_1(s) + \theta f_2(s), \tag{3}$$

where $f_1(s) = \mathbb{E}s^{X_{11}^{(1)}}$ and $f_2(s) = \mathbb{E}s^{X_{11}^{(2)}}$ are offspring probability generating functions of branching processes started by detected and undetected individuals, respectively. From here we easily find that the offspring mean $m =: \mathbb{E}X_{ni}$ and the variance $\sigma^2 =: \operatorname{var} X_{ni}$ are defined as

$$m = (1-\theta)\mu_1 + \theta\mu_2, \qquad \sigma^2 = (1-\theta)\sigma_1^2 + \theta\sigma_2^2 + (1-\theta)\theta(\mu_1-\mu_2)^2,$$

where $\mu_i = \mathbb{E}X_{11}^{(i)}$ and $\sigma_i^2 = \operatorname{var} X_{11}^{(i)}$ for $i = 1, 2$. We assume throughout the paper that the immigration mean $\lambda =: \mathbb{E}\nu_i$ and the variance $\gamma^2 =: \operatorname{var} \nu_i$ are finite.

We also denote $p_j^{(i)} = \mathbb{P}\{X_{11}^{(i)} = j\}$, $j = 0, 1, \ldots$, $i = 1, 2$. As in Pakes (1971) and Nanthi (1979), we assume throughout

$$p_0^{(i)} + p_1^{(i)} < 1, \qquad p_j^{(i)} \neq 1,\ j = 0, 1, \ldots$$

for $i = 1, 2$, just to exclude the trivial situations. Throughout the paper, $N(a, b^2)$ denotes a normal random variable with mean a and variance b^2. We use '$\xrightarrow{\mathrm{D}}$' for convergence in distribution and $X \overset{\mathrm{D}}{=} Y$ to mean that X and Y follow the same distribution.

As already mentioned, for the fully observed process Pakes (1971) has shown that

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=0}^{n} W_i$$

is a strongly consistent estimator for the mean of the stationary distribution $\mu = \lambda(1-m)^{-1}$. Under the conditions that $m < 1$ and both σ^2 and γ^2 are finite, the same author established that $\sqrt{n}(\hat{\mu}_n - \mu)$ as $n \to \infty$ converges in distribution to a normal random variable with mean 0 and variance $(\mu\sigma^2 + \gamma^2)/(1-m)^2$.

With regard to the offspring mean, in the case when the population size $\{W_i,\ i = 1, 2, \ldots, n\}$ and the number of immigrating individuals $\{\nu_i,\ i = 1, 2, \ldots, n\}$ are observable, it was shown in Nanthi (1979) (see also Venkataraman and Nanthi (1982)) that

$$\hat{m}_n = \frac{\sum_{i=1}^{n}(W_i - \nu_i)}{\sum_{i=1}^{n} W_{i-1}}$$

is a maximum likelihood estimator for m and $\sqrt{n}(\hat{m}_n - m)$ as $n \to \infty$ converges in distribution to a normal random variable mean 0 and variance σ^2/μ.

To estimate the mean of the limiting-stationary distribution, we assume that only $\{Y_i,\ 1 \leq i \leq n\}$ are observed. To estimate m based on a partial observation, we also need η_n, the number of observed individuals immigrating in the $(n-1)$th generation. Without restriction of generality, we define η_n as

$$\eta_{n+1} = \sum_{j=1}^{\nu_n} \xi_{nj}, \qquad n \geq 1,$$

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

and assume that $\{(Y_i,\ \eta_i), 1 \le i \le n\}$ are observed. We consider the following estimators for $\theta\mu$ and m:

$$\hat{t}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad \hat{a}_n = \frac{\sum_{i=2}^{n+1}(Y_i - \eta_i)}{\sum_{i=1}^{n} Y_i}$$

based on the partial observations of the reproduction and the immigration processes. As already mentioned, we also consider modified estimators for m defined as

$$\hat{b}_n = \frac{\sum_{i=1}^{n_o}(Y_{2i+1} - \eta_{2i+1})}{\sum_{i=1}^{n_o} Y_{2i}}, \qquad \hat{c}_n = \frac{\sum_{i=1}^{n_e}(Y_{2i} - \eta_{2i})}{\sum_{i=1}^{n_e} Y_{2i-1}},$$

where $n_o = [(n-1)/2]$ and $n_e = [n/2]$.

First we provide the result related to the estimator of the mean of the stationary distribution of the process.

Theorem 1. *Let $m < 1$ and $\lambda < \infty$.*

(i) *Then $\hat{t}_n$ is a strongly consistent estimator for $\theta\mu$, i.e. $\hat{t}_n \to \theta\mu$ almost surely as $n \to \infty$.*

(ii) *Assume that $X_{ni}^{(1)} = X_{ni}^{(2)}$ for all $(n,i) \in \mathcal{N}_0 \times \mathcal{N}$. If, in addition, $\sigma^2 < \infty$ and $\gamma^2 < \infty$ then $\sqrt{n}(\hat{t}_n - \theta\mu) \xrightarrow{\mathrm{D}} N(0, a^2)$ as $n \to \infty$, where*

$$a^2 = \frac{\theta^2}{(1-m)^2}(\mu\sigma^2 + \gamma^2) + \theta(1-\theta)\mu. \tag{4}$$

We now provide the main result of the paper related to the estimator of the offspring mean m.

Theorem 2. *Let $m < 1, \sigma^2 < \infty$ and $\lambda < \infty$.*

(i) *Then $\hat{a}_n$ is a strongly consistent estimator for m.*

(ii) *If, in addition, $\gamma^2 < \infty$ then $\sqrt{n}(\hat{b}_n - m) \xrightarrow{\mathrm{D}} N(0, 2b^2/\mu\theta)$ as $n \to \infty$, where*

$$b^2 = m(1-\theta) + m^2(1+\theta) + \sigma^2\theta - 2m\theta\mu_2. \tag{5}$$

In Section 3 in the proof of Theorem 2 we will see that Theorem 2(ii) also holds for the estimator $\hat{c}_n$ constructed by the 'complementary skipping scheme'. Let $\theta = 1$ in Theorem 2, which means that each individual in the process is observed with probability 1. In this case $m = \mu_2, \sigma^2 = \sigma_2^2$, and, therefore, $b^2 = \sigma^2$. Thus, in this case the variance of the limiting normal distribution is equal to the doubled variance in the Nanthi theorem. This is the result of the 'skipping' and the fact that the norming factor is $\sqrt{n}$. It is easy to see from the proof that the limiting variance would be $b^2/\mu\theta$, if we use a random norming instead of $\sqrt{n}$. If $\theta = 1$ in Theorem 1 then $a^2 = (\mu\sigma^2 + \gamma^2)/(1-m)^2$ and Theorem 1 coincides with Theorem 3 in Pakes (1971).

3. Proofs of theorems

We start with a simple but important result, which was first proved in Meester and Trapman (2006) for the process without immigration.

Proposition 1. *If $m < 1$ and $\lambda \in (0, \infty)$ then almost surely (a.s.)*

$$\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} Z_{i-1}} \to \theta \quad \text{as } n \to \infty. \tag{6}$$

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

Note that in Meester and Trapman (2006) the authors have shown (6) on the explosion set $\{\lim_{n\to\infty} X_n = \infty\}$, where X_n is the process without immigration. One can see that, if we repeat our proof for the process without immigration, it is possible to show that their result holds on the wider set $\{\lim_{n\to\infty} \sum_{i=1}^{n} X_i = \infty\}$.

Proof of Proposition 1. We use the traditional approach, based on a strong law of large numbers from Hall and Heyde (1980). Below we provide it in an appropriate form for ready reference.

Theorem 3. (See Hall and Heyde (1980, Theorem 2.18).) *Let* $\{T_n = \sum_{i=1}^{n} X_i,\ \mathfrak{F}_n, n \geq 1\}$ *be a martingale and* $\{U_n,\ n \geq 1\}$ *a nondecreasing sequence of positive random variables such that* U_n *is* $\mathfrak{F}_{n-1}$*-measurable for each* n*. If* $1 \leq p \leq 2$ *then* $\lim_{n\to\infty} T_n/U_n = 0$ *a.s. on the set*

$$\left\{\lim_{n\to\infty} U_n = \infty, \sum_{i=1}^{\infty} U_i^{-p}\mathbb{E}[|X_i|^p \mid \mathfrak{F}_{i-1}] < \infty\right\}.$$

It follows from the fact that $\hat{\mu}_n \to \mu$ a.s. as $n \to \infty$ (see Pakes (1971)) that under our conditions

$$\sum_{i=1}^{n} Z_{i-1} \to \infty \quad \text{a.s. as } n \to \infty.$$

Therefore, we can apply Theorem 3 to the sum $T_n = \sum_{i=1}^{n}[Y_i - \theta Z_{i-1}]$. Let $\mathfrak{F}_n$ be the σ-algebra generated by $\{Z_0, Z_1, \ldots, Z_n, Y_1, Y_2, \ldots, Y_n\}$ for each $n \geq 1$. Then, we have

$$\mathbb{E}[T_n \mid \mathfrak{F}_{n-1}] = T_{n-1} + \mathbb{E}[Y_n - \theta Z_{n-1} \mid \mathfrak{F}_{n-1}].$$

It follows from the definition of Y_n that the second term on the right-hand side of this equality is equal to 0. On the other hand, one can easily show that $\mathbb{E}T_n^2 = \theta(1-\theta)\sum_{i=0}^{n-1}\mathbb{E}Z_i$. Therefore, in our conditions $\mathbb{E}T_n^2 < \infty$ for each $n \geq 1$, i.e. $\{T_n,\ \mathfrak{F}_n, n \geq 1\}$ is a martingale.

Since

$$\mathbb{E}[(Y_j - \theta Z_{j-1})^2 \mid \mathfrak{F}_{j-1}] = \theta(1-\theta)Z_{j-1},$$

using the fact that, for any sequence $\{a_i,\ i \geq 1\}$ such that $a_i \geq 0$, the series $\sum_{i=1}^{\infty} a_i(\sum_{j=1}^{i} a_j)^{-2}$ is convergent whenever $\sum_{i=1}^{\infty} a_i = \infty$, which holds due to Dini's theorem (see Knopp (1956, p. 125)), we obtain

$$\sum_{j=1}^{\infty} \frac{\mathbb{E}[(Y_j - \theta Z_{j-1})^2 \mid \mathfrak{F}_{j-1}]}{(\sum_{i=1}^{j} Z_{i-1})^2} < \infty.$$

It is not difficult to see that

$$\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} Z_{i-1}} - \theta = \frac{T_n}{\sum_{i=1}^{n} Z_{i-1}}.$$

The assertion of Proposition 1 now follows, if we apply Theorem 3 to the right-hand side of this equality with $X_i = Y_i - \theta Z_{i-1}$ and $U_n = \sum_{i=1}^{n} Z_{i-1}$. □

Proposition 2. *If* $m < 1$, $\lambda < \infty$, *and* $\sigma^2 < \infty$ *then* $\hat{m}_n \to m$ *a.s. as* $n \to \infty$.

Proof. We refer the reader to Sriram (1991) for the proof. □

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

Proof of Theorem 1. (i) This is a direct consequence of Proposition 1 and consistentcy of the Pakes estimator $\hat{\mu}_n$.

(ii) We use a representation of the total progeny of the partially observed process with immigration in terms of processes without immigration. If we label the individuals immigrated to the same generation by $j = 1, 2, \dots$, then the pair (i, j) corresponds to the jth individual immigrated to ith generation. Each of these immigrants generates a branching process without immigration with the same offspring distribution given in (3) with $f_1(s) = f_2(s)$. Below we shall call the branching process generated by the individual (i, j) as a '(i, j)-process'.

Let $\{\zeta_{ij}(l),\ l \geq 0\}$ be the (i, j)-process and $\{V_{ij}(n),\ n \geq 0\}$ be its total progeny for each pair (i, j), i.e.

$$V_{ij}(n) = \sum_{l=0}^{n} \zeta_{ij}(l), \qquad \zeta_{ij}(0) = 1.$$

It is well known (see Pakes (1971)) that the total progeny of the process Z_n can be written as

$$\sum_{i=0}^{n} \sum_{j=1}^{\nu_i} V_{ij}(n-i),$$

where $\nu_0 = 1$ and $V_{01}(n)$ is the total progeny of the initial individual up to n. Each individual from the process Z_n will be observed (counted in process Y_n) with probability θ. Therefore, the total progeny of the partially observed process with immigration can be represented as

$$\sum_{i=1}^{n} Y_i \stackrel{\mathrm{D}}{=} \sum_{i=0}^{n} \sum_{j=1}^{\nu_i} W_{ij}(n-i),$$

where

$$W_{ij}(r) = \sum_{l=1}^{V_{ij}(r)} \psi_{ijl},$$

and $\{\psi_{ijl},\ (i, j, l) \in \mathcal{N}_0 \times \mathcal{N}^2\}$ is a family of i.i.d. Bernoulli random variables with the probability of success θ, which are independent of all other variables of the model. We use '$\stackrel{\mathrm{D}}{=}$' to mean equality of distribution. We also note that when $X_{ni}^{(1)} = X_{ni}^{(2)}$, random variables $\psi_{ijl},\ l \geq 1$ are independent of $V_{ij}(r),\ r \geq 0$ for each pair $(i, j) \in \mathcal{N}_0 \times \mathcal{N}$. Now we can write

$$\sqrt{n}(\hat{t}_n - \theta\mu) \stackrel{\mathrm{D}}{=} \sum_{i=0}^{n} \bar{\chi}_{ni}, \tag{7}$$

where

$$\bar{\chi}_{ni} = \frac{1}{\sqrt{n}}(\chi_{ni} - \theta\mu), \qquad \chi_{ni} = \sum_{j=1}^{\nu_i} W_{ij}(n-i).$$

To complete the proof, we use the following CLT for a scheme of a series of independent random variables from Petrov (1995).

Theorem 4. (See Petrov (1995, Theorem 4.1).) *Let $\{X_{nk},\ k = 1, \dots, k_n,\ n = 1, 2, \dots\}$ be a sequence of series, and let $F_{nk}(x)$ be the distribution function of X_{nk}. The condition of infinite*

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

smallness will be satisfied and $\sum_{k=1}^{k_n} X_{nk} \xrightarrow{\mathrm{D}} N(\alpha, \beta^2)$, *if and only if, for any* $\varepsilon > 0$ *as* $n \to \infty$,

$$\sum_{k=1}^{k_n} \mathbb{P}\{|X_{nk}| > \varepsilon\} \to 0, \tag{8}$$

$$\sum_{k=1}^{k_n} \left\{ \int_{|x|<\varepsilon} x^2 \, \mathrm{d}F_{nk}(x) - \left(\int_{|x|<\varepsilon} x \, \mathrm{d}F_{nk}(x) \right)^2 \right\} \to \beta^2, \tag{9}$$

$$\sum_{k=1}^{k_n} \int_{|x|<\varepsilon} x \, \mathrm{d}F_{nk}(x) \to \alpha. \tag{10}$$

We note that conditions (8)–(10) are necessary for the asymptotic normality and a CLT, which requires stronger but more convenient to check sufficient conditions would not be applicable in our proof.

Since $\{\bar{\chi}_{ni},\ i = 0, 1, 2, \ldots, n\}$ are independent for each n, we apply Theorem 4 to the sum in (7). As in Pakes (1971), we first need to verify that both $|\chi_{ni} - \theta\mu|$ and $(\chi_{ni} - \theta\mu)^2$, $i = 0, 1, 2, \ldots$ are uniformly integrable. It is known that for the total progeny of a subcritical branching process $V_{ij}(k) \xrightarrow{\mathrm{D}} V_{ij}$ as $k \to \infty$, where $g(s) = \mathbb{E}s^{V_{ij}}$ is a probability generating function and satisfies the functional equation $g(s) = f(g(s))$, $0 \le s \le 1$. Therefore, for each (i, j)-process,

$$W_{ij}(k) \xrightarrow{\mathrm{D}} W_{ij} = \sum_{l=1}^{V_{ij}} \psi_{ijl} \quad \text{as } k \to \infty$$

and

$$\mathbb{E}s^{W_{ij}} = g(\theta s + 1 - \theta).$$

It follows from the independence of ν_i and (i, j)-processes that

$$\sum_{j=1}^{\nu_i} W_{ij}(k) \xrightarrow{\mathrm{D}} \sum_{j=1}^{\nu_i} W_{ij} \quad \text{as } k \to \infty \text{ for each } i \ge 0.$$

We consider

$$\begin{aligned} \mathbb{E}(W_{ij} - W_{ij}(k))^2 &= \mathbb{E}\left(\sum_{l=V_{ij}(k)+1}^{V_{ij}} \psi_{ijl} \right)^2 \\ &= \mathbb{E}\left(\sum_{l=1}^{V_{ij}-V_{ij}(k)} \psi_{ijl} \right)^2 \\ &= \mathbb{E}\left[\sum_{l=1}^{V_{ij}-V_{ij}(k)} \mathbb{E}\psi_{ijl}^2 \right] + \mathbb{E}\left[\sum_{1 \le l < r \le V_{ij}-V_{ij}(k)} \mathbb{E}(\psi_{ijl}\psi_{ijr}) \right]. \end{aligned} \tag{11}$$

The first term on the right-hand side of (11) is equal to $\theta\mathbb{E}[V_{ij} - V_{ij}(k)] = \theta m^{k+1}/(1 - m)$ and the second term is not greater than $\theta^2\mathbb{E}(V_{ij} - V_{ij}(k))^2$. It was shown in Pakes (1971) that $\mathbb{E}(V_{ij} - V_{ij}(k))^2 \le C_1 m^k$, where C_1 is a positive constant. Taking this into account in (11), we obtain that, for some positive constant C_2 and for each (i, j)-process,

$$\mathbb{E}(W_{ij} - W_{ij}(k))^2 \le C_2 m^k. \tag{12}$$

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

We now consider

$$\mathbb{E}\Bigg[\sum_{j=1}^{\nu_i} W_{ij} - \sum_{j=1}^{\nu_i} W_{ij}(k)\Bigg]^2 = \mathbb{E}\Bigg[\sum_{1\le j\le l\le \nu_i} \mathbb{E}(W_{ij} - W_{ij}(k))(W_{il} - W_{il}(k))\Bigg].$$

Using the Cauchy–Schwarz inequality, we find that the expectation on the right-hand side of last equality is not greater than

$$\mathbb{E}\Bigg[\sum_{1\le j\le l\le \nu_i} [\mathbb{E}(W_{ij} - W_{ij}(k))^2 \mathbb{E}(W_{il} - W_{il}(k))^2]^{1/2}\Bigg].$$

Hence, using (12), we obtain, for some positive C_0 and any $i \ge 0$,

$$\mathbb{E}\Bigg[\sum_{j=1}^{\nu_i} W_{ij} - \sum_{j=1}^{\nu_i} W_{ij}(k)\Bigg]^2 \le C_0 m^k. \tag{13}$$

Thus, from estimate (13) we see that the convergence of $\sum_{j=1}^{\nu_i} W_{ij}(k)$ holds in L_2 and, since $m < 1$, it also converges a.s. as $k \to \infty$ for each $i \ge 0$. Hence, the desired uniform integrability follows (see, for example, Gut (2005, Theorem 5.4)).

We now show that the conditions of Theorem 4 are satisfied. Condition (8) is satisfied due to estimate

$$\sum_{i=0}^{n} \mathbb{P}(|\bar{\chi}_{ni}| > \varepsilon) \le \frac{1}{n\varepsilon^2} \sum_{i=0}^{n} \int_{|x|>\varepsilon\sqrt{n}} x^2 \,\mathrm{d}\mathbb{P}(\chi_{ni} - \theta\mu \le x)$$

and the uniform integrability of $(\chi_{ni} - \theta\mu)^2$, $i = 0, 1, 2, \ldots$.

To check (9), we first prove that, as $n \to \infty$,

$$\sum_{i=0}^{n} \int_{|x|<\varepsilon} x^2 \,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \le x) \to a^2, \tag{14}$$

where a^2 is defined in (4). In fact, in representation

$$\sum_{i=0}^{n} \int_{|x|<\varepsilon} x^2 \,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \le x) = \sum_{i=0}^{n} \mathbb{E}\bar{\chi}_{ni}^2 - \sum_{i=0}^{n} \int_{|x|>\varepsilon} x^2 \,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \le x)$$

the second term on the right-hand side tends to 0 as $n \to \infty$ due to the above uniform integrability. The first term can be written as

$$\sum_{i=0}^{n} \mathbb{E}\bar{\chi}_{ni}^2 = I_1(n) + I_2(n),$$

where $I_1(n) = n^{-1}\sum_{i=0}^{n} \operatorname{var}(\chi_{ni})$ and $I_2(n) = n^{-1}\sum_{i=0}^{n} (\mathbb{E}\chi_{ni} - \theta\mu)^2$. Using

$$\operatorname{var}\Bigg(\sum_{i=1}^{N} X_i\Bigg) = \operatorname{var}(X_1)\mathbb{E}N + \operatorname{var}(N)(\mathbb{E}X_1)^2, \tag{15}$$

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

which holds for the variance of a random sum of i.i.d. random variables X_i, $i \geq 1$ independent of the number of terms N, we find that

$$\operatorname{var}(\chi_{ni}) = \lambda\left[\theta(1-\theta)\kappa_n(i) + \frac{\theta^2\sigma^2}{(1-m)^2}\left(\frac{1-m^{2(n-i)+1}}{1-m} + m^{n-i}(1+2(n-i))\right)\right] + (\theta\gamma\kappa_n(i))^2,$$

where

$$\kappa_n(i) =: \frac{1-m^{n-i+1}}{1-m}.$$

In the last equality we show that $\operatorname{var}(\chi_{ni}) \to a^2$ as $n-i \to \infty$ and, consequently, $I_1(n) \to a^2$ as $n \to \infty$. Since $\mathbb{E}\chi_{ni} = \theta\mu(1-m^{n-i+1})$, we have $(\mathbb{E}\chi_{ni} - \theta\mu)^2 \to 0$ as $n-i \to \infty$. Therefore, $I_2(n) \to 0$ as $n \to \infty$ and we see that (14) holds as $n \to \infty$ for any $\varepsilon > 0$.

We now consider the second term in condition (9), i.e.

$$I(n) = \sum_{i=0}^{n}\left[\int_{|x|<\varepsilon} x\,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \leq x)\right]^2.$$

It is not difficult to see that

$$\left[\int_{|x|<\varepsilon} x\,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \leq x)\right]^2 = \frac{1}{n}[\mathbb{E}\chi_{ni} - \theta\mu]^2 - \frac{2}{n}\mathbb{E}(\chi_{ni} - \theta\mu)\int_{|x|>\varepsilon\sqrt{n}} x\,\mathrm{d}\mathbb{P}(\chi_{ni} - \theta\mu \leq x) + \frac{1}{n}\left[\int_{|x|>\varepsilon\sqrt{n}} x\,\mathrm{d}\mathbb{P}(\chi_{ni} - \theta\mu \leq x)\right]^2.$$

Using this equality and the estimate

$$|\mathbb{E}\chi_{ni} - \theta\mu| \leq \theta\mu m^{n-i+1}, \tag{16}$$

we see that $I(n) \to 0$ as $n \to \infty$ for any $\varepsilon > 0$. Thus, from this and (14), we see that (9) is satisfied with the limit $\beta^2 = a^2$.

To prove that condition (10) holds, we consider

$$\sum_{i=0}^{n}\int_{|x|<\varepsilon} x\,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \leq x) = \sum_{i=0}^{n}\mathbb{E}\bar{\chi}_{ni} - \sum_{i=0}^{n}\int_{|x|>\varepsilon} x\,\mathrm{d}\mathbb{P}(\bar{\chi}_{ni} \leq x).$$

Using (16), we see that the first term on the right-hand side of the last equality is not greater than $\theta\mu\sum_{i=0}^{n} m^i/\sqrt{n}$ and tends to 0 as $n \to \infty$. The same arguments as in the proof of (14) allow us to conclude that the second term also tends to 0 as $n \to \infty$. Hence, the assertion follows from Theorem 4. □

Proof of Theorem 2. First we prove the consistency of the estimator $\hat{a}_n$. We represent it as

$$\hat{a}_n = A_1(n) + A_2(n) + A_3(n), \tag{17}$$

where

$$A_1(n) = \frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n} Y_i}\left(\frac{\sum_{i=2}^{n+1} Y_i}{\sum_{i=1}^{n} Z_i} - \theta\right), \qquad A_2(n) = \frac{\theta\sum_{i=1}^{n}(Z_i - \nu_i)}{\sum_{i=1}^{n} Y_i},$$
$$A_3(n) = \frac{\sum_{i=1}^{n}(\theta\nu_i - \eta_{i+1})}{\sum_{i=1}^{n} Y_i}.$$

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

It follows from Theorem 1(i), Proposition 1 and the consistency of $\hat{\mu}_n$ that $A_1(n) \to 0$ a.s. as $n \to \infty$. Now we consider

$$A_3(n) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{\nu_i} (\theta - \xi_{ij})}{\sum_{i=1}^{n} Y_i}.$$

Since, due to the strong law of large numbers,

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{\nu_i} (\theta - \xi_{ij}) \to \mathbb{E}\left[\sum_{j=1}^{\nu_1} (\theta - \xi_{1j}) \right] = 0 \quad \text{a.s. as } n \to \infty,$$

using Theorem 1(i), we conclude that $A_3(n)$ a.s. converges to 0 as $n \to \infty$. With regard to $A_2(n)$, we can write it as

$$A_2(n) = \frac{\theta \sum_{i=1}^{n} Z_{i-1}}{\sum_{i=1}^{n} Y_i} \frac{\sum_{i=1}^{n} (Z_i - \nu_i)}{\sum_{i=1}^{n} Z_{i-1}}.$$

From here, using Proposition 1 and Proposition 2 (consistentcy of Nanthi's estimator $\hat{m}_n$), we obtain $A_2(n) \to m$ a.s. as $n \to \infty$. Theorem 2(i) now follows from (17).

We now prove (ii). For this, we consider

$$B(n) := (\hat{b}_n - m) \sum_{i=1}^{n_o} Y_{2i} = \sum_{k=1}^{n_o} (Y_{2k+1} - \eta_{2k+1} - m Y_{2k}).$$

We denote $L_k(m, n) = \sum_{l=m}^{n} X_{k-2,l} + \nu_{k-1}$, $k \geq 2$, where X_{kl}, $(k, l) \in \mathcal{N}^2$ is defined in (2). Then we have, for $k \geq 2$,

$$\begin{aligned}
Y_k - \eta_k &= \sum_{j=1}^{L_k(1, Z_{k-2})} \xi_{k-1j} - \sum_{j=1}^{\nu_{k-1}} \xi_{k-1j} \\
&= \sum_{j=\nu_{k-1}+1}^{L_k(1, Z_{k-2})} \xi_{k-1j} \\
&= \sum_{j=\nu_{k-1}+1}^{L_k(1,1)} \xi_{k-1j} + \sum_{j=L_k(1,1)+1}^{L_k(1,2)} \xi_{k-1j} + \cdots + \sum_{j=L_k(1, Z_{k-2}-1)+1}^{L_k(1, Z_{k-2})} \xi_{k-1j} \\
&= \sum_{i=1}^{Z_{k-2}} \sum_{j=1}^{X_{k-2,i}} \rho_{k-1ij},
\end{aligned}$$

where ρ_{kij}, $(k, i, j) \in \mathcal{N}^3$ are i.i.d. Bernoulli random variables with the probability of success θ, which are independent of all other random variables in the model. We note that the terms in the last double sum are the same $\sum_{i=1}^{Z_{k-2}} X_{k-2,i}$ Bernoulli random variables ξ_{k-1j}, $j = \nu_{k-1} + 1, \ldots, \nu_{k-1} + \sum_{i=1}^{Z_{k-2}} X_{k-2,i}$ rearranged accordingly.

Thus, $B(n)$ can be represented as

$$B(n) = \sum_{k=1}^{n_o} \sum_{i=1}^{Z_{2k-1}} \omega_{ki}, \tag{18}$$

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

where

$$\omega_{ki} = \sum_{j=1}^{X_{2k-1i}} \rho_{2k,ij} - m\xi_{2k-1i}, \qquad i \in \mathcal{N},$$

are i.i.d. random variables for each pair $(k, i) \in \mathcal{N}^2$. Since X_{2k-1i} and $\rho_{2k,ij}$, $j \geq 1$ are independent for each (k, i), using Wald's identity we find that $\mathbb{E}\omega_{ki} = 0$, $(k, i) \in \mathcal{N}^2$. To find the variance of ω_{ki}, we use

$$\operatorname{var}(\omega_{ki}) = \mathbb{E}\left[\left(\sum_{j=1}^{X_{2k-1i}} \rho_{2k,ij}\right)^2\right] - 2m\mathbb{E}\left[\xi_{2k-1i} \sum_{j=1}^{X_{2k-1i}} \rho_{2k,ij}\right] + m^2\mathbb{E}\xi_{2k-1i}^2. \tag{19}$$

Using (15) for the variance of the random sum, we have

$$\mathbb{E}\left[\left(\sum_{j=1}^{X_{2k-1i}} \rho_{2k,ij}\right)^2\right] = \operatorname{var}\left(\sum_{j=1}^{X_{2k-1i}} \rho_{2k,ij}\right) + (m\theta)^2 = m\theta(1-\theta) + (\sigma\theta)^2 + (m\theta)^2.$$

Taking into account (2) by a standard conditioning technique, we find that the second term in (19) is equal to $-2m\theta^2\mu$ and the third term is equal to $m^2\theta$. Hence, straightforward calculations yield $\operatorname{var}(\omega_{ki}) = \theta b^2$, where b^2 is defined in (5).

To complete the proof, we need the following result.

Lemma 1. *Let $m < 1, \lambda < \infty$. Then*

(i) $n^{-1}\sum_{k=1}^{n} Z_{2k-1} \to \mu$ *a.s. as* $n \to \infty$;

(ii) $\sum_{k=1}^{n} Y_{2k} / \sum_{k=1}^{n} Z_{2k-1} \to \theta$ *a.s. as* $n \to \infty$.

Proof. (i) It follows from the definition of the process $\{Z_n,\ n \geq 1\}$ that

$$Z_{2n+1} = \sum_{i=1}^{Z_{2n-1}} X_{2n-1i}(2) + \sum_{i=1}^{\nu_{2n}} X_{2n,i}(1) + \nu_{2n+1}, \qquad n \geq 1, \tag{20}$$

where $X_{ij}(n)$, $(i, j) \in \mathcal{N}^2$, $n \geq 0$, are i.i.d. branching processes without immigration, with the offspring generating function $f(s)$ defined in (3), and $X_{ij}(0) = 1$. From (20) we see that $\{\hat{Z}_n = Z_{2n-1},\ n \geq 1\}$ is a branching process with immigration whose the offspring generating function is $f(f(s))$ and the number of immigrating individuals has the probability generating function $h(s)h(f(s))$, where $h(s) = \mathbb{E}s^{\nu_1}$, $\hat{Z}_1 = X_{01} + \nu_1$. Therefore, $\{\hat{Z}_n,\ n \geq 1\}$ is a Markov chain defined on the state-space $\{0, 1, 2, \dots\}$ and its transition probabilities $\mathbb{P}_{ij} = \mathbb{P}\{\hat{Z}_{n+1} = j \mid \hat{Z}_n = i\}$, $i, j \geq 0$, have the probability generating function

$$\sum_{j=0}^{\infty} \mathbb{P}_{ij}s^j = h(s)h(f(s))f^i(f(s)), \quad i \geq 0.$$

Using the trivial equality $\lim_{n\to\infty} \mathbb{P}\{Z_n = i\} = \lim_{n\to\infty} \mathbb{P}\{Z_{2n-1} = i\}$, which holds for each $i \geq 0$, we see that $\hat{Z}_n$, $n \geq 1$, has the same limiting-stationary distribution with mean μ as that for Z_n. Thus, the assertion of Lemma 1(i) follows from the law of large numbers for Markov chains as in Pakes (1971).

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

(ii) We consider

$$\frac{\sum_{k=1}^{n} Y_{2k}}{\sum_{k=1}^{n} Z_{2k-1}} - \theta = \frac{1}{\sum_{k=1}^{n} Z_{2k-1}} \sum_{k=1}^{n} (Y_{2k} - \theta Z_{2k-1}).$$

If, by $\mathcal{A}_k$, we denote the σ-algebra generated by $(Z_1, Z_2, \ldots, Z_{2k}, Z_{2k+1}, Y_1, Y_2, \ldots, Y_{2k})$, we can easily prove that $\mathbb{E}(Y_{2k} - \theta Z_{2k-1} \mid \mathcal{A}_{k-1}) = 0$ and $(\sum_{k=1}^{n}(Y_{2k} - \theta Z_{2k-1}), \mathcal{A}_k)$ is a martingale. The assertion of Lemma 1(ii) can now be obtained as in the proof of Proposition 1. This completes the proof. □

We now continue the proof of Theorem 2. Taking into account (18), we obtain

$$\sqrt{n}(\hat{b}_n - m) \stackrel{\mathrm{D}}{=} \sqrt{\frac{n}{\hat{U}_n}} \frac{\hat{U}_n}{\sum_{k=1}^{n_o} Y_{2k}} \frac{1}{\sqrt{\hat{U}_n}} \sum_{j=1}^{\hat{U}_n} w_j, \tag{21}$$

where $\hat{U}_n = \sum_{k=1}^{n_o} Z_{2k-1}$, $n_o = [(n-1)/2]$, and w_j, $j \geq 1$ are i.i.d. random variables with $\mathbb{E}(w_j) = 0$ and $\mathrm{var}(w_j) = \theta b^2$. Owing to Lemma 1(i), we can apply Theorem 14.4 of Billingsley (1999) to the last random sum in (21) and obtain, as $n \to \infty$,

$$\frac{1}{\sqrt{\hat{U}_n}} \sum_{j=1}^{\hat{U}_n} w_j \stackrel{\mathrm{D}}{\to} \mathcal{N}(0, \theta b^2). \tag{22}$$

The assertion of Theorem 2(ii) now follows from (22), Lemma 1 and Slutsky's theorem. Theorem 2 is proved. □

Acknowledgements

My sincere thanks to the anonymous referee and the associate editor for their valuable comments and to the University College of Zayed University.

References

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and their Statistical Analysis*. Springer, New York.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.

Gut, A. (2005). *Probability: A Graduate Course*, 2nd edn. Springer, New York.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

Knopp, K. (1956). *Infinite Sequences and Series*. Dover, New York.

Kvitkovičová, A. and Panaretos, V. M. (2011). Asymptotic inference for partially observed branching processes. *Adv. Appl. Prob.* **43,** 1166–1190.

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. S. and Diekmann, O. (2002). Modeling and real-time prediction of classical swine fever epidemics. *Biometrics* **58,** 178–184.

Nanthi, K. (1979). Some limit theorems of statistical relevance on branching processes. Doctoral Thesis, University of Madras.

Pakes, A. G. (1971). Branching processes with immigration. *J. Appl. Prob.* **8,** 32–42.

Panaretos, V. M. (2007). Partially observed branching processes for stochastic epidemics. *J. Math. Biol.* **54,** 645–668.

Petrov, V. V. (1995). *Limit Theorems of Probability Theory* (Oxford Studies in Probability **4**). Oxford University Press.

Scott, D. J. (1977/78). A central limit theorem for martingales and an application to branching processes. *Stoch. Process. Appl.* **6,** 241–252.

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press

Sriram, T. N. (1991). On uniform strong consistency of an estimator of the offspring mean in a branching process with immigration. *Statist. Prob. Lett.* **12,** 151–155.

Venkataraman, K. N. and Nanthi, K. (1982). A limit theorem on subcritical Galton–Watson process with immigration. *Ann. Prob.* **10,** 1069–1074.

Yanev, N. M. (2008). Statistical inference for branching processes. In *Records and Branching Processes*, eds M. Ahsanullah and G. P. Yanev, Nova Science, New York.

https://doi.org/10.1017/jpr.2016.88 Published online by Cambridge University Press