# ESTIMATION IN BRANCHING PROCESSES WITH RESTRICTED OBSERVATIONS

RONALD MEESTER,* *Vrije Universiteit Amsterdam*

PIETER TRAPMAN,** *Utrecht University and Vrije Universiteit Amsterdam*

## Abstract

We consider an epidemic model where the spread of the epidemic can be described by a discrete-time Galton–Watson branching process. Between times $n$ and $n + 1$, any infected individual is detected with unknown probability $\pi$ and the numbers of these detected individuals are the only observations we have. Detected individuals produce a reduced number of offspring in the time interval of detection, and no offspring at all thereafter. If only the generation sizes of a Galton–Watson process are observed, it is known that one can only estimate the first two moments of the offspring distribution consistently on the explosion set of the process (and, apart from some lattice parameters, no parameters that are not determined by those moments). Somewhat surprisingly, in our context, where we observe a binomially distributed subset of each generation, we are able to estimate three functions of the parameters consistently. In concrete situations, this often enables us to estimate $\pi$ consistently, as well as the mean number of offspring. We apply the estimators to data for a real epidemic of classical swine fever.

*Keywords:* Galton–Watson branching process; consistent estimation; partial observation; epidemic; real-time estimation

2000 Mathematics Subject Classification: Primary 60J80; 62F12
Secondary 62P10; 60G42

## 1. Introduction and motivation

It is known that it is impossible to consistently estimate more than two moments of the offspring distribution of a supercritical Galton–Watson process if only the generation sizes, $X_n$, of the process are observed [4, Theorem 1.3]. However, it is not a-priori clear what can be estimated consistently in a situation where, between times $n$ and $n + 1$, any infected individual is detected with unknown probability $\pi$ and the numbers of these detected individuals are the only observations we have. Detected individuals produce a reduced number of offspring in the time interval of detection, and no offspring at all thereafter.

The kind of partial observation we are dealing with is especially interesting for estimation in epidemics of infectious diseases. If the number of susceptible individuals (where an 'individual' may also refer to a herd as well as to an individual animal) is very large, we may describe the start of an idealized epidemic using a Galton–Watson process where discrete points in time index the generations (see, e.g. [1, Chapter 3] or [10]). As soon as an infectious disease is observed in an individual, it stops being infective, because of either isolation (in the case of

human infections) or culling (in the case of very contagious animal diseases like classical swine fever, foot-and-mouth disease, or avian influenza).

In the time interval of detection, the individual is only infectious during a fraction of the interval length, which implies that observed individuals have a reduced number of offspring. In this epidemiological setting, an individual that is infectious but not detected in a certain generation will still be infectious in the next one, so a surviving individual will cause there to be at least one infective individual in the next generation, namely itself. Our aim is to estimate parameters of the offspring distribution using only this partial information. This interpretation should be compared with the work in [2]. In that paper, estimation takes place under the assumption that one also observes, in addition to individuals with no further offspring, the total number of infectious individuals at both the beginning and the end of the observation period.

Assuming that $X_n$ denotes the generation sizes of a branching process and $\pi_n$ denotes a known sequence converging to $\pi$, Jacob and Peccoud [6] showed that if the number of observations in generation $n + 1$ is binomially distributed with parameters $X_n$ and $\pi_n$ and the offspring distribution has a finite fourth moment, then it is also possible to estimate the first two moments of the offspring distribution consistently on the explosion set (i.e. the set where $\lim_{n \to \infty} X_n = \infty$). Under the assumptions of [6], the observed individuals may produce offspring, but the number of these offspring is supposed to be distributed in the same way as the number of offspring of $L$ unobserved individuals, where $L$ is a nonnegative integer.

Our set-up differs from that of [6] in two aspects. First, we are interested in the case where $\pi_n \equiv \pi$ is constant but unknown. Second, we assume that the offspring distributions of unobserved and observed individuals have a finite fourth moment, but make no further assumptions about them. Our methods are also quite different; our martingales are based on observable quantities. Besides proving some results analogous to [6], we show that $\pi$ can, in certain circumstances, be estimated consistently.

Our main interest is in estimating the offspring mean and the parameter $\pi$, because these parameters are extremely important for decisions about measures to be taken to stop an epidemic. We are able to estimate the offspring mean very efficiently. On the explosion set, we are (under certain conditions) able to consistently estimate two other functions of $\pi$ and the parameters of the offspring distribution. These three estimators will lead to a system of three equations. For many models we also have three unknowns, namely the offspring mean, the offspring variance, and $\pi$. In principle, we can therefore often estimate these quantities. However, it turns out that what is theoretically possible is not always practically feasible, due to extremely slow convergence of the second and third estimators. Note that we can only hope to obtain consistency if the process explodes; otherwise, the number of observations and the number of involved individuals will be finite.

In the next section we set up the problem formally. In Section 3 we give consistent estimators for three functions of the parameters and results about the rates of convergence of these estimators. We apply this to real data from the 1997 epidemic of classical swine fever in the Netherlands. In Section 4 we prove the consistency of two estimators for the offspring mean on the explosion set of the branching process. In Section 5 we estimate a second function of the parameters. This function can be interpreted as a second moment, as we will explain. In Section 6 we show that the estimator for a third function of the parameters is consistent.

## 2. Formal set-up

We let $G_n$ denote the collection of infected individuals at the discrete time instants $n$, $n = 0, 1, 2, \ldots$, and we let $X_n = |G_n|$ denote its cardinality. In our context, the $X_n$ are *not*

observable. The dynamics from time $n$ to time $n + 1$ is as follows. Between times $n$ and $n + 1$, a certain (random) number of the infected individuals in $G_n$ is detected; we assume that each infected individual is detected with probability $\pi$ during this time interval, independently of the others. The parameter $\pi$ is unknown. The collection of individuals detected between time $n$ and time $n + 1$ is denoted by $D_{n+1}$, and the number of individuals in $D_{n+1}$ is denoted by $Z_{n+1}$; this random quantity *is* observable. So, given $X_n$, $Z_{n+1}$ has a binomial distribution with parameters $X_n$ and $\pi$. Individuals in $\bigcup_n D_n$ produce no offspring.

An individual in $G_n$ which is detected (and which therefore produces an element in $D_{n+1}$) may also produce offspring in $G_{n+1}$ (in our terminology, 'offspring' always means 'direct offspring'). An individual in $G_n$ which is not detected will remain infective, and will possibly infect other individuals. The offspring of such an individual in $G_{n+1}$ therefore consists of at least one individual, namely itself. Note that, as a result, one physical individual corresponds to various individuals of the process. The whole process now constitutes a two-type branching process.

The offspring distributions of detected and undetected individuals are different. We denote by $m_+$ the expected number of $X$-offspring of an infected individual at time $n$ (that is, offspring in $G_{n+1}$), given that it is not detected between time $n$ and time $n + 1$. Similarly, $m_-$ is the expected number of $X$-offspring of an infected individual given that it is detected. The corresponding variances are denoted by $\sigma_+^2$ and $\sigma_-^2$, respectively. In formulae, this reads as follows:

$$m_+ := \mathrm{E}(X_1 \mid X_0 = 1,\ Z_1 = 0),$$
$$m_- := \mathrm{E}(X_1 \mid X_0 = 1,\ Z_1 = 1),$$
$$\sigma_+^2 := \mathrm{E}((X_1 - m_+)^2 \mid X_0 = 1,\ Z_1 = 0),$$
$$\sigma_-^2 := \mathrm{E}((X_1 - m_-)^2 \mid X_0 = 1,\ Z_1 = 1).$$

Finally, we write $m$ for the unconditional expected number of $X$-offspring of an infected individual:

$$m = \mathrm{E}(X_1 \mid X_0 = 1) = (1 - \pi)m_+ + \pi m_-.$$

Similarly, the unconditional variance is denoted by $\sigma^2$. From [4, Lemma 2.1], for a random variable $Y$ and an event $F$ (with complement $F^c$), we have

$$\mathrm{var}(Y) = \mathrm{P}(F)\,\mathrm{var}(Y \mid F) + \mathrm{P}(F^c)\,\mathrm{var}(Y \mid F^c)$$
$$+ (\mathrm{E}(Y \mid F) - \mathrm{E}(Y \mid F^c))^2\,\mathrm{P}(F)\,\mathrm{P}(F^c).$$

Applying this with $F = \{Z_1 = 0\}$ yields

$$\sigma^2 = \mathrm{var}(X_1 \mid X_0 = 1) = (1 - \pi)\sigma_+^2 + \pi\sigma_-^2 + (m_+ - m_-)^2\pi(1 - \pi).$$

We assume that the offspring distributions have finite fourth moments, i.e. for $i \in \{0, 1\}$,

$$\mathrm{E}([X_1 - \mathrm{E}(X_1 \mid X_0 = 1,\ Z_1 = i)]^4 \mid X_0 = 1,\ Z_1 = i) < \infty.$$

Define $A$ to be the explosion set, that is, the set where $\lim_{n \to \infty} X_n = \infty$. Because $A$ is a tail event, conditioning on this event is, strictly speaking, not correct for estimation purposes; but analysing the behaviour of the process on the set $A$ is necessary, for only on this set do we obtain infinitely many observations. In practical situations we can simply ignore this, since we only use observable quantities.

We are able to estimate several functions of $m_+, m_-, \sigma^2$, and $\pi$ on $A$. One of those functions is $m$, and the other two are given by

$$\gamma \equiv \gamma(m_+, m_-, \sigma^2, \pi) := (1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2 - 2\pi m m_-,$$

$$\gamma_* \equiv \gamma_*(m_+, m_-, \sigma^2, \pi) := (m^2 + m)\gamma - 2m^3 + 2\pi m^2 m_-.$$

The reason for these somewhat complicated expressions will become clear soon.

## 3. Results and application

### 3.1. The main result

In this section consistent estimators for three different functions of the parameters are given.

**Theorem 3.1.** *Using the notation and assumptions of Section 2, we have, as $n \to \infty$,*

(a) $\bar{m}_n \to m$ *almost surely (a.s.) on $A$,*

(b) $\tilde{m}_n \to m$ *a.s. on $A$,*

(c) $n^{-1}\tilde{S}_n(\tilde{m}_n) \to \gamma$ *in probability on $A$,*

(d) $n^{-1}\tilde{S}_n^*(\tilde{m}_n) \to \gamma_*$ *in probability on $A$,*

*where*

$$\bar{m}_n = \frac{Z_{n+1}}{Z_n}, \qquad \tilde{m}_n = \frac{\sum_{i=2}^{n+1} Z_i}{\sum_{i=1}^n Z_i},$$

$$\tilde{S}_n(m) = \sum_{i=1}^n (Z_i + 1)\left(\frac{Z_{i+1}}{Z_i + 1} - m\right)^2,$$

$$\tilde{S}_n^*(m) = \sum_{i=1}^n (Z_i + 1)\left(\frac{Z_{i+2}}{Z_i + 1} - m^2\right)^2.$$

In principle, this theorem gives three equations with four unknowns, namely $m$, $\pi$, $\sigma^2$, and $m_-$. If we have further information, or make further assumptions about the relationship between $m$ and $m_-$, then we can estimate all parameters consistently, in theory at least.

The speed of convergence of our estimators is given by the following two theorems.

**Theorem 3.2.** *The random variables $(\sum_{i=1}^n Z_i)^{1/2}(\tilde{m}_n - m)$ converge in distribution to a sum of three normal random variables with zero mean and finite variance.*

This theorem implies that $\tilde{m}_n - m$ converges to 0 with a rate of order $(\sum_{i=1}^n Z_i)^{-1/2}$.

**Theorem 3.3.** *As $n \to \infty$, for all $\delta > 0$ we have*

(a) $n^{1/2-\delta}(n^{-1}\tilde{S}_n(\tilde{m}_n) - \gamma) \to 0$,

(b) $n^{1/2-\delta}(n^{-1}\tilde{S}_n^*(\tilde{m}_n) - \gamma_*) \to 0$,

*in probability on $A$.*

This theorem implies that $n^{-1}\tilde{S}_n(\tilde{m}_n) - \gamma$ and $n^{-1}\tilde{S}_n^*(\tilde{m}_n) - \gamma_*$ converge with a rate of order at least $n^{-1/2+\delta}$, for any $\delta > 0$.

The proofs of Theorem 3.1(a), Theorem 3.1(b), and Theorem 3.2 are given in Section 4. The proofs of Theorem 3.1(c) and Theorem 3.3(a) are given in Section 5. Finally, the proof of Theorem 3.1(d) is given in Section 6. Theorem 3.3(b) can be proved in exactly the same way as Theorem 3.3(a), and the proof is thus omitted.

## 3.2. Application to an epidemic model

In this subsection we apply our results to a concrete example from epidemic theory. We analyse the discrete approximation of the standard SIR (susceptible–infective–removed) epidemic (see, e.g. [3]), where we take the number of susceptible individuals to be infinite.

3.2.1. *The model.* We assume that if an infective individual is not detected in a certain interval, then the number of new infections caused by this infective individual is Poisson distributed with parameter $\lambda$. Since an individual that is not detected remains infective itself, this implies that

$$m_+ = \lambda + 1, \qquad \sigma_+^2 = \lambda.$$

Next we need to choose what happens during the interval in which an individual is detected. In order to keep the model general, we assume that the detected individual is infective during a (known) fraction, $\phi$, of the detection interval. It then follows that

$$m_- = \phi\lambda, \qquad \sigma_-^2 = \phi\lambda,$$

and, hence,

$$m = (1 - \pi)(\lambda + 1) + \phi\lambda\pi,$$
$$\sigma^2 = (1 - \pi)\lambda + \phi\lambda\pi + \pi(1 - \pi)((1 - \phi)\lambda + 1)^2.$$

One remark: if we assume that the detection time is uniformly distributed over the interval of detection, then $m_-$ will equal $\lambda/2$. The variance of the offspring in the interval of detection will be slightly larger than $\lambda/2$ because the randomness of the detection time will cause some extra variance. In fact, the variance will be $\lambda/2 + \lambda^2/12$.

We are particularly interested in estimating $m$ (which describes the mean growth of the number of infectious individuals) and $\pi$ (needed to estimate the number of infectious individuals at a certain time, which is very important in order to make decisions about measures to stop the epidemic). In the context of the present example, Theorem 3.1 gives

$$
\begin{aligned}
\tilde{m}_n &\to (1 - \pi)(\lambda + 1) + \phi\lambda\pi & &\text{a.s. on } A, \\
n^{-1}\tilde{S}_n(\tilde{m}_n) &\to (1 - \pi)\lambda + (1 - \pi)^2 + (1 - \pi)(\lambda + 1)^2 + \phi\lambda\pi & &\text{in probability on } A, \\
n^{-1}\tilde{S}_n^*(\tilde{m}_n) &\to (m^2 + m)\gamma - 2m^2(1 - \pi)(\lambda + 1) & &\text{in probability on } A.
\end{aligned}
$$
$$(3.1)$$

For ease of notation, we have not expanded the quantity on the right-hand side of the final expression. In (3.1) we have, after substituting the estimates for $m$, $\gamma$, and $\gamma_*$, three equations with two unknowns, $\lambda$ and $\pi$. At this point, it seems that the third equation does not help us very much. There are two ways to proceed.

First, we can ignore the third equation and solve the other two for $\lambda$ and $\pi$. However, it turns out that $n^{-1}\tilde{S}_n(\tilde{m}_n)$ converges very slowly, meaning that we need a huge number of generations to obtain reliable estimates (see the next subsection). Second, we can use the information contained in the third equation in a meaningful way, to reparametrize the epidemic

TABLE 1: Maximum likelihood estimates (MLEs) from the 1997 Dutch classical swine fever outbreak.

| Stage | Duration (weeks) | Number of observations | MLEs | | Computed values | | | Estimated values | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\pi$ | $\lambda$ | $m$ | $\gamma$ | $\gamma_*$ | $m$ | $\gamma$ | $\gamma_*$ |
| 2 | 10 | 101 | 0.4 | 0.6 | 1.08 | 2.376 | 3.098 | 1.15 | 1.03 | 0.942 |
| 3 | 8 | 160 | 0.5 | 0.7 | 1.025 | 2.220 | 2.822 | 1.04 | 1.87 | 2.57 |
| 4 | 9 | 107 | 0.3 | 0.3 | 0.955 | 1.928 | 1.940 | 0.857 | 0.813 | 0.935 |
| 5 | 30 | 51 | 0.05 | 0.25 | 1.194 | 2.631 | 3.505 | 0.880 | 0.859 | 0.867 |

process using $m$ and $\pi$ instead of $\lambda$ and $\pi$. The parameter $m$ is estimated using $\tilde{m}_n$, and we may use the combination

$$(1 + \tilde{m}_n^{-1})n^{-1}\tilde{S}_n(\tilde{m}_n) - \tilde{m}_n^{-2}n^{-1}\tilde{S}_n^*(\tilde{m}_n)$$

to estimate $2(1 - \pi)(\lambda + 1)$. Since $m = (1 - \pi)(\lambda + 1) + \phi\lambda\pi$, we may write

$$\lambda + 1 = \frac{m + \phi\pi}{(1 - \pi) + \phi\pi},$$

so we again have a system of two equations with two unknowns. From simulation results, it turns out that we can give reasonable estimates for $\pi$ that converge much faster than in the case where we use only the estimators for $m$ and $\gamma$, but for practical purposes our new estimator for $\pi$ still converges too slowly.

3.2.2. *The data.* We have analysed the data from the 1997 Dutch classical swine fever outbreak as treated in [8]. In that paper, the outbreak was modelled using a Galton–Watson process. Because of changing measures taken by the Dutch government, the parameters $\pi$ and $\lambda$ differ at different stages of the epidemic. For this reason, we divide the epidemic into five stages. The time unit is one week, and it is assumed that, on average, detections take place in the middle of a time interval, so $\phi$ is set to 0.5. In [8] a time-consuming algorithm was used to find the maximum likelihood estimators for $\pi$ and $\lambda$ at the different stages. Klinkenberg [7, Chapter 6] has shown that these maximum likelihood estimates are not very reliable.

We will use our estimators to estimate $m$ and $\gamma$ at the different stages. We omit the first stage, because for that stage we have only one observation. We compare our estimates with the values for $m$ and $\gamma$ computed from the maximum likelihood estimates of $\lambda$ and $\pi$ given in [8]. The results are given in Table 1. We also give the duration of the stage (in weeks) and the number of observed individuals in a stage of the epidemic in this table.

In the second, third, and fourth stages of the epidemic, our estimates for $m$ seem to be rather good, as we might expect. The estimates for $\gamma$ do not seem to be very informative. In the final stage of the epidemic, significantly fewer cases were observed than in the second, third, and fourth stages, and there were many weeks in which no observations were made. For this reason, we may expect our estimators not to converge very quickly. In none of the stages could we estimate $\pi$ and $\lambda$ by using our estimated $m$ and $\gamma$ in (3.1), as the solutions of this system of equations gave no real solution for $\pi$ between 0 and 1. We have simulated the epidemic using the maximum likelihood estimates from [8] as the real parameter values to obtain some idea about the speed of convergence of the estimator of $\pi$ (see Table 2). We see that, even after 750 weeks, $\pi$ is not accurately estimated.

TABLE 2: Estimated values at different generations of simulated data, for $\lambda = 0.6$, $\pi = 0.4$, and $\phi = 0.5$.

| Generation ($n$) | $\tilde{m}_n$ | $n^{-1}\tilde{S}_n(\tilde{m}_n)$ | $(n-1)^{-1}\tilde{S}_{n-1}^*(\tilde{m}_n)$ | Estimated $\pi$ |
|---|---|---|---|---|
| 20 | 1.112 | 2.907 | 2.745 | — |
| 50 | 1.082 | 2.849 | 3.142 | — |
| 100 | 1.082 | 2.928 | 3.750 | — |
| 250 | 1.080 | 2.478 | 3.424 | 0.4601 |
| 500 | 1.080 | 2.522 | 3.288 | 0.2852 |
| 750 | 1.080 | 2.398 | 3.059 | 0.3318 |

## 4. Estimating the offspring mean

In this section we discuss the two consistent estimators for $m$ given in Section 3.1. We start with the estimator $\bar{m}_n := Z_{n+1}/Z_n$. Theorem 3.1(a) states that this estimator is indeed consistent on the explosion set $A$.

*Proof of Theorem 3.1(a).* The proof is based on a simple martingale argument. Let

$$M_n := \prod_{i=0}^{n} \frac{Z_{i+1} + 1}{\pi X_i + 1}.$$

Note that this is a (positive) martingale with respect to $\mathcal{F}_n$, the $\sigma$-algebra generated by $\{Z_{i+1}, X_i : 0 \le i \le n\}$. Since $\sup_n \mathrm{E}(M_n) \le 1$, the martingale convergence theorem implies that $M_n$ converges a.s. to an a.s. finite random variable $M$.

We need to show that $M$ is strictly positive on $A$. To do this, we define

$$\bar{M}_n := \prod_{i=0}^{n} \frac{\pi(X_i + 1)}{Z_{i+1} + 1}.$$

Elementary computations yield

$$\mathrm{E}\left(\frac{1}{Z_1 + 1} \,\middle|\, X_0 = k\right) = \frac{1}{\pi(k+1)}(1 - (1-\pi)^{k+1}) \le \frac{1}{\pi(k+1)}, \tag{4.1}$$

so $\bar{M}_n$ is a supermartingale with respect to $\mathcal{F}_n$. By the martingale convergence theorem, we know that $\bar{M}_n$ converges a.s. to an a.s. finite random variable $\bar{M}$. Now write

$$M_n \bar{M}_n = \prod_{i=0}^{n} \frac{\pi(X_i + 1)}{\pi X_i + 1} = \prod_{i=0}^{n}\left(1 - \frac{1-\pi}{\pi X_i + 1}\right).$$

The $X_i$ a.s. grow exponentially on $A$, so

$$\sum_{i=1}^{\infty} \frac{1-\pi}{\pi X_i + 1} < \infty$$

a.s. on $A$, which implies that

$$M\bar{M} = \lim_{n\to\infty} M_n \bar{M}_n = \prod_{i=0}^{\infty}\left(1 - \frac{1-\pi}{\pi X_i + 1}\right) > 0 \tag{4.2}$$

a.s. on $A$. Because $\bar{M}$ is a.s. finite, (4.2) is only possible if $M$ is a.s. positive on $A$.

Now, since $X_{i+1}/X_i \to m$ a.s. on $A$ [4, Theorem 2.1], we have

$$\frac{X_{n+1} + \pi^{-1}}{X_n + \pi^{-1}} \to m \quad \text{a.s. on } A.$$

Thus,

$$\frac{Z_{n+1} + 1}{Z_n + 1} = \frac{M_n/M_{n-1}}{M_{n-1}/M_{n-2}} \frac{\pi X_n + 1}{\pi X_{n-1} + 1} \to m \quad \text{a.s. on } A,$$

because, since $0 < M < \infty$ a.s. on $A$, it follows that

$$\frac{M_n}{M_{n-1}} \to \frac{M}{M} = 1 \quad \text{a.s. on } A.$$

On $A$, $Z_n$ will a.s. tend to infinity, so $(Z_{n+1} + 1)/(Z_n + 1)$ will have the same limit as $Z_{n+1}/Z_n$, proving the theorem.

Clearly, $\bar{m}_n$ does not use all the available information. To this end, we also considered the second estimator, namely

$$\tilde{m}_n := \frac{\sum_{i=2}^{n+1} Z_i}{\sum_{i=1}^{n} Z_i}.$$

Theorem 3.1(b) states that $\tilde{m}_n$ also is a consistent estimator for $m$. To prove this theorem, we start with a lemma from [5].

**Lemma 4.1.** ([5, Lemma 2.18].) *Let $(S_n = \sum_{i=1}^{n} \xi_i, \mathcal{F}_n : n \geq 1)$ be a martingale and let $(U_n : n \geq 1)$ be a nondecreasing sequence of positive random variables such that $U_n$ is $\mathcal{F}_{n-1}$-measurable. Then $U_n^{-1} S_n \to 0$ a.s. on the set*

$$\left\{ \lim U_n = \infty, \ \sum_{i=1}^{\infty} U_i^{-2} \, \mathrm{E}(\xi_i^2 \mid \mathcal{F}_{i-1}) < \infty \right\}.$$

**Corollary 4.1.** *On the explosion set $A$, we have, as $n \to \infty$,*

$$\frac{\sum_{i=1}^{n} Z_i}{\sum_{i=0}^{n-1} X_i} \to \pi \quad \text{a.s.}$$

This corollary is intuitively obvious because the denominator is the total number of individuals in the first $n$ generations (including generation 0), while the numerator is the number of observed individuals in these generations. Here is a formal proof.

*Proof of Corollary 4.1.* Write

$$\frac{\sum_{i=1}^{n} Z_i}{\sum_{i=0}^{n-1} X_i} = \frac{\sum_{i=1}^{n} (Z_i - \pi X_{i-1})}{\sum_{i=0}^{n-1} X_i} + \pi.$$

We define

$$U_n = \sum_{i=0}^{n-1} X_i, \qquad \xi_i = Z_i - \pi X_{i-1}, \qquad S_n = \sum_{i=1}^{n} \xi_i.$$

Note that $U_n$ is $\mathcal{F}_{n-1}$-measurable, where $\mathcal{F}_{n-1}$ is the $\sigma$-algebra generated by $X_0, X_1, \ldots, X_{n-1}$, $Z_1, \ldots, Z_{n-1}$. Furthermore,

$$
\begin{aligned}
\sum_{i=1}^{\infty} U_i^{-2} \, \mathrm{E}(\xi_i^2 \mid \mathcal{F}_{i-1}) &= \sum_{i=1}^{\infty} \frac{\mathrm{E}((Z_i - \pi X_{i-1})^2 \mid \mathcal{F}_{i-1})}{(\sum_{j=0}^{i-1} X_j)^2} \\
&= \sum_{i=1}^{\infty} \frac{\pi(1-\pi)X_{i-1}}{(\sum_{j=0}^{i-1} X_j)^2} \\
&\le \sum_{i=1}^{\infty} \frac{\pi(1-\pi)}{X_{i-1}}.
\end{aligned}
$$

The last sum is a.s. finite on $A$, because $X_i$ is strictly positive and a.s. grows exponentially in $i$. So the set

$$
\left\{ \lim U_n = \infty, \ \sum_{i=1}^{\infty} U_i^{-2} \, \mathrm{E}(\xi_i^2 \mid \mathcal{F}_{i-1}) < \infty \right\}
$$

contains $A$ up to a set of measure 0. Now we may apply Lemma 4.1, and conclude that

$$
\frac{\sum_{i=1}^{n} Z_i}{\sum_{i=0}^{n-1} X_i} \to 0 + \pi = \pi \quad \text{a.s. on } A,
$$

as $n \to \infty$.

*Proof of Theorem 3.1(b).* From [4, Theorem 2.1], we have $(\sum_{i=1}^{n} X_i)/(\sum_{i=0}^{n-1} X_i) \to m$ a.s. on $A$. We apply Corollary 4.1, giving

$$
\frac{\sum_{i=2}^{n+1} Z_i}{\sum_{i=1}^{n} Z_i} = \frac{(\sum_{i=2}^{n+1} Z_i)/(\sum_{i=1}^{n} X_i)}{(\sum_{i=1}^{n} Z_i)/(\sum_{i=0}^{n-1} X_i)} \frac{\sum_{i=1}^{n} X_i}{\sum_{i=0}^{n-1} X_i} \to \frac{\pi}{\pi} m = m \quad \text{a.s. on } A,
$$

which proves the theorem.

The rate of convergence follows from Theorem 3.2. We use part of [4, Theorem 2.3] as a lemma to prove Theorem 3.2.

**Lemma 4.2.** *Assume that $m > 1$ and let $Y$ be a standard normal random variable independent of $X_n$. For any $x$, we have*

$$
\mathrm{P}\left( \frac{1}{\sigma} \left( \sum_{i=1}^{n} X_{i-1} \right)^{1/2} \left( \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_{i-1}} - m \right) \le x \ \Big| \ X_n > 0 \right) \to \mathrm{P}(Y \le x).
$$

In the same way, we can prove that

$$
\mathrm{P}\left( \frac{1}{\sqrt{\pi(1-\pi)}} \left( \sum_{i=1}^{n} X_{i-1} \right)^{1/2} \left( \frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n} X_{i-1}} - \pi \right) \le x \ \Big| \ X_n > 0 \right) \to \mathrm{P}(Y \le x).
$$

*Proof of Theorem 3.2.* First we rewrite $(\sum_{i=1}^{n} Z_i)^{1/2}(\tilde{m}_n - m)$ as

$$
\left(\sum_{i=1}^{n} Z_i\right)^{1/2} (\tilde{m}_n - m) = \left(\sum_{i=1}^{n} Z_i\right)^{-1/2} \sum_{i=1}^{n} (Z_{i+1} - m Z_i)
$$

$$
= \left(\sum_{i=1}^{n} Z_i\right)^{-1/2} \left[\sum_{i=1}^{n} (Z_{i+1} - \pi X_i) + \pi \sum_{i=1}^{n} (X_i - m X_{i-1})\right.
$$

$$
\left. - m \sum_{i=1}^{n} (Z_i - \pi X_{i-1})\right]. \tag{4.3}
$$

We already know that, on $A$, $(\sum_{i=1}^{n} Z_i)/(\sum_{i=1}^{n} X_{i-1})$ converges a.s. to the constant $\pi$ and $(\sum_{i=1}^{n} X_i)/(\sum_{i=1}^{n} X_{i-1})$ converges a.s. to the constant $m$. Now, the second term on the right-hand side of (4.3) can be rewritten as

$$
\pi \left(\frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n} X_{i-1}}\right)^{-1/2} \left(\sum_{i=1}^{n} X_{i-1}\right)^{-1/2} \sum_{i=1}^{n} (X_i - m X_{i-1})
$$

$$
= \pi \left(\frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n} X_{i-1}}\right)^{-1/2} \left(\sum_{i=1}^{n} X_{i-1}\right)^{1/2} \left(\frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_{i-1}} - m\right).
$$

From Lemma 4.2, we see that the second term converges in distribution to a normal distribution with zero mean and finite variance. We can treat the other terms on the right-hand side in the same way, proving the theorem.

## 5. Estimating a second function of the parameters

Next we want to prove Theorem 3.1(c), which gives a consistent estimator for $\gamma$. We first do this for the special case where $m_- = \sigma_-^2 = 0$. Then we treat the general case, and finally we interpret the function of the parameters that we can estimate.

### 5.1. The case $m_- = \sigma_-^2 = 0$

If $m_- = \sigma_-^2 = 0$ then $\gamma = (1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2$. Theorem 3.1(c) now reads as follows.

**Theorem 5.1.** *As $n \to \infty$, we have*

$$
n^{-1} \sum_{i=1}^{n} (Z_i + 1) \left(\frac{Z_{i+1}}{Z_i + 1} - \tilde{m}_n\right)^2 \to (1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2
$$

*in probability on $A$.*

We first compute

$$
\mathrm{E}\left((Z_1 + 1)\left(\frac{Z_2}{Z_1 + 1} - m\right)^2 \,\Big|\, X_0 = k\right).
$$

To do this, we recall from (4.1) that

$$
\mathrm{E}\left(\frac{1}{Z_1 + 1} \,\Big|\, X_0 = k\right) = \frac{1}{\pi(k+1)}(1 - (1-\pi)^{k+1}),
$$

and note that elementary computations yield

$$\mathrm{E}\left(\frac{X_1}{Z_1 + 1} \,\bigg|\, X_0 = k\right) = \frac{m}{\pi}(1 - (1 - \pi)^k),$$

$$\mathrm{E}\left(\frac{(X_1)^2}{Z_1 + 1} \,\bigg|\, X_0 = k\right) = \frac{m^2 k}{\pi}(1 - (1 - \pi)^{k-1}) + \frac{m^2 + \sigma^2}{\pi}(1 - (1 - \pi)^k).$$

Now we can compute the desired expectation in a straightforward way:

$$\mathrm{E}\left((Z_1 + 1)\left(\frac{Z_2}{Z_1 + 1} - m\right)^2 \,\bigg|\, X_0 = k\right)$$

$$= \mathrm{E}\left(\frac{\pi(1 - \pi)X_1}{Z_1 + 1} \,\bigg|\, X_0 = k\right) + \mathrm{E}((Z_1 + 1)^{-1}(\pi X_1 - m(Z_1 + 1))^2 \mid X_0 = k)$$

$$= ((1 - \pi)m + \pi\sigma^2 + \pi m^2)(1 - (1 - \pi)^k) + m^2(1 - k\pi(1 - \pi)^{k-1})$$

$$= ((1 - \pi)m + \pi\sigma^2 + \pi m^2)\,\mathrm{P}(Z_1 \neq 0 \mid X_0 = k) + m^2\,\mathrm{P}(Z_1 \neq 1 \mid X_0 = k).$$

With this expression in hand, we can identify a suitable martingale. We denote by $\mathcal{F}_n$ the $\sigma$-algebra generated by $\{Z_i : 1 \leq i \leq 2n\}$.

**Lemma 5.1.** (The process of random variables.)

$$M_n := \sum_{j=1}^{n}\left((Z_{2j-1} + 1)\left(\frac{Z_{2j}}{Z_{2j-1} + 1} - m\right)^2\right.$$

$$\left. - [((1 - \pi)m + \pi\sigma^2 + \pi m^2)\,\mathbf{1}_{\{Z_{2j-1} > 0\}} + m^2\,\mathbf{1}_{\{Z_{2j-1} \neq 1\}}]\right)$$

*is a martingale with respect to $\mathcal{F}_n$.*

*Proof.* It is clear that $M_n$ is measurable with respect to $\mathcal{F}_n$. Let $\xi_{n+1} := M_{n+1} - M_n$ denote the increments and note that $\mathrm{E}(\xi_{n+1} \mid X_{2n}, \mathcal{F}_n) = \mathrm{E}(\xi_{n+1} \mid X_{2n}) = 0$, where the last equality follows from the computation just above this lemma. Hence,

$$\mathrm{E}(M_{n+1} \mid \mathcal{F}_n) = M_n + \mathrm{E}(\mathrm{E}(\xi_{n+1} \mid X_{2n}, \mathcal{F}_n) \mid \mathcal{F}_n) = M_n.$$

**Theorem 5.2.** *As $n \to \infty$, we have*

$$n^{-1}\sum_{j=1}^{n}(Z_{2j-1} + 1)\left(\frac{Z_{2j}}{Z_{2j-1} + 1} - m\right)^2 \to (1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2 \quad a.s.\ on\ A.$$

*Furthermore, writing*

$$\tilde{S}_n(m) := \sum_{i=1}^{n}(Z_i + 1)\left(\frac{Z_{i+1}}{Z_i + 1} - m\right)^2,$$

*we have*

$$n^{-1}\tilde{S}_n(m) \to (1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2 \quad a.s.\ on\ A.$$

*Proof.* From Lemma 5.1, it follows that $M_n$ is a martingale with respect to $\mathcal{F}_n$, with increments

$$
\begin{aligned}
\xi_j := (Z_{2j-1} + 1)\left(\frac{Z_{2j}}{Z_{2j-1}+1} - m\right)^2 \\
- [((1-\pi)m + \pi\sigma^2 + \pi m^2)\mathbf{1}_{\{Z_{2j-1}>0\}} + m^2\,\mathbf{1}_{\{Z_{2j-1}\neq 1\}}].
\end{aligned}
$$

Now we apply Lemma 4.1 with the given $\xi_j$ and $U_n = n$. On the set $A$, we have $U_n \to \infty$. To show that $\sum_{i=1}^{\infty} U_i^{-2} \, \mathrm{E}(\xi_i^2 \mid \mathcal{F}_{i-1}) < \infty$ on $A$, we claim that there exists a constant $C < \infty$ such that

$$
\mathrm{var}\left((Z_1 + 1)\left(\frac{Z_2}{Z_1+1} - m\right)^2 \,\Bigg|\, X_0 = k\right) < C,
$$

uniformly in $k$. The computations that justify this claim are lengthy but straightforward, and are omitted (see the appendix of [9, Chapter 3] for details). Now we have

$$
\begin{aligned}
\mathrm{E}(\xi_i^2 \mid \mathcal{F}_{i-1}) &= \mathrm{var}\left((Z_{2i-1} + 1)\left(\frac{Z_{2i}}{Z_{2i-1}+1} - m\right)^2 \,\Bigg|\, \mathcal{F}_{i-1}\right) \\
&= \mathrm{E}\left(\mathrm{var}\left((Z_{2i-1} + 1)\left(\frac{Z_{2i}}{Z_{2i-1}+1} - m\right)^2 \,\Bigg|\, X_{2(i-1)}\right)\,\Bigg|\, \mathcal{F}_{i-1}\right) \\
&< C,
\end{aligned}
$$

and we conclude that

$$
\frac{M_n}{n} \to 0.
$$

Now write $n^{-1}\bar{M}_n = n^{-1}\sum_{j=1}^{n}\bar{\xi}_j$, where

$$
\bar{\xi}_j = (Z_{2j} + 1)\left(\frac{Z_{2j+1}}{Z_{2j}+1} - m\right)^2 - [((1-\pi)m + \pi\sigma^2 + \pi m^2)\mathbf{1}_{\{Z_{2j}>0\}} + m^2\,\mathbf{1}_{\{Z_{2j}\neq 1\}}]
$$

are the martingale increments. Define $\bar{\mathcal{F}}_j$ to be the $\sigma$-algebra generated by $\{Z_1, \ldots, Z_{2j+1}\}$. Now, using the same arguments as for the a.s. convergence of $n^{-1}M_n$, we may prove that $n^{-1}\bar{M}_n \to 0$ a.s. on $A$. Finally, note that

$$
\frac{1}{2n}\sum_{i=1}^{2n}(Z_i + 1)\left(\frac{Z_{i+1}}{Z_i+1} - m\right)^2 - [((1-\pi)m + \pi\sigma^2 + \pi m^2)\mathbf{1}_{\{Z_{2j-1}>0\}} + m^2\,\mathbf{1}_{\{Z_{2j-1}\neq 1\}}]
$$

is equal to

$$
\frac{1}{2n}(M_n + \bar{M}_n).
$$

The second result of the theorem now follows.

We remark that

$$
\sum_{i=1}^{n}(Z_i + 1)\left(\frac{Z_{i+1}}{Z_i+1} - m\right)^2 - [((1-\pi)m + \pi\sigma^2 + \pi m^2)\mathbf{1}_{\{Z_i>0\}} + m^2\,\mathbf{1}_{\{Z_i\neq 1\}}]
$$

is not a martingale itself, so we cannot use Lemma 4.1 directly.

Because we do not know $m$, we cannot use $\tilde{S}_n(m)$ for estimation purposes, and we also need to analyse the behaviour of $\tilde{S}_n(\tilde{m}_n)$. Some algebra yields

$$n^{-1}(\tilde{S}_n(m) - \tilde{S}_n(\tilde{m}_n)) = n^{-1}(m - \tilde{m}_n)^2 \sum_{i=1}^{n} Z_i + (m^2 - \tilde{m}_n^2). \tag{5.1}$$

From Theorem 3.2, we know that the square root of $(m - \tilde{m}_n)^2 \sum_{i=1}^{n} Z_i$ is the sum of three random variables, each converging in distribution to a normally distributed random variable with finite variance. The square root of the first term on the right-hand side thus converges in distribution to 0. Because 0 is a constant, the convergence also holds in probability. If $A_n \to 0$ in probability then $A_n^2 \to 0$ in probability, so $n^{-1}(m - \tilde{m}_n)^2 \sum_{i=1}^{n} Z_i$ converges in probability to 0 on $A$. Together with Theorem 5.2, this proves Theorem 5.1.

### 5.2. The general case

Up until now, we have considered the situation where the observed individuals have no further offspring. We now allow observed individuals to have some $X$-offspring in the generation after the observation. Hence, in terms of epidemics, in this section we allow detected individuals to infect other individuals during the interval of detection. Theorem 3.1 gives us a consistent estimator (in probability) of $(1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2 - 2\pi m m_-$ on the explosion set $A$.

To prove the theorem, we can compute that (again omitting lengthy details)

$$
\begin{aligned}
\mathrm{E}\Bigg( (Z_1 + 1)&\left(\frac{Z_2}{Z_1 + 1} - m\right)^2 - [(1 - \pi)m + \pi\sigma^2 + \pi m^2]\mathbf{1}_{\{Z_1 > 0\}} \\
&\qquad\qquad - m^2\,\mathbf{1}_{\{Z_1 \neq 1\}} + 2\pi m(m_-)\,\mathbf{1}_{\{Z_1 > 1\}} \;\Bigg|\; X_0 = k \Bigg) \\
+ \mathrm{E}\Bigg( &\frac{\pi[\pi\sigma_-^2 + (1 - \pi)(m_-)]}{Z_1 + 1}\mathbf{1}_{\{Z_1 > 0\}} - \frac{\pi^2(m_-)^2}{Z_1 + 1}(\mathbf{1}_{\{Z_1 > 1\}} - \mathbf{1}_{\{Z_1 = 1\}}) \;\Bigg|\; X_0 = k \Bigg) \\
&= 0.
\end{aligned}
$$

This leads to the following lemma, which can be proved in the same way as Lemma 5.1.

**Lemma 5.2.** *Let $\mathcal{F}_n$ be the $\sigma$-algebra generated by $\{Z_i : 1 \leq i \leq 2n\}$. Then the sequence of random variables*

$$
\begin{aligned}
M_n := \sum_{j=1}^{n}\Bigg[ &(Z_{2j-1} + 1)\left(\frac{Z_{2j}}{Z_{2j-1} + 1} - m\right)^2 \\
&- [(1 - \pi)m + \pi\sigma^2 + \pi m^2]\mathbf{1}_{\{Z_{2j-1} > 0\}} - m^2\,\mathbf{1}_{\{Z_{2j-1} \neq 1\}} \\
&+ 2\pi m(m_-)\,\mathbf{1}_{\{Z_{2j-1} > 1\}} + \frac{\pi[\pi\sigma_-^2 + (1 - \pi)(m_-)]}{Z_{2j-1} + 1}\mathbf{1}_{\{Z_{2j-1} > 0\}} \\
&- \frac{\pi^2(m_-)^2}{Z_{2j-1} + 1}(\mathbf{1}_{\{Z_{2j-1} > 1\}} - \mathbf{1}_{\{Z_{2j-1} = 1\}})\Bigg]
\end{aligned}
$$

*is a martingale with respect to $\mathcal{F}_n$.*

Using this lemma, we can now prove our next theorem.

**Theorem 5.3.** *Let $\gamma$ and $\tilde{S}_n(m)$ be as in Theorem 3.1. As $n \to \infty$, we have*

$$n^{-1} \sum_{j=1}^{n} (Z_{2j-1} + 1) \left( \frac{Z_{2j}}{Z_{2j-1} + 1} - m \right)^2 \to \gamma \quad a.s. \text{ on } A.$$

*Furthermore, we have*

$$n^{-1} \tilde{S}_n(m) \to \gamma \quad a.s. \text{ on } A.$$

We can prove this in the same way as we did for the special case of the previous subsection. The only extra thing to prove is that there exists a $C < \infty$ such that

$$\text{var} \left( (Z_1 + 1) \left( \frac{Z_2}{Z_1 + 1} - m \right)^2 \,\bigg|\, X_0 = k \right) < C$$

for all $k$ in the general case, too. Again, the computations that verify this inequality are lengthy but straightforward, and are omitted (see the appendix of [9, Chapter 3] for details).

The remainder of the argument proving Theorem 3.1(c) now follows exactly as in the previous section.

It seems rather difficult to establish the rate of convergence for the estimator of the second function of the parameters, but Theorem 3.3(a) gives a bound for this rate.

*Proof of Theorem 3.3(a).* From [5, Corollary 3.1] it follows that if $\{S_{n,i}, \mathcal{F}_{n,i} : 1 \le i \le n\}$ is a square-integrable martingale array with differences $X_{n,i} := S_{n,i} - S_{n,i-1}$ (where $S_{n,0}$ is defined as 0) and such that

$$\sum_{i=1}^{n} \text{E}(X_{n,i}^2 \mid \mathcal{F}_{n,i-1}) \to 0 \quad \text{in probability}$$

and $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ for $1 \le i \le n$, then $S_{n,n} = \sum_{i=1}^{n} X_{n,i} \to 0$ in probability.

We use this with $S_{n,n} = n^{-(1/2+\delta)} M_n$, where $M_n$ is as in Lemma 5.2, and $X_{n,i} = n^{-(1/2+\delta)} T_i$, where $T_i$ are the summands of $M_n$. We define $\mathcal{F}_{n,i}$, $1 \le i \le n$, to be the $\sigma$-algebra generated by $\{Z_i : 1 \le i \le 2n\}$, so $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ holds for $1 \le i \le n$. We have already shown that $\text{E}(T_i^2) < C$ for some $C$, uniformly in $i$. Hence,

$$\sum_{i=1}^{n} \text{E}(X_{n,i}^2 \mid \mathcal{F}_{n,i-1}) \le n^{-(1+2\delta)} nC \to 0 \quad \text{in probability},$$

so $n^{-(1/2+\delta)} M_n \to 0$ in probability. In the same way, we can prove that $n^{-(1/2+\delta)} \bar{M}_n \to 0$, where $\bar{M}_n$ is as in the proof of Theorem 5.2. Now, by the definitions of $M_n$ and $\bar{M}_n$, we see that, on the explosion set $A$,

$$n^{1/2-\delta} (n^{-1} \tilde{S}_n(m) - \gamma) = n^{-(1/2+\delta)} M_n + n^{-(1/2+\delta)} \bar{M}_n + n^{-(1/2+\delta)} \sum_{i=1}^{n} f(Z_i),$$

where $f(x) = O(1/x)$. Since $Z_{i+1}/X_i \to \pi$ a.s. on $A$ and $m^{-i} X_i$ converges a.s. to a finite random variable, we know that $m^{-i} Z_i$ a.s. converges to an a.s. finite, positive random variable, $\bar{W}$ say. Now we use the Toeplitz lemma [4, Lemma 1.2] to see that

$$\sum_{i=1}^{\infty} f(Z_i) < \left( \sum_{i=1}^{\infty} m^{-i} \right) C \frac{1}{\bar{W}} < \infty,$$

for some $C$. Thus, $n^{1/2-\delta} (n^{-1} \tilde{S}_n(m) - \gamma) \to 0$ in probability on the explosion set $A$.

By using the rate of convergence of $\tilde{m}_n$, (5.1), and the arguments following (5.1), we see that, for all $\delta_1 > 0$, $n^{-\delta_1} \tilde{S}_n(m) - n^{-\delta_1} \tilde{S}_n(\tilde{m}_n) \to 0$ in probability on $A$. The theorem now follows, with $\delta_1 = \frac{1}{2} + \delta$.

### 5.3. Interpretation of $\gamma$

The expression

$$\gamma = (1 - \pi)m + \pi\sigma^2 + \pi m^2 + m^2 - 2\pi m m_-,$$

which appeared as the limit in the previous subsection, turns out to have a somewhat surprising interpretation: it appears as a second moment if we treat our sequence of observations itself as a Galton–Watson process. To explain what we mean by this, we compute $P(X_0 = l \mid Z_1 = k)$ when the a-priori distribution of $X_0$ is uniform on the integers between $k$ and $N$, $N \gg k$. After this we let $N$ tend to $\infty$. We obtain

$$P(X_0 = l \mid Z_1 = k) = \frac{P(Z_1 = k \mid X_0 = l)\, P(X_0 = l)}{\sum_{i=k}^{N} P(Z_1 = k \mid X_0 = i)\, P(X_0 = i)} = \frac{P(Z_1 = k \mid X_0 = l)}{\sum_{i=k}^{N} P(Z_1 = k \mid X_0 = i)}.$$

First we compute the denominator as $N \to \infty$:

$$\lim_{N \to \infty} \sum_{i=k}^{N} P(Z_1 = k \mid X_0 = i) = \sum_{i=k}^{\infty} \binom{i}{k} \pi^k (1 - \pi)^{i-k}$$

$$= \frac{1}{\pi} \sum_{j=0}^{\infty} \binom{k + j}{j} \pi^{k+1} (1 - \pi)^j.$$

The summands are exactly the probabilities of a negative binomial distribution with parameters $k + 1$ and $\pi$, so $\sum_{j=0}^{\infty} \binom{k+j}{j} \pi^{k+1} (1 - \pi)^j = 1$. Therefore, we know that the denominator converges to $\pi^{-1}$.

With some abuse of notation, we will henceforth write a superscript '$*$' when we discuss probabilities and the accompanying expectations and variances after taking the limit as $N \to \infty$. Thus,

$$P^*(X_0 = l \mid Z_1 = k) = \binom{l}{k} \pi^{k+1} (1 - \pi)^{l-k}.$$

It is now easy to compute that

$$E^*(X_0 \mid Z_1 = k) = \frac{k + 1}{\pi} - 1,$$

$$\text{var}^*(X_0 \mid Z_1 = k) = \frac{(1 - \pi)(k + 1)}{\pi^2},$$

and some straightforward computations yield

$$E^*(Z_2 \mid Z_1 = k) = (1 - \pi)m_+ + mk,$$

$$\text{var}^*(Z_2 \mid Z_1 = k) = [(1 - \pi)m + \pi\sigma^2 + (1 + \pi)m^2 - 2\pi m m_-](k + 1)$$
$$- (1 - \pi)\pi m_- - \pi^2 \sigma_-^2.$$

We see that $E^*(Z_2 \mid Z_1 = k) = mk + O(1)$ as $k \to \infty$ and $\text{var}^*(Z_2 \mid Z_1 = k) = \gamma k + O(1)$ as $k \to \infty$. In this sense, we again estimate a first and a second moment, just as in the classical case, where the full generation sizes are observed.

## 6. Estimating a third function of the parameters

From [4, Theorem 1.3], we know that we can estimate two moments of the offspring distribution of a Galton–Watson process and no other functions of the parameters consistently if only the generation sizes are observed. However, in our context, Theorem 3.1 implies that we can, under certain conditions, estimate a third function of the parameters.

We have already shown that, as $k \to \infty$,

$$\mathrm{E}\left(\frac{Z_2}{Z_1+1} \,\middle|\, X_0 = k\right) \to m,$$

$$\mathrm{E}\left((Z_1+1)\left(\frac{Z_2}{Z_1+1} - m\right)^2 \,\middle|\, X_0 = k\right) \to \gamma.$$

We next compute $\mathrm{E}((Z_1+1)(Z_3/(Z_1+1) - m^2)^2 \mid X_0 = k)$:

$$
\mathrm{E}\left((Z_1+1)\left(\frac{Z_3}{Z_1+1} - m^2\right)^2 \,\middle|\, X_0 = k\right)
$$

$$
= \mathrm{E}\left((Z_1+1)\left[\left(\frac{Z_3 - \pi X_2}{Z_1+1}\right)^2 + \pi^2\left(\frac{X_2 - mX_1}{Z_1+1}\right)^2 + m^2\left(\frac{\pi X_1}{Z_1+1} - m\right)^2\right] \,\middle|\, X_0 = k\right)
$$

$$
= \pi(1-\pi)m\,\mathrm{E}\left(\frac{X_1}{Z_1+1} \,\middle|\, X_0 = k\right) + \pi^2\sigma^2\,\mathrm{E}\left(\frac{X_1}{Z_1+1} \,\middle|\, X_0 = k\right)
$$

$$
+ m^2\,\mathrm{E}\left((Z_1+1)\left(\frac{Z_2}{Z_1+1} - m\right)^2 \,\middle|\, X_0 = k\right) - m^2\pi(1-\pi)\,\mathrm{E}\left(\frac{X_1}{Z_1+1} \,\middle|\, X_0 = k\right)
$$

$$
= ((1-\pi)m + \pi\sigma^2 - m^2(1-\pi))\,\mathrm{E}\left(\frac{Z_2}{Z_1+1} \,\middle|\, X_0 = k\right)
$$

$$
+ m^2\,\mathrm{E}\left((Z_1+1)\left(\frac{Z_2}{Z_1+1} - m\right)^2 \,\middle|\, X_0 = k\right).
$$

Now note that, on $A$,

$$
\mathrm{E}\left((Z_1+1)\left(\frac{Z_3}{Z_1+1} - m^2\right)^2 \,\middle|\, X_0 = k\right)
$$

$$
= ((1-\pi)m + \pi\sigma^2 - m^2(1-\pi))\,\mathrm{E}\left(\frac{Z_2}{Z_1+1} \,\middle|\, X_0 = k\right)
$$

$$
+ m^2\,\mathrm{E}\left((Z_1+1)\left(\frac{Z_2}{Z_1+1} - m\right)^2 \,\middle|\, X_0 = k\right)
$$

$$
\to (\gamma - 2m^2 + 2\pi mm_-)m + m^2\gamma
$$

$$
= (m^2 + m)\gamma - 2(1-\pi)m^2 m_+
$$

$$
= \gamma_*,
$$

where the convergence is almost sure. Recall that

$$
S_n^*(m) := \sum_{i=1}^{n}(Z_i+1)\left(\frac{Z_{i+2}}{Z_i+1} - m^2\right)^2.
$$

We can use the martingale argument used to prove Theorem 5.3 to prove the following result.

**Theorem 6.1.** *The quantity*

$$n^{-1} \sum_{i=1}^{n} \left[ \left( \left[ \frac{\sum_{j=2}^{n+1} Z_j}{\sum_{j=1}^{n} Z_j} \right]^2 + \frac{\sum_{j=2}^{n+1} Z_j}{\sum_{j=1}^{n} Z_j} \right) (Z_i + 1) \left( \frac{Z_{i+1}}{Z_i + 1} - \frac{\sum_{j=2}^{n+1} Z_j}{\sum_{j=1}^{n} Z_j} \right)^2 \right] - n^{-1} S_n^*(m)$$

*converges (on A) in probability to* $2(1 - \pi)m^2 m_+$. *Hence,*

$$n^{-1} S_n^*(m) \to \gamma_*$$

*in probability on A.*

In order to obtain an observable quantity ($m$ is not observable), we need to bound

$$n^{-1}(S_n^*(\tilde{m}_n) - S_n^*(m)).$$

We claim that

$$\tilde{S}_n^*(m) - \tilde{S}_n^*(\tilde{m}_n) \leq n(m^4 - (\tilde{m}_n)^4) + 2(Z_2 - \tilde{m}_{n+1} Z_1)[m^2 - (\tilde{m}_n)^2]$$

$$+ m(m + \tilde{m}_n)(m - \tilde{m}_n)^2 \left( \sum_{i=1}^{n} Z_i \right)$$

$$+ (m + \tilde{m}_n)(m - \tilde{m}_{n+1})^2 \left( \left( \sum_{i=1}^{n+1} Z_i \right) - \tilde{m}_n Z_1 \right).$$

This claim can be justified by a long but straightforward computation, which can be found in [9, Section 3.6].

Now, $n^{-1} Z_1$ and $n^{-1} Z_2$ converge a.s. to 0. We can use the arguments used to prove the convergence of $n^{-1}(S_n(\tilde{m}_n) - S_n(m))$ in Section 5 to see that $n^{-1}(S_n^*(\tilde{m}_n) - S_n^*(m))$ converges to 0 in probability, which proves Theorem 3.1(d).

Note that $2(1 - \pi)m^2 m_+$ is not necessarily determined by $m$ and $\gamma$, so we are able to estimate a third function of the parameters. Also note that if $m_- = 0$ then we are unable to estimate the third parameter in this way. However, if we look at the case where we observe a binomially distributed number of individuals from each generation, but observations do not influence the offspring distribution (that is, $m_+ = m_- = m$ and $\sigma_+^2 = \sigma_-^2 = \sigma^2$), then we have an estimator for $(1 - \pi)m^3$ and, because we have an estimator for $m$ that converges a.s., we are in theory able to estimate $\pi$ consistently.

## 7. Conclusions

(A) From [4, Theorem 1.3], we know that we cannot estimate more than two functions of the parameters (the first two moments) consistently if only the generation sizes, $X_n$, of a branching process are given. In [6] it was shown that if we observe only a binomial($X_n, \pi$)-distributed fraction of the generation sizes, then we can estimate two functions of the parameters consistently if $\pi$ is known. In this paper we have shown that, under certain conditions, we can estimate three functions of the parameters, even when we do not know $\pi$.

(B) For epidemiological purposes, we also want to estimate $\pi$, because this parameter gives an indication of how many individuals are infectious at a certain time, which may be important for implementing measures to counter the epidemic. In order to estimate this parameter in a

reasonable time, we apparently need more, and other, information. We can possibly obtain this information by using contact tracing, i.e. finding out what contacts were made by an individual before it was observed and which contact may have caused the infection. Sometimes it is possible to obtain experimental information about the time between the infection and the removal of an individual, from which we may also estimate $\pi$. Note that Becker and Hasofer were able to estimate $\pi$ and $\lambda$ in [2], but needed information about the number of infectious individuals at the time of estimation, which is typically not available.

## References

[1] ANDERSSON, H. AND BRITTON, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis* (Lecture Notes Statist. **151**). Springer, New York.

[2] BECKER, N. G. AND HASOFER, A. M. (1997). Estimation in epidemics with incomplete observations. *J. R. Statist. Soc. B* **59,** 415–429.

[3] DIEKMANN, O. AND HEESTERBEEK, J. A. P. (2000). *Mathematical Epidemiology of Infectious Diseases. Model Building, Analysis and Interpretation*. John Wiley, Chichester.

[4] GUTTORP, P. (1991). *Statistical Inference for Branching Processes*. John Wiley, New York.

[5] HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

[6] JACOB, C. AND PECCOUD, J. (1998). Estimation of the parameters of a branching process from migrating binomial observations. *Adv. Appl. Prob.* **30,** 948–967.

[7] KLINKENBERG, D. (2003). Mathematical epidemiology and the control of classical swine fever virus. Doctoral Thesis, Utrecht University. Available at http://igitur-archive.library.uu.nl/dissertations/2003-0828-125441/full.pdf.

[8] MEESTER, R., DE KONING, J., DE JONG, M. C. M. AND DIEKMANN, O. (2002). Modelling and prediction of classical swine fever epidemics. *Biometrics* **41,** 178–184.

[9] TRAPMAN, P. (2006). On stochastic models for the spread of infections. Doctoral Thesis, Vrije Universiteit Amsterdam. Available at http://dare.ubvu.vu.nl/bitstream/1871/10202/1/7546.pdf.

[10] TRAPMAN, P., MEESTER, R. AND HEESTERBEEK, H. (2004). A branching model for the spread of infectious animal diseases in varying environments. *J. Math. Biol.* **49,** 553–576.