

# A CENTRAL LIMIT THEOREM FOR A DISCRETE-TIME SIS MODEL WITH INDIVIDUAL VARIATION

R. MCVINISH<sup>\*\*</sup> AND

P. K. POLLETT,<sup>\*\*\*</sup> *University of Queensland*

## Abstract

A discrete-time SIS model is presented that allows individuals in the population to vary in terms of their susceptibility to infection and their rate of recovery. This model is a generalisation of the metapopulation model presented in McVinish and Pollett (2010). The main result of the paper is a central limit theorem showing that fluctuations in the proportion of infected individuals around the limiting proportion converges to a Gaussian random variable when appropriately rescaled. In contrast to the case where there is no variation amongst individuals, the limiting Gaussian distribution has a nonzero mean.

*Keywords:* Epidemic modelling; fixed point; metapopulation modelling; weak convergence

2010 Mathematics Subject Classification: Primary 60F05  
Secondary 60J10

## 1. Introduction

Discrete-time Markov chains have been used by many researchers for modelling epidemics (see, for example, [3], [4], and [6]). In this note we consider an extension of the discrete-time susceptible-infected-susceptible (SIS) model [1]. For each  $n$ , let  $\{X_t^n\}_{t=1}^\infty$  be a discrete-time homogeneous Markov chain where

$$X_t^n = (X_{1,t}^n, \dots, X_{n,t}^n) \quad \text{with} \quad X_{i,t}^n = \begin{cases} 1 & \text{if individual } i \text{ is infected at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $B(m, p)$  denote a binomial random variable with  $m$  trials with success probability  $p$ . Conditional on  $X_t^n$ , the  $X_{i,t+1}^n$  are independent with transitions given by

$$X_{i,t+1}^n \sim B(X_{i,t}^n, s_i) + B\left(1 - X_{i,t}^n, c_i f\left(n^{-1} \sum_{j=1}^n X_{j,t}^n\right)\right), \quad (1)$$

where  $f: [0, 1] \mapsto [0, 1]$  is an increasing concave function. The parameter  $s_i$  is the probability that individual  $i$  does not recover from the infection in one time step, given that he/she is infected. The parameter  $c_i$  relates to the susceptibility to infection of individual  $i$  and may take values in  $[0, 1/f(1)]$ . However, to simplify notation, we will assume that  $c_i$  takes values in  $[0, 1]$ . In the special case where  $f(x) = x$ ,  $s_i = s$ , and  $c_i = c$  for all  $i = 1, \dots, n$ , the

Received 9 December 2010; revision received 15 December 2011.

\* Postal address: School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia.

\*\* Email address: r.mcvinish@uq.edu.au

\*\*\* Email address: pkp@maths.uq.edu.au

model for the number of infected individuals reduces to the standard discrete-time SIS model. By allowing variation amongst individuals, we aim to have a more realistic model of epidemic dynamics.

Note that model (1) is also relevant to modelling metapopulation dynamics. In that context, the state vector  $X_t^n$  models the occupied patches in a metapopulation of  $n$  patches. The parameter  $s_i$  is the probability of the local population at patch  $i$  surviving and  $c_i$  is related to the probability of patch  $i$  being recolonised. By taking  $c_i = s_i$  for all  $i = 1, \dots, n$ , the model reduces to the colonisation-extinction (CE) model studied in [5].

For a model with this level of complexity, asymptotic analysis is a valuable tool for understanding the model behaviour. A basic assumption of the analysis is that the parameters  $\{s_i, c_i\}$  can be treated as random variables having properties similar to the weak law of large numbers and central limit theorem. Under a similar assumption, in [5] we derived the deterministic limit of the proportion of occupied habitat patches for the CE model and studied some of the properties of the limiting deterministic trajectory. The main contribution presented in this note is an analysis of the fluctuations of the proportion of infected individuals about the limiting deterministic function. We show that these fluctuations, appropriately scaled, converge in distribution to a Gaussian random variable. However, the mean of the Gaussian distribution is nonzero and depends largely on the distribution of  $\{s_i, c_i\}$ ,  $i = 1, \dots, n$ .

## 2. Central limit theorem

### 2.1. Preliminaries

We first extend some results from Section 2 of [5] to the more general setting of model (1). Consider the sequences of arrays

$$d_n(t; j, k) := n^{-1} \sum_{i=1}^n s_i^j c_i^k X_{i,t}^n \quad \text{and} \quad m_n(j, k) := n^{-1} \sum_{i=1}^n s_i^j c_i^k,$$

where  $j, k \in \{0, 1, 2, \dots\}$ . In epidemic modelling,  $d_n(t; 0, 0)$  is the proportion of infective individuals in a population of size  $n$  while, in the context of metapopulation modelling,  $d_n(t; 0, 0)$  is the proportion of occupied patches of a metapopulation comprising  $n$  patches. Throughout this note, we shall assume that there exists a probability measure  $\sigma$  on  $[0, 1]^2$  such that

$$m_n(j, k) \xrightarrow{P} m(j, k) := \int s^j c^k \sigma(ds, dc) \tag{2}$$

and

$$d_n(0; j, k) \xrightarrow{P} d(0; j, k), \tag{3}$$

where  $d(0; j, k)$  is an array of constants. Under assumptions (2) and (3),  $d_n(t; j, k)$  converges in probability to the solution of the recurrence equation

$$d(t + 1; j, k) = d(t; j + 1, k) + f(d(t; 0, 0))(m(j, k + 1) - d(t; j, k + 1)). \tag{4}$$

The proof of this result follows the same arguments as those used in the proof of Theorem 2.1 of [5]. The fixed points of the recurrence equation (4) are given by

$$d(j, k) = \int \frac{f(\psi)s^j c^{k+1}}{1 - s + f(\psi)c} \sigma(ds, dc), \tag{5}$$

where  $\psi$  solves

$$\psi = R(\psi) := \int \frac{f(\psi)c}{1 - s + f(\psi)c} \sigma(ds, dc). \tag{6}$$

If  $f(0) > 0$  then there exists a unique  $\psi > 0$  satisfying (6). If  $f(0) = 0$  and

$$f'(0) \int \frac{c}{1 - s} \sigma(ds, dc) \leq 1, \tag{7}$$

then  $\psi = 0$  is the unique solution to (6). Otherwise, (6) has two solutions of which one is  $\psi = 0$ . The derivation of the fixed points of the recurrence equation (4) follows the same arguments as those used in the proof of Theorem 2.2 of [5].

**2.2. Main result**

The main contribution of this note is to examine the fluctuations of  $d_n(t; j, k)$  about their deterministic limit  $d(t; j, k)$  given by (4). Define the sequences of arrays

$$z_n(t; j, k) := n^{1/2}(d_n(t; j, k) - d(t; j, k)) \quad \text{and} \quad \xi_n(j, k) := n^{1/2}(m_n(j, k) - m(j, k)).$$

The central limit theorem for  $z_n(t; j, k)$  given below differs from that presented in [2] due to the variations in  $s_i$  and  $c_i$ . If  $s_i = c_i = s$  for all  $i$  then  $\xi_n(j, k) = 0$  for all  $j, k, n$  and Theorem 4 (CE model) of [2] is recovered.

**Theorem 1.** *Assume that (2) and (3) hold and that  $f$  is a twice continuously differentiable function. Assume also that the infinite array  $\{z_n(0; j, k), \xi_n(j, k)\}$  converges in distribution to  $\{z(0; j, k), \xi(j, k)\}$ . Then  $z_n(t; j, k) \Rightarrow z(t; j, k)$ , where*

$$z(t + 1; j, k) = z(t; j + 1, k) + f(d(t; 0, 0))(\xi(j, k + 1) - z(t; j, k + 1)) + f'(d(t; 0, 0))(m(j, k + 1) - d(t; j, k + 1))z(t; 0, 0) + \varepsilon(t + 1; j, k) \tag{8}$$

and  $\varepsilon(t + 1; \cdot, \cdot)$  is a sequence of independent, Gaussian random arrays with mean 0 and covariance given by

$$\begin{aligned} E \varepsilon(t; j_1, k_1) \varepsilon(t; j_2, k_2) &= d(t; j_1 + j_2 + 1, k_1 + k_2) - d(t; j_1 + j_2 + 2, k_1 + k_2) \\ &\quad + f(d(t; 0, 0))(m(j_1 + j_2, k_1 + k_2 + 1) - d(t; j_1 + j_2, k_1 + k_2 + 1)) \\ &\quad - f^2(d(t; 0, 0))(m(j_1 + j_2, k_1 + k_2 + 2) - d(t; j_1 + j_2, k_1 + k_2 + 2)). \end{aligned} \tag{9}$$

*Proof.* Let  $Z_n(t; q)$  denote the array  $\{z_n(t; j, k)\}$ ,  $j = 0, \dots, q, k = 0, \dots, q$ , and let  $\Xi_n(q)$  denote the array  $\{\xi_n(j, k)\}$ ,  $j = 0, \dots, q, k = 0, \dots, q$ . For any  $q \geq 2t$ , let  $\Omega_n(t; q)$  denote

$$(Z_n(t; q - 2t), Z_n(t - 1; q - 2(t - 1)), \dots, Z_n(0; q), \Xi_n(q)).$$

The proof proceeds by showing that the characteristic function of  $\Omega_n(t; q)$  converges to the characteristic function of the desired Gaussian process given by (8) and (9). This is achieved using induction. By assumption,  $\Omega_n(0; q)$  converges in distribution for any  $q \geq 0$ . Now suppose that  $\Omega_n(T; q)$  converges in distribution for some  $T$  such that  $T < t$ . To compute the characteristic function of  $\Omega_n(T + 1; q)$ , we begin by computing the characteristic function of

$Z_n(T + 1; q - 2(T + 1))$  conditional on  $X_T^n$  and  $\{s_i, c_i\}$ ,  $i = 1, \dots, n$ . Let  $q' = q - 2(T + 1)$ . Then

$$\begin{aligned} & \mathbb{E}\left(\exp\left(i \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} z_n(T + 1; j, k)\right) \middle| X_T^n, \{s_i, c_i\}\right) \\ &= \mathbb{E}\left(\exp\left(-i \sum_{j=0}^{q'} \sum_{k=0}^{q'} n^{1/2} \beta_{jk} d(T + 1; j, k) \right. \right. \\ & \quad \left. \left. + i \sum_{i=1}^n \sum_{j=0}^{q'} \sum_{k=0}^{q'} n^{-1/2} \beta_{jk} s_i^j c_i^k X_{i,T+1}^n\right) \middle| X_T^n, \{s_i, c_i\}\right). \end{aligned} \tag{10}$$

Let  $\theta_i = \sum_{j=0}^{q'} \sum_{k=0}^{q'} n^{-1/2} \beta_{jk} s_i^j c_i^k$ , and let  $|\beta| = \max_{jk} |\beta_{jk}|$ . As the  $X_{i,T+1}^n$  are independent given  $X_T^n$ ,

$$\begin{aligned} & \mathbb{E}\left(\exp\left(i \sum_{i=1}^n \theta_i X_{i,T+1}^n\right) \middle| X_T^n, \{s_i, c_i\}\right) \\ &= \exp\left(\sum_{i=1}^n X_{i,T}^n \log(1 - s_i + s_i \exp(i\theta_i))\right) \\ & \quad \times \exp\left(\sum_{i=1}^n (1 - X_{i,T}^n) \log\left(1 - c_i f\left(n^{-1} \sum_{i=1}^n X_{i,T}^n\right) \right. \right. \\ & \quad \left. \left. + c_i f\left(n^{-1} \sum_{i=1}^n X_{i,T}^n\right) \exp(i\theta_i)\right)\right). \end{aligned} \tag{11}$$

We aim to express the two exponents in (11) in terms of the variables in  $\Omega_n(T; q)$  and other variables that converge in probability to constants as  $n \rightarrow \infty$ . The following estimates will be required:

$$\begin{aligned} & \left| \log(1 + z) - z + \frac{1}{2}z^2 \right| \leq |z|^3, \quad z \in \mathbb{C}, \quad |z| < \frac{1}{2}, \\ & \left| \exp(iz) - 1 - iz + \frac{1}{2}z^2 \right| \leq \min\left(|z|^2, \frac{|z|^3}{6}\right), \quad z \in \mathbb{R}. \end{aligned} \tag{12}$$

The second exponent in (11) can be expressed as

$$\begin{aligned} & \sum_{i=1}^n (1 - X_{i,T}^n) \log(1 - c_i f(d_n(T; 0, 0)) + c_i f(d_n(T; 0, 0)) \exp(i\theta_i)) \\ &= \sum_{i=1}^n (1 - X_{i,T}^n) \left( c_i f(d_n(T; 0, 0)) (\exp(i\theta_i) - 1) \right. \\ & \quad \left. - \frac{1}{2} c_i^2 f^2(d_n(T; 0, 0)) (\exp(i\theta_i) - 1)^2 + R_{i,n} \right), \end{aligned}$$

where, for sufficiently large  $n$ ,

$$|R_{i,n}| \leq |\exp(i\theta_i) - 1|^3 \leq |\theta_i|^3 \leq n^{-3/2} (q' + 1)^6 |\beta|^3.$$

Using the approximation of the exponential function, we can express the second exponent in (11) as

$$\sum_{i=1}^n (1 - X_{i,T}^n) \left( i c_i f(d_n(T; 0, 0)) \theta_i - \frac{1}{2} c_i f(d_n(T; 0, 0)) \theta_i^2 + \frac{1}{2} c_i^2 f^2(d_n(T; 0, 0)) \theta_i^2 + \tilde{R}_{i,n} \right), \tag{13}$$

where

$$\begin{aligned} \tilde{R}_{i,n} = & R_{i,n} + c_i f(d_n(T; 0, 0)) (\exp(i\theta_i) - 1 - i\theta_i + \frac{1}{2}\theta_i^2) \\ & - \frac{1}{2} c_i^2 f^2(d_n(T; 0, 0)) ((\exp(i\theta_i) - 1)^2 + \theta_i^2). \end{aligned}$$

Bound (12) implies that  $|\tilde{R}_{i,n}| \leq 3n^{-3/2}(q' + 1)^6 |\beta|^3$ . Expression (13) can now be written in terms of  $d_n(t; j, k)$  and  $m_n(j, k)$  as

$$\begin{aligned} & in^{1/2} f(d_n(T; 0, 0)) \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} (m_n(j, k + 1) - d_n(T; j, k + 1)) \\ & - \frac{1}{2} f(d_n(T; 0, 0)) \sum_{j_1=0}^{q'} \sum_{k_1=0}^{q'} \sum_{j_2=0}^{q'} \sum_{k_2=0}^{q'} \beta_{j_1 k_1} \beta_{j_2 k_2} (m_n(j_1 + j_2, k_1 + k_2 + 1) \\ & \qquad \qquad \qquad - d_n(T; j_1 + j_2, k_1 + k_2 + 1)) \\ & + \frac{1}{2} f^2(d_n(T; 0, 0)) \sum_{j_1=0}^{q'} \sum_{k_1=0}^{q'} \sum_{j_2=0}^{q'} \sum_{k_2=0}^{q'} \beta_{j_1 k_1} \beta_{j_2 k_2} (m_n(j_1 + j_2, k_1 + k_2 + 2) \\ & \qquad \qquad \qquad - d_n(T; j_1 + j_2, k_1 + k_2 + 2)) \\ & + R_n, \end{aligned} \tag{14}$$

where  $R_n = \sum_{i=1}^n (1 - X_{i,T}^n) \tilde{R}_{i,n}$  and  $|R_n| \leq 3n^{-1/2}(q' + 1)^6 |\beta|^3$ . Similar arguments yield the following expression for the first exponent in (11):

$$\begin{aligned} & in^{1/2} \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} d_n(T; j + 1, k) + R'_n \\ & - \frac{1}{2} \sum_{j_1=0}^{q'} \sum_{k_1=0}^{q'} \sum_{j_2=0}^{q'} \sum_{k_2=0}^{q'} \beta_{j_1 k_1} \beta_{j_2 k_2} (d_n(T; j_1 + j_2 + 1, k_1 + k_2) \\ & \qquad \qquad \qquad - d_n(T; j_1 + j_2 + 2, k_1 + k_2)). \end{aligned} \tag{15}$$

Here  $|R'_n| \leq 3n^{-1/2}(q' + 1)^3 |\beta|^3$ . Expressions (15) and (14) give the first and second exponents in (11), respectively. We now substitute these expressions into (10). We first group the terms in the exponent involving  $n^{1/2}$ . They are

$$\begin{aligned} & -i \sum_{j=0}^{q'} \sum_{k=0}^{q'} n^{1/2} \beta_{jk} d(T + 1; j, k) + in^{1/2} \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} d_n(T; j + 1, k) \\ & + in^{1/2} f(d_n(T; 0, 0)) \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} (m_n(j, k + 1) - d_n(T; j, k + 1)). \end{aligned} \tag{16}$$

Replacing  $d(T + 1; j, k)$  by its recursion (4) and taking a first-order Taylor expansion of  $f$ , expression (16) becomes

$$\begin{aligned}
 & i \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} n^{1/2} (d_n(T; j + 1, k) - d(T; j + 1, k)) \\
 & + i f(d(T; 0, 0)) \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} n^{1/2} (m_n(j, k + 1) - m(j, k + 1) - d_n(T; j, k + 1) \\
 & \qquad \qquad \qquad + d(T; j, k + 1)) \\
 & + i f'(d(T; 0, 0)) z_n(T; 0, 0) \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} (m_n(j, k + 1) - d_n(T; j, k + 1)) + R_n'', \tag{17}
 \end{aligned}$$

where

$$\begin{aligned}
 R_n'' &= i(n^{1/2}(f(d_n(T; 0, 0)) - f(d(T; 0, 0))) - f'(d(T; 0, 0))z_n(T; 0, 0)) \\
 & \quad \times \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} (m_n(j, k + 1) - d_n(T; j, k + 1)).
 \end{aligned}$$

Combining expression (17) with the expressions for the two exponents in (11) gives

$$\begin{aligned}
 & \log E \left( \exp \left( i \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} z_n(T + 1; j, k) \right) \middle| X_T^n, \{s_i, c_i\} \right) \\
 &= i \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} z_n(T; j + 1, k) \\
 & \quad + i f(d_n(T; 0, 0)) \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} (\xi_n(j, k + 1) - z_n(T; j, k + 1)) \\
 & \quad + i f'(d(T; 0, 0)) z_n(T; 0, 0) \sum_{j=0}^{q'} \sum_{k=0}^{q'} \beta_{jk} (m_n(j, k + 1) - d_n(T; j, k + 1)) \\
 & \quad - \frac{1}{2} \sum_{j_1=0}^{q'} \sum_{k_1=0}^{q'} \sum_{j_2=0}^{q'} \sum_{k_2=0}^{q'} \beta_{j_1 k_1} \beta_{j_2 k_2} (d_n(T; j_1 + j_2 + 1, k_1 + k_2) \\
 & \qquad \qquad \qquad - d_n(T; j_1 + j_2 + 2, k_1 + k_2)) \\
 & \quad - \frac{1}{2} f(d_n(T; 0, 0)) \sum_{j_1=0}^{q'} \sum_{k_1=0}^{q'} \sum_{j_2=0}^{q'} \sum_{k_2=0}^{q'} \beta_{j_1 k_1} \beta_{j_2 k_2} (m_n(j_1 + j_2, k_1 + k_2 + 1) \\
 & \qquad \qquad \qquad - d_n(T; j_1 + j_2, k_1 + k_2 + 1)) \\
 & \quad + \frac{1}{2} f^2(d_n(T; 0, 0)) \sum_{j_1=0}^{q'} \sum_{k_1=0}^{q'} \sum_{j_2=0}^{q'} \sum_{k_2=0}^{q'} \beta_{j_1 k_1} \beta_{j_2 k_2} (m_n(j_1 + j_2, k_1 + k_2 + 2) \\
 & \qquad \qquad \qquad - d_n(T; j_1 + j_2, k_1 + k_2 + 2)) \\
 & \quad + R_n + R_n' + R_n''. \tag{18}
 \end{aligned}$$

As  $f$  is twice, continuously differentiable and  $z_n(T; 0, 0)$  converges in distribution,  $R''_n$  converges in probability to 0. From (18) we see that the conditional characteristic function (10) is a bounded, continuous function of  $\Omega_n(T; q)$ ,  $d_n(T; j, k)$ ,  $m_n(j, k)$ , and the remainder terms  $R_n$ ,  $R'_n$ , and  $R''_n$ . As  $\Omega_n(T; q)$  converges in distribution and  $d_n(T; j, k)$ ,  $m_n(j, k)$ , and the remainder terms converge in probability to constants, these variables jointly converge in distribution. Therefore, the characteristic function of  $\Omega_n(T + 1; q)$  converges pointwise as  $n \rightarrow \infty$ . An application of Lévy’s continuity theorem completes the proof of convergence in distribution. The form of the limiting process given in (8) and (9) can be identified from the conditional characteristic function (18). This completes the proof.

**2.3. Mean of  $z$  conditional on  $\xi$**

Since  $\{s_i, c_i\}_{i=1}^n$  are essentially parameters of model (1), it seems natural to consider the distribution of  $z(t; j, k)$  conditional on  $\xi$ . Although properties of the process defined by (8) and (9), such as the mean and variance of  $z(t; j, k)$  conditional on  $\xi$ , can be investigated numerically, some analytical progress is possible if we impose the restriction that  $s_i = c_i$ . In this case  $z(t; k) = z(t; j, k - j)$ ,  $d(t; k) = d(t; j, k - j)$ ,  $d(k) = d(j, k - j)$ , and  $m(k) = m(j, k - j)$  for any  $j = 0, \dots, k$  and all  $k$ . Similar notation applies for all other arrays. Now suppose that the deterministic limit is in a nonzero equilibrium, that is,  $d(t; k) = d(k) > 0$  for all  $t$  and all  $k$ . Recall that  $d(0) = \psi$ , where  $\psi$  is the nonzero solution to (6). The mean of  $z(t; k)$  conditional on  $\xi$  is given by

$$E(z(t + 1; k) \mid \xi) = (1 - f(\psi))E(z(t; k + 1) \mid \xi) + f(\psi)\xi(k + 1) + f'(\psi)(m(k + 1) - d(k + 1))E(z(t; 0) \mid \xi). \tag{19}$$

From (5),

$$m(k) - d(k) = \int \frac{\lambda^k(1 - \lambda)}{1 - \lambda + \lambda f(\psi)}\sigma(d\lambda).$$

To determine the fixed points of (19), consider the difference equation

$$\mu(k) = (1 - f(\psi))\mu(k + 1) + f(\psi)\xi(k + 1) + \mu_0 f'(\psi) \int \frac{\lambda^{k+1}(1 - \lambda)}{1 - \lambda + \lambda f(\psi)}\sigma(d\lambda), \tag{20}$$

subject to the initial condition  $\mu(0) = \mu_0$ . Standard arguments show that the solution to (20) must be unique. It is easily verified that this solution is

$$\mu(k) = \chi(k) + \frac{\chi(0)}{1 - R'(\psi)} \int \frac{f'(\psi)(1 - \lambda)\lambda^{k+1}}{(1 - \lambda + f(\psi)\lambda)^2}\sigma(d\lambda), \tag{21}$$

where

$$\chi(k) = f(\psi) \sum_{r=1}^{\infty} (1 - f(\psi))^{r-1} \xi(k + r), \tag{22}$$

provided this infinite sum is well defined. Note that  $R'(\psi) < 1$  since  $R(x)$  is an increasing, concave function and  $\psi$  is the solution to  $\psi = R(\psi)$ . Therefore, the unique fixed point of (19) is given by the right-hand side of (21). Not only is (21) the unique fixed point, it is also a stable fixed point. To see this, substitute  $\mu(k) + \delta(t; k)$  for  $E(z(t; k) \mid \xi)$  in (19). This leads to the difference equation

$$\delta(t + 1; k) = (1 - f(\psi))\delta(t; k + 1) + f'(\psi)(m(k + 1) - d(k + 1))\delta(t; 0).$$

Define  $\hat{\delta}(t; x) = \sum_{k=0}^{\infty} |\delta(t; k)|x^k$  for  $x \in (0, 1]$ . Using the positivity of  $(m(k + 1) - d(k + 1))$

(see (5)) and the triangle inequality, we obtain the inequality

$$\hat{\delta}(t + 1; x) \leq \frac{(1 - f(\psi))\hat{\delta}(t; x)}{x} + |\delta(t; 0)| \left( f'(\psi) \left( \sum_{k=0}^{\infty} x^k (m(k + 1) - d(k + 1)) \right) - \frac{1 - f(\psi)}{x} \right).$$

From (5),

$$f'(\psi) \sum_{k=0}^{\infty} x^k (m(k + 1) - d(k + 1)) = f'(\psi) \int \frac{\lambda}{1 - \lambda x} \frac{1 - \lambda}{1 - \lambda + f(\psi)\lambda} \sigma(d\lambda),$$

which is equal to  $R'(\psi)$  for  $x = 1 - f(\psi)$ . As  $R'(\psi) < 1$ , it follows that

$$f'(\psi) \left( \sum_{k=0}^{\infty} x^k (m(k + 1) - d(k + 1)) \right) - \frac{1 - f(\psi)}{x} < 0$$

for  $x = 1 - f(\psi)$ . By continuity in  $x$ , there exists an  $x^* > 1 - f(\psi)$  such that

$$\hat{\delta}(t + 1; x^*) \leq \frac{(1 - f(\psi))\hat{\delta}(t; x^*)}{x^*}.$$

Letting  $t \rightarrow \infty$ ,  $\hat{\delta}(t; x^*) \rightarrow 0$ ; hence,  $\delta(t; k) \rightarrow 0$  for all  $k$ . We can conclude that  $\mu(k)$  is a stable fixed point of (19).

**2.4. Example**

We now illustrate the result of this note by comparing the distribution of the proportion of infected individuals in the simulated model (1) with the Gaussian distribution described in Theorem 1. Consider a population of size  $n = 50$ ,  $f(x) = 0.7x$ , and  $c_i = s_i$ , with the  $s_i$  generated independently from a Beta(0.6, 0.4) distribution. The initial state of the Markov chain is generated according to

$$P(X_{i,0}^n = 1 \mid s_i) = \frac{0.7\psi s_i}{1 - s_i + 0.7\psi s_i},$$

where  $\psi$  is the nonzero solution to (6). This method of generating the initial state ensures that the limiting deterministic process is in equilibrium.

Define  $z_n(t; 0) := z_n(t; 0, 0)$  and  $\xi_n(k) := \xi_n(k, 0)$ . To apply the Gaussian approximation, we need to evaluate the conditional mean  $E(z(t; 0) \mid \xi)$  and conditional variance  $\text{var}(z(t; 0) \mid \xi)$ . The conditional mean  $E(z(t; 0) \mid \xi)$  can be approximated by truncating the infinite sum in (22) and replacing the  $\xi(k)$  with  $\xi_n(k)$ . Note that the infinite sum defining  $\chi(k)$  converges quickly provided  $\psi$  is not very small. A recursion of the conditional variance can be derived from Theorem 1 and is given by

$$\begin{aligned} & \text{cov}(z(t + 1; j), z(t + 1; k) \mid \xi) \\ &= (1 - f(\psi))^2 \text{cov}(z(t; j + 1), z(t; k + 1) \mid \xi) \\ & \quad + (1 - f(\psi))f'(\psi)(m(j + 1) - d(j + 1)) \text{cov}(z(t; 0), z(t; k + 1) \mid \xi) \\ & \quad + (1 - f(\psi))f'(\psi)(m(k + 1) - d(k + 1)) \text{cov}(z(t; j + 1), z(t; 0) \mid \xi) \\ & \quad + (f'(\psi))^2(m(k + 1) - d(k + 1))(m(j + 1) - d(j + 1)) \text{cov}(z(t; 0), z(t; 0) \mid \xi) \\ & \quad + E \varepsilon(t + 1; j)\varepsilon(t + 1; k), \end{aligned}$$

where  $E \varepsilon(t; j)\varepsilon(t; k)$  is given by (9).

The process  $X_t^n$  was simulated and the behaviour of  $z_n(t; 0)$  compared with that predicted by the asymptotic theory. In Figure 1 the process  $z_n(t; 0)$  is plotted together with the approximation of the conditional mean and plus and minus two standard deviations. As the recursion for the conditional variance quickly converged, we simulated  $z_n(t; 0)$  up to time step  $10^3$  and compared its histogram with the approximating Gaussian distribution. This is given in Figure 2.

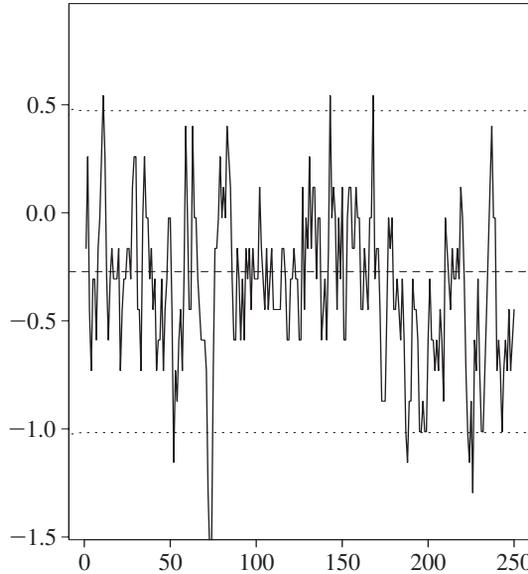


FIGURE 1: Plot of  $z_n(t; 0)$  for the first 250 time steps. The dashed line is the approximation to  $E(z(t; 0) | \xi)$ . The dotted lines are two standard deviations from the approximation to the conditional mean.

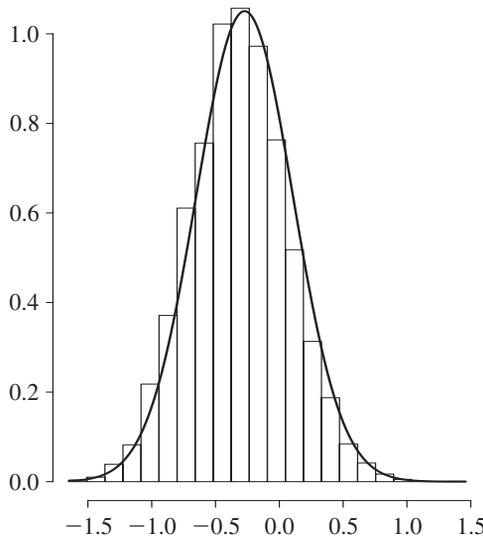


FIGURE 2: Histogram of  $z_n(t; 0)$  from the simulated process together with the approximating Gaussian density from Theorem 1.

From the figures we see that the asymptotic distribution provides a good approximation to the distribution of  $z_n(t; 0)$ .

### 3. Discussion

We have studied a discrete-time Markov chain that can be used to model SIS-type epidemics and also metapopulations. For this model, we have derived an asymptotic distribution for the fluctuations of the proportion of infected individuals/occupied patches around the deterministic limit. Our simulations suggest that the asymptotic distribution provides a good approximation even when the population size is small, provided  $\psi$  is not too small. When  $\psi$  is small, simulations not reported here indicate that the population size needs to be large for the asymptotic distribution to be a good approximation.

One problem for future study will be to examine the recursion for the asymptotic covariance matrix of  $z_n(t; 0)$ . Our numerical investigations suggest that this recursion will converge quickly to some limiting covariance matrix. It would be good to be able to show that the recursion has a unique stable fixed point.

### Acknowledgements

We are grateful to the anonymous referee for helpful comments and suggestions. This research was supported by the Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems.

### References

- [1] ALLEN, L. J. S. AND BURGIN, A. M. (2000). Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Math. Biosci.* **163**, 1–33.
- [2] BUCKLEY, F. M. AND POLLETT, P. K. (2010). Limit theorems for discrete-time metapopulation models. *Prob. Surveys* **7**, 53–83.
- [3] GANI, J., YAKOWITZ, Y. AND BLOUNT, M. (1997). The spread and quarantine of HIV infection in a prison system. *SIAM J. Appl. Math.* **57**, 1510–1530.
- [4] LEKONE, P. E. AND FINKENSTÄDT, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* **62**, 1170–1177.
- [5] MCVINISH, R. AND POLLETT, P. K. (2010). Limits of large metapopulations with patch-dependent extinction probabilities. *Adv. Appl. Prob.* **42**, 1172–1186.
- [6] TUCKWELL, H. C. AND WILLIAMS, R. J. (2007). Some properties of a simple stochastic epidemic model of SIR type. *Math. Biosci.* **208**, 76–97.