University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

School of Mathematical and Statistical
Sciences Faculty Publications and
Presentations

College of Sciences

10-31-2017

Multi-Type Branching Processes Modeling of Nosocomial Epidemics

Zeinab Mohamed

The University of Texas Rio Grande Valley

Tamer Oraby

The University of Texas Rio Grande Valley

Follow this and additional works at: https://scholarworks.utrgv.edu/mss_fac

Part of the [Mathematics Commons](#)

Recommended Citation

Mohamed, Z., & Oraby, T. (2017). Multi-Type Branching Processes Modeling of Nosocomial Epidemics. *Stochastics and Quality Control*, 32(2), 63–75. <https://doi.org/10.1515/eqc-2017-0026>

This Article is brought to you for free and open access by the College of Sciences at ScholarWorks @ UTRGV. It has been accepted for inclusion in School of Mathematical and Statistical Sciences Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

Research Article

Zeinab Mohamed and Tamer Oraby*

Multi-Type Branching Processes Modeling of Nosocomial Epidemics

<https://doi.org/10.1515/eqc-2017-0026>

Received October 30, 2017; revised October 31, 2017; accepted October 31, 2017

Abstract: Nosocomial epidemics are infectious diseases which spread among different types of susceptible individuals in a health-care facility. To model this type of epidemics, we use a multi-type branching process with a multivariate negative binomial offspring distribution. In particular, we estimate the basic reproduction number R_0 and study its relationship with the parameters of the offspring distribution. In case of a single-type epidemic, we investigate the effect of contact tracing on the estimates for R_0 .

Keywords: Branching Processes, Nosocomial Epidemics, Bivariate Negative Binomial, Basic Reproduction Number, Tracing

MSC 2010: 60G99

1 Introduction

Nosocomial epidemics, i.e., epidemics of disease spread in health care facility, cause burden on the health care systems. They also increase the risk of spreading communicable diseases through the visitors. Nosocomial epidemics involve three types of agents: patients (P), health care workers (HCW), and visitors (V). The latter type might be less common in epidemics, but cannot be ignored. Health care workers may also spread the disease between the wards and other units of hospitals if they do not properly follow the infection control procedures, like hand washing and mask usage. A detailed modeling of the epidemic is required to better understand the effect of those control measures. Of particular interest are the estimates of the basic reproduction number, R_0 . It is defined as the average number of secondary infections made by one infected person in a totally susceptible population. These estimates allow for a better assessment of both strength of the disease and effect of control measures on disease eradication.

Branching processes are widely used in modeling chemical [7], biological [8] and epidemiological systems [1]. Several researcher used spatial branching models to study species invasion, e.g., [11], and cancer growth [10]. Others used it in modeling disease epidemics, e.g., [1, 12–14]. Blumberg and Lloyd-Smith [3] studied the basic reproduction number using the distribution of the sizes of stuttering chains in which the negative binomial distribution is the offspring progeny distribution. Chowell, Abdirizak, Lee, Lee, Jung, Nishiura and Viboud [4] studied transmission trees of two nosocomial epidemics of MERS-CoV and SARS diseases using new stuttering trees and based on [3] in which there is an embedded assumption that $R_0 < 1$. Transmission tree reconstruction has also been studied by a number of researchers when the genetic data is available, e.g., [9, 15]. Genetic data, when available, can provide great information about the question of who infected whom [15].

Zeinab Mohamed, School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, Edinburg, USA,
e-mail: zienab.mohamed01@utrgv.edu

***Corresponding author: Tamer Oraby**, School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley,
Edinburg, USA, e-mail: tamer.oraby@utrgv.edu

In this paper, we consider the reconstruction of the transmission tree in the absence of genome data and for general values of R_0 . We still assume that progeny distribution is a negative binomial probability distribution. Yet, we discuss both single-type branching processes with univariate negative binomial distribution and two-type branching processes with bivariate negative binomial distribution. We are going to simulate both processes and discuss when the two-type epidemics dies out. We show an algorithm of back-tracing a single-type epidemic to reconstruct the tree. Two-type transmission tree reconstruction will be studied in the future.

2 Models and Methods

In this paper, we consider two kinds of nosocomial epidemics. The first kind is a single-type epidemic in which there is one type of infected people, who are patients in the health care facility. The second kind is a two-type epidemic in which there are two types of infected people, Type I (patients, P) and Type II (health-care workers, HCW). Both kinds can be modeled using single-type and two-type branching processes, respectively. Those branching processes are also called transmission trees; see Figure 1 for a depiction of a two-type transmission tree of a nosocomial epidemic. Usually those epidemics end as result of the intervention of the infection management in the health-care facility. We are interested in the epidemic all they way up to the last infection and exclude the zero infection due to intervention. We will present the model in the two-type nosocomial and then discuss the single-type at the end of the paper. In the analysis, we use probability generating functions, Monte Carlo simulations, and numerical algorithms.

Figure 1: Instance of a two-type nosocomial epidemic as a two-type transmission tree.

2.1 Modeling Two-Type Epidemics

We analyze the transmission dynamics of the infection under incomplete information, using branching processes. Denote the number type I and type II infectious in generation n by $X_1(n)$ and $X_2(n)$, respectively. The number of patients (offspring) infected by another patient is Y_{11} and by a HCW is Y_{12} . Finally, the number of HSW (offspring) infected by a patient is Y_{21} and by another HCW is Y_{22} .

2.1.1 Model

Suppose $X_1(0) = 1$, and $X_2(0) = 0$; that is, the epidemic begins with the admission of one infected patient into the hospital; called the primary case. Let $p_i(l_1, l_2)$ be the offspring probability distribution of an individual

of type i . That is, an individual of type i infects l_1 individuals of type I (patients) and l_2 individuals of type II (HCW) with probability $p_i(l_1, l_2) = \text{Prob}\{Y_{1i} = l_1, Y_{2i} = l_2\}$, where $l_j = 0, 1, 2, \dots$ and $j = 1, 2$. The offspring probability generating function (p.g.f.), assuming $X_i(0) = \delta_{i,1}$, is $f_i(\cdot, \cdot) : [0, 1]^2 \rightarrow [0, 1]$ given by

$$f_i(t_1, t_2) = \sum_{l_2=0}^{\infty} \sum_{l_1=0}^{\infty} p_i(l_1, l_2) t_1^{l_1} t_2^{l_2}, \quad i = 1, 2.$$

The choice of offspring distribution is important because it defines the relationship between the intensity and heterogeneity of the transmissions. We adopt a flexible framework by assuming that transmissions can be characterized by a bivariate negative binomial distribution [5] with p.g.f. for type I and type II given by

$$f_1(t_1, t_2) = \left(\frac{p_0}{1 - p_1 t_1 - p_2 t_2} \right)^k,$$

$$f_2(t_1, t_2) = \left(\frac{p'_0}{1 - p'_1 t_1 - p'_2 t_2} \right)^{k'}.$$

Parameter k is called the clustering or aggregation parameter. The larger its reciprocal is, the more clustered the infection will be. A key advantage of using negative binomial distribution [5, 6] rather than geometric or Poisson distributions is that the former distribution includes the later distributions as special or asymptotic cases, when $k = 1$ and as $k \rightarrow \infty$, respectively. The infection process ranges from $k = 0$ (absolute heterogeneity) to $k \rightarrow \infty$ (homogeneity). The case $k = 0$, assuming the random variable is positive, is the logarithmic distribution [2].

Environmental and behavioral characteristics of the health-care facility and agents (patients and HCW) are embedded in the probabilities and the clustering parameters. For instance, the probability of a patient infecting another patient (p_1) may be a function of the ventilation, distance between beds, cleaning, disposal of hazardous materials, social distance, etc. Parametric forms like a logistic function can be used for those probabilities making a hierarchical branching processes model of the epidemic; but we are not going to explore this here.

The mean (expected) matrix of the two-type branching process is given by

$$M = \begin{pmatrix} \frac{kp_1}{p_0} & \frac{kp_2}{p_0} \\ \frac{k'p'_1}{p'_0} & \frac{k'p'_2}{p'_0} \end{pmatrix},$$

where the (i, j) th element, M_{ij} , say, is the expected number of infected type j individuals by a type i individual.

The basic reproduction number R_0 which quantifies the expected number of infections caused by an infected individual is defined as the spectral radius of the mean matrix [1]. A transmission is called subcritical if $R_0 < 1$. In this case the epidemic will be eradicated with probability one [1].

Theorem 1. Consider a two-type branching process with a bivariate negative binomial offspring distribution. Then the process will die out with probability one if

$$\left(1 - \frac{k'p'_2}{p'_0}\right) \left(1 - \frac{kp_1}{p_0}\right) > \left(\frac{kp_2}{p_0} \cdot \frac{k'p'_1}{p'_0}\right),$$

where $k, k' > 0$, $0 < p_1, p_2, p'_1, p'_2 < 1$, $p_0 + p_1 + p_2 = 1$ and $p'_0 + p'_1 + p'_2 = 1$.

Proof. Since the mean matrix given by

$$M = \begin{pmatrix} \frac{kp_1}{p_0} & \frac{kp_2}{p_0} \\ \frac{k'p'_1}{p'_0} & \frac{k'p'_2}{p'_0} \end{pmatrix}$$

is regular, we have for its spectral radius $R_0 = \max\{|\lambda_1|, |\lambda_2|\}$, where λ_1 and λ_2 are the eigenvalues of the mean matrix. It is not difficult to calculate

$$R_0 = \frac{1}{2} \left\{ \frac{kp_1}{p_0} + \frac{k'p'_2}{p'_0} + \sqrt{\left(\frac{kp_1}{p_0} + \frac{k'p'_2}{p'_0} \right)^2 - 4 \left(\frac{kp_1}{p_0} \cdot \frac{k'p'_2}{p'_0} - \frac{kp_2}{p_0} \cdot \frac{k'p'_1}{p'_0} \right)} \right\},$$

which is less than one if and only if

$$\left(\frac{kp_1}{p_0} + \frac{k'p'_2}{p'_0}\right)^2 - 4\left(\frac{kp_1}{p_0} \cdot \frac{k'p'_2}{p'_0} - \frac{kp_2}{p_0} \cdot \frac{k'p'_1}{p'_0}\right) < \left(2 - \left(\frac{kp_1}{p_0} + \frac{k'p'_2}{p'_0}\right)\right)^2$$

or simply

$$\left(\frac{kp_1}{p_0} \cdot \frac{k'p'_2}{p'_0}\right) - \left(\frac{kp_2}{p_0} \cdot \frac{k'p'_1}{p'_0}\right) > \left(\frac{kp_1}{p_0} + \frac{k'p'_2}{p'_0}\right) - 1.$$

This yields the criterion

$$\left(1 - \frac{k'p'_2}{p'_0}\right)\left(1 - \frac{kp_1}{p_0}\right) > \left(\frac{kp_2}{p_0} \cdot \frac{k'p'_1}{p'_0}\right).$$

Note that

$$\left(\frac{kp_1}{p_0} + \frac{k'p'_2}{p'_0}\right)^2 - 4\left[\left(\frac{kp_1}{p_0} \cdot \frac{k'p'_2}{p'_0}\right) - \left(\frac{kp_2}{p_0} \cdot \frac{k'p'_1}{p'_0}\right)\right] > 0$$

and hence M has real eigenvalues. $\square$

This theorem implies that in a nosocomial epidemics both patients and HCW must be controlled applying appropriate measures. If the eigenvalue of either one is bigger than one, then there is a positive probability that the epidemic will not die out. Note that R_0 can be written as a function of the basic reproduction numbers of both patients and health-care workers as follows:

$$R_0 = \frac{1}{2} \left\{ R_0^{pp} + R_0^{hh} + \sqrt{(R_0^{pp} + R_0^{hh})^2 - 4(R_0^{pp} \cdot R_0^{hh} - R_0^{ph} \cdot R_0^{hp})} \right\},$$

where R_0^{ij} is the average number of type j infected from a single infected individual of type i and

$$(1 - R_0^{hh})(1 - R_0^{pp}) > R_0^{ph} R_0^{hp}.$$

The relationship between the parameters $k, k', p_1, p_2, p'_1, p'_2$ and R_0^{ph} are as follows:

$$p_1 = \frac{R_0^{pp}}{k + R_0^{pp} + R_0^{ph}},$$

$$p_2 = \frac{R_0^{ph}}{k + R_0^{pp} + R_0^{ph}}.$$

Similarly, we obtain

$$p'_1 = \frac{R_0^{hp}}{k' + R_0^{hp} + R_0^{hh}},$$

$$p'_2 = \frac{R_0^{hh}}{k' + R_0^{hp} + R_0^{hh}}.$$

The p.g.f. can also be expressed in terms of $R_0^{pp}, R_0^{ph}, R_0^{hp}$ and R_0^{hh} as follows:

$$f_1(t_1, t_2) = \left(1 + \frac{(1-t_1)R_0^{pp} + (1-t_2)R_0^{ph}}{k}\right)^{-k},$$

$$f_2(t_1, t_2) = \left(1 + \frac{(1-t_1)R_0^{hp} + (1-t_2)R_0^{hh}}{k'}\right)^{-k'}.$$

Note that

$$\lim_{k \rightarrow \infty} f_1(t_1, t_2) = \lim_{k \rightarrow \infty} \left(1 + \frac{(1-t_1)R_0^{pp} + (1-t_2)R_0^{ph}}{k}\right)^{-k} = \exp\{-(1-t_1)R_0^{pp} - (1-t_2)R_0^{ph}\},$$

which is the p.g.f. of bivariate Poisson distribution with independent marginals. This means that if we assume that there is no clustering in the disease, then the offspring distribution will be a Poisson distribution and the agents independently infect either type of individuals. If, in addition, $k = 1$ the offspring distribution is a bivariate geometric distribution.

The extinction probability vector (q_1^*, q_2^*) in a two-type branching process with negative binomial offspring distribution is the solution of the equations

$$f_1(q_1, q_2) = \left(\frac{p_0}{1 - p_1 q_1 - p_2 q_2} \right)^k = q_1, \quad (1)$$

$$f_2(q_1, q_2) = \left(\frac{p'_0}{1 - p'_1 q_1 - p'_2 q_2} \right)^{k'} = q_2. \quad (2)$$

Simplifying (1) and (2), we obtain

$$p_1 q_1^{\frac{k+1}{k}} + (p_2 q_2 - 1) q_1^{\frac{1}{k}} + p_0 = 0,$$

$$p'_2 q_2^{\frac{k'+1}{k'}} + (p'_1 q_1 - 1) q_2^{\frac{1}{k'}} + p'_0 = 0.$$

The probability that the nosocomial epidemic will be eradicated is given by

$$P_0 = q_1^{*X_1(0)} q_2^{*X_2(0)},$$

see [1].

Example 2.1. Consider a two-type branching process with bivariate negative binomial offspring distribution and p.g.f. given by

$$f_1(t_1, t_2) = \left(\frac{0.8}{1 - 0.1t_1 - 0.1t_2} \right)^{10},$$

$$f_2(t_1, t_2) = \left(\frac{0.9}{1 - 0.05t_1 - 0.05t_2} \right)^{15}.$$

For the mean matrix we have

$$M = \begin{pmatrix} \frac{5}{4} & \frac{5}{4} \\ \frac{5}{6} & \frac{5}{6} \end{pmatrix}.$$

Since M is regular with dominant eigenvalue $\lambda \approx 2.1 > 1$, there exist q_1 and $q_2 \in [0, 1)$ such that

$$f_1(q_1, q_2) = q_1,$$

$$f_2(q_1, q_2) = q_2.$$

Also the fixed point $(q_1^*, q_2^*) \approx (0.2, 0.3)$ yields the probability of eradication (extinction)

$$P_0 = 0.2^{X_1(0)} 0.3^{X_2(0)} = 0.2.$$

2.1.2 Estimation and Simulations

Before providing a procedure of estimating the basic reproduction number R_0 , we will discuss the simulation of a two-type branching processes and estimating its offspring parameters.

All transmission occurrences are assumed to be independent events. We also assume that nosocomial epidemics begin with a single primary infection case (patient). We fix the number of generations to be N , that is, assume that the hospital intervention will stop the progression of the epidemic after N generations. In each generation, the number of cases is a random vector with bivariate negative binomial distribution. The first coordinate of the random vector is the number of patients infected by the source case and the second coordinate is the number of health-care workers infected by the source case. The offspring distribution of the number of patients and health-care worker are modeled according to bivariate negative binomial with possibly different parameters. Algorithm 1 describes a simulation of the branching process.

After generating the transmission tree of the disease, we test the ability of the estimation procedures to retrieve the preset parameters. We use the maximum likelihood estimation (MLE) for the likelihood

$$L(k, p_1, p_2, k', p'_1, p'_2 | Y) = \prod_{j \in S_P} p_1(y_{11,j}, y_{21,j}) \prod_{j' \in S_H} p_2(y_{12,j'}, y_{22,j'}),$$

Algorithm 1. Simulating Two-Type Branching Process.*Input: Parameters r, k, p_1, p_2 .

Output: Random numbers follow bivariate negative binomial.

begin $\text{rmnbinom} = \text{function}(r, k, p_1, p_2)$ **for** $i = 1$ **to** r X is random number follows negative binomial $(k, 1 - \frac{p_2}{1-p_1})$ Y is random number follows negative binomial $(k + X, 1 - p_2)$ return (X, Y) **end***Input: Parameters $k, p_1, p_2, k', p'_1, p'_2, N$.

Output: Two-type branching process.

begin **for** $i = 2$ **to** N **for** $j = 1$ **to** length of previous level **if** j is even generating random numbers follow $\text{rmnbinom}(k', p'_1, p'_2)$ **else** generating random numbers follow $\text{rmnbinom}(k, p_1, p_2)$ **end**

where $Y = \{(y_{11,j}, y_{21,j}) : j \in S_P\} \cup \{(y_{12,j'}, y_{22,j'}) : j' \in S_H\}$ consists of the numbers of patients and HCW infected by patients, $(y_{11,j}, y_{21,j})$ for $j \in S_P$, and the numbers of patients and HCW infected by HCWs, $(y_{12,j'}, y_{22,j'})$ for $j' \in S_H$; here the sets of indexes for patients and HCWs are denoted by S_P and S_H . Maximization of the log-likelihood was implemented using the software R via the function “constrOptim” [16].

2.1.3 Applications

We simulated a two-type process with maximum number of generations $N = 3, 4, 5$, a reasonable setting for a nosocomial epidemic. The parameter values were set to be $k = 2, p_1 = 0.3, p_2 = 0.6, k' = 3, p'_1 = 0.5$ and $p'_2 = 0.3$. It can be seen from Figures 2, 3 and 4 that the estimates approach the preset parameter values as the number of generations increases. This illustrates the consistency of the MLE estimates.

Figure 2: Estimation of the cluster parameters k and k' for different maximum number of generations using a simulated two-type branching process with $k = 2, p_1 = 0.3, p_2 = 0.6, k' = 3, p'_1 = 0.5$ and $p'_2 = 0.3$.

Figure 3: Estimation of p_1 and p_2 for different maximum number of generations using a simulated two-type branching process with $k = 2$, $p_1 = 0.3$, $p_2 = 0.6$, $k' = 3$, $p'_1 = 0.5$ and $p'_2 = 0.3$.

Figure 4: Estimation of p'_1 and p'_2 for different maximum number of generations using a simulated two-type branching process with $k = 2$, $p_1 = 0.3$, $p_2 = 0.6$, $k' = 3$, $p'_1 = 0.5$ and $p'_2 = 0.3$.

Figure 5: Mean size of epidemic in patients and health-care workers for different maximum number of generations.

Figure 6: Relative errors for 500 replications of a branching process with maximum number of generations $N = 4$ and parameters $k = 5$, $k' = 6$, $p_1 = 0.4$, $p_2 = 0.4$, $p'_1 = 0.5$ and $p'_2 = 0.1$.

Another important characteristic of an epidemic is its size. We simulate and estimate the mean size of the epidemic for different maximum number of generations. It is clear in Figure 5 that the branching process models well an epidemic in its early stages only. So, in the sequel, we will perform all our simulations assuming the maximum number of generations to be three or four at the expense of the possibility of obtaining biased estimates.

The relative errors show small bias of $\hat{R}_0$ in estimating R_0 for 500 replications in Figure 6. Regarding the relationship among R_0 , k , and k' , as it can be seen in Figure 7, the expected number of infections caused by both patients and health-care workers increases as the clusters parameters k and k' increase. That is, if the infection spreads homogeneously in a hospital, then it spreads widely. Different insights could be found in Figures 8 and 9 when we change probabilities of producing same type versus the other type.

Figure 7: Estimation of R_0 for different values of k and k' with $p_1 = 0.45$, $p_2 = 0.45$, $p'_1 = 0.45$, $p'_2 = 0.45$.

Figure 8: Estimation of R_0 for different values of p_1 and p'_1 .

Figure 9: Estimation of R_0 for different values of p_2 and p'_2 .

2.2 Estimating R_0 in One-Type Epidemic

In this subsection, we discuss the method of back-tracing (or building up) the transmission tree assuming the following information is given: units visited by the index case; the times of diagnosis of each of infected individuals starting from the admission of the index case; the hospital units in which each case was diagnosed; and the transition matrix between units. The first three are about epidemic specific information while the fourth one is a hospital specific information. The mean length of the incubation period is also assumed to be epidemiologically known in advance.

2.2.1 Back-Tracing Algorithm

Let P be the transition matrix between the units of the health-care facility, let T be the time of diagnosis upon admission, let U be unit number where onset of symptoms of the index case and secondary cases happened, and let DT_i be the time of diagnosis for the i th infected patient upon admission of the index case.

Algorithm 2. Back-Tracing Algorithm.Input: Parameters s, P, U, DT, α .**begin** set initial k and p (1) For patient $i = 1, 2, \dots, s$:

- (i) Project back in time, patient i by generating t_i from a geometric distribution conditional on not exceeding neither DT_i nor T , given by

$$\text{Prob}(L = t | L \leq \min\{DT_i, T\}) = \frac{\alpha(1 - \alpha)^{t-1}}{1 - (1 - \alpha)^{\min\{DT_i, T\}}},$$

where $t = 1, 2, \dots, \min\{DT_i, T\}$.

- (ii) Project back in space, patient i for t_i steps between units using the transition matrix P after transposing and normalization. The projection is based on a multinomial distribution.

(2) Rearrange the s patients chronologically according to the generated times of their infections. Add the latent branches connecting patients according to the following rules: For patient i for $i = 1, 2, \dots, s$:

- (i) Add a branch between i and j if a patient j and i are in the same unit. Patient j is called the parent of i .
- (ii) If there are more than one patients in the same unit with i , then using negative binomial distribution (k, p) and Bayes' theorem we select the parent of i .

(3) Estimate k and p , and then calculate the estimate for $R_0 = \frac{k(1-p)}{p}$.

(4) Go back to Step (1).

Using this information, we create a Back-Tracing (Stochastic) Algorithm 2 that constructs the transmission tree as a single-type branching process. Next we obtain the MLEs for k and p and use them to calculate the estimate for R_0 . Each back-projection yields one transmission tree, so we run the back-tracing many times. As a result we obtain a histogram for the distribution of R_0 and use the mean (or median) as an estimate.

Example 2.2. Assume that there are twelve patients in the emergency unit that were diagnosed as secondary cases on days 10, 11, 8, 9, 9, 10, 10, 10, 11, 9, 12, and 9 upon the admission of the index case. The secondary cases were diagnosed with the disease within the hospital units number: 2, 2, 4, 3, 4, 3, 3, 3, 2, 4, 3, and 2. Also the movement of the index case between the four units is given by 1, 2, 2, 3, 3, 4, 4, and 4 till his/her diagnosis. Furthermore, we have the transition matrix

$$P = \begin{pmatrix} 0.9 & 0.02 & 0.03 & 0.05 \\ 0.01 & 0.99 & 0 & 0 \\ 0 & 0 & 0.95 & 0.05 \\ 0 & 0 & 0.01 & 0.99 \end{pmatrix}.$$

Assume that the mean length of the incubation period is $\alpha = 7$ days.

Using the proposed back-tracing algorithm, we estimate the parameters k, p and find the estimate for R_0 . Numerical summaries of $\hat{k}$ are shown in Table 1 for two traces for two different initial values of k . These traces are shown in black and red colors in Figure 10. Numerical summaries of $\hat{p}$ are shown in Table 2 for two traces using two different initial values of p . The traces are shown in red and black colors in Figure 11.

	Initial $k = 0.1$	Initial $k = 3$
Mean	0.4815958	0.4731787
Median	0.422367	0.4223887
Variance	0.05315066	0.05459733

Table 1: Numerical summaries of $\hat{k}$ for different initial values.

	Initial $p = 0.8$	Initial $p = 0.0001$
Mean	0.3906378	0.3899772
Median	0.3493233	0.3492362
Variance	0.02249573	0.02462301

Table 2: Numerical summaries of $\hat{p}$ for different initial values.

Figure 10: Two traces of k for two different initials values. Figure 11: Two traces of p using two different initials values.

Furthermore, numerical summaries of $\hat{R}_0$ are shown in Table 3 for two traces using two different initial values of R_0 . These traces are shown in red and black colors in Figure 12. In order to assess the convergence, we compare the two generated traces using Brooks–Gelman–Rubin test (BGR). The traces are compared in terms the variance of the sample median. The BGR tests for all three parameters confirm that both traces converge to the same limiting band, see Figures 13, 14, and 15. A histogram and density plot in the case of $\hat{R}_0$ are given in Figure 16.

	Initial $k = 0.1, p = 0.8$	Initial $k = 3, p = 0.0001$
Mean	0.4395707	0.4351105
Median	0.2299511	0.220787
Variance	0.2076249	0.2010339

Table 3: Numerical summaries of $\hat{R}_0$ for different initial values of k and p . Figure 12: Two traces of R_0 using two different initials values. Figure 13: Brooks–Gelman–Rubin (BGR) test for both chains of $\hat{k}$.

Figure 14: Brooks–Gelman–Rubin (BGR) test for both chains of $\hat{p}$.

Figure 15: Brooks–Gelman–Rubin (BGR) test for both chains of $\hat{R}_0$.

Figure 16: Histogram and density curve of $\hat{R}_0$.

2.3 One-Type Epidemic Model Simulation

In order to validate the back-tracing algorithm, we need to use a gold-standard data with accurate tracing. However, in the absence of such data, we will use simulated to validate the method, instead. We begin with simulating the transmission tree as one-type branching process. Then using the same disease and hospital specific parameters α and P , respectively, we project forward the offspring. We assume that the values of these two parameters are known exactly. They are not epidemic specific parameters and could be estimated independently of the underlying epidemic. At the end we retain the information for DT and U .

We performed one time simulation of the nosocomial epidemic with $k = 2$ and $p = 0.2$, and hence $R_0 = 8$. Then we ran the back-tracing with the $\alpha = 7$ transition matrix P , the same as in the previous example. We calculated the mean, median and standard deviation of R_0 to be 9.04, 8.82, and 2.15, respectively. A histogram for the distribution of $\hat{R}_0$ is shown in Figure 17.

Figure 17: Histogram for $\hat{R}_0$.

3 Discussion and Conclusion

In this paper we present models for nosocomial epidemics based on branching processes. Using the well-developed theory of branching processes, we study nosocomial epidemics. It is our intend to continue with further improvements and validation of the discussed here models. In particular, we plan to address the identifiability problem for estimating R_0 using k and p in one-type branching processes. We will also seek gold-standard data to validate the back-tracing of single-type epidemics. The back-tracing should be expanded to two-type epidemics. We anticipate that our methodology will draw some attention with respect to future applications in the stochastic modeling of nosocomial epidemics.

References

- [1] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, CRC Press, Boca Raton, 2010.
- [2] F. J. Anscombe, Sampling theory of the negative binomial and logarithmic series distributions, *Biometrika* **37** (1950), no. 3–4, 358–382.
- [3] S. Blumberg and J. O. Lloyd-Smith, Inference of $R(0)$ and transmission heterogeneity from the size distribution of stuttering chains, *PLoS Comput. Biol.* **9** (2013), no. 5, Article ID: e1002993.
- [4] G. Chowell, F. Abdirizak, S. Lee, J. Lee, E. Jung, H. Nishiura and C. Viboud, Transmission characteristics of MERS and SARS in the healthcare setting: A comparative study, *BMC Med.* **13** (2015), DOI 10.1186/s12916-015-0450-0.
- [5] J. E. Dunn, Characterization of the bivariate negative binomial distribution, *J. Arkansas Acad. Sci.* **21** (1967), 77–86.
- [6] W. Feller, *An Introduction to Probability Theory and Its Applications. Volume II*, John Wiley & Sons, New York, 1966.
- [7] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 2007.
- [8] M. Kimmel and D. E. Axelrod, *Branching Processes in Biology*, Springer, New York, 2016.
- [9] D. Klinkenberg, J. A. Backer, X. Didelot, C. Colijn and J. Wallinga, Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks, *PLoS Comput. Biol.* **13** (2017), no. 5, Article ID e1005495.
- [10] T. O. McDonald and M. Kimmel, A multitype infinite-allele branching process with applications to cancer evolution, *J. Appl. Probab.* **52** (2015), no. 3, 864–876.
- [11] S. J. Schreiber and J. O. Lloyd-Smith, Invasion dynamics in spatially heterogeneous environments, *Amer. Naturalist* **174** (2009), no. 4, 490–505.
- [12] S. Singh, *Branching processes in disease epidemics*, Ph.D. thesis, Cornell University, 2014.
- [13] D. J. A. Toth, W. D. Tanner, K. Khader and A. V. Gundlapalli, Estimates of the risk of large or long-lasting outbreaks of middle east respiratory syndrome after importations outside the arabian peninsula, *Epidemics* **16** (2016), 27–32.
- [14] P. Trapman, R. Meester and H. Heesterbeek, A branching model for the spread of infectious animal diseases in varying environments, *J. Math. Biol.* **49** (2004), no. 6, 553–576.
- [15] C. J. Worby, P. D. O'Neill, T. Kypraios, J. V. Robotham, D. De Angelis, E. J. P. Cartwright, S. J. Peacock and B. S. Cooper, Reconstructing transmission trees for communicable diseases using densely sampled genetic data, *Ann. Appl. Stat.* **10** (2016), no. 1, 395–417.
- [16] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2011.