SCITECH
RESEARCH ORGANISATION|

Volume 16, Issue 3

Published online: August 18, 2020|

Journal of Progressive Research in Mathematics

www.scitecresearch.com/journals

Effects of minimum epidemic and population sizes on a global epidemic in simulations of final size data

Umar M. Abdulkarim[a], and Husseini S. Ndakwo[b]

[a,b] Nasarawa State University, Keffi Nigeria,

Department of Mathematics

Abstract.

The stochastic SIR household epidemic model is well discussed in [2], [3] and [4]. The work of [1] also proposed maximum likelihood based algorithm for its inference by assuming independence of epidemic in each household, contrary to the dependency assumption in [4].

Using simulations, we examined the need for an appropriate choice of cut-off between small and large epidemics often referred to as minimum epidemic size, using rejection sampling, for a global infection to occur and then compared the estimates of the model parameters over a range of theoretical parameters, λ_L and λ_G with corresponding $z \in [0,1]$.

We found that with large population size, appropriate choice of the minimum epidemic size and $\lambda_G \neq 0$ facilitate the occurrence of a global epidemic.

Thus, given these scenarios, the adequacy of the model fitness to the final size epidemic data is then realised.

Keywords: Final size epidemic, Global Infection, Infectious Period distribution, Maximum likelihood estimates.

1 Introduction

Most often, in simulation of household epidemic, the target is to allow a global epidemic occur in which unimodal behaviour for the distribution of the number infected is observed. This is to provide enough information for estimation of the model parameters.

This work examined ways in which this can be realized and organized in the following form.

In section 2.1, we examined the model with reviews of related literature's, in section 3, we discuss the household structure, in section 4, we examined the properties of the epidemic in the early stages, in section 5, we examined the properties and conditions for the occurrence of a global epidemic. In section 6, we examined maximum likelihood estimation method for the model parameters. In section 7, we simulate and examined the epidemic for occurrence of global infection and the properties of the estimates, given small and large population sizes with different choice of minimum epidemic size, we present plots of the estimates with minimum epidemic size of 10, table of parameter estimates and other statistics when the minimum epidemic size is 10. Also, we present plots of the estimates and table of mean, standard deviation, mean square error and root mean square error with minimum epidemic size of 50 and those with minimum epidemic size of 1000.

In section 9, we discussed our results and in section 10, we present our conclusion.

2 Material and Methods

2.1 The model

The stochastic S-I-R household epidemic model is a generalization of the simple stochastic epidemic, called the general stochastic epidemic with exponentially distributed infectious period. Its development can be traced to the pioneering work of [2-4],[6] and other active contributors like, [9], [8] and [7], [12], [5], [13, 14], [11], [8], [9, 10] etc.

The model assumes a closed and finite population structured into households, each made of susceptibles, infectives and removed individuals, with homogeneous mixing between susceptibles and infectives, independently and at random at two levels, (locally and globally) within the households and individuals from different households, at the points of a homogeneous Poisson processes having rates, λ_L and $\frac{\lambda_G}{N}$ respectively [4], where N is the total population size, λ_G is the total rate that a given infective makes global contacts [3, 4].

Any susceptible contacted will immediately become infectious (since there is no latency for the disease) for period T_I, referred to as the infectious period after which it is removed (died or isolated or immune) at the end of the infectious period. The infectious period of each infective is assumed to be independent and identically distributed according to the random variable T_I which is arbitrary but must be specified [3, 4]. The Poisson processes describing contacts and the infectious period are assumed to be mutually independent.

However [12] and [1] proposed an extensions, which assumes, that infection is initiated from outside the population unlike [4] assumption in which infection is started by some initial number of infectives within the household.

3 Household Structure

The proportion or distribution of households of size $n = 1,2,...,$is given in [3, 4] and [6] as,

$$\alpha_n = \frac{M_n}{M}$$

where M_n is the number of household of size n and M, the total number of households.

The mean household size is then $h = \sum_j j\alpha_j$ and the Probability that global contact is with an individual residing in a household of size n is,

$$\tilde{\alpha}_n = nM_n/N$$

.

Where the total population size illustrated in plate I, is

$$N = \sum_{n=1}^{\infty} nM_n$$

4 Epidemic in the Early Stages.

Every global contact is with a completely susceptible household. The epidemic process behaves like a branching process, in which extinction of the population occur with probability one, if the expected number of offspring is less than one, and by translation from the branching process theorem, this means, the expected number of infected households emanating from a typical infected household, $R_* < 1$. Global epidemic occurs if in the limit as $m \to \infty$ the epidemic infects infinitely many groups, if the branching process does not go instinct [3, 4, 6].

By standard branching process theorem, global epidemic occur if and only if $R_* > 1$,

where $R_* = E(R)$, and R is the number of infected households emanating from a typical infected household.

5 Global epidemic

During the early stage of an epidemic, the number of initial infectives are small, if the number of households is large, then the probability that any global contact is with an infected household is also small. This means that, global contact is with a completely susceptible household.

This makes it possible to approximate an epidemic during its early stages, with a continuous times branching process, in which infected individuals independently infect susceptibles in completely susceptible households at random. However, it is showed by[3] that global epidemic occurs if in the limit as $m \to \infty$, the epidemic infects infinitely many household, ie, if the branching process does not go extinct. The probability of a global epidemic is then computed in line with [4] as,

$$1 - p^a \tag{1}$$

where p is a unique root of $f(s) = s$, in [0,1], representing the proportion of the susceptibles infected in the global epidemic a is the number of the initial infectives, $f(s) = E(s^R)$ is the probability generating function of R, the number of infected households emanating from a typical infected household also referred to as the offspring random variable for the approximating branching process. For other configurations of the initial infectives, the probability of a global epidemic is defined by conditioning on the size of the first generation in the branching process [4, 6].

We need to first determine the generating function $f(s)$ in line with [3, 4, 6] defined by,

$$f(s) = E(s^R) = \sum_{n=1}^{\infty} \tilde{\alpha}_n E(s^{R_n}) \tag{2}$$

where R_n is the total number of global contacts emanating from the household epidemic, assumed to follow a Poisson distribution with random mean, $\lambda_G A_n$, A_n is the sum of the infectious period of all the infectives and

$$E(s^{R_n}) = E(E(s^{R_n}|A_n)) \tag{3}$$

$$= E(\exp(-\lambda_G A_n(1-s))) \tag{4}$$

$$= \phi_{n-1,1}(\lambda_G(1-s)) \tag{5}$$

where R_n and A_n are defined in equation (2), λ_G and λ_L are the local and global contact rates. Also, $\phi_{n,a} = E(\exp(-\theta A_{n,a}))$ and $A_{n,a}$ is the sum of the infectious periods of the infectives in the household epidemic, called severity of the household epidemic with initially n susceptibles and a infectives, which is defined in [3, 6] as,

$$\phi_{n,a}(\theta) = \sum_{k=0}^{n} \binom{n}{k} \gamma_k(\theta)\phi(\theta + \lambda_L.k)^{n+a-k}. \tag{6}$$

Where $\gamma_i(\theta)$ for $i = 0,1,\dots,n$ are determined recursively by,

$$\sum_{i=0}^{k-1} \gamma_i(\theta)\phi(\theta + \lambda_L.i)^{k-i} + \binom{k}{k}\gamma_k(\theta) = 1.(k = 0,1 \dots n) \tag{7}$$

The gamma function in equation (7) can further be simplified for every $k = 0,1,\dots n$ as follows,

$k = 0$, gives,

$$\binom{0}{0}\gamma_0(\theta)\phi(\theta+\lambda_L.0)^{0-0}=1$$

$$\gamma_0(\theta)=1,\forall\theta\geq 0 \tag{8}$$

Thus, $\gamma_k(\theta)$ can be written as,

$$\gamma_k(\theta)=1-\sum_{i=0}^{k-1}\binom{k}{i}\gamma_i(\theta)\phi(\theta+\lambda_L.i)^{k-i},k=1,2\ldots n,\forall\theta\geq 0, \tag{9}$$

where $\phi(\theta)=E(\exp\{-\theta T_I\}),T_I$ is the infectious period of an infected individual.

If λ_L and θ approaches zero simultaneously, then $\phi(0)$ goes to 1. Then from equation (8), we can see that equation (9) reduces to,

$$\gamma_k(0)=1-\sum_{i=0}^{k-1}\binom{k}{i}\gamma_i(0),k=1,2\ldots n, \tag{10}$$

We prove by induction that $\gamma_k(0)=0,\forall k\in\mathbf{Z}_+-\{1\}$.

When $k=1$, equation (10), reduces to, $\gamma_1(0)=1-\binom{1}{0}\gamma_0(0)=0$.

When $k=2$, $\gamma_2(0)$ reduces to,

$$\gamma_2(0)=1-\left(\binom{2}{0}\gamma_0(0)+\binom{2}{1}\gamma_1(0)\right)$$

using $\gamma_0(0)=1$ and $\gamma_1(0)=0$, we get $\gamma_2(0)=1-1=0$.

We assumed, this expression holds for all $k\in\mathbf{Z}_+$ and show that it holds for $k+1$.

For $k+1$, equation (10) assumes,

$$\gamma_{k+1}(0)=1-\sum_{i=0}^{k}\binom{k+1}{i}\gamma_i(0) \tag{11}$$

Since, we can express $\binom{k+1}{i}$ as $\binom{k}{i}+\binom{k}{i-1}$, equation (11) can be written as,

$$\gamma_{k+1}(0)=1-\sum_{i=0}^{k}\left(\binom{k}{i}+\binom{k}{i-1}\right)\gamma_i(0)$$

$$=1-\left(\sum_{i=0}^{k-1}\binom{k}{i}\gamma_i(0)+\gamma_k(0)+\sum_{i=0}^{k}\binom{k}{i-1}\gamma_i(0)\right)$$

Also, from equation (10), we have $\sum_{i=0}^{k-1}\binom{k}{i}\gamma_i(0)=1-\gamma_k(0)$ so that

$$\gamma_{k+1}(0)=-\sum_{i=0}^{k}\binom{k}{i-1}\gamma_i(0)$$

Thus, $\gamma_{k+1}(0)=0$.

The hypothesis, holds for $k+1$, and so in general, $\gamma_k(0)=0,\forall k\in\mathbf{Z}_+-\{1\}$.

Putting these results in $\phi_{n,a}(0)$ we now have,

$$\phi_{n,a}(0)=\sum_{i=0}^{n}\binom{n}{i}\gamma_i(0). \tag{12}$$

From equation (12), we can see that the terms reduces to zero except the first one, i.e, $\binom{n}{0}\gamma_0(0) = 1$, since $\phi(0)^n = 1$. Hence equation (12) reduces to $1, \forall n \geq 0$.

The probability generating function of the number of infected households generated by a typical infected household is given by,

$$f(s) = E(s^R) = \sum_{n=1}^{\infty} \tilde{\alpha}_n = 1, \tag{13}$$

where, $\theta, \lambda_L \to 0, s = 1$ is the largest solution of $f(s) = s = p \in [0,1]$, for which there is no global epidemic. Where a global epidemic, is given by $1 - p = 0$, for an epidemic started by a single initial infective.

Thus, the probability of nonglobal epidemic corresponding to extinction of an approximating branching process to the epidemic process is given by 1.

This means, there can't be an epidemic in absence of contacts within and among members of the households.

If on the other hand $\lambda_L \to \infty$ and $\theta \geq 0$,

then $\phi(\theta + \lambda_L)^{k-i}$ in equation (9) reduces to 0, $\forall \theta \geq 0$, since $k - i \geq 1, \forall i = 0,1, \dots, k-1$. Also from equation (9), we see that $\gamma_k(\theta)$ reduces to 1, $\forall k \in \mathbf{Z}_+, \theta > 0$ and $\lambda_L \to \infty$.

Applying these results in equation (6), we have $\phi(\theta + \lambda_L.k)^{n+a-k} \to 0$, since $n + a - k \geq 1, \forall a \geq 1$, where a is the initial number of infectives.

It trivially follows that, if $\lambda_L \to \infty$, then $\phi_{n,a}(\theta) \to 0, \forall \theta > 0$.

From equation (2), we will have $s = 0$. Hence the extinction probability of the approximating branching process to the epidemic process is, $p = 0$. The probability of a global epidemic is 1.

This result also holds if $\theta \to \infty$ for some $\lambda_L \geq 0$. Since $k - i \geq 1$, in equation (9), we have, $\phi(\theta + \lambda_L.k)^{k-i} \to 0$, for $\lambda_L \geq 0$.

From equation (9), if $\theta \to \infty$, then $\gamma_k(\theta) \to 1, \forall \lambda_L \geq 0$.

Also, from equation (6), $\phi(\theta + \lambda_L.k)^{n+a-k} \to 0$, where $n + a - k \geq a$, $\forall a > 1$, and $n + a - k$ cannot be 0, $\forall k \in \mathbf{Z}_+$.

It follows that, if $\theta \to \infty$, then $\phi_{n,a}(\theta) \to 0, \forall \lambda_L$.

Hence, equation (6) reduces to 0, $f(s) = E(s^{R_n}) = 0$ and $s = 0$.

There will be a global epidemic with probability 1.

If $\theta, \lambda_L \to \infty$ simultaneously, then we see that equation (9) will assume, 1, since $\phi(\theta + \lambda_L.i)^{k-i} \to 0$, where $k - i \geq 1$.

Similarly, in equation (6), $\phi(\theta + \lambda_L.k)^{n+a-k}$ will be zero, where $n + a - k \geq a$, $a > 1$. Hence, equation (6) reduces to 0.

It follows from equation (2) that $f(s) = 0$, and $s = 0$. There will be a global epidemic with probability 1.

From $\theta = \lambda_G(1 - s)$ and in lines with [3, 4, 6], we see that $\gamma_k(\theta)$ is a function of both λ_L and λ_G respectively and $\theta \to 0$ will mean either $\lambda_G \to 0$ for $s \in (0,1]$, or $s = 0$ for $\lambda_G > 0$.

If $\lambda_G \to 0$ for some $s \in [0,1)$ and $\lambda_L \to 0$. Then $\gamma_k(0) = 0$ for $\forall k \in \mathbf{Z}_+ - \{0\}$, where $\gamma_0(0)$ is given in equation (8) and $\phi_{n,a}(0) = 1$, $f(s) = 1$. There will be nonglobal epidemic with probability 1. The probability of a global epidemic is 0.

However, if $s = 1$, $\lambda_G > 0$ and $\lambda_L \to 0$, then we will have $\gamma_k(0) = 0$ and $\gamma_0(0)$ in equation (8), $\phi_{n,a}(0) = 1$, and since $s = 1$. It follows that the probability of nonglobal epidemic is 1 and the probability of a global epidemic is 0.

Similarly, the case for which $\theta \to 0$ for some $\lambda_G > 0$, follows by allowing either $\lambda_G \to 0$ for some $s \in [0,1)$ or $\lambda_G > 0$ and $s = 1$.

Thus, $\lambda_G > 0$ for a global epidemic to occur in the household

6 Maximum likelihood estimation

If $X_{n,j}$ is the number of households of size n with j infectives, (total number of cases), and $P_{n,j}$ is the final size probabilities, (probability of j cases in a household of size n at the end of the epidemic), then each household size, has a separate multinomial distribution for $X_{n,0}, \dots, X_{n,j}, (j = 0, \dots, n, n = 1, \dots, max)$, given by [10] as,

$$P(X_{n,0} = x_{n,0}, \dots, X_{n,j} = x_{n,j}) = \frac{(M_n)!}{\prod_{j=1}^{max} (x_{n,j})!} \prod_{j=0}^{n} P_{n,j}^{x_{n,j}}, \tag{14}$$

where M_n is the number of household of type n among the infected households.

By assuming independence of epidemics in each household in accordance with [1], the likelihood function which is referred to as approximate likelihood function of the parameters, λ_L and π, [4] is given by,

$$L(\lambda_L, \pi) = \frac{(M_{n.})!}{\prod_{i=1}^{max} (x_{n,j})!} \prod_{i=1}^{max} \prod_{j=0}^{n} P_{n,j}(\lambda_L, \pi)^{x_{n,j}}, \tag{15}$$

where $P_{n,j}$ are the final size probabilities, n is the household size, π is the probability of avoiding infection from outside the household, λ_L is the local contact rate, $x_{n,j}$ is the final size data defined as the number of households of size n with j number of infectives, max is the maximum household size, and M_n is the number of households of size n among the infected households.

Using logarithm in equation (15) for ease of computation and simplification, we can express the approximate likelihood function in terms of its loglikelihood as,

$$l(\lambda_L, \pi) = \log(M_{n.})! - \sum_{i=1}^{max} \log(x_{n,j})! + \sum_{i=1}^{max} \sum_{j=0}^{n} x_{n,j} \log P_{n,j}. \tag{16}$$

The approximate loglikelihood function of the theoretical parameters, λ_L and π can then be computed using appropriate numerical optimization along the lines of the computational algorithm given in [1].

We have developed Matlab programs using the Nelder-Mead fminsearch simplex numerical algorithm referred to here as two dimensional numerical optimization to estimate the parameters.

7 Simulation and Inference

We studied the behavior of the epidemic and the model using simulations and then examined the place of minimum epidemic size, population size and the magnitude of λ_G on the occurrence of a global infection. Using appropriate cut-off between the small and large epidemics, we see that global infection is realized. We examined the estimates of the parameters, their mean, variance and mean square error given these scenarios.

Using the assumption of independence of epidemic between households in [1] and since each household size (number of cases) has separate multinomial distribution given in equation (14), we can express the approximate likelihood function as in equation (15).

The parameters of the approximate likelihood function, which are the local infection rate and the probability of avoiding infection from outside the households, λ_L and π are then estimated.

The process is such that the starting values for π and λ_L are obtained according to [9] from equations (17) and (18).

For example estimating π, requires equation (17) to be used in evaluating the starting value given as

$$\hat{\pi} = 1n \sum_{s=1}^{\max} n_s \left(\frac{n_{0,s}}{n_s}\right)^{1s}, \tag{17}$$

where n is the total number of households, max is the maximum household size, n_s is the number of households of sizes s and $n_{j,s}$ is the number of households of size s in which the size of the outbreak is $j = 0,1,\dots,s$. i.e. number infectives in the household of size s. Observe that $\sum_{j=0}^{s} n_{j,s} = n_s$ and $\sum_{s=1}^{\max} n_s = n$ respectively.

Here, $n_{0,s}n_s$ is an unbiased estimate of $P_0(s) = \pi^s$, where $P_0(s)$ is the probability of zero infectives in the household of size s, which can also be read as the probability that all the susceptibles in the household of size s avoid global infection.

Then $(n_{0,s}n_s)^{1s}$ provides estimates of π for the household sizes $s = 1,2,\dots,max$. Pooling the estimates together [9] gave the initial estimates in equation (17).

For the local infection rate, a reasonable estimate for λ_L for the household size s is given by [9] as, $(n_{1,s}(n_s - n_{0,s}))^{1(s-1)}$ and is unity when $n_{0,s} = n_s$.

Pooling the estimates together as in [9], the estimate of λ_L is started using,

$$\hat{\lambda}_L = \frac{1}{\sum_{s=2}^{\max}(n_s - n_{0,s})} \sum_{s=2}^{\max} (n_s - n_{0,s})\left(\frac{n_{1,s}}{n_s - n_{0,s}}\right). \tag{18}$$

Alternatively if we know the pair of parameters, (λ_L, λ_G), then by defining a new functional D, which is the sum of square difference between the old and new values of π and between the old and new values of z given as,

$$D = (\pi_{old} - \pi_{new})^2 + (z_{old} - z_{new})^2,$$

$$\pi_{New} = \exp(-\hat{\lambda}_G z_{old} E(T_I),$$

$$z_{New} = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^{n} \binom{n}{k} (1 - \pi_{old})^k \pi_{old}^{n-k} \mu_{n-k,k} \hat{\lambda}_L.$$

Then we can adopt the Nelder-Mead fminsearch simplex numerical algorithm on D to find the values of z and π respectively. Using these procedures, the parameters are estimated in table 1.

8 Table of Theoretical Parameters

	Corresponding theoretical parameter		
(λ_L, λ_G)	π	z	R_*
(0.3, 0.12)	0.84487	0.3426	1.2902
(0.13, 0.17)	0.74223	0.4275	1.1432
(0.1, 0.29)	0.4199	0.7298	2.2166
(0.25, 0.39)	0.2302	0.9185	4.0229

Table 1: Pairs of the local and global infection rates with their corresponding theoretical parameters.

With household structure and population size fifty times that of [1] given as $[133,189,108,106,31] \times 50$, minimum epidemic size of 1000 and simulation runs of 1000, in comparison with our studies in sections 8.1 and 8.5 for theoretical parameters corresponding to $z = 0.1775$ and population size in [1] for different choice of the minimum epidemic size and simulation runs of 1000.

This is done in order to study the influence of the minimum epidemic size and the population size on the occurrence of a global infection in the households and hence their effects on the estimates of the parameters.

These are implemented using program functions developed for this purpose with the theoretical parameters in [1, 4] and household structure in [1], population size of 1414, and simulation runs of 1000 for the following minimum epidemic sizes 10,100 respectively.

The scatter plots of the estimates and the histogram of the number infected are then presented to provide insights into their behaviors.

8.1 Plots of the estimates with minimum epidemic size of 10

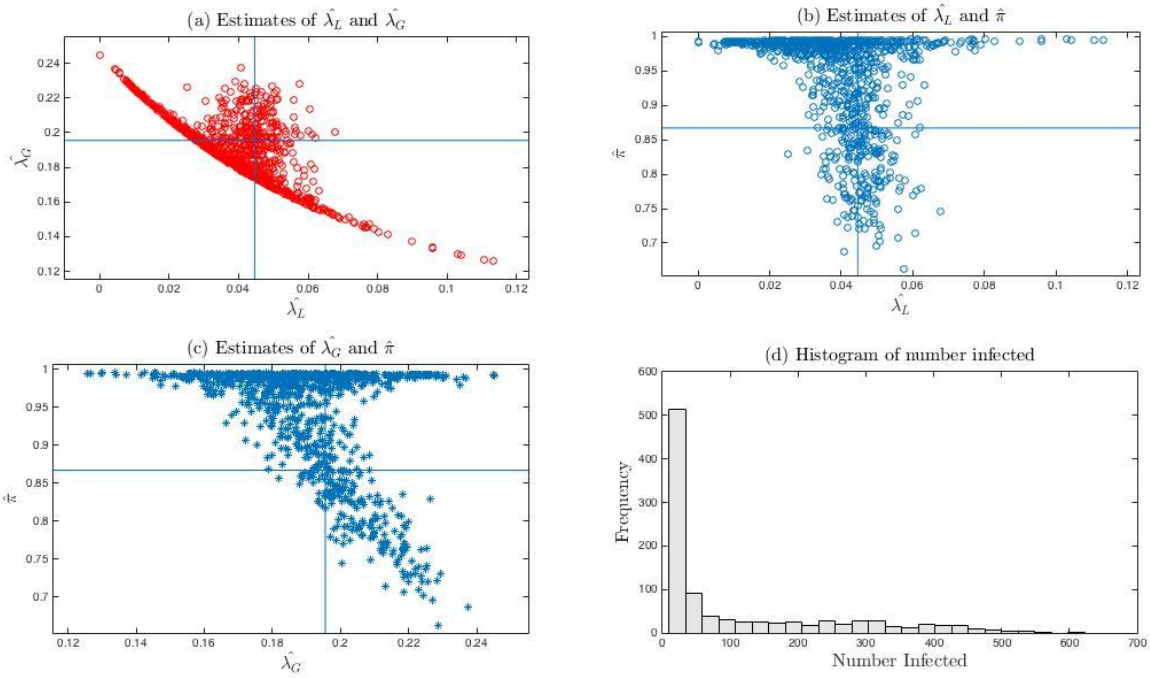

Figure 1: Plots of the estimates of $(\lambda_L, \lambda_G), (\lambda_L, \pi), (\lambda_G, \pi)$ and histogram of number infected with theoretical parameters corresponding to $z = 0.1775$ and minimum epidemic size of 10.

8.2 Table of parameter estimates and other statistics when the minimum epidemic size is 10.

	Parameter Estimates.				
Mean, SD, MSE, RMSE.	$\hat{\lambda}_L$	$\hat{\lambda}_G$	$\hat{\pi}$	$\hat{z}$	$\hat{R}_*$
Theoretical Parameters	0.0446	0.1955	0.8674	0.1775	1.0596
Mean	0.038025	0.19346	0.94022	0.07902	1.0661
Standard Deviation	0.01582	0.020356	0.075164	0.098512	0.081201
Mean Square Error	0.00029238	0.00041811	0.010968	0.019431	0.011596
Root Mean Square Error	0.017125	0.020448	0.10473	0.1394	0.10769

Table 2: Mean of the parameter estimates for theoretical parameters corresponding to $z = 0.1775$, household structure and size in [1, 4] and minimum epidemic size of 10.

8.3 Plots of the estimates and table of mean, standard deviation, mean square error and root mean square error with minimum epidemic size of 100.

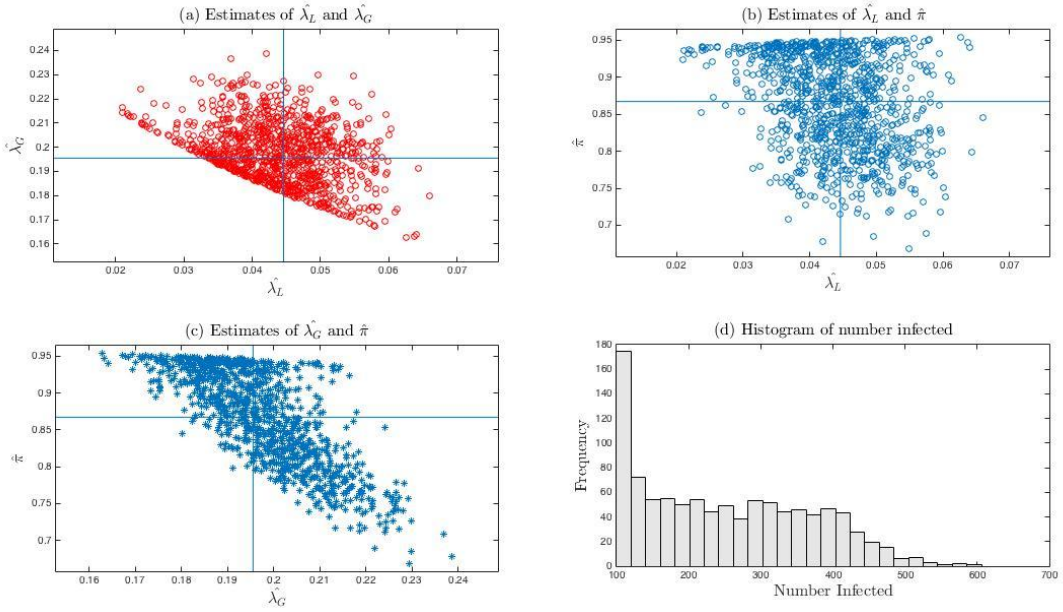

Figure 2: Plots of the estimates of $(\lambda_L, \lambda_G), (\lambda_L, \pi), (\lambda_G, \pi)$ and histogram of number infected with theoretical parameters corresponding to $z = 0.1775$ and Minimum Epidemic size of 100.

8.4 Table of parameter estimates and other statistics when the minimum epidemic size is 100.

	Parameter Estimates.				
Mean, SD, MSE, RMSE.	$\hat{\lambda}_L$	$\hat{\lambda}_G$	$\hat{\pi}$	$\hat{z}$	$\hat{R}_*$
Theoretical Parameters	0.0446	0.1955	0.8674	0.1775	1.1303
Mean	0.043791	0.19786	0.86498	0.17851	1.1362
Standard Deviation	0.0074468	0.012422	0.063977	0.082632	0.07399
Mean Square Error	5.61E-05	0.00015973	0.0040941	0.0068219	0.0055032
Root Mean Square Error	0.0074869	0.012638	0.063985	0.082595	0.074183

Table 3: Mean of the parameter estimates for theoretical parameters corresponding to $z = 0.1775$ and household structure and size in [1, 4] and minimum epidemic size of 100.

8.5 Plots of the estimates and table of mean, standard deviation, mean square error and root mean square error with minimum epidemic size of 1000

The behaviour of the estimates are further examined in table 4 with minimum epidemic sizes of 10 and 100 in figures 1 and 2 with corresponding tables of statistics, 2 and 3 respectively.

From table 3, we see that the estimates are unbiased given the population size in [1] and minimum epidemic size of 100 compared to the choice of minimum epidemic size less than 100. However, the question then is how precise are the estimates if the minimum epidemic size is extremely larger than 100, given the small population size of 1414 in [1] and also population size larger than 1414.

We explored these questions by assuming minimum epidemic sizes of 1000 for the small population size of 1414, which is far greater than 100, adopted in figures 3 (a)-(d). We employed the same minimum epidemic size of 1000 for the population of size of 70700, which is fifty times greater than the population size considered in [1] as in table 4.

In the case of the small population size of 1414, a minimum epidemic size of 1000, give estimates that are biased and imprecise compared to the choice of 100 as the minimum epidemic size in table 3 with the same population size. Unlike in table 3, we see significant difference between the mean of the parameter estimates and their true values.

The mean square error of the estimates does not satisfy the minimum mean square error criterion required of good estimates. With large population size of, 70700, and choice of minimum epidemic size 1000, the estimates are unbiased with insignificant difference from their true mean values compared to the former as shown in table 4.

There is no doubt that inappropriate choice of the minimum epidemic size below and above its threshold given small and large population sizes affects the precision and other properties of the estimates of the parameters. Hence, there is the need to apply a better strategy of choosing these parameters. These involve, firstly simulating the household epidemic with minimum epidemic size of 1 to understand the bimodal behavior of the distribution of the epidemic and hence locate the minimum cut-off of the number infected between the epidemics then followed by rejection sampling.

From the bimodal behaviors of the distribution of the number infected in figure 3, for the small and large population sizes, 1414 and 70700, the cut-off of 100 and 1000 respectively are reasonable.

Choice of extremely large value above the minimum epidemic size leads to loss of information in the final size epidemic data. This is because simulations with large number infected will be rejected and hence may result in estimates that are biased and imprecise as shown in table 4, with minimum epidemic size of, 1000, for population sizes, 1414, and, 70700, respectively. The choice of, 1000, for the small population size of 1414, is far above the required cut-off between the epidemics as shown in figure 3 for small and large population sizes and hence some of the large epidemics will be wrongly rejected. This then leads to loss of information required for inference from the final size epidemic data. Hence biased estimates are obtained unlike the case with 100, in table 4.

		Pop. size=1414			Pop. size=70700		
Par.	Estim.	mean	std	MSE	mean	std	MSE
$\hat{\lambda}_L$	0.0446	0.053486	0.0089206	0.00015846	0.0445	0.0010809	1.18E-06
$\hat{\lambda}_G$	0.1955	0.33199	0.012481	0.018786	0.19525	0.0028492	8.17E-06
$\hat{\pi}$	0.86725	0.38183	0.013799	0.23583	0.86946	0.018014	0.00032903
$\hat{z}$	0.1777	0.70781	0.0013614	0.28103	0.17469	0.023642	0.00056745
$\hat{R}_*$	1.1304	2.0239	0.033412	0.7995	1.1282	0.019158	0.00037142

Table 4: Table of comparison of the mean, standard deviation and mean square error of the estimates using the minimum epidemic size of 1000 and simulation runs of 1000.

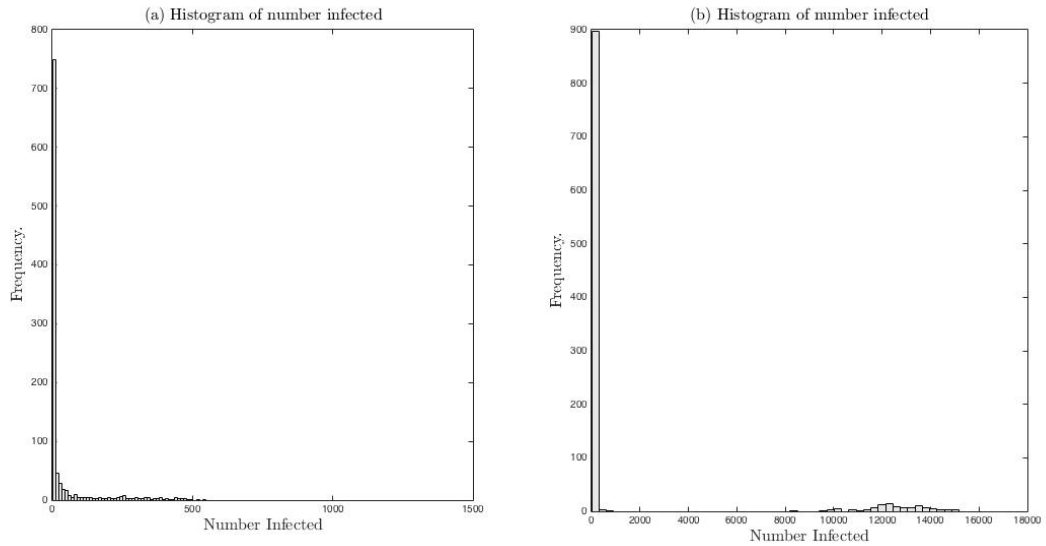

Figure 3: Histogram of number infected from simulations of household epidemic with population sizes of 1414 and 70700 respectively, minimum epidemic size of 1 and simulation runs of 1000.

8.6 Parameter estimates with minimum epidemic size of 1000

In section previous simulations, we found that with small population size and inappropriate choice of the minmimu epidemic, global infection failed to occur. Hence π, z and R_* are biased with imprecise estimates owing to lack of enough information in the final size data. With increasing minimum epidemic size, these estimates become less biased with improved estimates.

In order to overcome this estimation problem, we considered large population size with appropriate minimum epidemic size of 1000 and a range of theoretical parameters in table 1 to allow global epidemic and hence provide sufficient information for parameter estimation.

We considered pair of theoretical parameters (λ_L, λ_G) corresponding to $0 < z < 0.5$ and $0.5 < z < 1$ away from its boundaries and then studied the behavior of the estimates and the distribution of the number infected for these sets of theoretical parameters corresponding to z in the given sets.

Starting with $\lambda_L = 0.0446, \lambda_G = 0.1955$ and corresponding theoretical parameters, $\pi = 0.8674, z = 0.1775, R_* = 1.1303$, minimum epidemic size of 1000, to allow global epidemic to take off in each of the simulation runs. We simulate 1000 times household epidemic in a population of size 70700 which is fifty times that of [1] given as 1414, estimate and plot the parameters, $(\lambda_L, \lambda_G), (\lambda_L, \pi), (\lambda_G, \pi)$ and histogram of the distribution of number infected.

Table of mean, standard deviation and root mean square error of the estimates are presented.

8.7 Plots of the estimate of λ_L, λ_G and π when the theoretical parameters are $\lambda_L = 0.0446$ and $\lambda_G = 0.1955$ with minimum epidemic size of 1000.

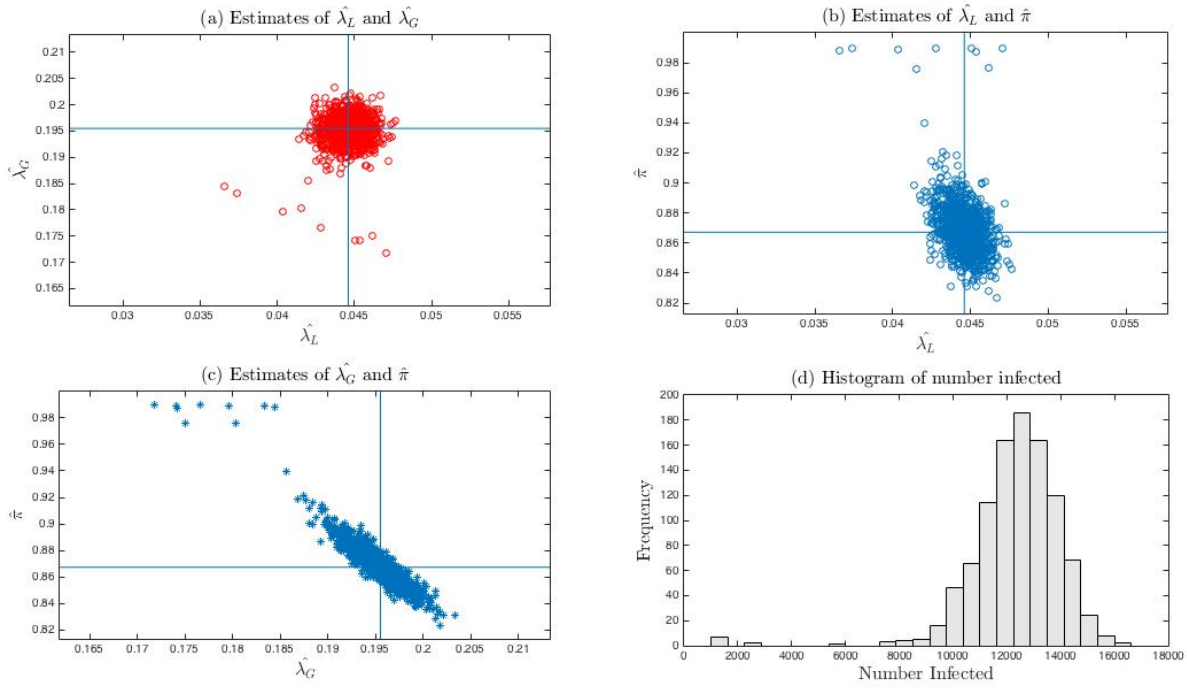

Figure 4: Plots of the Estimates of $(\lambda_L, \lambda_G), (\lambda_L, \pi), (\lambda_G, \pi)$ and histogram of number infected with theoretical parameters $\lambda_L = 0.0446, \lambda_G = 0.1955$ and minimum epidemic size of 1000.

8.8 Plots of the estimate of λ_L, λ_G and π when the theoretical parameters are $\lambda_L = 0.25$ and $\lambda_G = 0.39$ with minimum epidemic size of 1000.

Also with $\lambda_L = 0.25, \lambda_G = 0.39$ and corresponding theoretical parameters, $\pi = 0.2302, z = 0.9185, R_* = 4.0229$. We plot $(\lambda_L, \lambda_G), (\lambda_L, \pi), (\lambda_G, \pi)$ and the histogram of the distribution of number infected. Table of mean, standard deviation and root mean square error are presented.

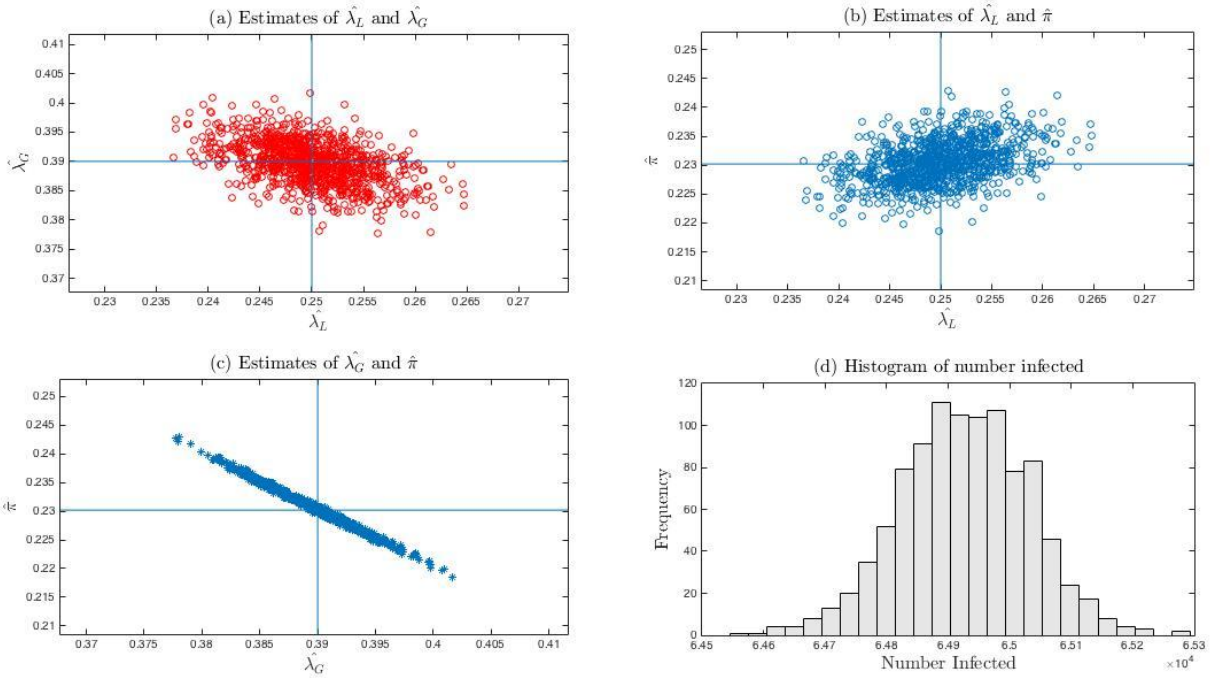

Figure 5: Plots of the estimates of $(\lambda_L, \lambda_G), (\lambda_L, \pi), (\lambda_G, \pi)$ and histogram of number infected with theoretical parameters $\lambda_L = 0.25, \lambda_G = 0.39$ and minimum epidemic size of 1000.

	Proportion Infected.							
Par.	z=0.1775	Theor. Par.	z=0.42757	Theor. Par.	z=0.7298	Theor. Par.	z=0.9185	Theor. Par.
$\hat{\lambda}_L$	0.044578	0.0446	0.13004	0.13	0.099901	0.1	0.24987	0.25
$\hat{\lambda}_G$	0.19515	0.1955	0.16997	0.17	0.28997	0.29	0.38983	0.39
$\hat{\pi}$	0.86956	0.8674	0.74247	0.7423	0.42011	0.4199	0.23046	0.23021
$\hat{z}$	0.17461	0.1775	0.42728	0.42757	0.72949	0.7298	0.91833	0.9185
$\hat{R}_*$	1.1282	1.1303	1.4315	1.4316	2.2154	2.2166	4.0203	4.0229

Table 5: Table of mean of the estimates from the two dimensional model and theoretical parameters in table 1.

	Proportion Infected.			
Par.	z=0.1775	z=0.42757	z=0.7298	z=0.9185
$\hat{\lambda}_L$	0.0010624	0.0015197	0.0015715	0.0047053
$\hat{\lambda}_G$	0.0030219	0.0016325	0.0023247	0.0036795
$\hat{\pi}$	0.018571	0.006892	0.0045573	0.0036487
$\hat{z}$	0.024377	0.0094885	0.0037947	0.0014917
R_*	0.019749	0.014713	0.017152	0.033281

Table 6: Table of the standard deviation of the estimates from the two dimensional model with theoretical parameters in table 1

	Proportion Infected.			
Parameter.	z=0.1775	z=0.42757	z=0.7298	z=0.9185
$\hat{\lambda}_L$	0.0010621	0.0015196	0.0015738	0.0047048
$\hat{\lambda}_G$	0.0030408	0.001632	0.0023238	0.0036814
$\hat{\pi}$	0.018705	0.0068906	0.004558	0.0036554
$\hat{z}$	0.024559	0.009487	0.0037996	0.0015001
$\hat{R}_*$	0.019861	0.014707	0.017183	0.033367

Table 7: Table of the root mean square error of the estimates from the two dimensional model with theoretical parameters in table 1. The estimates are precise.

9 Results and Discussion

In figure 1, we see positive and negative linear correlation between some of the parameter estimates for example increasing λ_L leads to decreasing λ_G. Generally, in most of the simulations few number of infections occurred, many suceptibles avoid global infection. Hence a global epidemic has not taken place.

In table 2, we see small difference between the mean of the estimates of λ_L, λ_G and their theoretical values. While those of π, z and R_* are significantly different from their theoretical mean and possess large standard deviation, which are the standard error of the estimates. These later three parameter estimates are biased owing to the choice of 10 as the minimum epidemic size with the small population size in [1].

Most of the simulations yielded small number of infections, as many susceptibes avoided global infection.

The estimates are less biased compared to those in table 2. This indicates that appropriate choice of the minimum epidemic size leads to the realization of a global infection in the households and hence the occurrence of a global epidemic in which there is enough information for the estimation of the parameters.

Also large number of simulations yielded few number infected with only small number of simulations with large number infected as shown by the bimodal behavior of the histogram of the distribution of the number infected associated with simulations with small population size.

In figures 4 (a)-(d), the estimates are unbiased and scattered around their true parameter values. The unimodal pattern of the distribution of the number infected by the histogram is indicative of the occurrence of a global epidemics.

In figures 5 (a)-(d), the estimates are precise and centered around the true parameter values. Also large number of susceptibles are infected.

10 Conclusion

In summary with appropriate choice of the minimum epidemic size, large population size and $\lambda_G \neq 0$, arbitrary choice of the local infection rate, λ_L, the histogram of the number infected exhibits unimodal behaviour indicative of global infection with high precision for the parameter estimates as expected.

Thus, the two dimensional stochastic SIR household epidemic model fit reasonably well given these scenarios, in which adequate choice of the minimum epidemic size and large population size are assumed.

References

[1] C. ADDY, I. M. LONGINI JR, AND M. HABER, *A Generalised Stochastic Model for the Analysis of Infectious Disease Final Size Data*. Biometrics, Vol. 47, No. 3, (1991), pp. 961-974.

[2] F. G. BALL, *The Threshold Behaviour of Epidemic Models*. Journal of Applied Probability, Vol. 20, No. 2, (1983), pp. 227-241.

[3] F. G. BALL, *A Unified Approach to the Distribution of the total size and Total Area under the Trajectory of Infection in Epidemic Models*. Advances in Applied Probability, Vol. 18, No. 2, (1986), pp. 289-310.

[4] F. G. BALL, D. MOLLISON AND G. SCALIA-TOMBA, *Epdemics with Two Levels of Mixing*. Annals of Applied Probability, Vol. 7, No. 1, (1997), pp. 46-89.

[5] F. BALL AND P. DONNELLY, *Strong Approximations for Epidemic Models*. Stochastic Processes and their Application, Vol. 55, (1995), pp. 1-21.

[6] F. G. BALL AND O. D. LYNE, *Epidemics Among A Population of Households*. Mathematical Approaches for the Emerging and Reemerging Infectious Disease: Models, Methods and Theory, (The IMA Volumes in Mathematics and its Applications), Springer, Editor: Castillo-Chavez, Vol. 126, (2000), pp. 115-125.

[7] F. G. BALL, P. O'NEILL AND J. PIKE, *Stochastic Epidemics in Structured Populations Featuring Dynamic Vaccination and Isolation*. Journal of Applied Probability, Vol. 44, No. 3 (Sept., 2007), pp. 571-585.

[8] F. G. BALL AND P. NEAL, *A general model for the stochastic SIR epidemic with two levels of mixing*. Journal of Math. Biosciences, Vol. 180, (2002), pp. 73-102.

[9] N. G. BECKER, *A stochastic Model for Interacting Population*. Journal of Applied Probability, Vol. 7, No. 3, (1970), pp. 544-564.

[10] N. G. BECKER, *Analysis of Infectious Disease Data: Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, (1989).

[11] D. Clancy and P. D. O'NEILL, *Exact Bayesian Inference and Model Selection for Stochastic Models of Epidemics Among a Community of Households*. Scandinavian Journal of Statistics, Vol. 34, No. 2, (2007), pp. 259-274.

[12] I. M. LONGINI, JR AND J.S. KOOPMAN, *Household and Community Transmission Parameters from Final Distribution of Infections in Households*. Biometrics, Vol. 38, No. 1, (1982), pp. 115-126.

[13] P. Neal, *Efficient Likelihood-free Bayesian Computation for Household Epidemics* . Journal of Statistics and computing, Vol. 22, No.6, (2012), pp. 1239-1256.

[14] P. Neal, *A Household SIR Epidemic Model Incorporating Time of Day Effects*. Journal of Applied Probability, Vol. 53, (2016), pp. 489-501.