

Research Article

Improved Estimation of the Initial Number of Susceptible Individuals in the General Stochastic Epidemic Model Using Penalized Likelihood

Changhyuck Oh

Department of Statistics, Yeungnam University, Gyeongsan, Gyeongbuk 712-749, Republic of Korea

Correspondence should be addressed to Changhyuck Oh; choh@yu.ac.kr

Received 24 March 2014; Revised 26 August 2014; Accepted 27 August 2014; Published 11 September 2014

Academic Editor: Getachew Dagne

Copyright © 2014 Changhyuck Oh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The initial size of a completely susceptible population in a group of individuals plays a key role in drawing inferences for epidemic models. However, this can be difficult to obtain in practice because, in any population, there might be individuals who may not transmit the disease during the epidemic. This short note describes how to improve the maximum likelihood estimators of the infection rate and the initial number of susceptible individuals and provides their approximate Hessian matrix for the general stochastic epidemic model by using the concept of the penalized likelihood function. The simulations of major epidemics show significant improvements in performance in averages and coverage ratios for the suggested estimator of the initial number in comparison to existing methods. We applied the proposed method to the Abakaliki smallpox data.

1. Introduction

In any epidemic in a group of individuals, there is a subgroup of individuals who are not susceptible to a disease, that is, those immune to the disease naturally or by vaccination, as well as those not exposed to the disease owing to physical separation or other reasons. Therefore, estimation of the size of the initially susceptible population in the group might be pivotal; for example, see [1]. For the general stochastic epidemic model, [2–4] have dealt with estimating the initial number.

For the case where an epidemic is observed fully over a given time interval such that all infection and removal times are known, [2] provided martingale estimating equations to propose an estimator called the M-estimator of the initial number of susceptible individuals and its approximate variance. However, the M-estimator does not have the property of consistent coverage ratios for confidence intervals of the initial number of susceptible individuals. For the same conditions, [4] derived a likelihood function using the counting process theory after [5] to yield the maximum likelihood estimator called the k-MLE and its approximate variance.

However, this likelihood function does not coincide with that given by [5] but with that obtained by using the survival function method of two earlier studies [6, 7]. The k-MLE better improves coverage ratios for confidence intervals than the M-estimator, but the problem of inconsistent coverage ratios remains. [3] extended the martingale procedure of [2] when only the removal times are observed. See [8] for a summary of the likelihoods of the completely observed data given parameters under the various model setups for the general stochastic epidemic model adopted by many researchers such as [5, 6, 9, 10].

Here, the first approach to improving the estimator of the initial number of susceptible individuals was to employ the likelihood function of [5]. A system of equations was derived from the log-likelihood function to find the MLEs of the infection rate and the initial number of susceptible individuals, and a normal limiting distribution was assumed to propose a corresponding approximate Hessian matrix. However, because simulations for the MLE give unstable results such as infinite values for the estimate of the initial number and low coverage ratios for confidence intervals such as the M-estimator, a method of penalized likelihood

function is proposed. See [11] for an example of estimation using a penalized likelihood function.

Simulations were conducted to compare the proposed maximum penalized likelihood estimator called the p-MLE with the k-MLE and the M-estimator of the initial number of susceptible individuals. Then, the proposed method was applied to the Abakaliki smallpox data from Nigeria to compare results with the findings of [2, 4].

The rest of this paper is organized as follows: Section 2 presents the notations and the general stochastic epidemic model. Section 3 describes the estimation methods. Section 4 presents the simulation results. Section 5 considers a numerical example, and Section 6 concludes with a discussion and concluding remarks.

2. Notations and the General Stochastic Epidemic Model

Notations very similar to those of [2, 4] were adopted for the spread of a susceptible-infected-removed infectious disease in a population whose individuals are mixing homogeneously. Suppose that the epidemic is observed over the time interval $[0, T]$ in the population, whose size is $\nu + a$ at time $t = 0$, where a indicates infectious individuals and ν susceptible individuals. Let $S(t)$ denote the number of susceptible individuals present at time t ; $X(t)$ the number of individuals infected up to and including time t , including the initial set of infectious individuals; $I(t)$ the number of infectious individuals present at time t ; $R(t)$ the number of infected individuals removed up to and including time t ; β the infection rate; γ the removal rate; and G_t the σ -algebra generated by the history $\{S(u), I(u); 0 \leq u \leq t\}$. The number of individuals who become infected and are removed by time T , respectively, is denoted by $n_I \leq \nu$ and $n_R \leq \nu$. Note that when an individual becomes infected, the individual is assumed to be immediately infectious. Given $S(t)$ and $I(t)$, assume that the probability of a susceptible individual becoming infected and that of an infectious individual being removed within a small time interval $(t, t + h)$ are given by $\beta/\nu S(t)I(t)h + o(h)$ and $\gamma I(t)h + o(h)$, respectively, such that the transition probability is

$$\begin{aligned} & \Pr \{X(t+h) - X(t) = 1, R(t+h) - R(t) = 0 \mid G_t\} \\ &= \frac{\beta}{\nu} S(t) I(t) h + o(h), \\ & \Pr \{X(t+h) - X(t) = 0, R(t+h) - R(t) = 1 \mid G_t\} \\ &= \gamma I(t) h + o(h), \\ & \Pr \{X(t+h) - X(t) = 0, R(t+h) - R(t) = 0 \mid G_t\} \\ &= 1 - \frac{\beta}{\nu} S(t) I(t) h - \gamma I(t) h + o(h). \end{aligned} \tag{1}$$

The correction term $o(h)$ becomes negligible when h is small; that is, $o(h)/h \rightarrow 0$.

The process $S(t)$, $0 \leq t \leq T$, is assumed not to be observed, such that ν is not observable. However, the

process $I(t)$, $0 \leq t \leq T$, is fully observable, such that the times at which individuals become infected are observable. The observation here includes the infection time for the infected individual and his or her removal time. Let $\phi = (\phi_1, \phi_2, \dots, \phi_{n_I})$ be the ordered successive infection times observed over $(0, T]$. As indicated in [2], the number of individuals infected in $[0, T]$ who are still susceptible at time t can be observed; that is, $S_T(t) = X(T) - X(t-)$. Note that although $I(t)$ depends only on infection and removal times, $S(t)$ depends on ν , as well as infection and removal times.

3. Derivation of the Estimators

First, consider the likelihood function of the parameters β and ν according to [5]

$$\begin{aligned} \mathcal{L}(\beta, \nu \mid \phi) &\propto \left(\prod_{j=1}^{n_I} \frac{\beta}{\nu} S(\phi_j^-) I(\phi_j^-) \right) \\ &\times \exp \left\{ -\frac{\beta}{\nu} \int_0^T S(t) I(t) dt \right\}, \end{aligned} \tag{2}$$

where $S(\phi_j^-)$ and $I(\phi_j^-)$ denote the situation just prior to time ϕ_j . The likelihood function (2) differs from that given by [4]

$$\left(\prod_{j=2}^{n_I} \frac{\beta}{\nu} S(\phi_j^-) I(\phi_j^-) \right) \exp \left\{ -\frac{\beta}{\nu} \int_{\phi_0}^T S(t) I(t) dt \right\}, \tag{3}$$

which is obtained under the same conditions as (2), except that the epidemic process is observed until all infectious individuals are removed, such that the likelihood function can be interpreted as obtained by observing the process over the interval (ϕ_1, ∞) for β and ν . Note that the likelihood functions (2) and (3) are both derived using the definition of a likelihood function in statistical physics based on the counting process theory.

Here, inferences are drawn for ν when the infection process $I(t)$ is observed over a fixed time interval $(0, T]$, which is a relaxation of the condition in [2, 4], where the epidemic is observed until it ceases.

The log-likelihood function is given by taking the logarithm of (2):

$$\begin{aligned} L(\beta, \nu \mid \phi) &\propto n_I (\log \beta - \log \nu) + \sum_{j=1}^{n_I} \log (S(\phi_j) - 1) \\ &- \frac{\beta}{\nu} \int_0^T S(t) I(t) dt. \end{aligned} \tag{4}$$

Here, we use the relation $S(\phi_j^-) = S(\phi_j) - 1$. Note that the total number of susceptible individuals at time u is just $S_T(u)$ plus the number of susceptible individuals not infected by T :

$$S(u) = S_T(u) + \nu + a - X(T), \tag{5}$$

where $S_T(u) = X(T) - X(u^-)$ denotes the number of individuals infected in $(0, T]$ and still susceptible at time u .

With (5) substituted into (4) to use information on those infected at the infection time, the following is obtained:

$$L(\beta, \nu | \phi) \propto n_I (\log \beta - \log \nu) + \sum_{j=1}^{n_I} \log(S_T(\phi_j) + \nu + a - X(T) - 1) - \frac{\beta}{\nu} \{A_1 + (\nu + a - X(T)) A_2\}, \quad (6)$$

where

$$A_1 = \int_0^T S_T(t) I(t) dt, \quad A_2 = \int_0^T I(t) dt. \quad (7)$$

Let the first partial derivatives of the log-likelihood function (4) with respect to β and ν be 0; thus, the system of two nonlinear equations is

$$\ell_1(\beta, \nu) \equiv \beta - \frac{n_I \nu}{g_1(\nu)} = 0, \quad (8)$$

$$\ell_2(\nu) \equiv -\frac{n_I A_2}{g_1(\nu)} + \sum_{j=1}^{n_I} \frac{1}{g_2(\nu, j)} = 0, \quad (9)$$

where

$$g_1(\nu) = A_1 + A_2(a + \nu - X(T)), \quad (10)$$

$$g_2(\nu, j) = S_T(\phi_j) + \nu + a - X(T) - 1.$$

Because there are no terms of β in (9), (9) can be solved with respect to ν to obtain the MLE $\hat{\nu}$ of ν . Then, the MLE $\hat{\nu}$ of ν can be plugged into (8) to get the maximum likelihood estimator of β :

$$\hat{\beta} = \frac{n_I \hat{\nu}}{g_1(\hat{\nu})}, \quad (11)$$

which is the same as the M-estimator of β in [2]. The two nonlinear equations are solved here separately to get solutions, whereas [4] maximizes his log-likelihood function of variables ν and β .

The Hessian matrix can be approximated by

$$H(\hat{\beta}, \hat{\nu}) = \begin{pmatrix} -\frac{n_I}{\hat{\beta}^2} & -\frac{A_2}{\hat{\nu}} + \frac{g_1(\hat{\nu})}{\hat{\nu}^2} \\ -\frac{A_2}{\hat{\nu}} + \frac{g_1(\hat{\nu})}{\hat{\nu}^2} & \frac{2A_2 \hat{\beta}}{\hat{\nu}^2} + \frac{n_I}{\hat{\nu}^2} - \frac{2\hat{\beta} g_1(\hat{\nu})}{\hat{\nu}^3} - \sum_{j=1}^{n_I} \frac{1}{(g_2(\hat{\nu}, j))^2} \end{pmatrix}, \quad (12)$$

assuming that the limiting distribution of $\hat{\beta}$ and $\hat{\nu}$ follows a normal distribution.

Equation (9) does not have a finite solution for some values of the observations of ϕ . To show this, we can first rewrite $\ell_2(\nu)$ as

$$\ell_2(\nu) = \sum_{j=1}^{n_I} \frac{j - n_I + A_1/A_2}{(\nu - j)^2 + (\nu - j)(j - n_I + A_1/A_2)}, \quad (13)$$

by using the relationships $S_T(\phi_j) = n_I - j + 1$ and $X(T) = n_I + a$. It is clear that when $A_1/A_2 > n_I - 1$, $\ell_2(\nu)$ becomes positive such that the solution of (9) should be infinity. Here, we assumed $\nu > n_I$ without loss of generality.

When the number of the initial susceptible individuals is not large enough, the number of simulated epidemics for which the estimates of ν that did not exist cannot be ignored (Table 1). Therefore a method for improving the maximum likelihood estimator $\hat{\nu}$ is proposed by considering a penalized likelihood of (6):

$$L_p(\beta, \nu | \phi) = L(\beta, \nu | \phi)^2 - \text{pen}(\nu | \phi), \quad (14)$$

where

$$\text{pen}(\nu | \phi) = n_I A_2 (\log g_1^+(\nu) - \log g_1(\nu)) + \log g_2(\nu, 1), \quad (15)$$

for $g_1^+(\nu) = A_1 + A_2(a + \nu - X(T) + 1)$, which modifies (9) to

$$\ell_2^p(\nu) \equiv -\frac{n_I A_2}{g_1^+(\nu)} + \sum_{j=2}^{n_I} \frac{1}{g_2(\nu, j)} = 0. \quad (16)$$

Note that $g_1^+(\nu)$ is a modification of $g_1(\nu)$ to make $\ell_2(\nu)$ slightly bigger, and $1/g_2(\nu, 1)$ is subtracted from $\ell_2(\nu)$ to make it slightly smaller. The penalty function $\text{pen}(\nu | \phi)$ is heuristically chosen to penalize a large value of the estimate of ν for the log-likelihood (6). It can be shown that the denominator of the first derivative of $\text{pen}(\nu | \phi)$ with respect to ν is positive and that the numerator is given in the quadratic equation form $\nu^2 + c_1 \nu + c_2$ for constants c_1 and c_2 such that $\text{pen}(\nu | \phi)$ increases as ν increases with $\nu > \nu_0$ for some finite value $\nu_0 > 0$. See [11] for a discussion in choosing a penalty function. Let the estimators of ν and β obtained by solving (8) and (16) be $\hat{\nu}_p$ and $\hat{\beta}_p$, respectively, and be called the p-MLE.

Note that the k-MLE $\hat{\nu}_k$ in [4] can be obtained by

$$\ell_3^p(\nu) \equiv -\frac{n_I A_2}{g_1^\phi(\nu)} + \sum_{j=1}^{n_I} \frac{1}{g_2(\nu, j)} = 0, \quad (17)$$

instead of (16), where $g_1^\phi(\nu) = A_1^\phi + A_2(a + \nu - X(T))$ with $A_1^\phi = \int_{\phi_1}^T S_T(t) I(t) dt$ to reduce the value of $\ell_2(\nu)$ in (9).

The Hessian matrix can be approximated to

$$H(\hat{\beta}_p, \hat{\nu}_p) = \begin{pmatrix} -\frac{n_I}{\hat{\beta}_p^2} & -\frac{A_2}{\hat{\nu}_p} + \frac{g_1^+(\hat{\nu}_p)}{\hat{\nu}_p^2} \\ -\frac{A_2}{\hat{\nu}_p} + \frac{g_1^+(\hat{\nu}_p)}{\hat{\nu}_p^2} & \frac{2A_2 \hat{\beta}_p}{\hat{\nu}_p^2} + \frac{n_I}{\hat{\nu}_p^2} - \frac{2\hat{\beta}_p g_1^+(\hat{\nu}_p)}{\hat{\nu}_p^3} - \sum_{j=2}^{n_I} \frac{1}{(g_2(\hat{\nu}_p, j))^2} \end{pmatrix}, \quad (18)$$

assuming that the limiting distribution of $\hat{\beta}_p$ and $\hat{\nu}_p$ follows a normal distribution. Therefore, the diagonal elements of $(-H(\hat{\beta}_p, \hat{\nu}_p))^{-1}$ can be used to give estimated standard errors $\widehat{\text{se}}(\hat{\nu}_p)$ and $\widehat{\text{se}}(\hat{\beta}_p)$, which may be used to construct approximate nominal 95% confidence intervals of ν and β as

$$\hat{\nu}_p \pm 1.96 \widehat{\text{se}}(\hat{\nu}_p), \quad \hat{\beta}_p \pm 1.96 \widehat{\text{se}}(\hat{\beta}_p), \quad (19)$$

respectively.

TABLE 1: The number of simulated epidemics for which the estimates of $\hat{\nu}$ did not exist for $a = 5$, $\gamma = 1.0$, replication = 1000, and endemic rate ϵ .

ν	(β, ϵ)		
	(1.5, 0.2)	(1.3, 0.2)	(1.3, 0.4)
100	41	85	28
250	3	26	6
1000	0	2	0
5000	0	0	0

4. Simulations

A simulation study very similar to that of [2, 4] was conducted to compare the efficiency of the p-MLE relative to the k-MLE and the M-estimate. Here, populations of $\nu = 100, 250, 1000$, and 5000 susceptible individuals and $a = 5$ initial infectious individuals are considered. In this simulation, $\gamma = 1$ and $\beta = 1.5$ and 1.3 were taken. Results were conditional on a major epidemic, and, therefore, following [2], only simulated epidemics with more than 20% of infected individuals were considered. For $\beta = 1.3$, epidemics with more than 40% of infected individuals were considered. The value of T was set to ∞ to compare simulation results with the findings of [2, 4]. For each combination of parameters, 1000 epidemics were simulated.

The number of simulated epidemics for which the maximum likelihood estimate $\hat{\nu}$ does not exist was counted (Table 1). The results resemble those of [2]. Furthermore, in each scenario, the following were computed: (i) $av(\hat{\beta}_p)$, $av(\hat{\nu}_p)$, $sd(\hat{\beta}_p)$, and $sd(\hat{\nu}_p)$ (averages and standard deviations of 1000 estimates of $\hat{\beta}_p$ and $\hat{\nu}_p$, resp.); (ii) $av(\widehat{se}(\hat{\beta}_p))$ and $av(\widehat{se}(\hat{\nu}_p))$ (averages of estimated approximate standard errors of 1000 estimates of $\hat{\beta}_p$ and $\hat{\nu}_p$, resp.); and (iii) $C(\nu)$ (percentage of 1000 nominal 95% confidence intervals containing ν), the coverage ratio, and $av(N(T))$ (average final size of simulated epidemics). The simulation results are presented in Table 2.

Among the 1000 simulations, there were no cases in which the estimate $\hat{\nu}_p$ was infinite; this held true for the k-MLE as well, as in [4]. In the comparison with the k-MLE and the M-estimator, there were substantial improvements in coverage ratios and averages of estimates of $\hat{\nu}_p$. The coverage ratios for the p-MLE were quite stable in comparison to those for the k-MLE and the M-estimator. An increase in the value of ν increased the coverage ratio such that all coverage ratios for $\nu = 5000$ were close to the true confidence coefficient 0.95.

The average estimates of $\hat{\nu}_p$ were closer to the true value than those of $\hat{\nu}_K$, the k-MLE, for each combination of parameters. The standard deviations $sd(\hat{\nu}_p)$ and the average of estimated approximate standard errors $av(\widehat{se}(\hat{\nu}_p))$ of 1000 estimates of $\hat{\nu}_p$ were quite close to each other in all cases. However, this was not the case for the k-MLE and the M-estimator.

In addition, the averages of 1000 estimates of $\hat{\beta}_p$ tended to increase to the true value with an increase in ν , and the standard deviations of estimates $sd(\hat{\beta}_p)$ and the average of

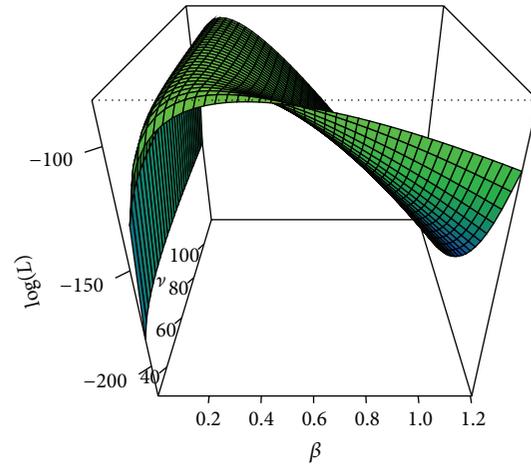


FIGURE 1: The log-likelihood function of β and ν for the Abakaliki smallpox data.

estimated approximate standard errors $av(\widehat{se}(\hat{\beta}_p))$ were also quite close to each other in all cases.

5. Application

The proposed method was applied to the Abakaliki smallpox data from Nigeria. [12] provided 29 infection times for infected individuals and the number of infectious individuals on each day of the epidemic in Abakaliki. The infectious period is assumed to be fixed at 7 days for every individual, and the latent period fixed at 13 days. The estimates $\hat{\beta}_p$ and $\hat{\nu}_p$, as well as corresponding standard errors, were obtained (Table 3). The function `nleqslv` in R was used to solve (16) for $\hat{\nu}_p$. The results indicated that both the estimate and estimated approximate standard error for $\hat{\nu}_p$ were less than those for $\hat{\nu}_K$ and $\hat{\nu}_M$. The estimate (33.88) and estimated approximate standard error for $\hat{\nu}_p$ (4.13) were close to those for $\hat{\nu}_K$ (35.27 and 6.70, resp.) but lower than those for $\hat{\nu}_M$ (42.12 and 37.15, resp.). The estimate of initial susceptible individuals was much lower than 120, the population size. Because of the assumption of homogeneous mixing between individuals, as in earlier studies [2, 4, 12, 13], this estimate was interpreted assuming that a number of individuals were not susceptible to the disease owing to natural immunity, vaccination, or isolation.

Figure 1 shows the shape of the log-likelihood function of β and ν , which is similar to that in [4], and Figure 2 presents the gradient of the likelihood function of ν for the smallpox data.

6. Discussion and Conclusions

In any epidemic, the estimation of the initial number of susceptible individuals is of great interest. Although [5] derived a log-likelihood function for the initial number and the infection rate when the epidemic is fully observed over a given time interval, no study has investigated the properties of the MLEs. [4] considered the log-likelihood function to

TABLE 2: Simulation results for $\alpha = 5, \gamma = 1.0$, replications = 1000, and $\epsilon =$ endemic rate.

(β, ϵ)	ν	$av(\hat{\beta}_p)$	$sd(\hat{\beta}_p)$	$av(\widehat{se}(\hat{\beta}_p))$	$av(\hat{\nu}_p)$	$sd(\hat{\nu}_p)$	$av(\widehat{se}(\hat{\nu}_p))$	$C(\nu)^a$	$av(N(T))$
(1.5, 0.2)	100	1.34	0.20	0.25	98.7	30.6	36.9	89.0	57.6
	250	1.45	0.16	0.19	247.6	58.1	60.8	89.4	143.4
	1000	1.49	0.10	0.10	1004.1	117.9	118.2	93.1	578.6
	5000	1.50	0.05	0.05	5011.0	249.3	247.7	94.7	2907.1
(1.3, 0.2)	100	1.21	0.18	0.24	93.4	34.0	43.7	86.5	48.4
	250	1.31	0.15	0.19	227.0	71.7	79.9	82.3	112.6
	1000	1.32	0.09	0.11	947.0	209.2	213.0	85.8	425.2
	5000	1.30	0.05	0.05	4965.1	540.7	536.1	92.5	2107.4
(1.3, 0.4)	100	1.24	0.19	0.23	104.6	30.5	42.8	92.5	57.2
	250	1.32	0.15	0.18	251.8	63.6	76.1	90.7	131.4
	1000	1.33	0.09	0.10	1002.9	178.3	190.9	91.7	473.9
	5000	1.31	0.05	0.05	5020.8	488.4	492.0	94.6	2205.1

^aCoverage ratio for ν .

TABLE 3: Estimates of the infection rate for initial susceptible individuals based on the Abakaliki smallpox data.

Parameter	Estimate	Standard error	Nominal 95% confidence interval
β	0.3133	0.0773	(0.1618, 0.4649)
ν	33.88	4.13	(25.79, 41.98)

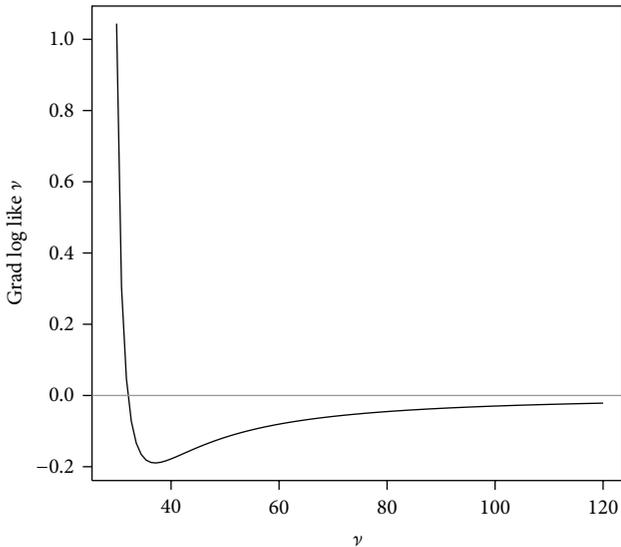


FIGURE 2: The gradient of the log-likelihood function of ν for the Abakaliki smallpox data.

be obtained when the epidemic process is observed from the time of the first infection to the time when all infectious individuals are removed and derived the MLE of parameters of interest. [2] used a martingale framework to propose an estimator. The present study used the log-likelihood function of [5] and the relationship between S and S_T of [2] to derive a system of equations for parameters and solved the system to obtain MLEs of parameters, instead of finding them as [4] did, by maximizing the log-likelihood function. An approximate

Hessian matrix of estimators $\hat{\nu}_p$ and $\hat{\beta}_p$ was derived based on the assumption that the limiting distribution of $\hat{\nu}_p$ and $\hat{\beta}_p$ follows a normal distribution. The derivation of the limiting distribution of $\hat{\beta}_p$ and $\hat{\nu}_p$ can be a challenge.

Our modification of (9) can be considered the same as penalizing the likelihood function (4) with a suitable penalizing factor. Because there are various penalizing methods, another penalized likelihood function can be attempted for better estimators in a future study.

The simulations for the p-MLE provide a more stable result than the M-estimator and the k-MLE for unbiasedness, standard errors, and coverage ratios, and therefore, the proposed method can be used as a more reliable tool for estimating the initial number of susceptible individuals in a population.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by the 2012 Yeungnam University Research Grant.

References

- [1] E. B. O’Dea, K. M. Pepin, B. A. Lopman, and C. O. Wilke, “Fitting outbreak models to data from many small norovirus outbreaks,” *Epidemics*, vol. 6, pp. 18–29, 2014.
- [2] R. M. Huggins, P. S. F. Yip, and E. H. Y. Lau, “A note on the estimation of the initial number of susceptible individuals in the general epidemic model,” *Statistics and Probability Letters*, vol. 67, no. 4, pp. 321–330, 2004.
- [3] E. H. Y. Lau and P. S. F. Yip, “Estimating the basic reproductive number in the general epidemic model with an unknown initial number of susceptible individuals,” *Scandinavian Journal of Statistics*, vol. 35, no. 4, pp. 650–663, 2008.

- [4] T. Kypraios, "A note on maximum likelihood estimation of the initial number of susceptibles in the general stochastic epidemic model," *Statistics and Probability Letters*, vol. 79, no. 18, pp. 1972–1976, 2009.
- [5] H. Andersson and T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis*, vol. 151, Springer, New York, NY, USA, 2000.
- [6] P. D. O'Neill and G. O. Roberts, "Bayesian inference for partially observed stochastic epidemics," *Journal of the Royal Statistical Society A*, vol. 162, no. 1, pp. 121–129, 1999.
- [7] M. Höhle and E. Jørgensen, "Estimating parameters for stochastic epidemics," Dina Research Report 102, Danish Institute of Agricultural Sciences, Tjele, Denmark, 2002.
- [8] T. Kypraios, *Efficient Bayesian inference for partially observed stochastic epidemics and a new class of semi-parametric time series models [Ph.D. thesis]*, Department of Mathematics and Statistics, Lancaster University, Lancaster, UK, 2007.
- [9] P. Neal and G. Roberts, "A case study in non-centering for data augmentation: stochastic epidemics," *Statistics and Computing*, vol. 15, no. 4, pp. 315–327, 2005.
- [10] N. G. Becker and T. Britton, "Statistical studies of infectious disease incidence," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 61, no. 2, pp. 287–307, 1999.
- [11] M. Moreno and S. R. Lele, "Improved estimation of site occupancy using penalized likelihood," *Ecology*, vol. 91, no. 2, pp. 341–346, 2010.
- [12] N. G. Becker, *Analysis of Infectious Disease Data*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, UK, 1989.
- [13] N. G. Becker and P. Yip, "Analysis of variations in an infection rate," *Australian Journal of Statistics*, vol. 31, pp. 42–52, 1989.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

