Mathematical Biosciences 246 (2013) 272–279

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Estimation of the Malthusian parameter in an stochastic epidemic model using martingale methods

David Lindenstrand, Åke Svensson *

Department of Mathematics, Stockholm university, SE-10691 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 26 March 2013
Received in revised form 18 September 2013
Accepted 14 October 2013
Available online 24 October 2013

Keywords:
Malthusian parameter
Stochastic epidemic model
Growth rate
Generation time distribution

ABSTRACT

Data, on the number of infected, gathered from a large epidemic outbreak can be used to estimate parameters related to the strength and speed of the spread. The Malthusian parameter, which determines the initial growth rate of the epidemic is often of crucial interest. Using a simple epidemic SEIR model with known generation time distribution, we define and analyze an estimate, based on martingale methods. We derive asymptotic properties of the estimate and compare them to the results from simulations of the epidemic. The estimate uses all the information contained in the epidemic curve, in contrast to estimates which only use data from the start of the outbreak.

1. Introduction

When considering possible effects of an epidemic outbreak of an infectious disease several characteristics are of interest. Traditionally the basic reproduction number, R_0, is considered important. It is defined as the mean number of persons an infected infects in a totally susceptible population. R_0 is closely related to the proportion of a population that finally will be infected during an epidemic outbreak, and can be seen as a primary measure of the infectivity of the infectious agent considered. Recently the speed at which the epidemic grows has attracted considerable interest. Together with R_0 the speed decides the possibilities to control spread of an infection.

Experience indicates that the number of infected persons initially grows at a constant exponential rate, i.e. proportional to $\exp(rt)$, where t is the time the spread has been going on in the population. The increase will slow down as the number of infected and immune in the population grows. Theoretically the exponential behavior, in the start, is a consequence of that an epidemic, in the initial phase, is well approximated by a branching process. Branching processes have been studied for long and much is known of their behavior cf e.g. [1]. The parameter r is often referred to as the Malthusian parameter. A relation between R_0 and r, which depends on the generation time *distribution*, is given by the Euler–Lotka equation (cf Section 3). A discussion of generation time distributions, the basic reproduction number and the Malthusian parameter in epidemic models can be found in [2].

The epidemic curve describes how the number of infected persons grows with time. The purpose of this paper is to suggest methods to derive estimates of both the basic reproduction number and the Malthusian parameter based on possibly partial observations of the epidemic curve. As regard estimates of r simple methods have been suggested that are restricted to using data from periods where the branching process approximation is applicable (see Section 2). In order to use the information contained in longer sections of the epidemic curve we will use assumptions on the generation time distribution.

The analysis is based on a class of simple models defined in Section 3. We will consider spread in a closed population with n persons in which infections are transmitted according to homogeneous mixing, i.e., when a transmission occurs the new infected is a susceptible member of the population chosen uniformly at random. The epidemic is started by one infected individual and a person may only be infected once.

We will preferably use martingale estimators. Becker, [3,4], gives a thorough introduction to martingale methods for estimation of infection rates. Becker also derives estimates of R_0, and their standard deviations, for the standard SIR epidemic model. In particular he studies models where the infectious period is assumed to be exponentially distributed. This assumption makes it possible to derive estimates that depend on the number of removed persons. He also considers models with other types of mixing.

In Section 4 we assume that each infected person is infectious during a fixed time interval. We derive the generation time distribution and find a relation between the basic reproduction number and the Malthusian parameter using the Euler–Lotka equation. With this assumption it is possible, knowing the epidemic curve,

* Corresponding author. Tel.: +46 8164569.
E-mail address: akes@math.su.se (Å. Svensson).

0025-5564/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2013.10.002

to calculate, at each time, the number of infectious persons. Using this we can define estimators of R_0 and r and derive their asymptotic properties. Using theoretical results and simulations we illustrate, what is gained by using this martingale estimator. In this special case we can also derive, as an alternative, Maximum Likelihood estimators.

The simple model with a fixed infectious period is used to illustrate the methods. In Section 5 more complicated generation time distributions are considered and it is illustrated how the estimation method can be generalized.

2. Estimates of the growth rate based on the initial phase of the epidemic

As pointed out in the introduction the number of infections grows initially proportional to $\exp(rt)$. Eventually the progress slows down, when a substantial proportion of the population is infected and therefore immune.

An obvious, and often used idea, (see e.g. [5]), is to use observations from the initial part of the epidemic to estimate r. Denote the (counting) process of *infected* individuals by $\{N_n(t); t \geqslant 0\}$, with $N_n(0) = 0$ where the subindex n refers to the size of initially susceptible population. If we observe N_n at times t_1 and t_2 we have

$$\tilde{r}_{simple} = \frac{log\{N_n(t_2)/N_n(t_1)\}}{t_2 - t_1} \tag{2.1}$$

This estimate will degenerate if the branching process approximation is not appropriate in at least one of the observation times. If this is the case may e.g. be judged by a simple inspection of the epidemic curve. Another possibility is to use a statistical test to see if the progress is slowing down. Such a method is suggested in [6]. Methods of this kind have the advantage that they do not use any assumptions of anything else than the initial part of the epidemic curve.

The alternative methods that are discussed in the remainder of this paper use data also from times when the growth of the epidemic has slowed down but require the use of assumptions on the generation time distribution.

3. Basic model and the Euler–Lotka equation

We introduce one *infectious* individual in a closed, homogeneously mixing, population of n susceptible individuals. All individuals make contact with other particular individuals according to independent Poisson processes with rate $\frac{\alpha}{n}$. Let Y be the length of the time interval during which an infected person is infectious. The infectious time may be proceeded by a latent time, with length X, during which the infection is not transmitted. In general both these times may be individually random.

Each infected individual has a random number of potentially contagious contacts. We denote this number by ξ. Now, given the length of the infectious time, ξ is Poisson distributed with parameter αY.

The basic reproduction number is given by $R_0 = E[\xi] = \alpha E(Y)$. A large epidemic outbreak is possible exactly when $R_0 > 1$. Let τ_n denote the final proportion infected individuals after a large epidemic outbreak, i.e. $\tau_n = N_n(\infty)/n$. It is well known that τ_n converges in probability to τ as $n \to \infty$, where τ is the solution to the equation (see e.g. [7])

$$1 - \tau = e^{-\tau R_0}, \tag{3.1}$$

or equivalently

$$-\log(1 - \tau) = R_0 \tau$$

The generation time distribution is defined by the density

$$g(t) = \frac{P(X \leqslant t < X + Y)}{E(Y)} = \frac{P(X + Y > t) - P(X > t)}{E(Y)}. \tag{3.2}$$

(cf [2]).

Let

$$h(r) = \int_0^\infty e^{-rt} P(X < t \leqslant X + Y) dt. \tag{3.3}$$

The Euler–Lotka equation which has the Malthusian parameter, r, as a solution can be written as

$$1 = R_0 \int_0^\infty e^{-rt} g(t) dt = \alpha h(r). \tag{3.4}$$

From Eq. (3.1) we observe that there is a one-to-one relation between α and the final size τ. Since (3.4) implies that there is a one-to-one relation between α and r there is also a relation between r and the final size:

$$-\log(1 - \tau) = \tau E(Y)/h(r). \tag{3.5}$$

4. Simple model

To illustrate how estimates of α and r can be derived we will study an extremely simple model that allows us to highlight main points of the analysis. In Section 5 we will discuss how the methods can be extended to more complicated models. For the moment we will assume that there is no latent time and that the infectious time Y is nonrandom and covers one time unit.

Let N_n be as defined in Section 2. At time t an infectious individual spreads the disease at the rate $\frac{\alpha}{n}(n - N_n(t-)) = \alpha(1 - N_n(t-)/n)$ where $(1 - N_n(t-)/n)$ is the proportion of susceptible individuals in the population. The number of infectious individuals at time t is $I(t) = (N_n(t-) - N_n((t-1) \vee 0) + \mathbf{1}_{(0,1]}$, where $\mathbf{1}_{(0,1]}$ is the contribution of the initial infected. Thus the total infectious pressure in the population at time t is $\alpha\lambda_n(t)$ where

$$\lambda_n(t) = (1 - N_n(t-)/n) I(t),$$

This model is a S (Susceptibles)-I (Infectious)-R (Removed) model with

$$\begin{aligned} S(t) &= n - N_n(t-) \\ I(t) &= N_n(t-) - N_n((t-1) \vee 0) + \mathbf{1}_{(0,1]} \\ R(t) &= N_n((t-1) \vee 0) + \mathbf{1}_{(1,\infty]}. \end{aligned}$$

Observe that this epidemic process is completely specified by the process N_n. It is possible to derive estimates of α and the Malthusian parameter observing only the process N_n. This is the reason why we choose to investigate this model first.

In this simple model $R_0 = \alpha$ and

$$h(r) = \int_0^1 e^{-rt} dt = \frac{1 - e^{-r}}{r} \tag{4.1}$$

According to Eq. (3.4),

$$R_0 = \alpha = \frac{r}{1 - e^{-r}}. \tag{4.2}$$

We will first estimate α and then use the relation (4.2) to obtain an estimate of r.

4.1. Inference

We will start by defining and analyzing the properties of an estimate of the contact intensity α. The discussion uses the