J. R. Statist. Soc. B (1998)
60, *Part* 4, *pp.* 663–679

Estimation in multitype epidemics

Tom Britton†

La Trobe University, Bundoora, Australia

[Received February 1997. Revised January 1998]

Summary. A multitype epidemic model is analysed assuming proportionate mixing between types. Estimation procedures for the susceptibilities and infectivities are derived for three sets of data: complete data, meaning that the whole epidemic process is observed continuously; the removal processes are observed continuously; only the final state is observed. Under the assumption of a major outbreak in a population of size n it is shown that, for all three data sets, the susceptibility estimators are always efficient, i.e. consistent with a $\sqrt{n}$ rate of convergence. The infectivity estimators are 'in most cases' respectively efficient, efficient and unidentifiable. However, if some susceptibilities are equal then the corresponding infectivity estimators are respectively barely consistent ($\sqrt{\log(n)}$ rate of convergence), *not* consistent and unidentifiable. The estimators are applied to simulated data.

Keywords: Consistent estimator; Counting processes; Estimating equations; Martingales; Multitype epidemics; Susceptibility and infectivity

1. Introduction

This paper concerns estimation procedures for epidemics where individuals are not homogeneous but may be classified into different types, assuming homogeneity in terms of susceptibility and infectivity within each type. Information about these parameters is important for better understanding the spreading mechanism of the disease, but also when aiming at controlling future epidemics in that groups with high infectivity and/or susceptibility should receive extra attention.

The disease is assumed to be an S–I–R (susceptible–infectious–removed) infectious disease (Lefèvre, 1990). The underlying epidemic model adopts the assumption of proportionate mixing between individuals, meaning that the rate of infection between two individuals is a product of two terms: the *infectivity* of the infectious individual and the *susceptibility* of the susceptible individual. In Section 2 the model is defined in detail.

In Section 3 we derive estimators of the parameters when the complete data are available, i.e. the time of infection and the time of removal are known for all infected individuals. For such data the likelihood is explicit and hence maximum likelihood (ML) estimation is feasible, also giving approximate confidence regions for the parameters. For a homogeneous population Rida (1991) studied properties of several estimators for this type of data.

Because complete data from epidemics are rarely available two different types of partial observation of the epidemic are treated in Section 4. The first type is where the removal process is observed continuously, i.e. for each individual we know whether he or she became

†*Address for correspondence*: Department of Mathematics, Uppsala University, Box 480, S-751 06 Uppsala, Sweden.
E-mail: tom.britton@math.uu.se

 1369–7412/98/60663

infected, and if he or she did we also know the time of removal; see Becker and Hasofer (1997) for estimation procedures in a homogeneous population for such data. The second set of partial data considered is the *final state* of the epidemic, i.e. a binary variable for each individual indicating whether or not he or she became infected. When an epidemic is only partially observed the likelihood is not explicit so estimation techniques that are *not* based on the likelihood must be applied. The susceptibility estimators for the two types of partial data coincide. They are obtained by using the method of moments by equating suitable martingales equal to their means, and martingale theory is used to obtain approximate confidence regions for the parameters. The infectivities are unidentifiable if only the final state is observed. For the removal data the infectivity estimators are constructed from the system of differential equations that is associated with the deterministic version of the epidemic model. Unfortunately, no explicit confidence regions are given, only their asymptotic order. In Section 5 the performances of the estimators are examined on simulated data. The paper concludes with a discussion on possible extensions in future research.

The main result of the paper shows that the susceptibilites can always be estimated efficiently ($\sqrt{n}$-convergence) whereas a type-specific infectivity can only be estimated efficiently if the corresponding susceptibility differs from all other susceptibilities. Otherwise the estimator is confounded with other infectivity estimators. In fact, if the complete data are observed such confounded estimators converge at the very slow rate $\sqrt{\log(n)}$, and the estimator suggested for the removal data is not even consistent. A consequence of the result is that, if individuals in a group are believed to be equally susceptible, then a postulated hypothesis claiming that a certain subgroup has higher or lower infectivity than the others, for natural reasons or because of preventive measures, has little chance of gaining empirical evidence or of being rejected.

Recently, Rhodes *et al.* (1996) studied estimation procedures in a similar model for several levels of informative data, mainly more detailed than those of the present paper in that they contain information about the actual contacts.

2. The model

Consider a closed population consisting of n individuals. Each individual is classified as one of k different *types* labelled $i = 1, \ldots, k$. Let n_i denote the number of i-individuals ($\Sigma_i n_i = n$) and $\pi_i = n_i/n$ denotes the corresponding population proportion.

We now define a continuous time S–I–R epidemic model for this population. Let $S_i(t)$, $I_i(t)$ and $R_i(t)$ denote the number of susceptible, infective and removed individuals of type i at time t, $i = 1, \ldots, k$, $0 \leqslant t < \infty$. Because the population is closed we always have $S_i(t) + I_i(t) + R_i(t) = n_i$. At the start of the epidemic ($t = 0$) all individuals are susceptible to the disease (if the population contains immune individuals these are assumed to be known and treated as not belonging to the population). At this time a small number of individuals are infected because of some external source. As the epidemic evolves, individuals behave in the following way. Whenever a susceptible individual becomes infected he or she immediately becomes infectious and remains so for an exponentially distributed time with mean $1/\gamma$, independent of which type he or she is. During a j-individual's infectious period this individual has 'close contact' with a given i-individual at rate $\alpha_j\beta_i/n$. A close contact is defined as a contact that will result in infection if the other individual is susceptible. Other close contacts have no effect. When the infectious period is over, the individual recovers, becomes immune and plays no further role in the epidemic. This state is called removed. The epidemic evolves until the first time τ when there are no infectious individuals in the

population. When this happens no-one can become infected and we say that the epidemic has terminated. All infectious periods and contact processes are defined to be mutually independent.

An equivalent definition of the model is by means of counting processes. Let $\mathcal{F}_t$ denote the history of the epidemic, i.e. the natural filtration generated by the epidemic: $\mathcal{F}_t = \sigma(S_i(u), I_i(u), R_i(u); 0 \leqslant u \leqslant t, i = 1, \ldots, k)$. The counting processes $N_i(t) = n_i - S_i(t)$ and $R_i(t)$ are adapted to $\mathcal{F}_t$. It should be clear that the model defined above is then equivalent to defining the following intensities of the counting processes:

$$\begin{aligned} N_i(t) = n_i - S_i(t) \quad &\text{has intensity} \quad \beta_i \bar{S}_i(t-)\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(t-), \quad \text{and} \\ R_i(t) \quad &\text{has intensity} \quad \gamma I_i(t-), \; i = 1, \ldots, k. \end{aligned} \tag{2.1}$$

where $\bar{S}_i(u) = S_i(u)/n$ (similar notation will be used in all processes).

As the model is defined it is overparameterized. This follows from the observation that multiplying all infectivities by a constant and dividing all susceptibilities by the same constant does not change the intensities in expression (2.1). Thus, one linear combination of the susceptibilities (or the infectivities) may be chosen arbitrarily and assumed to be fixed. As is customary, we call the coefficients $\{\alpha_j\}$ (relative) infectivities, reflecting both infectiousness and social activity, and the coefficients $\{\beta_i\}$ (relative) susceptibilities, depending on immunological factors as well as social activity.

An important parameter in epidemic theory is the basic reproduction number R_0, a parameter defined as the mean number of new infections that a *typical* individual causes in the initial part of the epidemic. For the model defined above

$$R_0 = \gamma^{-1} \sum_i \alpha_i \beta_i \pi_i$$

(e.g. Becker and Marschner (1990)). This can be explained from the fact that an infected i-individual on average infects $\gamma^{-1}\alpha_i \Sigma_k \beta_k \pi_k$ individuals when the whole population is susceptible, and the probability that an infected individual in the beginning of the epidemic is an i-individual is $\beta_i \pi_i / \Sigma_k \beta_k \pi_k$. It is well known that the probability of a major outbreak tends to 0 as $n \to \infty$ if and only if $R_0 \leqslant 1$ (e.g. Ball (1983)). The asymptotic results of Sections 3 and 4 are all on the part of the sample where a major outbreak occurs. Thus, for the statements not to be empty it is natural to assume that

$$R_0 = \gamma^{-1} \sum_i \alpha_i \beta_i \pi_i > 1.$$

3. Estimation under complete observation

In this section we derive ML estimators and investigate their properties, for the case that we observe the whole epidemic process, i.e. all of $(S_i(t), I_i(t), R_i(t); 0 \leqslant t \leqslant \tau, i = 1, \ldots, k)$. Such detailed data are rarely available in applications but these estimators can serve as a reference when studying properties of estimators obtained from more frequently available and less informative data.

3.1. Deriving estimators

As mentioned in the previous section we may without loss of generality choose one linear constraint on the infectivities. It turns out to be convenient to let $\boldsymbol{\alpha}$ satisfy the stochastic constraint

$$\sum_{i=1}^{k} \alpha_i \pi_i \tilde{p}_i = \gamma, \tag{3.1}$$

where $\tilde{p}_i = R_i(\infty)/n_i$ denotes the proportion of all i-types that were ultimately infected. This means that γ is expressed in terms of the other parameters and will hence not be shown in the log-likelihood below.

It follows from general theory for counting processes (e.g. Andersen *et al.* (1993), p. 402) that if the epidemic defined in Section 2 is observed up to some time t then the log-partial-likelihood of the data is given by

$$l_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{k} \int_0^t \log\{\beta_i \bar{S}_i(u-)\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u-)\}\, \mathrm{d}N_i(u) - \beta_i \bar{S}_i(u-)\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u-)\, \mathrm{d}u$$

$$+ \sum_{i=1}^{k} \int_0^t \log\{\gamma I_i(u-)\}\, \mathrm{d}R_i(u) - \gamma I_i(u-)\, \mathrm{d}u.$$

The ML estimator for these data is obtained by differentiating the log-partial-likelihood with respect to each parameter, setting each such derivative equal to 0, and solving the system of equations. When differentiating with respect to α_i equation (3.1) must be kept in mind. Straightforward calculations yield the following derivatives:

$$\frac{\partial l_t(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_j} = \sum_{i=1}^{k} \int_0^t \frac{I_j(u-)}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u-)} \{\mathrm{d}N_i(u) - \beta_i \bar{S}_i(u)\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u)\, \mathrm{d}u\} + \frac{\pi_j \tilde{p}_j}{\gamma} \sum_{i=1}^{k} \int_0^t \mathrm{d}R_i(u) - \gamma I_i(u)\, \mathrm{d}u$$

$$\frac{\partial l_t(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_i} = \int_0^t \frac{1}{\beta_i} \{\mathrm{d}N_i(u) - \beta_i \bar{S}_i(u)\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u)\, \mathrm{d}u\}. \tag{3.2}$$

These expressions can be simplified. Let $N(t) = \Sigma_i N_i(t)$ and $R(t) = \Sigma_i R_i(t)$, the total number of infected and removed individuals respectively,

$$G_{i,j}(t) = \int_0^t \bar{S}_i(u)\, I_j(u)\, \mathrm{d}u$$

and

$$H(t) = \sum_i \int_0^t I_i(u)\, \mathrm{d}u.$$

Then the equations above may be written as

$$\frac{\partial l_t(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_j} = \int_0^t \frac{I_j(u-)}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u-)}\, \mathrm{d}N(u) - \sum_i \beta_i G_{i,j}(t) + \frac{\pi_j \tilde{p}_j}{\gamma} R(t) - \pi_j \tilde{p}_j H(t), \tag{3.3}$$

$$\frac{\partial l_t(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_i} = \frac{N_i(t)}{\beta_i} - \sum_j \alpha_j G_{i,j}(t). \tag{3.4}$$

From equation (3.4) we see that the ML estimator of β_i when the epidemic is observed until the end ($t = \tau$), as a function of $\boldsymbol{\alpha}$, is given by

$$\hat{\beta}_i = \hat{\beta}_i(\boldsymbol{\alpha}) = N_i(\tau) \Big/ \sum_j \alpha_j G_{i,j}(\tau), \qquad i = 1, \ldots, k.$$

The ML estimator for $\boldsymbol{\alpha}$, and hence also for $\gamma = \gamma(\boldsymbol{\alpha})$ because of equation (3.1), is the solution to the equations

$$\int_0^\tau \frac{I_j(u-)}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u-)}\,\mathrm{d}N(u) - \sum_i \hat{\beta}_i(\boldsymbol{\alpha})\,G_{i,j}(\tau) + \frac{\pi_j\tilde{p}_j}{\gamma(\boldsymbol{\alpha})}R(\tau) = \pi_j\tilde{p}_j\,H(\tau), \qquad j = 1, \ldots, k.$$

This solution $\hat{\boldsymbol{\alpha}}$ must be obtained numerically. If the epidemic is only observed until $t < \tau$ then the corresponding estimators are the same, just replacing τ by t. In Section 5 these and other estimators are applied to simulated data.

3.2. Variance of estimators

Under some regularity conditions the (asymptotic) variance matrix of the ML estimators is given by the inverse of the *observed information matrix*, and the observed information matrix is defined as the matrix of second-order partial derivatives of the log-likelihood multiplied by -1. This matrix, which we denote by $\Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta})$, contains as elements

$$-\frac{\partial^2 l_\tau(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial\alpha_j\partial\alpha_{j'}} = \int_0^\tau \frac{I_j(u-)\,I_{j'}(u-)}{\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u-)\}^2}\,\mathrm{d}N(u) + \frac{\pi_j\tilde{p}_j\pi_{j'}\tilde{p}_{j'}}{\gamma}R(\tau), \tag{3.5}$$

$$-\frac{\partial^2 l_\tau(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial\alpha_j\partial\beta_i} = \int_0^\tau \bar{S}_i(u)\,I_j(u)\,\mathrm{d}u, \tag{3.6}$$

$$-\frac{\partial^2 l_\tau(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial\beta_i\partial\beta_{i'}} = \begin{cases} N_i(\tau)/\beta_i^2 & \text{if } i = i', \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Being a matrix of stochastic integrals, $\Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta})$ will have an inverse with probability 1. So, for a given realization of an epidemic which is observed continuously, the variance matrix may be estimated by numerically inverting the matrix $\Sigma(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$.

3.3. Consistency and rate of convergence of estimators

The derivatives of the log-partial-likelihood, equations (3.3) and (3.4), evaluated at the true parameter values $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are martingales, viewed as processes indexed by t. These processes are known as score processes. It is a general property that score processes are martingales but it can also be seen from equation (3.2): integrating predictable processes with respect to martingales gives us new martingales. The ML estimators are obtained by setting the martingales equal to their means (equal to 0). This enables us to use theory for martingales to obtain confidence intervals for the estimators. Theorems VI.1.1 and VI.1.2 in Andersen *et al.* (1993) give asymptotic properties of ML estimators in counting process models which are used in the theorem below. First we define the deterministic matrix Σ_{d} to which $n^{-1}\,\Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta})$ converges in probability. For this, let $\sigma_i(t)$, $\iota_i(t)$, $\rho_i(t)$ and $\nu_i(t)$ denote the deterministic counterparts of $\bar{S}_i(t)$, $\bar{I}_i(t)$, $\bar{R}_i(t)$ and $\bar{N}_i(t)$, i.e. let them be solutions to the system of differential equations

$$\begin{aligned} \nu_i'(t) &= \beta_i\,\sigma_i(t)\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\iota}(t), \\ \rho_i'(t) &= \gamma\,\iota_i(t), \end{aligned} \tag{3.8}$$

(to be compared with expression (2.1)) with initial condition $\sigma_i(0) = \pi_i(1 - \beta_i\epsilon)$, $\iota_i(0) = \pi_i\beta_i\epsilon$ and $\rho_i(0) = 0$, where ϵ is a very small number. The reason for choosing the initial proportions

infective proportional to $\pi_i\beta_i$ comes from the branching approximation of the start of the epidemic. The asymptotic-type proportions *alive* (which corresponds to infectious in epidemics) satisfy this relationship almost surely on the set of non-extinction (e.g. Jagers (1975), p. 95). Below we analyse certain integrals of these functions, and it is not difficult to show that these integrals converge as $\epsilon \to 0$. This is the solution that we consider because it is equivalent to starting with a few initially infective individuals and assuming a major outbreak (Ball and Clancy, 1993). Let $p_i = \nu_i(\infty)/\pi_i = \rho_i(\infty)/\pi_i$ denote the final proportion infected among i-individuals in the deterministic model and $p = \Sigma_i \pi_i p_i$ the corresponding overall proportion. Then $\{p_i\}$ is the unique positive solution to the system of equations given by

$$1 - p_i = \exp\left(-\frac{\beta_i}{\gamma} \sum_j \pi_j p_j \alpha_j \right) = \exp(-\beta_i), \qquad i = 1, \ldots, k, \tag{3.9}$$

e.g. Becker and Marschner (1990). The last equality follows from the linear constraint $\Sigma_j \pi_j p_j \alpha_j = \gamma$ corresponding to equation (3.1) for the stochastic model.

Define Σ_d as the symmetric $2k \times 2k$ matrix, indexed like $\Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta})$, with elements

$$\int_0^\infty \frac{\iota_j(t)\,\iota_{j'}(t)}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\iota}(t)} \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\sigma}(t)\,\mathrm{d}t + \frac{\pi_j p_j \pi_{j'} p_{j'}}{\gamma} \int_0^\infty \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\sigma}(t)\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\iota}(t)\,\mathrm{d}t = \int_0^\infty \iota_j(t)\,\iota_{j'}(t) \frac{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\sigma}(t)}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\iota}(t)}\,\mathrm{d}t + \frac{\pi_j p_j \pi_{j'} p_{j'}}{\gamma} p, \tag{3.10}$$

$$\int_0^\infty \sigma_i(t)\,\iota_j(t)\,\mathrm{d}t, \tag{3.11}$$

$$\begin{cases} (1/\beta_i) \displaystyle\int_0^\infty \sigma_i(t)\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\iota}(t)\,\mathrm{d}t & \\ 0 \qquad\qquad i \neq i', \end{cases} = \begin{cases} \pi_i p_i/\beta_i^2 & \\ 0 \qquad\qquad i \neq i', \end{cases} \tag{3.12}$$

where the equalities follow from equations (3.8) and the definition of $\{p_i\}$. Note the similarities with equations (3.5)–(3.7).

The proof of the following theorem is found in Appendix A.

Theorem 1. Assume that the whole epidemic process is observed until the end. Then, on the part of the sample space where there is a major epidemic,

$$\sqrt{n}\Sigma_d \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \end{pmatrix} \xrightarrow{\mathrm{d}} N(\mathbf{0}, \Sigma_d) \qquad \text{as } n \to \infty,$$

where Σ_d is the deterministic matrix with elements defined by expressions (3.10)–(3.12). Further, if the true underlying susceptibilities are all unique, i.e. $i \neq j \Longrightarrow \beta_i \neq \beta_j$, then the matrix Σ_d is invertible for almost all infectivity vectors $\boldsymbol{\alpha}$ (i.e. except a null set), and for such $\boldsymbol{\alpha}$

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \end{pmatrix} \xrightarrow{\mathrm{d}} N(\mathbf{0}, \Sigma_d^{-1}),$$

and the matrix Σ_d^{-1} may be estimated consistently by $n^{-1}\,\Sigma(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1}$.

Remark 1. When some susceptibilities are identical the matrix Σ_d is not invertible. If for example $\beta_1 = \beta_2$, it is easy to show that the first row in Σ_d is just a multiple of the second row. However, from the first part of the theorem we can still estimate all susceptibilities and

for each class of types with equal susceptibility we can estimate the corresponding sum of infectivities weighted by the type frequencies $\{\pi_i\}$ with $\sqrt{n}$-convergence. For example, suppose that $\beta_1 = \beta_2 \neq \beta_k$, $k > 2$. Then we can estimate $\pi_1\alpha_1 + \pi_2\alpha_2$ with $\sqrt{n}$-convergence, but not α_1 and α_2 separately.

Remark 2. The assumption of a major outbreak is natural for consistent estimation in epidemic models (e.g. Rida (1991)). As $n \to \infty$ the dynamics of the start of the epidemic tends to that of a multitype branching process (BP) with the same R_0. In particular if the BP dies out then, with a suitable coupling argument, so does the epidemic and then only a finite amount of information is ever available, thus giving bounds on the amount of information contained in a minor epidemic for any finite n.

A relevant question in light of the theorem above is the following: is it possible to estimate the separate infectivities consistently in case the susceptibilities are equal? The somewhat surprising answer is yes, but at the much slower rate of convergence $\sqrt{\log(n)}$.

Because the susceptibilities and certain linear combinations of the infectivities were possible to estimate with a faster rate of convergence we shall assume that these parameters are known when showing the slower rate of convergence for the separate infectivity estimators. To simplify matters even more we assume that there are only two different types of individual. Thus, assume that $\beta_1 = \beta_2 =: \beta$ and $\gamma = \alpha_1\pi_1 p + \alpha_2\pi_2 p$ are known, which means that we have only one parameter, α_1 say. We still write α_2 below to simplify the notation. In this simpler set-up the one-dimensional score process is

$$l_t'(\alpha_1) = \int_0^t \frac{I_1(u-) - (\pi_1/\pi_2)\, I_2(u-)}{\alpha_1\, I_1(u-) + \alpha_2\, I_2(u-)} [\mathrm{d}N(u) - \beta\, \bar{S}(u)\{\alpha_1\, I_1(u) + \alpha_2\, I_2(u)\}\, \mathrm{d}u].$$

The ML estimator $\hat{\alpha}_1$ based on the whole epidemic process is the solution to the equation $l_\tau'(\alpha_1) = 0$. Taylor expand $l_\tau'(\hat{\alpha}_1)$ around the true value α_1: $0 = l_\tau'(\hat{\alpha}_1) = l_\tau'(\alpha_1) + (\hat{\alpha}_1 - \alpha_1)\, l_\tau''(\alpha_1^*)$, where α_1^* lies between the estimate and the true value. Rearranging and dividing by $\sqrt{\log(n)}$ this is equivalent to

$$\frac{1}{\sqrt{\log(n)}}\, l_\tau'(\alpha_1) = \frac{-l_\tau''(\alpha_1^*)}{\log(n)}\, (\hat{\alpha}_1 - \alpha_1)\sqrt{\log(n)}, \tag{3.13}$$

and the observed information is

$$-l_\tau''(\alpha_1) = \int_0^\tau \frac{\{I_1(u-) - (\pi_1/\pi_2)\, I_2(u-)\}^2}{\alpha_1\, I_1(u-) + \alpha_2\, I_2(u-)} \frac{\mathrm{d}N(u)}{\alpha_1\, I_1(u-) + \alpha_2\, I_2(u-)}. \tag{3.14}$$

From equation (3.13) we are now ready to claim that the estimator converges at rate $\sqrt{\log(n)}$; the proof is found in Appendix A.

Theorem 2. Assume the set-up given above and that the epidemic is observed completely. Then, on the part of the sample space where there is a major epidemic, there is a consistent solution $\hat{\alpha}_1$ (the ML estimator) to the equation $l_\tau'(\alpha_1) = 0$, and $(\hat{\alpha}_1 - \alpha_1)\sqrt{\log(n)}$ is bounded in probability.

Remark 3. The general result is of course that, within each class of types with equal susceptibility, the infectivities can be estimated separately at the $\sqrt{\log(n)}$ rate of convergence.

A heuristic explanation for the observed phenomenon is that, if the type susceptibility is unique, then this type will have a *mean* proportion infective that varies over time differently

from all other types. By studying *when* susceptible individuals tend to become infected it should hence be possible to estimate the infectivities separately. But, if the susceptibility is equal to the susceptibility of some other type, then both types will have approximately equal proportions of infectives all through the epidemic, making it more difficult to estimate which type causes more infections. In this situation it is only from the stochastic fluctuations of the proportions of infective individuals around their means that the inference on the infectivities receives its information; a deterministic model would not be able to identify α_1 and α_2 separately.

It is worth mentioning that estimation of the infectivities separately is relevant even when the corresponding susceptibilities are equal. For example, individuals with high infectivity are more important to reach in vaccination programmes.

In the present paper the susceptibilities are assumed to be fixed, either different or identical, as $n \to \infty$. Of mathematical interest is the intermediate case where susceptibilities coincide as $n \to \infty$. If, for example $\beta_2 = \beta_1 + b_n$ where $b_n \to 0$, heuristic arguments indicate that the corresponding infectivity estimators converge at the rate $|b_n|\sqrt{n}$ if $|b_n|\sqrt{n} \to \infty$ and at the rate $\log(n)$ otherwise.

4. Estimation under partial observation

In applications the epidemic process is rarely observed completely. In particular the time of infection is often unknown. If any longitudinal information is available for an epidemic it is usually the show of symptoms or the time of diagnosis. It can be argued that this time is approximately the same as the mathematical term *removal times*. First, individuals are usually more infectious soon after their infection and by the time that they have shown symptoms the infectiousness has often been reduced. Second, and perhaps more important, the social activity — and hence the infectivity — is often reduced drastically on show of symptoms or diagnosis. This motivates the study below where we assume that the removal processes for each type is observed. We also treat the data consisting of the final state of the epidemic, perhaps the most commonly available data.

4.1. Deriving estimators

4.1.1. Susceptibility estimators

Equation (2.1) defines $2k$ martingales:

$$M_{i,1}(t) = N_i(t) - \int_0^t \beta_i \bar{S}_i(u)\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{I}(u)\,\mathrm{d}u, \qquad i = 1, \ldots, k, \tag{4.1}$$

$$M_{i,2}(t) = R_i(t) - \gamma \int_0^t I_i(u)\,\mathrm{d}u, \qquad i = 1, \ldots, k. \tag{4.2}$$

Unfortunately, none of these martingales are observable for the data at hand. Instead we construct k new observable martingales from equations (4.1) and (4.2). By equating the martingales to their means (equal to 0) this will give k estimating equations and thus enable estimation of k parameters, an estimation technique known as the method of moments.

Two well-known properties of martingales (e.g. Andersen *et al.* (1993)) are that integrating predictable (left continuous) processes with respect to martingales results in new martingales, and that linear combinations of martingales are also martingales. Hence

$$U_i(t;\,\boldsymbol{\alpha},\,\boldsymbol{\beta}) = \frac{\beta_i}{\gamma}\sum_j \frac{\alpha_j}{n}\,M_{j,2}(t) - \int_0^t \frac{1}{S_i(u-)}\,\mathrm{d}M_{i,1}(u), \qquad i = 1,\,\ldots,\,k, \tag{4.3}$$

defines k martingales $\{U_i(\cdot\,;\,\boldsymbol{\alpha},\,\boldsymbol{\beta})\} = \mathbf{U}(\cdot\,;\,\boldsymbol{\alpha},\,\boldsymbol{\beta})$ (as mentioned previously they are martingales only for the true parameters $(\boldsymbol{\alpha},\,\boldsymbol{\beta})$). The reason for choosing these particular martingales is that the unobservable du-terms appearing in equations (4.1) and (4.2) cancel out. In fact, noting that $\mathrm{d}N_i(u) = -\mathrm{d}S_i(u)$, we have

$$U_i(t;\,\boldsymbol{\alpha},\,\boldsymbol{\beta}) = \frac{\beta_i}{\gamma}\sum_j \frac{\alpha_j}{n}\,R_j(t) + \int_0^t \frac{\mathrm{d}S_i(u)}{S_i(u-)} = \frac{\beta_i}{\gamma}\sum_j \frac{\alpha_j}{n}\,R_j(t) - A_i(t),$$

where

$$A_i(t) = \frac{1}{S_i(0)} + \frac{1}{S_i(0)-1} + \,\ldots\, + \frac{1}{S_i(t)+1},$$

and $A_i(t) = 0$ if $S_i(t) = S_i(0)$ ($\mathbf{A}(t)$ denotes the corresponding vector). At $t = \tau$ this martingale only depends on the final state. In a fairly large population, and assuming that initially only few were infective and the rest susceptible (i.e. $S_i(0)/n_i \approx 1$), we have the well-known approximation $A_i(\tau) \approx -\log(1-\tilde{p}_i)$ where $\tilde{p}_i = 1 - S_i(\tau)/n_i = N_i(\tau)/n_i$ denotes the observed final proportion infected i-individuals. This gives us the estimating equations

$$\beta_i\gamma^{-1}\sum_j \alpha_j\pi_j\tilde{p}_j = -\log(1-\tilde{p}_i).$$

As mentioned previously the model is overparameterized and we assume that equation (3.1) holds giving us the estimator

$$\hat{\beta}_i = A_i(\tau) \approx -\log(1-\tilde{p}_i).$$

The estimator obtained when applying the approximation above is identical with the estimator derived from the corresponding deterministic model.

4.1.2. Infectivity estimators

If only the final state is observed we cannot estimate more than the k susceptibilities and linear combinations thereof. This should not be a surprise since data are only k dimensional: $(\tilde{p}_1,\,\ldots,\,\tilde{p}_k)$. In particular, it is not possible to estimate $R_0 = \gamma^{-1}\,\Sigma_i\,\alpha_i\beta_i\pi_i$ consistently since we have no information about the separate infectivities. However, if the removal processes are observed continuously (i.e. $\{R_1(u),\,\ldots,\,R_k(u);\,0 \leqslant u \leqslant \tau\}$) we can also obtain information about the infectivities, a situation we now treat.

The author has not been successful in the search for martingales that are suitable for the estimation of $\{\alpha_i\}$. Instead estimators are obtained by studying the deterministic system of differential equations (3.8) more closely. Since $\sigma_i(t) = \pi_i - \nu_i(t)$ it follows from equations (3.8) that

$$\nu_i(t) = \pi_i\left[1 - \exp\left\{-\beta_i\int_0^t \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\iota}(u)\,\mathrm{d}u\right\}\right] = \pi_i[1 - \exp\{-\beta_i\gamma^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\rho}(t)\}],$$

approximating the initial proportion infective, ϵ, to 0. Again using equations (3.8) we thus have

$$\rho_i'(t) = \gamma\,\iota_i(t) = \gamma\{\nu_i(t) - \rho_i(t)\} = \gamma(\pi_i[1 - \exp\{-\beta_i\gamma^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\rho}(t)\}] - \rho_i(t)). \tag{4.4}$$

We therefore pretend that $\{R_i(t)\}$ are counting processes with stochastic intensities

$$\eta_i(t;\, \boldsymbol{\alpha},\, \boldsymbol{\beta}) = 0 \vee n\gamma(\pi_i[1 - \exp\{-\beta_i\gamma^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\bar{\mathbf{R}}(t)\}] - \bar{R}_i(t))$$

(the deterministic functions in equation (4.4) are always non-negative but this may fail for the stochastic versions — the intensity is then set to 0). We hence have an approximate likelihood $\tilde{l}_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ which we may differentiate with respect to α_j to obtain approximate score processes as in Section 3. Divided by n, these are

$$V_j(t;\, \boldsymbol{\alpha},\, \boldsymbol{\beta}) = n^{-1}\frac{\partial\tilde{l}_t(\boldsymbol{\alpha},\, \boldsymbol{\beta})}{\partial\alpha_j} = \sum_{i=1}^{k}\int_0^t \frac{\eta_i^{(j)}(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})}{\eta_i(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})}\{\mathrm{d}\bar{R}_i(u) - \bar{\eta}_i(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})\,\mathrm{d}u\}, \tag{4.5}$$

where $\eta_i^{(j)}(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})$ is the derivative of $\eta_i(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})$ with respect to α_j and $\bar{\eta}_i(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta}) = n^{-1}\times$ $\eta_i(u;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})$. The estimating equations for $\hat{\boldsymbol{\alpha}}$ are then given by $V_j(t;\, \hat{\boldsymbol{\alpha}},\, \hat{\boldsymbol{\beta}}) = 0,\, j = 1,\, \ldots,\, k$, with $\hat{\boldsymbol{\beta}}$ defined in Section 4.1.1. The estimator $\hat{\boldsymbol{\beta}}$ only relies on the final state of the epidemic whereas $\hat{\boldsymbol{\alpha}}$ depends on the removal processes as well. Simulations in Section 5 illustrate the performance of the estimators.

4.2. Variance of estimators

Taylor expanding the estimators around the true values $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ yields

$$\begin{pmatrix} \mathbf{0} & I \\ \partial^{(1)}\mathbf{V}(\tau;\, \boldsymbol{\alpha}^*,\, \boldsymbol{\beta}^*) & \partial^{(2)}\mathbf{V}(\tau;\, \boldsymbol{\alpha}^*,\, \boldsymbol{\beta}^*) \end{pmatrix}\begin{pmatrix} \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}} \\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{U}(\tau;\, \boldsymbol{\alpha},\, \boldsymbol{\beta}) \\ \mathbf{V}(\tau;\, \boldsymbol{\alpha},\, \boldsymbol{\beta}) \end{pmatrix}, \tag{4.6}$$

where I denotes the identity matrix, $\partial^{(1)}\mathbf{V}$ and $\partial^{(2)}\mathbf{V}$ are the matrices of partial derivatives with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, and $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ are some points along the line between the true value and the estimates. For later use we define G^* as the matrix on the far left-hand side of equation (4.6) and its estimator $\hat{G}$ obtained by replacing $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ by the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$.

Since $\hat{\boldsymbol{\beta}}$ is obtained independently of $\hat{\boldsymbol{\alpha}}$ and because $\mathbf{U}$ is a martingale this suggests that the estimator is approximately Gaussian with a variance matrix which may be estimated by the observable optional variation matrix process $[\hat{\mathbf{U}}](\tau)$. From equations (4.1)–(4.3) and theory for counting processes (e.g. Andersen *et al.* (1993)) it follows that the diagonal and off-diagonal elements of the estimated variance matrix are given by

$$\widehat{[U_i,\, U_i]}(\tau) = \frac{\hat{\beta}_i^2}{\hat{\gamma}^2}\sum_k \frac{\hat{\alpha}_k^2}{n^2}\, R_k(\tau) + \frac{1}{S_i^2(0)} + \frac{1}{\{S_i(0)-1\}^2} + \ \ldots\ + \frac{1}{\{S_i(\tau)+1\}^2}$$

and

$$\widehat{[U_i,\, U_j]}(\tau) = \frac{\hat{\beta}_i\hat{\beta}_j}{\hat{\gamma}^2}\sum_k \frac{\hat{\alpha}_k^2}{n^2}\, R_k(\tau), \qquad i \neq j.$$

In the next subsection it is shown that the estimator $[\hat{\mathbf{U}}](\tau)$, which depends on the removal processes, is consistent only if the true susceptibilities are all different.

How to estimate the variance of $\hat{\boldsymbol{\alpha}}$ and the covariance matrix of $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ remains an open problem; in the next subsection we show the asymptotic order of the matrices. From the definition of $V_j(t;\, \boldsymbol{\alpha},\, \boldsymbol{\beta})$, equation (4.5), it follows that

$$V_j(t;\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^{k} \int_0^t \frac{\eta_i^{(j)}(u;\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_i(u;\boldsymbol{\alpha},\boldsymbol{\beta})} \{\mathrm{d}\bar{R}_i(u) - \gamma\,\bar{I}_i(u)\,\mathrm{d}u\}$$

$$+ \sum_{i=1}^{k} \int_0^t \frac{\eta_i^{(j)}(u;\boldsymbol{\alpha},\boldsymbol{\beta})}{\eta_i(u;\boldsymbol{\alpha},\boldsymbol{\beta})} \{\gamma\,\bar{I}_i(u) - \bar{\eta}_i(u;\boldsymbol{\alpha},\boldsymbol{\beta})\}\,\mathrm{d}u. \tag{4.7}$$

The first sum is a martingale, so the corresponding vector has an available variance estimate, but for a process like $\mathbf{V}$ it is more complicated to estimate the variance matrix. If an estimator of the full variance matrix of $\binom{\mathbf{U}(\tau)}{\mathbf{V}(\tau)}$ were available, $\hat{H}$ say, then the variance matrix of $\binom{\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}}}{\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}}$ would be estimated by $\hat{G}^{-1}\hat{H}(\hat{G}^{-1})^{\mathrm{T}}$. One way to estimate H is to neglect the second term on the right-hand side of equation (4.7) so that $\mathbf{V}$ becomes a martingale. However, this will most likely result in the variance matrix of $\mathbf{V}$ being underestimated.

4.3. Consistency and rate of convergence of estimators

Let G_{d} be the deterministic $2k \times 2k$ matrix with $k \times k$ submatrices

$$\begin{pmatrix} \mathbf{0} & I \\ D_1 & D_2 \end{pmatrix}.$$

The matrix D_1 has (j, j')-element defined by

$$-\sum_{i=1}^{k} \int_0^{\infty} f_i^{(j)}(t) f_i^{(j')}(t)/f_i(t)\,\mathrm{d}t,$$

where

$$f_i(t) = \gamma(\pi_i[1 - \exp\{-\beta_i\gamma^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\rho}(t)\}] - \rho_i(t))$$

(i.e. $f_i(t) = \rho_i'(t)$, the solution to equation (4.4)), and $f_i^{(j)}(t)$ is the derivative with respect to α_j. Let $f_i^{(*i)}(t)$ denote the derivative of $f_i(t)$ with respect to β_i. Then D_2 is the matrix with (j, i)-element

$$-\int_0^{\infty} f_i^{(j)}(t) f_i^{(*i)}(t)/f_i(t)\,\mathrm{d}t.$$

The matrix G_{d} will be the limit of the stochastic matrix G^* defined as the matrix on the far left-hand side of equation (4.6).

The proof of the following theorem is found in Appendix A.

Theorem 3. As $n \to \infty$, on the part of the sample space where there is a major epidemic,

$$G^*\sqrt{n}\begin{pmatrix} \boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}} \\ \boldsymbol{\beta}-\hat{\boldsymbol{\beta}} \end{pmatrix} \xrightarrow{\mathrm{d}} N(\mathbf{0},\,\Sigma),$$

where Σ is a non-zero matrix. Further, the matrix $\hat{G}$, and hence also G^*, converges in probability to G_{d}. If all β_i are distinct then G_{d} is invertible for almost all infectivity vectors $\boldsymbol{\alpha}$ (i.e. except a null set), and for such $\boldsymbol{\alpha}$

$$\sqrt{n}\begin{pmatrix} \boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}} \\ \boldsymbol{\beta}-\hat{\boldsymbol{\beta}} \end{pmatrix} \xrightarrow{\mathrm{d}} N(\mathbf{0},\, G_{\mathrm{d}}^{-1}\Sigma G_{\mathrm{d}}^{-1\,\mathrm{T}}),$$

and the variance matrix of $\hat{\boldsymbol{\beta}}$ may be estimated consistently with $[\hat{\mathbf{U}}](\tau)$. If some susceptibilities

are equal, then G_d is not invertible and the corresponding infectivity estimators are not consistent. Estimators of all susceptibilities and weighted sums of the infectivities with equal susceptibilities (weighted by their frequencies) still converge at rate $\sqrt{n}$.

5. Simulations

In this section we provide some numerical examples indicating the performance of the estimators. The hypothetical population consists of two types of individual, equally many of each (i.e. $\pi_1 = \pi_2 = 0.5$) and with true parameter values $\gamma = 1$, $\alpha_1 = \beta_1 = 1$ and $\alpha_2 = \beta_2 = 2$ with the interpretation that type 2 individuals are more susceptible *and* more infective when infected, perhaps because of higher social activity. Simulations were performed for populations of size $n = 500, 2000, 8000$; a factor 4 was chosen so that the standard deviation of the estimators should reduce by 50%. The simulations were initiated with one infectious individual of each type, and the remaining population susceptible. Results in the present paper rely on a major outbreak, so simulations resulting in less than 10% infected were rejected, and 100 major epidemics were simulated for each population.

As mentioned previously the model is overparameterized. Since there are only two types of individual it was decided to estimate the relative susceptibility β_2/β_1, the relative infectivity α_2/α_1, the basic reproduction number $R_0 = (\alpha_1\beta_1 + \alpha_2\beta_2)/2$ and γ^{-1}, the average length of the infectious period. As it turned out, the estimators for the relative infectivity had a much larger variation. For this reason, and the fact that α_2/α_1 is always non-negative, $\ln(\alpha_2/\alpha_1)$ was estimated instead.

In Table 1 we show the simulation averages of the estimates for different detailed data and different population sizes. Within parentheses are the standard errors for the estimates; the standard errors of the averages are 10 ($\sqrt{100}$) times smaller. As pointed out in Section 4 it is not possible to estimate the relative infectivity nor the average length of the infectious period when only the final state of the epidemic is observed. The important parameter R_0 is not estimable either. However, one more function of the parameters is estimable besides the relative susceptibility β_2/β_1: $\beta_1(0.5\alpha_1 p_1 + 0.5\alpha_2 p_2)\gamma^{-1}$, where p_1 and p_2 are defined from the parameters by equation (3.9). This means that we could estimate $\beta_i(0.5\alpha_1 p_1 + 0.5\alpha_2 p_2)\gamma^{-1}$, $i = 1, 2$, when the final state is observed, quantities interpreted as the accumulated infection force acting on i-individuals. Being less central the corresponding estimates are not given.

Table 1. Parameter estimates†

Population size n	*Data*	*Estimates for the following parameters and true values:* β_1/β_2	γ^{-1}	R_0	$\ln(\alpha_2/\alpha_1)$
		2	*1*	*2.5*	*ln(2) = 0.693*
500	Complete	2.003 (0.20)	1.004 (0.03)	2.532 (0.22)	0.827 (1.05)
	Removal	2.018 (0.24)	0.895 (0.18)	2.388 (0.32)	0.103 (1.47)
	Final state	2.018 (0.24)	—	—	—
2000	Complete	2.001 (0.09)	1.001 (0.02)	2.509 (0.13)	0.782 (0.66)
	Removal	2.002 (0.11)	0.935 (0.11)	2.423 (0.24)	0.444 (1.13)
	Final state	2.002 (0.11)	—	—	—
8000	Complete	2.000 (0.05)	1.000 (0.01)	2.506 (0.06)	0.730 (0.31)
	Removal	2.001 (0.06)	0.974 (0.07)	2.468 (0.17)	0.672 (0.91)
	Final state	2.001 (0.06)	—	—	—

†Averages from 100 simulated major epidemics; standard errors of the estimates are given in parentheses.

In Table 1 it is seen that the relative susceptibility is estimated accurately for all types of data, even in small populations. The gain in precision having the complete data is moderate. When observing the complete data or the removal processes continuously γ^{-1} and R_0 are estimated quite accurately even in small populations, only now estimation based on complete observation has better precision than removal estimates, in particular when estimating γ^{-1}. The estimators for the relative infectivity have much larger variation than the other estimators. Here also the removal process performs worse than when the complete data are available. It is seen that the standard errors are reduced by approximately a half when the population size is increased by a factor 4 as the limit theorems of Sections 3 and 4 suggest. This is not as evident for the estimates of R_0 and $\ln(\alpha_2/\alpha_1)$ based on removal data where asymptotics seem to catch in in larger populations.

Simulations for populations where the types have equal susceptibility have also been performed but are not given explicitly. As expected, the first three parameters are estimated accurately like in the examples above but the estimators of the relative infectivity α_2/α_1 have no accuracy at all, not even for complete data and a population size of many thousands.

6. Discussion

In Sections 3 and 4 it was shown that the rates of convergence of the infectivity estimators were of different order depending on whether the corresponding susceptibilities were unique or not. In applications this may not be known in advance. Still, there is no need to test the hypothesis that any susceptibilities are equal. Instead, this is reflected in that the elements of the variance matrix converge at the correct rate, independently of the underlying susceptibilities. For the complete data the variance matrix was estimated by $n^{-1}\Sigma(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1}$, defined in Section 3, and in the case of equal susceptibilities this matrix will contain elements of different order. The same is true for the variance matrix of the estimator obtained from the removal data (Section 4), only this time no estimate of the matrix was readily available.

There are several ways to generalize the multitype epidemic model of the present paper; see for example Ball and Clancy (1993). First, the assumption of exponentially distributed infectious periods is not suitable for some applications. Several results could be extended to a general parameterized distribution of the infectious period and also if a latency period was introduced in the model. Loosely speaking, the qualitative statements should remain unchanged; it is only the expressions for the variance matrices that change. An exception is if the distribution of infectiousness is a constant whence there is no difference between observing the complete data and only observing the removal processes. A theoretically interesting open problem is to characterize the class of distributions of the infectious period for which the infectivities may be estimated consistently from the removal data in the case that the susceptibilities are equal (because of the previous observation about constant infectious period the class is non-empty). The assumption of equally distributed infectious periods, i.e. equal mean γ^{-1}, for different types is also a restriction. In applications this might be easier to cope with since the infectious period is an immunological quantity which usually does not vary much between different individuals, and the alternative to bring in more parameters also has its drawbacks.

The contact matrix was assumed to be of the form $\Lambda = \boldsymbol{\beta}\boldsymbol{\alpha}^{\mathrm{T}}/n$, a restriction of the most general form being an arbitrary matrix $\Lambda = (\lambda_{ji})$. This assumption is known as *proportionate mixing* (e.g. Hethcote and Van Ark (1987) and Becker and Marschner (1990)). The close contact rate between different types is actually a factor of three ingredients: how infectious a j-type individual is, how susceptible type i-individuals are and the rate at which the two have

contact. If this contact rate splits into a product of the 'social activity' of the two types, then Λ reduces to the form $\beta\boldsymbol{\alpha}^{\mathrm{T}}/n$ as has been assumed in this paper. This assumption is not suitable when certain community structures, such as households, play an important role in the spread of disease.

Acknowledgements

I would like to thank Niels Becker for introducing me to the problem. Financial support from the Swedish Natural Science Research Council is gratefully acknowledged.

Appendix A: Proofs

In this section we outline the proofs of the theorems given in the main text. Several details are left out to reduce the length of the proofs.

A.1. Proof of theorem 1

The statement of theorem 1 follows from theorem VI.1.2 in Andersen *et al.* (1993) if we can verify the five assumptions denoted A–E. Conditions A, C and E, concerning the parameter space and regularities on the likelihood, are easily verified and will not be shown. Condition B, for the application at hand, assumes that

$$n^{-1}\Sigma(\boldsymbol{\alpha}, \beta) \xrightarrow{\mathrm{p}} \Sigma_{\mathrm{d}}$$

where $\Sigma(\boldsymbol{\alpha}, \beta)$ is given by equations (3.5)–(3.7) and Σ_{d} is the deterministic matrix defined by expressions (3.10)–(3.12). Condition D assumes that Σ_{d} is positive definite, i.e. invertible, and is only needed in the second part of the theorem.

Start with condition B. From Ball and Clancy (1993) it follows that the diagonal elements appearing in equation (3.7) divided by n converge in probability to $\pi_i p_i/\beta_i^2$, where $\{p_i\}$ are defined by equation (3.9). For the same reason, applying Slutsky's theorem, the second term in equation (3.5) divided by n converges in probability to $(\pi_j p_j \pi_{j'} p_{j'}/\gamma)p$. For the integrals appearing in equations (3.5) and (3.6) a little more effort is needed. Define the *beginning* of the epidemic to be from $t = 0$ until the first time η when there are ϵn infected individuals, i.e. the first time $N(t) = \epsilon n$, where ϵ is a small number. The *final part* of the epidemic is defined as the time from which $R(t) \geqslant n(p - \epsilon)$ until τ (for finite n this period may be empty). By majorizing the beginning and the final part of the epidemic with suitable multitype branching processes it can be shown that the contributions from these parts, divided by n, to the integrals can be made arbitrarily small by choosing ϵ sufficiently small. For the middle part of the epidemic process, whose time duration is bounded in probability, we may for example apply theorem 11.2.1 in Ethier and Kurtz (1986). This theorem shows that, if the beginning of the epidemic is overlooked and the time clock is started at η, then the 'bar processes' converge in probability, uniformly on bounded intervals, to the deterministic functions defined by equations (3.8). From this it follows that the middle parts of the integrals converge in probability. Finally, by choosing ϵ small the limit can be made arbitrarily close to the solutions of expressions (3.10) and (3.11).

We now argue that, for fixed and mutually distinct $\{\beta_i\}$, Σ_{d} is invertible (i.e. condition D) for almost all $\{\alpha_j\}$ obeying the linear constraint $\Sigma_j \alpha_j \pi_j p_j = \gamma$, which was assumed without loss of generality. It follows from equation (3.9) that $p_i = \exp(-\beta_i)$ for all such $\boldsymbol{\alpha}$. Thus, as we vary $\boldsymbol{\alpha}$ over the allowed configurations, expression (3.12) and the second term of equation (3.10) remain constant. The non-linearity of the system of differential equations (3.8) defining the deterministic solution $\{\sigma_i(t), \iota_i(t), \rho_i(t), \nu_i(t)\}$ implies that each (i, j)-term in expresssion (3.11) and (j, j')-term of equation (3.10) change differently as we vary $\boldsymbol{\alpha}$. This implies that the subspace for which Σ_{d} is not invertible has a lower dimension.

A.2. Proof of theorem 2

The statement for theorem 2 follows from equation (3.13) if we can show that $l'_\tau(\alpha_1)/\sqrt{\log(n)}$ is bounded in probability and that $l''_\tau(\alpha_1^*)/\log(n)$ is bounded away from 0 in probability.

Start with $l'_\tau(\alpha_1)/\sqrt{\log(n)}$. It follows from likelihood theory for counting processes (e.g. Andersen *et al.* (1993)) that $l'_t(\alpha_1)/\sqrt{\log(n)}$ is a zero-mean martingale indexed by t, and its variance is the expectation of equation (3.14) divided by $\log(n)$. This expectation is unchanged if $\mathrm{d}N(u)$ is replaced by its intensity

$$\{\alpha_1 I_1(u) + \alpha_2 I_2(u)\}\boldsymbol{\beta}^{\mathrm{T}}\bar{\mathbf{S}}(u)\,\mathrm{d}u.$$

Because τ is of the order $\log(n)$ and the first ratio under the integral sign has bounded expectation this implies that $l'_\tau(\alpha_1)/\sqrt{\log(n)}$ has bounded variance which in turn implies boundedness in probability.

Now we present a heuristic argument that $l''_\tau(\alpha_1^*)/\log(n)$ is bounded away from 0 in probability. The first factor under the integral sign of equation (3.14) is approximately a squared normal process (except during the beginning and the final part of the epidemic which may be omitted), and the second factor is larger than a constant times $\mathrm{d}N(u)/N(u-)$. Thus, $-l''_\tau(\alpha_1^*)/\log(n)$ is minorized by some constant with arbitrarily large probability. A simple proof of this is not known to the author. Instead we outline a rather long proof which consists of three parts:

(a) to *minorize* $-l''_\tau(\alpha_1^*)$ by a similar expression for a continuous time multitype Markov branching process;
(b) to approximate this new expression with a similar expression now containing a multitype Galton–Watson branching process;
(c) to show that the corresponding expression is larger than something sufficiently small (but positive) with arbitrarily large probability.

Along the line of strong approximations of epidemics with branching processes (e.g. Ball and Clancy (1993)) step (a) is performed by only integrating until the first time when the epidemic and branching process differ (the appearance of the first 'ghost'), a time of order $\log(n)$. In part (b) we replace the integration with respect to real time by integration of the induced Galton–Watson process with respect to generation as a discrete time parameter. This is possible because all individuals have equally distributed life lengths (equivalent to infectious periods), so individuals of the same generation live approximately around the same time. By choosing c_1 and c_2 small $-l''_\tau(\alpha_1^*)$ can thus be minorized by

$$c_1 \sum_{k=1}^{[c_2\log(n)]} \frac{\{Z_1(k) - (\pi_1/\pi_2)\,Z_2(k)\}^2}{\alpha_1^* Z_1(k) + \alpha_2^* Z_2(k)} \frac{Z_1(k+1) + Z_2(k+1)}{\alpha_1^* Z_1(k) + \alpha_2^* Z_2(k)}, \tag{A.1}$$

where $Z_i(k)$ denotes the number of i-individuals in generation k in the multitype Galton–Watson process. The term $Z_1(k) - (\pi_1/\pi_2)\,Z_2(k)$ consists of the sum of $Z_1(k-1)$ independent and identically distributed variables with mean 0 and finite variance plus the sum of $Z_2(k-1)$ other independent and identically distributed random variables, also with mean 0 and finite variance. Its square is thus stochastically increasing in $Z_1(k-1)$ and $Z_2(k-1)$. For multitype Galton–Watson processes $Z_i(k)\rho^{-k} \to W_i$ almost surely, and the random variable W_i is strictly positive on the set of non-extinction (e.g. Jagers (1975), p. 95). Let $X_{k,i}$ be independent and identically distributed random variables with mean 0 and small positive variance. Then these two facts together imply that the quantity in expression (A.1) can be minorized by

$$c_3 \sum_{k=1}^{[c_2\log(n)]} \left(\sum_{i=1}^{[c_4\rho^k]} X_{k,i} \Big/ \sqrt{\rho^k} \right)^2$$

with arbitrarily large probability, by choosing c_3 and c_4 sufficiently small. The outer sum contains $[c_2\log(n)]$ *independent* terms, and all except the first few terms can be approximated by the square of a Gaussian random variable with mean 0 and fixed positive variance. If we divide the sum by $\log(n)$ it hence converges in probability to a strictly positive constant.

A.3. Proof of theorem 3

To show the first statement of theorem 3 we note from equation (4.6) that this is equivalent to showing that

$$\sqrt{n}\begin{pmatrix}\mathbf{U}(\tau)\\ \mathbf{V}(\tau)\end{pmatrix} \xrightarrow{\mathrm{d}} N(\mathbf{0},\,\boldsymbol{\Sigma}), \qquad \text{as } n \to \infty$$

($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are omitted in the notation). Because $\mathbf{U}(t)$ is a martingale we can apply Rebolledo's theorem (e.g. Andersen *et al.* (1993), p. 83), to conclude that

$$\mathbf{U}(\tau)\sqrt{n} \xrightarrow{\mathrm{d}} N(\mathbf{0}, \Sigma_{11}).$$

For showing that $\mathbf{V}(\tau)\sqrt{n}$ converges to a Gaussian variable we use equation (4.7). The first sum is a martingale so properly normed it converges to a Gaussian vector; also from Rebolledo's theorem, the conditions of the theorem are not difficult to verify. To show that the second sum on the right-hand side of equation (4.7) multiplied by $\sqrt{n}$ converges to a Gaussian random variable we first note that the integral may be approximated by omitting integration over the beginning and final parts of the epidemic, as defined in the proof of theorem 1. This is true because the factors $\gamma \bar{I}_i(u)$ and $\bar{\eta}_i(u)$ are negligible for such us, implying that $\{\gamma \bar{I}_i(u) - \bar{\eta}_i(u)\}\sqrt{n}$ is negligible on these parts. On the central parts of the epidemic we apply theorem 11.2.1 in Ethier and Kurtz (1986) to conclude that $\eta_i^{(j)}(u)/\eta_i(u)$ converges in probability, uniformly on finite intervals, to $f_i^{(j)}(u)/f_i(u)$, and theorem 11.2.3 of Ethier and Kurtz (1986) to conclude that the process $\{\gamma \bar{I}_i(u) - \bar{\eta}_i(u)\}\sqrt{n}$ converges weakly to a Gauss–Markov process. From these observations it follows that $\mathbf{V}(\tau)\sqrt{n}$ converges to a Gaussian vector. This proves the first statement of the theorem.

We now show that

$$\hat{G} \xrightarrow{\mathrm{p}} G_{\mathrm{d}}.$$

From previous results it follows that

$$\mathbf{U}(\tau) \xrightarrow{\mathrm{p}} \mathbf{0},$$

so by equation (4.6)

$$\hat{\boldsymbol{\beta}} \xrightarrow{\mathrm{p}} \boldsymbol{\beta}.$$

The same is not always true for $\hat{\boldsymbol{\alpha}}$. However, $V_j(\tau; \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$, and its partial derivatives, only depend on $\boldsymbol{\alpha}^*$ through $(\boldsymbol{\alpha}^*)^{\mathrm{T}}\bar{\mathbf{R}}(u)$, $0 \leqslant u \leqslant \tau$. Besides the beginning of the epidemic, which may be omitted in the integrals, it can be shown that $\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\bar{\mathbf{R}}(u)$ converges in probability to $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\rho}(u)$, so the same result holds for $(\boldsymbol{\alpha}^*)^{\mathrm{T}}\bar{\mathbf{R}}(u)$. This implies that

$$\partial^{(i)}\mathbf{V}(\tau; \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \xrightarrow{\mathrm{p}} D_i, \qquad i = 1, 2,$$

as was claimed.

If the susceptibilities $\{\beta_i\}$ are all distinct it can be verified that G_{d} is invertible for almost all $\boldsymbol{\alpha}$ by using a similar argument as when showing that Σ_{d} was invertible in the proof of theorem 1. When this is the case the second central limit theorem of the theorem and the consistency statement immediately follow.

When some susceptibilities are equal the corresponding rows of D_1 and D_2 are multiples of each other, so G_{d} is not invertible. For the same reason the corresponding estimating equations, $V_j(\tau; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$, are different only because of stochastic deviations in the susceptibility estimators $\{\hat{\beta}_i\}$, so the corresponding infectivity estimators are not consistent. Still, $(\beta_i - \hat{\beta}_i)\sqrt{n}$ is asymptotically Gaussian because of equation (4.6). Finally, by multiplying G_{d} from the left by suitable matrices, it is possible to show that the sum of the infectivity estimators (weighted by the type frequencies), for which the corresponding susceptibilities are identical, converges to the corresponding true sum at rate $\sqrt{n}$.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models Based on Counting Processes*. Berlin: Springer.

Ball, F. G. (1983) The threshold behaviour of epidemic models. *J. Appl. Probab.*, **20**, 227–241.

Ball, F. and Clancy, D. (1993) The final size and severity of a generalised stochastic multitype epidemic model. *Adv. Appl. Probab.*, **25**, 721–736.

Becker, N. G. and Hasofer, A. M. (1997) Estimation in epidemics with incomplete observations. *J. R. Statist. Soc.* B, **59**, 415–429.

Becker, N. G. and Marschner, I. C. (1990) The effect of heterogeneity on the spread of disease. *Lect. Notes Biomath.*, **86**, 90–103.

Ethier, S. N. and Kurtz, T. G. (1986) *Markov Processes, Characterization and Convergence*. New York: Wiley.

Hethcote, H. W. and Van Ark, J. W. (1987) Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation and immunization programs. *Math. Biosci.*, **84**, 85–118.

Jagers, P. (1975) *Branching Processes with Biological Applications*. London: Wiley.

Lefèvre, C. (1990) Stochastic epidemic models for S-I-R infectious diseases: a brief survey of the recent theory. *Lect. Notes Biomath.*, **86**, 1–12.

Rhodes, P. H., Halloran, M. E. and Longini, Jr, I. M. (1996) Counting process models for infectious disease data: distinguishing exposure to infection from susceptibility. *J. R. Statist. Soc.* B, **58**, 751–762.

Rida, W. N. (1991) Asymptotic properties of some estimators for the infection rate in the general stochastic epidemic model. *J. R. Statist. Soc.* B, **53**, 269–283.