STEIN'S METHOD FOR EPIDEMIC PROCESSES

Gesine Reinert
King's College Research Centre
and Statistical Laboratory
UK - Cambridge CB2 1ST
G.Reinert@statslab.cam.ac.uk
http://www.statslab.cam.ac.uk/Dept/People/reinert.html

Abstract. A General Stochastic Epidemic with non-Markovian transition behaviour is considered. At time $t = 0$, the population of total size K consists of aK individuals that are infected by a certain disease (and infectious); the remaining bK individuals are susceptible with respect to that disease. An initially susceptible individual i, when infected (call A_i^K its time of infection), stays infectious for a period of length r_i, until it is removed. An initially infected individual i stays infected for a period of length $\hat{r}_i$, until it is removed. Removed individuals can no longer be affected by the disease. A bound to the distance of the empirical measure

$$\xi_K = \frac{1}{K}\sum_{i=1}^{aK}\delta_{(0,\hat{r}_i)} + \frac{1}{K}\sum_{i=1}^{bK}\delta_{(A_i^K, A_i^K + r_i)},$$

describing the average path behaviour, to its mean-field limit is established, using Stein's method. The bound, being in fact the first bound available for this mean-field approximation for epidemics, gives explicit constants depending on the time length that the epidemic is observed, and on the total population size.

Introduction

In 1949, Bartlett [13] introduced the General Stochastic Epidemic (GSE). This is a birth-death-process in a closed population where the temporal evolution of one individual depends "uniformly" on those of the others. In a closed population with K individuals, at time $t = 0$ a proportion a of the individuals is infected by a certain disease (and infective); the remaining $bK = (1-a)K$ individuals are susceptible to that disease. Infectious individuals will get removed after some time, e.g., by lifelong immunity or death, and are then no longer affected by that disease. Thus we have an SIR model.

In Bartlett [13], the general stochastic epidemic is defined as a Markov process $(X_K(t))_{t\geq 0}$, where $X_K(t) = (Y_K(t), Z_K(t))$, taking values in the set $S = \{(r,s) : r \leq (1-a)K, r+s \leq K, r,s \in I\!N\}$, with $X_K(0) = (K-a, a)$ and transition probabilities $(\alpha, \beta > 0)$

$$\begin{aligned} I\!P[X_K(t+\Delta t) = (r-1, s+1) | X_K(t) = (r,s)] &= \alpha r s \Delta t + o(\Delta t) \\ I\!P[X_K(t+\Delta t) = (r, s-1) | X_K(t) = (r,s)] &= \beta s \Delta t + o(\Delta t) \qquad (0.1) \\ I\!P[X_K(t+\Delta t) = (r, s) | X_K(t) = (r,s)] &= 1 - (\alpha r + \beta) s \Delta t + o(\Delta t) \end{aligned}$$

for $(r,s), (r, s-1), (r-1, s+1) \in S$. Here, $Y_K(t)$ is interpreted as the number of susceptibles, and $Z_K(t)$ is the number of infectives, respectively, at time t. An infective individual is assumed to be infectious during its infected period, and no multiple infections are allowed.

Often it is convenient to use the time-scale given by $\tau = \beta t$; the quantity $\rho = \frac{\beta}{\alpha}$, the ratio of removal-rate to infection-rate, is called the *relative removal-rate* and is a critical parameter for the epidemic; see Bailey [4].

In the following, similar to the construction of Bartlett's GSE by Sellke [42], we will consider a generalization that was first studied in Reinert [37]; we also apply the term "GSE" to this more general model.

From the viewpoint of an individual, the epidemic process can be described as follows. Let $(l_i, r_i)_{i \in I\!N}$ be a family of positive i. i. d. random vectors, and let $(\hat{r}_i)_{i \in I\!N}$ be a family of positive, i. i. d. random variables. Assume that the families $(l_i, r_i)_{i \in I\!N}$ and $(\hat{r}_i)_{i \in I\!N}$ are mutually independent. An initially infected individual i stays infectious for a period of length $\hat{r}_i$; then it is removed. (That the $\hat{r}_i$ need not have the same distribution as the r_i reflects the possibility that an infected individual has already been infectious for a certain period before, at time $t = 0$, it is observed.) An initially susceptible individual i, once infected, stays infectious for a period of length r_i, until it is removed. Furthermore, an initially susceptible individual i accumulates exposure to infection with a rate that depends on the evolution of the epidemic; if the total exposure reaches l_i, the individual i becomes infected. The possible dependence between l_i and r_i for each fixed i reflects the fact that both the resistance to infection and the duration of the infection may, for a fixed individual, depend on its physical constitution.

More precisely, an initially susceptible individual i gets infected as soon as a certain functional, depending on the course of the epidemic, exceeds the individual's level l_i of resistance; denote its infection time by A_i^K. If $I_K(t)$ denotes the proportion of infected individuals present in the population at time $t \in I\!R_+$, then A_i^K is given by

$$A_i^K = \inf \left\{ t \in I\!R_+ : \int_{(0,t]} \lambda(s, I_K) ds = l_i \right\},$$

for a certain function λ, acting on time as one variable and on a function path as the second variable; this function will be specified later.

Since, for epidemics, the length of the infectious period of an individual is usually very small compared to its life length, we neglect births and removals that are not caused by the disease, as well as any age-dependence of the infectivity or the susceptibility. Thus the model studied is a closed epidemic. Furthermore, the population is idealized to be homogeneously mixing.

From this construction, Bartlett's GSE can be recovered by choosing $\lambda(t,x) = x(t)$, (l_i) being i. i. d. $\exp(\alpha)$, and $(\hat{r}_i),(r_i)$ being i. i. d. $\exp(\beta)$ for some $\alpha > 0, \beta > 0$; for each i, l_i and r_i are independent. The same model is obtained by choosing $\lambda(t,x) = \alpha x(t)$, (l_i) being i. i. d. $\exp(1)$, and $(\hat{r}_i),(r_i)$ being i. i. d. $\exp(\beta)$ (in this sense, Sellke's construction is ambiguous.) From now on, for Bartlett's GSE we shall use the parametrization $\lambda(t,x) = \alpha x(t)$, (l_i) being i. i. d. $\exp(1)$, and $(\hat{r}_i),(r_i)$ being i. i. d. $\exp(\beta)$.

For applications see, e.g., Berard *et al.* [14], Swinton *et al.* [46], to name but a few; often latency periods are included. This model is rather rigid in assuming a Markovian structure. In particular, in fungi infections for instance, the accumulation function should not be modeled as directly proportional to the number of infectives present at any given time, but should rather involve the entire history of the process. Moreover it is often not reasonable to assume the duration of the infectious period to be exponentially distributed (the memoryless assumption is obviously frequently not appropriate); see, e.g., Keeling and Grenfell [26]. This motivated the above generalization.

In the literature, a generalization in the same direction has been investigated by Wang [47], [48], still assuming Markovian transition behaviour,

i. e. the law of l_1 being $\mathcal{L}(l_1) = \exp(\alpha)$ for some $\alpha > 0$. Moreover, he assumed that l_i and r_i are independent for each individual, and he chose the special type of accumulation function $\lambda(t, x) = \lambda(x(t))$. For general λ, Solomon [43] has discussed a related, age-dependent population model that deals only with one class of individuals, with an easier dependence structure. The case of variability among susceptibles, for a Markovian model, has been studied by Picard and Lefevre [35], for example.

Typically, for epidemics the vector of the proportion of susceptibles, infected and removed individuals is studied. The information in this vector is limited; though. For instance, one might also be interested in the proportion of individuals that were infected before time s and are not removed before time t, that is the infectivity at time t in the population resulting from individuals that were infected before time s. This quantity is of interest if, for example, an immunization campaign was started in the population at time s.

For large populations, stochastic epidemic models are often approximated by their mean-field limits, yielding deterministic epidemic models for the dynamic variables. This approximation being warranted only for large populations, bounds on the rate of convergence of the stochastic model towards the deterministic model are needed to make the mean-field approximation useful in practice. Yet, so far there are no bounds given in the literature. Here we will derive explicit bounds for any finite time interval. We will show that there is an explicit constant C, depending on the length of the time interval for which we observe the epidemic, such that the distance (in a metric based on smooth test functions) will never be larger than this constant times $K^{-1/2}$. This bound is valid for any finite population.

The quantity we will approximate is not only the vector of the proportion of susceptibles, infected and removed, but rather the whole average path behaviour of the epidemic process, given by the empirical measure

$$\xi_K = \frac{1}{K} \sum_{i=1}^{aK} \delta_{(0, \hat{r}_i)} + \frac{1}{K} \sum_{i=1}^{bK} \delta_{(A_i^K, A_i^K + r_i)}.$$

As not necessarily (and hopefully) not every individual in the population gets infected by the disease, the total mass of ξ_K can be less than 1. Thus ξ_K is a substochastic measure on $[0, \infty)^2$, where the half-open interval $[u, v) \subset [0, \infty)$ is represented by the point $(u, v) \in [0, \infty)^2$. (In general, δ_x shall denote the Dirac measure at the point x). We could think of a point process with values in the positive quadrant $[0, \infty) \times [0, \infty)$, each point marking infection time and removal time for one individual. For a Borel set B, the empirical measure ξ_K then returns for B the number of points of this process that are in B.

From the empirical measure the proportion of infected and the proportion of susceptibles can easily be derived. For instance, if $t \geq 0$,

$$\xi_K([0,t] \times (t,\infty)) = \frac{1}{K}\sum_{i=1}^{aK} 1_{[0,\hat{r}_i)}(t) + \frac{1}{K}\sum_{i=1}^{bK} 1_{[A_i^K, A_i^K + r_i)}(t) =: I_K(t)$$

describes the proportion of infected present at time t. Moreover, we can also investigate quantities like

$$\xi_K([0,s] \times (t,\infty]), \ t > s,$$

giving the proportion of individuals that were infected before time s and are not removed before time t. Thus we gain new insights concerning the behavior of the epidemic.

As a tool for deriving this result, we will employ Stein's method. In the context of empirical measures it has been derived in Reinert [36]. In Reinert [37] it has been used to prove convergence of the average path behaviour to its mean field limit, but no bound on the rate of convergence has been given. This void will be filled in the present paper.

The novelties of the approach described here are thus the generality of the considered model (not assuming any Markovian structure), the generality of the quantities that can be approximated, and, perhaps most importantly, the first known bound on the quality of mean-field approximations in epidemics.

For didactical purposes, the paper is organized as follows. In Section 1 we give a brief review of Stein's method. For measure-valued random elements, some technical background is described, and the main results of Stein's method for mean-field limits are recalled. In Section 2 we apply Stein's method to the GSE, yielding an explicit bound on the distance of the GSE to its mean-field limit. In Section 3 we discuss the above results and we point out directions of future research.

CHAPTER 1

A brief introduction to Stein's method

Stein's method has first been introduced by Stein [44] for proving normal approximations. This method avoids characteristic functions, but instead relies on a characterizing equation for the normal distribution. The distribution of any random variable would then be approximately normal if it satisfies the characterizing equation approximately. Indeed, the deviation from satisfying the characterizing equation exactly turns out to provide a measure on the distance to the normal distribution. (Lately, Langevin equations have been used in a similar spirit.)

1.1 Stein's method for normal approximations

As this chapter is designed to give a brief introduction to Stein's method, we will first illustrate this method by its original example - normal approximations. First described in Stein [44], in more detail it can be found in Stein [45]; improved bounds are in Baldi *et al.* [5]. It can be sketched as follows.

1. A random variable Z has standard normal distribution, that is, $\mathcal{L}(Z) = \mathcal{N}(0,1)$, if and only if for all smooth f,
$$\mathbb{E}\{f'(Z) - Zf(Z)\} = 0.$$
2. Let $\mathcal{L}(Z) = \mathcal{N}(0,1)$. For any smooth h there is a function $f = f_h$ solving
$$h(x) - \mathbb{E}h(Z) = f'(x) - xf(x) \quad (\textit{Stein equation})$$
such that
$$\begin{aligned} \|f\| &\leq \sqrt{\frac{\pi}{2}}\|h - \mathbb{E}h(Z)\| \\ \|f'\| &\leq (\sup h - \inf h) \\ \|f^{(2)}\| &\leq 2\|h'\|. \end{aligned}$$
Here, $\| f \| = \sup_{x\in\mathbb{R}} |f(x)|$ denotes the supremum norm.
3. So for any random variable W, any smooth h, substituting W for x in the Stein equation and taking expectations on both sides gives
$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}f'(W) - \mathbb{E}Wf(W). \tag{1.1}$$

Now if $\mathcal{H}$ is a convergence-determining class for weak convergence such that, for all $h \in \mathcal{H}$ we have that $f_h \in \mathcal{H}$, then taking absolute values in Equation (1.1) gives

$$\sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| = \sup_{f\in\mathcal{H}} |\mathbb{E}f'(W) - \mathbb{E}Wf(W)|,$$

and if the r.h.s., for some quantity $W = W_n$, tends to zero as $n \to \infty$, then we obtain a central limit theorem. In particular, we can choose $\mathcal{H}$ as the space of all continuous, bounded functions with piecewise continuous, bounded first derivatives. Similar results for nonsmooth test functions have been obtained by Bolthausen [15], by Götze [23], and by Rinott and Rotar [40].

To see why this approach might be useful at all, consider a very classical example. Let $X, X_1, \ldots, X_n$ i.i.d. random variables with $\mathbb{E}X = 0$ and $\mathrm{Var}X = \frac{1}{n}$. Then

$$W = \sum_{i=1}^n X_i$$

has mean zero and variance 1. Put

$$W_i = W - X_i = \sum_{j\neq i} X_j.$$

To evaluate the right-hand side of Equation (1.1) we first calculate

$$\begin{aligned}
\mathbb{E}Wf(W) &= \sum_{i=1}^n \mathbb{E}X_i f(W) \\
&= \sum_{i=1}^n \mathbb{E}X_i f(W_i) + \sum_{i=1}^n \mathbb{E}X_i^2 f'(W_i) + R \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}f'(W_i) + R,
\end{aligned}$$

where we used Taylor expansion, and for the Taylor remainder term R we have

$$|R| \leq \|f''\| \sum_{i=1}^n \mathbb{E}|X_i|^3.$$

So

$$\mathbb{E}f'(W) - \mathbb{E}Wf(W) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{f'(W) - f'(W_i)\} + R.$$

Applying Taylor expansion again we obtain the following result (see, e.g. Stein [45]).

Theorem 1.1.1 *Let $\mathcal{L}(Z) = \mathcal{N}(0,1)$. For any smooth h*

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \quad \leq \quad \|h'\| \left(\frac{2}{\sqrt{n}} + \sum_{i=1}^{n} \mathbb{E}|X_i^3| \right).$$

Note that the bound on the r.h.s. does not involve any asymptotic statement; rather, it is valid for any number n. Thus, even if convergence might not hold, a bound on the distance is still be obtainable.

Stein's method has been proven particularly useful for proving results for sums of dependent observations. Consider, for example, the case that $X, X_1, \ldots, X_n$ are random variables with $\mathbb{E}X = 0$ and $\mathrm{Var}X = \frac{1}{n}$, such that for each X_i there is a set S_i such that X_i is independent of $\sigma(X_j, j \notin S_i)$. (A special case would be m-dependent random variables.) For simplicity assume that $|S_i| = \gamma$ for some γ, and that $X_i \leq \frac{C}{\sqrt{n}}$ for some constant C that does not depend on i. Again put

$$W = \sum_{i=1}^{n} X_i.$$

If γ is small, then the summands are approximately independent, so that a normal approximation should hold. Indeed, we will prove the following result.

Theorem 1.1.2 *Let $X, X_1, \ldots, X_n$ be random variables with $\mathbb{E}X = 0$ and $VarX = \frac{1}{n}$, such that for each X_i there is a set S_i such that X_i is independent of $\sigma(X_j, j \notin S_i)$. Assume that $|S_i| = \gamma$ for some γ. Assume that $X_i \leq \frac{C}{\sqrt{n}}$ for some constant C that does not depend on i. Let $\mathcal{L}(Z) = \mathcal{N}(0,1)$. For all continuous, bounded functions h with piecewise continuous, bounded first derivatives,*

$$\begin{aligned} & |\mathbb{E}h(W) - \mathbb{E}h(Z)| \\ \leq \quad & \|h'\| \frac{10\gamma^2 C^3}{\sqrt{n}} + (\sup h - \inf h) \sum_{i=1}^{n} \sum_{j \in S_i, j \neq i} \mathbb{E}|X_i X_j|. \end{aligned}$$

Note that Theorem 1.1.2 gives an explicit bound in terms of neighborhood size and number of observations, as well as correlations. If the covariance term is large, one would rather approximate the sum with a normal distribution having as variance $\mathrm{Var}(\sum_{i=1}^{n} X_i)$; then the second error term in Theorem 1.1.2 vanishes.

Proof of Theorem 1.1.2.

Put

$$W_i = \sum_{j \notin S_i} X_j$$

and

$$W_{i,j} = \sum_{k \notin S_i \cup S_j} X_k.$$

Then W_i is independent of X_i, and $W_{i,j}$ is independent of X_i and X_j. Similarly as above, we expand the right-hand side of Equation (1.1). We have

$$\begin{aligned}
\mathbb{E} W f(W) &= \sum_{i=1}^n \mathbb{E} X_i f(W) \\
&= \sum_{i=1}^n \mathbb{E} X_i f(W_i) + \sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j f'(W_i) + R_1 \\
&= \sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j f'(W_i) + R,
\end{aligned}$$

where

$$\begin{aligned}
|R_1| &\leq \|f''\| \sum_{i=1}^n \sum_{j \in S_i} \sum_{k \in S_i} \mathbb{E} |X_i X_j X_k| \\
&\leq \|f''\| \frac{\gamma^2 C^3}{\sqrt{n}}.
\end{aligned}$$

Moreover,

$$\sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j f'(W_i) = \sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j f'(W_{i,j}) + R_2,$$

where

$$\begin{aligned}
|R_2| &\leq \|f''\| \sum_{i=1}^n \sum_{j \in S_i} \sum_{k \in S_i \cup S_j} \mathbb{E} |X_i X_j| \mathbb{E} |X_k| \\
&\leq \|f''\| \frac{2\gamma^2 C^3}{\sqrt{n}}.
\end{aligned}$$

Using Taylor expansion again, we obtain

$$\begin{aligned}
\sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j f'(W_{i,j}) &= \sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j \mathbb{E} f'(W_{i,j}) \\
&= \sum_{i=1}^n \sum_{j \in S_i} \mathbb{E} X_i X_j \mathbb{E} f'(W) + R_3,
\end{aligned}$$

with

$$|R_3| \leq \|f''\| \frac{2\gamma^2 C^3}{\sqrt{n}};$$

thus,

$$I\!\!EWf(W) = Ef'(W) + \sum_{i=1}^{n} \sum_{j \in S_i, j \neq i} I\!\!EX_i X_j I\!\!Ef'(W) + R_1 + R_2 + R_3.$$

This yields the result. $\square$

A more detailed survey on Stein's method for normal approximations is in Reinert [38].

1.2 Stein's method in general

Stein's method in general can briefly be described as follows. Find a good characterization (that is, an operator $\mathcal{A}$) of the target distribution μ that is of the type

$$\mathcal{L}(X) = \mu \iff I\!\!E\mathcal{A}f(X) = 0, \text{ for all smooth functions } f,$$

where X is a random element, $\mathcal{L}(X)$ denotes the law of X. and $\mathcal{A}$ is an operator associated with the distribution μ. Then assume X to have distribution μ, and consider the *Stein equation*

$$h(x) - I\!\!Eh(X) = \mathcal{A}f(x), \quad x \in I\!\!R. \tag{1.2}$$

For every smooth h, find a corresponding solution f of this equation. Then, for any random element W, we obtain

$$I\!\!Eh(W) - I\!\!Eh(X) = I\!\!E\mathcal{A}f(W).$$

Hence, to estimate the proximity of W and X, it is sufficient to estimate $I\!\!E\mathcal{A}f(W)$ for all possible solutions f of (1.2). Moreover, bounding $I\!\!E\mathcal{A}f(W)$ for all smooth f automatically yields a bound on the distance of $\mathcal{L}(W)$ to μ in a smooth metric, regardless of asymptotics.

For μ being the standard normal distribution, the corresponding operator is $\mathcal{A}f(x) = f'(x) - xf(x)$. Of course the operator could also be defined as a second-order operator, namely $\mathcal{A}f(x) = f''(x) - xf'(x)$, the Ornstein-Uhlenbeck generator.

Barbour [8] suggests employing as operator $\mathcal{A}$ in equation (1.2) the generator of a Markov process, as this provides a way to look for solutions of (1.2). This is what in the following will be called the generator method. Suppose we can find a Markov process $(X(t))_{t \geq 0}$ with generator $\mathcal{A}$ and unique stationary distribution μ, such that $\mathcal{L}(X(t)) \stackrel{w}{\Rightarrow} \mu \ (t \to \infty)$; here, $\stackrel{w}{\Rightarrow}$ denotes weak convergence. Then, if a random variable X has distribution μ,

$$I\!\!E\mathcal{A}f(X) = 0$$

for all $f \in \mathcal{D}(\mathcal{A})$. Now a method for solving equation (1.2) is provided by Proposition 1.5 of Ethier and Kurtz ([20], p. 9; for the argument, see

Barbour [8]). Let $(T_t)_{t\geq 0}$ be the transition semigroup of the Markov process $(X(t))_{t\geq 0}$. Then

$$T_t h - h = \mathcal{A}\left(\int_0^t T_u h\, du\right).$$

As $(T_t)_{t\geq 0}$ is a strongly continuous contraction semigroup, $\mathcal{A}$ is closed (Ethier and Kurtz [20], Corollary 1.6), and we could formally take limits:

$$h(x) - \mathbb{E}h(X) = -\mathcal{A}\left(\int_0^\infty T_u h\, du\right).$$

Thus $f = -\int_0^\infty T_u h\, du$ would be a solution of (1.2), if this expression exists and if $f \in \mathcal{D}(\mathcal{A})$. This will in general be the case only for a certain class of functions h. However, the latter conditions can usually be checked.

Stein's method has been generalized to many other distributions, foremost the Poisson distribution (see Chen [16], Arratia *et al.* [3], Barbour *et al.* [10], Aldous [1], to cite but a few). Other distributions include the uniform distribution (Diaconis [18]), the binomial distribution (Ehm [22]), the compound Poisson distribution (Barbour *et al.* [9], Barbour and Utev [12], Roos [41]), the multinomial distribution (Loh [27]), the gamma distribution (Luk [29]; for the χ^2 distribution see also Mann [30]), the geometric distribution (Peköz [34]) and, more generally, Pearson curves (Diaconis and Zabell [19], Loh [28]).

The most obvious advantage of Stein's method is that it yields immediate bounds. Moreover in many situations where dependence comes into play the application is straightforward; many examples are of combinatorial nature. An early success of Stein's method is the work by Bolthausen [15] for a combinatorial central limit theorem; he was the first to obtain the correct order for this approximation. In examples from random graph theory, where the method of moments used to be the most popular technique, Stein's method allowed not only to provide rates of convergence for the first time, but also to weaken conditions; see, for instance, Barbour *et al.* [11]. Another advantage of Stein's method is that it can also be used to derive lower bounds for the approximations; Hall and Barbour [24] applied it to give lower bounds for the rate of convergence in the central limit theorem for independent random variables.

Unfortunately such a straightforward application of Stein's method may not yield the correct order for the rate of convergence; for an example, see, e.g. Reinert [38].

1.3 Stein's method for the weak law of large numbers

Here we will be interested in applying Stein's method to point mass in measure space, as developed in Reinert [36], resulting in weak laws of large numbers. Point mass can be seen as an extreme case of the normal distribution with zero variance. Hence we put

$$\mathcal{A}f(x) = -xf'(x), \quad x \in \mathbb{R}.$$

Note that $\mathcal{A}$ is the generator of the deterministic Markov process $(Y(t))_{t\geq 0}$ given by

$$\mathbb{P}[Y(t) = xe^{-t} \mid Y(0) = x] = 1, \quad x \in \mathbb{R}.$$

The corresponding transition semigroup is

$$T_t h(x) = h(xe^{-t}),$$

and the unique stationary distribution is δ_0.

According to the general equation (1.2), the Stein equation in this context is

$$h(x) - h(0) = -xf'(x), \quad x \in \mathbb{R}. \tag{1.3}$$

Let $C_b^2(\mathbb{R})$ be the space of all bounded, twice continuously differentiable real-valued functions on $\mathbb{R}$ with bounded first and second derivatives, and let $D_b^2(\mathbb{R})$ be the space of all twice continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with bounded first and second derivatives. Using the semigroup approach the following proposition is easy to derive.

Proposition 1.3.1 *For any $h \in C_b^2(\mathbb{R})$, there is a function $f = \phi(h) \in D_b^2(\mathbb{R})$ that solves the Stein equation (1.3) for h. Furthermore, for the derivatives, $\|f'\| \leq \|h'\|$, and $\|f''\| \leq \|h''\|$.*

Now we have all the ingredients to derive weak laws of large numbers.

Theorem 1.3.2 *Let $(X_i)_{i\in\mathbb{N}}$ be a family of random variables on $\mathbb{R}$, defined on the same probability space, with finite variances. Put*

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i).$$

Then, for all $h \in C_b^2(\mathbb{R})$

$$|\mathbb{E}h(Y_n) - h(0)| \leq \|h''\| \, Var\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big).$$

The proof of Theorem 1.3.2 follows easily using Taylor expansion (indeed, it is not even necessary to employ Stein's method).

As $C_b^2(\mathbb{R})$ is convergence-determining for weak convergence of the laws of real-valued random variables, a weak law of large numbers follows from Theorem 1.3.2 provided that

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) \to 0 \ \ (n \to \infty).$$

Moreover we obtain an explicit bound on the distance to point mass.

1.4 Stein's method for the weak law in measure space

As the quantity ξ_K we want to approximate is a random element taking values in the space of substochastic measures, some technical details on convergence on this space are needed (see, e.g., Reinert [36], [39]). Let E be a locally compact Hausdorff space with a countable basis (for instance, $E = I\!R^2$), let $\mathcal{E} = \mathcal{B}(E)$ be the Borel-σ-field of E, and let $M^{\mathrm{b}}(E)$ the space of all bounded Radon measures on E, equipped with the vague topology. Let $C_c(E)$ be the space of real-valued continuous functions on E with support contained in a compact set. Abbreviate the integral

$$\langle \mu, \phi \rangle = \int_E \phi d\mu,$$

Convergence in the vague topology is defined as

$$\nu_n \overset{v}{\Rightarrow} \nu \iff \text{for all } f \in C_c(E) : \langle \nu_n, f \rangle \to \langle \nu, f \rangle \quad (n \to \infty).$$

For probability measures this differs from weak convergence in that the class of test function is not $C_b(E)$, the class of bounded continuous functions on E. Thus, for example, if $E = I\!R$ and $\nu_n = \delta_n$, then $\nu_n \overset{v}{\Rightarrow} 0$, but $(\nu_n)_n$ does not converge weakly. Here, 0 denotes the measure that assigns measure 0 to any set. In what follows, weak convergence is denoted by $\overset{w}{\Rightarrow}$.

For $\mu \in M^{\mathrm{b}}(E)$, set $\|\mu\| = \sup_{A \in \mathcal{E}} |\mu(A)|$. Let

$$M_1(E) = \{\mu \in M^{\mathrm{b}}(E) : \mu \text{ positive}, \|\mu\| \leq 1\}$$

be the space of all positive Radon measures with total mass smaller or equal to 1 (the space of substochastic measures). As E has a countable basis, $M_1(E)$ is Polish with respect to the vague topology. Moreover $M_1(E)$ is vaguely compact with a countable basis. Thus the considerations about vague convergence are valid for both $E = I\!R^2$ and $E = M_1(I\!R^2)$.

The next ingredient needed is a convergence-determining class of functions. Put

$$\begin{aligned} \mathcal{F} \quad := \quad & \{F \in C_b(M_1(I\!R^2)) : F \text{ has the form} \\ & F(\mu) = f(\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle) \\ & \text{for an } m \in I\!N, f \in C_b^\infty(I\!R^m) \\ & \text{with } \|f'\| \leq 1, \|f''\| \leq 1, \|f'''\| \leq 1, \text{ and for} \\ & \phi_i \in C_b^\infty(I\!R^2) \text{ with } \|\phi_i\| \leq 1, \|\phi_i'\| \leq 1, i = 1, \ldots, m\}. \end{aligned} \tag{1.4}$$

Here the superscript $'$ indicates the total derivative, and the norm $\| \cdot \|$ indicates the sum of the supremum norms of the components, so that, for $f \in C_b^\infty(I\!R^m)$, the space of infinitely often differentiable continuous functions from $I\!R^m$ to $I\!R$ with bounded derivatives, $\| f' \| = \sum_{i=1}^m \| f_{(j)} \|$, where $f_{(j)}$ is the partial derivative of f in direction x_j.

This construction is similar to the algebra of polynomials used by Dawson

[17]. In Reinert [36] it is shown that this class of functions is convergence-determining for vague convergence. Thus it is a suitable class for Stein's method. One could even introduce a Zolotarev-type metric using $\mathcal{F}$; see Reinert [39] for details.

Now we have all the ingredients needed to set up Stein's method on the space of substochastic measures. Following Reinert [36], the corresponding generator for the weak law of large numbers for random measures with target measure δ_μ, for some $\mu \in M_1(I\!R^2)$, is, for $F \in \mathcal{F}$ of the form (1.4)

$$\mathcal{A}F(\nu) = \sum_{i=1}^{m} f_{(i)}(\langle \nu, \phi_j \rangle, j = 1, \dots, m)\langle \nu - \mu, \phi_i \rangle. \tag{1.5}$$

(This can be seen as a Gâteaux differential operator.) Moreover, the Stein equation has smooth solutions; we have

Proposition 1.4.1 *For any function $H \in \mathcal{F}$, there is a function $\psi(H) \in \mathcal{F}$ that solves the Stein equation with the operator (1.5) for H. If*

$$H(\nu) = h(\langle \nu, \phi_1 \rangle, \dots, \langle \nu, \phi_m \rangle),$$

then

$$\psi(H)(\nu) = f(\langle \nu, \phi_1 \rangle, \dots, \langle \nu, \phi_m \rangle)$$

for a function $f \in C_b^\infty(I\!R^m)$ with $\|f^{(k)}\| \le \|h^{(k)}\|$ for all $k \in I\!N$.

Stein's method then immediately yields the following theorem.

Theorem 1.4.2 *Let ξ_n be a random element with values in $M_1(I\!R^2)$, let $\mu \in M_1(I\!R^2)$, and let $\mathcal{A}$ be as in (1.5). Then*

$$\sup\{|EH(\xi_n) - H(\mu)| : H \in \mathcal{F}\} \quad \le \quad \sup\{|I\!E\mathcal{A}F(\xi_n)| : F \in \mathcal{F}\}.$$

Many examples on how to apply this theorem are given in Reinert [36]. Here we will apply it to the GSE as constructed above.

CHAPTER 2

A bound on the distance of the GSE to its mean-field limit

2.1 Assumptions

As described in the introduction, let $(l_i, r_i)_{i \in \mathbb{N}}$ be a family of positive i. i. d. random vectors, let Ψ be the common distribution function of the $(l_i)_{i \in \mathbb{N}}$, let Φ be the common distribution function of the $(r_i)_{i \in \mathbb{N}}$, and let $(\hat{r}_i)_{i \in \mathbb{N}}$ be a family of positive, i. i. d. random variables with distribution function $\hat{\Phi}$ and law $\hat{\mu}$. Assume the $(l_i, r_i)_{i \in \mathbb{N}}$, $(\hat{r}_i)_{i \in \mathbb{N}}$ to be mutually independent (whereas, for each fixed i, l_i and r_i may be dependent). (In Reinert [37], the $\hat{r}_i$ were allowed to have different distributions; however, although it would not be difficult to incorporate this inhomogeneity, for the sake of presentation it has been omitted.) Let $D_+ = \{x : [0, \infty) \to [-1, 1]$ right continuous with left-hand limits$\}$, and let $\lambda : \mathbb{R}_+ \times D_+ \to \mathbb{R}_+$ be the "accumulation" function. We use the notation $1_C(t)$ to denote the indicator function on the set C; the notation $I[t \in C]$ refers to the indicator of a set, not considered as a function. Then, for an initially susceptible individual i, its infection time A_i^K is given by

$$A_i^K = \inf \Big\{ t \in \mathbb{R}_+ : \int_{(0,t]} \lambda(s, I_K) ds = l_i \Big\}, \tag{2.1}$$

with

$$I_K(t) = \frac{1}{K} \sum_{j=1}^{aK} 1_{[0, \hat{r}_j)}(t) + \frac{1}{K} \sum_{j=1}^{bK} 1_{[A_j^K, A_j^K + r_j)}(t)$$

being the proportion of infected individuals present in the population at time $t \in \mathbb{R}_+$. This gives a recursive definition of the A_i^K's: If $l_{(j)}$ is the jth order statistics of $l_1, ..., l_{bK}$, corresponding to the individual i_j, say, then

$$\begin{aligned} A_{i_j}^K &= \inf \Bigg\{ t > 0 : \quad \int_{(0,t]} \lambda \Bigg(s, \frac{1}{K} \sum_{m=1}^{aK} 1_{[0, \hat{r}_m)} + \frac{1}{K} \sum_{k=1}^{j-1} 1_{[A_{i_k}^K, A_{i_k}^K + r_{i_k})} \Bigg) ds \\ &\qquad = l_{(j)} \Bigg\}. \end{aligned}$$

This completes the description of the model. Furthermore, we make some technical assumptions.

- The function $\lambda : I\!\!R_+ \times D_+ \to I\!\!R_+$ satisfies for all $t \in I\!\!R_+$, $x, y \in D_+$
  1. it is non-anticipating: $\lambda(t, x) = \lambda(t, x_t)$, where, for $t, u \in I\!\!R_+$, $x \in D_+$, $x_t(u) = x(t \wedge u)$;
  2. it is Lipschitz: there is a constant $\alpha > 0$ such that, for all t,
  $$|\lambda(t, x) - \lambda(t, y)| \leq \alpha \sup_{0 \leq s \leq t} |x(s) - y(s)|; \tag{2.2}$$
  3. it is bounded: there is a constant $\gamma > 0$ such that, for all t,
  $$\sup_{0 \leq s \leq t} \lambda(s, x) \leq \gamma;$$
  4. we have that, for all t,
  $$\lambda(t, x) = 0 \iff x(t) = 0.$$
- There is a constant $\beta > 0$ such that, for each $x \in I\!\!R_+$, the conditional distribution function $\Psi_x(t) := I\!\!P[l_1 \leq t | r_1 = x]$ has a density $\psi_x(t)$ that is uniformly bounded from above by β;
  $$\psi_x(t) \leq \beta \text{ for all } x \in I\!\!R_+, t \in I\!\!R_+. \tag{2.3}$$
  We assume that Ψ has a density ψ. (In Reinert [37] it was only assumed that the conditional distribution function is Lipschitz continuous; however, imposing the existence of a density is more convenient.)
- We assume that $\hat{\Phi}(0) = 0$ and $\Phi(0) = 0$, so that infected individuals do not immediately get removed.

For Bartlett's GSE defined in Equations (0.1), the above assumptions are satisfied. Choosing $\lambda(t, x) = \alpha x(t)$, (l_i) being i. i. d. exp(1), and $(\hat{r}_i), (r_i)$ being i. i. d. exp(ρ), where for each i, l_i and r_i are independent, we obtain

$$|\lambda(t, x) - \lambda(t, y)| \leq \alpha \sup_{0 \leq s \leq t} |x(s) - y(s)|.$$

Moreover λ is bounded by $\gamma = 1$, and $I\!\!P[l_1 \leq t | r_1 = x] = 1 - e^{-t}$ has density that is uniformly bounded by 1.

Note that we think of x as resembling the proportion of infected individuals. Thus the Condition 4. on λ guarantees that, once there are no infectives left, the epidemic stops. In Reinert [37] this assumption was not made, which resulted in lengthy calculations.

2.2 Heuristics

To understand the argument for the limiting distribution, here is a heuristic explanation. Put

$$F_K(t) = \int_0^t \lambda(s, I_K) ds, \tag{2.4}$$

then

$$A_i^K = F_K^{-1}(l_i).$$

Moreover, the proportion $I_K(t)$ of infectives present at time t depends itself on F_K; we have

$$I_K(t) = \frac{1}{K}\sum_{i=1}^{aK} \mathbf{1}(\hat{r}_i > t) + \frac{1}{K}\sum_{j=1}^{bK} \mathbf{1}(F_K^{-1}(l_j) \le t < F_K^{-1}(l_j) + r_j).$$

For $f \in D(I\!R_+)$, the space of right-continuous functions from $I\!R_+$ to $I\!R$ with left-hand limits, and for $t \in I\!R_+$, we define operators

$$\mathcal{Z}_K f(t) = \frac{1}{K}\sum_{i=1}^{aK} \mathbf{1}(\hat{r}_i > t) + \frac{1}{K}\sum_{j=1}^{bK} \mathbf{1}(f(t - r_j) < l_j \le f(t)) \quad (2.5)$$

$$L_K f(t) = \int_{(0,t]} \lambda(s, \mathcal{Z}_K f) ds. \quad (2.6)$$

Then $F_K = L_K F_K$; thus F_K can be described as a fixed point of an operator. Moreover, for the operator $\mathcal{Z}_K$ the weak law of large numbers suggests as limiting operator, for $f \in C(I\!R_+), t \in I\!R_+$,

$$\mathcal{Z}f(t) = a(1 - \hat{\Phi}(t)) + bI\!P(f(t - r_1) < l_1 \le f(t)). \quad (2.7)$$

Accordingly we define

$$Lf(t) = \int_{(0,t]} \lambda(s, \mathcal{Z}f) ds. \quad (2.8)$$

Using the Contraction Theorem it can be shown (see Reinert [37]) that on each finite time interval the equation $Lf = f$ has a unique solution G. It thus heuristically follows that

$$F_K \approx G.$$

As G is deterministic, the desired limiting distribution of the process can now easily be derived, and is given in the next section.

2.3 Previous results

Under the above assumptions, in Reinert [37] the following results were obtained. For the required measure μ, observe that during the course of the epidemic not necessarily (hopefully) every susceptible will get infected; $A_i^K = \infty$ for some i is possible. Therefore, if such a μ exists, it will in general not be a probability measure but a positive measure with total mass ≤ 1. Furthermore, as the existence of $I\!Er_i$ or $I\!E\hat{r}_i$, $i \in I\!N$, is not assumed, we restrict the observations to finite intervals $[0,T] \times [0,T]$ for a $T \in I\!R_+$ arbitrary, fixed. This leads to some notation. For $T \in I\!R_+$, put

$[0,T]^2 = [0,T] \times [0,T]$, and $\mathcal{B}_T = \mathcal{B}([0,T]^2)$. Let $\nu \in M_1(\mathbb{R}_+^2)$, then

$$\nu^T \quad = \quad \nu|_{\mathcal{B}_T}$$

is the restriction of ν on $\mathcal{B}_T$ (hence, $\nu^T \in M_1([0,T]^2)$). For $A \in \mathcal{B}(\mathbb{R}^2)$, put

$$\nu^T(A) = \nu(A \cap [0,T]^2);$$

this defines ν^T also on $\mathcal{B}(\mathbb{R}^2)$. If in addition X is a random element with $\mathcal{L}(X) = \nu$, then, for all $T \in \mathbb{R}_+, f \in L_1(\nu), A \in \mathcal{B}(\mathbb{R}^2)$, the corresponding restrictions are

$$\begin{aligned} \mathbb{E}^T f(X) &= \int f(x)\nu^T(\,dx) \\ \mathbb{P}^T[X \in A] &= \int 1_A(x)\nu^T(\,dx) \\ \mathcal{L}^T f(X) &= \mathcal{L}(f(X))|_{\mathcal{B}_T}. \end{aligned}$$

Moreover, for any $T > 0$ we use the notation

$$\| f \|_T \quad = \quad \sup_{t \in [0,T]} |f(t)|.$$

Theorem 2.3.1 *For $T \in \mathbb{R}_+$, the equation*

$$f(t) = \int_{(0,t]} \lambda(s, \mathcal{Z}f)ds, \quad 0 \le t \le T, \tag{2.9}$$

has a unique solution G_T. This solution can be obtained by an iteration procedure: Choose an arbitrary $f_0 \in C([0,T])$, put $f_1 = Lf_0$, $f_n = Lf_{n-1}$ for $n \in \mathbb{N}$. Then,

$$\begin{aligned} &\| f_n(t) - G_T(t) \|_T \\ &\le \frac{(\frac{b}{2})^n}{1 - \frac{b}{2}} \frac{(1 + 4\alpha\beta T(n+1))^{\frac{T}{\eta}+2} - 1}{4\alpha\beta T(n+1)} \| f_0 - Lf_0 \|_T, \end{aligned}$$

where

$$\eta = \sup\left\{ t \le T : \int_{(0,T]} (1 + \Phi(s))\, ds \le \frac{1}{2\alpha\beta} \right\}.$$

Theorem 2.3.2 *For $T \in \mathbb{R}_+$, let G_T be the solution of (2.9) and $\tilde{\mu}^T \in M_1(\mathbb{R}_+^2)$ be given for $r, s \in (0,T]$ by*

$$\begin{aligned} \tilde{\mu}^T([0,r] \times [0,s]) \quad = \quad & \int_{(0,(s-r)\vee 0]} \Psi_x(G_T(r)) d\Phi(x) \\ & + \int_{((s-r)\vee 0, s]} \Psi_x(G_T(s-x)) d\Phi(x). \end{aligned}$$

Put

$$\mu^T = a(\delta_0 \times \hat{\mu})^T + b\tilde{\mu}^T.$$

Then

$$\frac{1}{K}\sum_{i=1}^{aK} \mathcal{L}^T((0,\hat{r}_i)) + \frac{1}{K}\sum_{i=1}^{bK} \mathcal{L}^T((A_i^K, A_i^K + r_i)) \stackrel{v}{\Rightarrow} \mu^T \quad (K \to \infty).$$

Note that it might be more intuitive to think of $\tilde{\mu}^T([0,r] \times [0,s])$ as

$$\tilde{\mu}^T([0,r] \times [0,s]) = I\!P^T[l_1 \leq G_T(r), l_1 \leq G_T(s - r_1)].$$

In practice, Theorem 2.3.1 thus provides a numerical iteration procedure for finding the limiting function G_T; from Theorem 2.3.2 it follows that this function suffices to find the deterministic approximation.

Theorem 2.3.3 *Let μ^T be as in Theorem 2.3.2. Then, for all $T \in I\!R_+$,*

$$\mathcal{L}(\xi_K^T) \stackrel{w}{\Rightarrow} \delta_{\mu^T} \quad (K \to \infty).$$

However, in Reinert [37], no bound on the rate of convergence was given. As the proof employed the Glivenko-Cantelli theorem and thus almost-sure convergence results, below we will give a different proof of Theorem 2.3.3. First note, though, that Theorem 2.3.3 gives an approximation of the GSE by its mean-field limit. To make this more obvious, in the next subsection we consider the proportion of infected individuals and the proportion of susceptible individuals.

2.3.1 The proportion of infected individuals and the proportion of susceptible individuals

As mentioned in the introduction, the proportion of infected individuals and the proportion of susceptible individuals can be reconstructed via ξ_K. For $t > 0$,

$$\xi_K([0,t] \times (t, \infty)) =: I_K(t)$$

gives the proportion of infected present at t,

$$\xi_K((t,\infty] \times [0,\infty]) = \frac{1}{K}\sum_{i=1}^{bK} I[A_i^K > t] =: S_K(t)$$

gives the proportion of susceptibles present at time t, and

$$\xi_K([0,t] \times [0,t]) =: R_K(t)$$

is the proportion of removed at time t. We obtain as weak limits for all $t \in I\!R_+$ with $\hat{\mu}(\{t\}) = 0$ and for all $T > t$ with $I\!P[r_1 = T - t] = 0$

$$\begin{aligned}
R(t) &= I\!P\!-\!\lim_{K\to\infty} R_K(t) \\
&= a\hat{\Phi}(t) + b\int_{(0,t]} \Psi_x(G_T(t-x))d\Phi(x) \\
S(t) &= I\!P\!-\!\lim_{K\to\infty} S_K(t) \\
&= b(1-\Psi(G_T(t)) \\
I(t) &= I\!P\!-\!\lim_{K\to\infty} I_K(t) \\
&= a(1-\hat{\Phi}(t)) + b\Big(\Psi(G_T(t)) - \int_{(0,t]} \Psi_x(G_T(t-x))d\Phi(x)\Big).
\end{aligned}$$

Moreover, additional information about the epidemic is provided. Suppose, for example, as mentioned in the introduction, that an epidemic is known to be taking place in a region, and that after some time t_0 every remaining susceptible in that region is immunized. Thus there are no new cases, though infectives may still be present. To decide at what time the region, which was probably put under quarantine, can be opened to the public again, we are interested in estimating the remaining infectivity in the population at times $s > t_0$. This is given by $\xi_K([0,t_0] \times (s,\infty))$, which is the proportion of individuals that were infected before the time t_0 and are still present in the population at time s. For large K and T,

$$\begin{aligned}
&\xi_K([0,t_0] \times (s,\infty)) \\
&\quad\approx\; a(1-\hat{\Phi}(s)) + b\Big(\Psi(G_T(t_0)) - \int_{(0,s]} \Psi_x(G_T(s-x))d\Phi(x)\Big).
\end{aligned}$$

Thus, as soon as this expression is smaller than a certain critical level, we can abandon the isolation.

The above expressions for the limiting quantities can be translated into a system of deterministic differential equations. In the next subsection we will see that this system coincides with results previously obtained by Wang [47], [48] in the special case that he considers. Yet, without an explicit bound on the distance to the limit these approximations remain naive.

2.3.2 Example: A Markovian epidemic

The above asymptotic results can be compared with those obtained by Wang [47], which seem to be the most general ones for the GSE known so far. As is made more explicit in Wang [48], Wang [47] considers a population of total size $N = K$ and assumed, in our notation,

1. $I\!P$ [a particular susceptible individual becomes infected during the time

interval $[t, t+\delta t]] = \lambda(I_N(t))\Delta t + o(\Delta t)$; for a function λ that is positive, bounded and Lipschitz on $[0,1]$, and $\lambda(0) = 0$.

2. $I\!P$[a particular infected individual stays infected for at least a period of length t] $= F(t)$; for a function F with $F(0) = 1$ and $F(t) \searrow 0$ as $t \to \infty$.

3. At time 0 there are $NI_N(0) = x(N)$ infected individuals $s_1, \cdots, s_{x(N)}$ present, where s_i represents also the total time that the ith individual has been infected up to time 0. There is assumed to exist a positive density $q \in L_1(I\!R_+)$ such that for all $s \in I\!R_+$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{x(N)} 1_{[0,s]}(s_i) = \int_0^s q(u)du.$$

With

$$\begin{aligned} g(s,t) &= \frac{F(t+s)}{F(s)} I[F(s) > 0], \\ \gamma(t) &= \int_0^\infty g(s,t)q(s)ds, \end{aligned}$$

Wang proves that $(I_N(t), I_N(t) + R_N(t))$ converges to the unique positive solution $(P(t), B(t))$ of the system

$$\begin{aligned} P(t) &= \gamma(t) + \int_{(0,t]} \lambda(P(u))(1 - B(u))F(t-u)du \\ B(t) &= P(0) + \int_{(0,t]} \lambda(P(u))(1 - B(u))du, \end{aligned} \tag{2.10}$$

in the sense that for every $\epsilon > 0$,

$$\lim_{N\to\infty} I\!P[\sup_{u\in[0,T]} |I_N(u) - P(u)| + |I_N(u) + R_N(u) - B(u)| > \epsilon] = 0.$$

Now consider in our model the special case that the (l_i) have an exp(1)-distribution, that l_i and r_i are independent for each i, that there are $s_i \in I\!R_+, i \in I\!N$ such that, with $\Phi(t) = 1 - F(t)$,

$$I\!P[\hat{r}_i > t] = \frac{1 - \Phi(t + s_i)}{1 - \Phi(s_i)}, \quad i \in I\!N,$$

that $\lambda(t,x) = \lambda(x(t))$, and that there is a positive density $q \in L_1(I\!R_+)$ with

$$\lim_{K\to\infty} \frac{1}{K} \sum_{i=1}^{aK} 1_{[0,s]}(s_i) = \int_0^s q(u)du.$$

Under these assumptions, we recover Wang's model, as well as his characterization of the deterministic limit; a derivation can be found in Reinert

[37]. However, our formulation covers a much wider class of models, and gives much more detailed information about the process.

In Bartlett's GSE, we moreover have $\lambda(t,x) = \alpha x(t)$, and that $s_i = 0$ for all i, so that $\gamma(t) = 0$ for all t. The above deterministic system then leads to the classical deterministic approximation; see, for example, Bailey [4] or Isham [25].

2.4 A bound on the distance to the mean-field limit

In this subsection we will prove the following refinement of Theorem 2.3.3.

Theorem 2.4.1 *Let μ^T be as in Theorem 2.3.2. Then, for all $T \in \mathbb{R}_+$, and for all $H \in \mathcal{F}$,*

$$\begin{aligned} &|\mathbb{E}^T H(\xi_K) - H(\mu)| \\ &\leq \frac{\sqrt{a}+\sqrt{b}}{\sqrt{K}} + \alpha b \beta T(T+2)\exp(b\lceil 2\alpha\beta T\rceil)\left\{(1+b)\sqrt{\frac{1}{K}} + \frac{2}{K}\right\}, \end{aligned}$$

where α and β are as in (2.2) and (2.3), and $\lceil x \rceil$ is the smallest integer larger than x.

Remarks. Note that, in Theorem 2.4.1, the constants α and β always occur as a combination. This relates to the ambiguity in Sellke's construction, referred to before: we could choose $l_i \sim \exp(\beta)$ and $\lambda(t,x) = \alpha x(t)$, or, equally, choose $l_i \sim \exp(1)$ and $\lambda(t,x) = \alpha\beta x(t)$, for example. Furthermore, the bound does not depend on the distribution of the infectious period. This is due to Lemma 2.4.2, a uniform convergence result. Of course the distribution of the infectious period is reflected in μ.

Theorem 2.4.1 gives the result in terms of expectations. Using Markov's inequality, we also obtain information for the distribution function. Namely,

$$\begin{aligned} &P\left(\left|\mathbb{E}^T H(\xi_K) - H(\mu)\right| > \frac{\log K}{\sqrt{K}}\right) \\ &\leq \frac{1}{\log K}\left\{\sqrt{a} + \sqrt{b} + \alpha b\beta T(T+2)\exp(b\lceil 2\alpha\beta T\rceil)\left\{(1+b) + \frac{2}{\sqrt{K}}\right\}\right\}. \end{aligned}$$

As the proof of Theorem 2.4.1 employs uniform convergence for the empirical distribution function, we first prove the following bound.

Lemma 2.4.2 *Let $X_1, \ldots X_n$ be i. i. d. real-valued random variables from a distribution with distribution function F, and let F_n denote the empirical distribution function*

$$F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i \leq t).$$

Then

$$\mathbb{E} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leq \frac{1}{\sqrt{n}}.$$

Proof of Lemma 2.4.2

To prove this lemma, we employ the following bound from Massart [31]. For all $\epsilon > 0$,

$$\mathbb{P}(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \epsilon) \leq 2e^{-2n\epsilon^2}.$$

The above bound is trivial for $\epsilon \leq \sqrt{\frac{\ln 2}{2n}}$. Thus we have

$$\begin{aligned}\mathbb{E} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| &= \int_0^\infty \mathbb{P}(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \epsilon)d\epsilon \\ &\leq \sqrt{\frac{\ln 2}{2n}} + \int_{\sqrt{\frac{\ln 2}{2n}}}^1 2e^{-2n\epsilon^2} d\epsilon.\end{aligned}$$

Using a change of variable we obtain

$$\begin{aligned}\mathbb{E} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| &< \sqrt{\frac{\ln 2}{n}} + \sqrt{\frac{2\pi}{n}}(1 - \Phi_{\mathcal{N}(0,1)}(\sqrt{2\ln 2})) \\ &< \frac{1}{\sqrt{n}},\end{aligned}$$

where $\Phi_{\mathcal{N}(0,1)}$ denotes the standard normal distribution function. $\square$

Now we have all the ingredients to prove Theorem 2.4.1.

Proof of Theorem 2.4.1

From Theorem 1.4.2, it suffices to bound, for all $m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m)$, and for all $\phi_1, \ldots, \phi_m \in C_b^\infty([0,T]^2)$ satisfying (1.4)

$$\sum_{j=1}^m \mathbb{E} f_{(j)}(\langle \xi_K^T, \phi_k\rangle, k = 1, \ldots, m)\langle \mu^T - \xi_K^T, \phi_j\rangle.$$

We abbreviate

$$\zeta_K = \frac{1}{bK} \sum_{i=1}^{bK} \delta_{(A_i^K, A_i^K + r_i)},$$

so that

$$\zeta_K^T = \frac{1}{bK} \sum_{i=1}^{bK} \delta_{(A_i^K, A_i^K + r_i)} \mathbf{1}(A_i^K + r_i \leq T).$$

Then we have

$$\begin{aligned}
&\left|\sum_{j=1}^{m} I\!E f_{(j)}(\langle \xi_K^T, \phi_k\rangle, k=1,\ldots,m)\langle \mu^T - \xi_K^T, \phi_j\rangle\right| \\
= &\left| a\sum_{j=1}^{m} I\!E f_{(j)}(\langle \xi_K^T, \phi_k\rangle, k=1,\ldots,m)\langle (\delta_0 \times \hat{\mu})^T - \frac{1}{aK}\sum_{i=1}^{aK} \delta^T_{(0,\hat{r}_i)}, \phi_j\rangle \right.\\
&\left. + b\sum_{j=1}^{m} I\!E f_{(j)}(\langle \xi_K^T, \phi_k\rangle, k=1,\ldots,m)\langle \tilde{\mu}^T - \zeta_K^T, \phi_j\rangle \right| \\
\leq\ & a\sum_{j=1}^{m} \| f_{(j)} \| I\!E \left| \frac{1}{aK}\sum_{i=1}^{aK}(\phi_j(0,\hat{r}_i)\mathbf{1}(\hat{r}_i \leq T) - I\!E\phi_j(0,\hat{r}_i)\mathbf{1}(\hat{r}_i \leq T))\right| \\
& + b\sum_{j=1}^{m} \| f_{(j)} \| I\!E \left| \frac{1}{bK}\sum_{i=1}^{bK}(\phi_j(A_i^K, A_i^K + r_i)\mathbf{1}(A_i^K + r_i \leq T) \right.\\
& \left. - I\!E\phi_j(G_T^{-1}(l_i), G_T^{-1}(l_i) + r_i)\mathbf{1}(G_T^{-1}(l_i) + r_i \leq T)\right|.
\end{aligned}$$

For the first summand, using the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
& a\sum_{j=1}^{m} \| f_{(j)} \| I\!E \left| \frac{1}{aK}\sum_{i=1}^{aK}(\phi_j(0,\hat{r}_i))\mathbf{1}(\hat{r}_i \leq T) - I\!E\phi_j(0,\hat{r}_i))\mathbf{1}(\hat{r}_i \leq T))\right| \\
\leq\ & a\sum_{j=1}^{m} \| f_{(j)} \| \left(\mathrm{Var}\left(\frac{1}{aK}\sum_{i=1}^{aK}\phi_j(0,\hat{r}_i)\mathbf{1}(\hat{r}_i \leq T)\right)\right)^{\frac{1}{2}} \\
\leq\ & \frac{\sqrt{a}}{\sqrt{K}},
\end{aligned}$$

where we used the boundedness assumptions from (1.4) in the last step. Thus it remains to bound

$$\begin{aligned}
& b\sum_{j=1}^{m} \| f_{(j)} \| I\!E \left| \frac{1}{bK}\sum_{i=1}^{bK}(\phi_j(A_i^K, A_i^K + r_i)\mathbf{1}(A_i^K + r_i \leq T) \right.\\
& \left. - I\!E\phi_j(G_T^{-1}(l_i), G_T^{-1}(l_i) + r_i)\mathbf{1}(G_T^{-1}(l_i) + r_i \leq T))\right|.
\end{aligned}$$

From (1.4) it suffices to bound, for any ϕ as in (1.4),

$$\begin{aligned}
& bI\!E \left| \frac{1}{bK}\sum_{i=1}^{bK}(\phi(A_i^K, A_i^K + r_i)\mathbf{1}(A_i^K + r_i \leq T) \right.\\
& \left. - I\!E\phi(G_T^{-1}(l_i), G_T^{-1}(l_i) + r_i)\mathbf{1}(G_T^{-1}(l_i) + r_i \leq T))\right|.
\end{aligned}$$

Similarly as above, we obtain

$$
\begin{aligned}
b\mathbb{E}\Big|\frac{1}{bK}\sum_{i=1}^{bK}&(\phi(G_T^{-1}(l_i),G_T^{-1}(l_i)+r_i)\mathbf{1}(G_T^{-1}(l_i)+r_i\le T)\\
&-\mathbb{E}\phi(G_T^{-1}(l_i),G_T^{-1}(l_i)+r_i)\mathbf{1}(G_T^{-1}(l_i)+r_i\le T))\Big|\\
\le\ &\frac{\sqrt{b}}{\sqrt{K}}.
\end{aligned}
$$

Bounding

$$
\begin{aligned}
b\mathbb{E}\Big|\frac{1}{bK}\sum_{i=1}^{bK}&(\phi(A_i^K,A_i^K+r_i)\mathbf{1}(A_i^K+r_i\le T)\\
&-\phi(G_T^{-1}(l_i),G_T^{-1}(l_i)+r_i)\mathbf{1}(G_T^{-1}(l_i)+r_i\le T)\Big|
\end{aligned}
$$

is more complicated, partly because the derivative of G_T^{-1} is not necessarily bounded. Instead, similarly as in Reinert [37] we mimic differentiation, making use of the additional stochasticity introduced by the random point l_1. First observe that, as λ is non-anticipating, if we omit individual 1 from the population, the course of the epidemic up to time $F_K^{-1}(l_1)$ is not affected. To make this precise, put

$$\mathcal{H}_T = D([0,T]),$$

the space of right-continuous functions with left-hand limits from $[0,T]$ to $\mathbb{R}$. Similarly to (2.4), (2.5), and (2.6), define for $h\in\mathcal{H}_T$ the operators

$$
\begin{aligned}
\mathcal{Z}_{K,1}h(t) &= \frac{1}{K}\sum_{i=1}^{aK}\mathbf{1}(\hat{r}_i>t)+\frac{1}{K}\sum_{j=2}^{bK}\mathbf{1}(h(t-r_j)<l_j\le h(t))\\
L_{K,1}h(t) &= \int_{(0,t]}\lambda(s,\mathcal{Z}_{K,1}h)\,ds.
\end{aligned}
$$

and let $F_{K,1}$ be the unique fixed point of $L_{K,1}h = h$. Then we have $F_K^{-1}(l_1)=F_{K,1}^{-1}(l_1)$ by construction. Thus

$$
\begin{aligned}
\mathbb{E}\Big|\frac{1}{bK}\sum_{i=1}^{bK}&(\phi(A_i^K,A_i^K+r_i)\mathbf{1}(A_i^K+r_i\le T)\\
&-\phi(G_T^{-1}(l_i),G_T^{-1}(l_i)+r_i)\mathbf{1}(G_T^{-1}(l_i)+r_i\le T))\Big|\\
\le\ &\mathbb{E}\left|\phi(A_1^K,A_1^K+r_1)\mathbf{1}(A_1^K+r_1\le T)\right.\\
&\left.-\phi(G_T^{-1}(l_1),G_T^{-1}(l_1)+r_1)\mathbf{1}(G_T^{-1}(l_1)+r_1\le T)\right|.
\end{aligned}
$$

Expanding the indicators, we have

$$
\begin{aligned}
&\phi(A_1^K,A_1^K+r_1)\mathbf{1}(A_1^K+r_1\le T)\\
&\quad-\phi(G_T^{-1}(l_1),G_T^{-1}(l_1)+r_1)\mathbf{1}(G_T^{-1}(l_1)+r_1\le T)
\end{aligned}
$$

$$
\begin{aligned}
&= \phi(F_{K,1}^{-1}(l_1), F_{K,1}^{-1}(l_1)+r_1)\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T) \\
&\quad -\phi(G_T^{-1}(l_1), G_T^{-1}(l_1)+r_1)\mathbf{1}(G_T^{-1}(l_1)+r_1 \le T)t \\
&= \phi(F_{K,1}^{-1}(l_1), F_{K,1}^{-1}(l_1)+r_1)\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 \le T) \\
&\quad +\phi(F_{K,1}^{-1}(l_1), F_{K,1}^{-1}(l_1)+r_1)\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 > T) \\
&\quad -\phi(G_T^{-1}(l_1), G_T^{-1}(l_1)+r_1)\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 \le T) \\
&\quad -\phi(G_T^{-1}(l_1), G_T^{-1}(l_1)+r_1)\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 > T, G_T^{-1}(l_1)+r_1 \le T).
\end{aligned}
$$

(For simplicity of notation we omit the superscript T for $F_{K,1}$ and for the expectation.) Using Taylor expansion and (1.4) we thus may bound

$$
\begin{aligned}
&\mathbb{E}\left|\frac{1}{bK}\sum_{i=1}^{bK}(\phi(A_i^K, A_i^K+r_i)\mathbf{1}(A_i^K+r_i \le T)\right. \\
&\qquad \left. -\phi(G_T^{-1}(l_i), G_T^{-1}(l_i)+r_i)\mathbf{1}(G_T^{-1}(l_i)+r_i \le T))\right| \\
&\le \mathbb{E}\left|((F_{K,1})^{-1}(l_1) - G_T^{-1}(l_1))\right. \\
&\qquad \left.\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 \le T)\right| \\
&\qquad +\mathbb{P}(F_{K,1}^{-1}(l_1)+r_1 > T, G_T^{-1}(l_1)+r_1 \le T) \\
&\qquad +\mathbb{P}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 > T).
\end{aligned}
$$

Firstly,

$$
\begin{aligned}
&\mathbb{P}(F_{K,1}^{-1}(l_1)+r_1 > T, G_T^{-1}(l_1)+r_1 \le T) \\
&\le \beta\mathbb{E} \parallel F_{K,1} - G_T \parallel_T .
\end{aligned}
$$

Thus, by symmetry,

$$
\begin{aligned}
&\mathbb{E}\left|((F_{K,1})^{-1}(l_1) - G_T^{-1}(l_1))\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 \le T)\right| \\
&+ \mathbb{P}(F_{K,1}^{-1}(l_1)+r_1 > T, G_T^{-1}(l_1)+r_1 \le T) \\
&\quad +\mathbb{P}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 > T) \\
&\le \mathbb{E}\left|(F_{K,1})^{-1}(l_1) - G_T^{-1}(l_1))\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 \le T)\right| \\
&\quad +2\beta\mathbb{E} \parallel F_{K,1} - G_T \parallel_T .
\end{aligned}
$$

We will bound $\mathbb{E} \parallel F_{K,1} - G_T \parallel_T$ later. Using the existence of the density ψ, for the first term we have

$$
\begin{aligned}
&\mathbb{E}\left|((F_{K,1})^{-1}(l_1) - G_T^{-1}(l_1))\mathbf{1}(F_{K,1}^{-1}(l_1)+r_1 \le T, G_T^{-1}(l_1)+r_1 \le T)\right| \\
&\le \mathbb{E}\left|(F_{K,1})^{-1}(l_1) - G_T^{-1}(l_1)\right|\mathbf{1}(F_{K,1}^{-1}(l_1) \le T, G_T^{-1}(l_1) \le T) \\
&= \mathbb{E}\int_0^{F_{K,1}(T)\wedge G_T(T)} \left|(F_{K,1})^{-1}(x) - G_T^{-1}(x)\right| \\
&\qquad \mathbf{1}(F_{K,1}^{-1}(x) \le T, G_T^{-1}(x) \le T)\psi(x)dx
\end{aligned}
$$

$$
\begin{aligned}
= \quad & I\!E \int_0^{F_{K,1}(T)\wedge G_T(T)} \left| G_T^{-1}(G_T((F_{K,1})^{-1}(x))) - G_T^{-1}(x) \right| \\
& \mathbf{1}(F_{K,1}^{-1}(x) \le T, G_T^{-1}(x) \le T)\psi(x)dx.
\end{aligned}
$$

From Condition 4. on λ we have that $F_{K,1}$ and G_T are strictly increasing until they stay constant forever after. Thus the derivatives $(G_T^{-1})'$ and $(F_{K,1}^{-1})'$ exist for all t for which G_T and $F_{K,1}$ are not constant; in particular they exist for $t < F_{K,1}^{-1}(x)$ when $F_{K,1}^{-1}(x) \le T$, and for $t < G_T^{-1}(x)$ when $G_T^{-1}(x) \le T$. Thus we may integrate as follows.

$$
\begin{aligned}
& I\!E \int_0^{F_{K,1}(T)\wedge G_T(T)} \left| G_T^{-1}(G_T((F_{K,1})^{-1}(x))) - G_T^{-1}(x) \right| \\
& \quad \mathbf{1}(F_{K,1}^{-1}(x) \le T, G_T^{-1}(x) \le T)\psi(x)dx \\
& = \quad I\!E \int_0^{F_{K,1}(T)\wedge G_T(T)} \left| \int_x^{G_T(F_{K,1})^{-1}(x))} (G_T^{-1})'(y)dy \right| \psi(x)dx \\
& = \quad I\!E \int_0^{F_{K,1}(T)\wedge G_T(T)} \int_x^{G_T((F_{K,1})^{-1}(x))} (G_T^{-1})'(y)dy \\
& \quad \mathbf{1}(x < G_T((F_{K,1})^{-1}(x)))\psi(x)dx \\
& \quad + I\!E \int_0^{F_{K,1}(T)\wedge G_T(T)} \int_{G_T((F_{K,1})^{-1}(x))}^{x} (G_T^{-1})'(y)dy \\
& \quad \mathbf{1}(x > G_T((F_{K,1})^{-1}(x)))\psi(x)dx.
\end{aligned}
$$

As all integrals are finite, we may interchange the order of integration and obtain for the above expression

$$
\begin{aligned}
& I\!E \int_0^{G_T(F_{K,1}^{-1}(F_{K,1}(T)\wedge G_T(T)))} \int_{F_{K,1}(G_T^{-1}(y))}^{G_T(G_T^{-1}(y))} \mathbf{1}(x < G_T((F_{K,1})^{-1}(x))) \\
& \quad \psi(x)dx(G_T^{-1})'(y)dy \\
& + \quad I\!E \int_0^{G_T(F_{K,1}^{-1}(F_{K,1}(T)\wedge G_T(T)))} \int_{G_T(G_T^{-1}(y))}^{F_{K,1}(G_T^{-1}(y))} \mathbf{1}(x > G_T((F_{K,1})^{-1}(x))) \\
& \quad \psi(x)dx(G_T^{-1})'(y)dy \\
& \le \quad \beta T I\!E \parallel F_{K,1} - G_T \parallel_T .
\end{aligned}
$$

Thus we have derived that

$$
\begin{aligned}
& b I\!E \left| \frac{1}{bK} \sum_{i=1}^{bK} (\phi(A_i^K, A_i^K + r_i)\mathbf{1}(A_i^K + r_i \le T) \right. \\
& \quad \left. -\phi(G_T^{-1}(l_i), G_T^{-1}(l_i) + r_i)\mathbf{1}(G_T^{-1}(l_i) + r_i \le T) \right| \\
& \le \quad b\beta(T+2) I\!E \parallel F_{K,1} - G_T \parallel_T .
\end{aligned}
$$

Bound on $\| F_{K,1}^T - G_T \|_T$

To bound this quantity, we proceed by an inductive argument, similarly to the proof of Theorem 2.3.1 in Reinert [37]. For any t we have

$$\begin{aligned} F_{K,1}(t) - G_T(t) &= L_{K,1}F_{K,1}(t) - LG_T(t) \\ &= (L_{K,1}F_{K,1}(t) - LF_{K,1}(t)) + (LF_{K,1}(t) - LG_T(t)), \end{aligned}$$

so that, for all $t \leq T$,

$$\mathbb{E} \| F_{K,1} - G_T \|_t \leq \mathbb{E} \sup_{h \in \mathcal{H}_t} \| L_{K,1}h - Lh \|_T + \mathbb{E} \| LF_{K,1} - LG_T \|_t .$$

First we bound $\mathbb{E} \sup_{h \in \mathcal{H}_t} \| L_{K,1}h - Lh \|_T$. For $h \in \mathcal{H}_T$ we have, due to the Lipschitz property of λ, that

$$\begin{aligned} \| L_{K,1}h - Lh \|_T &\leq \alpha \int_0^T \sup_{s \leq x} |\mathcal{Z}_{K,1}h(s) - \mathcal{Z}h(s)| dx \\ &\leq \alpha T \left(aR_2 + 2bR_3 + \frac{2}{K} \right), \end{aligned}$$

where

$$R_2 = \sup_s \left| \frac{1}{aK} \sum_{i=1}^{aK} \mathbf{1}(\hat{r}_i \leq s) - \hat{\Phi}(s) \right|$$

and

$$R_3 = \sup_s \left| \frac{1}{bK-1} \sum_{i=2}^{bK} \mathbf{1}(l_i \leq s) - \Psi(s) \right| .$$

From Lemma 2.4.2 we have that for both $u = 1$ and $u = 2$, $\mathbb{E}R_u \leq \sqrt{\frac{1}{K}}$. As $a + b = 1$, this yields

$$\begin{aligned} \mathbb{E} \sup_{h \in \mathcal{H}_t} \| L_{K,1}h - Lh \|_T &\leq \alpha T \left\{ (1+b)\sqrt{\frac{1}{K}} + \frac{2}{K} \right\} \\ &= : S(K). \end{aligned} \tag{2.11}$$

Now we bound $\mathbb{E} \| LF_{K,1} - LG_T \|_t$. We have

$$\begin{aligned} &|LF_{K,1}(t) - LG_T(t)| \\ &\leq \alpha b \int_0^t \sup_{s \leq x} \Big| \Psi(F_{K,1}(s)) - \Psi(G_T(s)) \\ &\qquad + \int_0^s \{ \Psi_u(F_{K,1}(s-u)) - \Psi_u(G_T(s-u)) \} \mathbb{P}(r_1 \in du) \Big| dx \\ &\leq \alpha b \beta \int_0^t \| F_{K,1} - G_T \|_x \, (1 + \Phi(x)) dx. \end{aligned}$$

Thus we obtain

$$\begin{aligned} & \mathbb{E} \parallel F_{K,1} - G_T \parallel_t \\ \leq \quad & S(K) + \alpha b \beta \int_0^t \mathbb{E} \parallel F_{K,1} - G_T \parallel_x (1 + \Phi(x)) dx. \end{aligned} \tag{2.12}$$

Fix $c \geq b$ arbitrary, and define

$$\eta = \frac{1}{2c\alpha\beta}. \tag{2.13}$$

Then we have that

$$2\eta\alpha\beta b \leq \frac{b}{c}.$$

Hence, as $\Phi(s) \leq 1$ always,

$$\mathbb{E} \parallel F_{K,1} - G_T \parallel_\eta \quad \leq \quad S(K) + \frac{b}{c} \mathbb{E} \parallel F_{K,1} - G_T \parallel_\eta;$$

yielding

$$\begin{aligned} \mathbb{E} \parallel F_{K,1} - G_T \parallel_\eta \quad & \leq \quad \frac{1}{1 - \frac{b}{c}} S(K) \\ & = \quad \frac{c}{c-b} S(K). \end{aligned}$$

We now prove by induction that, for any $k \in \mathbb{N}$,

$$\mathbb{E} \parallel F_{K,1} - G_T \parallel_{k\eta} \quad \leq \quad \left(\frac{c}{c-b} \right)^k S(K). \tag{2.14}$$

The case $k = 1$ has already been proven above. Suppose (2.14) is true for k. Then, from (2.12),

$$\begin{aligned} & \mathbb{E} \parallel F_{K,1} - G_T \parallel_{(k+1)\eta} \\ \leq \quad & S(K) + \alpha b \beta \int_0^{(k+1)\eta} \mathbb{E} \parallel F_{K,1} - G_T \parallel_x (1 + \Phi(x)) dx \\ \leq \quad & S(K) + \alpha b \beta \sum_{l=1}^{k} \int_{(l-1)\eta}^{l\eta} \mathbb{E} \parallel F_{K,1} - G_T \parallel_x (1 + \Phi(x)) dx \\ & + \alpha b \beta \int_{k\eta}^{(k+1)\eta} \mathbb{E} \parallel F_{K,1} - G_T \parallel_x (1 + \Phi(x)) dx \\ \leq \quad & S(K) + \alpha b \beta \eta \sum_{l=1}^{k} \left(\frac{c}{c-b} \right)^l S(K) \\ & + \alpha b \beta \int_{k\eta}^{(k+1)\eta} \mathbb{E} \parallel F_{K,1} - G_T \parallel_x (1 + \Phi(x)) dx, \end{aligned}$$

where we used the induction assumption. From the definition (2.13) of η we now obtain

$$\begin{aligned}
&I\!E \parallel F_{K,1} - G_T \parallel_{(k+1)\eta} \\
&\quad\leq \quad S(K) + \frac{b}{c}\sum_{l=1}^{k}\left(\frac{c}{c-b}\right)^{l} S(K) + \frac{b}{c} I\!E \parallel F_{K,1} - G_T \parallel_{(k+1)\eta};
\end{aligned}$$

thus

$$\begin{aligned}
I\!E \parallel F_{K,1} - G_T \parallel_{(k+1)\eta} &\leq \frac{1}{1-\frac{b}{c}} S(K)\left(1 + \frac{b}{c}\sum_{l=1}^{k}\left(\frac{c}{c-b}\right)^{l}\right) \\
&= \left(\frac{c}{c-b}\right)^{k+1} S(K).
\end{aligned}$$

This proves (2.14). Denoting by $\lceil \frac{T}{\eta} \rceil$ the smallest integer at least as large as $\frac{T}{\eta}$, we have hence shown that, with (2.13),

$$\begin{aligned}
I\!E \parallel F_{K,1} - G_T \parallel_T &\leq \left(\frac{c}{c-b}\right)^{\lceil \frac{T}{\eta} \rceil} S(K) \\
&= \exp(\lceil 2c\alpha\beta T \rceil (\ln(c) - \ln(c-b)) S(K).
\end{aligned}$$

As $c > b$ was arbitrarily chosen, we may take the limit $c \to \infty$ and obtain

$$\begin{aligned}
I\!E \parallel F_{K,1} - G_T \parallel_T &\leq \exp(b\lceil 2\alpha\beta T \rceil) S(K) \\
&= \alpha T \exp(b\lceil 2\alpha\beta T \rceil)\left\{(1+b)\sqrt{\frac{1}{K}} + \frac{2}{K}\right\},
\end{aligned}$$

using (2.11). Combining these steps we obtain

$$\begin{aligned}
& bI\!E \left| \frac{1}{bK}\sum_{i=1}^{bK}(\phi(A_i^K, A_i^K + r_i)\mathbf{1}(A_i^K + r_i \leq T) \right. \\
&\qquad \left. -\phi(G_T^{-1}(l_i), G_T^{-1}(l_i) + r_i)\mathbf{1}(G_T^{-1}(l_i) + r_i \leq T) \right| \\
&\quad \leq \quad \alpha b\beta T(T+2)\exp(b\lceil 2\alpha\beta T \rceil)\left\{(1+b)\sqrt{\frac{1}{K}} + \frac{2}{K}\right\}.
\end{aligned}$$

This completes the proof. □

In practice, the above bound may become rather large. If K is very large, and $b \approx 1$, then the bound is less than 1 only if

$$K > 4(\alpha\beta)^2 T^4 e^{4\alpha\beta T},$$

which often would be valid only for the initial stages of an epidemic, or if $\alpha\beta$ is tiny. In the latter case, most of the susceptibles become infected almost instantly, so that the epidemic process behaves nearly as a simple death process. In this case, the deterministic approximation should be good, as uniform approximations could be in place, see Lemma 2.4.2.

The fact that the deterministic approximation is not as good as one would expect when inspecting epidemic curves has already been noticed by Metz [33]. He shows heuristically that epidemic curves all have essentially the same form, differing only by a random transformation. The offset caused by this random transformation might be the major source of error in the deterministic approximation. In the following subsection, we will make this heuristic precise, for a special case.

2.5 A special case: $\lambda(t, x) = \alpha x(t)$

In the case that $\lambda(t, x) = x(t)$, much more can be said. Here we consider this special case, and we furthermore assume that l_i and r_i are independent, and that r_i possesses a density ϕ. (Note that this case includes Bartlett's GSE as a special case.) Much of the uncertainty in approximating an epidemic lies in the initial stages. Initially, only few individuals would typically be infected, so it might take a long time until the epidemic takes off. If, instead, we use a deterministic approximation only after the epidemic has acquired a substantial size, the approximation would be much improved.

With the notation above, choose a "threshold" value d and let t_0 be such that $G_T(t_0) = d$. Assume that, for some $\rho > 0$, $I(t_0 + t) \geq \rho$ and $I_K(\tau + t) \geq \rho$ for all $t \leq T$. Define

$$\tau = \inf\{t : F_K(t) = d\}.$$

Put, for all $t > 0$,

$$G_d(t) = G_T(t_0 + t)$$

and

$$F_d(t) = F_K(\tau + t).$$

Thus we only approximate after the total infectivity in the population has reached the level d. Our new empirical measure is

$$\xi^d = \frac{1}{bK} \sum_{i=1}^{bK} \delta_{(F_d^{-1}(l_i), F_d^{-1}(l_i) + r_i)},$$

considered on $\mathbb{R}_+ \times \mathbb{R}_+$, and the approximating measure $\tilde{\mu}^d$ is given by

$$\tilde{\mu}^d([0, r] \times [0, s]) = \mathbb{P}[l_1 \leq G_d(r), l_1 \leq G_d(s - r_1)]. \tag{2.15}$$

Furthermore assume that

$$\alpha b \psi(d) \mathbb{E}^T(r_1) < 1. \tag{2.16}$$

We obtain the following proposition.

Proposition 2.5.1 *Suppose that $\lambda(t,x) = \alpha x(t)$. Assume that ψ decreases monotonically on $[t_0, T]$. Let $\tilde{\mu}^d$ be given in (2.15), and let d and T be such that (2.16) holds. Then, for all $T \in \mathbb{R}_+$, and for all $H \in \mathcal{F}$,*

$$\begin{aligned} & |\mathbb{E}^T H(\xi^d) - H(\tilde{\mu}^d)| \\ \leq \quad & \frac{1}{\sqrt{bK}} + \frac{\alpha b T}{\rho - \rho\alpha b \psi(d) \mathbb{E}^T(r_1)} \left\{ (1+b)\sqrt{\frac{1}{K}} + \frac{2}{K} \right\}. \end{aligned}$$

Remark. Note that the bound is now only linear in T, instead of exponential in T. Moreover, in the next subsection some plots will illustrate that the assumptions can be fulfilled in reasonable cases.

The proof of Proposition 2.5.1 is rather similar to the proof of Theorem 2.4.1. Due to the additional assumptions, though, the contraction argument is much simplified.

Proof of Proposition 2.5.1

From Theorem 1.4.2, it suffices to bound, for all $m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m)$, and for all $\phi_1, \ldots, \phi_m \in C_b^\infty([0,T]^2)$ satisfying (1.4)

$$\sum_{j=1}^m \mathbb{E}^T f_{(j)}(\langle \xi^d, \phi_k \rangle, k = 1, \ldots, m) \langle \tilde{\mu}^d - \xi^d, \phi_j \rangle.$$

We have

$$\begin{aligned} & \left| \sum_{j=1}^m \mathbb{E}^T f_{(j)}(\langle \xi^d, \phi_k \rangle, k = 1, \ldots, m) \langle \tilde{\mu}^d - \xi^d, \phi_j \rangle \right| \\ \leq \quad & \sum_{j=1}^m \| f_{(j)} \| \, \mathbb{E}^T \left| \frac{1}{bK} \sum_{i=1}^{bK} (\phi_j(F_d^{-1}(l_i), F_d^{-1}(l_i) + r_i) \right. \\ & \left. - \mathbb{E}\phi_j(G_d^{-1}(l_i), G_d^{-1}(l_i) + r_i) \right| \\ \leq \quad & \mathbb{E}^T \left| \frac{1}{bK} \sum_{i=1}^{bK} (\phi(G_d^{-1}(l_i), G_d^{-1}(l_i) + r_i) \right. \\ & \left. - \mathbb{E}^T \phi(G_d^{-1}(l_i), G_d^{-1}(l_i) + r_i)) \right| \\ & + \mathbb{E}^T \left| \frac{1}{bK} \sum_{i=1}^{bK} (\phi(F_d^{-1}(l_i), F_d^{-1}(l_i) + r_i) - \phi(G_d^{-1}(l_i), G_d^{-1}(l_i) + r_i) \right| \\ \leq \quad & \frac{1}{\sqrt{bK}} + \mathbb{E}^T \left| F_d^{-1}(l_i) - G_d^{-1}(l_i) \right|, \end{aligned}$$

where we used the Cauchy-Schwarz inequality for the first summand, as in the proof of Theorem 2.4.1.

The difficulty lies again in bounding

$$I\!E^T \left| F_d^{-1}(l_i) - G_d^{-1}(l_i) \right|.$$

In contrast to the proof of Theorem 2.4.1, we now differentiate, giving that

$$\begin{aligned} |F_d^{-1}(l_i) - G_d^{-1}(l_i)| &= |G_d^{-1}(G_d(F_d^{-1}(l_i))) - G_d^{-1}(F_d(F_d^{-1}(l_i)))| \\ &= |(G_d^{-1})'(l_i + \theta)(G_d(F_d^{-1}(l_i))) - F_d(F_d^{-1}(l_i)))|, \end{aligned}$$

for some θ. As we assumed that $G_d' = I(t_0 + t) \geq \rho$ on the interval considered, we may bound

$$I\!E^T \left| F_d^{-1}(l_i) - G_d^{-1}(l_i) \right| \leq \frac{1}{\rho} I\!E^T |G_d(F_d^{-1}(l_i))) - l_i|.$$

Observe that, if we omit individual 1 from the population, the course of the epidemic up to time $F_d^{-1}(l_1)$ is not affected. As in the proof of Theorem 2.4.1, define $\mathcal{H}_T = D([0,T])$, and, for $h \in \mathcal{H}_T$

$$\begin{aligned} \mathcal{Z}_{K,1} h(t) &= \frac{1}{K} \sum_{i=1}^{aK} \mathbf{1}(\hat{r}_i > t) + \frac{1}{K} \sum_{j=2}^{bK} \mathbf{1}(h(t - r_j) < l_j \leq h(t)) \\ L_{K,1} h(t) &= \int_{(0,t]} \lambda(s, \mathcal{Z}_{K,1} h)\, ds. \end{aligned}$$

Let $F_{K,1}$ denote the unique fixed point of $L_{K,1} h = h$. Then $F_K(t) = F_{K,1}(t)$ for all $t \leq l_1$ by construction. Moreover, put

$$F_{d,1}(t) = F_{K,1}(\tau + t).$$

Then

$$\begin{aligned} F_{d,1}^{-1}(l_1) &= \inf\{t : F_{d,1} = l_1\} \\ &= \inf\{t : F_{K,1}(d + t) = l_1\} \\ &= \inf\{t : F_K(d + t) = l_1\} \\ &= F_d^{-1}(l_1). \end{aligned}$$

Similarly, $F_{d,1}^{-1}(s) = F_d^{-1}(s)$ for all $s \leq l_1$. Furthermore, for any t we have

$$\begin{aligned} & F_{d,1}(t) - G_d(t) \\ &= L_{K,1} F_{K,1}(\tau + t) - L G_T(t_0 + t) \\ &= L_{K,1} F_{K,1}(\tau + t) - L F_{K,1}(\tau + t) + L F_{K,1}(\tau + t) - L G_T(t_0 + t) \\ &= L_{K,1} F_{K,1}(\tau + t) - L F_{K,1}(\tau + t) + L F_{d,1}(t) - L G_d(t). \end{aligned}$$

Hence, for all $s \leq F_d^{-1}(l_1)$,

$$\begin{aligned} & G_d(s) - F_d(s) \\ &= L_{K,1} F_{K,1}(\tau + s) - L F_{K,1}(\tau + s) + L F_{d,1}(s) - L G_d(s) \\ &= L_{K,1} F_{K,1}(\tau + s) - L F_{K,1}(\tau + s) + L F_d(s) - L G_d(s). \end{aligned}$$

Thus

$$\begin{aligned}
&\sup_{s\le F_d^{-1}(l_i)} |G_d(F_d^{-1}(l_i)) - l_i| \\
&\quad \le \sup_{h\in\mathcal{H}_t} \| L_{K,1}h - Lh \|_T + \sup_{s\le F_d^{-1}(l_i)} |LF_d(t) - LG_d(t)|.
\end{aligned}$$

As before, using Inequality (2.11),

$$\begin{aligned}
I\!\!E \sup_{h\in\mathcal{H}_t} \| L_{K,1}h - Lh \|_T &\le \alpha T \left\{ (1+b)\sqrt{\frac{1}{K}} + \frac{2}{K} \right\} \\
&= S(K).
\end{aligned}$$

Now we bound $|LF_d(F_d^{-1}(l_1)) - LG_d(F_d^{-1}(l_1))|$. Recall that, for any function f, as $\lambda(t,x) = \alpha x(t)$,

$$\begin{aligned}
&Lf(t) \\
&= a\int_0^t (1-\hat{\Phi}(s))ds + \alpha b \int_0^t \left(\Psi(f(s)) - \int_0^s \Psi(f(s-u))\phi(u)du \right) ds.
\end{aligned}$$

For the last integral, interchanging the order of integration gives

$$\begin{aligned}
&\int_0^t \int_0^s \Psi(f(s-u))\phi(u)duds \\
&\quad = \int_0^t \int_0^s \Psi(f(y))\phi(s-y)dyds \\
&\quad = \int_0^t \Psi(f(y)) \int_y^t \phi(s-y)dsdy \\
&\quad = \int_0^t \Psi(f(y)) \int_0^{t-y} \phi(s)dsdy \\
&\quad = \int_0^t \Psi(f(y))\Phi(t-y)dy.
\end{aligned}$$

Thus

$$\begin{aligned}
&Lf(t) \\
&\quad = a\int_0^t (1-\hat{\Phi}(s))ds + \alpha b \int_0^t \Psi(f(s))(1-\Phi(t-s))ds.
\end{aligned}$$

Hence,

$$\begin{aligned}
&|LF_d(F_d^{-1}(l_1)) - LG_d(F_d^{-1}(l_1))| \\
&\quad \le \alpha b \left| \int_0^{F_d^{-1}(l_1)} (\Psi(F_d(s)) - \Psi(G_d(s)))(1-\Phi(F_d^{-1}(l_1)-s))ds \right| \\
&\quad \le \psi(d) \| F_d - G_d \|_{F_d^{-1}(l_1)} \int_0^{F_d^{-1}(l_1)} (1 - -\Phi(F_d^{-1}(l_1)-s)ds
\end{aligned}$$

$$= \quad \psi(d)\mathbb{E}^T(r_1) \parallel F_d - G_d \parallel_{F_d^{-1}(l_1)},$$

noting that $\psi(t) < \psi(d)$ on the interval considered, as $F_d \geq d$ and $G_d \geq d$. This yields

$$\begin{aligned} &\mathbb{E}^T \sup_{s \leq F_d^{-1}(l_1)} |LF_d(s) - LG_d(s)| \\ &\leq \quad \alpha b\psi(d)\mathbb{E}^T(r_1) \parallel G_d - F_d \parallel_{F_d^{-1}(l_1)} . \end{aligned}$$

By Assumption (2.16), we may apply the contraction argument without having to dissect the target interval. Thus

$$\begin{aligned} &\mathbb{E}^T \parallel F_d - G_d \parallel_{F_d^{-1}(l_1)} \\ &\leq \quad S(K) + \alpha b\psi(d)\mathbb{E}^T(r_1)\mathbb{E}^T \parallel G_d - F_d \parallel_{F_d^{-1}(l_1)}, \end{aligned}$$

yielding that

$$\begin{aligned} &\mathbb{E}^T \parallel G_d - F_d \parallel_{F_d^{-1}(l_1)} \\ &\leq \quad \frac{1}{1 - \alpha b\psi(d)\mathbb{E}^T(r_1)} S(K) \\ &= \quad \frac{1}{1 - \alpha b\psi(d)\mathbb{E}^T(r_1)} \alpha bT \left\{ (1+b)\sqrt{\frac{1}{K}} + \frac{2}{K} \right\}. \end{aligned}$$

This completes the proof. □

Typically, $I(t)$ would be unimodal on a large interval, as would be $I_K(t)$, so that the restriction on being at least as large as ρ would be natural. To see this, in the next section we show some plots.

2.6 Some plots of the limiting expression

For simplicity, here we consider Bartlett's GSE. Here we have the case that $\Phi(x) = 1 - e^{-\beta x}$, and that $\phi(x) = e^{-x}$; if $\alpha = 1$ and $b \approx 1$, say, then, for $\beta \geq 1$ any $d > 0$ would satisfy Assumption (2.16).

Here, β can be interpreted as the relative removal rate. It is well-known (see Bailey [4] that, if $\beta < 1$, then the chance of ultimate extinction of the epidemic is less than unity, whereas for $\beta \geq 1$, the chance of ultimate extinction of the epidemic is unity. In the latter case, only a minor outbreak of the epidemic would be expected, whereas in the first case, a major build-up may occur. For the case of a minor outbreak, Proposition 2.5.1 is suitable.

Firstly we choose as parameters $a = 0.01$, $\alpha = 1$, as usual infection rate 1, and removal rate $\beta = 2$. The first plot, Figure 2.1, is the asymptotic expression for the proportion of removed individuals. It shows the characteristic S-shape that one would expect, see, e.g. Bailey [4]. Figure 2.2 shows the proportion of infectives. The proportion of infectives first decreases, corresponding to the initially infected individuals in the population, until enough infectivity in the population has accumulated; then it increases to a peak, and then decreases until it reaches 0; then the epidemic dies out. Thus the unimodality is valid for a large time interval. Figure 2.3 shows the cumulative infectivity function G. As mentioned above, any value $d > 0$ would be admissible. The bound in Proposition 2.5.1 improves with larger ρ, which corresponds to a shorter time interval. For $\rho = .0001$, for example, dependent on the observations, the time interval could be chosen as $[17, 134]$. As $G(17) = 0.00459$, choosing $d = 0.00459$, Proposition 2.5.1 provides as bound, for any $t \in [17, 134]$, $\frac{1}{\sqrt{K}}\left(1.006 + 19,517(t-17) + 2 + \frac{2}{\sqrt{K}}\right)$. For $\rho = .001$, in comparison, the time interval could be chosen as $[50, 100]$; as $G(50) = 0.009451$, Proposition 2.5.1 provides as bound, for any $t \in [50, 100]$, $\frac{1}{\sqrt{K}}\left(1.006 + 1,942(t-50) + 2 + \frac{2}{\sqrt{K}}\right)$. So, in practice, K has to be very large to make this bound useful.

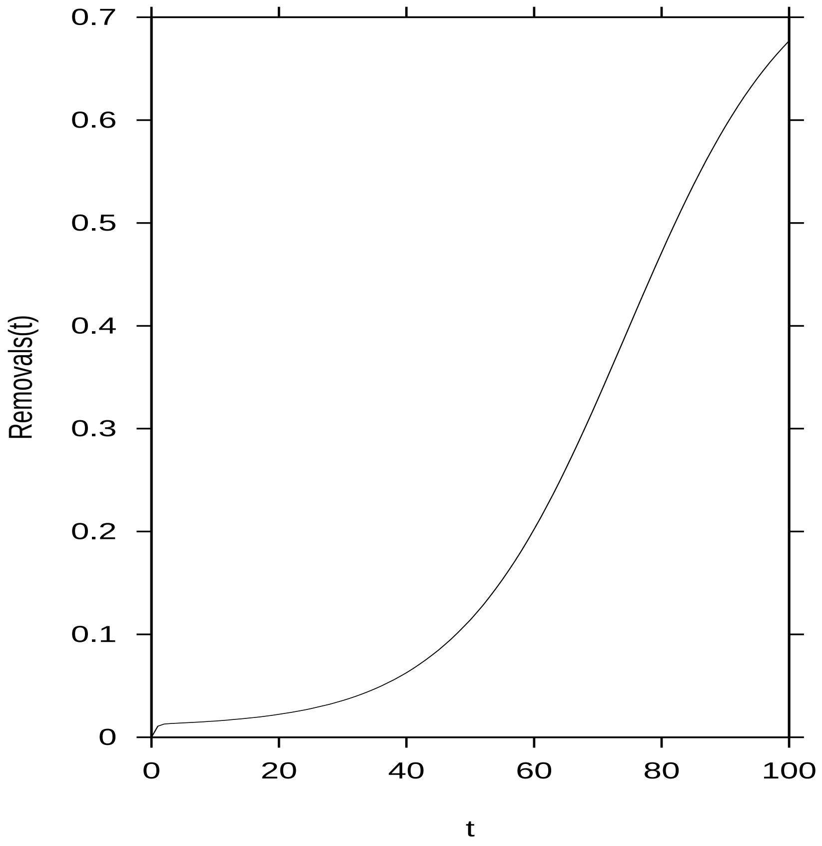

Figure 2.1 *Proportion removals in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 2$

The next series of plots illustrate the "critical" case $\beta = 1$ all the other parameters are as above. For $\rho = .01$, for example, dependent on the observations, the time interval could be chosen as $[31, 40]$. As $G(16) = 0.040931$, Proposition 2.5.1 provides as bound, for any $t \in [31, 40]$, the value $\frac{1}{\sqrt{K}}\left(1.006 + 200(t-40) + 2 + \frac{2}{\sqrt{K}}\right)$. Still, K has to be very large to make this bound useful in practice.

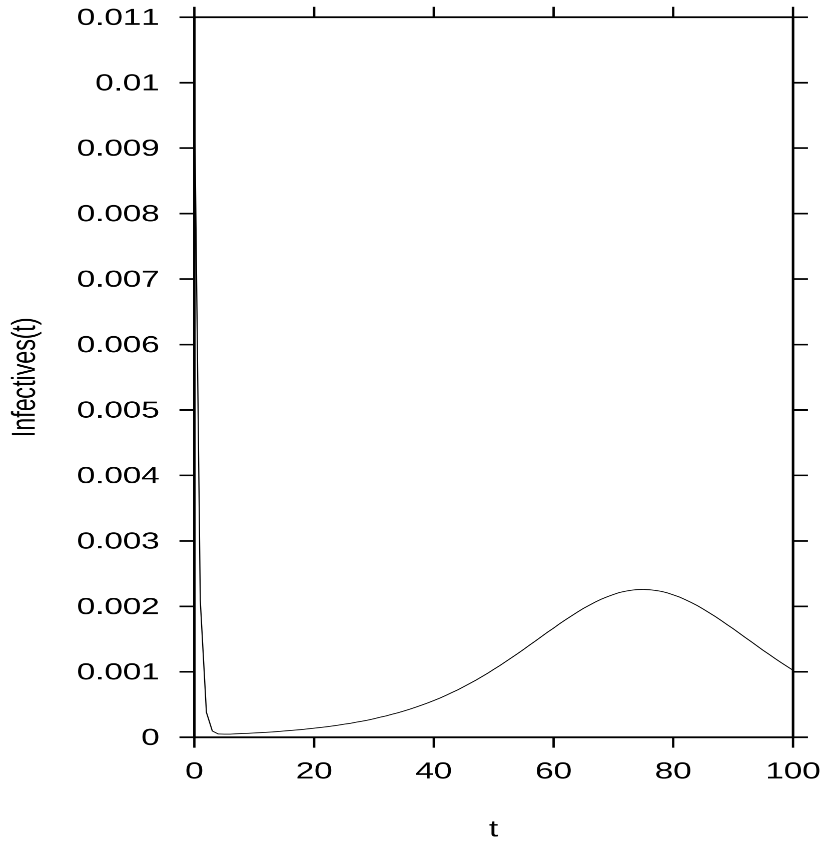

Figure 2.2 *Proportion infectives in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 2$

Lastly, Figure 2.7, Figure 2.8 and Figure 2.9 show the corresponding plots with $\beta = 0.3$. In this case, Assumption (2.16) translates to $e^{-d} < 0.3$, or $d > \ln(10/2.97) \approx 1.21$, a case that is clearly not of interest, as the function G never reaches that level. Thus the scope of Proposition 2.5.1 is limited. However, it would be possible to refine the contraction approach, by separating the time interval into only a few intervals, for which the contraction property then would hold again. This would estend the scope of Proposition 2.5.1, but would yield bounds that are higher order polynomials in T.

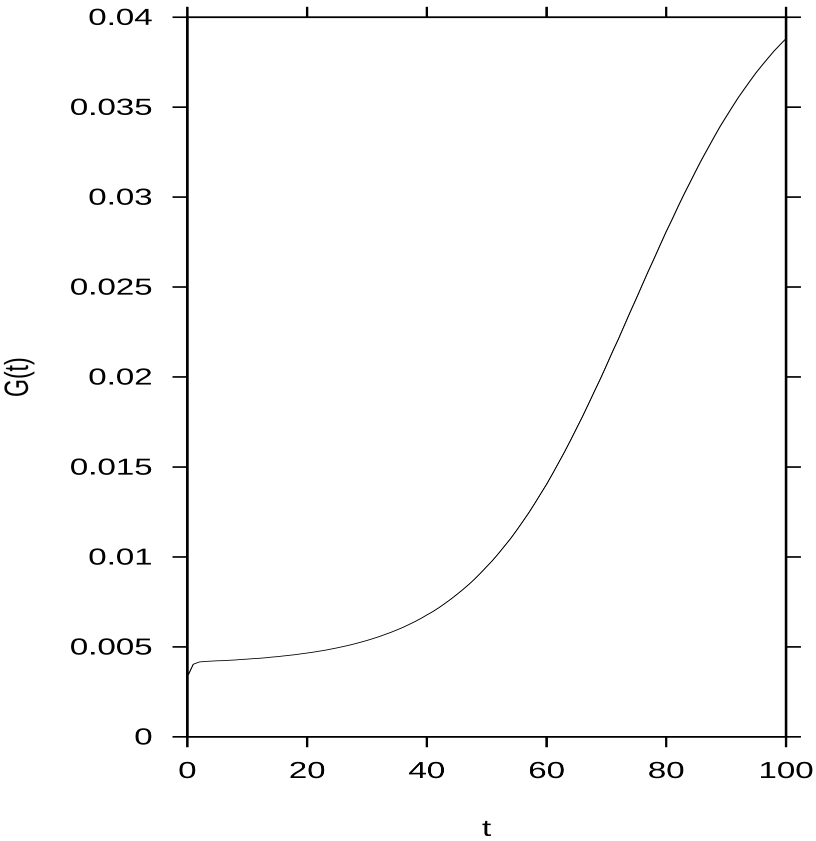

Figure 2.3 *Cumulative infectivity function G in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 2$

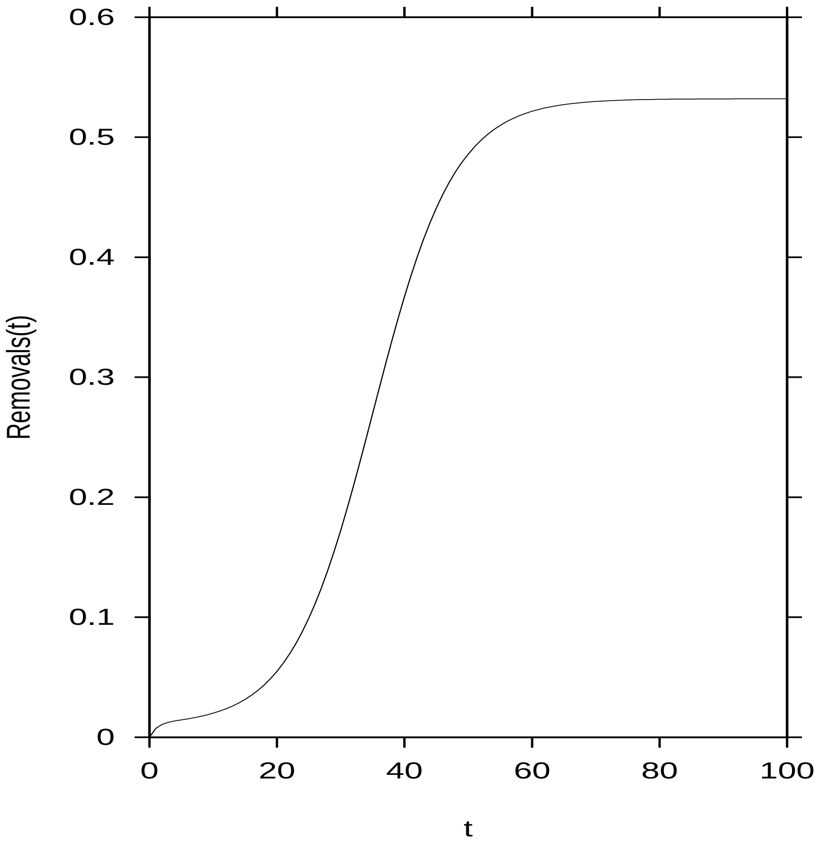

Figure 2.4 *Proportion removals in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 1$

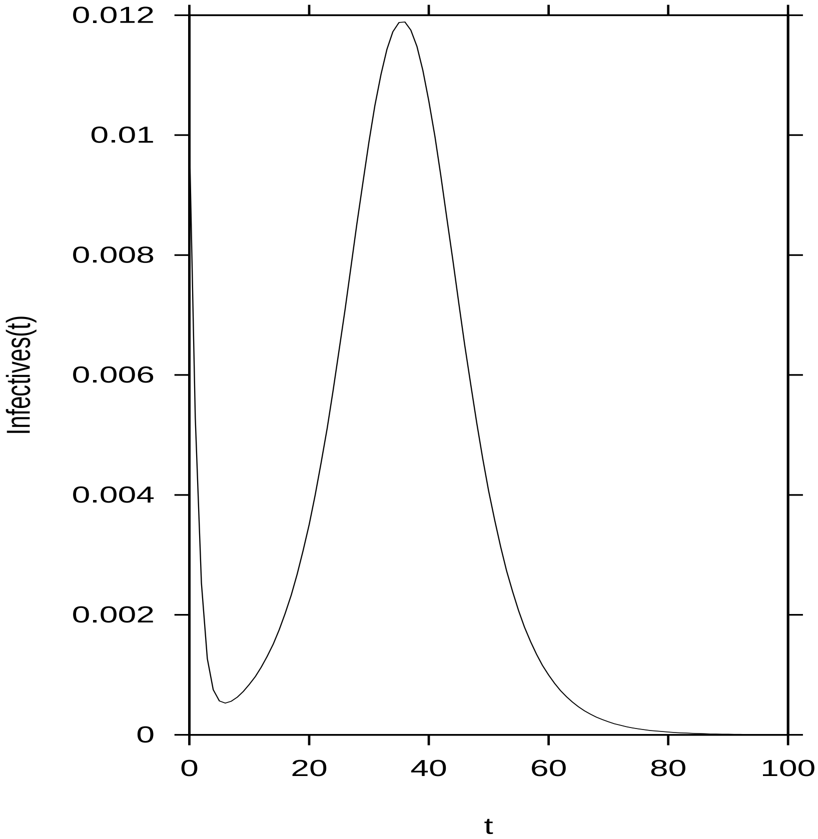

Figure 2.5 *Proportion infectives in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 1$

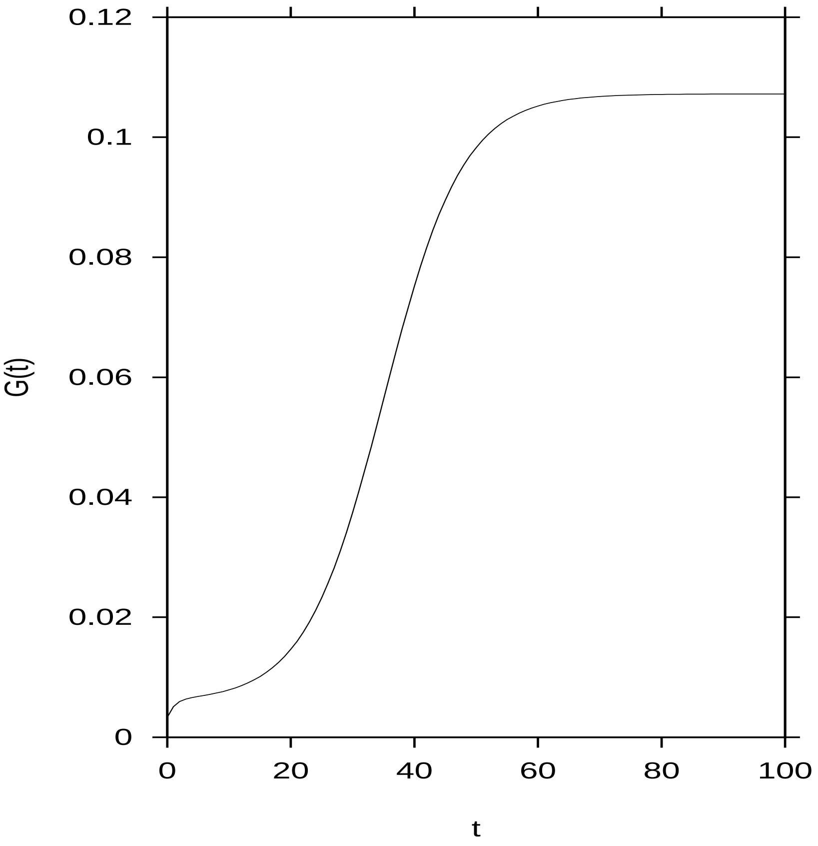

Figure 2.6 *Cumulative infectivity function G in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 1$

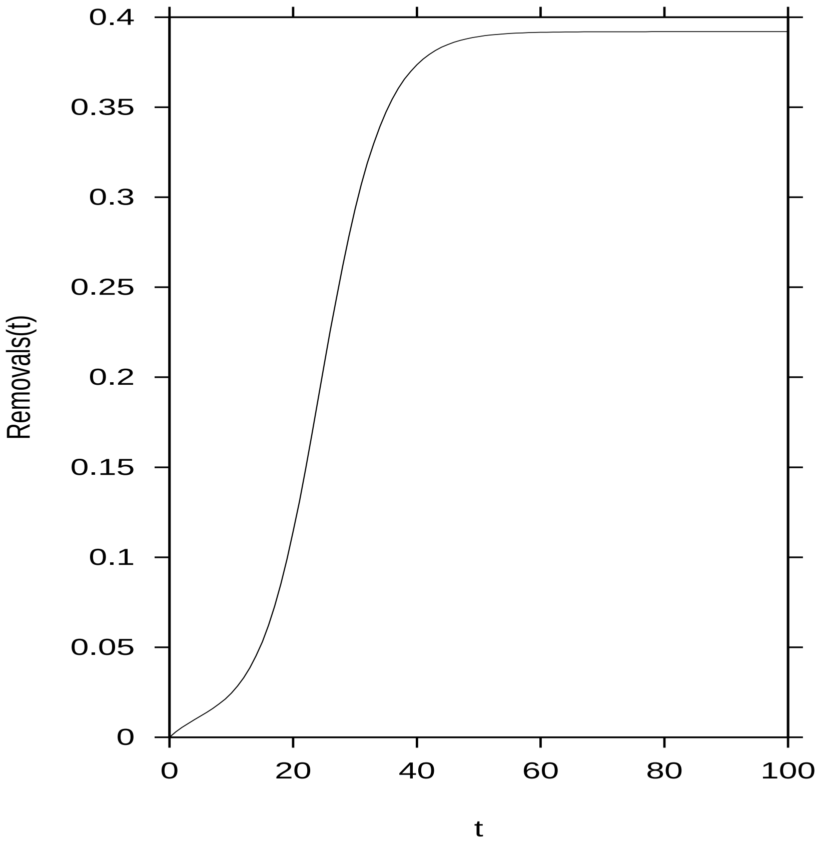

Figure 2.7 *Proportion removals in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 0.3$

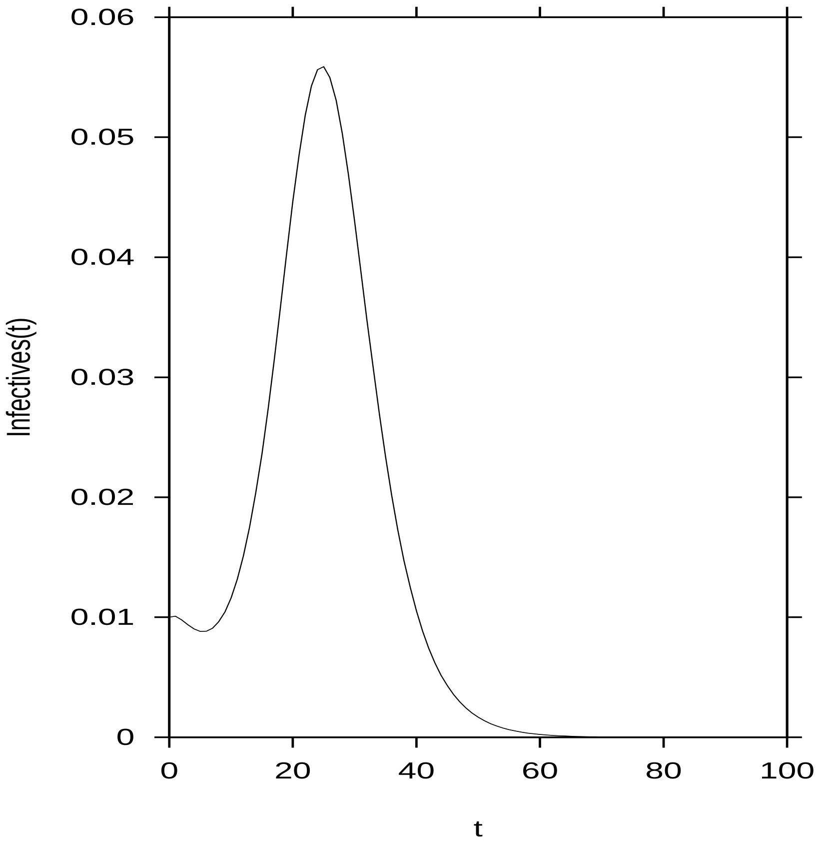

Figure 2.8 *Proportion infectives in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 0.3$

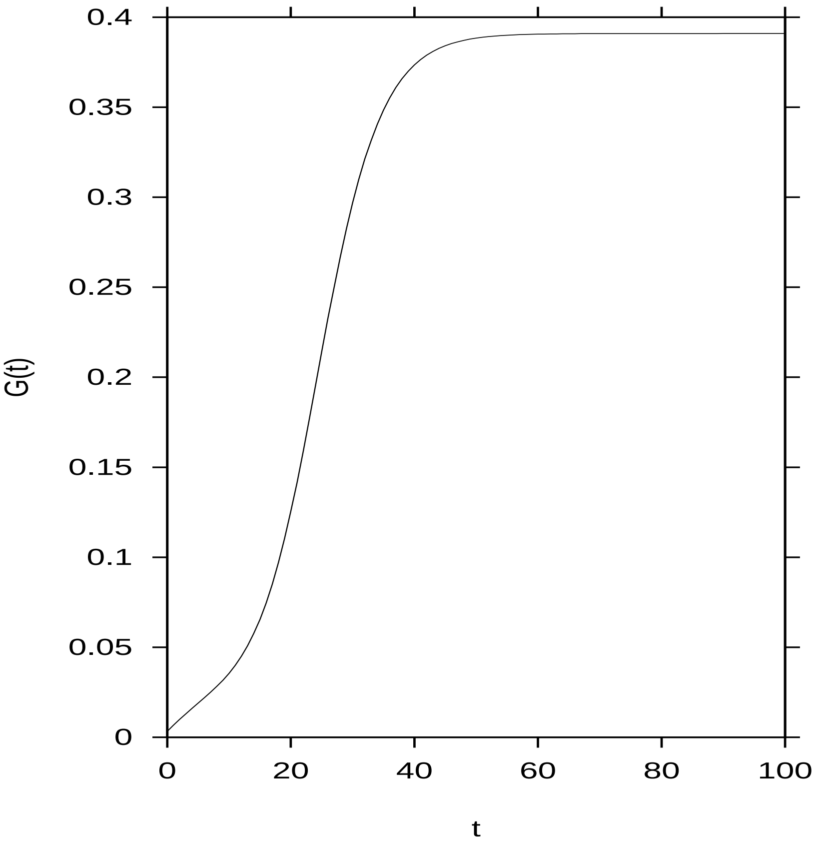

Figure 2.9 *Cumulative infectivity function G in Bartlett's GSE;* $a = 0.01$, $\alpha = 1$, *and* $\beta = 0.3$

CHAPTER 3

Discussion

In the above we have derived an explicit bound on the distance of the time course of the GSE over a finite time interval $[0, T]$ to its mean-field limit. This is the first explicit bound known for this problem. Of course normal approximations, as given by Barbour [6] provide the order of the distance, but only in the Markovian case, and do not give an explicit expression. The results above not only give an explicit bound, but also a numerically fast procedure (using a contraction construction) to derive the approximating deterministic system. Moreover, we confirmed a heuristic by Metz by showint that the deterministic approximation is linear in time if the approximation is started at a random time τ where the epidemic has taken off, and if the approximation is only derived on an interval where the epidemic grows strictly.

Note that considering only a finite time interval is not a strong restriction, as all our observations from an epidemic process will necessarily be over a finite time interval. For Bartlett's GSE, from Barbour [7] it is known that the duration of the epidemic is a nontrivial random quantity with center of order $\log K$, so that our bounds will be expected to be satisfactory for a relatively long time period. This fact also illustrates why the approximation breaks down when used for a very large time period; as the duration of the epidemic is truly random, no good deterministic approximation for it exists. Moreover, in Bartlett's GSE the parameter $\alpha\beta$ corresponds to the transmission rate. As a rule of thumb, the faster the transmission rate, the shorter the duration of the epidemic; in this sense $\alpha\beta$ scales with T.

The deterministic approximation agrees, for Bartlett's GSE, with the ones obtained in the literature. Isham [25] shows how to derive the approximation using moment closure methods (although her focus is on stochastic approximations). Moment closure methods have been proven a powerful approach for Markovian models, in particular for spatial models. However, I am not aware of any extension to non-Markovian settings. Furthermore, the moment closure approach has heuristic character, to determine the covariance, for example, when a Gaussian approximation is known to be valid. It does not give a bound on the distance to the approximation.

The model discussed here is a closed epidemic, without reinfection, and not admitting birth of individuals, nor removals not due to the infection. These assumptions exclude many biologically relevant cases. However, independent birth events and independent death events would only result in a slightly modified empirical measure, for which the above arguments, in particular the contraction argument, should still be valid. In this extension we would expect a recurrent epidemic, due to the possibility of births after the epidemic has taken off; some complex dynamics could then occur, see, e.g., Earn *et al.* [21] and Andersson and Britton [2]. Moreover it would be possible to treat a stratified population, with different types of individuals (the (l_i, r_i)'s may come from different distributions); for a motivation, see, e.g., May [32].

Also it would be desirable to generalize the above for accumulation functions λ that might not involve just the proportion of infectives over the time course, but also some neighborhood structure. This would allow to treat spatial epidemic models in a similar way.

The above bounds have only been derived for smooth test functions. Deriving bounds for nonsmooth test functions is not a fundamental problem but a rather tedious enterprise; see, for example, Rinott and Rotar [40], Götze [23], or Bolthausen [15]. Lastly, a Gaussian approximation together with error bounds would be a natural next step to investigate.

Acknowledgement. This work has been motivated by many fruitful discussions with Andrew Barbour. John Einmahl was very helpful in connection with the Glivenko-Cantelli bound; Guy Gielis and Nico Stollenwerk made many useful comments on the manuscript. Frank Ball drew my attention to Metz' work and intuition. Finally, I would like to thank the organizers for the invitation, and the referees for their comments, which yielded a substantial improvement of the paper.

References

Aldous, D. (1989). Stein's method in a two-dimensional coverage problem. *Statist. Probab. Lett.* **8**, 307–314.

Andersson, H. and Britton, T. (2000). Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. Preprint.

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17**, 9–25.

Bailey, N.T.J. (1975). *The mathematical theory of infectious diseases and its applications.*(2nd ed.) Griffin, London.

Baldi, P., Rinott, Y., and Stein, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics and Mathematics, Papers in Honor of Samuel Karlin.* T.W. Anderson, K.B. Athreya and D.L. Iglehart, eds., Academic Press, 59–81.

Barbour, A.D. (1974). On a functional central limit theorem for Markov population processes. *Adv. Appl. Probab.* **6**, 21–39.

Barbour, A.D. (1975). The duration of the closed stochastic epidemic. *Biometrika* **62**, 477–482.

Barbour, A.D. (1990). Stein's method for diffusion approximations. *Probability Theory and Related Fields* **84**, 297–322.

Barbour, A.D., Chen, L.H.Y., and Loh, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20**, 1843–1866.

Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford Science Publications.

Barbour, A.D., Karoński, M., and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Comb. Theory*, Ser. B, **47**, 125–145.

Barbour, A.D., and Utev, S. (1998). Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Prob.* **30**, 449–475.

Bartlett, M.S. (1949) Some evolutionary stochastic processes. *J. R. Statist. Soc., Ser. B* **11**, 211–229.

Berard, D.R., Burgess, J.W. and Coss, R.G. (1981). Plasticity of dendritic spine formation: a state-dependent stochastic process. *Intern. J. Neuroscience* **13**, 93–98.

Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **66**, 379–386.

Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

Dawson, D.A. (1993). Measure-valued Markov processes. In *Ecole d'Eté de*

Probabilité de Saint-Flour 1991, LNM 1541, Hennequin, P.L., eds., Springer-Verlag, 1–260.

Diaconis, P. (1989). An example for Stein's method. Stanford Stat. Dept. Tech. Rep.

Diaconis, P., and Zabell, S. (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statistical Science* **6**, 284-302.

Ethier, S.N. and Kurtz, T.G. (1986). *Markov processes: characterization and convergence.* Wiley, New York.

Earn, D., Rohani, P., Bolker, B.M., and Grenfell, B. (2000). A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670.

Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11**, 7–16.

Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.

Hall, P., and Barbour, A.D. (1984). Reversing the Berry-Esséen theorem. *Proc. AMS* **90**, 107–110.

Isham, V. (1991). Assessing the variability of stochastic epidemics. *Math. Biosci.* **107**, 209–224.

Keeling, M., and Grenfell, B.Y. (1998). Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Math. Biosciences* **147**, 207–226.

Loh, W.-L. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2**, 536–554.

Loh, W.-L. (1997). Stein's method and Pearson Type 4 approximations. Research report.

Luk, H.M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis. University of Southern California, Los Angeles.

Mann, B. (1997). Stein's method for χ^2 of a multinomial. Preprint.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18**, 1269–1283.

May, R. (2000). Simple rules with complex dynamics. *Science* **287**, 601–602.

Metz, J.A.J. (1978). The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheoretica.* **27**, 75–123.

Peköz, E. (1995). Stein's method for geometric approximation. Tech. Report #225, Stat. Dept. U. California. Berkeley.

Picard, P. and Lefevre, C. (1991). The dimension of Reed-Frost epidemic models with randomized suscpetibility levels. *Math. Biosci.* **107**, 225–233.

Reinert, G. (1994). A weak law of large numbers for empirical measures via Stein's method. *Ann. Probab.* **23**, 334–354.

Reinert, G. (1995) The asymptotic evolution of the General Stochastic Epidemic. *Ann. Appl. Probab.* **5**, 1061–1086.

Reinert, G. (1998). Couplings for normal approximations with Stein's method. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **41**, 193–207.

Reinert, G. (1998). Stein's method in application to empirical measures. In

Modelos Estocasticos. J.M. Gonzales Barrios and L.G.Gorostiza, eds. Sociedad Matematica Mexicana, 65–120.

RINOTT, Y., AND ROTAR, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Analysis.* **56**, 333–350.

ROOS, M. (1993). *Stein-Chen method for compound Poisson approximation.* Ph.D. thesis, Zürich.

SELLKE, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Probab.* **20**, 390–394.

SOLOMON. W. (1987). Representation and approximation of large population age distributions using Poisson random measures. *Stoch. Proc. Appl.* **26**, 237–255.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583–602. Univ. California Press, Berkeley.

STEIN, C. (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. **7**, Hayward, California.

SWINTON, J., HARWOOD, J., GRENFELL, B.T., AND GILLIGAN, C. (1998). Persistence thresholds for phocine distember virus infection in harbour seal *Phoca vitulina* metapopulations. *Journ. Animal Ecology* **67**, 54–68.

WANG, F.S.J. (1975) Limit theorems for age and density dependent stochastic population models. *J. Math. Bio.* **2**, 373–400

WANG, F.S.J. (1977) A central limit theorem for age-and-density-dependent population processes. *Stoch. Proc. Appl.* **5**, 173–193.

Index

Section 0

Section 1

Section 2

Section 3