

Tom Britton
Etienne Pardoux
(Eds.)

Stochastic Epidemic Models with Inference

With Contributions by:

Frank Ball, Tom Britton, Catherine Larédo, Etienne Pardoux,
David Sirl and Viet Chi Tran

Contents

Part I Stochastic epidemics in a homogeneous community Tom Britton and Etienne Pardoux

Introduction	3
1 Stochastic Epidemic Models	5
1.1 The stochastic SEIR epidemic model in a closed homogeneous community	5
1.1.1 Model definition	5
1.1.2 Some remarks, submodels and model generalizations	6
1.1.3 Two key quantities: R_0 and the escape probability	8
1.2 The early stage of an outbreak	9
1.3 The final size of the epidemic in case of no major outbreak	14
1.4 Vaccination	17
2 Markov Models	21
2.1 The deterministic SEIR epidemic model	21
2.2 Law of Large Numbers	25
2.3 Central Limit Theorem	34
2.4 Diffusion Approximation	40
3 General Closed Models	43
3.1 Exact results for the final size in small communities	43
3.2 The Sellke construction	47
3.3 LLN and CLT for the final size of the epidemic	49
3.3.1 Law of Large Numbers	50
3.3.2 Central Limit Theorem	53
3.4 The duration of the stochastic SEIR epidemic	55
4 Open Markov Models	59
4.1 Open populations: time to extinction and critical population size	59
4.2 Large deviations and extinction of an endemic disease	63

4.2.1	Introduction	63
4.2.2	The rate function	64
4.2.3	The lower bound	68
4.2.4	The upper bound	75
4.2.5	Time of extinction in the SIRS model	82
4.2.6	Time of extinction in the SIS model	92
4.2.7	Time of extinction in the SIR model with demography ...	94
Appendix	97
A.1	Branching processes	97
A.1.1	Discrete time branching processes	97
A.1.2	Continuous time branching processes	99
A.2	The Poisson process and Poisson point process	102
A.3	Cramér's theorem for Poisson random variables	105
A.4	Martingales	107
A.4.1	Martingales in discrete time	107
A.4.2	Martingales in continuous time	110
A.5	Tightness and weak convergence in path space	111
A.6	Pontryagin's maximum principle	112
A.7	Semi- and equicontinuity	114
A.8	Solutions to selected exercises	114
References for Part I	119

Part I
**Stochastic epidemics in a homogeneous
community**

Tom Britton and Etienne Pardoux

Tom Britton
e-mail: `tom.britton@math.su.se`
Etienne Pardoux
e-mail: `etienne.pardoux@univ-amu.fr`

Introduction

In this Part I of the lecture notes our focus lies exclusively on stochastic epidemic models for a homogeneously mixing community of individuals being of the same type. The important extensions allowing for different types of individuals and allowing for non-uniform mixing behaviour in the community is left for later parts in the Notes.

In Chapter 1, we present the stochastic SEIR epidemic model, derive some important properties of it, in particular for the beginning of an outbreak. Motivated by mathematical tractability rather than realism we then study in Chapter 2 the special situation where the model is Markovian, and derive additional results for this sub-model.

What happens later on in the outbreak will depend on our model assumptions, which in turn depend on the scientific questions. In Chapter 3 we focus on short-term outbreaks, when it can be assumed that the community is fixed and constant during the outbreak; we call these models closed models. In Chapter 4 we are more interested in long-term behaviour, and then it is necessary to allow for influx of new individuals and that people die, or to include return to susceptibility. Such so-called open population models are harder to analyse – for this reason we stick to the simpler class of Markovian models. In this chapter we consider situations where the deterministic model has a unique stable equilibrium, and use both the central limit theorem and large deviation techniques to predict the time at which the disease goes extinct in the population.

The Notes end with an extensive Appendix, giving some relevant probability theory used in the main part of the Notes and also solutions to most of the exercises being scattered out in the different chapters.

Chapter 1

Stochastic Epidemic Models

This first chapter introduces some basic facts about stochastic epidemic models. We consider the case of a closed community, i.e. without influx of new susceptibles or mortality. In particular, we assume that the size of the population is fixed, and that the individuals who recover from the illness are immune and do not become susceptible again. We describe the general class of stochastic epidemic models, and define the basic reproduction number, which allows one to determine whether or not a major epidemic may start from the initial infection of a small number of individuals. We then approximate the early stage of an outbreak with the help of a branching process, and from this obtain the distribution of the final size (i.e. the total number of individuals who ever get infected) in case of a minor outbreak. Finally we discuss the impact of vaccination.

The important problem of estimating model parameters from (various types of) data is left to Part IV of the current volume (also discussed in Chapter 4 of Part III). Here we assume the model parameters to be known.

1.1 The stochastic SEIR epidemic model in a closed homogeneous community

1.1.1 Model definition

Consider a closed population of $N + 1$ individuals (N is the number of initially susceptible). At any point in time each individual is either susceptible, exposed, infectious or recovered. Let $S(t)$, $E(t)$, $I(t)$ and $R(t)$ denote the numbers of individuals in the different states at time t (so $S(t) + E(t) + I(t) + R(t) = N + 1$ for all t). The epidemic starts at $t = 0$ in a specified state, often the state with one infectious individual, called the index case and thought of as being externally infected, and the rest being susceptible: $(S(0), E(0), I(0), R(0)) = (N, 0, 1, 0)$.

Definition 1.1.1. While infectious, an individual has infectious contacts according to a Poisson process with rate λ . Each contact is with an individual chosen uniformly at random from the rest of the population, and if the contacted individual is susceptible he/she becomes infected – otherwise the infectious contact has no effect. Individuals that become infected are first latent (called exposed) for a random duration L with distribution F_L , then they become infectious for a duration I with distribution F_I , after which they become recovered and immune for the remaining time. All Poisson processes, uniform contact choices, latent periods and infectious periods of all individuals are defined to be mutually independent.

The epidemic goes on until the first time τ when there are no exposed or infectious individuals, $E(\tau) + I(\tau) = 0$. At this time no further individuals can get infected so the epidemic stops. The final state hence consists of susceptible and recovered individuals, and we let Z denote the *final size*, i.e. the number of infected (by then recovered) individuals at the end of the epidemic excluding the index case(s): $Z = R(\tau) - I(0) = N - S(\tau)$. The possible values of Z are hence $0, \dots, N$.

1.1.2 Some remarks, submodels and model generalizations

Quite often the rate of “infectious contacts” λ can be thought of as a product of a rate c at which the infectious individual has contact with others, and the probability p that such a contact results in infection given that the other person is susceptible, so $\lambda = cp$. As regards to the propagation of the disease it is however only the product λ that matters and since fewer parameters is preferable we keep only λ .

The rate of infectious contacts is λ , so the rate at which one infectious has contact with a specific other individual is λ/N since each contact is with a uniformly chosen other individual.

First we will look what happens in a very small community/group, but the main focus of these notes is for a large community, and the asymptotics are hence for $N \rightarrow \infty$. The parameters of the model, the infection rate λ , and the latent and infectious periods L and I , are defined independently of N , but the epidemic is highly dependent on N so when this needs to be emphasized we equip the corresponding notation with an N -index, e.g. $S^N(t)$ and τ^N which hence is not a power.

Some special cases of the model have received special attention in the literature. If both L and I are exponentially distributed (with rates ν and γ say), the model is Markovian which simplifies the mathematical analysis a great deal. This model is called the *Markovian SEIR*. If $L \equiv 0$ and $I \sim \text{Exp}(\gamma)$ then we have the *Markovian SIR* (whenever there is no latency period the model is said to be SIR) which is better known under the unfortunate name the *General stochastic epidemic*. Another special case of the stochastic SEIR model is where the infectious period I is non-random. Also here there is a underlying mathematical reason – when the duration of the infectious period is non-random and equal to t say, then an infectious individual has infectious contacts with each other individual at rate λ/N during a non-random

time implying that the number of contacts with different individuals are *independent*. Consequently, an infectious individual has infectious contacts with each other individual independently with probability $p = 1 - e^{-(\lambda/N)t}$, so the total number of contacts is Binomially distributed, and in the limit as $N \rightarrow \infty$ the number of infectious contacts an individual has is Poisson distributed with mean λt . If further the latent period is long in comparison to the infectious period then it is possible to identify the infected individuals in terms of *generations*: the first generation are the index cases, the second generation those who were infected by the index case(s), and so on. When the model is described in this discrete time setting and individuals infect different individuals independently with probability p , this model is the well-known Reed–Frost model named after its inventors Reed and Frost.

The two most studied special cases are hence when the infectious period is exponentially distributed and when it is nonrandom. For real infectious diseases none of these two extremes apply, for influenza for example, the infectious period is believed to be about 4 days, plus or minus one or two days. If one has to choose between these choices a nonrandom infectious period is probably closer to reality.

The stochastic SEIR model in a closed homogeneous community may of course also be generalized towards more realism. Two such extensions have already been mentioned: allowing for individuals to die and new ones to be born, and allowing for some social structures. Some such extensions will be treated in the other articles of the current lecture notes but not here. But even when assuming a closed homogeneously mixing community of homogeneous individuals it is possible to make the model more realistic. The most important such generalization is to let the rate of infectious contact vary with time since infection. The current model assumes there are no infectious contacts during the latent state, and then, suddenly when the latent period ends, the rate of infectious contact becomes λ until the infectious period ends when it suddenly drops down to 0 again. In reality, the infectious rate is usually a function $\lambda(s)$ s time units after infection. In most situations $\lambda(s)$ is very small initially (corresponding to the latency period) followed by a gradual increase for some days, and then $\lambda(s)$ starts decaying down towards 0 which it hits when the individual has recovered completely (see Figure 1.1.1 for an example where infectivity starts growing after one day and is more or less over after one week). The function $\lambda(s)$ could be the same for all individuals, or it may be random and hence a stochastic process, i.i.d. for different individuals. As regards to the temporal dynamics of the epidemic process, the functional form of $\lambda(s)$ is important, and also its random properties in case it is random. If one is only interested in the final size τ , it is however possible to show that all that affects the final size is the accumulated force of infection, i.e. the distribution of $\int_0^\infty \lambda(s) ds$. In particular, if we let λI in the stochastic SEIR model have the same distribution as $\int_0^\infty \lambda(s) ds$ in the more general model, then the two models have the same final size distribution. In that sense, the extended model can be included in the stochastic SEIR model.

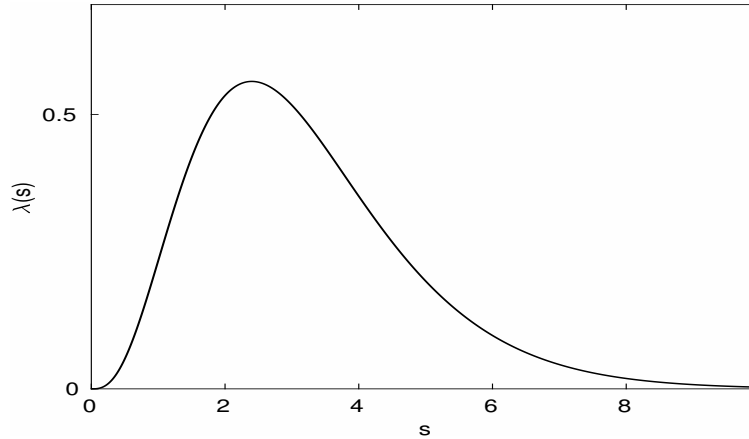


Fig. 1.1.1 Plot of a possible infectivity curve $\lambda(s)$. The time s denotes the time since infection in unit of days.

1.1.3 Two key quantities: R_0 and the escape probability

The most important quantity for this, as well as most other epidemic models, is the *basic reproduction number* (sometimes “number” is replaced by “ratio”) and denoted by R_0 . In more complicated models its definition and interpretation are sometimes debated, but for the present model it is quite straightforward: R_0 denotes the mean number of infectious contacts a typical infected has during the early stage of an outbreak. As the population under consideration becomes large, this number will coincide with the mean number of infections caused by a typical infected during the early stages of an outbreak. We derive an expression for R_0 , but before that we should consider its important threshold value of 1. If $R_0 > 1$ this means that on average an infected infects more than one individual in the beginning of an epidemic. Then the index case on average is replaced by more than one infected, who in turn each are replaced by more than one infected and so on. This clearly suggests that a big community fraction can become infected. If on the other hand $R_0 \leq 1$, then the same reasoning suggests that there will never be a big community outbreak. Those results hold true which we prove in Section 1.2 (Corollaries 1.2.6 and 1.2.7).

In applications the basic reproduction number R_0 is a central quantity of interest. Many studies of disease outbreaks contain estimates of R_0 for a specific disease and community, together with modeling conclusions about preventive measures which, if put into place, will reduce the reproduction number R down to below the critical value of 1 when an outbreak is no longer possible (e.g. Fraser et al. [12]).

Let us now derive an expression for R_0 . An infected individual has infectious contacts only when infectious, and when in this state the individual has infectious contacts at rate λ . This means that the expected number of infectious contacts equals

$$R_0 = \mathbb{E}(\lambda I) = \lambda \iota. \quad (1.1.1)$$

Sometimes the rate λ of having infectious contacts is replaced by an over-all rate of contact c multiplied by the probability p of a contact leading to infection, so $\lambda = c p$ and $R_0 = c p \iota$ (cf. the first lines of the above Subsection 1.1.2, Anderson and May [1] and Giesecke [14]).

Another key quantity appearing later several times is the probability for a given susceptible to escape getting infected from a specific infective. The instantaneous infectious force from the infective to this specific susceptible is λ/N , and the random duration of the infectious period is I . Conditional upon $I = x$, the escape probability is hence $e^{-(\lambda/N)x}$, and the unconditional probability to escape infection is therefore

$$\mathbb{P}(\text{escape infection from an infective}) = \mathbb{E}(e^{-\lambda I/N}) = \psi_I(-\lambda/N), \quad (1.1.2)$$

where $\psi_I(b) = \mathbb{E}(e^{bI})$ is the moment generating function of the infectious period (so $\psi(-b)$ is the Laplace transform – in Part II in this volume the Laplace transform has a separate notation, ϕ , so $\phi(b) = \psi(-b)$).

Exercise 1.1.2. Consider the Markovian SEIR epidemic in which $\lambda = 1.8$, $\nu = 2$ and $\gamma = 1$ in a village of size $N = 100$, (parameters inspired by Ebola with weeks as time unit). Compute R_0 and the escape probability.

Exercise 1.1.3. Repeat the previous exercise, but now for the Reed–Frost epidemic with $\lambda = 1.8$, $L \equiv 2$ and $I \equiv \iota = 1$ in a village of size $N = 100$, (perhaps having more realistic distributions than in the previous exercise).

1.2 The early stage of an outbreak

We now consider the situation where the community size N is large and study the stochastic SEIR epidemic in the beginning of an outbreak. By “beginning” we mean that less than $k = k(N)$ individuals have been infected. Recall from the model definition that infectious individuals have infectious contacts with others independently, each infective at rate λ . The dependence only appears because individuals can only get infected once, so if an individual has already received an infectious contact, then future infectious contacts with that individual no longer result in someone getting infected. However, in the beginning of an outbreak in a large community it is very unlikely that two infectives happen to have infectious contacts with the same individual. This suggests that during the early phase of an outbreak, infectives infect new individuals more or less independently. This implies that the number of infected can be approximated by a branching process in the beginning of an outbreak, where “being born” corresponds to having been infected, and “giving birth” corresponds to infecting someone. The current section is devoted to making this approximation rigorous, and thus obtaining asymptotic results for the epidemic in regards to having a minor versus a major outbreak. In the next section this approximation is exploited

in order to determine the distribution of the final size in the case of a minor outbreak. If the epidemic takes off, which happens in the case of a major outbreak, then the approximation that individuals infect others independently breaks down. What happens in this situation is treated in later sections.

First we define the approximating branching process and derive some properties of it. After this we show rigorously that, as $N \rightarrow \infty$, the initial phase of the epidemic process converges to the initial phase of the branching process by using an elegant coupling technique.

The approximating branching process is defined similarly to the epidemic. A newborn individual is first unable to give birth to new individuals for a period with duration L (this period might be denoted childhood in the branching process setting). After this childhood, the individual enters the reproductive stage which last for I units of time. During this period individuals give birth to new individuals at rate λ (randomly in time according to a Poisson process with rate λ). Once the reproductive stage has terminated the individual dies (or at least cannot reproduce and hence plays no further role).

The number of offspring of an individual, X , depends on the duration of the reproductive stage I . Conditional upon $I = y$, the number of births follow the Poisson distribution $\text{Poi}(\lambda y)$, so the unconditional distribution of number of offspring is mixed-Poisson, written as $X \sim \text{MixPoi}(\lambda I)$, where I has distribution F_I .

If we forget calendar time, and simply study the number of individuals born in each generation, then our branching process is a Bienaymé–Galton–Watson process with offspring distribution being $\text{MixPoi}(\lambda I)$. The mean number of children/offspring equals $m = \mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|I)) = \mathbb{E}(\lambda I) = \lambda \iota$.

Exercise 1.2.1. Compute the offspring distribution $\mathbb{P}(X = x)$ explicitly for the two cases: (i) where the infectious period is non-random, $I \equiv \iota$, corresponding to the continuous–time version of the Reed–Frost epidemic; and (ii) for the Markovian SEIR where I is exponential with mean ι .

We now show an elegant coupling construction which we will use to show that the epidemic and branching process have similar distributions in the beginning. To this end we define the approximating branching process as well as all epidemics, i.e. for each $N = 1, 2, \dots$, on the same probability space. To this end, let L_0, L_1, \dots be i.i.d. latent periods having distribution F_L , and similarly let I_0, I_1, \dots be i.i.d. infectious periods having distribution F_I . Further, let $\xi_0(\cdot), \xi_1(\cdot), \dots$ be i.i.d. Poisson processes having intensity λ , and let U_1, U_2, \dots be i.i.d. $U(0, 1)$ random variables. All random variables and Poisson processes are assumed to be mutually independent. These will be used to construct the branching process as well as the stochastic SEIR epidemic for each N as follows.

Definition 1.2.2. The approximating branching process. At time $t = 0$ there is one new born ancestor having label 0. Let the ancestor have childhood length L_0 and reproductive stage for a duration I_0 (so the ancestor dies at time $L_0 + I_0$), during which the ancestor gives birth at the time points of the Poisson process $\xi_0(\cdot)$. If the jump times of this Poisson process are denoted $T_{0,1} < T_{0,2} < \dots$ and X_0 denotes

the number of jumps prior to I_0 , then the ancestor gives birth at the time points $L_0 + T_{0,1}, \dots, L_0 + T_{0,X_0}$ (the set is empty if $X_0 = 0$). The first born individual is given label 1, and having childhood period L_1 , reproductive period I_1 and birth process $\xi_1(\cdot)$. This individual gives birth according to the same rules (starting the latency period at time $L_0 + T_{0,1}$), and the next individual born, either to individual 0 or 1, is given label 2 and variables L_2, I_2 and birth process $\xi_2(\cdot)$, and so on. This defines the branching process, and we let $L(t), I(t), R(t)$ respectively denote the numbers of individuals in the childhood state, in the reproductive state and dead, respectively, at time t . The total number of individuals born up to time t , excluding the ancestor/index case, is denoted by $Z(t) = L(t) + I(t) + R(t) - 1$ in the branching process, and the ultimate number ever born, excluding the ancestor, is denoted by Z which may be finite or infinite.

We now define the epidemic for any fixed N (in the epidemic childhood corresponds to latent and reproductive stage to being infectious). This is done similarly to the branching process with the exception that we now keep track of which individuals who get infected using the uniform random variables U_1, U_2, \dots .

Definition 1.2.3. The stochastic SEIR epidemic with N initial susceptibles. We label the $N + 1$ individuals $0, 1, \dots, N$, with the index case having label 0 and the others being labelled arbitrarily. As for the branching process, the index case is given latency period L_0 , infectious period I_0 and contact process $\xi_0(\cdot)$ and the epidemic is started at time $t = 0$. The infectious contacts of the index case occur at the time points $L_0 + T_{0,1}, \dots, L_0 + T_{0,X_0}$. The first infectious contact is with individual $[U_1 N] + 1$, the integer part of NU_1 plus 1 (this picks an individual uniformly among $1, \dots, N$). This individual, k say, then becomes infected (and latent) and is given latent period, infectious period and contact process L_1, I_1 and $\xi_1(\cdot)$. The next infectious contact (from either the index case or individual k) will be with individual $[U_2 N] + 1$. If the contacted person is individual k then nothing happens, but otherwise this new individual gets infected (and latent), and so on. Infectious contacts only result in infection if the contacted individual is still susceptible. When a contact is with an already infected individual the branching process has a birth whereas there is no infection in the epidemic – we say a “ghost” was infected when comparing with the branching process. Descendants of all ghosts are also ignored in the epidemic. The epidemic goes on until there are no latent or infectious individuals. This will happen within a finite time (bounded by $\sum_{j=0}^N (L_j + I_j)$). The final number of infected individuals excluding the index case is as before denoted $Z^N \in [0, \dots, N]$. Similar to before we let $L^N(t), I^N(t), R^N(t)$ denote the numbers of latent, infectious and recovered individuals at time t , and now we can also define the number of susceptibles $S^N(t) = N + 1 - L^N(t) - I^N(t) - R^N(t)$.

In our model the index case cannot be contacted. This is of course unrealistic but simplifies notation. In the limit as N gets large this assumption has no effect. We now state two important results for these constructions of the branching process and epidemics.

Theorem 1.2.4. *The definition above agrees with the earlier definition of the Stochastic SEIR epidemic in a homogeneous community.*

Proof. The latent and infectious periods have the desired distributions, and an infective has infectious contacts with others at overall rate λ , and each time such a contact is with a uniformly selected individual as desired. \square

We now prove that the branching process and the epidemic process (with population size N) are identical up to a time point which tends to infinity in probability as $N \rightarrow \infty$. To this end, we let M^N denote the number of infections prior to the first ghost (i.e. how many uniformly selected individuals $[U_k N]$ there were before someone was reselected. If this never happens we set $M^N = \infty$. Let T^N denote the time at which the first ghost appears (and if this never happens we also set $T^N = \infty$).

Theorem 1.2.5. *The branching process and N -epidemic agree up until T^N : $(L^N(t), I^N(t), R^N(t)) = (L(t), I(t), R(t))$ for all $t \in [0, T^N)$. Secondly, $T_N \rightarrow \infty$ and $M^N \rightarrow \infty$ in probability as $N \rightarrow \infty$.*

Proof. The first statement of the proof is obvious. The only difference between the epidemic and the branching process in our construction is that specific individuals are contacted in the epidemic, and up until the first time when some individual is contacted again, each infectious contact results in infection just as in the branching process.

As for the second part of the theorem we first compute the probability that M^N will tend to infinity, and then that the time T^N until the first ghost appears also tends to infinity. It is easy to compute $\mathbb{P}(M^N > k)$ since this will happen if and only if all the first k contacts are with distinct individuals:

$$\mathbb{P}(M^N > k) = 1 \times \frac{N-1}{N} \times \cdots \times \frac{N-k}{N} = \prod_{j=0}^{k-1} \left(1 - \frac{j}{N}\right).$$

(This formula is identical to the celebrated (...) birthday problem if $N+1 = 365$ and k is the size of the class.) For fixed k we see that this probability tends to 1 as $N \rightarrow \infty$. We can in fact say more. We have the following lower bound (which is easily proved by recurrence):

$$\mathbb{P}(M^N > k) = \prod_{j=0}^{k-1} \left(1 - \frac{j}{N}\right) \geq 1 - \sum_{j=1}^k \frac{j}{N} = 1 - \frac{(k+1)k}{2N}.$$

As a consequence, we see that $\mathbb{P}(M^N > k(N)) \rightarrow 1$ as long as $k = k(N) = o(\sqrt{N})$. In particular $M^N \rightarrow \infty$ in probability as $N \rightarrow \infty$. In what follows we write w.l.p. for “with large probability”, meaning with a probability tending to 1 as $N \rightarrow \infty$. The consequence hence implies that all infectious contacts up to $k(N)$ will w.l.p. be with distinct individuals and thus will result in infections. So, up until $k(N)$ individuals have been infected, the epidemic can be approximated by a branching process for any $k(N) = o(\sqrt{N})$. Let $Z(t)$ denote the number of individuals born before t in the branching process (excluding the ancestor) and $Z^N(t) = N - S^N(t)$ the

number of individuals that have been infected before t (excluding the index case) in the N -epidemic. Since the epidemic and branching process agree up until T^N it follows that $Z(t) = Z^N(t)$ for $t < T^N$. But, since $k(N) < M^N$ w.l.p. it follows that $\inf\{t; Z(t) = k(N)\} \leq T^N$ w.l.p. If the branching process is (sub)critical, then $Z(t)$ remains bounded as $t \rightarrow \infty$, so $T^N = +\infty$ w.l.p. Consider now the supercritical case. From Section A.1.2 (Proposition A.1.4) we know that $Z(t) = O_p(e^{rt})$ where the Malthusian parameter r solves the equation

$$\int_0^\infty e^{-rs} \lambda(s) ds = 1. \quad (1.2.1)$$

The function $\lambda(s)$ is the rate at which an individual gives birth s time units after being born, so $\lambda(s) = \lambda \mathbb{P}(\text{infectious at } s)$ and hence $\lambda(s) = \lambda \mathbb{P}(L < s < L + I)$ for our model. We thus have that $k(N) \leq ce^{rT^N}$ w.l.p., which implies that $T^N \geq \log k(N)/r - \log c$. So if for example $k(N) = N^{1/3}$, which clearly satisfies $k(N) = o(\sqrt{N})$, it follows that $T^N \rightarrow \infty$ in probability. \square

Theorem 1.2.5 shows that the epidemic behaves like the branching process up to a time point tending to infinity as $N \rightarrow \infty$, and that the number of infections/births by then also tends to infinity. This implies that we can use theory for branching processes to obtain results for the early part of the epidemic. We state these important results in the following corollaries; the first corollary is for the subcritical and critical cases and the second corollary is for the supercritical case. Recall that $R_0 = \lambda \mathbb{E}(I)$, the basic reproduction number in the epidemic and the mean offspring number in the branching process.

Corollary 1.2.6. *If $R_0 \leq 1$, then $(L^N(t), I^N(t), R^N(t)) = (L(t), I(t), R(t))$ for all $t \in [0, \infty)$ w.l.p. As a consequence, $\mathbb{P}(Z^N = k) \rightarrow \mathbb{P}(Z = k)$ as $N \rightarrow \infty$, and in particular Z^N is bounded in probability.*

Proof. In Theorem 1.2.5 it was shown that the epidemic and branching process agree up until there has been M^N births, where $M^N > N^{1/3}$ w.l.p. for example. But from branching process theory (Proposition A.1.1) we know that this will happen with a probability tending to 0 with N when $R_0 \leq 1$, implying that $T^N = \infty$ w.l.p. \square

Corollary 1.2.7. *If $R_0 > 1$, then for finite k : $\mathbb{P}(Z^N = k) \rightarrow \mathbb{P}(Z = k)$ as $N \rightarrow \infty$. Further, $\{Z^N \rightarrow \infty\}$ with the same probability as $\{Z = \infty\}$, which is the complement to the extinction probability, the latter being the smallest solution to the equation $z = g(z)$ described in Proposition A.1.1.*

Proof. Also this corollary is a direct consequence of Theorem 1.2.5 and properties of branching processes. If only k births occur, then there will be no ghost w.l.p., implying that the epidemic and the branching process agree forever w.l.p. On the other hand, the coupling construction showed that $M^N \rightarrow \infty$ on the other part of the sample space, and $Z \geq Z^N \geq M^N$ which completes the proof. \square

The two corollaries state that the epidemic and branching process coincide forever as long as the branching process stays finite. If the branching process grows

beyond all limits (only possible when $R_0 > 1$) then the epidemic and branching process will not remain identical even though also the epidemic tends to infinity with N . For any fixed N we have $0 \leq Z^N \leq N$ which clearly is different from $Z = \infty$ in that case. The distribution of Z^N on the part of the sample space where $Z^N \rightarrow \infty$ is treated below in Section 3.3.

The two corollaries show that the final number infected Z^N will be small with a probability equal to the extinction probability of the approximating branching process, and it will tend to infinity with the remaining (explosion) probability. In Section 3.3 we study the distribution of Z^N (properly normed) and then see that the distribution is clearly bimodal with one part close to 0 and the other part being $O(N)$. These two parts are referred to as *minor outbreak* and *major outbreak* respectively.

What happens during the early stage of an outbreak is particularly important when considering so-called *emerging epidemic outbreaks*. Then statistical inference based on this type of branching process approximation is often used. For example, in [38] a branching process approximation that is very similar to the SEIR branching process of Definition 1.2.2 is used for modelling the spread of Ebola during the early stage of the outbreak in West Africa in 2014.

Exercise 1.2.8. Use the branching process approximation of the current section to compute the probability of a major outbreak of the SEIR epidemic assuming that $I \equiv \iota$ (the continuous time Reed–Frost case), and $I \sim \text{Exp}(\gamma)$ (the Markovian SIR) with $\gamma = 1/\iota$. Only one of them will be explicit. Compute things numerically for $R_0 = 1.5$ and $\iota = 1$.

Exercise 1.2.9. Use the branching process approximation of the current section to compute the exponential growth rate r for the following two cases: $L \equiv 0$ and $I \equiv \iota$ (the continuous time Reed–Frost), and $L \equiv 0$ and $I \sim \text{Exp}(\gamma = 1/\iota)$ (the Markovian SIR). Compute r numerically for the two cases when $R_0 = 1.5$ and $\iota = \gamma = 1$.

1.3 The final size of the epidemic in case of no major outbreak

Let Z^N denote the final size of the epidemic (i.e. the total number of individuals that get infected during the outbreak) but now also including the initially infected individual. In the case of no major outbreak, if the total population size N is large enough, Z^N is well approximated by the total number of individuals in a branching process, as we saw in the previous section. Hence we consider Z as the total number of individuals ever born in a branching process (including the ancestor), where the number of offspring of the k -th individual is X_k . Let X_1, X_2, \dots be i.i.d. \mathbb{N} -valued random variables. We start by establishing an identity which is an instance of Kemperman’s formula, see e.g. Pitman [28] page 123.

Proposition 1.3.1. For all $k \geq 1$,

$$\mathbb{P}(Z = k) = \frac{1}{k} \mathbb{P}(X_1 + X_2 + \dots + X_k = k - 1).$$

Proof. Consider the process of depth-first search of the genealogical tree of the infected individuals. This procedure can be defined as follows. The tree is explored starting from the root. Suppose we have visited k vertices. The next visit will be to the leftmost still unexplored son of this individual, if any; otherwise to the leftmost unexplored son of the most recently visited node among those having not yet visited son(s), see Figure 1.3.1. X_1 is the number of sons of the root, who is the first visited individual. X_k is the number of sons of the k -th visited individual. This exploration of the tree ends at step k if and only if $X_1 \geq 1$, $X_1 + X_2 \geq 2$, $X_1 + X_2 + X_3 \geq 3$, ... $X_1 + X_2 + \dots + X_{k-1} \geq k-1$, and $X_1 + X_2 + \dots + X_k = k-1$. Let us rewrite those conditions. Define

$$\begin{aligned} Y_i &= X_i - 1, \quad i \geq 1, \\ S_k &= Y_1 + Y_2 + \dots + Y_k. \end{aligned}$$

A trajectory $\{Y_i, 1 \leq i \leq k\}$ explores a tree of size k if and only if the following conditions are satisfied

$$(C_k) \quad S_0 = 0, S_1 \geq 0, S_2 \geq 0, \dots, S_{k-1} \geq 0, S_k = -1.$$

Indeed, it is easy to convince oneself that it is the case if there is only one generation: if the ancestor has $k-1$ children, then $Y_1 = k-2$, and $Y_2 = \dots = Y_k = -1$, hence (C_k) holds. If one attaches one generation trees to some of the leaves of the previous tree, then one replaces a unique -1 step by an excursion upwards which finishes at the same level as the replaced step. Iterating this procedure, we see that the exploration of a general tree with k nodes satisfies (C_k) .

The statement of the proposition is equivalent to

$$\mathbb{P}(Z = k) = \frac{1}{k} \mathbb{P}(Y_1 + Y_2 + \dots + Y_k = -1).$$

Denote by V_k the set of sequences of k integers ≥ -1 which satisfy conditions (C_k) , and U_k the set of sequences of k integers ≥ -1 which satisfy the unique condition $S_k = -1$. We use circular permutations operating on the Y_i 's. For $1 \leq i, \ell \leq k$, let

$$(i + \ell)_k = \begin{cases} i + \ell, & \text{if } i + \ell \leq k; \\ i + \ell - k, & \text{if } i + \ell > k. \end{cases}$$

For each $1 \leq \ell \leq k$, let $Z_i^\ell = Y_{(i+\ell)_k}$, $S_j^\ell = \sum_{i=1}^j Z_i^\ell$ for $1 \leq i \leq k$. Clearly $S_k^\ell = -1$ for all ℓ as soon as (C_k) is satisfied. On the other hand $S^k \equiv S$ is the only trajectory which satisfies conditions (C_k) . The other S^ℓ hit the value -1 before rank k , see Figure 1.3.1. The Z^ℓ 's are sequences of integers ≥ -1 of length k , whose sum equals -1 . Finally to each element of V_k we have associated k distinct elements of U_k , all having the same probability.

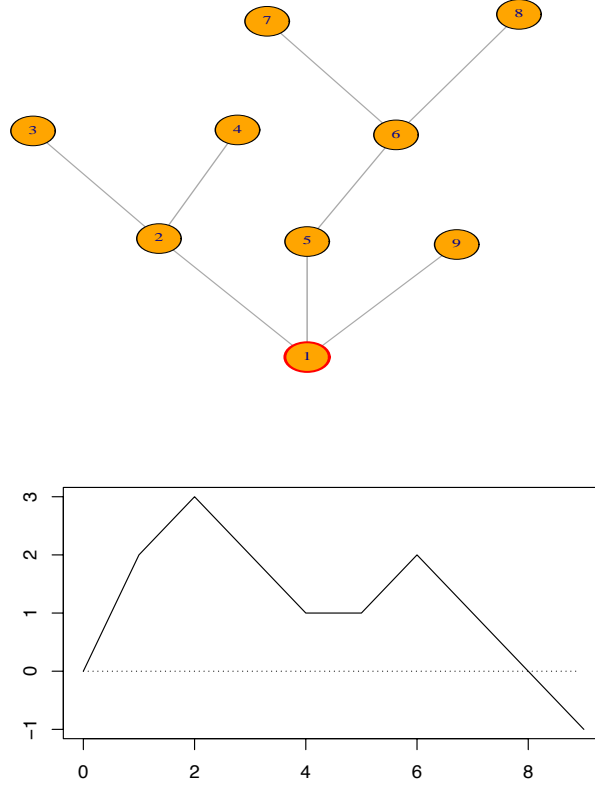


Fig. 1.3.1 Top: the tree. Bottom: the random walk S_k . Here $X_1 = 3, X_2 = 2, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 2, X_7 = X_8 = X_9 = 0, Y_1 = 2, Y_2 = 1, Y_3 = Y_4 = -1, Y_5 = 0, Y_6 = 1, Y_7 = Y_8 = Y_9 = -1$.

Reciprocally, to one element S of $U_k \setminus V_k$, choosing $\ell = \operatorname{argmin}_{1 \leq i \leq k} S_i$ and using the above transformation, we deduce that $S^\ell \in V_k$.

Finally, to each trajectory of V_k , we associate k trajectories of U_k , who all have the same probability, and which are such that the inverse transformation gives back the same trajectory of V_k . The result is proved. \square

Note that from branching process theory (Proposition A.1.1), we have clearly

$$\sum_{k \geq 1} \mathbb{P}(Z = k) \begin{cases} = 1, & \text{if } \mathbb{E}R_0 \leq 1; \\ < 1, & \text{if } \mathbb{E}R_0 > 1, \end{cases}$$

which is not so obvious from the proposition.

We now deduce the exact law of Z from Proposition 1.3.1 in two cases which are probably the two most interesting cases for epidemics models. First we consider the case where the X_i s are Poisson, which is the situation of the continuous time Reed–Frost model, where the infectious period is non-random. Second we consider the case where the X_i s are geometric, which is the case in the Markovian model.

Example 1.3.2. Suppose that the joint law of the X_i s is $\text{Poi}(\mu)$, with $0 < \mu < 1$. Then $X_1 + \dots + X_k \sim \text{Poi}(k\mu)$, and consequently

$$\begin{aligned}\mathbb{P}(Z = k) &= \frac{1}{k} \mathbb{P}(X_1 + \dots + X_k = k - 1) \\ &= e^{-\mu k} \frac{(\mu k)^{k-1}}{k!}.\end{aligned}$$

This law of Z is called the Borel distribution with parameter μ . Note that

$$\begin{aligned}\mathbb{E}Z &= 1 + \mu + \mu^2 + \dots \\ &= \frac{1}{1 - \mu}.\end{aligned}$$

Example 1.3.3. Consider now the case where $X_i \sim \mathcal{G}(p)$, where we mean here that $\mathbb{P}(X_i = k) = (1 - p)^k p$, $k = 0, 1, \dots$. The law of $X_i + 1$ is the geometric distribution with parameter p whose support is \mathbb{N} , in other words $\mathbb{P}(X_i + 1 > k) = (1 - p)^k$. Then $k + X_1 + \dots + X_k$ follows the negative binomial distribution with parameters (k, p) . Hence

$$\begin{aligned}\mathbb{P}(Z = k) &= \frac{1}{k} \mathbb{P}(k + X_1 + \dots + X_k = 2k - 1) \\ &= \frac{1}{k} \binom{2k - 2}{k - 1} p^k (1 - p)^{k-1} \\ &= \frac{(2k - 2)!}{k!(k - 1)!} p^k (1 - p)^{k-1}.\end{aligned}$$

In the case $p > 1/2$, $\mathbb{E}Z = (2p - 1)^{-1} p$.

1.4 Vaccination

One important reason for modelling the spread of infectious diseases is to better understand effects of different preventive measures, such as for example vaccination, isolation and school closure. When a new outbreak occurs, epidemiologists (together with mathematicians and statisticians) estimate model parameters and then use these to predict effects of various preventive measures, and based on these pre-

dictions, health authorities decide upon which preventive measures to put in place, cf. [38].

We refer the reader to Part IV in this volume for estimation methods, but in the current section we touch upon the area of modeling prevention. Our focus is on vaccination, and we consider only vaccination prior to the arrival of an outbreak; the situation where vaccination (or other preventive measures) are put into place *during* the outbreak is not considered. “Vaccination” can be interpreted in a wider sense. From a mathematical and spreading point of view, the important feature is that the individual cannot spread the disease further, which could also be achieved by e.g. isolation or medication. Modelling effects of vaccination is also considered in Part II, Section 2.4, and in Part III, Section 2.6, in the current volume.

Suppose that a fraction v of the community is vaccinated prior to the arrival of the disease. We assume that the vaccine is perfect in the sense that it gives 100% protection from being infected and hence of spreading the disease (but see the exercise below). This implies that only a fraction $1 - v$ are initially susceptible, and the remaining fraction v are immunized (as discussed briefly in Section 2.1). Hence we can neglect the latter fraction and consider only the initial susceptible part of the community of size $N' = N(1 - v)$. However, it is not only the number of initially susceptibles that changes, the rate of having contact with initial susceptibles has also changed to $\lambda' = \lambda(1 - v)$, since a fraction v of all contacts are “wasted” on vaccinated people. The spread of disease in a partly-vaccinated community can therefore be modelled using exactly the same SEIR stochastic model with the only difference being that we have a different population size N' and a different contact rate parameter λ' .

From this we conclude the new reproduction number, which we denote R_v to show the dependence on v , satisfies

$$R_v = \lambda' \mathbb{E}(I) = \lambda(1 - v) \mathbb{E}(I) = (1 - v)R_0.$$

As a consequence, a major outbreak in the community is not possible if $R_v \leq 1$, which (when $R_0 > 1$) is equivalent to $v \geq 1 - 1/R_0$. This limit, called the *critical vaccination coverage* and denoted

$$v_c = 1 - \frac{1}{R_0}, \quad (1.4.1)$$

is hence a very important quantity: if more than this fraction is vaccinated before an outbreak, then the whole community is protected from a major outbreak and not only the vaccinated, a situation called *herd immunity*. Equation (1.4.1) is well known among infectious disease epidemiologists (e.g. Giesecke [14]) and is used by public health authorities all over the world to determine the minimal yearly vaccination coverage in vaccination programs of childhood diseases.

If $v < v_c$ there is still a possibility of a major outbreak. The probability for such an outbreak is obtained using earlier results with λ replaced by $\lambda' = \lambda(1 - v)$: the probability of a minor outbreak is the solution s_v to the equation $s = g_v(s)$, where $g_v(\cdot)$ is the probability generating function of $X_v \sim \text{MixPoi}(\lambda(1 - v)I)$, the num-

ber of offspring (= new infections) in the case that a fraction v are immunized by vaccination.

In the case when there is a major outbreak, the relative size z_v of the outbreak (among the initially susceptible!) is given by the unique positive solution to the equation

$$1 - z = e^{-R_0 z}, \text{ or equivalently } 1 - z = e^{-(1-v)R_0 z}, \quad (1.4.2)$$

this result is shown in later sections, cf. Equation (2.1.3). The community fraction getting infected is hence $(1 - v)z_v$.

We summarize our result in the following theorem where we let Z_v^N denote the final number infected when a fraction v are vaccinated prior to the outbreak.

Theorem 1.4.1. *If $v \geq v_c = 1 - 1/R_0$, then $Z_v^N/N \rightarrow 0$ in probability. If $v < v_c = 1 - 1/R_0$, then $Z_v^N/N \Rightarrow Z_v^\infty$ which has a two-point distribution: $\mathbb{P}(Z_v^\infty = 0) = s_v$ and $\mathbb{P}(Z_v^\infty = (1 - v)z_v) = 1 - s_v$, where s_v and z_v have been defined above.*

Exercise 1.4.2. Consider the Markovian SEIR epidemic with $\lambda = 2$, $L \sim \text{Exp}(2)$ and $I \sim \text{Exp}(1)$. Compute the critical vaccination coverage v_c . Compute also numerically the probability of a major outbreak, and the community-fraction that will get infected in the case of a major outbreak when $v = 0.333$.

Exercise 1.4.3. Suppose that the vaccine gives only partial protection to catching and spreading the disease. Suppose that the vaccine has the effect the risk of getting infected by a contact is only 20% of the risk of getting infected when not vaccinated, but that the vaccine has no effect on infectivity if the person gets infected (such a vaccine is said to be a “leaky vaccine” having 80% efficacy on susceptibility and 0% efficacy on infectivity). Compute the reproduction number R_v in the case that a fraction v is vaccinated with such a vaccine. (Another vaccine response model is “all-or-nothing” where a fraction is assumed to receive 100% effect and the remaining fraction receive no effect from vaccination, for example due to the cold chain being broken for a live vaccine.)

Chapter 2

Markov Models

This chapter describes the important class of Markov models. It starts with a presentation of the deterministic ODE models. We then formulate precisely the random Markov epidemic model as a Poisson process driven stochastic differential equation, and establish the law of large numbers (later referred to as LLN), whose limit is precisely the already described ODE model. The next section studies the fluctuations around this LLN limit, which is described by the central limit theorem. Finally we give a diffusion approximation result, i.e. a diffusion process (solution of a Brownian motion driven stochastic differential equation) which, again in the case of a large population, is a good approximation of our Poisson process driven model. One of the earliest references for those three approximation theorems is Kurtz [22]. See also chapter 11 of Ethier and Kurtz [11].

2.1 The deterministic SEIR epidemic model

Before analysing the stochastic SEIR model assuming $N \rightarrow \infty$ in greater detail in the following subsections, we first derive heuristically a deterministic counterpart for the Markovian version and study some of its properties, which are relevant also for the asymptotic case of the stochastic model.

Consider the Markovian stochastic SEIR model. There are three types of events: a susceptible gets infected and becomes exposed, an exposed becomes infectious when the latent period terminates, and an infectious individual recovers and becomes immune. Since the model is Markovian all these events happen at rates depending only on the current state, and these rates are respectively given by: $\lambda S(t)I(t)/N$, $\nu E(t)$ and $\gamma I(t)$. When an infection occurs, the number of susceptibles decreases by 1 and the number of exposed increases by 1; when a latency period ends, the number of exposed decreases by 1 and the number of infectives increases by 1; and finally when there is a recovery, the number of infectives decreases by 1 and the number of recovered increases by 1. If we instead look at “proportions” (to simplify notation we divide by N rather than the more appropri-

ate choice $N + 1$), the corresponding changes are $-1/N$ and $+1/N$. This reasoning justifies a deterministic model for proportions where one should think of an infinite population size allowing the proportions to be continuous. **The deterministic SEIR epidemic** $(s(t), e(t), i(t), r(t))$ is given by

$$\begin{aligned} s'(t) &= -\lambda s(t)i(t), \\ e'(t) &= \lambda s(t)i(t) - \nu e(t), \\ i'(t) &= \nu e(t) - \gamma i(t), \\ r'(t) &= \gamma i(t). \end{aligned}$$

We start with all fractions being non-negative and summing to unity, which implies that $s(t) + e(t) + i(t) + r(t) = 1$ and all being nonnegative for all t . It is important to stress that this system of differential equations only approximates the *Markovian* SEIR model. If for example the latent and infectious stages are non-random, then a set of differential-delay equations would be the appropriate approximation. If these durations are random but not exponential one possible pragmatic assumption is to use a gamma distribution where the shape parameter is an integer (so it can be seen as a sum of i.i.d. exponentials). Then the deterministic approximation would be a set of differential equations where the state space has been expanded. Just like for the stochastic SEIR model, the deterministic model has to start with a positive fraction of exposed and/or infectives for anything to happen. Most often it is assumed that there is a very small fraction ε of latent and/or infectives.

The case where there is no latent period meaning that $\nu \rightarrow \infty$, the **deterministic SIR epidemic** (or deterministic general epidemic), sometimes called the Kermack–McKendrick equations, has perhaps received more attention in the literature:

$$\begin{cases} s'(t) = -\lambda s(t)i(t), \\ i'(t) = \lambda s(t)i(t) - \gamma i(t), \\ r'(t) = \gamma i(t). \end{cases} \quad (2.1.1)$$

This system of differential equations (and the SEIR system on the previous page) are undoubtedly the most commonly analysed epidemic models (e.g. Anderson and May [1]), and numerous related extended models, capturing various heterogeneous aspects of disease spreading, are published every year in mathematical biology journals.

The deterministic SEIR and SIR share the two most important properties in that they have the same basic reproduction number R_0 and give the same final size (assuming the initial number of infectives/exposed are positive but negligible in both cases), which we now show. In Figure 2.1.1 both the SEIR and SIR systems are plotted for the same values of $\lambda = 1.5$ and $\gamma = 1$ (so $R_0 = 1.5$), and with $\nu = 1$ in the SEIR system.

From the differential equations we see that $s(t)$ is monotonically decreasing and $r(t)$ monotonically increasing. The differential for $i(t)$ in the SIR model can be written $i'(t) = \gamma i(t) \left(\frac{\lambda}{\gamma} s(t) - 1 \right)$. The initial value is $i(0) = \varepsilon \approx 0$ and $s(0) = 1 - \varepsilon \approx$

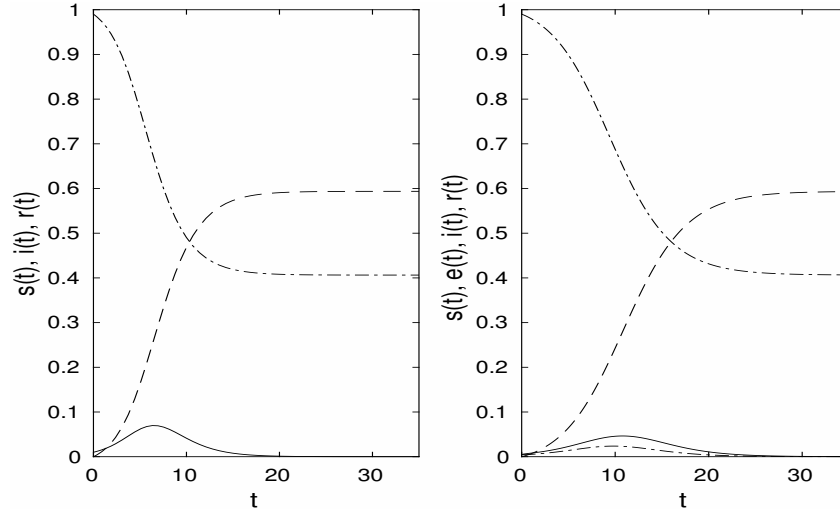


Fig. 2.1.1 Plot of the deterministic SIR (left) and SEIR (right) systems for $\lambda = 1.5$ and $\gamma = 1$, and with $v = 1$ in the SEIR model. The dash-dotted curve is the fraction of susceptibles, the solid curve the fraction of infectives, the dashed curve the fraction of recovered, and the lowest curve in the right figure is the fraction of exposed (latent).

1. From this we see that for having $i'(0) > 0$ we need that $\lambda/\gamma > 1$. If this holds, $i(t)$ grows up until $s(t) < \gamma/\lambda$ after which $i(t)$ decays down to 0. If on the other hand $\lambda/\gamma \leq 1$, then $i(t)$ is decreasing from the start and since its initial value is $\varepsilon \approx 0$, nothing much will happen so $s(\infty) \approx s(0) \approx 1$ and $r(\infty) \approx r(0) = 0$. We hence see that also in the deterministic model, $R_0 = \lambda/\gamma$ plays an important role in that whether or not R_0 exceeds 1 determines whether there will be a substantial or a negligible fraction getting infected during the outbreak. Note that this is the same R_0 as for the Markovian SEIR epidemic. There the infectious period is exponentially distributed with parameter γ , so $\tau := E(I) = 1/\gamma$.

An important difference between deterministic and stochastic epidemic models lies in the initial values. Stochastic models usually start with a small *number* of infectious individuals (in the model of the current Notes we assumed one initial infective: $I(0) = 1$). This implies that the initial *fraction* of infectives tend to 0 as $N \rightarrow \infty$. In the deterministic setting we however have to assume a fixed and strictly positive fraction ε of initially infectives (if we start with a fraction 0 of infectives nothing happens in the deterministic model). This implicitly implies that the deterministic model starts to approximate the stochastic counterpart only when the number of infectives in the stochastic model has grown up to a *fraction* ε , so a number $N\varepsilon$. The earlier part of the stochastic model cannot be approximated by this deterministic model, and as we have seen it might in fact never reach this level (if there is only a minor outbreak).

In order to derive an expression for the ultimate fraction getting infected we use the differential for $s(t)$ (and below also the one for $r(t)$). Dividing by s and

multiplying by dt gives the following differential: $ds/s = -\lambda i dt$. Integrating both sides and recalling that $R_0 = \lambda/\gamma$, we obtain

$$\begin{aligned}\log s(t) - \log s(0) &= -\lambda \int_0^t i(t) dt \\ &= -R_0 \int_0^t r'(s) ds \\ &= -R_0(r(t) - r(0)) = -R_0 r(t).\end{aligned}$$

And since $s(0) = 1 - \varepsilon \approx 1$ and $r(\infty) = 1 - s(\infty)$ we obtain the following equation for the final size $z = r(\infty) = 1 - s(\infty)$:

$$1 - z = e^{-R_0 z}. \quad (2.1.2)$$

In Section 3.3.1 we show that this final size equation coincides with that of the LLN limit of the final fraction getting infected in the stochastic model (cf. Equation (3.3.2), which is identical to (2.1.2)).

The equation always has a root at $z = 0$ corresponding to no (or minor) outbreak. It can be shown (cf. Exercise 2.1.1) that if and only if $R_0 > 1$ there is a second solution to (2.1.2), corresponding to the size of a major outbreak, and this solution z^* is strictly positive and smaller than 1. For a given value of $R_0 > 1$ the solution z^* has to be computed numerically. In Figure 2.1.2 the solution is plotted as a function of R_0 .

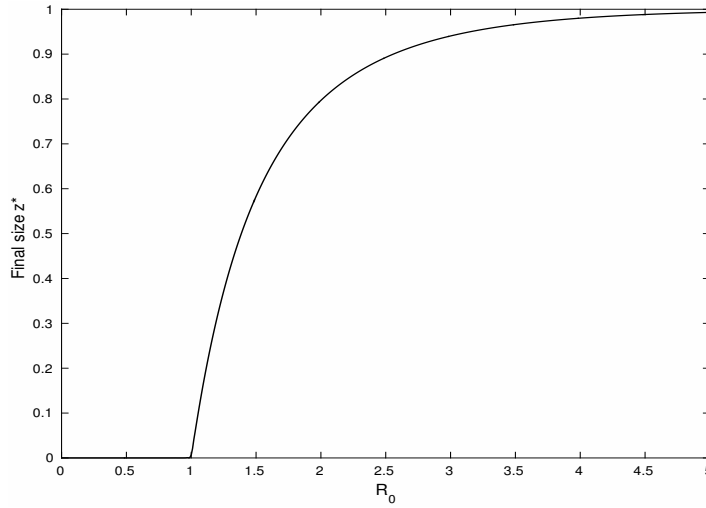


Fig. 2.1.2 Plot of the final size solution z^* to Equation (2.1.2) as a function of R_0 .

It is important to point out that the final size equation (2.1.2) assumes that, at $t = 0$, all individuals (except the very few initially latent and infectives) are susceptible. If a fraction v is initially immune (perhaps due to natural immunity, or vaccination as described in Section 1.4) then $r(0) = v$ and $s(0) = 1 - v$, resulting in the equation

$$1 - z = e^{-R_0 z(1-v)}, \quad (2.1.3)$$

where its solution z_v now is interpreted as the fraction among the initially susceptible that get infected. The overall fraction getting infected is hence $z_v(1 - v)$. Using the same argument as for the final size without immunity, we conclude that $z = 0$ is the only solution if $R_0(1 - v) \leq 1$. This is equivalent to $v \geq 1 - 1/R_0$. If immunity was caused by vaccination, this hence suggests that a fraction exceeding $v_c = 1 - 1/R_0$ should be vaccinated; then there will be no outbreak! For this reason, the quantity $v_c = 1 - 1/R_0$ is often called the *critical vaccination coverage*, and if this coverage is reached, so-called *herd immunity* is achieved. Herd immunity implies that not only the vaccinated are protected, but so are also the unvaccinated, since the community is protected from epidemic outbreaks.

Exercise 2.1.1. Show that $z = 0$ is the only solution to (2.1.2) when $R_0 \leq 1$ and that there is a unique positive solution if $R_0 > 1$. (*Hint:* Study suitable properties of the function $f(z) = e^{-R_0 z} + z - 1$.)

Exercise 2.1.2. Compute the final size numerically for $R_0 = 1.5$ (e.g. influenza), $R_0 = 3$ (e.g. rubella) and $R_0 = 15$ (e.g. measles).

2.2 Law of Large Numbers

Consider a general compartmental model, which takes the form

$$\mathcal{Z}_t^N = z_N + \sum_{j=1}^k h_j P_j \left(\int_0^t \beta_{N,j}(s, \mathcal{Z}_s^N) ds \right),$$

where the P_j s are mutually independent standard (i.e. unit rate) Poisson processes, and $\beta_{N,j}(t, \mathcal{Z}_t^N)$ is the rate of jumps in the direction h_j at time t , h_j being a d -dimensional vector. \mathcal{Z}_t^N takes values in \mathbb{Z}_+^d . The i -th component of \mathcal{Z}_t^N is the number of individuals in the i -th compartment at time t . N is a scale parameter. In the case of models with fixed total population size, $N = \sum_{i=1}^d Z_t^{N,i}$ is the total population at any time t . Note that the above formula for \mathcal{Z}_t^N can be rewritten equivalently, following the comments at the end of Section A.2 in the Appendix below, as

$$\mathcal{Z}_t^N = z_N + \sum_{j=1}^k h_j \int_0^t \int_0^{\beta_{N,j}(s, \mathcal{Z}_s^N)} Q_j(ds, du),$$

where Q_1, \dots, Q_k are mutually independent Poisson random measures on \mathbb{R}_+^2 , with mean measure $ds du$.

We now define

$$Z_t^N = N^{-1} \mathcal{Z}_t^N$$

the vector of rescaled numbers of individuals in the various compartments. In the case of a constant population size equal to N , the components of the vector Z_t^N are the proportions of the total population in the various compartments at time t . The equation for Z_t^N reads, with $x_N = N^{-1}z_N$,

$$Z_t^N = x_N + \sum_{j=1}^k \frac{h_j}{N} P_j \left(\int_0^t \beta_{N,j}(s, N Z_s^N) ds \right).$$

Example 2.2.1. The SIR model.

One important example is that of the *SIR* model with constant population size. Suppose there is no latency period and that the duration of infection satisfies $I \sim \text{Exp}(\gamma)$. In that case, let $S(t)$, $I(t)$, and $R(t)$ denote respectively the number of susceptibles, infectives and recovered at time t .

In this model, two types of events happen:

1. infection of a susceptible (such an event decreases $S(t)$ by one, and increases $I(t)$ by one, so $h_1 = (-1, 1, 0)$); these events happen at rate

$$\beta_{N,1}(t, \mathcal{Z}_t) = \frac{\lambda}{N} S(t) I(t), \quad \text{where } \lambda = cp;$$

2. recovery of an infective (such an event decreases $I(t)$ by one, and increases $R(t)$ by one, so $h_2 = (0, -1, 1)$); these events happen at rate

$$\beta_{N,2}(t, \mathcal{Z}_t) = \gamma I(t).$$

Hence we have the following equations, with $P_1(t)$ and $P_2(t)$ two standard mutually independent Poisson processes:

$$\begin{aligned} S(t) &= S(0) - P_1 \left(\frac{\lambda}{N} \int_0^t S(r) I(r) dr \right), \\ I(t) &= I(0) + P_1 \left(\frac{\lambda}{N} \int_0^t S(r) I(r) dr \right) - P_2 \left(\gamma \int_0^t I(r) dr \right), \\ R(t) &= R(0) + P_2 \left(\gamma \int_0^t I(r) dr \right). \end{aligned}$$

We can clearly forget about the third equation, since $R(t) = N - S(t) - I(t)$.

We now define $(S^N(t), I^N(t)) = (N^{-1}S(t), N^{-1}I(t))$. We have

$$S^N(t) = S^N(0) - \frac{1}{N} P_1 \left(N \lambda \int_0^t S^N(r) I^N(r) dr \right),$$

$$I^N(t) = I^N(0) + \frac{1}{N}P_1 \left(N\lambda \int_0^t S^N(r)I^N(r)dr \right) - \frac{1}{N}P_2 \left(N\gamma \int_0^t I^N(r)dr \right).$$

The above model assumes that λ and γ are constant, but in applications at least λ may depend upon t .

Example 2.2.2. The SEIRS model with demography.

We now describe one rather general example. We add to the preceding example the state E and the fact that removed individuals lose their immunity at a certain rate, which gives the SEIRS model. In addition, we add demography. There is an influx of susceptible individuals at rate μN , and each individual, irrespective of its type, dies at rate μ . This gives the following stochastic differential equation

$$\begin{aligned} S(t) &= S(0) - P_{se} \left(\frac{\lambda}{N} \int_0^t S(r)I(r)dr \right) + P_{rs} \left(\rho \int_0^t R(r)dr \right) \\ &\quad + P_b(\mu Nt) - P_{ds} \left(\mu \int_0^t S(r)dr \right), \\ E(t) &= E(0) + P_{se} \left(\frac{\lambda}{N} \int_0^t S(r)I(r)dr \right) - P_{ei} \left(\nu \int_0^t E(r)dr \right) \\ &\quad - P_{de} \left(\mu \int_0^t E(r)dr \right), \\ I(t) &= I(0) + P_{ei} \left(\nu \int_0^t E(r)dr \right) - P_{ir} \left(\gamma \int_0^t I(r)dr \right) - P_{di} \left(\mu \int_0^t I(r)dr \right), \\ R(t) &= R(0) + P_{ir} \left(\gamma \int_0^t I(r)dr \right) - P_{rs} \left(\rho \int_0^t R(r)dr \right) - P_{dr} \left(\mu \int_0^t R(r)dr \right). \end{aligned}$$

In this system, the various Poisson processes are standard and mutually independent. The indices should be self-explanatory. Note that the rate of births is $\mu \times N$ rather than $\mu \times$ the actual number of individuals in the population, in order to avoid the pitfalls of branching processes (either exponential growth or extinction). Also, the probability $S(t)/N(t)$ that an infective meets a susceptible (where $N(t)$ denotes the total population at time t) is approximated by $S(t)/N$ for the sake of mathematical simplicity. Note however that $\frac{N(t)}{N} \rightarrow 1$ a.s. as $N \rightarrow \infty$, see Exercise 4.1.1 below. The equations for the proportions in the various compartments read

$$\begin{aligned} S^N(t) &= S^N(0) - \frac{1}{N}P_{se} \left(N\lambda \int_0^t S^N(r)I^N(r)dr \right) + \frac{1}{N}P_{rs} \left(N\rho \int_0^t R^N(r)dr \right) \\ &\quad + \frac{1}{N}P_b(\mu Nt) - \frac{1}{N}P_{ds} \left(\mu N \int_0^t S^N(r)dr \right), \\ E^N(t) &= E^N(0) + \frac{1}{N}P_{se} \left(N\lambda \int_0^t S^N(r)I^N(r)dr \right) \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{N}P_{ei}\left(vN\int_0^t E^N(r)dr\right) - \frac{1}{N}P_{de}\left(\mu N\int_0^t E^N(r)dr\right), \\
I^N(t) &= I^N(0) + \frac{1}{N}P_{ei}\left(vN\int_0^t E^N(r)dr\right) - \frac{1}{N}P_{ir}\left(N\gamma\int_0^t I^N(r)dr\right) \\
& - \frac{1}{N}P_{di}\left(\mu N\int_0^t I^N(r)dr\right), \\
R^N(t) &= R^N(0) + \frac{1}{N}P_{ir}\left(N\gamma\int_0^t I^N(r)dr\right) - \frac{1}{N}P_{rs}\left(N\rho\int_0^t R^N(r)dr\right) \\
& - \frac{1}{N}P_{dr}\left(\mu N\int_0^t R^N(r)dr\right).
\end{aligned}$$

Example 2.2.3. A variant of the SEIRS model with demography.

In the preceding example, we decided to replace the true proportion of susceptibles by its approximation $S(t)/N$, in order to avoid complications. There is another option, which is to force the population to remain constant. The most natural way to achieve this is to assume that each death event coincides with a birth event. Every susceptible, exposed, infected, removed individual dies at rate μ . Each death is compensated by the birth of a susceptible. The equation for the evolution of $(S(t), E(t), I(t), R(t))$ reads

$$\begin{aligned}
S(t) &= S(0) - P_{se}\left(\frac{\lambda}{N}\int_0^t S(r)I(r)dr\right) + P_{rs}\left(\rho\int_0^t R(r)dr\right) \\
& + P_{ds}\left(\mu\int_0^t S(r)dr\right) + P_{de}\left(\mu\int_0^t E(r)dr\right) + P_{di}\left(\mu\int_0^t I(r)dr\right) \\
& + P_{dr}\left(\mu\int_0^t R(r)dr\right) - P_{ds}\left(\mu\int_0^t S(r)dr\right), \\
E(t) &= E(0) + P_{se}\left(\frac{\lambda}{N}\int_0^t S(r)I(r)dr\right) - P_{ei}\left(v\int_0^t E(r)dr\right) \\
& - P_{de}\left(\mu\int_0^t E(r)dr\right), \\
I(t) &= I(0) + P_{ei}\left(v\int_0^t E(r)dr\right) - P_{ir}\left(\gamma\int_0^t I(r)dr\right) - P_{di}\left(\mu\int_0^t I(r)dr\right), \\
R(t) &= R(0) + P_{ir}\left(\gamma\int_0^t I(r)dr\right) - P_{rs}\left(\rho\int_0^t R(r)dr\right) - P_{dr}\left(\mu\int_0^t R(r)dr\right).
\end{aligned}$$

The equations for the proportions in the various compartments read

$$S^N(t) = S^N(0) - \frac{1}{N}P_{se}\left(N\lambda\int_0^t S^N(r)I^N(r)dr\right)$$

$$\begin{aligned}
& + \frac{1}{N} P_{rs} \left(N \rho \int_0^t R^N(r) dr \right) + \frac{1}{N} P_{ds} \left(N \mu \int_0^t S^N(r) dr \right) \\
& + \frac{1}{N} P_{de} \left(N \mu \int_0^t E^N(r) dr \right) + \frac{1}{N} P_{di} \left(N \mu \int_0^t I^N(r) dr \right) \\
& + \frac{1}{N} P_{dr} \left(N \mu \int_0^t R^N(r) dr \right) - \frac{1}{N} P_{ds} \left(N \mu \int_0^t S^N(r) dr \right), \\
E^N(t) &= E^N(0) + \frac{1}{N} P_{se} \left(N \lambda \int_0^t S^N(r) I^N(r) dr \right) - \frac{1}{N} P_{ei} \left(N \nu \int_0^t E^N(r) dr \right) \\
& - \frac{1}{N} P_{de} \left(N \mu \int_0^t E^N(r) dr \right), \\
I^N(t) &= I^N(0) + \frac{1}{N} P_{ei} \left(\nu \int_0^t E^N(r) dr \right) - \frac{1}{N} P_{ir} \left(\gamma \int_0^t I^N(r) dr \right) \\
& - \frac{1}{N} P_{di} \left(N \mu \int_0^t I^N(r) dr \right), \\
R^N(t) &= R^N(0) + \frac{1}{N} P_{ir} \left(N \gamma \int_0^t I^N(r) dr \right) - \frac{1}{N} P_{rs} \left(N \rho \int_0^t R^N(r) dr \right) \\
& - \frac{1}{N} P_{dr} \left(N \mu \int_0^t R^N(r) dr \right).
\end{aligned}$$

In the three above examples, for each j , $\beta_{N,j}(t, Nz) = N\beta_j(t, z)$, for some $\beta_j(t, z)$ which does not depend upon N . We shall assume from now on that this is the case in our general model, namely that

$$\beta_{N,j}(t, Nz) = N\beta_j(t, z), \quad \text{for all } 1 \leq j \leq k, N \geq 1, z \in \mathbb{R}_+^d.$$

Remark 2.2.4. We could assume more generally that

$$\beta_{N,j}(t, Nz) = N\tilde{\beta}_{N,j}(t, z), \quad \text{where } \tilde{\beta}_{N,j}(t, z) \rightarrow \beta_j(t, z),$$

locally uniformly as $N \rightarrow \infty$.

Finally our model reads

$$Z_t^N = x_N + \sum_{j=1}^k \frac{h_j}{N} P_j \left(\int_0^t N\beta_j(s, Z_s^N) ds \right). \quad (2.2.1)$$

We note that in the first example above, $0 \leq Z_j^N(t) \leq 1$ for all $1 \leq j \leq k, t \geq 0, N \geq 1$. In the second example however, such a simple upper bound does not hold, but a much weaker assumption will suffice.

We assume that all β_j are locally bounded, which is clearly satisfied in all examples we can think of, so that for any $K > 0$,

$$C(T, K) := \sup_{1 \leq j \leq k} \sup_{0 \leq t \leq T} \sup_{|z| \leq K} \beta_j(t, z) < \infty. \quad (2.2.2)$$

We first prove the Law of Large Numbers for Poisson processes.

Proposition 2.2.5. *Let $\{P(t), t \geq 0\}$ be a rate λ Poisson process. Then*

$$t^{-1}P(t) \rightarrow \lambda \text{ a.s. as } t \rightarrow \infty.$$

Proof. Consider first for $n \in \mathbb{Z}_+$

$$\begin{aligned} n^{-1}P(n) &= n^{-1} \sum_{i=1}^n [P(i) - P(i-1)] \\ &\rightarrow \lambda \text{ a.s. as } n \rightarrow \infty \end{aligned}$$

from the standard strong **Law of Large Numbers**, since the random variables $P(i) - P(i-1)$, $1 \leq i \leq n$ are i.i.d. Poisson with parameter λ . Now

$$\begin{aligned} t^{-1}P(t) &= \frac{[t]}{t} [t]^{-1}P([t]) + t^{-1}\{P(t) - P([t])\}, \\ \text{so } |t^{-1}P(t) - \lambda| &\leq \left| \frac{[t]}{t} [t]^{-1}P([t]) - \lambda \right| + t^{-1}\{P([t] + 1) - P([t])\}. \end{aligned}$$

But

$$t^{-1}\{P([t] + 1) - P([t])\} = t^{-1}P([t] + 1) - t^{-1}P([t])$$

is the difference of two sequences which converge a.s. towards the same limit, hence it converges to 0 a.s. \square

Define the continuous time martingales (see Section A.4.2 in Appendix A) $M_j(t) = P_j(t) - t$, $1 \leq j \leq k$. We have

$$Z_t^N = x_N + \int_0^t b(s, Z_s^N) ds + \sum_{j=1}^k \frac{h_j}{N} M_j \left(\int_0^t N \beta_j(s, Z_s^N) ds \right),$$

where

$$b(t, x) = \sum_{j=1}^k h_j \beta_j(t, x).$$

Consider the k -dimensional process $\mathcal{M}^N(t)$ whose j -th component is defined as

$$\mathcal{M}_j^N(t) := \frac{1}{N} M_j \left(N \int_0^t \beta_j(r, Z_r^N) dr \right).$$

From the above, we readily deduce the following.

Proposition 2.2.6. *For any $K > 0$, let $\tau_K := \inf\{t > 0, |Z_t^N| \geq K\}$. As $N \rightarrow \infty$, for all $T > 0$, provided (2.2.2) holds,*

$$\sup_{0 \leq t \leq T \wedge \tau_K} |\mathcal{M}^N(t)| \rightarrow 0 \text{ a.s.}$$

Proof. In order to simplify the notation we treat the case $d = 1$. It follows from (2.2.2) that, if $M(t) = P(t) - t$ and N is large enough,

$$\sup_{0 \leq t \leq T \wedge \tau_K} |\mathcal{M}^N(t)| \leq \frac{1}{N} \sup_{0 \leq r \leq NTC(T,K)} |M(r)|.$$

From the previous proposition, for all $t > 0$,

$$\frac{P(Nt)}{N} \rightarrow t \quad \text{a.s. as } N \rightarrow \infty.$$

Note that we have pointwise convergence of a sequence of increasing functions towards a continuous (and of course increasing) function. Consequently from the second Dini Theorem (see e.g. pages 81 and 270 in Polya and Szegő [30]), this convergence is uniform on any compact interval, hence for all $T > 0$,

$$\frac{1}{N} \sup_{0 \leq r \leq NTC(T,K)} |M(r)| \rightarrow 0 \quad \text{a.s.}$$

□

Concerning the initial condition, we assume that for some $x \in [0, 1]^d$, $x_N = [Nx]/N$, where $[Nx]$ is of course a vector of integers. We can now prove the following theorem.

Theorem 2.2.7. Law of Large Numbers *Assume that the initial condition is given as above, that $b(t, x) = \sum_{j=1}^k \beta_j(t, x) h_j$ is locally Lipschitz as a function of x , locally uniformly in t , that (2.2.2) holds and that the unique solution of the ODE*

$$\frac{dz_t}{dt} = b(t, z_t), \quad z_0 = x$$

does not explode in finite time. Let Z_t^N denote the solution of the SDE (2.2.1). Then $Z_t^N \rightarrow z_t$ a.s. locally uniformly in t , where $\{z_t, t \geq 0\}$ is the unique solution of the above ODE.

Needless to say, our theorem applies to the general model (2.2.1). We shall describe below three specific models to which we can apply it. Note that if the initial fraction of infected is zero, then the fraction of infected is zero for all $t \geq 0$.

Proof. We have

$$Z_t^N = x_N + \int_0^t b(s, Z_s^N) ds + \sum_{j=1}^k h_j \mathcal{M}_j^N(t).$$

Let us fix an arbitrary $T > 0$. We want to show uniform convergence on $[0, T]$. Let $K := \sup_{0 \leq t \leq T} |z_t| + C$, where $C > 0$ is arbitrary, and let $\tau_K = \inf\{t > 0, |Z_t^N| \geq K\}$. Since $b(t, \cdot)$ is locally Lipschitz,

$$c_{T,K} := \sup_{0 \leq t \leq T, x \neq x', |x|, |x'| \leq K} \frac{|b(t, x) - b(t, x')|}{|x - x'|} < \infty.$$

For any $0 \leq t \leq T$, if we define $Y_t^N = \sum_{j=1}^k h_j \cdot \mathcal{M}_j^N(t)$, we have

$$\begin{aligned} |Z_{t \wedge \tau_K}^N - z_{t \wedge \tau_K}| &\leq |x_N - x| + c_{T,K} \int_0^{t \wedge \tau_K} |Z_s^N - z_s| ds + |Y_{t \wedge \tau_K}^N| \\ &\leq \varepsilon_N \exp(c_{T,K} t), \end{aligned}$$

where $\varepsilon_N := |x_N - x| + \sup_{0 \leq t \leq T \wedge \tau_K} |Y_t^N|$ and we have used Gronwall's Lemma 2.2.9 below. It follows from our assumption on x_N and Proposition 2.2.6 that $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$. The result follows, since as soon as $\varepsilon_N \exp(c_{T,K} T) \leq C$, $\tau_K \geq T$. \square

Remark 2.2.8. Showing that a stochastic epidemic model (for population proportions) converges to a particular deterministic process is important also for applications. This motivates the use of deterministic models, which are easier to analyse, in the case of large populations.

Lemma 2.2.9. Gronwall *Let $a, b \geq 0$ and $\varphi : [0, T] \rightarrow \mathbb{R}$ be such that for all $0 \leq t \leq T$,*

$$\varphi(t) \leq a + b \int_0^t \varphi(r) dr.$$

Then $\varphi(t) \leq ae^{bt}$.

Proof. We deduce from the assumption that

$$e^{-bt} \varphi(t) - be^{-bt} \int_0^t \varphi(r) dr \leq ae^{-bt},$$

or in other words

$$\frac{d}{dt} \left(e^{-bt} \int_0^t \varphi(r) dr \right) \leq ae^{-bt}.$$

Integrating this inequality, we deduce

$$e^{-bt} \int_0^t \varphi(r) dr \leq a \frac{1 - e^{-bt}}{b}.$$

Multiplying by be^{bt} and exploiting again the assumption yields the result. \square

Example 2.2.10. The SIR model. It is clear that Theorem 2.2.7 applies to Example 2.2.1. The limit of $(S^N(t), I^N(t))$ is the solution $(s(t), i(t))$ of the ODE

$$\begin{aligned} s'(t) &= -\lambda s(t)i(t), \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t). \end{aligned}$$

Example 2.2.11. The SEIRS model with demography (continued). Again Theorem 2.2.7 applies to Example 2.2.2. The limit of $(S^N(t), E^N(t), I^N(t), R^N(t))$ is the solution $(s(t), e(t), i(t), r(t))$ of the ODE

$$\begin{aligned}
s'(t) &= \mu(1 - s(t)) - \lambda s(t)i(t) + \rho r(t), \\
e'(t) &= \lambda s(t)i(t) - (\nu + \mu)e(t), \\
i'(t) &= \nu e(t) - (\gamma + \mu)i(t), \\
r'(t) &= \gamma i(t) - (\rho + \mu)r(t).
\end{aligned}$$

Note that if we define the total renormalized population as $n(t) = s(t) + e(t) + i(t) + r(t)$, then it is easy to deduce from the above ODE that $n'(t) = \mu(1 - n(t))$, consequently $n(t) = 1 + e^{-\mu t}(n(0) - 1)$. If $n(0) = 1$, then $n(t) \equiv 1$, and we can reduce the above model to a three-dimensional model (and to a two-dimensional model as in the previous example if we are treating the SIR or the SIRS model with demography).

We note that this “Law of Large Numbers” approximation is only valid when $s, i > 0$, i.e. when significant fractions of the population are infective and are susceptible, in particular at time 0. The ODE is of course of no help to compute the probability that the introduction of a single infective results in a major epidemic.

The vast majority of the literature on mathematical models in epidemiology considers ODEs of the type of equations which we have just obtained. The probabilistic point of view is more recent.

Exercise 2.2.12. Let us consider Ross’s model of malaria, which we write in a stochastic form. Denote by $H(t)$ the number of humans (hosts) who are infected by malaria, and by $V(t)$ the number of mosquitos (vectors) who are infected by malaria at time t . Let N_H denote the total number of humans, and N_V denote the total number of mosquitos, which are assumed to be constant in time. The humans (resp. the mosquitos) which are not infected are all supposed to be susceptibles. Let $m = N_V/N_H$ and denote by a the mean number of bites of humans by one mosquito per time unit, p_{VH} the probability that the bite of a susceptible human by an infected mosquito infects the human, and by p_{HV} the probability that a susceptible mosquito gets infected while biting an infected human. We assume that the infected humans (resp. mosquitos) recover at rate γ (resp. at rate μ).

1. What is the mean number of bites that a human suffers per time unit?
2. Given 4 mutually independent standard Poisson processes $P_1(t), P_2(t), P_3(t)$ and $P_4(t)$, justify the following as a stochastic model of the propagation of malaria.

$$\begin{aligned}
H(t) &= H(0) + P_1 \left(a p_{VH} \int_0^t V(s) \frac{N_H - H(s)}{N_H} ds \right) - P_2 \left(\gamma \int_0^t H(s) ds \right) \\
V(t) &= V(0) + P_3 \left(a m p_{HV} \int_0^t H(s) \frac{N_V - V(s)}{N_V} ds \right) - P_4 \left(\mu \int_0^t V(s) ds \right).
\end{aligned}$$

3. Define now (with $N_H = N, N_V = mN$)

$$h_N(t) = \frac{H(t)}{N_H}, \quad v_N(t) = \frac{V(t)}{N_V}.$$

Write the equation for the pair $(h_N(t), v_N(t))$. Show that as $N \rightarrow \infty$, with m constant, $(h_N(t), v_N(t)) \rightarrow (h(t), v(t))$, the solution of Ross's ODE:

$$\begin{aligned}\frac{dh}{dt}(t) &= ap_{VH}mv(t)(1-h(t)) - \gamma h(t), \\ \frac{dv}{dt}(t) &= ap_{HV}h(t)(1-v(t)) - \mu v(t).\end{aligned}$$

2.3 Central Limit Theorem

In the previous section we have shown that the stochastic process describing the evolution of the proportions of the total population in the various compartments converges, in the asymptotic of large population, to the deterministic solution of a system of ODEs. In the current section we look at fluctuations of the difference between the stochastic epidemic process and its deterministic limit.

We now introduce the rescaled difference between Z_t^N and z_t , namely

$$U_t^N = \sqrt{N}(Z_t^N - z_t).$$

We wish to show that U_t^N converges in law to a Gaussian process. It is clear that

$$U_t^N = \sqrt{N}(x_N - x) + \sqrt{N} \int_0^t [b(s, Z_s^N) - b(s, z_s)] ds + \sum_{j=1}^k h_j \cdot \widetilde{\mathcal{M}}_j^N(t),$$

where for $1 \leq j \leq k$,

$$\widetilde{\mathcal{M}}_j^N(t) = \frac{1}{\sqrt{N}} M_j \left(N \int_0^t \beta_j(r, Z_r^N) dr \right).$$

We certainly need to find the limit in law of the k dimensional process $\widetilde{\mathcal{M}}_t^N$, whose j -th coordinate is $\widetilde{\mathcal{M}}_j^N(t)$. We prove the following proposition below.

Proposition 2.3.1. *As $N \rightarrow \infty$,*

$$\{\widetilde{\mathcal{M}}_t^N, t \geq 0\} \Rightarrow \{\widetilde{\mathcal{M}}_t, t \geq 0\}$$

meaning weak convergence for the topology of locally uniform convergence, where for $1 \leq j \leq k$, $\widetilde{\mathcal{M}}_j(t) = \int_0^t \sqrt{\beta_j(s, z_s)} dB_j(s)$ and the processes $B_1(t), \dots, B_k(t)$ are mutually independent standard Brownian motions.

Let us first show that the main result of this section is indeed a consequence of this proposition.

Theorem 2.3.2. Central Limit Theorem *In addition to the assumptions of Theorem 2.2.7, we assume that $x \rightarrow b(t, x)$ is of class C^1 , locally uniformly in t . Then, as $N \rightarrow \infty$, $\{U_t^N, t \geq 0\} \Rightarrow \{U_t, t \geq 0\}$, where*

$$U_t = \int_0^t \nabla_x b(s, z_s) U_s ds + \sum_{j=1}^k h_j \int_0^t \sqrt{\beta_j(s, z_s)} dB_j(s), \quad t \geq 0. \quad (2.3.1)$$

Proof. We shall fix an arbitrary $T > 0$ throughout the proof. Let $V^N(s) := \sqrt{N}[b(s, Z_s^N) - b(s, z_s)]$ and $\widetilde{\mathcal{N}}_t^N := \sum_{j=1}^k h_j \widetilde{\mathcal{M}}_j^N(t)$. We have

$$U_t^N = U_0^N + \int_0^t V^N(s) ds + \widetilde{\mathcal{N}}_t^N.$$

Let us admit for the moment the following lemma.

Lemma 2.3.3. *For each $N \geq 1$, $0 \leq t \leq T$ there exists a random $d \times d$ matrix A_t^N such that*

$$V_t^N = \nabla b(t, z_t) U_t^N + A_t^N U_t^N.$$

Moreover, $\sup_{0 \leq t \leq T} \|A_t^N\| \rightarrow 0$, a.s., as $N \rightarrow \infty$.

We clearly have

$$U_t^N = U_0^N + \int_0^t [\nabla b(s, z_s) + A_s^N] U_s^N ds + \widetilde{\mathcal{N}}_t^N.$$

It then follows from Gronwall's Lemma that

$$\sup_{0 \leq t \leq T} |U_t^N| \leq \left(|U_0^N| + \sup_{0 \leq t \leq T} |\widetilde{\mathcal{N}}_t^N| \right) \exp \left(\sup_{0 \leq t \leq T} \|\nabla b(t, z_t) + A_t^N\| T \right).$$

The right-hand side of this inequality is tight,¹ hence the same is true for the left-hand side. From this and Lemma 2.3.3 it follows that $R_t^N := A_t^N U_t^N$ tends to 0 in probability as $N \rightarrow \infty$, uniformly for $0 \leq t \leq T$. Consequently

$$U_t^N = \int_0^t \nabla_x b(s, z_s) U_s^N ds + W_t^N, \quad \text{where}$$

$$W_t^N = U_0^N + \int_0^t R_s^N ds + \widetilde{\mathcal{N}}_t^N.$$

The following two hold

1. $\sup_{0 \leq t \leq T} |U_0^N + \int_0^t R_s^N ds| \rightarrow 0$ in probability, and from Proposition 2.3.1 $\widetilde{\mathcal{N}}_t^N \Rightarrow \widetilde{\mathcal{N}}_t$, hence $W_t^N \Rightarrow \widetilde{\mathcal{N}}_t$ for the topology of uniform convergence on $[0, T]$.
2. The mapping $y \mapsto \Phi(y)$, which to $y \in C([0, T]; \mathbb{R}^d)$ associates $x \in C([0, T]; \mathbb{R}^d)$, the solution of the ODE

$$x(t) = \int_0^t \nabla b(s, z_s) x(s) ds + y(t),$$

is continuous.

¹ A sequence ξ_n of \mathbb{R}_+ -valued random variables is tight if for any $\varepsilon > 0$, there exists an M_ε such that $\mathbb{P}(\xi_n > M_\varepsilon) \leq \varepsilon$, for all $n \geq 1$, see Section A.5 in the Appendix.

Indeed, we can construct this mapping by first solving the ODE

$$\dot{z}(t) = \nabla b(t, z_t)[z(t) + y(t)], \quad z(0) = 0,$$

and then defining $x(t) = z(t) + y(t)$.

Since

$$U^N = \Phi(W^N),$$

the result follows from 1. and 2., and the fact that T is arbitrary. \square

Proof of Lemma 2.3.3. For $1 \leq i \leq d, 0 \leq t \leq T$, define the random function $\rho_{i,t}(\theta) = b_i(t, z_t + \theta(Z_t^N - z_t))$, $0 \leq \theta \leq 1$. The mean value theorem applied to the function $\rho_{i,t}$ implies that for all $0 \leq i \leq d, 0 \leq t \leq T$, there exists a random $0 < \bar{\theta}_{i,t} < 1$ such that

$$b_i(t, Z_t^N) - b_i(t, z_t) = \langle \nabla b_i(t, z_t + \bar{\theta}_{i,t}(Z_t^N - z_t)), Z_t^N - z_t \rangle.$$

Applying the same argument for all $1 \leq i \leq d$ yields the first part of the Lemma. Theorem 2.2.7 and the continuity in z of $\nabla b(t, z)$ uniformly in t imply that $\nabla b_i(t, z_t + \bar{\theta}_{i,t}(Z_t^N - z_t)) - \nabla b_i(t, z_t) \rightarrow 0$ a.s., uniformly in t , as $N \rightarrow \infty$. \square

It remains to prove Proposition 2.3.1. Let us first establish a central limit theorem for standard Poisson processes. Let $\{P_j(t), t \geq 0\}_{1 \leq j \leq k}$ be k mutually independent standard Poisson processes and $M(t)$ denote the k -dimensional process whose j -th component is $P_j(t) - t$.

Lemma 2.3.4. As $N \rightarrow \infty$,

$$\frac{M(Nt)}{\sqrt{N}} \Rightarrow B(t),$$

where $B(t)$ is a k -dimensional standard Brownian motion (in particular $B(t) \sim \mathcal{N}(0, tI)$, with I the $d \times d$ identity matrix) and the convergence is in the sense of convergence in law in $D([0, +\infty); \mathbb{R}^k)$.

For a definition of the space $D([0, +\infty); \mathbb{R}^k)$ of the \mathbb{R}^k -valued càlàg functions of $t \in [0, \infty)$ and its topology, see section A.5 in the Appendix.

Proof. It suffices to consider each component separately, since they are independent. So we do as if $k = 1$. We first note that our process is a martingale, whose associated predictable increasing process is given by $\langle N^{-1/2}M(N\cdot), N^{-1/2}M(N\cdot) \rangle_t = t$. Hence it is tight.

Let us now compute the characteristic function of the random variable $N^{-1/2}M(Nt)$. We obtain

$$\mathbb{E} \left(\exp \left[iu N^{-1/2}M(Nt) \right] \right) = \exp \left(Nt \left[e^{i \frac{u}{\sqrt{N}}} - 1 - i \frac{u}{\sqrt{N}} \right] \right) \rightarrow \exp \left(-t \frac{u^2}{2} \right),$$

as $N \rightarrow \infty$. This shows that $N^{-1/2}M(Nt)$ converges in law to an $\mathcal{N}(0, t)$ r.v.

Now let $n \geq 1$ and $0 < t_1 < \dots < t_n$. The random variables $N^{-1/2}M(Nt_1)$, $N^{-1/2}M(Nt_2) - N^{-1/2}M(Nt_1), \dots, N^{-1/2}M(Nt_n) - N^{-1/2}M(Nt_{n-1})$ are mutually independent and, if $B(t)$ denotes a standard one dimensional Brownian motion, the previous argument shows that, with $M(0) = B(0) = 0$, for any $1 \leq k \leq n$, $N^{-1/2}(M(Nt_k) - M(Nt_{k-1})) \Rightarrow B(t_k) - B(t_{k-1})$. Thus, since the random variables $B(t_1), B(t_2) - B(t_1), \dots, B(t_n) - B(t_{n-1})$ are mutually independent, we have shown that

$$\begin{aligned} & \left(\frac{M(Nt_1)}{\sqrt{N}}, \frac{M(Nt_2) - M(Nt_1)}{\sqrt{N}}, \dots, \frac{M(Nt_n) - M(Nt_{n-1})}{\sqrt{N}} \right) \\ & \Rightarrow (B(t_1), B(t_2) - B(t_1), \dots, B(t_n) - B(t_{n-1})) \end{aligned}$$

as $N \rightarrow \infty$. This proves that the finite dimensional distributions of the process $N^{-1/2}M(Nt)$ converge to those of $B(t)$. Together with tightness, this shows the lemma. \square

Proof of Proposition 2.3.1. With the notation of the previous lemma,

$$\mathcal{M}_j^N(t) = N^{-1/2}M_j \left(N \int_0^t \beta_j(s, Z_s^N) ds \right).$$

We write

$$\mathcal{M}_j^N(t) = N^{-1/2}M_j \left(N \int_0^t \beta_j(s, z_s) ds \right) + \widetilde{\mathcal{M}}_j^N(t),$$

where

$$\widetilde{\mathcal{M}}_j^N(t) = N^{-1/2}M_j \left(N \int_0^t \beta_j(s, Z_s^N) ds \right) - N^{-1/2}M_j \left(N \int_0^t \beta_j(s, z_s) ds \right).$$

For $C > 0$, let $\tau_{N,C} = \inf\{t > 0, |Z_t^N| > C\}$. We assume for a moment the identity

$$\mathbb{E} \left(\left| \widetilde{\mathcal{M}}_j^N(t \wedge \tau_{N,C}) \right|^2 \right) = \mathbb{E} \left(\left| \int_0^{t \wedge \tau_{N,C}} \beta_j(s, Z_s^N) ds - \int_0^t \beta_j(s, z_s) ds \right|^2 \right). \quad (2.3.2)$$

The above right-hand side is easily shown to converge to 0 as $N \rightarrow \infty$. Jointly with Doob's inequality from Proposition A.4.8 in the Appendix, this shows that for all $T > 0, \varepsilon > 0$,

$$\begin{aligned} \mathbb{P} \left(\sup_{0 \leq t \leq T} \left| \widetilde{\mathcal{M}}_j^N(t) \right| > \varepsilon \right) & \leq \mathbb{P}(\tau_{N,C} < T) + \mathbb{P} \left(\sup_{0 \leq t \leq T \wedge \tau_{N,C}} \left| \widetilde{\mathcal{M}}_j^N(t) \right| > \varepsilon \right) \\ & \leq \mathbb{P}(\tau_{N,C} < T) \\ & \quad + \frac{4}{\varepsilon^2} \mathbb{E} \left(\left| \int_0^{T \wedge \tau_{N,C}} \beta_j(s, Z_s^N) ds - \int_0^T \beta_j(s, z_s) ds \right|^2 \right). \end{aligned}$$

It follows from Theorem 2.2.7 that for $C > 0$ large enough, both terms on the right tend to 0, as $N \rightarrow \infty$. Consequently $\sup_{0 \leq t \leq T} \left| \widetilde{\mathcal{M}}_j^N(t) \right| \rightarrow 0$ in probability as $N \rightarrow \infty$.

It remains to note that an immediate consequence of Lemma 2.3.4 is that

$$N^{-1/2} M_j \left(N \int_0^t \beta_j(s, z_s) ds \right) \Rightarrow B_j \left(\int_0^t \beta_j(s, z_s) ds \right)$$

in the sense of weak convergence in the space $D((0, +\infty); \mathbb{R})$, and the coordinates are mutually independent. However the two processes $B_j(\int_0^t \beta_j(s, z_s) ds)$ and $\int_0^t \sqrt{\beta_j(s, z_s)} dB_j(s)$ are two centered Gaussian processes which have the same covariance functions. Hence they have the same law.

We finally need to establish (2.3.2). Following the development in Section A.2 in the Appendix, we can rewrite the local martingale $\widetilde{\mathcal{M}}_j^N(t)$ as follows, forgetting the index j , and the time parameter of β for the sake of simplifying notations

$$\begin{aligned} \widetilde{\mathcal{M}}^N(t) &= N^{-1/2} \int_0^t \int_0^\infty \mathbf{1}_{\{N\beta(z_s) \leq u \leq N\beta(Z_{s-}^N)\}} \bar{Q}(ds, du) \\ &\quad - N^{-1/2} \int_0^t \int_0^\infty \mathbf{1}_{\{N\beta(Z_{s-}^N) \leq u \leq N\beta(z_s)\}} \bar{Q}(ds, du), \end{aligned}$$

where $\bar{Q}(ds, du) = Q(ds, du) - dsdu$ and Q is a standard Poisson point measure on \mathbb{R}_+^2 . Noting that the square of each jump of the above martingale equals N^{-1} , we deduce from Proposition A.4.9 in the Appendix that

$$\begin{aligned} \mathbb{E} \left(\left| \widetilde{\mathcal{M}}^N(t \wedge \tau_{N,C}) \right|^2 \right) &= N^{-1} \mathbb{E} \int_0^{t \wedge \tau_{N,C}} \int_{N[\beta(Z_{s-}^N \vee \beta(z_s))]}^{N[\beta(Z_s^N \vee \beta(z_s))]} Q(ds, du) \\ &= N^{-1} \mathbb{E} \int_0^{t \wedge \tau_{N,C}} \int_{N[\beta(Z_s^N \wedge \beta(z_s))]}^{N[\beta(Z_s^N \vee \beta(z_s))]} dsdu, \end{aligned}$$

which yields (2.3.2). \square

Example 2.3.5. The *SIR* model. It is clear that Theorem 2.3.2 applies to Example 2.2.1. If we define $\begin{pmatrix} U_t \\ V_t \end{pmatrix} = \lim_{N \rightarrow \infty} \sqrt{N} \begin{pmatrix} S^N(t) - s(t) \\ I^N(t) - i(t) \end{pmatrix}$, we have

$$\begin{aligned} U_t &= -\lambda \int_0^t [i(r)U_r + s(r)V_r] dr - \int_0^t \sqrt{\lambda s(r)i(r)} dB_1(r), \\ V_t &= \int_0^t [\lambda(i(r)U_r + s(r)V_r) - \gamma V_r] dr + \int_0^t \sqrt{\lambda s(r)i(r)} dB_1(r) \\ &\quad - \int_0^t \sqrt{\gamma i(r)} dB_2(r). \end{aligned}$$

Remark 2.3.6. Consider now the *SIR* model, started with a fixed small number of infectious individuals, all others being susceptible, so that $(S^N(0), I^N(0)) \rightarrow (1, 0)$, as

$N \rightarrow \infty$. The solution of the ODE from Example 2.2.10 starting from $(s(0), i(0)) = (1, 0)$ is the constant $(s(t), i(t)) \equiv (1, 0)$. So in that case the coefficients of the noise in the last example are identically 0, and, the initial condition of the stochastic model being deterministic, it is natural to assume that $(U_0, V_0) = (0, 0)$. Then $(U_t, V_t) \equiv (0, 0)$. Consequently Theorem 2.3.2 tells us that, as $N \rightarrow \infty$, for any $T > 0$,

$$\sqrt{N} \begin{pmatrix} S^N(t) - 1 \\ I^N(t) - 0 \end{pmatrix} \rightarrow 0, \text{ in probability, uniformly w.r.t. } t \in [0, T].$$

In the case $R_0 > 1$, i.e. $\lambda > \gamma$, with positive probability the epidemic gets off. However, as we shall see in Section 3.4 below, this take time of the order of $\log(N)$, and there is no contradiction with the present result.

We close this section by a discussion of some of the properties of solutions of linear SDEs of the above type, following some of the developments in section 5.6 of Karatzas and Shreve [17]. Suppose that $\{A(t), t \geq 0\}$ and $\{C(t), t \geq 0\}$ are $d \times d$ matrix-valued measurable and locally bounded deterministic functions of t . With $\{B(t), t \geq 0\}$ being a d -dimensional Brownian motion, we consider the SDE

$$dX_t = A(t)X_t dt + C(t)dB_t, t \geq 0,$$

X_0 being a given d -dimensional Gaussian random vector independent of the Brownian motion $\{B(t)\}$. The solution to this SDE is the \mathbb{R}^d -valued process given by the explicit formula

$$X(t) = \Gamma(t, 0)X_0 + \int_0^t \Gamma(t, s)C(s)dB_s,$$

where the $d \times d$ matrix $\Gamma(t, s)$ is defined for all $0 \leq s \leq t$ as follows. For each fixed $s \geq 0$, $\{\Gamma(t, s), t \geq s\}$ solves the linear ODE

$$\frac{d\Gamma(t, s)}{dt} = A(t)\Gamma(t, s), \quad \Gamma(s, s) = I,$$

where I denotes the $d \times d$ identity matrix. It follows that $\{X_t, t \geq 0\}$ is a Gaussian process, and for each $t > 0$, the mean and the covariance matrix of the Gaussian random vector X_t are given by (denoting by C^* the transpose of the matrix C)

$$\mathbb{E}(X_t) = \Gamma(t, 0)\mathbb{E}(X_0),$$

$$\text{Cov}(X_t) = \Gamma(t, 0)\text{Cov}(X_0)\Gamma^*(t, 0) + \int_0^t \Gamma(t, s)C(s)C^*(s)\Gamma^*(t, s)ds.$$

Assume now that $A(t) \equiv A$ and $C(t) \equiv C$ are constant matrices. Then $\Gamma(t, s) = \exp((t-s)A)$. If we define $V(t) := \text{Cov}(X_t)$, we have that

$$V(t) = e^{tA} \left[V(0) + \int_0^t e^{-sA} C C^* e^{-sA^*} ds \right] e^{tA^*}.$$

If we assume moreover that all the eigenvalues of A have negative real parts, then it is not hard to show that as $t \rightarrow \infty$,

$$V(t) \rightarrow V := \int_0^\infty e^{sA} C C^* e^{sA^*} ds.$$

In that case the Gaussian law with mean zero and covariance matrix V is an invariant distribution of Gauss–Markov process X_t . This means in particular that if X_0 has that distribution, then the same is true for X_t for all $t > 0$. We now show the following result, which is often useful for computing the covariance matrix V in particular cases.

Lemma 2.3.7. *Under the above assumptions on the matrix A , V is the unique $d \times d$ positive semidefinite symmetric matrix which satisfies*

$$AV + VA^* + CC^* = 0.$$

Proof. Uniqueness follows from the fact that the difference \tilde{V} of two solutions satisfies $A\tilde{V} + \tilde{V}A^* = 0$. This implies that for all $x \in \mathbb{R}^d$, $\langle A\tilde{V}x, x \rangle = 0$. Since none of the eigenvalues of A^* is zero, this implies that $\langle \tilde{V}x, x \rangle = 0$ for all eigenvectors x of A^* , hence for all $x \in \mathbb{R}^d$. Since \tilde{V} is symmetric, this implies that $\tilde{V} = 0$.

To show that V satisfies the wished identity, assume that the law of X_0 is Gaussian with mean 0 and Covariance matrix V . Then V is also the covariance matrix of X_t . Consequently

$$V = e^{tA} V e^{tA^*} + \int_0^t e^{(t-s)A} C C^* e^{(t-s)A^*} ds.$$

Differentiating with respect to t , and letting $t = 0$ yields the result. \square

We leave the last result as an exercise for the reader.

Exercise 2.3.8. Consider again the case of time varying matrices $A(t)$ and $C(t)$. We assume that $A(t) \rightarrow A$ and $C(t) \rightarrow C$ as $t \rightarrow \infty$, and moreover that the real parts of all the eigenvalues of A are negative. Conclude that the law of X_t converges to the Gaussian law with mean 0 and covariance matrix V defined as above.

2.4 Diffusion Approximation

We consider again the vector of proportions in our model as

$$Z^N(t) = x + \frac{1}{N} \sum_{j=1}^k h_j P_j \left(\int_0^t N \beta_j(Z^N(s)) ds \right). \quad (2.4.1)$$

From the strong law of large numbers, $\sup_{0 \leq t \leq T} \|Z^N(t) - z_t\| \rightarrow 0$ almost surely as $N \rightarrow \infty$, for all $T > 0$, where z_t solves the ODE

$$\dot{z}_t = b(z_t), \quad z_0 = x; \quad \text{where } b(x) = \sum_{j=1}^k h_j \beta_j(x).$$

We now consider a diffusion approximation X_t^N of the above model, which solves the SDE

$$X_t^N = x + \int_0^t b(X_s^N) ds + \sum_{j=1}^k \frac{h_j}{\sqrt{N}} \int_0^t \sqrt{\beta_j(X_s^N)} dB_s^j,$$

where B^1, \dots, B^k are mutually independent standard Brownian motions. Let us define the Wasserstein-1 distance on the interval $[0, T]$ between two \mathbb{R}^d -valued processes U_t and V_t as

$$W_{1,T}(U, V) = \inf \mathbb{E}(\|U - V\|_T),$$

where, if $x : [0, T] \rightarrow \mathbb{R}^d$, $\|x\|_T = \sup_{0 \leq t \leq T} \|x(t)\|$, and the above infimum is over all couplings of the two processes $U(t)$ and $V(t)$, i.e. over all ways of defining jointly the two processes, while respecting the two marginal laws of U and V . We shall use the two following well-known facts about the Wasserstein distance: it is a distance (and satisfies the triangle inequality); if U_n is a sequence of random elements of $D([0, T]; \mathbb{R}^d)$ which converges in law to a continuous process U , and is such that the sequence of random variables $\|U_n\|_T$ is uniformly integrable, then $W_{1,T}(U_n, U) \rightarrow 0$ as $n \rightarrow \infty$.

The aim of this section is to establish the following theorem.

Theorem 2.4.1. *For all $T > 0$, as $N \rightarrow \infty$,*

$$\sqrt{N} W_{1,T}(Z^N, X^N) \rightarrow 0,$$

or in other words, $W_{1,T}(Z^N, X^N) = o(N^{-1/2})$.

Proof. We have proved in Theorem 2.2.7 that $\sup_{0 \leq t \leq T} \|Z_t^N - z_t\|_T \rightarrow 0$ almost surely, as $N \rightarrow \infty$, and moreover $\sqrt{N}(Z^N - z) \Rightarrow U$ as $N \rightarrow \infty$, where the above convergence holds for the topology of uniform convergence on the interval $[0, T]$, and U is the Gaussian process solution of the SDE

$$U_t = \int_0^t \nabla b(z_s) U_s ds + \sum_{j=1}^k h_j \int_0^t \sqrt{\beta_j(z_s)} dB_s^j.$$

It is not hard to prove the following.

Exercise 2.4.2. As $N \rightarrow \infty$, $\sup_{0 \leq t \leq T} \|X_t^N - z_t\|_T \rightarrow 0$ almost surely, and moreover $\sqrt{N}(X^N - z) \Rightarrow U$.

We first note that from the triangle inequality

$$\begin{aligned} W_{1,T}(\sqrt{N}(Z^N - z), \sqrt{N}(X^N - z)) &\leq W_{1,T}(\sqrt{N}(Z^N - z), U) \\ &\quad + W_{1,T}(\sqrt{N}(X^N - z), U) \\ &\rightarrow 0, \end{aligned}$$

as $N \rightarrow \infty$. Moreover

$$\begin{aligned}
 W_{1,T}(\sqrt{N}(Z^N - z), \sqrt{N}(X^N - z)) &= \inf_{\text{couplings}} \mathbb{E} \|\sqrt{N}(Z^N - z) - \sqrt{N}(X^N - z)\|_T \\
 &= \inf_{\text{couplings}} \sqrt{N} \mathbb{E} \|(Z^N - z) - (X^N - z)\|_T \\
 &= \inf_{\text{couplings}} \sqrt{N} \mathbb{E} \|Z^N - X^N\|_T \\
 &= \sqrt{N} W_{1,T}(Z^N, X^N).
 \end{aligned}$$

Theorem 2.4.1 follows from the two last computations. \square

Remark 2.4.3. If we combine the law of large numbers and the central limit theorem which have been established in the previous two sections, we conclude that $Z_t^N - z_t - N^{-1/2}U_t = o(N^{-1/2})$. In other words, if we replace Z_t^N by the Gaussian process $z_t + N^{-1/2}U_t$, the error we make, at least on any given finite time interval, is small compared to $N^{-1/2}$. The same is true for the diffusion approximation X_t^N .

Chapter 3

General Closed Models

In this chapter we go back to the general model, i.e. not assuming exponential latent and infectious periods implying that the epidemic process is Markovian. We consider models which are closed in the sense that there is no influx of new susceptibles during the epidemic. No birth, no immigration, and the removed individual are either dead or recovered, with an immunity which they do not lose in the considered time frame.

In this context, the epidemic will stop sooner or later. The questions of main interest are: the evaluation of the duration of the epidemic, and the total number of individuals which are ever infected. The first section gives exact results concerning the second issue in small communities. The rest of the chapter is concerned with large communities. We present the Sellke construction, and then use it to give a law of large number and a central limit theorem for the number of infected individuals. Finally we study the duration of the epidemic.

3.1 Exact results for the final size in small communities

In earlier sections it is often assumed that the population size N is large. In other situations this is not the case, for example in planned infectious disease experiments in veterinary science the number of studied animals is of the order 5–20 (e.g. Quenee et al. [29]), and in such cases law of large numbers and central limit theorems have not yet kicked in, which motivates the current section about exact results in small populations.

It turns out that it is quite complicated to derive expressions for the distribution of the final size, even when N is quite small. The underlying reason for this is that there are many ways in which an outbreak can result in exactly k initially susceptible individuals getting infected. We illustrate this by computing the final size distribution $\{p_k^{(N)}\}$ for the Reed–Frost model for $N = 1, 2$ and 3 . We then derive a recursive formula for the final outcome of the full model valid for general N and k (but numerically unstable for N larger than, say, 40).

Consider the Reed–Frost epidemic where the probability to infect a given susceptible equals p ($= 1 - e^{-\lambda t/N}$). And let $N = 1$, one susceptible and one infectious individual to start with. The possible values of Z are then 0 and 1, and obviously we have $p_0^{(1)} = \mathbb{P}(Z = 0|N = 1) = 1 - p$ and $p_1^{(1)} = p$. For $N = 2$ things are slightly more complicated. No one getting infected is easy: $p_0^{(2)} = (1 - p)^2$, since both individuals have to escape infection from the index case. For $Z = 1$ to occur, the index case must infect exactly one of the two remaining, but further, this individual must not infect the third person: $p_1^{(2)} = \binom{2}{1} p(1 - p) * (1 - p)$. Finally, the probability of $Z = 2$ is of course the complimentary probability, but it can also be obtained by considering the two possibilities for this to happen: either the index case infects both, or else the index case infects exactly one of the two, and that individual in turn infects the remaining individual: $p_2^{(2)} = p^2 + 2p(1 - p) * p$.

For $N = 3$ initial susceptibles the situation becomes even more complicated. It is best to write down the different epidemic generation chains at which individuals get infected. We always have one index case. The chain in which the index case infects two individuals who in turn together infect the last individual, is denoted $1 \rightarrow 2 \rightarrow 1 \rightarrow 0$. The probability for such a chain can be computed sequentially for each generation keeping in mind: how many susceptibles there are at risk, how many that get infected and what is the risk of getting infected (the complimentary probability of escaping infection). The probability for the chain just mentioned is given by

$$\mathbb{P}(1 \rightarrow 2 \rightarrow 1 \rightarrow 0|N = 3) = \binom{3}{2} p^2 (1 - p)^1 * (1 - (1 - p)^2).$$

The last factor comes from the final individual getting infected when there were two infected individuals in the previous generation (so the escape probability equals $(1 - p)^2$). We hence see that the probability of a chain is the product of (different) binomial probabilities. The final size probabilities are then obtained by writing down the different possible chains giving the desired final outcome:

$$\begin{aligned} p_0^{(3)} &= \mathbb{P}(1 \rightarrow 0) = (1 - p)^3 \\ p_1^{(3)} &= \mathbb{P}(1 \rightarrow 1 \rightarrow 0) = \binom{3}{1} p(1 - p)^2 * (1 - p)^2 \\ p_2^{(3)} &= \mathbb{P}(1 \rightarrow 2 \rightarrow 0) + \mathbb{P}(1 \rightarrow 1 \rightarrow 1 \rightarrow 0) \\ &= \binom{3}{2} p^2 (1 - p) * ((1 - p)^2) + \binom{3}{1} p(1 - p)^2 * \binom{2}{1} p(1 - p) * (1 - p) \\ p_3^{(3)} &= \mathbb{P}(1 \rightarrow 3 \rightarrow 0) + \mathbb{P}(1 \rightarrow 2 \rightarrow 1 \rightarrow 0) + \mathbb{P}(1 \rightarrow 1 \rightarrow 2 \rightarrow 0) \\ &\quad + \mathbb{P}(1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 0) \\ &= \dots \end{aligned}$$

Exercise 3.1.1. Compute $p_3^{(3)}$ explicitly by computing the probabilities of the different chains. Check that $\sum_{k=0}^3 p_3^{(k)} = 1$ for any $p \in [0, 1]$.

For general N it is possible to write down the outcome probability for a specific chain as follows. If we denote the number of susceptibles and infectives in generation k by (S_k, I_k) , then the epidemic starts with $(S_0, I_0) = (s_0, i_0) = (N, 1)$. From a chain $1 \rightarrow i_1 \rightarrow \dots i_j \rightarrow 0$ (so $i_{j+1} = 0$) the number of susceptibles in generation k is also known from the relation $s_k = s_0 - \sum_{j=1}^k i_j$. We use this when we compute the binomial probabilities of a given generation of the chain, these binomial probabilities depend on: how many were at risk, how many infectives there were in the previous generation, and how many to be infected in the current. Finally, the probability of a chain is the product of the different binomial probabilities of the different generations. From this we obtain the following so called chain-binomial probabilities

$$\mathbb{P}(1 \rightarrow i_1 \rightarrow \dots i_j \rightarrow 0) = \prod_{k=1}^{j+1} \binom{s_{k-1}}{i_k} (1 - (1-p)^{i_{k-1}})^{i_k} ((1-p)^{i_{k-1}})^{s_{k-1}-i_k}.$$

As seen, these expression are quite long albeit explicit. However, computing the final outcome probabilities $p_N(k)$, $k = 0, \dots, N$, is still tedious since there are many different possible chains resulting in exactly k getting infected at the end of the epidemic. Further, things become even more complicated when considering different distributions of the infectious period than a constant infectious period as is assumed for the Reed–Frost epidemic model.

However, it is possible to derive a recursive formula for the final number infected $p_N(k)$, see e.g. Ball [5], which we now show. The derivation of the recursion of the final size uses two main ideas: a Wald’s identity for the final size and the total infection pressure, and the exchangeability of individuals making it possible to express the probability of having k infections among the initially N susceptibles in terms of the probability of getting all k infected in the subgroup containing those k individuals and the index case, and the probability that the remaining $N - 1 - k$ individuals escape infection from that group.

Let us start with the latter. Fix N and write $\bar{\lambda} = \lambda/N$. As before we let Z^N denote the total number infected excluding the index case(s), explicitly showing the dependence on the number of initially susceptible N . Since individuals are exchangeable we can label the individuals according to the order in which they get infected. The index case is labelled 0, the individuals who get infected during the outbreak are labelled: $1, \dots, Z^N$, and those who avoid infection according to any order $Z^N + 1, \dots, N$. With this labelling we define the total infection pressure A^N by

$$A^N = \bar{\lambda} \sum_{i=0}^{Z^N} I_i \quad (3.1.1)$$

i.e. the infection pressure, exerted on any individual, during the complete outbreak (sometimes referred to as the “total cost” or the “severity” of the epidemic).

As earlier we let $p_i^{(N)} = \mathbb{P}(Z^N = i)$ denote the probability that exactly k initial susceptibles out of N get infected during the outbreak. Reasoning in terms of subsets among the initial susceptibles as described earlier, and using the exchangeability of individuals, it can be shown ([5]) that for any $i \leq k \leq N$,

$$\frac{p_i^{(N)}}{\binom{N}{i}} = \frac{p_i^{(k)}}{\binom{k}{i}} \mathbb{E} \left(e^{-(N-k)A^k} | Z^k = i \right). \quad (3.1.2)$$

The equation is explained as follows. On the left-hand side is the probability that a specific group of size i (out of N) get infected and no one else. On the right-hand side this event is divided into two sub events. This is done by considering another group of size $k \geq i$, containing the earlier specified group of size i as a subset. The first factor is then the probability that exactly the subgroup of size i get infected within the bigger group of size k . The second factor, the expectation, is the probability that all individuals outside the bigger subgroup avoid getting infected. The notation A^k and Z^k hence denote the total pressure and final size starting with k susceptibles.

We use the following steps to show Wald's identity recalling that $\psi_I(b) = \mathbb{E}(e^{bI})$ is the moment generating function of the infectious period (so $\psi_I(-b)$ is the Laplace transform)

$$\begin{aligned} (\psi_I(-\theta\bar{\lambda}))^{k+1} &= \mathbb{E} \left[\exp \left(-\theta\bar{\lambda} \sum_{i=0}^k I_i \right) \right] \\ &= \mathbb{E} \left[\exp \left(-\theta \left(A^k + \bar{\lambda} \sum_{i=Z^k+1}^k I_i \right) \right) \right] \\ &= \mathbb{E} \left[e^{-\theta A^k} (\psi_I(-\theta\bar{\lambda}))^{k-Z^k} \right]. \end{aligned}$$

The last identity follows since the $k - Z^k$ infectious periods I_{Z^k+1}, \dots, I_k , are mutually independent and also independent jointly of the total pressure A^k (which only depends on the first Z^k infectious periods and the contact processes of these individuals). If we now divide both sides by $(\psi_I(-\theta\bar{\lambda}))^{k+1}$ we obtain Wald's identity for Z^k and A^k :

$$\mathbb{E} \left(\frac{e^{-\theta A^k}}{(\psi_I(-\theta\bar{\lambda}))^{1+Z^k}} \right) = 1, \quad \theta \geq 0. \quad (3.1.3)$$

If we apply Wald's identity with $\theta = N - k$ and condition on the value of Z^k we get

$$\sum_{i=0}^k \frac{\mathbb{E} \left(e^{-(N-k)A^k} | Z^k = i \right)}{(\psi_I(-(N-k)\bar{\lambda}))^{i+1}} p_i^{(k)} = 1. \quad (3.1.4)$$

If we now use Equation (3.1.2) in the equation above we get

$$\sum_{i=0}^k \frac{\binom{k}{i} p_i^{(N)}}{\binom{N}{i} (\psi_I(-(N-k)\bar{\lambda}))^{i+1}} = 1.$$

Simplifying the equation, returning to $\lambda = \bar{\lambda}N$ and putting $p_k^{(N)}$ on one side, we obtain the recursive formula for the final size distribution $p_k^{(N)}, k = 0, \dots, N$.

Theorem 3.1.2. *The exact final size distribution is given by the recursive formula*

$$p_k^{(N)} = \binom{N}{k} [\psi_I(-(N-k)\lambda/N)]^{k+1} - \sum_{i=0}^{k-1} \binom{N-i}{k-i} [\psi_I(-(N-k)\lambda/N)]^{k-i} p_i^{(N)}. \quad (3.1.5)$$

For example, solving Equation (3.1.5) for $k = 0$ (when the sum is vacuous) and then for $k = 1$ gives, after some algebra,

$$\begin{aligned} p_0^{(N)} &= \psi_I(\lambda), \\ p_1^{(N)} &= N \psi_I\left(\frac{(N-1)\lambda}{N}\right) \\ &\quad \times \left[\psi_I\left(\frac{(N-1)\lambda}{N}\right) - \psi_I(\lambda) \right]. \end{aligned}$$

In order to compute $p_k^{(N)}$ using (3.1.5) it is required to sequentially compute $p_0^{(N)}$ up to $p_{k-1}^{(N)}$. Further, the formula is not very enlightening and it may be numerically very unstable when k (and hence $N \geq k$) is large. For this reason we devote the major part of these notes to approximations assuming N is large.

In Section 1.9 of Part II of this volume the exact results above are generalized to a model allowing for heterogeneous spreading, meaning that the transmission rate depends on the two individuals involved.

Exercise 3.1.3. Compute the final size distribution $\{p_k^{(N)}\}$ numerically using some suitable software for $N = 10, 50$ and 100 , for $\lambda = 2$ and $I \equiv 1$ (the Reed–Frost model) and $I \sim \Gamma(3, 1/3)$ (having mean 1 and variance $1/3$).

3.2 The Sellke construction

We now present the Sellke construction (Sellke [36]), which is an ingenious way to define the epidemic outbreak in continuous time using two sets of i.i.d. random variables. This elegant construction is made use of in many new epidemic models, as proven by having more than 50 citations in the past decade.

We number the individuals from 0 to N : 0 1 2 3 \dots N . Index 0 denotes the initially infected individual, and the individuals numbered from 1 to N are all susceptible at time 0.

Let

Q_1, Q_2, \dots, Q_N be i.i.d. random variables, with the law $\text{Exp}(1)$;

$(L_0, I_0), (L_1, I_1), \dots, (L_N, I_N)$ be i.i.d. random variables, with the law $\mathbb{P}_{(L,I)}$.

In the Markov model, L_i and I_i are independent, hence $\mathbb{P}_{(L,I)} = \mathbb{P}_L \otimes \mathbb{P}_I$,² where \mathbb{P}_L is the law of the latency period and \mathbb{P}_I that of the infectious period. But this need not be the case in more general non-Markov models.

Individual 0 has the latency period L_0 and the infectious period I_0 . We denote below

$L(t)$ the number of individuals in state E at time t ;

$I(t)$ the number of individuals in state I at time t .

Note that for each i , the two random variables L_i and I_i could be dependent, which typically is not the case in a Markov model.

We define the cumulative force of infection experienced by an individual, between times 0 and t as

$$\Lambda_C(t) = \frac{\lambda}{N} \int_0^t I(s) ds.$$

For $i = 1, \dots, N$, individual i is infected at the time when $\Lambda_C(t)$ achieves the value Q_i (which might be considered as the “level of resistance to infection of individual i ”). The j -th infected susceptible has the latency period L_j and the infectious period I_j . The epidemic stops when there is no individual in either the latent or infectious state, after which $\Lambda_C(t)$ does not grow any more, $\Lambda_C(t) = \Lambda_C(\infty)$. The individuals such that $Q_i > \Lambda_C(\infty)$ escape infection.

We put the Q_i s in increasing order: $Q_{(1)} < Q_{(2)} < \dots < Q_{(N)}$. It is the order in which individuals are infected in Sellke’s model. Note that Sellke’s model respects the durations of latency and infection. In order to show that Sellke’s construction gives a process which has the same law as the process from Definition 1.1.1, it remains to verify that the rates at which infections happen are the correct ones.

In the initial model, we assume that each infectious meets other individuals at rate c . Since each individual has the same probability of being the one who is met, the probability that a given individual is that one is $1/N$. Hence the rate at which a given individual is met by a given infectious one is c/N . Each encounter between a susceptible and an infectious individual achieves an infection with probability p . Hence the rate at which a given individual is infected by a given infectious individual is λ/N , where we have set $\lambda = cp$. The rate at which an infectious individual infects susceptibles is then $\lambda S(t)/N$. Finally the epidemic propagates at rate $\lambda S(t)I(t)/N$.

Let us go back to Sellke’s construction. At time t , $S(t)$ susceptibles have not yet been infected. Each of those corresponds to a $Q_i > \Lambda_C(t)$. At time t , the slope of the curve which represents the function $t \mapsto \Lambda_C(t)$ is $\lambda I(t)/N$. If $Q_i > \Lambda_C(t) = x$, then

² This notation stands for the product of the two probability measures \mathbb{P}_L and \mathbb{P}_I . The fact that the law of the pair is the product of the two marginals is equivalent to the fact that the two random variables L and I are independent.

$$\begin{aligned}
\mathbb{P}(Q_i > x+y | Q_i > x) &= e^{-y}, \\
\text{hence } \mathbb{P}(Q_i > \Lambda_C(t+s) | Q_i > \Lambda_C(t)) &= \exp\left(-\frac{\lambda}{N} \int_t^{t+s} I(r) dr\right) \\
&= \exp\left(-\frac{\lambda}{N} I(t)s\right),
\end{aligned}$$

if I is constant on the interval $[t, t+s]$. Consequently, conditionally upon $Q_i > \Lambda_C(t)$,

$$Q_i - \Lambda_C(t) \sim \text{Exp}\left(\frac{\lambda}{N} I(t)\right).$$

The same is true for all $S(t)$ of those Q_i which are $> \Lambda_C(t)$. The next individual to get infected corresponds to the minimum of those Q_i , hence the waiting time after t for the next infection follows the law $\text{Exp}\left(\frac{\lambda}{N} I(t) S(t)\right)$, if no removal of an infectious individual happens in the mean time, which would modify $I(t)$.

Thus in Sellke's construction, at time t the next infection comes at rate

$$\frac{\lambda}{N} I(t) S(t),$$

as in the model described above.

3.3 LLN and CLT for the final size of the epidemic

Define, for $0 \leq w \leq N+1$, with the notation $[w]$ = integer part of w , and the convention that a sum over an empty index set is zero,

$$\mathcal{J}(w) = \frac{\lambda}{N} \sum_{i=0}^{[w]-1} I_i.$$

Note that $i=0$ is the index of the initially infected individual, I_i denotes here the length of the infectious period of individual whose resistance level is $Q_{(i)}$ (who is not that of the i -th individual of the original list, but of the individual having the i -th smallest resistance).

$\mathcal{J}(w)$ is the infection pressure produced by the first $[w]$ infected individuals (including number 0). For any integer k , \mathcal{J} is of course constant on the interval $[k, k+1)$. Define for $v > 0$ the number of individuals who do not resist to the infectious pressure v :

$$\bar{q}(v) = \sum_{i=1}^N \mathbf{1}_{\{Q_i \leq v\}}.$$

The total number of infected individuals in the epidemic is

$$\begin{aligned}
Z &= \min \left\{ k \geq 0; Q_{(k+1)} > \frac{\lambda}{N} \sum_{i=0}^k I_i \right\} \\
&= \min \left\{ k \geq 0; Q_{(k+1)} > \mathcal{J}(k+1) \right\} \\
&= \min \left\{ w \geq 0; \bar{q}(\mathcal{J}(w+1)) = w \right\}.
\end{aligned} \tag{3.3.1}$$

Suppose indeed that $Z = i$. Then according to (3.3.1),

$$\begin{aligned}
&\mathcal{J}(j) > Q_{(j)}, \text{ hence } \bar{q}(\mathcal{J}(j)) \geq j, \text{ for all } j \leq i, \\
&\text{and } \mathcal{J}(i+1) < Q_{(i+1)} \text{ hence } \bar{q}(\mathcal{J}(i+1)) < i+1.
\end{aligned}$$

In other words $Z = i$ if and only if i is the smallest integer such that

$$\bar{q}(\mathcal{J}(i+1)) < i+1, \text{ hence } \bar{q}(\mathcal{J}(i+1)) = i.$$

3.3.1 Law of Large Numbers

Let us index \mathcal{J} and \bar{q} by N , the population size, so that they become \mathcal{J}_N and \bar{q}_N . We now define

$$\begin{aligned}
\overline{\mathcal{J}}_N(w) &= \mathcal{J}_N(Nw) \\
\bar{\bar{q}}_N(v) &= \frac{\bar{q}_N(v)}{N}.
\end{aligned}$$

It follows from the strong law of large numbers that as $N \rightarrow \infty$,

$$\begin{aligned}
\overline{\mathcal{J}}_N(w) &\rightarrow \lambda \mathbb{E}(I)w = R_0 w \text{ almost surely, and} \\
\bar{\bar{q}}_N(v) &\rightarrow 1 - e^{-v} \text{ a.s.}
\end{aligned}$$

Hence, with the notation $f \circ g(u) := f(g(u))$, as $N \rightarrow \infty$,

$$\bar{\bar{q}}_N \circ \overline{\mathcal{J}}_N(w) \rightarrow 1 - e^{-R_0 w}$$

a.s., uniformly on $[0, 1]$ (the uniformity in w follows from the second Dini theorem, as in the proof of Proposition 2.2.6). We have (replacing now Z by Z^N)

$$\begin{aligned}
\frac{Z^N}{N} &= \min \left\{ \frac{w}{N} \geq 0; \bar{q}_N(\mathcal{J}_N(w+1)) = w \right\} \\
&= \min \left\{ s \geq 0; \frac{1}{N} \bar{q}_N \left(\mathcal{J}_N \left(N \left(s + \frac{1}{N} \right) \right) \right) = s \right\} \\
&= \min \left\{ s \geq 0; \bar{\bar{q}}_N \left(\overline{\mathcal{J}}_N \left(s + \frac{1}{N} \right) \right) = s \right\}.
\end{aligned}$$

Recall from (1.1.1) that $R_0 = \lambda \iota$, where $\iota = \mathbb{E}(I)$. Note that when $R_0 > 1$, the equation

$$z = 1 - e^{-R_0 z} \quad (3.3.2)$$

(which is equation (2.1.2) from Section 2.1) has a unique solution $z^* \in (0, 1)$ (besides the zero solution). Indeed, $f(z) = 1 - e^{-R_0 z}$ is concave, $f(1) < 1$, and $f'(0) = R_0$. For the proof of the next theorem, we follow an argument from Andersson and Britton [2] (see also Ball and Clancy [6]).

Theorem 3.3.1. *If $R_0 \leq 1$, then $Z^N/N \rightarrow 0$ a.s., as $N \rightarrow \infty$.*

If $R_0 > 1$, as $N \rightarrow \infty$, Z^N/N converges in law to the random variable ζ which is such that $\mathbb{P}(\zeta = 0) = z_\infty = 1 - \mathbb{P}(\zeta = z^)$, where*

z_∞ , the probability of a minor outbreak (i.e. that the epidemic does not get off), is the solution in $(0, 1)$ of (3.3.3) below, and z^ is the positive solution of (3.3.2).*

Let us explain how one can characterize z_∞ . It follows from Theorem 1.2.5 that the probability z_∞ that the epidemic does not get off equals the probability that the associated branching process goes extinct, which is the probability that the associated discrete time branching process (where we consider the infected by generation) goes extinct. According to Proposition A.1.1 from Appendix A, the probability that this happens is the solution in the interval $(0, 1)$ of the equation $g(s) = s$, where g is the generating function of the random number ξ of individuals that one infected infects. As explained in Section 1.2, the law of ξ is $\text{MixPoi}(\lambda I)$, so if we denote by $\psi_I(\mu) = \mathbb{E}[\exp(-\mu I)]$ the Laplace transform of I , which is well defined for $\mu > 0$, then $g(s) = \psi_I(\lambda(1-s))$. Hence z_∞ is the unique solution in $(0, 1)$ of the equation

$$\psi_I(\lambda(1-s)) = s. \quad (3.3.3)$$

Proof. If $R_0 \leq 1$, then from Corollary 1.2.6, Z^N remains bounded, hence $Z^N/N \rightarrow 0$.

If $R_0 > 1$, then Z^N remains bounded with probability z_∞ , which is the probability of extinction in the branching process which approximates the early stage of the epidemic. We now need to see what happens on the complementary event. For that sake, we first choose an arbitrary sequence of integers t_N , which satisfies both $t_N/N \rightarrow 0$ and $t_N/\sqrt{N} \rightarrow \infty$, as $N \rightarrow \infty$. We note that on the event $\{Z^N \leq t_N\}$, each infective infects susceptibles at a rate which is bounded below by $\lambda_N = \lambda \frac{N+1-t_N}{N}$. Let $Z(\lambda_N, I)$ denote the total progeny of a single ancestor in a branching process, where each individual has children according to a rate λ_N Poisson process, during his life whose length is I . It is plain that for any $t \in \mathbb{Z}_+$, and N large enough such that $t \leq t_N$,

$$\mathbb{P}(B(\lambda, I) \leq t) \leq \mathbb{P}(Z^N \leq t) \leq \mathbb{P}(Z^N \leq t_N) \leq \mathbb{P}(B(\lambda_N, I) < \infty).$$

Define as in the statement $z_\infty = \mathbb{P}(B(\lambda, I) < \infty)$ the probability of extinction of the branching process approximating the early stage of the epidemic, and $z_{N,\infty} = \mathbb{P}(B(\lambda_N, I) < \infty)$. It is not hard to show that $z_{N,\infty} \rightarrow z_\infty$ as $N \rightarrow \infty$, as a consequence of the fact that $\lambda_N \rightarrow \lambda$ (since $t_N/N \rightarrow 0$). Hence for any $\varepsilon > 0$, we can choose t large enough such that $\mathbb{P}(B(\lambda, I) \leq t) \leq z_\infty - \varepsilon$, and N large enough such that

$z_{N,\infty} \leq z_\infty + \varepsilon$. We have shown that

$$\mathbb{P}(Z^N \leq t_N) \rightarrow z_\infty, \text{ as } N \rightarrow \infty. \quad (3.3.4)$$

This shows that a.s. on the event that the epidemic goes off, Z^N tends to ∞ faster than t_N . We will next prove that

$$\lim_{c \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{P} \left(\left\{ t_N < Z^N < Nz^* - c\sqrt{N} \right\} \cup \left\{ Z^N > Nz^* + c\sqrt{N} \right\} \right) = 0. \quad (3.3.5)$$

Recalling the last formula preceding the statement of the present theorem,

$$\begin{aligned} & \left\{ \frac{Z^N}{N} \in \left(t_N, z^* - \frac{c}{\sqrt{N}} \right) \cup \left(z^* - \frac{c}{\sqrt{N}}, 1 \right] \right\} \\ & \subset \left\{ \exists s \in \left(t_N, z^* - \frac{c}{\sqrt{N}} \right) \cup \left(z^* - \frac{c}{\sqrt{N}}, 1 \right] ; \bar{q}_N \left(\mathcal{J}_N \left(s + \frac{1}{N} \right) \right) = s \right\} \\ & \subset \left\{ \sup_{0 \leq s \leq 1} \left| \bar{q}_N \left(\mathcal{J}_N \left(s + \frac{1}{N} \right) \right) - 1 + e^{-R_0 s} \right| > \frac{\phi(c)}{\sqrt{N}} \right\}, \end{aligned} \quad (3.3.6)$$

where $\phi(c) \rightarrow \infty$, as $c \rightarrow \infty$, for N large enough. We have exploited the facts that $t_N/\sqrt{N} \rightarrow \infty$ as $N \rightarrow \infty$, and $f'(0) > 1$, $f'(z^*) < 1$. However, we shall see in the next subsection (see (3.3.7)) that

$$\left\{ \sqrt{N} \left(\bar{q}_N \left(\mathcal{J}_N \left(s + \frac{1}{N} \right) \right) - 1 + e^{-R_0 s} \right), s \in [0, 1] \right\}$$

converges weakly, for the sup-norm topology, to a centred Gaussian process with finite covariance, hence the limit as $N \rightarrow \infty$ of the probability of the event (3.3.6) tends to 0, as $c \rightarrow \infty$, which establishes (3.3.5). It is easily seen that the second part of the Theorem follows from the combination of (3.3.4) and (3.3.5). \square

We see that z^* is the size, measured as the proportion of the total population, of a “significant” epidemic, if it takes off, which happens with probability $1 - z_\infty$.

We notice that z^* depends on the particular model only through the quantity R_0 . In particular it depends on the law of the infectious period I only through its mean. In the case where both E and I are exponential random variables, we know from Section 2.2 that the model has a law of large numbers limit, which is a system of ODEs. The same value for z^* has been deduced from an analysis of this deterministic model in Section 2.1. The last theorem holds for a larger class of models.

3.3.2 Central Limit Theorem

From the classical CLT, as $N \rightarrow \infty$,

$$\begin{aligned} A_N(w) &:= \sqrt{N}(\overline{\mathcal{J}}_N(w) - R_0 w) = \frac{\lambda \sqrt{w}}{\sqrt{N} w} \sum_{i=0}^{[Nw]} [I_i - \mathbb{E}(I_i)] + O(1/\sqrt{N}) \\ &\Rightarrow A(w), \end{aligned}$$

where $A(w) \sim \mathcal{N}(0, p^2 c^2 \text{Var}(I)w)$. One can in fact show that, as processes

$$\{\sqrt{N}(\overline{\mathcal{J}}_N(w) - R_0 w), 0 \leq w \leq 1\} \Rightarrow \{A(w), 0 \leq w \leq 1\}$$

for the topology of uniform convergence, where $\{A(w), 0 \leq w \leq 1\}$ is a Brownian motion (i.e. a centered Gaussian process with independent increments and continuous trajectories) such that $\text{Var}(A(w)) = r^2 R_0^2 w$, where $r^2 = (\mathbb{E}I)^{-2} \text{Var}(I)$. It is easy to show that for all $k \geq 1$, all $0 < w_1 < \dots < w_k \leq 1$, if we define $A_N(w) := \sqrt{N}(\overline{\mathcal{J}}_N(w) - R_0 w)$,

$$(A_N(w_1), \dots, A_N(w_k)) \Rightarrow (A(w_1), \dots, A(w_k)).$$

This means the convergence of the finite dimensional distributions. Combining this with the techniques exposed in Section A.5 of the Appendix yields the above functional weak convergence.

Consider now \overline{q}_N . Again from the usual CLT,

$$\begin{aligned} B_N(v) &= \sqrt{N}(\overline{q}_N(v) - [1 - e^{-v}]) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N [\mathbf{1}_{\{Q_i \leq v\}} - (1 - e^{-v})] \\ &\Rightarrow B(v), \end{aligned}$$

where $B(v) \sim \mathcal{N}(0, e^{-v}(1 - e^{-v}))$. We have again a functional convergence, according to the Kolmogorov–Smirnov theorem, towards a time changed Brownian bridge. In simpler words, $\{B(v), v \geq 0\}$ is a centred Gaussian process with continuous trajectories whose covariance function is specified by the identity $\mathbb{E}[B(u)B(v)] = e^{-u \vee v} - e^{-(u+v)}$, where $u \vee v := \sup(u, v)$.

Let us now combine the two functional central limit theorems which we have just derived. We have

$$\begin{aligned} &\sqrt{N}(\overline{q}_N(\overline{\mathcal{J}}_N(w)) - 1 + e^{-R_0 w}) \\ &= \sqrt{N}(\overline{q}_N(\overline{\mathcal{J}}_N(w)) - 1 + \exp(-\overline{\mathcal{J}}_N(w))) + \sqrt{N}(e^{-R_0 w} - e^{-\overline{\mathcal{J}}_N(w)}) \\ &\sim B_N(\overline{\mathcal{J}}_N(w)) - R_0 e^{-R_0 w} A_N(w). \end{aligned}$$

Consequently

$$\sqrt{N} (\bar{q}_N(\bar{\mathcal{J}}_N(w)) - 1 + e^{-R_0 w}) B(R_0 w) - R_0 e^{-R_0 w} A(w), \quad (3.3.7)$$

which is the functional central limit theorem which was used in the proof of Theorem 3.3.1.

Recall that the above Law of Large Numbers has been obtained by taking the limit in the equation

$$\bar{q}_N(\bar{\mathcal{J}}_N(z + N^{-1})) = z.$$

Making use of the above two CLTs, we get

$$\begin{aligned} z &= 1 - e^{-\bar{\mathcal{J}}_N(z + N^{-1})} + N^{-1/2} B_N(\bar{\mathcal{J}}_N(z + N^{-1})) \\ &= 1 - \exp\left(-R_0(z + N^{-1}) - N^{-1/2} A_N(z + N^{-1})\right) \\ &\quad + N^{-1/2} B_N\left(R_0(z + N^{-1}) + N^{-1/2} A_N(z + N^{-1})\right). \end{aligned}$$

Let $z = z^* + z_N N^{-1/2} + o(N^{-1/2})$, where z^* satisfies $e^{-R_0 z^*} = 1 - z^*$. We obtain

$$\begin{aligned} &z^* + z_N N^{-1/2} + o(N^{-1/2}) \\ &= 1 - \exp\left(-R_0 z^* - R_0 z_N N^{-1/2} - A_N(z^*) N^{-1/2} + o(N^{-1/2})\right) \\ &\quad + N^{-1/2} B_N(R_0 z^*) + o(N^{-1/2}) \\ &= 1 - e^{-R_0 z^*} + N^{-1/2} e^{-R_0 z^*} (R_0 z_N + A_N(z^*)) + N^{-1/2} B_N(R_0 z^*) + o(N^{-1/2}). \end{aligned}$$

We simplify this relation by making use of the equation which specifies z^* . Multiplying the remaining terms by $N^{1/2}$, we deduce

$$[1 - (1 - z^*) R_0] z_N = B_N(R_0 z^*) + (1 - z^*) A_N(z^*) + o(1).$$

Hence $z_N \Rightarrow \Xi$, where (note that $e^{-R_0 z^*} (1 - e^{-R_0 z^*}) = z^* (1 - z^*)$)

$$\Xi \sim \mathcal{N}\left(0, \frac{z^* (1 - z^*)}{(1 - (1 - z^*) R_0)^2} (1 + r^2 (1 - z^*) R_0^2)\right),$$

where we have exploited the independence of the two processes $A(\cdot)$ and $B(\cdot)$, which follows from that of the two collections of random variables $(I_i, i \geq 0)$ and $(Q_i, i \geq 1)$.

Finally we can conclude with the following theorem. We refer to Scalia-Tomba [32] and [33] for a more complete justification.

Theorem 3.3.2. *As $N \rightarrow \infty$, conditionally upon the event that the epidemic takes off, the law of $N^{-1/2}(Z_N - N z^*)$ converges towards the Gaussian distribution*

$$\mathcal{N}\left(0, \frac{z^* (1 - z^*)}{(1 - (1 - z^*) R_0)^2} (1 + r^2 (1 - z^*) R_0^2)\right).$$

Exercise 3.3.3. Compute numerically the limiting mean and standard deviation of the final size Z^N in case of a major outbreak and $N = 1000$, $\lambda = 1.5$ and $\iota = 1$, for the following two situations. The first scenario is when $I \equiv 1$ (fixed infectious period), and the second when $I \sim \text{Exp}(1)$ (Markovian SIR).

3.4 The duration of the stochastic SEIR epidemic

Recall that $L^N(t)$ and $I^N(t)$ denote the numbers of latent and infectious individuals at time t respectively, and introduce $Z^N(t) = N - L^N(t) - I^N(t) - R^N(t)$ to denote the number of individuals who have been infected by time t (i.e. who are no longer susceptible). We now study how long it takes for the epidemic to first grow big, and then later to end, i.e. for the end of the epidemic we will study properties of $\tau^N = \inf\{t; L^N(t) + I^N(t) = 0\}$ as $N \rightarrow \infty$. It will only be a sketch since it is quite technical to prove the results rigorously. For detailed results we refer to Barbour [7]. From an applied point of view, this question has clear practical relevance, since for instance hospitals are on highest pressure when the epidemic peaks, and knowing how long until the outbreak is over indicates how long preventive measures should be enforced.

If the epidemic does not take off we know from branching process theory that the time to extinction is finite, so $\tau^N = O_p(1)$ on this part of the sample space ($O_p(1)$ denotes bounded in probability). We hence focus on the situation where the epidemic takes off resulting in a major outbreak, hence implicitly assuming that $R_0 > 1$.

We divide the duration of the whole epidemic τ^N into three parts: the beginning, the main part and the end of the epidemic. Pick $\varepsilon > 0$ small. Formally we define these parts by defining two intermediate times (inspired by Sir Winston Churchill): the end of the beginning $\tau_{Beg}^N = \inf\{t \leq \tau^N; Z^N(t) \geq \varepsilon N\}$, and the beginning of the end $\tau_{End}^N = \inf\{t \leq \tau^N; Z^N(t) \geq (1 - \varepsilon)z^*N\}$, where z^* is the positive solution to the final size equation from Section 2.1. Each of these times are equal to τ^N in the case when the event never occurs.

With these definitions the beginning of the epidemic is the time interval $[0, \tau_{Beg}^N)$, the main part $[\tau_{Beg}^N, \tau_{End}^N)$ and the end part $[\tau_{End}^N, \tau^N]$.

During the beginning we can sandwich the epidemic between two branching processes. The upper bound is the branching process $Z(t)$ described in Section 1.2. Similarly, we can construct a lower bound using a very similar branching process $Z^-(t)$, the only difference being that the birth rate is $\lambda(1 - \varepsilon)$ as opposed to λ for the upper branching process. This is true because before τ_{Beg}^N the rate of new infections in the epidemic equals $\lambda(1 - Z^N(t)/N)$ which lies between $\lambda(1 - \varepsilon)$ and λ . Since $Z^-(t) \leq Z^N(t) \leq Z(t)$ for $t \leq \tau_{Beg}^N$ it follows that $\tau_+^N \leq \tau_{Beg}^N \leq \tau_-^N$, where $\tau_+^N = \inf\{t; Z(t) \geq \varepsilon N\}$ and $\tau_-^N = \inf\{t; Z^-(t) \geq \varepsilon N\}$.

From Section A.1.2 we know the rate at which a branching process grows. More specifically, we know that when a branching process $Z'(t)$ takes off, it grows ex-

ponentially: $Z'(t) \sim e^{r't}$, where r' is the unique solution to $1 = \int_0^\infty e^{-r's} \lambda(s) ds = 1$, where $\lambda(s)$ is the average (expected) rate at which an individual gives birth at age s (cf. Equation (1.2.1)). For our two branching processes $Z(t)$ and $Z^-(t)$ we have $\lambda(s) = \lambda \mathbb{P}(L < s < L + I)$ and $\lambda^-(s) = \lambda(1 - \varepsilon) \mathbb{P}(L < s < L + I)$ respectively. From this it follows that the exponential growth rates r and r^- can be made arbitrary close to each other by choosing ε small enough ($r^- = r(1 + o(\varepsilon))$). The particular form of r and r^- depends on the distribution of L and I (see Exercise 3.4.2 below). Recall that $\tau_+^N = \inf\{t; Z(t) \geq \varepsilon N\}$, so the fact that $Z(t) \sim e^{rt}$ implies that $\tau_+^N = \frac{\log(\varepsilon N)}{r} + O_p(1)$. Similarly, $\tau_-^N = \frac{\log(\varepsilon N)}{r^-} + O_p(1)$. As a consequence, the two stopping times are arbitrary close to each other on the logarithmic scale. From this we have $\tau_{Beg}^N = \frac{\log(N)}{r} (1 + o(\varepsilon)) + O_p(1)$.

We now turn to the duration of the main part of the epidemic: $\tau_{End}^N - \tau_{Beg}^N$ which is positive only if the epidemic takes off, which we hence condition upon. During this part of the epidemic, the Markovian SEIR epidemic can be approximated by the deterministic SEIR model. This means that for the Markovian SEIR model, the duration of the main part of the epidemic $\tau_{End}^N - \tau_{Beg}^N$ can be well approximated by the corresponding duration of the deterministic system $\tau_{End}^{Det} - \tau_{Beg}^{Det}$. The deterministic system is started at $\tau_{Beg}^{Det} = 0$ with initial conditions $(s(0), e(0), i(0), r(0)) = (1 - \varepsilon, a\varepsilon, b\varepsilon, (1 - a - b)\varepsilon)$ for some positive numbers a and b with $0 < a + b \leq 1$ (there is no closed form expression for how the infected individuals are divided into exposed, infectives and recovered). The system is then run until $\tau_{End}^{Det} = \inf\{t; e(t) + i(t) + r(t) \geq (1 - \varepsilon)z\}$. We know that $z(t) = e(t) + i(t) + r(t) \rightarrow z$ and $z(t)$ is monotonically increasing (since $s(t) = 1 - z(t)$ is decreasing). This implies that $\tau_{End}^{Det} - \tau_{Beg}^{Det} = \tau_{Beg}^{Det}$ is just a constant for any fixed positive ε . It will depend slightly on a and b , but when ε is small the dependence is weak and there is a uniform bound. From this we conclude that the main part of the epidemic is bounded:

$$\tau_{End}^N - \tau_{Beg}^N = \tau_{Beg}^{Det} + o_p(\varepsilon) = O_p(1).$$

If the latent and infectious periods are not exponentially distributed, then the stochastic SEIR epidemic is not Markovian, and the deterministic approximating system is a difference-delay-system which we will not study more closely. The qualitative properties of this system coincide with those of the Markovian SEIR system; in particular, the duration of the main part is bounded in probability.

Just like the main part of the epidemic the duration of end of the epidemic, $\tau^N - \tau_{End}^N$ is only positive if the epidemic takes off, which we hence condition upon. At the beginning of the end part, the number of infected (either exposed, infectious or recovered) equals $Z^N(\tau_{End}^N) = (1 - \varepsilon)z^*N$ and $S^N(\tau_{End}^N) = (1 - z^*)N + \varepsilon z^*N$. Since ε is assumed to be small, infectious individuals give birth at rate $\lambda(1 - z^* + \varepsilon z^*) \approx (1 - z^*)$ during the rest of the epidemic (we know the final fraction infected converges to z^* in probability). Further, at the start of the beginning the fractions exposed and infectious will both close to that of the deterministic system which are both small, having size $c_E \varepsilon$ and $c_I \varepsilon$ say (cf. Figure 2.1.1 where it is seen that $e(t)$ and $i(t)$ are both small for large t). So, from the beginning of the end part, the

epidemic behaves like a branching process with childhood duration L , adult duration I and birth rate $\lambda(1 - z^*)$ during the adult life stage, and this part is started with $c_E \varepsilon N$ children (exposed) and $c_I \varepsilon N$ adults (infectious). The mean off-spring distribution for this branching process equals $\lambda \mathbb{E}(I)(1 - z^*) = R_0(1 - z)$ where z^* is the positive solution to $1 - z = e^{R_0 z}$. It can be shown (cf. Exercise 3.4.1 below) that $R_0(1 - z^*) < 1$ implying that the branching process is subcritical (otherwise the epidemic would not be on decline).

The duration $\tau^N - \tau_{End}^N$ of the end part can hence be approximated by the time until extinction of a subcritical branching process, starting with $c_E \varepsilon N$ children (exposed) and $c_I \varepsilon N$ adults (infectious). This branching process will have *negative* drift $r^* < 0$ being the solution to the corresponding equation $\int_0^\infty e^{-rs} \lambda(s) ds = 1$ where now $\lambda(s) = \lambda(1 - z^*) \mathbb{P}(L < s < L + I)$. So, $E(t) + I(t) \sim (E(0) + I(0))e^{r^* t} = (c_E + c_I) \varepsilon N e^{r^* t}$. The time until this branching process goes extinct (i.e. $E(t) + I(t) < 1$) is hence of order $-\log((c_E + c_I) \varepsilon N) / r^* = -\log N / r^* + O_p(1)$.

To sum up, the duration of the epidemic $\tau^N = O_p(1)$ if the epidemic does not take off, whereas it has the following structure in case it does take off:

$$\tau^N = \tau_{Beg}^N + (\tau_{End}^N - \tau_{Beg}^N) + (\tau^N - \tau_{End}^N) = \frac{\log N}{r} + O_p(1) + \frac{-\log N}{r^*}. \quad (3.4.1)$$

Note that the last term is also positive since $r^* < 0$.

Exercise 3.4.1. Show that $R_0(1 - z^*) < 1$ and compute it numerically for $R_0 = 1.5$.

Exercise 3.4.2. Consider the stochastic SEIR epidemic with infection rate $\lambda = 1.5$ per time unit. Compute the two leading terms of the duration of a major outbreak for the following three case: $L \equiv 0$ and $I \sim \text{Exp}(1)$ (Markovian SIR), $L \equiv 0$ and $I \equiv 1$ (continuous time Reed–Frost), and $L \sim \text{Exp}(1)$ and $I \sim \text{Exp}(1)$ (Markovian SEIR).

Chapter 4

Open Markov Models

In this chapter, contrary to the situation considered in earlier chapters, we study models where there is a constant supply of susceptibles (either by births, immigration or loss of immunity of the removed individuals) giving rise to endemic type situations. We study how the random fluctuations in the model can drive the system out of the basin of attraction of the stable endemic equilibrium of the deterministic model, such that the disease goes extinct.

As we shall see in Section 4.1, in the case of a moderate population size, one may expect that the Gaussian fluctuations described by the central limit theorem are strong enough to stop the endemy in a SIR model with demography. For larger population sizes, following Freidlin and Wentzell [13], we describe in Section 4.2 how long it will take for the random perturbations to stop the endemy. We apply this approach successively to the SIRS, the SIS and the SIR model with demography. In the case of the SIS model, we compute explicitly the constant which appears in the Freidlin–Wentzell theory, see Proposition 4.2.29 below. This is unfortunately the only case where we have such a simple and explicit formula in terms of the coefficients of the model.

4.1 Open populations: time to extinction and critical population size

Up until now we have (mainly) considered the stochastic SEIR epidemic model in a fixed community of size N , where N has been assumed large (except in Section 3.1 when N was assumed small). This is of course an approximation of reality, but when considering outbreaks of a few months (e.g. influenza outbreaks) it seems like a fair approximation; recall that the time to extinction of our model was $O_p(\log N)$. For other diseases including childhood diseases, the disease is present in the community constantly – such diseases are said to be *endemic*. When trying to understand the behaviour of such diseases it is necessary to also allow people to die and new people entering the population (by birth or immigration). In the current section we do

this and derive approximations for two important quantities: the time to extinction of the endemic disease T_E^N , and the critical community size N_c . These two quantities have received much attention in the literature over the years. In particular, the critical community size N_c and how it depends on properties of the disease and the community have been studied both in the mathematical and applied communities (e.g. Lindholm and Britton [23] and Keeling and Grenfell [18]).

Let us first describe the population model, which is the simplest model for a population which fluctuates randomly in time with a mean size of N individuals, and where individuals have life time distribution with mean $1/\mu$ (cf. Example 2.2.2). The population $N(t)$ is defined to be a Markovian birth-death process with constant birth rate μN and linear death rate, all individuals dying at rate μ . This process $N(t)$ will fluctuate around N , a parameter we denote by the mean population size. If N is large, it is known that $N(t)$ will be approximately normally distributed with mean N and standard deviation proportional to \sqrt{N} , so for practical purposes we will later approximate $N(t)$ by N .

Exercise 4.1.1. Assuming that $N(0) = N$, write $N(t)$ as the solution of an SDE of the same form as the SDE appearing at the beginning of Section 2.2. Define $Q_t^N = N(t)/N$ and show that, as a consequence of Theorem 2.2.7, $Q_t^N \rightarrow 1$ a.s., locally uniformly in t . Then deduce from Theorem 2.3.2 that $\sqrt{N}(Q_t^N - 1)$ converges weakly, as $N \rightarrow \infty$, towards an Ornstein–Uhlenbeck process of the form

$$U_t = \sqrt{2\mu} \int_0^t e^{-\mu(t-s)} dB_s,$$

where B_t is a standard Brownian motion. Prove that $\mathbb{E}(U_t) = 0$ and $\text{Var}(U_t) \rightarrow 1$ as $t \rightarrow \infty$. Deduce that for large N and t , $N(t)$ is approximately normally distributed with mean N and standard deviation proportional to \sqrt{N} .

For this population model, we assume that the Markovian SIR epidemic spreads (this can easily be extended to the Markovian SEIR model). By this we mean that individuals who get infected immediately become infectious and remain so for an $\text{Exp}(\gamma)$ time, unless they happen to die before by chance. In the fixed population size model, the contact rate was λ which implied that it was λ/N to each specific individual. Now, in the open population model, we assume that the infection rate to a specific individual is unchanged, λ/N . More appropriate would perhaps have been to instead have $\lambda/N(t)$ but since $N(t) \approx N$ for all t we use the simpler choice λ/N . So, new individuals enter the community at constant rate μN and all individuals die, irrespective of being susceptible, infectious or recovered, at rate μ , susceptible individuals get infected at rate $\lambda I^N(t)/N$, and infectious individuals recover at rate γ . The rate at which susceptibles get infected and infected recover hence equals $\lambda I^N(t)S^N(t)/N$, and $\gamma I^N(t)$ respectively. If we study the limiting deterministic system for the *fractions* in each state we get the following system of differential equations:

$$\begin{aligned}
s'(t) &= \mu - \lambda s(t)i(t) - \mu s(t), \\
i'(t) &= \lambda s(t)i(t) - \gamma i(t) - \mu i(t), \\
r'(t) &= \gamma i(t) - \mu r(t),
\end{aligned} \tag{4.1.1}$$

which is identical to those of Example 2.2.11 with $\rho = 0$ and $v = +\infty$. From this we can compute the endemic state where all derivatives are 0. First we note that the basic reproduction number R_0 (the expected number of infectious contact while infectious and alive) and the expected relative time of a life an individual is infected, ε , are given by

$$R_0 = \frac{\lambda}{\gamma + \mu} \quad \varepsilon = \frac{1/(\gamma + \mu)}{1/\mu} = \frac{\mu}{\gamma + \mu}. \tag{4.1.2}$$

The rate of recovery γ is much larger than the death rate μ (52 compared to 1/75 for a one week infectious period and 75 year life length) so for all practical purposes the two expressions can be approximated by $R_0 \approx \lambda/\gamma$ and $\varepsilon \approx \mu/\gamma$.

If we solve the system of differential equations (4.1.1) by setting all derivatives equal to 0, and replace μ , λ and γ by the dimensionless quantities R_0 and ε (three parameters can be replaced by two because the unit of time for the rates is arbitrary and one rate can be set to unity), we obtain the endemic level which is given by

$$(\hat{s}, \hat{i}, \hat{r}) = \left(\frac{1}{R_0}, \varepsilon \left(1 - \frac{1}{R_0} \right), 1 - \frac{1}{R_0} - \varepsilon \left(1 - \frac{1}{R_0} \right) \right) \tag{4.1.3}$$

Exercise 4.1.2. Show that this is the endemic level, i.e. that the solution solves Equation (4.1.1) with all derivatives being 0.

This state is only meaningful if $R_0 > 1$ (otherwise some fraction is negative), so the endemic level only exists if $R_0 > 1$. Another solution to the equation system is of course the disease free equilibrium $(s, i, r) = (1, 0, 0)$. It is well known that when $R_0 > 1$ (which we from now on assume), then the endemic state is globally stable whereas the disease free state is locally unstable, meaning the system converges to the endemic level irrespective of starting value as long as $i(0) > 0$.

Using the theory of Section 2.2 it can be shown that the current Markov model (for an open population) converges to the above deterministic model as $N \rightarrow \infty$, if the starting point is such that the fraction initially infectious is strictly positive ($I^N(0)/N \rightarrow i(0) > 0$).

This suggests that the stochastic model (for the fractions in different states) can be approximated by the corresponding deterministic function

$$(S^N(t)/N, I^N(t)/N, R^N(t)/N) \approx (s(t), i(t), r(t))$$

which solves Equation (4.1.1) and having the same initial condition as the stochastic system. And, since we know that $(s(t), i(t), r(t)) \rightarrow (\hat{s}, \hat{i}, \hat{r})$ as $t \rightarrow \infty$ this suggests that $(S^N(t)/N, I^N(t)/N, R^N(t)/N) \approx (\hat{s}, \hat{i}, \hat{r})$ when N and t are large. This is indeed true in some sense, but it is only true depending on the relation between N and

t . For any finite N , the stochastic epidemic, which fluctuates randomly around the endemic equilibrium, will eventually go extinct, meaning that for some random T_{Ext}^N (the extinction time) it will happen that $I^N(T_{Ext}^N) = 0$. When this happens the rate of new infections is 0 so the stochastic epidemic will remain disease free ever after (and eventually all removed will have died so all individuals are susceptible). Using large deviation theory (cf. Section 4.2 below) it can be shown that the time to extinction grows exponentially with N , $T_{Ext}^N \approx e^{cN}$ for some $c > 0$ as $N \rightarrow \infty$.

On the other hand, for any arbitrary but fixed time horizon $[0, t_{\max}]$ the stochastic epidemic will converge to the deterministic process as $N \rightarrow \infty$. It also follows from Theorem 2.3.2 that the scaled process

$$\sqrt{N}(S^N(t)/N - s(t), I^N(t)/N - i(t), R^N(t)/N - r(t))$$

converges to an Ornstein–Uhlenbeck process $(\tilde{S}(t), \tilde{I}(t), \tilde{R}(t))$. This Ornstein–Uhlenbeck process is a Gaussian process with stationary distribution being Normally distributed. In particular, the variance of $\tilde{I}(t)$ in stationarity is well approximated by $1/R_0 - 1/R_0^2$, see Nåsell [24].

Exercise 4.1.3. Show this as a consequence of Theorem 2.3.2, Lemma 2.3.7, and Exercise 2.3.8

This suggests that $I^N(t)$ will be approximately Gaussian with mean $N\hat{i}$ and standard deviation $\sqrt{N/R_0}$ when N is large and t is moderately large (smaller than T_{Ext}^N but still large since we assume the Ornstein–Uhlenbeck is close to stationary).

From above we know that T_{Ext}^N will grow exponentially with N as $N \rightarrow \infty$. On the other hand, if N is small or moderate, the disease will go extinct very quickly, e.g. within a year. We now use the Gaussian approximation above to define a sort of threshold, the *critical population size* N_c , between these two scenarios (quick extinction and very long time before extinction). Of course, there is no unique exact such value, so it will involve some arbitrary choice(s).

Above we noted that $I^N(t)$ was approximately Gaussian with mean $N\hat{i}$ and standard deviation $\sqrt{N/R_0}$. If we want to be above the critical population size, then we want to avoid quick extinction for which it is necessary that this approximately Gaussian process avoids extinction for a fairly long time. Extinction occurs when $I^N(t) = 0$, and if we want to avoid this we want the value 0 to be far enough away from the mean, e.g. at least 3 standard deviations away. The choice 3 is of course arbitrary but if we instead choose 2 the *process* will hit 0 fairly quickly with large enough probability, and if we choose 4 it seems extremely unlikely that it will hit extinction within e.g. a life time, so 3 seems like a reasonable compromise when it is unlikely but not completely impossible. This choice then suggests that the threshold is for the case $N\hat{i} - 3\sqrt{N/R_0} = 0$. This is equivalent to $\sqrt{N} = 3/\hat{i}\sqrt{R_0}$, i.e. $N = 9/\hat{i}^2 R_0$. Inserting that $\hat{i} = \varepsilon(1 - 1/R_0)$ (remember that $\varepsilon = \mu/(\gamma + \mu)$ is the relative length of the infectious period compared to life-length), then we arrive at our definition of the *critical population size* N_c :

$$N_c = \frac{9}{\varepsilon^2(1 - \frac{1}{R_0})^2 R_0}. \quad (4.1.4)$$

The conclusion is that, for a given infectious disease, i.e. given R_0 and ε , the disease will die out quickly in a community of size $N \ll N_c$ whereas it will persist for a very long time if $N \gg N_c$, during which the disease is endemic. As an illustration, consider measles prior to vaccination. If we assume that $R_0 \approx 15$ and the infectious period is 1 week (1/52 years) and life duration 75 years, implying that $\varepsilon \approx \frac{1/75}{1/(1/52)+1/75} \approx 1/3750$ we arrive at $N_c \approx 9(3750)^2/15 \approx 8 \cdot 10^6$. So, if the population is a couple of million (or less) the disease will go extinct quickly, whereas the disease will become endemic (for a very long time) in a population being larger than e.g. 20 million people. This confirms the empirical observation that measles was continuously endemic in UK whereas it died out quickly in Iceland (and was later reintroduced by infectious people visiting the country).

Exercise 4.1.4. Which parameter affects N_c the most? Compute N_c using the measles example but making R_0 50% bigger/smaller and the same for the duration of the infectious period (assuming we live equally long).

Exercise 4.1.5. Suppose that a vaccine giving 100% life long immunity is available, and that a fraction v of all infants are continuously vaccinated. How does this affect the critical community size, i.e. give an expression for N_c also containing v . (*Hint:* Vaccinating people affects both the relevant population size N_v , the non-vaccinated population, and the reproduction number R_v , but other than that nothing has changed.)

4.2 Large deviations and extinction of an endemic disease

4.2.1 Introduction

In Section 2.2, we have proved that, under appropriate conditions, the solution of the SDE

$$Z_t^N = x_N + \sum_{j=1}^k \frac{h_j}{N} P_j \left(N \int_0^t \beta_j(s, Z_s^N) ds \right) \quad (4.2.1)$$

converges a.s., locally uniformly in t , towards the unique solution of the ODE

$$\frac{dz_t}{dt} = b(t, z_t), \quad z_0 = x, \quad (4.2.2)$$

see Theorem 2.2.7, where $b(t, x) = \sum_{j=1}^k h_j \beta_j(t, x)$. Consequently the above SDE (4.2.1) can be considered for large N as a small random perturbation of the ODE (4.2.2). Small random perturbations of ODEs by Brownian motion have been studied by many authors, starting with Freidlin and Wentzell [13]. Our aim is to study the above type of random perturbations of an ODE like (4.2.2). The starting point is the estimation of a large deviation from the law of large numbers, which has been studied for our type of Poisson driven SDEs by Shwartz and Weiss [34]. The

difficulty is the fact that some of the rates in the SDE (4.2.1) vanish when the solution hits part of the boundary. This makes the estimate a bit delicate, since the logarithms of the rates enter the rate function in our large deviations estimate. This situation has been addressed first by Shwartz and Weiss [35], but their assumptions are not quite satisfied in our framework. Recently Kratz and Pardoux [21] and Pardoux and Samegni-Kepgnou [26] have developed an approach to Large Deviations which is well adapted to the epidemics models which are considered in these Notes. In fact the main difficulty concerns the lower bound. In the following, we present a new approach to the lower bound, based upon a quasi-continuity result, Proposition 4.2.4 below, which mimics a similar result for Brownian motion driven SDEs due to Azencott [4]. The same approach, for other types of Poisson driven SDEs, will soon appear in Kouegou-Kamen and Pardoux [19], [20].

The main application we have in mind is to estimate the time needed for the small random perturbations to drive the system from a stable endemic equilibrium to the disease free equilibrium (i.e. extinction). This applies to the classical SIS and SIRS models, as well as to an SIR model with demography, as well as to models with vaccination and to models with several levels of susceptibility, thus predicting the time it will take for the random perturbation to end an endemic disease.

We rewrite our model as

$$Z_t^{N,x_N} = x_N + \sum_{j=1}^k h_j \int_0^t \int_0^{\beta_j(s, Z_s^{N,x_N}-)} Q_j^N(ds, du),$$

where

$$Q_j^N(ds, du) = \frac{1}{N} Q_j(ds, Ndu),$$

and the Q_j 's are i.i.d. Poisson random measures on $[0, T] \times \mathbb{R}_+$, with mean λ^2 , the 2-dimensional Lebesgue measure.

(A.1) We shall assume in all of this section that the β_j 's are locally Lipschitz with respect to x , uniformly for $t \in [0, T]$.

4.2.2 The rate function

We want to establish a large deviations principle for trajectories in the space $D([0, T]; \mathbb{R}^d)$ of \mathbb{R}^d -valued right-continuous functions which have a left limit at any time $t \in (0, T]$. We shall also consider the sets $C([0, T]; \mathbb{R}^d)$ of continuous functions from $[0, T]$ into \mathbb{R}^d , and the subset of absolutely continuous functions, which we will denote $\mathcal{AC}_{T,d}$. For any $\phi \in \mathcal{AC}_{T,d}$, let $\mathcal{A}_k(\phi)$ denote the (possibly empty) set of functions $c \in L^1(0, T; \mathbb{R}_+^k)$ such that $c_j(t) = 0$ a.e. on the set $\{t, \beta_j(\phi_t) = 0\}$ and

$$\frac{d\phi_t}{dt} = \sum_{j=1}^k c_j(t) h_j, \quad \text{t a.e.}$$

We define the rate function

$$I_T(\phi) := \begin{cases} \inf_{c \in \mathcal{A}_k(\phi)} I_T(\phi|c), & \text{if } \phi \in \mathcal{AC}_{T,A}; \\ \infty, & \text{otherwise.} \end{cases}$$

where as usual the infimum over an empty set is $+\infty$, and

$$I_T(\phi|c) = \int_0^T \sum_{j=1}^k g(c_j(t), \beta_j(\phi_t)) dt$$

with $g(v, \omega) = v \log(v/\omega) - v + \omega$. We assume in the definition of $g(v, \omega)$ that for all $v > 0$, $\log(v/0) = \infty$ and $0 \log(0/0) = 0 \log(0) = 0$.

We consider I_T as a functional defined on the space $D([0, T]; \mathbb{R}^d)$ equipped with Skorokhod's topology. We first give two other possible definitions of the functional I_T . Let $\ell : \mathbb{R}^{3d} \mapsto \mathbb{R}$ be defined as

$$\ell(x, y, \theta) = \langle y, \theta \rangle - \sum_{j=1}^k \beta_j(x) \left(e^{\langle h_j, \theta \rangle} - 1 \right).$$

We define the map $L : \mathbb{R}^{2d} \mapsto (-\infty, +\infty]$ as

$$L(x, y) = \sup_{\theta \in \mathbb{R}^d} \ell(x, y, \theta).$$

We let

$$\hat{I}_T(\phi) = \int_0^T L(\phi_t, \dot{\phi}_t) dt.$$

It is not hard to see that the following is an equivalent definition of $\hat{I}_T(\phi)$:

$$\hat{I}_T(\phi) = \sup_{\theta \in C^1([0, T]; \mathbb{R}^d)} \int_0^T \ell(\phi_t, \dot{\phi}_t, \theta_t) dt.$$

We first establish

Proposition 4.2.1. *For any $\phi \in D([0, T]; \mathbb{R}^d)$, $I_T(\phi) = \hat{I}_T(\phi)$.*

Proof. We note that if $y = \sum_{j=1}^k c_j h_j$ with some $c \in \mathbb{R}_+^k$,

$$\ell(x, y, \theta) = \sum_{j=1}^k \left[c_j \langle h_j, \theta \rangle - \beta_j(x) \left(e^{\langle h_j, \theta \rangle} - 1 \right) \right].$$

But for any $1 \leq j \leq k$,

$$c_j \langle h_j, \theta \rangle - \beta_j(x) \left(e^{\langle h_j, \theta \rangle} - 1 \right) \leq \sup_{r \in \mathbb{R}} [c_j r - \beta_j(x) (e^r - 1)]$$

$$\begin{aligned}
&= c_j \log \left(\frac{c_j}{\beta_j(x)} \right) - c_j + \beta_j(x) \\
&= g(c_j, \beta_j(x)).
\end{aligned}$$

The inequality $\hat{I}_T(\phi) \leq I_T(\phi)$ for any $\phi \in D([0, T]; \mathbb{R}^d)$ follows readily.

In order to prove the converse inequality, we fix $x, y \in \mathbb{R}^d$ such that $L(x, y) < \infty$ (otherwise there is nothing to prove). Let θ_n be a sequence in \mathbb{R}^d such that $L(x, y) = \lim_{n \rightarrow \infty} \ell(x, y, \theta_n)$. It is clear that for any $1 \leq j \leq k$ such that $\beta_j(x) > 0$, the sequence $\langle \theta_n, h_j \rangle$ is bounded from above. Hence we can and do assume that, after the extraction of a subsequence, for any $1 \leq j \leq k$ such that $\beta_j(x) > 0$, the sequence $e^{\langle \theta_n, h_j \rangle} \rightarrow s_j$, for some $s_j \geq 0$. Consequently, as $n \rightarrow \infty$,

$$\langle \theta_n, y \rangle \rightarrow L(x, y) + \sum_{j=1}^k \beta_j(x) (s_j - 1). \quad (4.2.3)$$

Differentiating $\ell(x, y, \theta_n)$ with respect to its last variable, we get

$$\begin{aligned}
\nabla_{\theta} \ell(x, y, \theta_n) &= y - \sum_{j=1}^k \beta_j(x) e^{\langle h_j, \theta_n \rangle} h_j \\
&\rightarrow y - \sum_{j=1}^k \beta_j(x) s_j h_j,
\end{aligned}$$

as $n \rightarrow \infty$. But since θ_n is a maximizing sequence and the gradients converge, then since $L(x, y) < \infty$, their limit must be zero. Consequently

$$y = \sum_{j=1}^k \beta_j(x) s_j h_j.$$

Hence, with $c_j = \beta_j(x) s_j$, we have

$$\begin{aligned}
\langle \theta_n, y \rangle &= \sum_{j=1}^k c_j \langle \theta_n, h_j \rangle \\
&\rightarrow \sum_{j=1}^k c_j \log(s_j),
\end{aligned}$$

with the convention that $c_j \log(s_j) = 0$ if both $c_j = 0$ and $s_j = 0$. This, combined with (4.2.3), yields that

$$L(x, y) = \sum_{j=1}^k g(c_j, \beta_j(x))$$

which entails that $\hat{I}_T(\phi) \geq I_T(\phi)$. The proposition is established. \square

We have the

Proposition 4.2.2. *For any $T > 0$, $\phi \in D([0, T]; \mathbb{R}^d)$, $I_T(\phi) \geq 0$, and $I_T(\phi) = 0$ iff ϕ solves the ODE (4.2.2).*

Proof. It suffices to show that $L(x, y) \geq L(x, \sum_j \beta_j(x) h_j) = 0$, with strict inequality if $y \neq \sum_j \beta_j(x) h_j$. We first note that

$$L\left(x, \sum_j \beta_j(x) h_j\right) = \sup_{\theta} \left\{ \sum_j \beta_j(x) (\langle h_j, \theta \rangle - \exp \langle h_j, \theta \rangle + 1) \right\} = 0,$$

since $z - e^z + 1 \leq 0$, with equality at $z = 0$. Let now y be such that $L(x, y) = 0$. Then

$$\langle y, \theta \rangle - \sum_j \beta_j(x) (\exp \langle h_j, \theta \rangle - 1) \leq 0 \text{ for all } \theta \in \mathbb{R}^d.$$

Choosing $\theta = \varepsilon e_i$ (where e_i is the i -th basis vector of \mathbb{R}^d) yields

$$\varepsilon y_i \leq \sum_j \beta_j(x) (\exp(\varepsilon h_j^i) - 1).$$

Dividing by ε , then letting $\varepsilon \rightarrow 0$ yields $y_i \leq \sum_j \beta_j(x) h_j^i$, while the opposite inequality follows if we start with $\theta = -\varepsilon e_i$. The result follows. \square

In the next statement, we use the notion of a lower semi-continuous real-valued function, which is defined in Definition A.7.1 below. In the proof we use the notion of an equicontinuous collection of functions, which is defined in Definition A.7.2.

Theorem 4.2.3. *$\phi \rightarrow I_T(\phi)$ is lower semi-continuous on $D([0, T]; \mathbb{R}^d)$, and for any $R, K > 0$, the set $\{\phi \in D([0, T]; \mathbb{R}^d), \sup_{0 \leq t \leq T} |\phi_t| \leq R, I_T(\phi) \leq K\}$ is compact.*

Proof. The lower semicontinuity property is an immediate consequence of the fact that, from its second definition, \hat{I}_T is a supremum over continuous functions. To finish the proof, it suffices from the Arzelà–Ascoli theorem (see e.g. Theorem 7.2 in Billingsley [8]) to show that the set of functions satisfying $\sup_{0 \leq t \leq T} |\phi_t| \leq R$ and $I_T(\phi) \leq K$ is equicontinuous. It is clear that if $\bar{h} = \sup_{1 \leq j \leq k} |h_j|$ and $\bar{\beta}_R = \sup_{1 \leq j \leq k} \sup_{0 \leq t \leq T, |x| \leq R} \beta_j(t, x)$,

$$\begin{aligned} L(x, y) &\geq \ell\left(x, y, \frac{y \log(|y|)}{\bar{h}|y|}\right) \\ &\geq \frac{|y| \log(|y|)}{\bar{h}} - k \bar{\beta}_R |y|. \end{aligned}$$

Now let $0 \leq s < t \leq T$, with $t - s \leq \delta$.

$$\begin{aligned} |\phi_t - \phi_s| &\leq \int_s^t |\dot{\phi}_r| dr \\ &\leq \delta^{-1/2} \int_s^t \mathbf{1}_{|\dot{\phi}_r| \leq \delta^{-1/2}} dr + \int_s^t \mathbf{1}_{|\dot{\phi}_r| > \delta^{-1/2}} \frac{L(\phi_r, \dot{\phi}_r)}{L(\phi_r, \dot{\phi}_r)/|\dot{\phi}_r|} dr \end{aligned}$$

$$\leq \delta^{1/2} + \frac{K}{f(\delta^{-1/2})},$$

where $f(a) = \inf_{|x| \leq R, |y| \geq a} \frac{L(x,y)}{|y|}$. The result follows from the fact that from the above lower bound of $L(x,y)$, $f(a) \rightarrow \infty$, as $a \rightarrow \infty$. \square

4.2.3 The lower bound

Let $\eta = (\eta_1, \dots, \eta_k)$ be a vector of locally finite measures on $[0, T] \times \mathbb{R}_+$. We shall say that $\eta \in \mathcal{M}^k$. To $x \in \mathbb{R}^d$ and $\eta \in \mathcal{M}^k$, we associate $\Phi_t^x(\eta)$, solution (if it exists) of the ODE

$$\Phi_t^x(\eta) = x + \sum_{j=1}^k h_j \int_0^t \int_0^{\beta_j(s, \Phi_s^x)} \eta_j(ds, du).$$

If $\eta_j(ds, du) = f_j(s, u)dsdu$, $1 \leq j \leq k$, the above ODE has at least one solution (possibly up to an explosion time, as the solution of an ODE with continuous coefficients). If moreover

$$\sup_{u \geq 0} f_j(\cdot, u) \in L^1[0, T], \quad 1 \leq j \leq k,$$

then the above ODE has a unique solution (as the solution of an ODE with locally Lipschitz coefficients).

Let $\phi \in C([0, T]; \mathbb{R}^d)$ be an absolutely continuous function. We define

$$K_\phi := \inf_{c \in \mathcal{A}_k(\phi)} \sum_{j=1}^k \int_0^T \frac{c_j(t)}{\beta_j(t, \phi_t)} dt. \quad (4.2.4)$$

To a pair (ϕ, c) with $c \in \mathcal{A}_k(\phi)$, we associate for $1 \leq j \leq k$ the measure $\eta_j(ds, du)$ with the density

$$f_j(s, u) = \frac{c_j(s)}{\beta_j(s, \phi_s)} \mathbf{1}_{[0, \beta_j(s, \phi_s)]}(u) + \mathbf{1}_{(\beta_j(s, \phi_s), +\infty)}(u).$$

Then, with $x = \phi_0$, $\phi_t = \Phi_t^x(\eta)$.

Moreover, given $\phi \in C([0, T]; \mathbb{R}^d)$ and $L > 0$, we consider the set

$$A_{\phi, L} = \{(t, x), \quad 0 \leq t \leq T, \quad |x - \phi_t| \leq L + 1\},$$

and define

$$\bar{\beta}(\phi, L) = \sup_{1 \leq j \leq k} \sup_{(t, x) \in A_{\phi, L}} \beta_j(t, x).$$

We can now prove the following.

Proposition 4.2.4. *Let $T > 0$ be arbitrary. Given (ϕ, η) as above, such that in particular $K_\phi < \infty$, if $x_N = Z_0^N$, for any $R, L > 0$, there exists a $\delta, r > 0$ (depending upon K_ϕ) and N_0 such that whenever $|x - x_N| \leq r$, $N \geq N_0$,*

$$\mathbb{P} \left(\|Z^N - \phi\|_T > L, d_{T, \bar{\beta}}(Q^N, \eta) \leq \delta \right) \leq e^{-NR},$$

where

$$d_{T, \bar{\beta}}(\nu, \eta) = \sum_{j=1}^k \sup_{0 \leq t \leq T, 0 \leq u \leq \bar{\beta}} |\nu_j([0, t] \times [0, u]) - \eta_j([0, t] \times [0, u])|,$$

and $\bar{\beta} := \bar{\beta}(\phi, L)$.

Proof. It is clear that

$$\begin{aligned} |Z_t^N - \phi_t| &\leq |x_N - x| + \sum_{j=1}^k |h_j| \left| \int_0^t \int_0^{\beta_j(s, Z_{s-}^N)} [Q_j^N(ds, du) - \eta_j(ds, du)] \right| \\ &\quad + \sum_{j=1}^k |h_j| \left| \int_0^t \int_{\beta_j(s, Z_{s-}^N) \wedge \beta_j(s, \phi_s)}^{\beta_j(s, Z_{s-}^N) \vee \beta_j(s, \phi_s)} f_j(s, u) \vee 1 \, du \, ds \right| \\ &\leq r + \sum_{j=1}^k |h_j| \left| \int_0^t \int_0^{\beta_j(s, Z_{s-}^N)} [Q_j^N(ds, du) - \eta_j(ds, du)] \right| \\ &\quad + \sum_{j=1}^k |h_j| C \int_0^t \left(\frac{c_j(s)}{\beta_j(s, \phi_s)} \vee 1 \right) |Z_s^N - \phi_s| \, ds, \end{aligned}$$

where C is an upper bound of the Lipschitz constants of the β_j 's in $[0, T] \times [0, \bar{\beta}]$. Subdividing $[0, T]$ into $\left\lceil \frac{T}{\rho} \right\rceil + 1$ intervals of the form $[(i-1)\rho, i\rho \wedge T]$ and denoting

$$\bar{\beta}_j^i := \sup_{(i-1)\rho \leq s \leq i\rho} \beta_j(s, Z_{s-}^{N, x^N}), \quad \underline{\beta}_j^i := \inf_{(i-1)\rho \leq s \leq i\rho} \beta_j(s, Z_{s-}^{N, x^N}),$$

we define the random sets

$$A_j^{\rho, i} := [(i-1)\rho, i\rho] \times [0, \underline{\beta}_j^i], \quad B_j^{\rho, i} := [(i-1)\rho, i\rho] \times [\underline{\beta}_j^i, \bar{\beta}_j^i].$$

For all i and j ,

$$\sum_{j=1}^k |Q_j^N(A_j^{\rho, i}) - \eta_j(A_j^{\rho, i})| \leq 2d_{T, \bar{\beta}}(Q^N, \eta), \quad \sum_{j=1}^k |Q_j^N(B_j^{\rho, i}) - \eta_j(B_j^{\rho, i})| \leq 4d_{T, \bar{\beta}}(Q^N, \eta).$$

Consequently for all $0 \leq t \leq T$, if $\bar{h} := \sup_{1 \leq j \leq k} |h_j|$, then on the event $\{d_{T, \bar{\beta}}(Q^N, \eta) \leq \delta\}$,

$$\begin{aligned}
& \sum_{j=1}^k |h_j| \left| \int_0^t \beta_j(s, Z^{N, s-}) [\mathcal{Q}_j^N(ds, du) - \eta_j(ds, du)] \right| \\
& \leq \bar{h} \sum_{j=1}^k \left(\sum_{i=1}^{\lfloor \frac{t}{\rho} \rfloor + 1} |\mathcal{Q}_j^N(A_j^{\rho, i}) - \eta_j(A_j^{\rho, i})| + \sum_{i=1}^{\lfloor \frac{t}{\rho} \rfloor + 1} \{ \mathcal{Q}_j^N(B_j^{\rho, i}) + \eta_j(B_j^{\rho, i}) \} \right) \\
& \leq \bar{h} \sum_{j=1}^k \left(\sum_{i=1}^{\lfloor \frac{t}{\rho} \rfloor + 1} |\mathcal{Q}_j^N(A_j^{\rho, i}) - \eta_j(A_j^{\rho, i})| + \sum_{i=1}^{\lfloor \frac{t}{\rho} \rfloor + 1} |\mathcal{Q}_j^N(B_j^{\rho, i}) - \eta_j(B_j^{\rho, i})| \right. \\
& \quad \left. + 2 \sum_{i=1}^{\lfloor \frac{t}{\rho} \rfloor + 1} \eta_j(B_j^{\rho, i}) \right) \\
& \leq 6 \left(\frac{t}{\rho} + 1 \right) \bar{h} \delta + 2\bar{h} \sum_{j=1}^k \sum_{i=1}^{\lfloor \frac{t}{\rho} \rfloor + 1} \eta_j(B_j^{\rho, i}).
\end{aligned}$$

It follows from the two above inequalities and Gronwall's Lemma that

$$\sup_{0 \leq t \leq T} |Z_t^N - \phi_t| \leq \left(r + 6 \left(\frac{T}{\rho} + 1 \right) \bar{h} \delta + 2\bar{h} \sum_{j=1}^k \sum_{i=1}^{\lfloor \frac{T}{\rho} \rfloor + 1} \eta_j(B_j^{\rho, i}) \right) \exp [C(K_\phi + kT)\bar{h}]. \quad (4.2.5)$$

Since the $(B_j^{\rho, i})_i$ are disjoint we have for all j

$$\begin{aligned}
\sum_{i=1}^{\lfloor \frac{T}{\rho} \rfloor + 1} \eta_j(B_j^{\rho, i}) &= \eta_j \left(\bigcup_{i=1}^{\lfloor \frac{T}{\rho} \rfloor + 1} B_j^{\rho, i} \right) \leq \sum_{i=1}^{\lfloor \frac{T}{\rho} \rfloor + 1} (\bar{\beta}_j^i - \underline{\beta}_j^i) \int_{(i-1)\rho}^{i\rho} \frac{c_j(s)}{\beta_j(\phi_s)} \vee 1 \, ds \\
&\leq \max_{1 \leq i \leq \lfloor \frac{T}{\rho} \rfloor + 1} (\bar{\beta}_j^i - \underline{\beta}_j^i) \int_0^T \frac{c_j(s)}{\beta_j(\phi_s)} \vee 1 \, ds \\
&\leq (K_\phi + T) \max_{1 \leq i \leq \lfloor \frac{T}{\rho} \rfloor + 1} (\bar{\beta}_j^i - \underline{\beta}_j^i).
\end{aligned}$$

We note that for every i, j

$$\bar{\beta}_j^i - \underline{\beta}_j^i \leq C \frac{X_i}{N}$$

where X_i is a Poisson random variable of mean $\rho N \bar{\beta}$. For any $a > 0$, we have with $\bar{a} = \frac{a}{k(K_\phi + T)}$, using Cramér's Theorem A.3.1 for the fourth inequality,

$$\begin{aligned}
\mathbb{P} \left[\sum_{j=1}^k \sum_{i=1}^{\left[\frac{T}{\rho}\right]+1} \eta_j(B_j^{\rho,i}) > a \right] &\leq k \max_j \mathbb{P} \left[\sum_{i=1}^{\left[\frac{T}{\rho}\right]+1} \eta_j(B_j^{\rho,i}) > \frac{a}{k} \right] \\
&\leq k \mathbb{P} \left[\max_{1 \leq i \leq \left[\frac{T}{\rho}\right]+1} \frac{X_i}{N} > \bar{a} \right] \\
&\leq k \mathbb{P} \left[\bigcup_{1 \leq i \leq \left[\frac{T}{\rho}\right]+1} \left\{ \frac{X_i}{N} > \bar{a} \right\} \right] \\
&\leq k \left(\frac{T}{\rho} + 1 \right) \exp \left(-N \left[\bar{a} \log \frac{\bar{a}}{\rho \bar{\beta}} + \bar{a} - \rho \bar{\beta} \right] \right) \\
&= \exp \left(-N \left[\bar{a} \log \frac{\bar{a}}{\rho \bar{\beta}} + \bar{a} - \frac{1}{N} \log \left(k \left[\frac{T}{\rho} + 1 \right] \right) - \rho \bar{\beta} \right] \right).
\end{aligned} \tag{4.2.6}$$

We choose $\rho = \sqrt{\delta}$. Let δ_0 be such that

$$\begin{aligned}
6 \left(T \sqrt{\delta_0} + \delta_0 \right) \bar{h} &\leq \frac{L}{3} \exp \left[-C(K_\phi + kT) \bar{h} \right], \text{ and} \\
r &= \frac{L}{3} \exp \left[-C(K_\phi + kT) \bar{h} \right], \\
a &= \frac{L}{6\bar{h}} \exp \left[-C(K_\phi + kT) \bar{h} \right],
\end{aligned}$$

so that from (4.2.5),

$$\left\{ \sum_{j=1}^k \sum_{i=1}^{\left[\frac{T}{\rho}\right]+1} \eta_j(B_j^{\rho,i}) \leq a \right\} \subset \left\{ \|Z^N - \phi\|_T \leq L \right\}. \tag{4.2.7}$$

$R > 0$ being arbitrary, we now choose

$$\begin{aligned}
\delta &= \min \left\{ \delta_0, \left(\frac{\bar{a}}{\bar{\beta}} \right)^2 e^{-2R/\bar{a}}, \frac{\bar{a}}{2\bar{\beta}} \right\}, \text{ and} \\
N_0 &= \left\lceil \frac{2}{\bar{a}} \log \left(k \left[\frac{T}{\rho} + 1 \right] \right) \right\rceil.
\end{aligned}$$

The result follows from those choices, (4.2.6) and (4.2.7). \square

Before we establish the lower bound, we need to formulate an assumption.

(A.2) We assume that for any $\phi \in C([0, T]; \mathbb{R}^d)$ such that $I_T(\phi) < \infty$ and any $\varepsilon > 0$, there exists a ϕ^ε such that $\phi_0^\varepsilon = \phi_0$, $K_{\phi^\varepsilon} < \infty$, $\|\phi - \phi^\varepsilon\|_T \leq \varepsilon$ and $I_T(\phi^\varepsilon) \leq I_T(\phi) + \varepsilon$.

Exercise 4.2.5. Consider the SIRS model with fixed population size, and let $A := \{(x, y), 0 \leq x, 0 \leq y, x + y \leq 1\}$. Show that if $\phi \in C([0, T]; A)$ hits the boundary, then for any $\varepsilon > 0$, one can find ϕ^ε such that $\phi_0^\varepsilon = \phi_0$, $K_{\phi^\varepsilon} < \infty$, $\|\phi - \phi^\varepsilon\| \leq \varepsilon$ and $I_T(\phi^\varepsilon) \leq I_T(\phi) + \varepsilon$, where ϕ^ε can either remain in the interior of A , or else can hit the boundary.

We now have, with the notation $I_{T,x}(O) = \inf_{\phi \in O, \phi_0=x} I_T(\phi)$,

Theorem 4.2.6. *If the assumptions (A.1) and (A.2) are satisfied, then for any open subset $O \subset D([0, T]; \mathbb{R}^d)$, if $x_N \rightarrow x$ as $N \rightarrow \infty$,*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^{N, x_N} \in O) \geq -I_{T,x}(O).$$

Proof. It clearly suffices to treat the case where $I_{T,x}(O) < \infty$. Then for any $\varepsilon > 0$ there exists a $\phi \in O$ such that $\phi_0 = x$ and

$$I_T(\phi) \leq I_{T,x}(O) + \frac{\varepsilon}{4}.$$

It follows from assumption (A.2) that there exists a $\hat{\phi} \in O$ such that $\hat{\phi}_0 = \phi_0$, $K_{\hat{\phi}} < \infty$, $\|\hat{\phi} - \phi\|_T \leq \varepsilon$ and

$$I_T(\hat{\phi}) \leq I_T(\phi) + \frac{\varepsilon}{4}.$$

Now there exists a $c \in \mathcal{A}_k(\phi)$ such that $\sum_{j=1}^k \int_0^T \frac{c_j(t)}{\beta_j(t, \phi_t)} dt < \infty$, and

$$I_T(\hat{\phi}|c) \leq I_T(\hat{\phi}) + \frac{\varepsilon}{4}.$$

If ε has been chosen small enough, there exists an $L > 0$ be such that $\{\psi; \|\psi - \hat{\phi}\|_T < L\} \subset O$. From Proposition 4.2.4, if η^c denotes the vector of measures associated to c , $|x - x_N|$ is small enough and N large enough, for any $R > 0$, there exists a $\delta > 0$ such that with $\hat{\beta} = \bar{\beta}(\hat{\phi}, L)$,

$$\begin{aligned} \mathbb{P}(Z^{N, x_N} \in O) &\geq \mathbb{P}(\|Z^{N, x_N} - \phi\|_T < L) \\ &\geq \mathbb{P}(d_{T, \hat{\beta}}(Q^N, \eta^c) < \delta) \\ &\quad - \mathbb{P}(\|Z^{N, x_N} - \phi\|_T > L, d_{T, \hat{\beta}}(Q^N, \eta^c) < \delta) \\ &\geq \mathbb{P}(d_{T, \hat{\beta}}(Q^N, \eta^c) < \delta) - e^{-NR}. \end{aligned} \tag{4.2.8}$$

Let us admit for a moment the next lemma.

Lemma 4.2.7. *There exists a sequence of partitions $\{A_n^i, 1 \leq i \leq a_n\}$ of $[0, T] \times [0, \hat{\beta}]$ such that $\sup_i \lambda^2(A_n^i) \rightarrow 0$ as $n \rightarrow \infty$, and a sequence $\delta_n \downarrow 0$ and n_0 such that for all $n \geq n_0$,*

$$\bigcap_{j=1}^k \bigcap_{i=1}^{a_n} \{Q_j^N(A_n^i) \in (\eta_j^c(A_n^i) - \delta_n, \eta_j^c(A_n^i) + \delta_n)\} \subset \{d_{T,\hat{\beta}}(Q^N, \eta^c) < \delta\}.$$

As a consequence of this lemma, making use of Cramér's Theorem A.3.1 for the second inequality,

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left(d_{T,\hat{\beta}}(Q^N, \eta^c) < \delta \right) \\ &\geq \sum_{j=1}^k \sum_{i=1}^{a_n} \liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} (Q_j^N(A_n^i) \in (\eta_j^c(A_n^i) - \delta_n, \eta_j^c(A_n^i) + \delta_n)) \\ &\geq - \sum_{j=1}^k \sum_{i=1}^{a_n} \left(\eta_j^c(A_n^i) \log \frac{\eta_j^c(A_n^i)}{\lambda^2(A_n^i)} - \eta_j^c(A_n^i) + \lambda^2(A_n^i) \right) \\ &\geq - \sum_{j=1}^k \int_0^T \int_0^{\bar{\beta}} [f_j^c(s, u) \log[f_j^c(s, u)] - f_j^c(s, u) + 1] ds du - \frac{\varepsilon}{4} \\ &= - \sum_{j=1}^k \int_0^T \left[c_j(s) \log \frac{c_j(s)}{\beta_j(s, \phi_s)} - c_j(s) + \beta_j(s, \phi_s) \right] ds - \frac{\varepsilon}{4} \\ &= -I_T(\hat{\phi}|c) - \frac{\varepsilon}{4} \\ &\geq -I_{T,x}(O) - \varepsilon, \end{aligned}$$

where

$$f_j^c(s, u) = \frac{c_j(s)}{\beta_j(s, \phi_s)} \mathbf{1}_{[0, \beta_j(s, \phi_s)]}(u) + \mathbf{1}_{(\beta_j(s, \phi_s), +\infty)}(u)$$

and the second inequality holds true for n chosen large enough as a function of ε . We let $\varepsilon \rightarrow 0$, and to combine the resulting inequality with (4.2.8), hence

$$\begin{aligned} -I_{T,x}(O) &\leq \liminf_{N \rightarrow \infty} \frac{1}{N} \log (\mathbb{P}(Z^{N,x_N} \in O) + e^{-NR}) \\ &\leq \left(\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^{N,x_N} \in O) \right) \vee (-R). \end{aligned}$$

The result finally follows by letting $R \rightarrow \infty$. \square

We now need to pass to the

Proof of Lemma 4.2.7. For convenience, we replace the partition $\{A_n^i, 1 \leq i \leq a_n\}$ by a partition $\{A_n^{i,j}, 1 \leq i, j \leq n\}$, which we construct as follows. We first choose $0 = \beta_n^0 < \beta_n^1 < \dots < \beta_n^n = \hat{\beta}$ such that

$$\sup_{1 \leq j \leq n} \eta^c([0, T] \times (\beta_n^{j-1}, \beta_n^j]) \leq \frac{2}{n} \eta^c([0, T] \times [0, \hat{\beta}]).$$

We next choose a sequence $0 = t_n^0 < t_n^1 < \dots < t_n^n = T$ such that, if $A_n^{i,j} = (t_n^{i-1}, t_n^i] \times (\beta_n^{j-1}, \beta_n^j]$,

$$\sup_{1 \leq i \leq n} \eta^c(A_n^{i,j}) \leq \frac{2}{n} \eta^c([0, T] \times (\beta_n^{j-1}, \beta_n^j]) \leq \frac{4}{n^2} \eta^c([0, T] \times [0, \hat{\beta}]) := \frac{C}{n^2}.$$

For an arbitrary $0 \leq t \leq T$ and $0 \leq \alpha \leq \hat{\beta}$, we define the set

$$\partial_{t,\alpha} = \{t\} \times [0, \hat{\beta}] \cup [0, T] \times \{\alpha\},$$

which is the “boundary” of $[0, t] \times [0, \alpha]$. We note that $|\{i, j, A_n^{i,j} \cap \partial_{t,\alpha} \neq \emptyset\}| \leq 2n$. We need to bound

$$\begin{aligned} & \left| Q^N([0, t] \times [0, \alpha]) - \eta^c([0, t] \times [0, \alpha]) \right| \\ & \leq \sum_{i,j, A_n^{i,j} \subset [0, t] \times [0, \alpha]} |Q^N(A_n^{i,j}) - \eta^c(A_n^{i,j})| + \sum_{i,j, A_n^{i,j} \cap \partial_{t,\alpha} \neq \emptyset} (Q^N(A_n^{i,j}) + \eta^c(A_n^{i,j})) \\ & \leq n^2 \delta_n + 2n \left(\frac{2C}{n^2} + \delta_n \right) \\ & \leq \delta, \end{aligned}$$

for all $n \geq n_0$, provided we choose first $n_0 \geq \frac{8C}{\delta}$, and then a sequence δ_n such that $\delta_n \leq [2(n^2 + 2n)]^{-1} \delta$ for each $n \geq n_0$. \square

We now establish a slightly stronger result. Here and below we shall use the following notation concerning the initial condition of Z^N . We fix $x \in \mathbb{R}^d$ and start Z^N from the point $Z_0^N = x_N$, where the i -th coordinate x_N^i of x_N is given by $x_N^i = \frac{[x^i N]}{N}$. Here we assume that the process Z^N lives in a closed subset $A \subset \mathbb{R}^d$. We shall need the following

Definition 4.2.8. We shall say that the compact set of initial conditions \mathcal{K} is **adapted** to the open set of trajectories $O \subset D([0, T]; A)$ if

1. $\mathcal{K} \subset \{\phi_0, \phi \in O\}$.
2. For any $\varepsilon > 0$, the following holds. For any $x \in \mathcal{K}$, there exists a $\phi^x \in O$ such that $\phi_0^x = x$, $I_T(\phi^x) \leq I_{T,x}(O) + \varepsilon$ and moreover $\sup_{x \in \mathcal{K}} K_{\phi^x} < \infty$.

It follows readily from the proof of Theorem 4.2.6 that the following reinforced version holds.

Theorem 4.2.9. For any open subset $O \subset D([0, T]; A)$ and any compact subset \mathcal{K} of initial conditions which is adapted to O ,

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \inf_{x \in \mathcal{K}} \mathbb{P}(Z^{N, x_N} \in O) \geq - \sup_{x \in \mathcal{K}} I_{T,x}(O).$$

4.2.4 The upper bound

In this subsection, we shall again use the notation x_N for the vector whose i -th coordinate is given by $x_N^i = \frac{[x^i N]}{N}$. We want to prove that for any closed F , $F \subset D([0, T]; \mathbb{R}^d)$,

$$\limsup_{N \rightarrow \infty} \log \mathbb{P}(Z^{N, x_N} \in F) \leq -I_{T, x}(F). \quad (4.2.9)$$

Let us recall the concept of exponential tightness.

Definition 4.2.10. The sequence Z^N is said to be exponentially tight if for any $\alpha > 0$, there exists a compact K^α such that

$$\limsup_N \frac{1}{N} \log \mathbb{P}(Z^N \in K_\alpha^c) \leq -\alpha.$$

We have the following lemma.

Lemma 4.2.11. If (4.2.9) holds for any compact subset $F = K \subset D([0, T]; A)$, and Z^N is exponentially tight, then (4.2.9) holds for any closed subset $F \subset D([0, T]; A)$.

Proof. Let F be closed and $\alpha := I_{T, x}(F)$. We assume w.l.o.g. that $\alpha > 0$ (unless the conclusion below would be obvious). Let K_α be the compact set associated to α by Definition 4.2.10. It is clear that $F \cap K_\alpha$ is compact and $I_{T, x}(F \cap K_\alpha) \geq \alpha$. Hence from our assumption

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^N \in F \cap K_\alpha) \leq -\alpha.$$

Also from the choice of K_α ,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^N \in K_\alpha^c) \leq -\alpha.$$

But $\mathbb{P}(Z^N \in F) \leq \mathbb{P}(Z^N \in F \cap K_\alpha) + \mathbb{P}(Z^N \in K_\alpha^c)$, hence

$$\log \mathbb{P}(Z^N \in F) \leq \log 2 + \sup(\log \mathbb{P}(Z^N \in F \cap K_\alpha), \log \mathbb{P}(Z^N \in K_\alpha^c)),$$

and we clearly deduce that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^N \in F) \leq -\alpha,$$

as desired. □

Let us first establish

Theorem 4.2.12. Let $T > 0$ and $x \in \mathbb{R}^d$ be fixed. Let $x_N \rightarrow x$ as $N \rightarrow \infty$. For any compact set $K \subset D([0, T]; \mathbb{R}^d)$,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^{N,x_N} \in K) \leq -I_{T,x}(K).$$

Proof. Recall the formula

$$\begin{aligned} I_T(\phi) &= \sup_{\theta \in C^1([0,T];\mathbb{R}^d)} \int_0^T \ell(\phi_t, \dot{\phi}_t, \theta_t) dt \\ &= \sup_{\theta \in C^1([0,T];\mathbb{R}^d)} \mathcal{L}(\phi, \theta), \end{aligned}$$

where

$$\mathcal{L}(\phi, \theta) = \langle \phi_T, \theta_T \rangle - \langle \phi_0, \theta_0 \rangle - \int_0^T \langle \phi_t, \dot{\theta}_t \rangle dt - \sum_{j=1}^k \int_0^T \beta_j(\phi_t) \left[e^{\langle h_j, \theta_t \rangle} - 1 \right] dt.$$

For any $\theta \in C^1([0,T];\mathbb{R}^d)$, $0 \leq s < t \leq T$, we define

$$\begin{aligned} M_{s,t}^{N,\theta} &= \langle Z_t^{N,x_N}, \theta_t \rangle - \langle Z_s^{N,x_N}, \theta_s \rangle - \int_s^t \langle Z_r^{N,x_N}, \dot{\theta}_r \rangle dr - \sum_{j=1}^k \int_s^t \langle h_j, \theta_r \rangle \beta_j(r, Z_r^{N,x_N}) dr, \\ \Xi_{s,t}^{N,\theta} &= \exp \left(NM_{s,t}^{N,\theta} - N \sum_{j=1}^k \int_s^t \tau(\langle h_j, \theta_r \rangle) \beta_j(r, Z_r^{N,x_N}) dr \right), \end{aligned}$$

where $\tau(a) = e^a - 1 - a$, are such that $M_{0,t}^{N,\theta}$ and $\Xi_{0,t}^{N,\theta}$ are local martingales, the second being also a supermartingale such that $\mathbb{E}[\Xi_{0,t}^{N,\theta}] \leq 1$.

We assume that $I_{T,x}(K) > 0$, since otherwise the result is trivial. We also assume that $I_{T,x}(K) < \infty$. The case $I_{T,x}(K) = \infty$ can be treated in a way which is very similar to what follows, and we will not repeat the argument. Since $\phi \mapsto I_T(\phi)$ is lower semicontinuous and $K_x = \{\phi \in K, \phi_0 = x\}$ is compact, there exists a $\hat{\phi} \in K$ such that $\hat{\phi}_0 = x$ and $I_T(\hat{\phi}) = I_{T,x}(K)$. Let now $\phi \in K_x$ be arbitrary. First assume that $I_T(\phi) < \infty$. Then there exists a $\theta_\phi \in C^1([0,T];\mathbb{R}^d)$ such that

$$I_T(\phi) \leq \mathcal{L}(\phi, \theta_\phi) + \frac{\varepsilon}{2}.$$

Since $\psi \mapsto \mathcal{L}(\psi, \theta_\phi)$ is continuous on $D([0,T];\mathbb{R}^d)$ equipped with the Skorokhod topology, there exists a neighbourhood $\mathcal{V}_{\phi, \theta_\phi}(\varepsilon)$ of ϕ in $D([0,T];\mathbb{R}^d)$ such that for any $\psi \in \mathcal{V}_{\phi, \theta_\phi}(\varepsilon)$,

$$|\mathcal{L}(\phi, \theta_\phi) - \mathcal{L}(\psi, \theta_\phi)| \leq \frac{\varepsilon}{2}.$$

Now

$$\mathbb{P}(Z^{N,x_N} \in \mathcal{V}_{\phi, \theta_\phi}(\varepsilon)) = \mathbb{E} \left(\mathbf{1}_{Z^{N,x_N} \in \mathcal{V}_{\phi, \theta_\phi}(\varepsilon)} \right)$$

$$\begin{aligned}
&= e^{-N\mathcal{L}(\phi, \theta_\phi)} \mathbb{E} \left(e^{N\mathcal{L}(\phi, \theta_\phi)} \mathbf{1}_{Z^{N, x_N} \in \mathcal{V}_{\phi, \theta_\phi}(\varepsilon)} \right) \\
&\leq e^{-N[\mathcal{L}(\phi, \theta_\phi) - \frac{\varepsilon}{2}]} \mathbb{E} \left(e^{N\mathcal{L}(Z^{N, x_N}, \theta_\phi)} \right) \\
&\leq e^{-N[\mathcal{L}(\phi, \theta_\phi) - \frac{\varepsilon}{2}]} \\
&\leq e^{-NI_T(\phi) + N\varepsilon}, \tag{4.2.10}
\end{aligned}$$

where the before last inequality follows the fact that $N\mathcal{L}(Z^{N, x_N}, \theta_\phi) = \log(\Xi_T^{N, \theta_\phi})$ and $\mathbb{E}[\Xi_T^{N, \theta_\phi}] \leq 1$.

The second case is the one where $I_T(\phi) = +\infty$. Then there exists $M > I_{T, x}(K) + 1$ and $\theta_\phi \in C^1([0, T]; \mathbb{R}^d)$ such that $\mathcal{L}(\phi, \theta_\phi) > M + \varepsilon$. From the same argument as above, we deduce that

$$\mathbb{P}(Z^{N, x_N} \in \mathcal{V}_{\phi, \theta_\phi}(\varepsilon)) \leq e^{-NM}.$$

Let $K_x = \{\phi \in K, \phi_0 = x\}$. Since $K_x \subset \bigcup_{\phi \in K, \phi_0 = x} \mathcal{V}_{\phi, \theta_\phi}(\varepsilon)$ and K_x is compact, there exists $m = m(\varepsilon) \geq 1$ and $\phi_1, \dots, \phi_m \in K_x$ where we assume that $\phi_1 = \hat{\phi}$, such that

$$K_x \subset \bigcup_{i=1}^m \mathcal{V}_{\phi_i, \theta_{\phi_i}}(\varepsilon).$$

Now there exists a finite set of functions $\{\phi_{m+1}, \dots, \phi_{m+n}\} \subset K \setminus K_x$, such that

$$K \subset \bigcup_{i=1}^{m+n} \mathcal{V}_{\phi_i, \theta_{\phi_i}}(\varepsilon).$$

We choose ε small enough for $i \geq m+1$ such that $x \notin \mathcal{V}_{\phi_i, \theta_{\phi_i}}(\varepsilon)$. Then for N large enough, $\mathbb{P}(Z^{N, x_N} \in \mathcal{V}_{\phi_i, \theta_{\phi_i}}(\varepsilon)) = 0$ if $i \geq m+1$. Hence

$$\begin{aligned}
\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^{N, x_N} \in K) &\leq \limsup_{N \rightarrow \infty} \frac{1}{N} \log \left(\sum_{i=1}^{m+n} \mathbb{P}(Z^{N, x_N} \in \mathcal{V}_{\phi_i, \theta_{\phi_i}}(\varepsilon)) \right) \\
&\leq \max_{1 \leq i \leq m} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(Z^{N, x_N} \in \mathcal{V}_{\phi_i, \theta_{\phi_i}}(\varepsilon)) \\
&\leq - \inf_{1 \leq i \leq m} I_T(\phi_i) + \varepsilon \\
&\leq -I_{T, x}(K) + \varepsilon,
\end{aligned}$$

where we have used (4.2.10) in the third inequality. It remains to let $\varepsilon \rightarrow 0$. \square

It remains to establish exponential tightness. Now we need to impose a growth condition on the β_j 's. One natural assumption would be to assume that for some $C > 0$, all $1 \leq j \leq m$ and $x \in \mathbb{R}^d$, $\beta_j(t, x) \leq C(1 + |x|)$. However, this condition is not satisfied in most of our examples, because one of the β_j 's is quadratic. We shall

instead formulate an assumption which is satisfied in our epidemic models. We shall write $\mathbf{1}$ for the vector in \mathbb{R}^d whose coordinates are all equal to 1, and we exploit the fact that for those j 's such that β_j is quadratic, $\langle h_j, \mathbf{1} \rangle = 0$.

(A.3) We assume that for all starting points $x_N \in \mathbb{Z}_+^d/N$, Z^{N,x_N} takes its values in \mathbb{R}_+^d a.s., and moreover that there exists a $C_\beta > 0$ such that for any $0 \leq j \leq k$ such that $\langle h_j, \mathbf{1} \rangle \neq 0$, $\beta_j(t, x) \leq C_\beta(1 + |x|)$, $0 \leq t \leq T$, $x \in \mathbb{R}^d$.

We now prove

Proposition 4.2.13. *Assume that Conditions (A.1) and (A.3) are satisfied. Let $T > 0$ and $x \in \mathbb{R}^d$ be given, as well as a sequence $x_N \rightarrow x$ as $N \rightarrow \infty$, such that for all $N \geq 1$, $x_N \in \mathbb{Z}_+^d/N$. Then for all $\xi > 0$,*

$$\lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left(\sup_{0 \leq s, t \leq T, |t-s| \leq \delta} |Z_t^{N,x_N} - Z_s^{N,x_N}| > \xi \right) = -\infty.$$

Proof. $\xi > 0$ and $T > 0$ will be fixed throughout this proof. Consider the stopping time

$$\sigma_R^{N,x_N} = \inf\{t \in [0, T], |Z_t^{N,x_N}| > R\}.$$

It is clear that

$$\begin{aligned} \mathbb{P} \left(\sup_{|t-s| \leq \delta} |Z_t^{N,x_N} - Z_s^{N,x_N}| > \xi \right) &\leq \mathbb{P} \left(\sup_{|t-s| \leq \delta} |Z_{t \wedge \sigma_R^{N,x_N}}^{N,x_N} - Z_{s \wedge \sigma_R^{N,x_N}}^{N,x_N}| > \xi \right) \\ &\quad + \mathbb{P} \left(\sigma_R^{N,x_N} < T \right). \end{aligned}$$

We first consider the first term of the above right-hand side. For that purpose, we divide $[0, T]$ into subintervals of length δ , and let $\underline{i}(s) \leq s < \bar{i}(s)$ denote the points of the grid nearest to s .

$$\begin{aligned} &\mathbb{P} \left[\sup_{|s-t| \leq \delta} \left| Z_{t \wedge \sigma_R^{N,x_N}}^{N,x_N} - Z_{s \wedge \sigma_R^{N,x_N}}^{N,x_N} \right| > \xi \right] \\ &= \mathbb{P} \left[\exists 0 \leq s < t \leq T, t-s \leq \delta, \left| Z_{t \wedge \sigma_R^{N,x_N}}^{N,x_N} - Z_{s \wedge \sigma_R^{N,x_N}}^{N,x_N} \right| > \xi \right] \\ &\leq \mathbb{P} \left[\exists 0 \leq s < t \leq T, t-s \leq \delta, \left| Z_{t \wedge \sigma_R^{N,x_N}}^{N,x_N} - Z_{\underline{i}(s)}^{N,x_N} \right| + \left| Z_{\underline{i}(s)}^{N,x_N} - Z_{s \wedge \sigma_R^{N,x_N}}^{N,x_N} \right| > \xi \right] \\ &\leq 2 \left(\frac{T}{\delta} + 1 \right) \sup_{s \in [0, T]} \mathbb{P} \left[\sup_{t \in [s, s+2\delta]} \left| Z_{t \wedge \sigma_R^{N,x_N}}^{N,x_N} - Z_{s \wedge \sigma_R^{N,x_N}}^{N,x_N} \right| > \xi/2 \right]. \end{aligned}$$

Let $\{\theta_i, 1 \leq i \leq d\}$ (resp. $\{\theta_i, d+1 \leq i \leq 2d\}$) denote the standard basis of \mathbb{R}_+^d (resp. of \mathbb{R}_-^d). Thus for every $\lambda > 0$, assuming w.l.o.g. that $|z|$ stands here for $\sup_{1 \leq i \leq d} |z_i|$,

$$\begin{aligned}
& \mathbb{P} \left[\sup_{t \in [s, s+2\delta[} \left| Z_{t \wedge \sigma_R^{N, x_N}}^{N, x_N} - Z_{s \wedge \sigma_R^{N, x_N}}^{N, x_N} \right| > \xi/2 \right] \\
& \leq \sum_{i=1}^{2d} \mathbb{P} \left[\sup_{t \in [s, s+2\delta[} \langle Z_t^{N, x_N} - Z_s^{N, x_N}, \lambda \theta_i \rangle > \lambda \xi/2 \right] \\
& \leq \sum_{i=1}^{2d} \mathbb{P} \left[\sup_{t \in [s, s+2\delta[} \mathbf{M}_{(s,t) \wedge \sigma_R^{N, x_N}}^{N, \lambda \theta_i} + \sum_{j=1}^k \int_{s \wedge \sigma_R^{N, x_N}}^{t \wedge \sigma_R^{N, x_N}} \langle h_j, \lambda \theta_i \rangle \beta_j(r, Z_r^{N, x_N}) dr > \lambda \xi/2 \right] \\
& \leq \sum_{i=1}^{2d} \mathbb{P} \left[\sup_{t \in [s, s+2\delta[} \exp \left(N \mathbf{M}_{(s,t) \wedge \sigma_R^{N, x_N}}^{N, \lambda \theta_i} + N \sum_{j=1}^k \int_{s \wedge \sigma_R^{N, x_N}}^{t \wedge \sigma_R^{N, x_N}} \langle h_j, \lambda \theta_i \rangle \beta_j(r, Z_r^{N, x_N}) dr \right) > e^{N \lambda \xi/2} \right] \\
& \leq \sum_{i=1}^{2d} \mathbb{P} \left[\sup_{t \in [s, s+2\delta[} \Xi_{(s,t) \wedge \sigma_R^{N, x_N}}^{N, \lambda \theta_i} > \exp \left(N \lambda \xi/2 - N \sum_{i=1}^k \left(e^{\langle h_j, \lambda \theta_i \rangle} - 1 \right) \int_{s \wedge \sigma_R^{N, x_N}}^{t \wedge \sigma_R^{N, x_N}} \beta_j(s, Z_r^{N, x_N}) dr \right) \right] \\
& \leq \sum_{i=1}^{2d} \mathbb{P} \left[\sup_{t \in [s, s+2\delta[} \Xi_{(s,t) \wedge \sigma_R^{N, x_N}}^{N, \lambda \theta_i} > \exp \left(N \lambda \xi/2 - 2\delta N k \bar{\beta}_R e^{\lambda \bar{h}} \right) \right] \\
& \leq 2d \exp \left(-N \lambda \xi/2 + 2\delta N k \bar{\beta}_R e^{\lambda \bar{h}} \right),
\end{aligned}$$

where $\bar{\beta}_R = \sup_{1 \leq j \leq k} \sup_{0 \leq t \leq T, |x| \leq R} \beta_j(t, x)$. Optimizing over $\lambda > 0$ yields

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left[\sup_{|s-t| \leq \delta} \left| Z_{t \wedge \sigma_R^{N, x_N}}^{N, x_N} - Z_{s \wedge \sigma_R^{N, x_N}}^{N, x_N} \right| > \xi \right] \leq -\frac{\xi}{2\bar{h}} \left(\log \left(\frac{\xi \delta^{-1}}{4\bar{h}k\bar{\beta}_R} \right) - 1 \right).$$

Consequently for any fixed $R > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left[\sup_{|s-t| \leq \delta} \left| Z_{t \wedge \sigma_R^{N, x_N}}^{N, x_N} - Z_{s \wedge \sigma_R^{N, x_N}}^{N, x_N} \right| > \xi \right] = -\infty.$$

It remains to show that

$$\lim_{R \rightarrow \infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left(\sigma_R^{N, x_N} < T \right) = -\infty. \quad (4.2.11)$$

Combing the fact that for all $t \leq T$

$$\sup_{s \leq t} |Z_s^{N, x_N}| \leq \sup_{s \leq t} |\langle Z_s^{N, x_N}, \mathbb{1} \rangle|$$

and that by Gronwall's Lemma 2.2.9, with $\bar{h} = \sup_{1 \leq j \leq k} |h_j|$ and C_β the constant from assumption (A.3),

$$\sup_{s \leq t} |\langle Z_s^{N, x_N}, \mathbb{1} \rangle| \leq \left(|\langle x, \mathbb{1} \rangle| + k\bar{h}Ct + \sup_{s \leq t} |M_s^{N, \mathbb{1}}| \right) e^{k\bar{h}C_\beta t},$$

we deduce that

$$\sup_{s \leq t} |Z_s^{N, x_N}| \leq \left(|\langle x, \mathbb{1} \rangle| + k\bar{h}Ct + \sup_{s \leq t} |M_s^{N, \mathbb{1}}| \right) e^{k\bar{h}C\beta t}. \quad (4.2.12)$$

By Itô's formula we have, with $\mathcal{M}_t^{N, \mathbb{1}}$ a local martingale, and defining $A_s^N := 1 \vee (\sup_{0 \leq r \leq s} |M_r^{N, \mathbb{1}}|)$,

$$\begin{aligned} & \left(M_t^{N, \mathbb{1}} \right)^{2N} \\ &= N \sum_{j: \langle h_j, \mathbb{1} \rangle \neq 0} \int_0^t \beta_j(s, Z_s^{N, x_N}) \left[\left(M_s^{N, \mathbb{1}} + \frac{\langle h_j, \mathbb{1} \rangle}{N} \right)^{2N} - \left(M_s^{N, \mathbb{1}} \right)^{2N} - 2N \left(M_s^{N, \mathbb{1}} \right)^{2N-1} \frac{\langle h_j, \mathbb{1} \rangle}{N} \right] ds + \mathcal{M}_t^{N, \mathbb{1}} \\ &\leq NC_\beta \sum_j \frac{N(2N-1)}{N^2} \langle h_j, \mathbb{1} \rangle^2 \int_0^t (1 + |Z_s^{N, x_N}|) \left(|M_s^{N, \mathbb{1}}| + \frac{\langle h_j, \mathbb{1} \rangle}{N} \right)^{2N-2} ds + \mathcal{M}_t^{N, \mathbb{1}} \\ &\leq NC_\beta C_T \left(1 + \frac{\bar{h}}{N} \right)^{2N} \int_0^t \frac{1 + |Z_s^{N, x_N}|}{A_s^N} (A_s^N)^{2N-1} ds + \mathcal{M}_t^{N, \mathbb{1}} \\ &\leq NC_T' \int_0^t (A_s^N)^{2N} ds + \mathcal{M}_t^{N, \mathbb{1}}, \end{aligned} \quad (4.2.13)$$

where we have used (4.2.12) and the inequality $a + b \leq a(1 + b)$ for $a \geq 1$, $b \geq 0$. From Doob's inequality,

$$\mathbb{E} \left[\sup_{s \leq t \wedge \sigma_R^{N, x_N}} (M_s^{N, \mathbb{1}})^{2N} \right] \leq \left(\frac{2N}{2N-1} \right)^{2N} \mathbb{E} \left[(M_{t \wedge \sigma_R^{N, x_N}}^{N, \mathbb{1}})^{2N} \right]. \quad (4.2.14)$$

Since $\mathcal{M}_{t \wedge \sigma_R^{N, x_N}}^{N, \mathbb{1}}$ is a martingale, we can take the expectation in the inequality (4.2.13) at time $t \wedge \sigma_R^{N, x_N}$, and deduce from the resulting inequality, (4.2.14) and $\sup_{N \geq 1} \left(\frac{2N}{2N-1} \right)^{2N} < \infty$

$$\mathbb{E} \left[\sup_{s \leq t \wedge \sigma_R^{N, x_N}} (M_s^{N, \mathbb{1}})^{2N} \right] \leq NC_T'' \int_0^t \mathbb{E} \left[\left(A_{s \wedge \sigma_R^{N, x_N}}^N \right)^{2N} \right] ds.$$

Since for $a \geq 0$, $(1 \vee a)^{2N} \leq 1 + a^{2N}$, it follows that for all $0 \leq t \leq T$,

$$\mathbb{E} \left[\left(A_{t \wedge \sigma_R^{N, x_N}}^N \right)^{2N} \right] \leq 1 + NC_T'' \int_0^t \mathbb{E} \left[\left(A_{s \wedge \sigma_R^{N, x_N}}^N \right)^{2N} \right] ds.$$

Hence it follows from Gronwall's lemma that

$$\mathbb{E} \left[\sup_{t \leq T \wedge \sigma_R^{N, x_N}} (M_t^{N, \mathbb{1}})^{2N} \right] \leq \exp(C_T N T). \quad (4.2.15)$$

For any $0 < \kappa < R$, denoting

$$C(R, \kappa) := (R - \kappa)e^{-k\bar{h}C_T T} - |\langle x, \mathbb{1} \rangle| - k\bar{h}C_T T,$$

we have

$$\begin{aligned} \limsup_{N \rightarrow +\infty} \frac{1}{N} \log \mathbb{P} \left[\sigma_R^{N, x_N} \leq T \right] &\leq \limsup_{N \rightarrow +\infty} \frac{1}{N} \log \mathbb{P} \left[\sup_{t \leq T \wedge \sigma_R^{N, x_N}} |Z_t^{N, x}| > R - \kappa \right] \\ &\leq \limsup_{N \rightarrow +\infty} \frac{1}{N} \log \mathbb{P} \left[\sup_{t \leq T \wedge \sigma_R^{N, x_N}} |M_t^{N, \mathbb{1}}| > C(R, \kappa) \right] \\ &\leq \limsup_{N \rightarrow +\infty} \frac{1}{N} \log \mathbb{P} \left[\sup_{t \leq T \wedge \sigma_R^{N, x_N}} (M_t^{N, \mathbb{1}})^{2N} > [C(R, \kappa)]^{2N} \right] \\ &\leq -2 \log [C(R, \kappa)] + \limsup_{N \rightarrow +\infty} \frac{1}{N} \log \mathbb{E} \left[\sup_{t \leq T \wedge \sigma_R^{N, x_N}} \left(M_t^{N, \mathbb{1}} \right)^{2N} \right] \\ &\leq -2 \log [C(R, \kappa)] + CT, \end{aligned}$$

where we have used (4.2.15) for the last inequality. We deduce (4.2.11) by letting R tend to $+\infty$. \square

We shall also need the following lemma, where we use the notation

$$w'_{Z^N}(\delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq n} w_x([t_{i-1}, t_i]),$$

with $w_x([t_{i-1}, t_i]) = \sup_{t_{i-1} \leq s < t_i} |x_t - x_s|$ and the infimum is taken over all sequences $0 = t_0 < t_1 < \dots < t_n = T$ satisfying $\inf_{1 \leq i \leq n} (t_i - t_{i-1}) \geq \delta$.

Lemma 4.2.14. *If Conditions (A.1) and (A.3) are satisfied, then for any $N \geq 1$, $\rho > 0$,*

$$\lim_{\delta \rightarrow 0} \mathbb{P}(w'_{Z^N}(\delta) > \rho) = 0.$$

Proof. Since the space $D([0, T]; \mathbb{R}^d)$ is separable and complete, the law of Z^N on this space is tight, see Theorem 1.3 in Billingsley [8], which implies the lemma, from Theorem 13.2 of the same reference. \square

We can now deduce the following theorem from Proposition 4.2.13 and Lemma 4.2.14.

Theorem 4.2.15. *If Conditions (A.1) and (A.3) is satisfied, then the sequence $\{Z^{N, z_N}, N \geq 1\}$ is exponentially tight in $D([0, T]; \mathbb{R}^d)$.*

Proof. Given $R > 0$ and a sequence $\{\delta_\ell > 0, \ell \geq 1\}$ the following is a compact subset of $D([0, T]; \mathbb{R}^d)$ (see Theorem 12.3 in Billingsley [8]):

$$K_{R,\{\delta_\ell\}} = \{x, \|x\|_T \leq R\} \bigcap_{\ell \geq 1} \{x, w'_x(\delta_\ell) \leq \ell^{-1}\}.$$

For any $\alpha > 0$, we need to find R_α and $\{\delta_\ell^\alpha, \ell \geq 1\}$ such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left(\{ \|Z^{N,z_N}\|_T > R_\alpha \} \bigcup_{\ell \geq 1} \{ w'_{Z^N}(\delta_\ell^\alpha) > \ell^{-1} \} \right) \leq -\alpha. \quad (4.2.16)$$

It is not hard to find R_α such that $\mathbb{P}(\|Z^N\|_T > R_\alpha) \leq e^{-N\alpha}$, for all $N \geq 1$. Since $w'_x(\delta) \leq w_x(2\delta)$, it follows from Proposition 4.2.13 that for each $\ell \geq 1$, there exists a $\delta_\ell > 0$ such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}(w'_{Z^N}(\delta_\ell) > \ell^{-1}) \leq -(\alpha + \ell).$$

Consequently, there exists an N_ℓ such that for $N \geq N_\ell$,

$$\mathbb{P}(w'_{Z^N}(\delta_\ell) > \ell^{-1}) \leq e^{-N(\alpha + \ell)}.$$

Combining this with Lemma 4.2.14, we deduce that there exists $0 < \delta_\ell^\alpha \leq \delta_\ell$ such that for all $N \geq 1$,

$$\mathbb{P}(w'_{Z^N}(\delta_\ell^\alpha) > \ell^{-1}) \leq e^{-N(\alpha + \ell)}.$$

It follows that for all $N \geq 1$,

$$\begin{aligned} \mathbb{P} \left(\{ \|Z^{N,z_N}\|_T > R_\alpha \} \bigcup_{\ell \geq 1} \{ w'_{Z^N}(\delta_\ell^\alpha) > \ell^{-1} \} \right) &\leq e^{-N\alpha} \sum_{\ell \geq 0} e^{-N\ell} \\ &\leq (1 - e^{-N})^{-1} e^{-\alpha N}, \end{aligned}$$

from which (4.2.16) follows. \square

It is not hard to see that a combination of the exact same arguments as used in the proofs of Theorem 4.2.12, Proposition 4.2.13 and Theorem 4.2.15 yields the following result.

Theorem 4.2.16. *Assume that assumptions (A.1) and (A.3) are satisfied. Then for any closed subset $F \subset D([0, T]; \mathbb{R}^d)$ and any compact $\mathcal{K} \subset \mathbb{R}^d$, we have*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{x \in \mathcal{K}} \mathbb{P}(Z^{N,x_N} \in F) \leq - \inf_{x \in \mathcal{K}} I_{T,x}(F).$$

4.2.5 Time of extinction in the SIRS model

We shall denote by T_{Ext}^N the time of extinction of the disease, and we want to learn what large deviations can tell us about it. In order to simplify the presentation, we start with the two most simple examples of the SIRS model and the SIS model.

These are models with fixed population size N . We treat the SIRS model in this section, and the SIS model in the next one. In this section, we shall follow the arguments from Kratz and Pardoux [21], which itself follows closely the arguments in Dembo and Zeitouni [9].

The deterministic SIRS Model can be reduced to a 2-dimensional ODE for the pair $(s(t), i(t))$ which reads

$$\begin{cases} i'(t) = \lambda s(t)i(t) - \gamma i(t), \\ s'(t) = -\lambda s(t)i(t) + \rho(1 - s(t) - i(t)). \end{cases} \quad (4.2.17)$$

This process lives in the compact set $A = A_{SIRS} = \{(x, y), 0 \leq x, y, x + y \leq 1\}$. Provided again $R_0 = \frac{\lambda}{\gamma} > 1$, there is a unique stable endemic equilibrium $(i^*, s^*) = \left(\frac{\rho}{\lambda} \frac{\lambda - \gamma}{\rho + \gamma}, \frac{\gamma}{\lambda}\right) \in A$, while the disease free equilibrium $(1, 0)$ is unstable. Here $h_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\beta_1(x, y) = \lambda xy$, $h_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, $\beta_2(x, y) = \gamma y$, $h_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\beta_3(x, y) = \rho(1 - x - y)$.

The stochastic process $(I^N(t), S^N(t))$ may hit $\{0\} \times [0, 1]$, and then stays there for ever (this is how the disease goes extinct). On the other hand, if it hits $\partial A \setminus \{0\} \times [0, 1]$, the process comes back to \dot{A} . Similarly, starting from $\{0\} \times [0, 1]$, the ODE stays there for ever (and converges to $(0, 1)$), while starting from $\partial A \setminus \{0\} \times [0, 1]$, it enters \dot{A} instantaneously. We thus define

$$T_{\text{Ext}}^N = \inf\{t \geq 0, I^N(t) = 0\}.$$

Unfortunately, the theory of Large Deviations will not give us directly results on T_{Ext}^N , but rather on

$$T_\delta^N = \inf\{t \geq 0, I^N(t) \leq \delta\}, \text{ for any } \delta > 0.$$

An ad hoc argument, which we shall present at the end, allows us to deduce the desired result concerning T_{Ext}^N . We are interested in the exit time from $A_\delta := \{(x, y) \in A, x \geq \delta\}$ through the boundary $\partial A_\delta := \{(x, y) \in A, x = \delta\}$.

We shall write $D_{T,A} := D([0, T]; A)$. In order to formulate our results, we shall need the following notations (below z stands for (x, y))

$$\begin{aligned} V(z, z', T) &= \inf_{\phi \in D_{T,A}, \phi_0 = z, \phi_T = z'} I_T(\phi) \\ V(z, z') &= \inf_{T > 0} V(z, z', T) \\ \bar{V}_\delta &= \inf_{z \in \partial A_\delta} V(z^*, z), \\ \bar{V} &= \inf_{z \in \{0\} \times [0, 1]} V(z^*, z). \end{aligned}$$

We want to prove the

Theorem 4.2.17. Let $T_{Ext}^{N,z}$ denote the extinction time in the SIRS model starting from $z_N = \frac{[zN]}{N}$. Given $\eta > 0$, for all $z \in A$,

$$\lim_{N \rightarrow \infty} \mathbb{P}(\exp\{N(\bar{V} - \eta)\} < T_{Ext}^{N,z} < \exp\{N(\bar{V} + \eta)\}) = 1.$$

Moreover, for all $\eta > 0$, $z \in A$ and N large enough,

$$\exp\{N(\bar{V} - \eta)\} \leq \mathbb{E}(T_{Ext}^{N,z}) \leq \exp\{N(\bar{V} + \eta)\}.$$

We shall first establish

Proposition 4.2.18. Given $\eta > 0$, for all $z \in \mathring{A}_\delta$,

$$\lim_{N \rightarrow \infty} \mathbb{P}(\exp\{N(\bar{V}_\delta - \eta)\} < T_\delta^{N,z} < \exp\{N(\bar{V}_\delta + \eta)\}) = 1.$$

Moreover, for all $\eta > 0$, $z \in \mathring{A}_\delta$ and N large enough,

$$\exp\{N(\bar{V}_\delta - \eta)\} \leq \mathbb{E}(T_\delta^{N,z}) \leq \exp\{N(\bar{V}_\delta + \eta)\}.$$

Let us now formulate a set of assumptions which are satisfied in our case, under which we will prove Proposition 4.2.18. For that sake, we shall rewrite the ODE (4.2.17) as

$$\frac{dz_t}{dt} = b(z_t), \quad z_0 = z. \quad (4.2.18)$$

Assumption 4.2.19.

(E1) z^* is the only stable equilibrium point of (4.2.18) in A_δ and the solution z_t^x of (4.2.18) satisfies, for all $z_0 = z \in A_\delta$,

$$z_t^z \in \mathring{A}_\delta \text{ for all } t > 0 \text{ and } \lim_{t \rightarrow \infty} z_t^z = z^*.$$

(E2) $\bar{V} < \infty$.

(E3) For all $\rho > 0$ there exist constants $T(\rho), \varepsilon(\rho) > 0$ with $T(\rho), \varepsilon(\rho) \downarrow 0$ as $\rho \downarrow 0$ such that for all $z \in \partial A_\delta \cup \{z^*\}$ and all $x, y \in \bar{B}(z, \rho) \cap A$ there exists

$$\phi = \phi(\rho, x, y) : [0, T(\rho)] \mapsto A \text{ with } \phi_0 = x, \phi_{T(\rho)} = y \text{ and } I_{T(\rho)}(\phi) < \varepsilon(\rho).$$

(E4) For all $z \in \partial A_\delta$ there exists an $\eta_0 > 0$ such that for all $\eta < \eta_0$ there exists a $\tilde{z} = \tilde{z}(\eta) \in A \setminus A_\delta$ with $|z - \tilde{z}| > \eta$.

Note that the conditions (E1) and (E4) would not be satisfied if we replace A_δ by A .

The proof of Proposition 4.2.18 relies upon the following sequence of lemmas, whose proofs will be given below, after the proof of the proposition.

Lemma 4.2.20. *For any $\varepsilon > 0$, there exists a $\rho_0 > 0$ such that for all $\rho < \rho_0$,*

$$\sup_{z \in \partial A_\delta \cup \{z^*\}} \sup_{|z' - z| \vee |z'' - z| \leq \rho} \inf_{0 \leq T \leq 1} V(z', z'', T) < \varepsilon.$$

Lemma 4.2.21. *For any $\eta > 0$, there exists a $\rho_0 > 0$ such that for all $\rho < \rho_0$, there exists a $T_0 < \infty$ such that*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \inf_{|z - z^*| \leq \rho} \mathbb{P}(T_\delta^{N,z} \leq T_0) \geq -(\bar{V} + \eta).$$

Let us define for some $\rho > 0$ small enough, $B_\rho := \overline{B(z^*, \rho)}$ and

$$\sigma_\rho^N = \inf\{t \geq 0, Z_t^N \in B_\rho \cup \{z, z_1 \leq \delta\}\}.$$

Lemma 4.2.22. *If $\rho > 0$ is such that $B_\rho \subset \mathring{A}_\delta$, then*

$$\lim_{t \rightarrow \infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{x \in A_\delta} \mathbb{P}(\sigma_\rho^{N,z} > t) = -\infty.$$

Lemma 4.2.23. *Let C be a closed subset of $A \setminus \mathring{A}_\delta$. Then*

$$\lim_{\rho \rightarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{2\rho \leq |z - z^*| \leq 3\rho} \mathbb{P}(Z_{\sigma_\rho^N}^{N,z} \in C) \leq -\inf_{z' \in C} V(z^*, z').$$

Lemma 4.2.24. *If $\rho > 0$ is such that $B_\rho \subset \mathring{A}_\delta$ and $z \in \mathring{A}_\delta$,*

$$\lim_{N \rightarrow \infty} \mathbb{P}(Z_{\sigma_\rho^N}^{N,z} \in B_\rho) = 1.$$

Lemma 4.2.25. *For all $\rho, c > 0$, there exists a constant $T = T(c, \rho) < \infty$ such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{z \in A_\delta} \mathbb{P}(\sup_{0 \leq t \leq T} |Z_t^{N,z} - z| \geq \rho) \leq -c.$$

We first give the

Proof of Proposition 4.2.18. STEP 1: UPPER BOUND OF T_δ^N We choose $\eta = \varepsilon/2$, and ρ, T_0 as in Lemma 4.2.21. By Lemma 4.2.22, for any arbitrarily fixed $a > 0$, there exists a T_1 such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{z \in A_\delta} \mathbb{P}(\sigma_\rho^{N,z} > T_1) < -2a < 0.$$

Let $T = T_0 + T_1$. There exists an $N_0 \geq 1$ such that for all $N \geq N_0$,

$$\begin{aligned}
q &:= \inf_{z \in A_\delta} \mathbb{P}(T_\delta^{N,z} \leq T) \geq \inf_{z \in A_\delta} \mathbb{P}(\sigma_\rho^{N,z} \leq T_1) \inf_{z \in B_\rho} \mathbb{P}(T_\delta^{N,z} \leq T_0) \\
&\geq e^{-N(\bar{V}_\delta + \eta)}, \tag{4.2.19}
\end{aligned}$$

since the second factor is bounded from below by say $e^{-N(\bar{V}_\delta + \eta/2)}$ from Lemma 4.2.21, and from the previous estimate, we deduce that for N large enough,

$$\begin{aligned}
\inf_{z \in A_\delta} \mathbb{P}(\sigma_\rho^{N,z} \leq T_1) &= 1 - \sup_{z \in A_\delta} \mathbb{P}(\sigma_\rho^{N,z} > T_1) \\
&\geq 1 - e^{-Na} \\
&\geq e^{-N\eta/2}.
\end{aligned}$$

Next, by the strong Markov property,

$$\begin{aligned}
\mathbb{P}(T_\delta^{N,z} > (k+1)T) &= [1 - \mathbb{P}(T_\delta^{N,z} \leq (k+1)T | T_\delta^{N,z} > kT)] \mathbb{P}(T_\delta^{N,z} > kT) \\
&\leq (1-q) \mathbb{P}(T_\delta^{N,z} > kT).
\end{aligned}$$

Iterating, we get

$$\sup_{z \in A_\delta} \mathbb{P}(T_\delta^{N,z} > kT) \leq (1-q)^k.$$

Therefore

$$\sup_{z \in A_\delta} \mathbb{E}(T_\delta^{N,z}) \leq T[1 + \sum_{k=1}^{\infty} \sup_{z \in A_\delta} \mathbb{P}(T_\delta^{N,z} > kT)] \leq T \sum_{k=0}^{\infty} (1-q)^k = \frac{T}{q},$$

so from (4.2.19),

$$\sup_{x \in A_\delta} \mathbb{E}[T_\delta^{N,z}] \leq T e^{N(\bar{V}_\delta + \eta)}, \tag{4.2.20}$$

and the upper bound for $\mathbb{E}[T_\delta^{N,z}]$ follows. From Chebycheff,

$$\mathbb{P}(T_\delta^{N,z} \geq e^{N(\bar{V}_\delta + \varepsilon)}) \leq e^{-N(\bar{V}_\delta + \varepsilon)} \mathbb{E}[T_\delta^{N,z}] \leq T e^{-N\varepsilon/2},$$

which tends to 0 as $N \rightarrow \infty$, hence the upper bound for T_δ^N .

STEP 2: LOWER BOUND OF T_δ^N Let $\rho > 0$ be small enough such that $B_{2\rho} := B(z^*, 2\rho) \subset \mathring{A}_\delta$. We define a sequence of stopping times as follows. $\theta_0 = 0$ and for $m \geq 0$,

$$\begin{aligned}
\tau_m &= \inf\{t \geq \theta_m, Z_t^N \in B_\rho \cup \{z, z_1 \leq \delta\}\}, \\
\theta_{m+1} &= \inf\{t > \tau_m, Z_t^N \in (B_{2\rho})^c\},
\end{aligned}$$

with the convention that $\theta_{m+1} = \infty$ in case $Z_{\tau_m}^N \in \{z, z_1 \leq \delta\}$.

In case $\bar{V}_\delta = 0$, the lower bound is an easy consequence of Lemmas 4.2.24 and 4.2.25. So we assume from now on that $\bar{V}_\delta > 0$ and fix $\varepsilon > 0$ arbitrarily small. Since

$\{z, z_1 \leq \delta\}$ is a closed set, from Lemma 4.2.23, for $\rho > 0$ small enough,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{2\rho \leq |z - z^*| \leq 3\rho} \mathbb{P}(Z_{\sigma_\rho^N}^{N,z} \in \{z, z_1 \leq \delta\}) \leq -\bar{V}_\delta + \frac{\varepsilon}{3}.$$

Now with $c = \bar{V}_\delta$, we let $T_0 = T(c, \rho)$ be as in Lemma 4.2.25. Then there exists an N_0 such that for $N \geq N_0$, and all $m \geq 1$,

$$\sup_{z \in A_\delta} \mathbb{P}(T_\delta^{N,z} = \tau_m) \leq \sup_{2\rho \leq |z - z^*| \leq 3\rho} \mathbb{P}(Z_{\sigma_\rho^N}^{N,z} \in \{z, z_1 \leq \delta\}) \leq e^{-N(\bar{V}_\delta - \varepsilon/2)},$$

while

$$\sup_{z \in A_\delta} \mathbb{P}_z(\theta_m - \tau_{m-1} \leq T_0) \leq \sup_{z \in A_\delta} \mathbb{P}(\sup_{0 \leq t \leq T_0} |Z_t^{N,z} - z| \geq \rho) \leq e^{-N(\bar{V}_\delta - \varepsilon/2)}.$$

The event $\{T_\delta^N \leq kT_0\}$ implies that either one of the first $k+1$ events $\{T_\delta^N = \tau_m\}$ occurs, or else at least one of the first k excursions $[\tau_m, \tau_{m+1}]$ away from B_ρ is of length at most T_0 . Consequently, from the two preceding estimates,

$$\begin{aligned} \mathbb{P}(T_\delta^N \leq kT_0) &\leq \sum_{m=0}^k \mathbb{P}(T_\delta^N = \tau_m) + \mathbb{P}(\min_{1 \leq m \leq k} (\theta_m - \tau_{m-1}) \leq T_0) \\ &\leq \mathbb{P}(T_\delta^N = \tau_0) + 2ke^{-N(\bar{V}_\delta - \varepsilon/2)}. \end{aligned}$$

Choosing now $k = \lceil T_0^{-1} e^{N(\bar{V}_\delta - \varepsilon)} \rceil + 1$ yields

$$\mathbb{P}(T_\delta^N \leq e^{N(\bar{V}_\delta - \varepsilon)}) \leq \mathbb{P}(Z_{\sigma_\rho^N}^N \notin B_\rho) + 3T_0^{-1} e^{-N\varepsilon/2}.$$

By Lemma 4.2.24, the right-hand side tends to 0 as $N \rightarrow \infty$. We have completed the proof of the first statement in Proposition 4.2.18. This result combined with Chebycheff's inequality and (4.2.20) yields the second result. \square

We now turn to the proofs of the lemmas.

Proof of Lemma 4.2.20. This lemma is a direct consequence of the assumption (E3). \square

Proof of Lemma 4.2.21. We make use of Lemma 4.2.20 with $\varepsilon = \eta/4$ and choose $\rho < \rho_0$. Let $z \in B_\rho$. There exists a continuous path ψ^z such that $\psi_0^z = z$, $\psi_{t_z}^z = z^*$ for some $t_z \leq 1$ and $I_{t_z}(\psi^z) \leq \eta/4$. From assumption (E2), there exists a continuous path $\phi \in C([0, T_1]; A)$ such that $\phi_0 = z^*$, $\phi_{T_1} = z' \in \partial A_\delta$, and $I_{T_1}(\phi) \leq \bar{V} + \eta/4$. From Lemma 4.2.20, there exists a continuous path $\tilde{\psi}$ such that $\tilde{\psi}_0 = z'$ and $\tilde{\psi}_{s_{z'}} = z'' \in A \setminus A_\delta$, with $s_{z'} \leq 1$, $I_{s_{z'}}(\tilde{\psi}) \leq \eta/4$ and $d(z'', A_\delta) = \Delta > 0$, where $\Delta < \bar{\delta}$. Finally let $\{\xi_t, 0 \leq t \leq 2 - t_z - s_{z'}\}$ be a solution of (4.2.18) starting from $\xi_0 = z''$. From Proposition 4.2.2, $I(\xi) = 0$. Concatenating the paths ψ^z , ϕ , $\tilde{\psi}$ and ξ , we obtain a

path $\phi^z \in C([0, T_0]; A)$ (with $T_0 = T_1 + 2$) starting from z , with $I_{T_0}(\phi^z) \leq \bar{V} + 3\eta/4$.
Let now

$$\Psi = \bigcup_{z \in B_\rho} \{\psi \in D([0, T_0]; A), \|\psi - \phi^z\|_{T_0} < \Delta/2\}.$$

Ψ is an open subset of $D([0, T_0]; A)$, such that B_ρ is adapted to Ψ in the sense of Definition 4.2.8. Hence we can make use of Theorem 4.2.9, hence

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} \log \inf_{z \in B_\rho} \mathbb{P}(Z^{N,z} \in \Psi) &\geq - \sup_{z \in B_\rho} \inf_{\phi \in \Psi, \phi_0 = z} I_{T_0}(\phi) \\ &\geq - \sup_{z \in B_\rho} I_{T_0}(\phi^z) \\ &> -(\bar{V} + \eta). \end{aligned}$$

The results follows from this and $\{Z^N \in \Psi\} \subset \{T_\delta^N \leq T_0\}$. \square

Proof of Lemma 4.2.22. Since $\sigma_\rho^{N,z} = 0$ if $z \in B_\rho$, it suffices to restrict ourselves to $z \in A_\delta \setminus B_\rho$. For each $t > 0$, we define the closed set

$$\Psi_t := \{\phi \in D([0, t]; A), \phi_s \in \overline{A_\delta \setminus B_\rho} \text{ for all } 0 \leq s \leq t\},$$

so that $\{\sigma_\rho^{N,z} > t\} \subset \{Z^{N,z} \in \Psi_t\}$. Hence by Theorem 4.2.16,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{z \in A_\delta \setminus B_\rho} \mathbb{P}(\sigma_\rho^{N,z} > t) \leq - \inf_{\phi \in \Psi_t} I_t(\phi).$$

It then suffices to show that

$$\inf_{\phi \in \Psi_t} I_t(\phi) \rightarrow \infty \text{ as } t \rightarrow \infty. \quad (4.2.21)$$

Starting from any $z \in \overline{A_\delta \setminus B_\rho}$, the solution z_t^z of (4.2.18) hits $\mathring{B}_{\rho/2}$ in finite time T_z which is upper semicontinuous (recall Definition A.7.1) in z , so by compactness $T := \sup_{z \in \overline{A_\delta \setminus B_\rho}} T_z < \infty$, and $z. \notin \Psi_T$ as soon as $z.$ solves (4.2.18).

Now if (4.2.21) does not hold, there would exist $M > 0$ and for each $n \geq 1$ $\phi_n \in \Psi_{nT}$ such that $I_{nT}(\phi_n) \leq M$ or all $n \geq 1$. Now if $\phi_{n,k}(t) = \phi_n(kT + t)$, $0 \leq t \leq T$, $0 \leq k \leq n-1$, we have that

$$n \min_{0 \leq k \leq n-1} I_T(\phi_{n,k}) \leq \sum_{k=0}^{n-1} I_T(\phi_{n,k}) = I_{nT}(\phi_n) \leq M.$$

Hence we would produce a sequence $\psi_n \in \Psi_T$ such that $I_T(\psi_n) \rightarrow 0$ as $n \rightarrow \infty$. From Theorem 4.2.3, the sequence ψ_n belongs to a compact set, and I_T is lower semicontinuous (recall Definition A.7.1), so that along a subsequence, $\psi_n \rightarrow \psi^*$, where $\psi^* \in \Psi_T$ and $I_T(\psi^*) \leq \liminf_n I_T(\psi_n) = 0$, and those two last statements are

contradictory from Proposition 4.2.2 and the fact that Ψ_T contains no solution of (4.2.18). \square

Proof of Lemma 4.2.23. We need only consider the case $\inf_{z \in C} V(z^*, z) > 0$, since in the other case the result is trivial. So we can choose $\varepsilon > 0$ such that

$$V_C^\varepsilon := \left(\inf_{z \in C} V(z^*, z) - \varepsilon \right) \wedge \varepsilon^{-1} > 0.$$

By Lemma 4.2.20, there exists a $\rho_0 > 0$ such that for all $0 < \rho < \rho_0$,

$$\sup_{z \in B_{3\rho} \setminus B_{2\rho}} V(z^*, z) < \varepsilon,$$

hence

$$\inf_{z' \in B_{3\rho} \setminus B_{2\rho}, z \in C} V(z', z) \geq \inf_{z \in C} V(z^*, z) - \sup_{z' \in B_{3\rho} \setminus B_{2\rho}} V(z^*, z') > V_C^\varepsilon.$$

For $T > 0$, consider the closed set $\Phi^T \subset D([0, T]; A)$ defined as

$$\Phi^T = \{\phi \in D([0, T]; A), \phi_t \in C \text{ for some } 0 \leq t \leq T\}.$$

For $z' \in \overline{B_{3\rho} \setminus B_{2\rho}}$,

$$\mathbb{P}(Z_{\sigma_\rho^N}^{N, z'} \in C) \leq \mathbb{P}(\sigma_\rho^{N, z'} > T) + \mathbb{P}(Z^{N, z'} \in \Phi^T). \quad (4.2.22)$$

We next bound from above the two terms of the last right-hand side. Concerning the second term,

$$\inf_{\phi \in \Phi^T, \phi_0 \in B_{3\rho} \setminus B_{2\rho}} I_T(\phi) \geq \inf_{z' \in B_{3\rho} \setminus B_{2\rho}, z \in C} V(z', z) \geq V_C^\varepsilon.$$

Hence from Theorem 4.2.16,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{z' \in B_{3\rho} \setminus B_{2\rho}} \mathbb{P}(Z^{N, z'} \in \Phi^T) \leq -V_C^\varepsilon.$$

For the first term, we deduce from Lemma 4.2.22 that for some $T_0 > 0$, all $T \geq T_0$,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{z' \in B_{3\rho} \setminus B_{2\rho}} \mathbb{P}(\sigma_\rho^{N, z'} > T) < -V_C^\varepsilon.$$

(4.2.22) together with the last two estimates produces an inequality which, after letting $\varepsilon \rightarrow 0$, yields the result. \square

Proof of Lemma 4.2.24. z_t^z denoting the solution of (4.2.18) starting from $z \in A_\delta^\circ$, let $T_\rho = \inf\{t > 0, z_t^z \in B_{\rho/2}\}$. From (E1) it follows that $T_\rho < \infty$ and $\Delta := \inf_{0 \leq t \leq T_\rho} d(z_t^z, \partial A_\delta) >$

0. Consequently

$$\mathbb{P}\left(Z_{\sigma_p^N}^{N,z} \notin B_\rho\right) \leq \mathbb{P}\left(\sup_{0 \leq t \leq T_p} |Z_t^{N,z} - z_t^z| \geq \frac{\Delta \wedge \rho}{2}\right),$$

which tends to 0 as $N \rightarrow \infty$ from Theorem 2.2.7. Q.E.D. \square

Proof of Lemma 4.2.25. Let $\rho, c > 0$ be fixed. For $T > 0$, $N \geq 1$ and $z \in A_\delta$,

$$\begin{aligned} \mathbb{P}\left(\sup_{0 \leq t \leq T} |Z_t^{N,z} - z| \geq \rho\right) &= \mathbb{P}\left(\sup_{0 \leq t \leq T} \left| \sum_j h_j P_j \left(N \int_0^t \beta_j(Z_s^x) ds \right) \right| \geq \rho\right) \\ &\leq \mathbb{P}\left(\sum_j P_j(N\bar{\beta}T) \geq N\rho\bar{h}^{-1}\right) \\ &\leq k\mathbb{P}(P(N\bar{\beta}T) \geq N\rho\bar{h}^{-1}k^{-1}). \end{aligned}$$

Now from Cramér's Theorem A.3.1

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{z \in A_\delta} \mathbb{P}\left(\sup_{0 \leq t \leq T} |Z_t^{N,z} - z| \geq \rho\right) \leq -\frac{\rho}{hk} \log\left(\frac{\rho}{hk\bar{\beta}T}\right) + \frac{\rho}{hk} - \bar{\beta}T,$$

and the absolute value of the right-hand side can be made arbitrarily large by choosing T arbitrarily small. \square

It remains finally to turn to the

Proof of Theorem 4.2.17. Since $V_\delta \uparrow V$ as $\delta \downarrow 0$, it is clear that the lower bounds for T_{Ext}^N and its expectation follow from Proposition 4.2.18. It remains to establish the upper bound. Analyzing carefully the proof of the upper bound, we notice that the key step, which relies upon Lemmas 4.2.21 and 4.2.22 whose proof do not extend to our new situation, is the derivation of the inequality (4.2.19). The upper bound both for the time of exit and its expectation are a direct consequence of (4.2.19), without any further reference to those assumptions which are not valid any more. We fix $\eta > 0$. Let $t > 0$ be arbitrary. From Lemma 4.2.26 below, if $c_t := \log\left(\frac{\lambda - \gamma e^{(\gamma-\lambda)t}}{\gamma - \gamma e^{(\gamma-\lambda)t}}\right)$,

$$\inf_{z \in A \setminus A_\delta} \mathbb{P}(T_{\text{Ext}}^{N,z} \leq t) \geq e^{-\lceil N\delta \rceil c_t} \geq e^{-N(\delta + N_0^{-1})c_t}, \quad (4.2.23)$$

provided $N \geq N_0$. Choose N_0 large enough and $\delta > 0$ small enough such that $(\delta + N_0^{-1})c_t \leq \eta/2$. From (4.2.19), there exists a $T_\delta > 0$ such that, possibly increasing N_0 if necessary, if $N \geq N_0$,

$$\inf_{z \in A_\delta} \mathbb{P}(T_\delta^{N,z} \leq T_\delta) \geq e^{-N(\bar{V}_\delta + \eta/2)} \geq e^{-N(\bar{V} + \eta/2)}. \quad (4.2.24)$$

We deduce from (4.2.23), (4.2.24) and the strong Markov property that, with $T = T_\delta + t$,

$$\inf_{z \in A} \mathbb{P}(T_{Ext}^{N,z} \leq T) \geq e^{-N(\bar{V} + \eta)},$$

which is the wished extension of (4.2.19). \square

Lemma 4.2.26. *For any $t > 0$, if $c_t := \log \left(\frac{\lambda - \gamma e^{(\gamma - \lambda)t}}{\gamma - \gamma e^{(\gamma - \lambda)t}} \right)$,*

$$\inf_{z \in A \setminus A_\delta} \mathbb{P}(T_{Ext}^{N,z} < t) \geq \exp \{ -\lceil N\delta \rceil c_t \}.$$

Proof. Since $z \in A \setminus A_\delta$ implies that $z_1 \leq \delta$, the first component of the process $NZ^{N,x}(t)$ is dominated by the process

$$\lceil N\delta \rceil + P_1 \left(N\lambda \int_0^t Z_1^{N,z}(s) ds \right) - P_1 \left(N\gamma \int_0^t Z_1^{N,z}(s) ds \right),$$

which is a continuous time binary branching process with birth rate λ and death rate γ . This process goes extinct before time t with probability

$$\left(\frac{\gamma - \gamma e^{(\gamma - \lambda)t}}{\lambda - \gamma e^{(\gamma - \lambda)t}} \right)^{\lceil N\delta \rceil},$$

as can be seen by combining formula (1) from section III.4 of Athreya and Ney [3] with the formula in section 5 for $F(0, t)$ in the birth and death case. The result follows readily. \square

We shall need below the following additional results.

Proposition 4.2.27. *Under the assumptions of Proposition 4.2.18, if $C \subset \partial A_\delta$ is a closed set such that $V_C := \inf_{z \in C} V(z^*, z) > \bar{V}_\delta$, then for any $z \in \mathring{A}_\delta$, all $\varepsilon > 0$ small enough*

$$\lim_{N \rightarrow \infty} \mathbb{P}(d(Z_{T_\delta^N}^{N,z_N}, C) \leq \varepsilon) = 0.$$

Proof. Fix $\eta < (V_C - \bar{V}_\delta)/3$. From Lemma 4.2.23, for $\varepsilon > 0$ small enough, there exists a $\rho > 0$ small enough and N_0 large enough such that for all $N \geq N_0$,

$$\sup_{2\rho \leq |z - z^*| \leq 3\rho} \mathbb{P}(d(Z_{\sigma_\rho^N}^{N,z_N}, C) \leq \varepsilon) \leq e^{-N(V_C - \eta)}.$$

Let $c = V_C - \eta$ and $T_0 = T(c, \rho)$ given by Lemma 4.2.25. Then, increasing N_0 if necessary, we deduce from that Lemma that for any $N \geq N_0$, $\ell \geq 1$,

$$\mathbb{P}(\tau_\ell \leq \ell T_0) \leq \ell \sup_{z \in A_\delta} \mathbb{P} \left(\sup_{0 \leq t \leq T_0} |Z_t^{N,z_N} - z| \geq \rho \right) \leq \ell e^{-N(V_C - \eta)}.$$

For all $z \in B_\rho$, $\ell \geq 1$,

$$\mathbb{P}(d(Z_{T_\delta^N}^{N,z_N}, C) \leq \varepsilon)$$

$$\begin{aligned}
&\leq \mathbb{P}(T_\delta^{N,z} > \tau_\ell) + \sum_{m=1}^{\ell} \mathbb{P}(T_\delta^{N,z} > \tau_{m-1}) \mathbb{P}(d(Z_{\tau_m}^{N,z_N}, C) \leq \varepsilon | T_\delta^{N,z} > \tau_{m-1}) \\
&\leq \mathbb{P}(T_\delta^{N,z} > \ell T_0) + \mathbb{P}(\tau_\ell \leq \ell T_0) \\
&\quad + \sum_{m=1}^{\ell} \mathbb{P}(T_\delta^{N,z} > \tau_{m-1}) \mathbb{E}[\mathbb{P}(d(Z_{\sigma_\rho^N}^{N,z_N}, C) \leq \varepsilon | T_\delta^{N,z} > \tau_{m-1})] \\
&\leq \mathbb{P}(T_\delta^{N,z} > \ell T_0) + \mathbb{P}(\tau_\ell \leq \ell T_0) + \ell \sup_{2\rho \leq |z-z^*| \leq 3\rho} \mathbb{P}(d(Z_{\sigma_\rho^N}^{N,z_N}, C) \leq \varepsilon) \\
&\leq \mathbb{P}(T_\delta^{N,z} > \ell T_0) + 2\ell e^{-N(V_C - \eta)}.
\end{aligned}$$

Increasing further N_0 if necessary, we have that (4.2.20) holds for some $T > 0$ and all $N \geq N_0$. We choose $\ell = \left\lceil e^{N(\bar{V}_\delta + 2\eta)} \right\rceil$, hence from our choice of η ,

$$\limsup_{N \rightarrow \infty} \sup_{z \in B_\rho} \mathbb{P}(d(Z_{T_\delta}^{N,z_N}, C) \leq \varepsilon) \leq \limsup_{N \rightarrow \infty} \left(\frac{T}{\ell T_0} e^{N(\bar{V}_\delta + \eta)} + 2\ell e^{-N(V_C - \eta)} \right) = 0.$$

It remains to combine Lemma 4.2.24 and the inequality

$$\mathbb{P}(d(Z_{T_\delta}^{N,z_N}, C) \leq \varepsilon) \leq \mathbb{P}(Z_{\sigma_\rho^N}^{N,z_N} \notin B_\rho) + \sup_{y \in B_\rho} \mathbb{P}(d(Z_{T_\delta}^{N,y_N}, C) \leq \varepsilon).$$

□

The proof of the next important result is a bit lengthy, and we refer to Pardoux and Samegni-Kepgnou [27] for it.

Corollary 4.2.28. *If $C \subset \{z, z_1 = 0\}$ is such that $V_C := \inf_{z \in C} V(z^*, z) > \bar{V}$, then for any $z \in \bar{A}$,*

$$\lim_{N \rightarrow \infty} \mathbb{P}(Z_{T_{Ext}^N}^{N,z_N} \in C) = 0.$$

4.2.6 Time of extinction in the SIS model

While the above results are rather precise, it is frustrating that it does not seem possible to express the important constant \bar{V} explicitly in terms of the few constants of the model. One can only do a numerical evaluation of \bar{V} . We now simplify the problem, and consider the SIS model, where when an infectious individual cures, he immediately becomes susceptible again: there is no immunity. The advantage of this simplified model is that it can be written in dimension one and, as we shall see now, we can deduce from the Pontryagin maximum principle, see Section A.6 in the Appendix, a very simple explicit formula for \bar{V} .

The deterministic SIS model can be reduced to the following one-dimensional equation for the proportion of infected individuals

$$\dot{x}_t = \lambda x_t (1 - x_t) - \gamma x_t.$$

Here the process lives in the interval $A_{SIS} = [0, 1]$. Provided $R_0 = \frac{\lambda}{\gamma} > 1$, there is a unique stable endemic equilibrium $x^* = 1 - \frac{\gamma}{\lambda} \in (0, 1)$, while the disease free equilibrium $x^0 = 0$ is unstable. Here $h_1 = 1$, $\beta_1(x) = \lambda x(1-x)$, $h_2 = -1$, $\beta_2(x) = \gamma x$.

We assume that $\lambda > \gamma$, i.e. $R_0 > 1$. As the reader can easily verify, Theorem 4.2.17 applies to this situation, and now \bar{V} is the minimal value of the following control problem. With the notations of Section A.6 below, we are in the situation $d = 1$, $k = 2$, $\beta_1(x) = \lambda x(1-x)$, $\beta_2(x) = \gamma x$, $B = (1 \ -1)$. The identity (A.6.1) reads here

$$\lambda x_t(1-x_t)(1-e^{p_t}) + \gamma x_t(1-e^{-p_t}) = 0.$$

Hence either $p_t = 0$, or else $p_t = \log \frac{\gamma}{\lambda(1-x_t)}$. It is easy to convince oneself that $p_t = 0$ does not produce a control which does the wished job. Hence $p_t = \log \frac{\gamma}{\lambda(1-x_t)}$, $\hat{u}_1(t) = e^{p_t} \beta_1(x_t) = \gamma x_t$, $\hat{u}_2(t) = e^{-p_t} \beta_2(x_t) = \lambda x_t(1-x_t)$. The optimal trajectory reads

$$\dot{x}_t = \gamma x_t - \lambda x_t(1-x_t). \quad (4.2.25)$$

From the right-hand side of the identity (A.6.2),

$$\begin{aligned} \bar{V} &= \int_0^{\hat{T}} [\gamma x_t - \lambda x_t(1-x_t)] \log \frac{\gamma}{\lambda(1-x_t)} dt \\ &= \int_0^{\hat{T}} \log \frac{\gamma}{\lambda(1-x_t)} \dot{x}_t dt \\ &= \int_0^{\frac{\lambda-\gamma}{\lambda}} \log \frac{\gamma}{\lambda(1-x)} dx \\ &= \log \frac{\lambda}{\gamma} - 1 + \frac{\gamma}{\lambda}. \end{aligned}$$

Finally

Proposition 4.2.29. *We have the identities*

$$\bar{V} = \log R_0 - 1 + R_0^{-1}, \quad e^{N\bar{V}} = R_0^N e^{-N(R_0-1)/R_0}.$$

Combining this result with Theorem 4.2.17 adapted to the SIS model yields the following.

Corollary 4.2.30. *Suppose that $R_0 > 1$ and define*

$$T_{Ext}^{N,z} = \inf\{t > 0, Z_t^{N,z} = 0\}.$$

Then for any $0 < z \leq 1$, and $c > 1$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left((R_0/c)^N e^{-N(R_0-1)/R_0} < T_{Ext}^{N,z} < (cR_0)^N e^{-N(R_0-1)/R_0} \right) = 1,$$

$$\text{and } (R_0/c)^N e^{-N(R_0-1)/R_0} \leq \mathbb{E}(T_{Ext}^{N,z}) \leq (cR_0)^N e^{-N(R_0-1)/R_0}$$

for N large enough.

Remark 4.2.31. In fact, the pair $(\hat{u}_1(t), \hat{u}_2(t))$ is not an optimal control for the above control problem. Such an optimal control does not exist! The optimal trajectory, which is the original ODE time reversed, would take an infinite time to leave x^* , and an infinite time to reach 0. However, our $(\hat{u}_1(t), \hat{u}_2(t))$ is the limit of a minimizing sequence obtained by choosing a suboptimal control to drive the system from x^* to $x^* - \delta$, then the optimal control to drive the system from $x^* - \delta$ to δ , and finally a suboptimal control to drive the system from δ to 0. $\log \frac{\lambda}{\gamma} - 1 + \frac{\gamma}{\lambda}$ is indeed the minimal cost. Note that $\hat{T} = +\infty$.

4.2.7 Time of extinction in the SIR model with demography

We now turn to the SIR model with demography, which is the model which has been formally presented in Example 2.2.2, but where we let $\nu = +\infty$ (we suppress the stage E between S and I), and $\gamma = 0$ (there is no loss of immunity). The limiting ODE reads

$$\begin{aligned}\dot{x}_t &= \lambda x_t y_t - \gamma x_t - \mu x_t, \\ \dot{y}_t &= -\lambda x_t y_t + \mu - \mu y_t.\end{aligned}$$

We assume that $\lambda > \gamma + \mu$, in which case there is a unique stable endemic equilibrium, namely $z^* = (x^*, y^*) = (\frac{\mu}{\gamma + \mu}, \frac{\mu}{\lambda} - \frac{\gamma + \mu}{\lambda})$. The extinction in such a model has been studied using the Central Limit Theorem for moderate population size in Section 4.1. We now finally apply Large Deviations to this model. In this model, $Z_t^N = (I_t^N, S_t^N)$ lives in all of \mathbb{R}_+^2 . We note that in the proof of Proposition 4.2.18, the compactness of the set of possible values for Z_t^N has played a crucial role, especially in the proof of Lemma 4.2.22. However, if we define for each $R > 0$

$$T_{\text{Ext}}^{N,R} = T_{\text{Ext}}^N \wedge \sigma_R^N,$$

where $\sigma_R^N = \inf\{t > 0, I_t^N + S_t^N \geq R\}$, it is clear that we have reduced our situation to a bounded state space, and the exact same proofs leading to Proposition 4.2.18 and Theorem 4.2.17, which easily adapted to this new situation. Moreover, we have the

Lemma 4.2.32. As $R \rightarrow \infty$, $V_R := \inf_{z=x+y \geq R} V(z^*, z) \rightarrow \infty$.

Proof. We use the Pontryagin maximum principle and refer to the notations in Section A.6. Here $d = 2$ and $k = 5$, $B = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 & -1 \end{pmatrix}$, $\beta_1(x, y) = \lambda xy$, $\beta_2(x, y) = \gamma x$, $\beta_3(x, y) = \mu x$, $\beta_4(x, y) = \mu$, $\beta_5(x, y) = \mu y$. The forward-backward ODE system reads

$$\dot{x}_t = \lambda x_t y_t e^{p_t - q_t} - (\gamma + \mu) x_t e^{-p_t}, \quad x_0 = \frac{\mu}{\gamma + \mu} - \frac{\mu}{\lambda}$$

$$\begin{aligned}
\dot{y}_t &= -\lambda x_t y_t e^{p_t - q_t} + \mu e^{q_t} - \mu y_t e^{-q_t}, \quad y_0 = \frac{\gamma + \mu}{\lambda} \\
\dot{p}_t &= \lambda y_t + \gamma + \mu - \lambda y_t e^{p_t - q_t} - \gamma e^{-p_t} - \mu e^{-p_t}, \\
\dot{q}_t &= \lambda x_t + \mu - x_t e^{p_t - q_t} - \mu e^{-q_t}, \quad p_{\hat{T}} = q_{\hat{T}}.
\end{aligned}$$

Condition (A.6.1) at time \hat{T} together with the condition $p_{\hat{T}} = q_{\hat{T}}$ allows us to conclude that

$$p_{\hat{T}} = q_{\hat{T}} = \log \left(R + \frac{\gamma}{\mu} x \right).$$

It is clear that $\dot{p}_{\hat{T}} > \dot{q}_{\hat{T}}$. In fact it is not hard to show that, as long as $p_t \geq 0$, $p_t < q_t$. However, $\dot{p}_t \leq \lambda y_t + \gamma + \mu \leq \lambda R + \gamma + \mu$. Let $a = \frac{1}{2} \frac{\log R}{\lambda R + \gamma + \mu}$. For any $\hat{T} - a \leq t \leq \hat{T}$, $p_t \geq \log(R + \frac{\gamma}{\mu} x_{\hat{T}}) - \frac{1}{2} \log R \geq \frac{1}{2} \log R > 0$. Next we notice that $\dot{x}_t + \dot{y}_t \leq \mu e^{q_t}$. As long as $\hat{T} - a \leq t \leq \hat{T}$, we both have that $p_t \leq q_t$ and $q_t \geq 0$, hence $\dot{q}_t \geq 0$, and $0 < q_t \leq q_{\hat{T}} = \log \left(R + \frac{\gamma}{\mu} x_{\hat{T}} \right)$. Consequently $\dot{x}_t + \dot{y}_t \leq \mu \left(R + \frac{\gamma}{\mu} x_{\hat{T}} \right) \leq (\gamma + \mu) R$. Finally, for $\hat{T} - a \leq t \leq \hat{T}$, $x_t + y_t \geq R - \frac{(\gamma + \mu) R \log R}{2(\lambda R + \gamma + \mu)} \geq \frac{1}{2} R$, for R large enough.

We can now lower bound V_R . We use the expression on the left of (A.6.2) for the instantaneous cost. We have

$$\begin{aligned}
V_R &= \int_0^{\hat{T}} \left[\lambda x_t y_t (1 - e^{p_t - q_t} + (p_t - q_t) e^{p_t - q_t}) + (\mu + \gamma) x_t (1 - e^{-p_t} - p_t e^{-p_t}) \right. \\
&\quad \left. + \mu (1 - e^{q_t} + q_t e^{q_t}) + \mu y_t (1 - e^{-q_t} - q_t e^{-q_t}) \right] dt \\
&\geq \int_{\hat{T}-a}^{\hat{T}} \mu (x_t + y_t) \inf \{ 1 - e^{-p_t} - p_t e^{-p_t}, 1 - e^{-q_t} - q_t e^{-q_t} \} dt \\
&\geq \frac{\mu}{8} \frac{R \log R}{\lambda R + \mu + \gamma} \\
&\rightarrow +\infty,
\end{aligned}$$

as $R \rightarrow \infty$. □

It follows from Corollary 4.2.28 that as soon as $V_R > \hat{V}$, the probability that Z^N exits the truncated domain through the “extinction boundary” $\{z^1 = 0\}$ goes to 1 as $N \rightarrow \infty$. Also, for fixed N , $\mathbb{P}(T_{\text{Ext}}^N < \sigma_R^N) \rightarrow 1$, as $R \rightarrow \infty$.

Theorem 4.2.33. *Let $T_{\text{Ext}}^{N,z}$ denote the extinction time in the N -SIR model with demography starting from $z_N = \frac{[zN]}{N}$. Given $\eta > 0$, for all $z \in \mathbb{R}_+^2$ with $z_1 > 0$,*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\exp\{N(\bar{V} - \eta)\} < T_{\text{Ext}}^{N,z} < \exp\{N(\bar{V} + \eta)\}) = 1.$$

Moreover, for all $\eta > 0$, $z \in \mathbb{R}_+^2$ with $z_1 > 0$ and N large enough,

$$\exp\{N(\bar{V} - \eta)\} \leq \mathbb{E}(T_{\text{Ext}}^{N,z}) \leq \exp\{N(\bar{V} + \eta)\}.$$

Appendix

This Appendix presents several mathematical notions, mostly from the theory of stochastic processes, as well as a couple of notions related to continuity of real-valued functions, which are used in the previous chapters. Most proofs are given. Otherwise we refer to existing monographs.

A.1 Branching processes

We present the basic facts about branching processes, which are useful in these Notes. We give most of the proofs. Those which are missing can be found in classical monographs on branching processes, see e.g. Athreya and Ney [3] or Jagers [16], unless we give a precise reference in the text.

A.1.1 Discrete time branching processes

Consider an ancestor (at generation 0) who has ξ_0 children, such that

$$\mathbb{P}(\xi_0 = k) = q_k, \quad k \geq 0 \quad \text{and} \quad \sum_{k \geq 0} q_k = 1.$$

Define $m = \mathbb{E}[\xi_0] = \sum_{k \geq 1} k q_k$ and $g(s) = \mathbb{E}[s^{\xi_0}]$.

Each child of the ancestor belongs to generation 1. The i -th of those children has himself $\xi_{1,i}$ children, where the random variables $\{\xi_{k,i}, k \geq 0, i \geq 1\}$ are i.i.d., all having the same law as ξ_0 . If we define X_n as the number of individuals in generation n , we have

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{n,i}.$$

We have $g(0) = q_0$, $g(1) = 1$, $g'(1) = m$, $g'(s) > 0$, $g''(s) > 0$, for all $0 \leq s \leq 1$ (we assume that $q_0 > 0$ and $q_0 + q_1 < 1$). Let us compute the generating function of X_n : $g_n(s) = \mathbb{E}[s^{X_n}]$.

$$\begin{aligned} g_n(s) &= \mathbb{E} \left[s^{\sum_{i=1}^{X_{n-1}} \xi_{n-1,i}} \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[s^{\sum_{i=1}^{X_{n-1}} \xi_{n-1,i}} \middle| X_{n-1} \right] \right] \\ &= \mathbb{E} [g(s)^{X_{n-1}}] \\ &= g_{n-1} \circ g(s). \end{aligned}$$

If we iterate this argument, we obtain

$$g_n(s) = g \circ \cdots \circ g(s),$$

and also

$$\begin{aligned} \mathbb{P}(X_n = 0) &= g^{\circ n}(0) \\ &= g \left[g^{\circ(n-1)}(0) \right]. \end{aligned}$$

Hence if $z_n = \mathbb{P}(X_n = 0)$, $z_n = g(z_{n-1})$, and $z_1 = q_0$. We have $z_n \uparrow z_\infty$, where $z_\infty =$

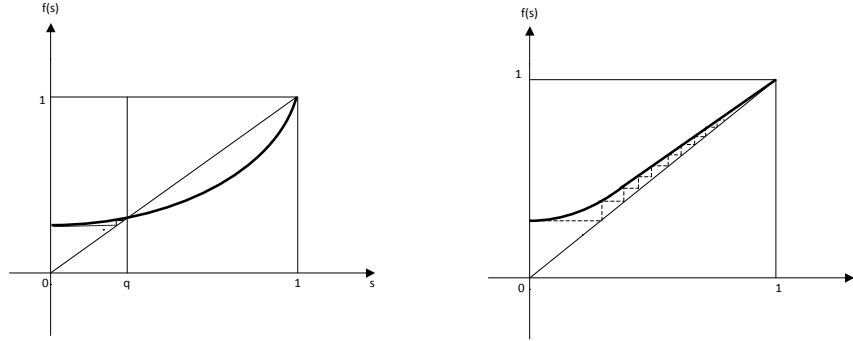


Fig. A.1.1 Graphs of g in case $m > 1$ (left) and in case $m \leq 1$ (right). The successive heights of the dashed line are the successive values of $\mathbb{P}(X_n = 0)$.

$\mathbb{P}(X_n = 0 \text{ from some } n)$. The proof of the following Proposition is essentially clear from Figure A.1.1.

Proposition A.1.1. *If $m \leq 1$, then $\mathbb{P}(X_n = 0) \rightarrow 1$ as $n \rightarrow \infty$, and $z_\infty = 1$.*

If $m > 1$, $\mathbb{P}(X_n = 0) \rightarrow z_\infty = q$ as $n \rightarrow \infty$, where q is the smallest solution of the equation $z = g(z)$.

Note that on the event $\cup_{n=0}^{\infty} \{X_n = 0\}$, which has probability one in the first case, the population goes extinct after a finite number of generations, and the total progeny is finite.

In the second case, with probability $1 - z_{\infty}$, the branching process does not go extinct.

Let us show that $W_n = m^{-n} X_n$ is a martingale.

$$\begin{aligned} \mathbb{E}(W_{n+1} | X_n) &= m^{-n} \mathbb{E} \left(m^{-1} \sum_{i=1}^{X_n} \xi_{n,i} | X_n \right) \\ &= m^{-n} X_n \\ &= W_n. \end{aligned}$$

One can show that $W_n \rightarrow W$ a.s. as $n \rightarrow \infty$, and moreover, provided $\sum_{j \geq 1} q_j j \log j < \infty$,

$$\mathbb{E}[W] = 1, \text{ and } \mathbb{P}(W > 0) = \mathbb{P}(\{\text{the branching process does not go extinct}\}).$$

In the case $\sum_{j \geq 1} q_j j \log j = \infty$, then $\mathbb{P}(W = 0) = 1$.

A.1.2 Continuous time branching processes

We shall consider only binary continuous time branching processes, i.e. where at most one child is born at a given time. This process starts with a single ancestor born at time $t = 0$. This ancestor is characterized by a pair $(L_0, \{N_0(t), t \geq 0\})$, where L_0 is the life length of the ancestor, and $N_0(t)$ is the number of children of the ancestor born on the time interval $[0, t]$. We assume that $N_0(\infty) = N_0(L_0)$, that is the ancestor does not give birth to offspring after his death. We now assume that the individuals are numbered in the order of their birth. To the individual i is attached a pair $(L_i, \{N_i(t)\})$, such that the sequence of pairs $\{(L_i, \{N_i(t)\})\}_{i \geq 0}$ is i.i.d. If the individual i is born at time B_i , the offspring of individual i are born at the jump times of the process $\{N_i(t - B_i), B_i \leq t \leq B_i + L_i\}$. Note that since B_i depends only upon the pairs $\{(L_j, \{N_j(t)\})\}_{0 \leq j < i}$, B_i and $(L_i, \{N_i(t)\})$ are independent.

Let X_t denote the number of individuals in the population alive at time t . This process is Markovian if and only if the law of the pair $(L_i, N_i(t))$ is such that L_i and $\{N_i(t), t \geq 0\}$ are independent, L_i is an exponential random variable with parameter d , and $N_i(t)$ is a rate b Poisson process. We first assume that we are in this situation. We shall denote by X_t^k the number of descendants at time t of k ancestors at time 0. The branching property implies that $\{X_t^k, t \geq 0\}$ is the sum of k independent copies of $\{X_t, t \geq 0\}$. We have the following result.

Proposition A.1.2. *The generating function of the process X is given by*

$$\mathbb{E}(s^{X_t^k}) = \psi_t(s)^k, \quad s \in [0, 1], \quad k \geq 1,$$

where

$$\frac{\partial \psi_t(s)}{\partial t} = \Phi(\psi_t(s)), \quad \psi_0(s) = s,$$

and the function Φ is defined by

$$\begin{aligned} \Phi(s) &= d(1-s) + b(s^2 - s) \\ &= (b+d)(h(s) - s), \quad s \in [0, 1], \end{aligned}$$

where h is the generating function of the probability measure $\frac{d}{b+d}\delta_0 + \frac{b}{d+b}\delta_2$.

Proof. The process X_t is a continuous time \mathbb{Z}_+ -valued jump Markov process. Denote by Q its infinitesimal generator. The non-zero elements of the n -th row of Q are given by

$$Q_{n,m} = \begin{cases} nd, & \text{if } m = n-1, \\ -n(b+d), & \text{if } m = n; \\ nb, & \text{if } m = n+1. \end{cases}$$

Define $f : \mathbb{N} \rightarrow [0, 1]$ by $f(k) = s^k$, $s \in [0, 1]$. Then $\psi_t(s) = P_t f(1) := \mathbb{E}[f(X_t^1)]$ (we use the unusual notation $X^1 = X$ to stress the fact that the process starts from $X_0 = 1$). It follows from the backward Kolmogorov equation for the process X (see e.g. Theorem 3.2, Chapter 7 in Pardoux [25]) that

$$\begin{aligned} \frac{dP_t f(1)}{dt} &= (QP_t f)(1) \\ \frac{\partial \psi_t(s)}{\partial t} &= Q_{1,0} + Q_{1,1}\psi_t(s) + Q_{1,2}\psi_t(s)^2 \\ &= d - (b+d)\psi_t(s) + b\psi_t(s)^2 \\ &= \Phi(\psi_t(s)). \end{aligned}$$

□

Corollary A.1.3. *We have*

$$\mathbb{E}[X_t^k] = ke^{rt}, \quad \text{where } r = b - d.$$

Proof. Differentiating with respect to s the above equation for $\psi_t(s)$ yields

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\partial}{\partial s} \psi_t(s) \right) &= \Phi'(\psi_t(s)) \frac{\partial}{\partial s} \psi_t(s) \\ &= (b+d)(h'(s) - 1) \frac{\partial}{\partial s} \psi_t(s). \end{aligned}$$

The last equation at $s = 1$ yields

$$\frac{d}{dt}\mathbb{E}[X_t] = r\mathbb{E}[X_t],$$

where $X_t = X_t^1$. The result follows for $k = 1$, and then the general case, since the mean number of offspring of k ancestors equals k times the mean number of offspring of one ancestor. \square

The quantity r is often referred to as the Malthusian parameter. It is the mean number of births minus the mean number of death per unit time. Another important quantity is the mean number of offspring of each individual, which is equal to $m = b/d$. The process X_t^k is said to be subcritical if $m < 1$, i.e. $r < 0$. In that case $X_t^k \rightarrow 0$ in $L^1(\Omega)$, and it is easy to show that $X_t^k = 0$ for t large enough. This last conclusion holds in the critical case ($m = 1$, i.e. $r = 0$) as well. In those two cases, the total progeny is finite a.s. We now study the large time behaviour of X_t^k in the supercritical case. In the next proposition, we again write X_t for X_t^1 .

Proposition A.1.4. *If $m > 1$, or equivalently $r > 0$, there exists a non-negative random variable W such that $X_t \sim We^{rt}$ almost surely, as $t \rightarrow \infty$. Moreover $\{W = 0\} = \{\exists t > 0 \text{ such that } X_t = 0\}$ and*

$$\mathbb{P}(W = 0) = \mathbb{P}(\{\exists t > 0 \text{ such that } X_t = 0\}) = \frac{d}{b}.$$

Proof. The first part of the result follows readily from the fact that $e^{-rt}X_t$ is a positive martingale, which converges a.s. to a limit W as $t \rightarrow \infty$. Moreover it is not hard to show that $\sup_{t>0} \mathbb{E}[e^{-2rt}(X_t)^2] < \infty$, hence the convergence holds in $L^1(\Omega)$, so $\mathbb{E}[W] = 1$. Now clearly $\{\exists t > 0 \text{ s.t. } X_t = 0\} \subset \{W = 0\}$. If we start with k ancestors, the limiting W is clearly the sum of k i.i.d. copies of W when starting with one ancestor, and $\mathbb{P}_k(W = 0) = (\mathbb{P}_1(W = 0))^k$. It is now easy to deduce that

$$\mathbb{P}_1(W = 0|X_t) = \mathbb{P}_1(W = 0)^{X_t}.$$

Taking the expectation in this identity and writing $q = \mathbb{P}_1(W = 0)$, we obtain $q = \mathbb{E}[q^{X_t}]$. Differentiating that identity at $t = 0$ and taking advantage of Proposition A.1.2, we deduce that q solves $bq^2 - (b+d)q + d = 0$. Moreover since $\mathbb{E}(W) = 1$, $q < 1$, hence $q = d/b$. Finally $\mathbb{P}(X_t = 0) = \psi_t(0)$ is the solution of the ODE $\dot{x}(t) = bx(t)^2 - (b+d)x(t) + d$, $x(0) = 0$. It is clear that as $t \rightarrow \infty$, $\psi_t(0)$ increases to the smallest solution of the equation $bs^2 - (b+d)s + d = 0$, again d/b . \square

We now consider non-Markovian continuous time binary branching processes. The non-Markovian continuous time branching processes which we have described at the beginning of this section are called Crump–Mode–Jagers processes. Now the law of the pairs $(L_i, \{N_i(t)\})$ can be quite general. For the application to epidemics models, we can consider the case where L_i and $\{N_i(t)\}$ are independent, N_i being a Poisson process, but the law of L_i is no longer exponential. We denote again by $m = \mathbb{E}[N_0(L_0)]$ the mean number of offspring of each individual. Of course, the process is

subcritical, critical, or supercritical according as $m < 1$, $m = 1$ or $m > 1$. We denote again by X_t the number of individuals alive at time t . We define $F(t) = \mathbb{E}[N(t)]$ and $G(t) = \mathbb{P}(L \leq t)$. We assume that F is non-lattice, and $F(0^+) < 1$. Doney [10] showed the following two results.

Proposition A.1.5. *If $1 < m < \infty$, then there exists a unique $r > 0$ such that*

$$\int_0^\infty e^{-rt} F(dt) = 1$$

and $\mathbb{E}[X_t] \sim ae^{rt}$, where

$$0 < a = \frac{\int_0^\infty (1 - G(t))e^{-rt} dt}{\int_0^\infty te^{-rt} F(dt)} < \infty.$$

Again r is called the Malthusian parameter. In the next statement, we use the notation

$$Y = \int_0^\infty e^{-rt} N(dt).$$

It is clear that

$$\mathbb{E}[Y] = \int_0^\infty e^{-rt} F(dt) = 1.$$

Theorem A.1.6. *Suppose that $1 < m < \infty$. Then, as $t \rightarrow \infty$*

$$\frac{X_t}{\mathbb{E}[X_t]} \rightarrow W \quad \text{in law.}$$

W is not identically 0 if and only if $\mathbb{E}[Y \log(Y)] < \infty$, in which case $\mathbb{E}[W] = 1$ and $\mathbb{P}(W = 0) = \mathbb{P}(\{\exists t > 0 \text{ s.t. } X_t = 0\})$. Moreover, the law of W has an atom at 0 and is absolutely continuous on $(0, \infty)$.

A.2 The Poisson process and Poisson point process

The Poisson process is central in this whole volume. Let $\lambda > 0$ be given. A rate λ Poisson (counting) process is defined as

$$P_t = \sup\{k \geq 1, T_k \leq t\},$$

where $0 = T_0 < T_1 < T_2 < \dots < T_k < \dots < \infty$, the random variables $\{T_k - T_{k-1}, k \geq 1\}$ being independent and identically distributed, each following the law $\text{Exp}(\lambda)$. We have

Proposition A.2.1. *For all $n \geq 1$, $0 < t_1 < t_2 < \dots < t_n$, the random variables $P_{t_1}, P_{t_2} - P_{t_1}, \dots, P_{t_n} - P_{t_{n-1}}$ are independent, and for all $1 \leq k \leq n$, $P_{t_k} - P_{t_{k-1}} \sim \text{Poi}[\lambda(t_k - t_{k-1})]$.*

Proof. Let us first prove that for all $t, s > 0$,

$$\mathbb{P}(P_{t+s} - P_t = 0 | P_t = k, T_1, T_2, \dots, T_k) = \exp(-\lambda s).$$

Indeed

$$\begin{aligned} \mathbb{P}(P_{t+s} - P_t = 0 | P_t = k, T_1, T_2, \dots, T_k) \\ &= \mathbb{P}(T_{k+1} > t + s | P_t = k, T_k) \\ &= \mathbb{P}(T_{k+1} - T_k > t + s - T_k | T_{k+1} - T_k > t - T_k > 0) \\ &= \mathbb{P}(T_{k+1} - T_k > s) \\ &= \exp(-\lambda s). \end{aligned}$$

Let now $n \geq 1$. For $1 \leq i \leq n$, we define $X_{n,i} = \mathbf{1}_{\{P_{t+is/n} - P_{t+(i-1)s/n} \geq 1\}}$, and finally $S_n = X_{n,1} + X_{n,2} + \dots + X_{n,n}$. It follows from the first part of the proof that conditionally upon $\sigma\{P_r, 0 \leq r \leq t\}$, the random variables $X_{n,1}, X_{n,2}, \dots, X_{n,n}$ are i.i.d., each Bernoulli with parameter $1 - e^{-\lambda s/n}$. Then conditionally upon $\sigma\{P_r, 0 \leq r \leq t\}$, S_n is binomial with parameters $(n, 1 - e^{-\lambda s/n})$. But $S_n \rightarrow P_{t+s} - P_t$ a.s. as $n \rightarrow \infty$, while its conditional law given $\sigma\{P_r, 0 \leq r \leq t\}$ converges towards the Poisson distribution with parameter λs , according to the following lemma. The proposition follows. \square

We have used the following well-known result. Recall the notation $\text{Bin}(n, p)$ for the binomial law with parameters n and p , where $n \geq 1$ and $0 < p < 1$.

Lemma A.2.2. *For all $n \geq 1$, let U_n be a $\text{Bin}(n, p_n)$ random variable. If $np_n \rightarrow \lambda$ as $n \rightarrow \infty$, with $\lambda > 0$, then U_n converges in law towards $\text{Poi}(\lambda)$.*

A Poisson process will be called standard if its rate is 1. If P is a standard Poisson process, then $\{P(\lambda t), t \geq 0\}$ is a rate λ Poisson process.

We will also use the following

Exercise A.2.3. Let $\{P_t, t \geq 0\}$ be a rate λ Poisson process, and $\{T_k, k \geq 1\}$ the random points of this Poisson process, i.e. for all $t > 0$, $P_t = \sup\{k \geq 1, T_k \leq t\}$. Let $0 < p < 1$. Suppose that each T_k is selected with probability p , not selected with probability $1 - p$, independently from the others. Let P'_t denote the number of selected points on the interval $[0, t]$. Then $\{P'_t, t \geq 0\}$ is a rate λp Poisson process.

A rate λ Poisson process ($\lambda > 0$) is a counting process $\{R_t, t \geq 0\}$ such that $R_t - \lambda t$ is a martingale. Let $\{P(t), t \geq 0\}$ be a standard Poisson process (i.e. with rate 1). Then $P(\lambda t) - \lambda t$ is martingale, and it is not hard to show that $\{P(\lambda t), t \geq 0\}$ is a rate λ Poisson process. Let now $\{\lambda(t), t \geq 0\}$ be a measurable and locally integrable \mathbb{R}_+ -valued function. Then the process $\{R_t := P(\int_0^t \lambda(s) ds), t \geq 0\}$ is called a rate $\lambda(t)$ Poisson process. Clearly $R_t - \int_0^t \lambda(s) ds$ is a martingale.

We now want to consider the case where λ is random. For that purpose, it is convenient to give an alternative definition of the above process R_t .

Consider a standard Poisson random measure Q on \mathbb{R}_+^+ , which is defined as follows. M is the counting process associated to a random cloud of points in \mathbb{R}_+^2 . One

way to construct that cloud of points is as follows. We can consider $\mathbb{R}_+^2 = \cup_{i=1}^\infty A_i$, where the A_i 's are disjoint squares with Lebesgue measure 1. Let $K_i, i \geq 1$ be i.i.d. mean one Poisson random variables. Let $\{X_j^i, j \geq 1, i \geq 1\}$ be independent random points of \mathbb{R}_+^2 , which are such that for any $i \geq 1$, the X_j^i 's are uniformly distributed in A_i . Then

$$Q(dx) = \sum_{i=1}^\infty \sum_{j=1}^{K_i} \delta_{X_j^i}(dx).$$

$\lambda(t)$ denoting a positive-valued measurable function, the above $\{R_t, t \geq 0\}$ has the same law as

$$R_t = \int_0^t \int_0^{\lambda(s)} Q(ds, du).$$

Now let $\{\lambda(t), t \geq 0\}$ be an \mathbb{R}_+ -valued stochastic process, which is assumed to be predictable, in the following sense. Let for $t \geq 0$,

$$\mathcal{F}_t = \sigma\{Q(A), A \text{ Borel subset of } [0, t] \times \mathbb{R}_+\},$$

and consider the σ -algebra of subset of $[0, \infty) \times \Omega$ generated by the subsets of the form $\mathbf{1}_{(s,t]} \mathbf{1}_F$, where $0 \leq s < t$ and $F \in \mathcal{F}_s$, which is called the predictable σ -algebra. Note that if X_t is \mathcal{F}_t -progressively measurable and left-continuous, then it is predictable. If X_t is progressively measurable and right-continuous, then X_{t-} is predictable.

We assume moreover that $\mathbb{E} \int_0^t \lambda(s) ds < \infty$ for all $t > 0$. We now define the process R_t as above:

$$R_t = \int_0^t \int_0^{\lambda(s)} Q(ds, du).$$

We have (see the next subsection for the definition of a martingale)

Lemma A.2.4. $R_t - \int_0^t \lambda(s) ds$ is a martingale.

Proof. For any $\delta > 0$, let

$$R_t^\delta = \int_0^t \int_0^{\lambda(s-\delta)} Q(ds, du),$$

where $\lambda(s) = 0$ for $s < 0$. It is not hard to show that $R_t^\delta - \int_0^t \lambda(s-\delta) ds$ is a martingale which converges in $L^1(\Omega)$ to $R_t - \int_0^t \lambda(s) ds$. Indeed, it suffices to show that if $0 < s < t$ with $t-s \leq \delta$, the restriction of the random measure M to $(s, t] \times (0, +\infty)$ is independent of $\{\lambda(r-\delta), s < r \leq t\}$, which is \mathcal{F}_s measurable hence

$$\mathbb{E}^{\mathcal{F}_s}(R_t^\delta - R_s^\delta) = \mathbb{E}^{\mathcal{F}_s} \int_s^t \lambda(r-\delta) dr.$$

The result follows. \square

The process R_t is sometimes called “a doubly stochastic Poisson process” or a Cox process. Of course the increments of R_t are not Poisson distributed. If we let

$\sigma(t) = \inf\{r > 0, \int_0^r \lambda(s)ds > t\}$, we have that $P(t) := R_{\sigma(t)}$ is a standard Poisson process, and it is clear that $R_t = P(\int_0^t \lambda(s)ds)$.

In particular, the process which counts the new infections, which appears in Section 2.2, takes the form

$$P\left(\frac{\lambda}{N} \int_0^t I(r)S(r)dr\right) = \int_0^t \int_0^\infty \mathbf{1}_{u \leq \frac{\lambda}{N} I(r-)S(r-)} Q(ds, du).$$

If we let $\bar{Q}(ds, du) = Q(ds, du) - ds \times du$ and $M(t) := P(t) - t$, it is clear that, as a consequence of the above Lemma, we have

Corollary A.2.5. *Define $M(\cdot)$ by*

$$\begin{aligned} M\left(\frac{\lambda}{N} \int_0^t I(r)S(r)dr\right) &= \int_0^t \int_0^\infty \mathbf{1}_{u \leq \frac{\lambda}{N} I(r-)S(r-)} \bar{Q}(ds, du) \\ &= \int_0^t \int_0^\infty \mathbf{1}_{u \leq \frac{\lambda}{N} I(r-)S(r-)} Q(ds, du) - \frac{\lambda}{N} \int_0^t I(r)S(r)dr. \end{aligned}$$

Then $M(t)$ is a martingale (see Definition A.4.7 below).

Note that $\int_0^t I(r-)S(r-)dr = \int_0^t I(r)S(r)dr$ since the two integrands coincide dr a.e. since they differ on each interval $[0, t]$ at most at finitely many points. We use the second formulation, since it is simpler.

A.3 Cramér's theorem for Poisson random variables

In order to explain what Large Deviations is about, let us first establish Cramér's Theorem, in the particular case of Poisson random variables. Let $X_1, X_2, \dots, X_n, \dots$ be mutually independent $\text{Poi}(\mu)$ random variables. The Law of Large Numbers tells us that

$$\frac{1}{N} \sum_{i=1}^N X_i \rightarrow \mu \quad \text{a.s. as } N \rightarrow \infty.$$

Let us first define, for $X \sim \text{Poi}(\mu)$ the logarithm of its Laplace transform

$$\Lambda(\lambda) = \log \mathbb{E}[\exp(\lambda X)] = \mu(e^\lambda - 1),$$

and the Fenchel–Legendre transform of the latter

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\} = x \log \left(\frac{x}{\mu} \right) - x + \mu.$$

Note that the minimum of Λ^* is achieved at $x = \mu$, and Λ^* is zero at that point.

Let v_N denote the law of the random variable $\frac{1}{N} \sum_{i=1}^N X_i$. We can now state Cramér's theorem.

Theorem A.3.1. *Let $F \subset \mathbb{R}$ be a closed set.*

$$\text{For any } N \geq 1, \quad v_N(F) \leq \exp \left(-N \inf_{x \in F} \Lambda^*(x) \right).$$

$$\text{Hence } \limsup_{N \rightarrow \infty} \frac{1}{N} \log v_N(F) \leq - \inf_{x \in F} \Lambda^*(x).$$

Let $G \subset \mathbb{R}$ be an open set.

$$\text{For any } N \geq 1, \quad v_N(G) \geq \exp \left(-N \inf_{x \in G} \Lambda^*(x) \right).$$

$$\text{Hence } \liminf_{N \rightarrow \infty} \frac{1}{N} \log v_N(G) \geq - \inf_{x \in G} \Lambda^*(x).$$

Proof. FIRST STEP. PROOF OF THE UPPER BOUND Let $X_1, X_2, \dots, X_n, \dots$ be mutually independent $\text{Poi}(\mu)$ random variables. For $\sigma > \mu$, we want to estimate

$$\mathbb{P} \left(\frac{1}{N} \sum_{i=1}^N X_i \geq \sigma \right),$$

which is the probability of a Large Deviation from the LLN, since we know that for large N , $\frac{1}{N} \sum_{i=1}^N X_i \simeq \mu$.

For any $\lambda > 0$, using Chebycheff's inequality,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{N} \sum_{i=1}^N X_i \geq \sigma \right) &= \mathbb{P} \left(\exp \left\{ \lambda \left(\sum_{i=1}^N X_i - N\sigma \right) \right\} \geq 1 \right) \\ &\leq \mathbb{E} \exp \left\{ \lambda \left(\sum_{i=1}^N X_i - N\sigma \right) \right\} \\ &= \exp [-N(\lambda\sigma - \Lambda(\lambda))]. \end{aligned}$$

The best possible upper bound is then (since with $\sigma > \mu$, $\Lambda^*(\sigma)$ is obtained by taking the supremum over $\lambda > 0$)

$$\begin{aligned} \mathbb{P} \left(\frac{1}{N} \sum_{i=1}^N X_i \geq \sigma \right) &\leq e^{-N\Lambda^*(\sigma)} \\ &= \exp \left[-N \left(\sigma \log \left(\frac{\sigma}{\mu} \right) - \sigma + \mu \right) \right]. \end{aligned}$$

Similarly, if $\sigma < \mu$, for any $\lambda < 0$,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{N} \sum_{i=1}^N X_i \leq \sigma \right) &\leq \mathbb{E} \exp \left\{ \lambda \left(\sum_{i=1}^N X_i - N\sigma \right) \right\} \\ &= \exp [-N(\lambda\sigma - \Lambda(\lambda))]. \end{aligned}$$

Since with $\sigma < \mu$, $\Lambda^*(\sigma)$ is obtained by taking the supremum over $\lambda < 0$, the above computation leads again to

$$\mathbb{P}\left(\frac{1}{N}\sum_{i=1}^N X_i \leq \sigma\right) \leq \exp\left[-N\left(\sigma \log\left(\frac{\sigma}{\mu}\right) - \sigma + \mu\right)\right].$$

It is not hard to see that the upper bound follows from the two above estimates.

SECOND STEP. PROOF OF THE LOWER BOUND For any $\delta > 0$,

$$\nu_N((-\delta, \delta)) \geq \nu_N(\{0\}) = e^{-N\mu}, \quad \text{hence } \frac{1}{N} \log \nu_N((-\delta, \delta)) \geq -\mu = -\Lambda^*(0).$$

Since transforming X into $Y = X - x$ results in Λ and Λ^* being transformed into $\Lambda_Y(\lambda) = \Lambda(\lambda) - \lambda x$ and $\Lambda_Y^*(\cdot) = \Lambda^*(\cdot + x)$, the above yields that for all $x > 0$,

$$\frac{1}{N} \log \nu_N((x - \delta, x + \delta)) \geq -\Lambda^*(x).$$

The lower bound follows readily. □

A.4 Martingales

A.4.1 Martingales in discrete time

$(\Omega, \mathcal{F}, \mathbb{P})$ being our standing probability space, let be given an increasing sequence $\{\mathcal{F}_n, n \geq 0\}$ of sub- σ -algebras of \mathcal{F} .

Definition A.4.1. A sequence $\{X_n, n \geq 0\}$ of random variables is called a martingale if

1. For all $n \geq 0$, X_n is \mathcal{F}_n -measurable and integrable,
2. For all $n \geq 0$, $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$ a. s.

A sub-martingale is a sequence which satisfies the first condition and $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \geq X_n$. A super-martingale is a sequence which satisfies the first condition and $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \leq X_n$.

It follows readily from Jensen's inequality for conditional expectations the

Proposition A.4.2. *If $\{X_n, n \geq 0\}$ is a martingale, and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function such that $\varphi(X_n)$ is integrable for all $n \geq 0$, then $\{\varphi(X_n), n \geq 0\}$ is a sub-martingale.*

We shall need the notion of stopping time

Definition A.4.3. A stopping time τ is an $\mathbb{Z}_+ \cup \{+\infty\}$ -valued random variable which satisfies $\{\tau = n\} \in \mathcal{F}_n$, for all $n \geq 0$.

We also let

$$\mathcal{F}_\tau = \{B \in \mathcal{F}, B \cap \{\tau = n\} \in \mathcal{F}_n, \forall n \in \mathbb{Z}_+\}.$$

We have Doob's optional sampling theorem:

Theorem A.4.4. *If $\{X_n, n \geq 0\}$ is a martingale (resp. a sub-martingale), and τ_1, τ_2 two stopping times s.t. $\tau_1 \leq \tau_2 \leq N$ a.s., then X_{τ_i} is \mathcal{F}_{τ_i} measurable and integrable, $i = 1, 2$, and moreover*

$$\begin{aligned} \mathbb{E}(X_{\tau_2} | \mathcal{F}_{\tau_1}) &= X_{\tau_1} \\ (\text{resp. } \mathbb{E}(X_{\tau_2} | \mathcal{F}_{\tau_1}) &\geq X_{\tau_1}). \end{aligned}$$

Proof. For all $A \in \mathcal{B}, n \geq 0$,

$$\{X_{\tau_i} \in A\} \cap \{\tau_i = n\} = \{X_n \in A\} \cap \{\tau_i = n\} \in \mathcal{F}_n,$$

and moreover

$$|X_{\tau_i}| \leq \sum_{k=1}^N |X_k|,$$

which establishes the first part of the statement.

Let $A \in \mathcal{F}_{\tau_1}$. Then

$$A \cap \{\tau_1 < k \leq \tau_2\} = A \cap \{\tau_1 \leq k-1\} \cap \{\tau_2 \leq k-1\}^c \in \mathcal{F}_{k-1}.$$

Indeed, we have

$$A \cap \{\tau_1 \leq k-1\} = \bigcup_{j=1}^{k-1} A \cap \{\tau_1 = j\} \in \mathcal{F}_{k-1}, \text{ and } \{\tau_2 \leq k-1\}^c \in \mathcal{F}_{k-1}.$$

Let $\Delta_k = X_k - X_{k-1}$. We have, with $A \in \mathcal{F}_{\tau_1}$,

$$\begin{aligned} \int_A (X_{\tau_2} - X_{\tau_1}) d\mathbb{P} &= \int_A \sum_{k=1}^n \mathbf{1}_{\{\tau_1 < k \leq \tau_2\}} \Delta_k d\mathbb{P} \\ &= \sum_{k=1}^n \int_{A \cap \{\tau_1 < k \leq \tau_2\}} \Delta_k d\mathbb{P} \\ &= 0 \end{aligned}$$

or else ≥ 0 , depending upon whether $\{X_n, n \geq 0\}$ is a martingale or a sub-martingale. \square

We have a first Doob's inequality

Proposition A.4.5. *If X_1, \dots, X_n is a sub-martingale, then for all $\alpha > 0$,*

$$\mathbb{P}\left(\max_{1 \leq i \leq n} X_i \geq \alpha\right) \leq \frac{1}{\alpha} \mathbb{E}(X_n^+).$$

Proof. Define the stopping time $\tau = \inf\{0 \leq k \leq n, X_k \geq \alpha\}$ and let $M_k = \max_{1 \leq i \leq k} X_i$. We have

$$\{M_n \geq \alpha\} \cap \{\tau \leq k\} = \{M_k \geq \alpha\} \in \mathcal{F}_k.$$

Hence $\{M_n \geq \alpha\} \in \mathcal{F}_\tau$. From the optional sampling Theorem,

$$\begin{aligned} \alpha \mathbb{P}(M_n \geq \alpha) &\leq \int_{\{M_n \geq \alpha\}} X_\tau d\mathbb{P} \\ &\leq \int_{\{M_n \geq \alpha\}} X_n d\mathbb{P} \\ &\leq \int_{\{M_n \geq \alpha\}} X_n^+ d\mathbb{P} \\ &\leq \mathbb{E}(X_n^+). \end{aligned}$$

□

We have finally a second Doob's inequality

Proposition A.4.6. *If M_1, \dots, M_n is a martingale, then*

$$\mathbb{E} \left[\sup_{0 \leq k \leq n} |M_k|^2 \right] \leq 4 \mathbb{E} [|M_n|^2].$$

Proof. Let $X_k = |M_k|$. From Proposition A.4.2, X_1, \dots, X_n is a sub-martingale. It follows from the proof of Proposition A.4.5 that, with the notation $X_k^* = \sup_{0 \leq k \leq n} X_k$,

$$\mathbb{P}(X_n^* > \lambda) \leq \frac{1}{\lambda} \mathbb{E} (X_n \mathbf{1}_{X_n^* > \lambda}).$$

Consequently

$$\begin{aligned} \int_0^\infty \lambda \mathbb{P}(X_n^* > \lambda) d\lambda &\leq \int_0^\infty \mathbb{E} (X_n \mathbf{1}_{X_n^* > \lambda}) d\lambda \\ \mathbb{E} \left(\int_0^{X_n^*} \lambda d\lambda \right) &\leq \mathbb{E} \left(X_n \int_0^{X_n^*} d\lambda \right) \\ \frac{1}{2} \mathbb{E} [|X_n^*|^2] &\leq \mathbb{E} (X_n X_n^*) \\ &\leq \sqrt{E(|X_n|^2)} \sqrt{E(|X_n^*|^2)}, \end{aligned}$$

from which the result follows. □

A.4.2 Martingales in continuous time

We are now given an increasing collection $\{\mathcal{F}_t, t \geq 0\}$ of sub- σ -algebras in continuous time.

Definition A.4.7. A process $\{X_t, t \geq 0\}$ is called a martingale if

1. for all $t \geq 0$, X_t is \mathcal{F}_t -measurable and integrable;
2. for all $0 \leq s < t$, $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$ a. s.

A sub-martingale is a process which satisfies the first condition and $\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s$. A super-martingale is a process which satisfies the first condition and $\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$.

Suppose $\{M_t, t \geq 0\}$ is a right-continuous martingale. For any $n \geq 1$, $0 = t_0 < t_1 < \dots < t_n$, $(M_{t_0}, M_{t_1}, \dots, M_{t_n})$ is a discrete time martingale, to which Proposition A.4.6 applies. Since

$$\sup_{0 \leq s \leq t} |M_s| = \sup_{\text{Partitions of } [0, t]} \sup_{1 \leq k \leq n} |M_{t_k}|,$$

Proposition A.4.6 implies readily

Proposition A.4.8. *If $\{M_t, t \geq 0\}$ is a right-continuous martingale,*

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |M_s|^2 \right] \leq 4 \mathbb{E} [|M_t|^2].$$

We now establish a particular (essentially obvious) instance of Itô's formula. Recall that an \mathbb{R} -valued function of t has locally bounded variations if and only if it is the difference of an increasing and a decreasing function. This class of functions excludes all non-zero continuous martingales, e.g. Brownian motion. But all processes considered in these Notes, except for the limit in the functional central limit theorem, are locally of bounded variations. Given such a locally bounded variation right-continuous 1-dimensional process X_t , we define the bracket $[X, X]_t = \sum_{0 \leq s \leq t} |\Delta X_s|^2$, where $\Delta X_s = X_s - X_{s-}$ is the jump of X at time s . It follows from the fact that X has bounded variation on any compact interval that the set $\{s \geq 0, \Delta X_s \neq 0\}$ is at most countable, hence the above sum makes sense. If X and Y are two processes of the above type, then

$$[X, Y]_t = \sum_{0 \leq s \leq t} \Delta X_s \Delta Y_s = \frac{1}{2} ([X + Y, X + Y]_t - [X, X]_t - [Y, Y]_t).$$

Now we have what we call Itô's formula. If X_t and Y_t are right-continuous and have left limits at any t , have bounded variations on any compact interval, then for any $t > 0$,

$$X_t Y_t = X_0 Y_0 + \int_0^t X_{s-} dY_s + \int_0^t Y_{s-} dX_s + [X, Y]_t. \quad (\text{A.4.1})$$

In case all jumps of X and Y are isolated, which is the only situation treated in these Notes, the result follows clearly by analyzing the evolution of both sides of the identity between the jumps, and at the jump times. The result in the more general situation is easily deduced by approximation.

If M_t is a right-continuous \mathbb{R} -valued martingale with locally bounded variation, we define as above its quadratic variation as

$$[M, M]_t = \sum_{0 \leq s \leq t} |\Delta M_s|^2,$$

and $\langle M, M \rangle_t$ as the unique increasing predictable process such that $[M, M]_t - \langle M, M \rangle_t$ is a martingale. Note that both $M_t^2 - [M, M]_t$ and $M_t^2 - \langle M, M \rangle_t$ are martingales. Consequently, we have in particular

Proposition A.4.9. *Let M_t be a square-integrable right-continuous \mathbb{R} -valued martingale with finite variation such that $M_0 = 0$. Then for all $t > 0$,*

$$\mathbb{E}(M_t^2) = \mathbb{E}\left(\sum_{0 \leq s \leq t} |\Delta M_s|^2\right).$$

A.5 Tightness and weak convergence in path space

In these Notes we consider continuous time processes with values in \mathbb{R}^d . Most of our processes are discontinuous. Their trajectories belong to the set $D([0, +\infty); \mathbb{R}^d)$ of functions which are right continuous and have left limits at any point $t \in [0, +\infty)$. It is not very convenient to use the topology of locally uniform convergence on this set, since we would like for instance the two Heaviside type functions $\mathbf{1}_{[1, +\infty)}(t)$ and $\mathbf{1}_{[1+\varepsilon, +\infty)}(t)$ to be close for ε small. The Skorokhod topology essentially says that two functions are close if after a time change which is close to the identity, they are (at least locally) close in the supremum topology. The only weak convergence (i.e. convergence in law) results we consider in these Notes are convergence results towards a continuous process. In this case, convergence in the sense of the Skorokhod topology is equivalent to locally uniform convergence.

Note also that weak convergence of a sequence of processes X^n towards X is equivalent to the two following facts:

1. The sequence $\{X^n\}_{n \geq 1}$ is tight, as a sequence of random elements of $D([0, +\infty); \mathbb{R}^d)$ equipped with the Skorokhod topology.
2. For any $k \geq 1$, $0 \leq t_1 < t_2 < \dots < t_k$, $(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow (X_{t_1}, \dots, X_{t_k})$, in the sense of weak convergence in $\mathbb{R}^{d \times k}$.

If only 2 is satisfied, then one has convergence in the sense of finite-dimensional distributions.

What do we mean by tightness? A sequence $\{X_n\}_{n \geq 1}$ of random variables with values in a topological space S is said to be tight if for any $\varepsilon > 0$, there exists a compact set $K \subset S$ such that $\mathbb{P}(X_n \in K) \geq 1 - \varepsilon$ for all $n \geq 1$.

Consider the product $X_n Y_n$, where X_n and Y_n are real-valued. If one of the two sequences is tight and the other tends to 0 in probability, then $X_n Y_n \rightarrow 0$ in probability. This easy result is used in the proof of Theorem 2.3.2.

In the proof of Lemma 2.3.4, we use the following argument: a sequence of continuous time martingales M_t^n satisfying $M_0^n = 0$ is tight as soon as the associated sequence of predictable increasing processes $\langle M^n, M^n \rangle_t$ is C -tight, in the sense that both it is tight, and any weak limit of a converging sub-sequence is continuous, see e.g. Theorem VI.4.13 in Jacod and Shiryaev [15]. In the situation of Lemma 2.3.4, $\langle M^n, M^n \rangle_t = t$ which is C -tight, since it does not depend upon n and is continuous.

A.6 Pontryagin's maximum principle

In this section, we present the Pontryagin maximum principle in optimal control, which is useful in order to compute or give some estimates for the exponent in the asymptotic evaluation of the time to extinction derived from large deviation theory. We refer the reader for a more general presentation, proofs and references to Trélat [37] and Pontryagin et al. [31].

The quantity of interest, denoted by \bar{V} in Section 4.2.5 and the following pages, is the value function of an optimal control problem which is of the following type. $x \in C([0, \infty); \mathbb{R}^d)$ solves the controlled ODE

$$\dot{x}_t = Bu_t, \quad x_0 = x^*,$$

where B is a $d \times k$ matrix, and $u \in L^1([0, \infty); \mathbb{R}_+^k)$ is to be chosen together with the final time T such as to minimize a cost functional

$$C(u) = \sum_{j=1}^k \int_0^T g(u_j(t), \beta_j(x_t)) dt,$$

while the following constraint must be satisfied: $x_T \in M_1$, where M_1 is some affine subspace of \mathbb{R}^d . The function g is the one which appears in Section 4.2.2, namely $g(a, b) = a \log(a/b) - a + b$, while the β_j 's are some mappings from \mathbb{R}^d into \mathbb{R}_+ , which, like the matrix B , depend upon the particular model we consider. Note that in our case all entries of B are either 1, 0, or -1 .

We associate to this optimal control problem a Hamiltonian which takes the form

$$H(x, p, u) = \langle p, Bu \rangle - \sum_{j=1}^k g(u_j, \beta_j(x)),$$

where $p \in C([0, T]; \mathbb{R}^d)$ is the adjoint state. The next statement constitutes Pontryagin maximum principle, applied to our particular situation.

Theorem A.6.1. *If (\hat{u}, \hat{T}) is an optimal pair, then there exists an adjoint state, such that the following is satisfied*

$$\begin{aligned} \dot{x}_t &= B\hat{u}_t, \quad x_0 = x^*, \quad x_{\hat{T}} \in M_1, \\ \dot{p}_t &= \sum_{j=1}^k \left[\nabla \beta_j(x_t) - \hat{u}_j(t) \frac{\nabla \beta_j(x_t)}{\beta_j(x_t)} \right], \quad p_{\hat{T}} \perp M_1, \\ H(x_t, p_t, \hat{u}_t) &= \max_{v \in \mathbb{R}_+^k} H(x_t, p_t, v) = 0, \quad 0 \leq t \leq \hat{T}. \end{aligned}$$

Of course, the first equation could be of the more general form $\dot{x} = f(x, u)$. The general form of the adjoint equation reads $\dot{p} = -\nabla_x H$. The Hamiltonian is zero at time \hat{T} since the final time is not fixed and there is no final cost. The Hamiltonian is constant along the optimal trajectory because none of the coefficients depends upon t .

Since $u \rightarrow (B^* p)_j u - g(u, \beta_j(x))$ is concave, the maximum is the zero of its derivative if it is non-negative. Hence

$$\hat{u}_j = e^{(B^* p)_j} \beta_j(x),$$

and the two above equations can be written as

$$\dot{x}_t = \sum_{j=1}^k e^{(B^* p)_j} \beta_j(x_t) h_j, \quad \dot{p}_t = \sum_{j=1}^k (1 - e^{(B^* p)_j}) \nabla \beta_j(x_t),$$

and the Hamiltonian along the optimal trajectory reads

$$H(x_t, p_t, \hat{u}_t) = \sum_{j=1}^k \beta_j(x_t) (e^{(B^* p)_j} - 1) = 0. \quad (\text{A.6.1})$$

Finally the instantaneous cost takes the form

$$\sum_{j=1}^k \left(1 - e^{(B^* p)_j} + (B^* p)_j e^{(B^* p)_j} \right) \beta_j(x_t) = \sum_{j=1}^k (B^* p)_j e^{(B^* p)_j} \beta_j(x_t), \quad (\text{A.6.2})$$

where this identity follows from (A.6.1).

A.7 Semi- and equicontinuity

Let \mathcal{X} be a metric space, equipped with a distance d , and f be a mapping from \mathcal{X} into $\mathbb{R} \cup \{-\infty, \infty\}$.

Definition A.7.1. f is said to be lower (resp. upper) semi-continuous if for any $x_0 \in \mathcal{X}$,

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0) \quad (\text{resp. } \limsup_{x \rightarrow x_0} f(x) \leq f(x_0)).$$

Clearly f is continuous if and only if it is both lower and upper semi-continuous.

A lower (resp. upper) semi-continuous $(-\infty, \infty]$ -valued (resp. $[-\infty, \infty)$ -valued) function achieves its minimum (resp. maximum) on a compact subset of \mathcal{X} .

The pointwise supremum (resp. infimum) of a collection of continuous functions is lower (resp. upper) semi-continuous.

Let now $\{f_i, i \in I\}$ be a collection of elements of $C(\mathcal{X})$ (i.e. of continuous functions from \mathcal{X} into \mathbb{R}), where I is an arbitrary index set.

Definition A.7.2. The collection $\{f_n, n \geq 1\}$ is said to be equicontinuous if for any $x_0 \in \mathcal{X}$, $\sup_{i \in I} |f_i(x) - f_i(x_0)| \rightarrow 0$, as $x \rightarrow x_0$. The same collection is said to be uniformly equicontinuous if $\sup_{i \in I} \sup_{d(x,y) \leq \delta} |f_i(x) - f_i(y)| \rightarrow 0$, as $\delta \rightarrow 0$.

Note that when \mathcal{X} is compact, equicontinuity and uniform equicontinuity are equivalent.

A.8 Solutions to selected exercises

Solution to Exercise 1.1.2. $R_0 = \lambda E(I) = \lambda/\gamma = 1.8$. The escape probability from a given under infected individual equals $E(e^{-\lambda I/N}) = \gamma/(\gamma + \lambda/N)$, since $\psi_I(-\lambda/N) = \gamma/(\gamma + \lambda/N)$ when $I \sim \text{Exp}(\gamma)$. For $\lambda = 1.8$, $\gamma = 1$, $N = 100$ we get 0.9823.

Solution to Exercise 1.1.3. For the Reed–Frost epidemic we hence have the same $R_0 = \lambda E(I) = 1.8$. As for the escape probability we get

$$P(\text{avoid infection from an infective}) = e^{-\lambda \iota/N} = 0.9822.$$

The escape probabilities are not identical, but very similar for the two models.

Solution to Exercise 1.2.1. If $I \equiv 1$, then $X \sim \text{Poi}(R_0)$. If $I \sim \text{Exp}(1/\iota)$, then $X \sim \text{MixPoi}(\lambda I)$. So

$$P(X = k) = \int_0^\infty P(X = k | I = s) e^{-s/\iota} / \iota ds = (R_0/(R_0 + 1))^k (1/(R_0 + 1)),$$

so $X \sim \text{Geo}(p = 1/(R_0 + 1))$.

Solution to Exercise 1.2.8. The probability of a minor outbreak corresponds to the probability of extinction in the approximating branching process. This probability q was derived in Section A.1 by conditioning on the number k infected in the first generation, the offspring distribution: if k get infected these all start new independent branching processes so the probability that all go extinct equals q^k . The general equation is hence

$$q = \sum_{k=0}^{\infty} q^k P(X = k).$$

The offspring distribution X depends on the infectious period distribution I . Given that $I = s$, X has a Poisson distribution with mean λs , so $X \sim \text{MixPoi}(\lambda I)$. In situation 2 (cont-time R-F) $I \equiv 1$ so $X \sim \text{Poi}(\lambda = 1.5)$. This gives the following equation

$$q = \sum_{k=0}^{\infty} q^k \frac{\lambda^k e^{-\lambda}}{k!} = \dots = e^{-R_0(1-q)} = e^{-1.5(1-q)}.$$

If this equation is solved numerically it gives the result that $q = 1 - 0.583 = 0.417$. So for the Reed–Frost case the probability of a major outbreak, equals 0.583.

As for the Markovian SIR, where $I \sim \text{Exp}(1)$ we get

$$\begin{aligned} P(X = k) &= \int_0^{\infty} P(X = k | I = s) f_I(s) ds \\ &= \int_0^{\infty} \frac{(\lambda s)^k e^{-\lambda s}}{k!} e^{-s} ds = \dots = \frac{1}{1 + \lambda} \left(\frac{\lambda}{1 + \lambda} \right)^k \end{aligned}$$

i.e. the geometric distribution, which should not come as a surprise (each time, the event is either infection or recovery, and the latter has probability $1/(\lambda + 1)$). We then get

$$q = \sum_{k=0}^{\infty} q^k P(X = k) = q^k \left(\frac{\lambda}{\lambda + 1} \right)^k \frac{1}{\lambda + 1} = \frac{1}{1 + (1 - q)\lambda}.$$

As a consequence, the probability of a minor outbreak for the Markovian SIR hence equals $q = 1/\lambda = 1/R_0 = 1/1.5 = 0.67$. The probability of a major outbreak is hence only 0.33. The randomness of the infectious period hence reduces the risk for a major outbreak. It can actually be proven that having a constant infectious period maximizes the outbreak probability among all distributions of the infectious period.

Solution to Exercise 1.2.9. The exponential growth rate (or decay rate if $R < 1$) r is the solution to Equation (1.2.1), where $h(s)$ is the average rate of infectious contacts s units after infection: $h(s) = \lambda P(L \leq s \leq L + I)$. For the Markovian SIR (for which $L \equiv 0$ and $I \sim \text{Exp}(\gamma = 1/\tau)$) we hence have $h(s) = \lambda e^{-s/\tau} = 1.5e^{-s}$, and

the solution equals $r = \lambda - 1/\iota$. For $R_0 = 1.5$ and $\gamma = \iota = 1$ this gives the exponential growth rate $r = 0.5$.

For the continuous time Reed–Frost model we have $h(s) = \lambda 1_{(s < \iota)}$. The equation then becomes $\int_0^1 e^{-rs} \lambda ds = \frac{\lambda}{r} (1 - e^{-r\iota}) = 1$. The equation is hence $r/\lambda = 1 - e^{-(r/\lambda)R_0}$. When $R_0 = 1.5$ we numerically get $r/\lambda = 0.583$, so $r = 0.874$ for the continuous time Reed–Frost model. This epidemic hence grows quicker than the Markovian SIR epidemic with the same parameters. The main reason for this is that even if the two infectious periods have equal mean $\iota = 1$, the average time of the infectious contacts are not the same. For the Reed–Frost the mean time to a randomly selected infectious contact (the mean of the generation time distribution) is of course 0.5 (the generation time distribution is uniform on $[0, 1]$, whereas for the Markovian SIR it equals 1 (the generation time distribution is $\text{Exp}(1)$)).

For the third case, with exponentially distributed latency and infectious periods, we have $h(s) = P(L < s < L + I) = \frac{\lambda\nu}{\gamma - \nu} (e^{-\nu s} - e^{-\gamma s})$. Solving $\int_0^\infty e^{-rs} h(s) ds = 1$ gives the solution

$$r = \sqrt{\nu(\lambda - \gamma) + \left(\frac{\gamma + \nu}{2}\right)^2} - \frac{\gamma + \nu}{2} \approx 0.2247.$$

Of course, adding a latency period before the infectious period will reduce the growth rate r of the epidemic.

Solution to Exercise 1.4.2. $\nu_c = 1 - 1/R_0 = 0.5$. When $\nu = 0.33$, z_ν solves the equation $1 - z_\nu = e^{-(1-\nu)R_0 z_\nu}$, and the numerical solution equals $z_\nu = 0.4544$. The over-all fraction infected is hence $(1 - \nu)z_\nu = 0.3029$. As for the probability of a major outbreak we have that for the Markovian SIR $P(\text{major outbreak}) = 1 - 1/R_\nu = 0.25$, since $R_\nu = (1 - \nu)R_0 = 0.67 \cdot 2 = 1.33$.

Solution to Exercise 1.4.3. The new rate at which an infectious individual makes infectious contacts when $\nu = 33\%$ are vaccinated is $\lambda' = \lambda p \nu + \lambda(1 - \nu)$ where $p = 0.2$ (this is true irrespective of whether the infector was vaccinated or not). Since the average infectious period equals $E(I) = 1$ we have $R_\nu = \lambda' E(I) = 1.467$ (instead of $R_0 = 2$ when no one is vaccinated).

Solution to Exercise 2.1.2. $R_0 = 1.5$: 0.583, $R_0 = 3$: 0.940, $R_0 = 15$: 1.000 (of course not exactly, but to this precision).

Solution to Exercise 3.1.1.

$$\begin{aligned} p_3^{(3)} &= p^3 + \binom{3}{2} p^2 (1 - p) * (1 - (1 - p)^2) \\ &\quad + \binom{3}{1} p (1 - p)^2 * p^2 + \binom{3}{1} p (1 - p)^2 * \binom{2}{1} p (1 - p) * p. \end{aligned}$$

Solution to Exercise 3.3.3 The limiting mean equals Nz where z solves $1 - z = e^{-R_0 z}$ so with $R_0 = \lambda t = 1.5$ we get $z = 0.583$ and the limiting mean equals 583 for both scenarios. The limiting variance of Z^N equals $N \frac{z(1-z)(1+r^2(1-z)R_0^2)}{(1-(1-z)R_0)^2}$, where r is the coefficient of variation of the infectious period. For the Reed–Frost case with non-random infectious period we have $r = 0$ implying that the limiting variance equals 1737, so the standard deviation equals 41.7, so one can expect that the final size will be somewhere in the interval 583 ± 80 with about 95% probability. The Markovian SIR has exponential infectious period which has $r = 1$ giving a variance of 3367 and standard deviation 58.0. So, the fact that the infectious period is exponential as compared to fixed makes the standard deviation of the final size increase by close to 50%.

Solution to Exercise 3.4.1. The numerical values are: the final size equals $z = 0.583$ and $R_0(1 - z) = 0.626 < 1$.

Solution to Exercise 3.4.2. Computing the two leading terms is equivalent to computing r and r^* . For the Markovian SIR we have $r = 0.5$ and $r^* = -0.3742$, for the continuous time Reed–Frost we get $r = 0.8742$ and $r^* = -0.8741$, and for the Markovian SEIR we have $r = 0.2247$ and $r^* = -0.2089$.

Solution to Exercise 4.1.3. Denoting by $U(t)$ the vector of the Gaussian fluctuations around $\begin{pmatrix} s(t) \\ i(t) \end{pmatrix}$, deduce from Theorem 2.3.2 that this vector solves the linear SDE

$$U(t) = \int_0^t A(r)U(r)dr + \int_0^t C(r)dB_r,$$

where $B(t)$ is a standard five-dimensional Brownian motion and

$$A(t) = \mu \begin{pmatrix} -1 - \frac{R_0}{\varepsilon} i(t) & -\frac{R_0}{\varepsilon} s(t) \\ \frac{R_0}{\varepsilon} i(t) & \varepsilon^{-1}(R_0 s(t) - 1) \end{pmatrix},$$

$$C(t) = \mu \begin{pmatrix} \sqrt{\mu} - \sqrt{\frac{\mu R_0}{\varepsilon} s(t) i(t)} - \sqrt{\mu s(t)} & 0 \\ 0 & \sqrt{\frac{\mu R_0}{\varepsilon} s(t) i(t)} & 0 & -\sqrt{\frac{\mu}{\varepsilon} i(t)} \end{pmatrix}.$$

Show that, as $t \rightarrow \infty$,

$$A(t) \rightarrow \mu \begin{pmatrix} -R_0 & -1/\varepsilon \\ R_0 - 1 & 0 \end{pmatrix}, \quad C(t)C^*(t) \rightarrow \frac{\mu}{R_0} \begin{pmatrix} 2R_0 & -(R_0 - 1) \\ -(R_0 - 1) & 2(R_0 - 1) \end{pmatrix}.$$

Show that the eigenvalues of $A = \lim_{t \rightarrow \infty} A(t)$ are complex, as soon as $\varepsilon < 4/R_0$, and that the real parts of those eigenvalues are negative. Conclude from a combination of Exercise 2.3.8 and Lemma 2.3.7 that the covariance matrix of the stationary distribution of $U(t)$ reads

$$\begin{pmatrix} \frac{1}{R_0} + \frac{1}{\varepsilon R_0^2} & -\frac{1}{R_0} \\ -\frac{1}{R_0} & \frac{1}{R_0} - \frac{1}{R_0^2} + \varepsilon \end{pmatrix}.$$

Conclude by taking into account that we expect to have $\varepsilon \ll R_0^{-1}$.

Solution to Exercise 4.1.4. The relative length of the infectious period ε affects the critical community size N_c much more than R_0 does, since it is squared in the approximation of N_c . As an illustration, if the infectious period is doubled (with half infectivity per unit of time thus keeping R_0 fixed) N_c will decrease by a factor 4, whereas if the basic reproduction number is doubled (keeping everything else fixed) only decreases N_c by a factor close to 2.

Solution to Exercise 4.1.5. There are two effects of this vaccination strategy. The first is that vaccinated individuals can be ignored, so the relevant population (of unvaccinated people) is now $N^{(unvacc)} = N(1 - v)$. Secondly, since infected individuals have contact with both types of individuals, the rate of having contact with the population of interest is reduced to $\lambda(1 - v)$ implying that the reproduction number is changed to $R_v = R_0(1 - v)$. The critical population size of unvaccinated people $N_c^{(unvacc)}$ is then simply obtained in the same way, but for these new parameters, so

$$N_c^{(unvacc)} = \frac{9}{\varepsilon^2(1 - \frac{1}{R_v})^2 R_v} = \frac{9}{\varepsilon^2(1 - \frac{1}{(1-v)R_0})^2 (1-v)R_0}.$$

However, a more interesting quantity is the critical community size counting all individuals, hence also vaccinated. Since $N = N^{(unvacc)} / (1 - v)$, the critical community size for a population in which a fraction v of the new-born are continuously being vaccinated is given by

$$N_c^{(v)} = \frac{9}{(1 - v)^2 \varepsilon^2 (1 - \frac{1}{(1-v)R_0})^2 R_0}.$$

By numerical studies it is easily shown that the critical community size grows very big with v , also agreeing with empirical evidence since e.g. measles is no longer endemic in England (or anywhere else in the world having high vaccination coverage).

References for Part I

1. R.M. Anderson and R.M. May, *Infectious diseases of humans; dynamic and control*, Oxford: Oxford University Press, 1991.
2. H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*, Springer Lecture Notes in Statistics. New York: Springer Verlag, 2000.
3. K.B. Athreya and P.E. Ney, *Branching Processes*, Grundlehren der Mathematischen Wissenschaften **196**, Springer, 1972. Reprinted in 2004 by Dover.
4. R. Azencott, Grandes déviations et applications. (French) *Eighth Saint Flour Probability Summer School - 1978*, pp. 1-176, Lecture Notes in Math. **774**, Springer, Berlin, 1980.
5. F.G. Ball, A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models, *Adv. Appl. Prob.* **18**, 289-310, 1986.
6. F.G. Ball and D. Clancy, The final size and severity of a generalised stochastic multitype epidemic model, *Adv. Appl. Prob.* **25**, 721-736, 1993.
7. A.D. Barbour, The duration of the closed stochastic epidemic, *Biometrika* **62**, 477-482, 1975.
8. P. Billingsley, *Convergence of Probability Measures*, 2d ed., J. Wiley & Sons, Inc., 1999.
9. A. Dembo and O. Zeitouni, *Large deviations, techniques and applications*, 2nd ed., Applications of Mathematics **38**, Springer, 1998.
10. R.A. Doney, A limit theorem for a class of supercritical branching processes, *J. Appl. Prob.* **9**, 707-724, 1972.
11. S.N. Ethier and T.G. Kurtz, *Markov Processes, Characterization and Convergence*, J. Wiley & Sons, Inc., 1986.
12. C. Fraser, C.A. Donnelly, S. Cauchemez, W.P. Hanage, M.D. Van Kerkhove, T.D. Hollingsworth, J. Griffin, R.F. Baggaley, H.E. Jenkins, E.J. Lyons, et al., Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings, *Science*, **324**: 1557-1561, 2009.
13. M. Freidlin and A. Wentzell, *Random perturbations of dynamical systems*, 3d ed., Grundlehren des Mathematischen Wissenschaften **260**, Springer, 2012.
14. J. Giesecke, *Modern infectious disease epidemiology*, 3rd Ed, London: CRC Press, 2017.
15. J. Jacod and A. Shiryaev, *Limit Theorems for Stochastic Processes*, Grundlehren des Mathematischen Wissenschaften **288**, Springer, 1987.
16. P. Jagers, *Branching Processes with Biological Applications*, J. Wiley & Sons, Ltd., 1975.
17. I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, 2d ed. 1991, Springer Verlag, 1988.
18. M.J. Keeling and B.T. Grenfell, Disease extinction and community size: modeling the persistence of measles, *Science* **275**(5295): 65-67, 1997.
19. B. Kouegou-Kamen and E. Pardoux, Large deviations of the small jumps asymptotic of the moving-optimum model, to appear.
20. B. Kouegou-Kamen and E. Pardoux, Large deviations for jump Markov processes : the lower bound, to appear.

21. P. Kratz and E. Pardoux, Large deviations for infectious diseases, Chapter 7 of *Séminaire de Probabilités XLIX*, Lecture Notes in Math. **2215**, pp. 221–327, 2018.
22. T.G. Kurtz, Strong approximation theorems for density dependent Markov chains, *Stochastic Processes and their Applications*, **6**(3): 223–240, 1978.
23. M. Lindholm and T. Britton, Endemic persistence or disease extinction: the effect of population separation into subcommunities, *Theor. pop. biol.*, **72**: 253–263, 2007.
24. I. Nåsell, On the time to extinction in recurrent epidemics, *J.R. Statist. Soc. B* **61**(2): 309–330, 1999.
25. E. Pardoux, *Markov processes and Applications*, J. Wiley, 2008.
26. E. Pardoux and B. Samegni-Kepgnou, Large deviation principle for Epidemic models, *J. Applied Probab.* **54** 905–920, 2017.
27. E. Pardoux and B. Samegni-Kepgnou, Large deviation of the exit measure through a characteristic boundary for a Poisson driven SDE, <http://arxiv.org/abs/1808.04991>, submitted 2018.
28. J. Pitman, *Combinatorial Stochastic Processes*, Ecole d'Eté de Probabilité de Saint-Flour XXXII - 2002, Lecture Notes in Mathematics **1875**, Springer Verlag, 2006.
29. L.E. Quenee, N.A. Ciletti, D. Elli, T.M. Hermanas and O. Schneewind, Prevention of pneumonic plague in mice, rats, guinea pigs and non-human primates with clinical grade rV10, rV10-2 or F1-V vaccines, *Vaccine* **29**(38): 6572–6583, 2011.
30. G. Polya and G. Szegő, *Problems and Theorems in Analysis I*, Springer Verlag, Berlin, Heidelberg, 1978.
31. L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mishchenko, *The mathematical theory of optimal processes*, Transl. by K.N. Trirogoff; ed. by L.W. Neustadt, John Wiley & Sons, 1962.
32. G. Scalia-Tomba, Asymptotic final size distribution for some chain-binomial processes, *Adv. Appl. Prob.* **17** 477–495, 1985.
33. G. Scalia-Tomba, On the asymptotic final size distribution of epidemics in heterogeneous populations, in *Stochastic processes in epidemic theory*, J.P. Gabriel, C. Lefèvre and P. Picard, *Lecture notes in Biomathematics* **86**, 189–196, 1990.
34. A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis*, Chapman Hall, London, 1995.
35. A. Shwartz and A. Weiss, Large deviations with diminishing rates, *Mathematics of Operations Research* **30** 281–310, 2005.
36. T. Sellke, On the asymptotic distribution of the size of a stochastic epidemic, *J. Appl. Prob.* **20**: 390–394, 1983.
37. E. Trélat, *Contrôle optimal, Théorie et Applications*, 2d ed. Vuibert, 2008.
38. WHO Ebola Response Team, Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections, *New Engl. J. Med.* **371**: 1481–1495, 2014.