Deviations from the Law of Large Numbers and Extinction of an Endemic Disease

Étienne Pardoux

Abstract

Consider an endemic disease, which corresponds to an epidemic model with a constant flux of susceptibles, in a situation where the corresponding deterministic epidemic model has a unique stable endemic equilibrium. If we consider the associated stochastic model, whose law of large numbers limit is the deterministic model, the disease free equilibrium is an absorbing state, which is reached soon or later by the process. However, for a large population size, i.e. when the stochastic model is close to its deterministic limit, the time needed for the stochastic perturbations to stop the epidemic may be enormous. In this presentation, we discuss how the Central Limit Theorem, Moderate and Large Deviations allow us to try to estimate the extinction time of the epidemic.

1 Introduction

We consider epidemic models where there is a constant flux of susceptibles, either because the infected individuals become susceptible immediately after healing, or after some time during which the individual is immune to the illness, or because there is a constant flux of newborn or immigrant susceptibles.

In the above three cases, for certain values of the parameters, there is an endemic equilibrium, which is a stable equilibrium of the associated deterministic epidemic model. The deterministic model can be considered as the Law of Large Numbers limit (as the size of the population tends to ∞) of a stochastic model, where infections, healings, births and deaths happen according to Poisson processes whose rates depend upon the numbers of individuals in each compartment.

Since the disease free states are absorbing, it follows from an irreducibility property which is clearly valid in our models, that the epidemic will stop soon or later in the more realistic stochastic model. However, the time which the stochastic perturbances will need to stop the epidemic may be enormous when the size N of the population is large. The aim of this paper is to describe, based upon the Central Limit Theorem, Large and Moderate Deviations, the time it takes for the epidemic to stop in the stochastic model.

The paper is organized as follows. In section 2, we describe the three deterministic and stochastic models which we have in mind, namely the SIS, SIRS and SIR model with demography. In section 3, we give the general formulation of the stochastic models, and recall the Law of Large Numbers, the Central Limit Theorem and the Large Deviations, and their application to the time of extinction of an epidemic. Finally in section 4, we present the moderate deviations result for the SIS model (which is the simplest of our three models), and explain how it can be used to predict the time taken for an epidemic to cease. Those results will be proved in more generality, with full details of the proofs in [13].

The results concerning the Law of Large Numbers and the Large Deviations can be found in Kratz and Pardoux [12], Pardoux and Samegni–Kepgnou [14], and Britton and Pardoux [5], where the Central Limit Theorem is also established. Note that the three above references present different approaches to the Large Deviations results. The moderate deviations results will appear in [13].

We conclude this introduction with a short history and a few references to books and lecture notes which describe models of infectious diseases and epidemics. Mathematical modelling of infectious diseases has a long history of being useful. The first such mathematical model was probably the one proposed by Bernoulli in [4] with a model of smallpox. A little more than one hundred years ago, Sir Ronald Ross, a british medical doctor and Nobel laureate who contributed to the understanding of malaria wrote : *"As a matter of fact all epidemiology, concerned as it is with variation of disease from time to time and from place to place, must be considered mathematically (...) and the mathematical method of treatment is really nothing but the application of careful reasoning to the problems at hand"*. As a matter of fact, Ross deduced from mathematical arguments conclusions concerning malaria, which his physician colleagues found hard to accept. One of the first books devoted to mathematical modelling of infectious diseases is [3]. A book which has had huge impact is [1], which exclusively deals with deterministic models. Since then, there has been steady production of new research monographs, e.g. [2] also looking at inference methodology, [7] focusing mainly on stochastic models, [11] dealing also with animal populations, and [9] covering both deterministic and stochastic modelling. Finally [6] will soon present the broadest treatment of stochastic epidemic models ever published in one volume, covering both classical and new results and methods, from mathematical models to statistical procedures.

2 The three models

2.1 The SIS model

The deterministic SIS model is the following. Let $s(t)$ (resp. $i(t)$) denote the proportion of susceptible (resp. infectious) individuals in the population. Given an infection parameter λ, and a recovery parameter γ, the deterministic SIS model reads

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

Since clearly $s(t) + i(t) \equiv 1$, the system can be reduced to a one dimensional ordinary differential equation. If we let $z(t) = i(t)$, we have $s(t) = 1 - z(t)$,and we obtain the ordinary differential equation

$$z'(t) = \lambda z(t)(1 - z(t)) - \gamma z(t)\,.$$

It is easy to verify that this ordinary differential equation has a so–called "disease free equilibrium", which is $z(t) = 0$. If $\lambda > \gamma$, this equilibrium is unstable, and there is an endemic stable equilibrium $z(t) = 1 - \gamma/\lambda$.

The corresponding stochastic model is as follows. Let S_t^N (resp. I_t^N) denote the proportion of susceptible (resp of infectious) individuals in a population of total size N.

$$\begin{cases} S_t^N = S_0^N - \dfrac{1}{N} P_{inf}\left(\lambda N \displaystyle\int_0^t S_r^N I_r^N dr\right) + \dfrac{1}{N} P_{rec}\left(\gamma N \displaystyle\int_0^t I_r^N dr\right), \\ I_t^N = I_0^N + \dfrac{1}{N} P_{inf}\left(\lambda N \displaystyle\int_0^t S_r^N I_r^N dr\right) - \dfrac{1}{N} P_{rec}\left(\gamma N \displaystyle\int_0^t I_r^N dr\right). \end{cases}$$

Here $P_{inf}(t)$ and $P_{rec}(t)$ are two mutually independent standard (i.e. rate 1) Poisson processes. Let us give some explanations, first concerning the modeling, then concerning the mathematical formulation.

Let $\mathcal{S}_t^N$ (resp. $\mathcal{I}_t^N$) denote the number of susceptible (resp. infectious) individuals in the population. The equations for those quantities are the above equations, multiplied by N. The argument of $P_{inf}(t)$ reads

$$\lambda \int_0^t \frac{\mathcal{S}_r^N}{N} \mathcal{I}_r^N dr\,.$$

The argument for such a rate of infections in the total population can be explained as follows. Each infectious individual meets other individuals in the population at some rate β. The encounter results in a new infection with probability p if the partner of the encounter is susceptible, which happens with probability S_t^N/N, since we assume that each individual in the population has the same probability of being that partner, and with probability 0 if the partner is an infectious individual. Letting $\lambda = \beta p$ and summing over the infectious at time t gives the above rate. Concerning recovery, it is assumed that each infectious recovers at rate γ, independently of the others.

Remark 1. *Let us comment about the fact that we write our stochastic models in terms of Poisson processes. The fact that the infection events happen according to a Poisson process is a rather natural assumption. However, concerning the recovery from infection, our model assumes that the duration of the infectious period follows an exponential distribution. This is not realistic. We are forced to make such an assumption if we want to have a Markov model. We must confess that this assumption is done for mathematical convenience.*

Note that there is an equivalent, but slightly more complicated way of writing the Poisson terms, which we now present. Let $\mathcal{M}_{inf}$ and $\mathcal{M}_{rec}$ denote two mutually independent Poisson random measures on $(0,+\infty)^2$, with mean measure the Lebesgue measure.

$$P_{inf}\left(\lambda N \int_0^t S_r^N I_r^N dr\right) \text{ can be rewritten as } \int_0^t \int_0^\infty \mathbf{1}_{u \le \lambda N S_r^N I_r^N dr} \mathcal{M}_{inf}(dr,du)$$

and

$$P_{rec}\left(\gamma N \int_0^t I_r^N dr\right) \text{ can be rewritten as } \int_0^t \int_0^\infty \mathbf{1}_{u \le \gamma N I_r^N dr} \mathcal{M}_{rec}(dr,du)\,.$$

Again we have $S_t^N + I_t^N = 1$, and $Z_t^N = I_t^N$ satisfies

$$Z_t^N = Z_0^N + \frac{1}{N} P_{inf}\left(\lambda N \int_0^t (1 - Z_r^N) Z_r^N dr\right) - \frac{1}{N} P_{rec}\left(\gamma N \int_0^t Z_r^N dr\right).$$

2.2 The SIRS model

In the SIRS model, contrary to the SIS model, an infectious who heals is first immune to the illness, he is "recovered", and only after some time does he loose his immunity and turn to susceptible. The deterministic SIRS model reads

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

while the stochastic SIRS model reads

$$\begin{cases} S_t^N = S_0^N - \frac{1}{N} P_{inf}\left(\lambda N \int_0^t S_r^N I_r^N dr\right) + \frac{1}{N} P_{loim}\left(\rho N \int_0^t R_r^N dr\right), \\ I_t^N = I_0^N + \frac{1}{N} P_{inf}\left(\lambda N \int_0^t S_r^N I_r^N dr\right) - \frac{1}{N} P_{rec}\left(\gamma N \int_0^t I_r^N dr\right) \\ R_t^N = R_0^N + \frac{1}{N} P_{rec}\left(\gamma N \int_0^t I_r^N dr\right) - \frac{1}{N} P_{loim}\left(\rho N \int_0^t R_r^N dr\right). \end{cases}$$

These two models could be reduced to two–dimensional models for $z(t) = (i(t), s(t))$ (resp. for $Z_t^N = (I_t^N, S_t^N)$).

2.3 The SIR model with demography

In this model, recovered individuals remain immune for ever, but there is a flux of susceptibles by births at rate μN, while individuals from each of the three compartments die at rate μ. Thus the deterministic model

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

whose stochastic variant reads

$$\begin{cases} S_t^N = S_0^N - \frac{1}{N} P_{inf}\left(\lambda N \int_0^t S_r^N I_r^N dr\right) + \frac{1}{N} P_{birth}(\rho N t) - \frac{1}{N} P_{ds}\left(\mu N \int_0^t S_r^N dr\right), \\ I_t^N = I_0^N + \frac{1}{N} P_{inf}\left(\lambda N \int_0^t S_r^N I_r^N dr\right) - \frac{1}{N} P_{rec}\left(\gamma N \int_0^t I_r^N dr\right) \\ \qquad - \frac{1}{N} P_{di}\left(\mu N \int_0^t I_r^N dr\right), \\ R_t^N = R_0^N + \frac{1}{N} P_{rec}\left(\gamma N \int_0^t I_r^N dr\right) - \frac{1}{N} P_{dr}\left(\mu N \int_0^t R_r^N dr\right). \end{cases}$$

Remark 2. *One may think that it would be more natural to decide that births happen at rate μ times the total population. Then the total population process would be a critical branching process, which would go extinct in finite time a.s., which we do not want. Next it might seem more natural to replace in the infection rate the ratio $\mathcal{S}_t^N/N$ by $S_t^N/(S_t^N + I_t^N + R_t^N)$, which is the actual ratio of susceptibles in the population at time t. It is easy to show that $S_t^N + I_t^N + R_t^N$ is close to 1, so we choose the simplest formulation.*

Again, we can reduce these models to two–dimensional models for $z(t) = (i(t), s(t))$ (resp. for $Z_t^N = (I_t^N, S_t^N)$), by deleting the r (resp. R^N) component.

3 The stochastic model, LLN, CLT and LD

3.1 The stochastic model

The three above stochastic models are of the following form.

$$\begin{aligned} Z_t^N &= z_N + \frac{1}{N} \sum_{j=1}^k h_j P_j\left(N \int_0^t \beta_j(Z_s^N) ds\right) \\ &= z_N + \int_0^t b(Z_s^N) ds + \frac{1}{N} \sum_{j=1}^k h_j M_j\left(N \int_0^t \beta_j(Z_s^N) ds\right), \end{aligned} \tag{1}$$

where $\{P_j(t),\, t \geq 0\}_{0 \leq j \leq k}$ are mutually independent standard Poisson processes, $M_j(t) = P_j(t) - t$, and $b(z) = \sum_{j=1}^k \beta_j(z) h_j$. Z_t^N takes its values in $\mathbb{R}^d$.

In the case of the SIS model, $d = 1$, $k = 2$, $h_1 = 1$, $\beta_1(z) = \lambda z(1-z)$, $h_2 = -1$ and $\beta_2(z) = \gamma z$.

In the case of the SIRS model, $d = 2$, $k = 3$, $h_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\beta_1(z) = \lambda z_1 z_2$, $h_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\beta_2(z) = \gamma z_1$ and $h_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\beta_3(z) = \rho(1 - z_1 - z_2)$.

In the case of the SIR model with demography, we can restrict ourselves to $d = 2$, while $k = 4$, $h_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\beta_1(z) = \lambda z_1 z_2$, $h_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\beta_2(z) = (\gamma + \mu) z_1$, $h_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\beta_3(z) = \mu$, $h_4 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, $\beta_4(z) = \mu z_2$.

While the above formulation has the advantage of being concise, it is for certain purposes more convenient to rewrite (1) using the equivalent formulation already described in case of the SIS model. Let $\{\mathcal{M}_j,\ 1 \le j \le k\}$ be mutually independent Poisson random measures on $\mathbb{R}_+^2$ with mean measure the Lebesgue measure, and let $\overline{\mathcal{M}}_j(ds, du) = \mathcal{M}_j(ds, du) - ds\, du$, $1 \le j \le k$. We can rewrite (1) in the form

$$
\begin{aligned}
Z_t^N &= z_N + \frac{1}{N} \sum_{j=1}^{k} h_j \int_0^t \int_0^{N\beta_j(Z_s^N)} \mathcal{M}_j(ds, du) \\
&= z_N + \int_0^t b(Z_s^N) ds + \frac{1}{N} \sum_{j=1}^{k} h_j \int_0^t \int_0^{N\beta_j(Z_s^N)} \overline{\mathcal{M}}_j(ds, du),
\end{aligned}
\tag{2}
$$

in the sense that the joint law of $\{Z^N,\ N \ge 1\}$ is the same law of a sequence of random elements of the Skorohod space $D([0,T]; \mathbb{R}^d)$, whether we use (1) or (2) for its definition.

We will now state a few results, without specifying particular assumptions. Those results are valid at least in the case of the three above examples. See [5] for details of the proofs, and precise assumptions under which those results hold true.

Concerning the initial condition, we assume that for some $z \in [0,1]^d$, $z_N = [Nz]/N$, where $[Nz] \in \mathbb{Z}_+^d$ is the vector whose i–th component is the integer part of the real number Nz^i.

3.2 Law of Large Numbers

We have a Law of Large Numbers

Theorem 3. *Let Z_t^N denote the solution of the stochastic differential equation* (1)*. Assume that the β_j are locally bounded, b is locally Lipschitz, and the unique solution of equation* (3) *does not explode in finite time. Then $Z_t^N \to z_t$ a.s. locally uniformly in t, where $\{z_t,\, t \geq 0\}$ is the unique solution of the ordinary differential equation*

$$\frac{dz_t}{dt} = b(z_t), \quad z_0 = x. \tag{3}$$

The main argument in the proof of the above theorem is the fact that, locally uniformly in t,

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

3.3 Central Limit Theorem

We also have a Central Limit Theorem. Let $U_t^N := \sqrt{N}(Z_t^N - z(t))$.

Theorem 4. *Assume in addition to the hypotheses of Theorem 3 that b is of class C^1. Then, as $N \to \infty$, $\{U_t^N,\, t \geq 0\} \Rightarrow \{U_t,\, t \geq 0\}$ for the topology of locally uniform convergence, where $\{U_t,\, t \geq 0\}$ is a Gaussian process of the form*

$$U_t = \int_0^t \nabla_x b(z_s) U_s ds + \sum_{j=1}^k h_j \int_0^t \sqrt{\beta_j(z_s)} dB_j(s),\ \ t \geq 0\,, \tag{4}$$

where $\{(B_1(t), B_2(t), \ldots, B_k(t)),\, t \geq 0\}$ are mutually independent standard Brownian motions.

3.4 Large Deviations, and extinction of an epidemic

We denote by $\mathcal{AC}_{T,d}$ the set of absolutely continuous functions from $[0, T]$ into $\mathbb{R}^d$. 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 $g(\nu, \omega) = \nu \log(\nu/\omega) - \nu + \omega$. We assume in the definition of $g(\nu, \omega)$ that for all $\nu > 0$, $\log(\nu/0) = \infty$ and $0 \log(0/0) = 0 \log(0) = 0$. It is not hard to verify that $I_T(\phi) = 0$ if and only if ϕ solves the ordinary differential equation (3). $I_T(\phi)$ can be interpreted as an energy needed for letting ϕ deviate from being a solution of (3).

The collection Z^N obeys a Large Deviations Principle, in the sense that

Theorem 5. *For any open subset $O \subset D([0, T]; \mathbb{R}^d)$,*

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(Z^{N,z_N} \in O\right) \geq -I_{T,z}(O).$$

For any closed subset $F \subset D([0, T]; \mathbb{R}^d)$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Z^{N,z_N} \in F) \leq -I_{T,z}(F)\,,$$

where for any $z \in \mathbb{R}^d$, $A \subset D([0, T]; \mathbb{R}^d)$,

$$I_{T,z}(AZ) := \inf_{\phi\in A, \phi(0)=z} I_T(\phi)\,.$$

A slight reinforcement of this theorem allows us to conclude a Wentzell–Freidlin type of result. Wentzell and Freidlin have studied small random perturbations of an ordinary differential equation like (3), see [10]. One of their main results is to compute asymptotically the time needed for a small random perturbation of such an equation to drive the solution outside of the basin of attraction of a stable equilibrium. The theory has been originally developed for Brownian perturbations. Here we give a statement of the same type, for a Poissonian perturbation. In what follows, we assume that the first component of Z_t^N (resp. of $z(t)$) is I_t^N (resp. $i(t)$). Assume that the deterministic ordinary differential equation (3) has a unique stable equilibrium z^* whose first component satisfies $z_1^* > 0$. We define

$$\overline{V} := \inf_{T>0} \inf_{\phi \in \mathcal{AC}_{T,d}, \phi(0)=z^*, \phi_1(T)=0} I_T(\phi).$$

Let now

$$T_{\text{Ext}}^{N,z} = \inf\{t > 0,\ Z_1^N(t) = 0,\ \text{if } Z^N(0) = z_N\}.$$

We have the

Theorem 6. *Given any $\eta > 0$, for any z with $z_1 > 0$,*

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

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

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

It is important to evaluate the quantity $\overline{V}$. Note that it is the value function of an optimal control problem. In case of the SIS model, which is one dimensional, one can solve this control problem explicitly with the help of Pontryagin's maximum principle[1], see [15] or for a concise introduction adapted to this application section A.6 in [5], and deduce in that case that $\overline{V} = \log \frac{\lambda}{\gamma} - 1 + \frac{\gamma}{\lambda}$. For other models, one can compute numerically the value of $\overline{V}$ for each given value of the parameters.

4 Moderate deviations

4.1 CLT and extinction of an endemic disease

Consider the SIR with demography.

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

[1]Pontryagin's maximum principle states sufficient conditions for a control to be optimal. In the case of the SIS model, the corresponding control problem is one dimensional, and Pontryagin's conditions allow to compute explicitly the optimal trajectory.

We assume that $\lambda > \gamma + \mu$, in which case there is a unique stable endemic equilibrium, namely $z^* = (i^*, s^*) = (\frac{\mu}{\gamma+\mu} - \frac{\mu}{\lambda}, \frac{\gamma+\mu}{\lambda})$. Following Section 4.1 in [5], we can study the extinction of an epidemic in the above model using the CLT. We note that the basic reproduction number R_0 (the expected number of infectious contacts by one infectious at the start of the epidemic, i.e. when $s(t) \simeq 1$) and the expected relative time of a life an individual is infected, ε, are given by

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

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 $\varepsilon \approx \mu/\gamma$. Denote again by I_t^N the fraction of the population which is infectious in a population of size N. The law of large numbers tells us that for N and t large, I_t^N is close to i^*. The CLT tell us that $\sqrt{N}(I_t^N - i^*)$ converges to a Gaussian process, whose asymptotic variance can be shown to well approximated by $R_0^{-1} - R_0^{-2} \sim R_0^{-1}$. This suggests that for large t, the number of infectious in the population is approximately Gaussian, with mean Ni^* and standard deviation $\sqrt{N/R_0}$. Since we expect a Gaussian process with marginal $N(0,1)$ to hit -2 fairly quickly, we expect that if $Ni^* - 2\sqrt{N/R_0} \leq 0$, then the epidemic will stop rather quickly, while if $Ni^* - 4\sqrt{N/R_0} \geq 0$, it is not clear that the time of extinction will be of order 1 (as a function of N). This gives a critical population size roughly of the order of

$$N_c = \frac{9}{(i^*)^2 R_0} = \frac{9}{\epsilon^2 (1 - R_0^{-1})^2 R_0}.$$

Note that the factor 9 is rather arbitrary. This N_c is rather large since i^* is relatively small. Clearly, even if everybody in the population gets ill at some point, being ill one week in a life of average length 75 years gives a small fraction of infectious in the population.

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 at most a couple of million, we expect that the disease will go extinct quickly, whereas the disease will become endemic (for a rather long time) in a population being larger than e.g. 20 million people. This confirms the empirical observation that measles was continously endemic in UK whereas it died out quickly in Iceland (and was later reintroduced by infectious people visiting the country), see [1].

4.2 Moderate Deviations

If the CLT allows to predict extinction of an endemic disease for population sizes under a given threshold N_c, and Large Deviations gives predictions for arbitrarily large population sizes, it is fair to look at Moderate Deviations, which describes ranges of fluctuations between those of the CLT and those of the LD. We shall present the Moderate Deviations approach in the specific case of the SIS model. In other words, our model from now on is 1–dimensional, and it reads

$$\begin{aligned}
Z_t^N &= Z_0^N + \int_0^t b(Z_s^N)ds + Y_t^N, \quad \text{where} \\
b(z) &= [\lambda(1-z)-\gamma]z, \quad \text{and} \\
Y_t^N &= \frac{1}{N}\int_0^t \int_0^{\lambda N Z_s^N(1-Z_s^N)} \overline{\mathcal{M}}_1(ds,du) - \frac{1}{N}\int_0^t \int_0^{\gamma N Z_s^N} \overline{\mathcal{M}}_2(ds,du).
\end{aligned}$$

We consider the case $\lambda > \gamma$ and recall that the unique stable equilibrium of the deterministic model then is $z^* = 1 - \frac{\gamma}{\lambda}$. We assume that $Z_0^N = z_N^* := [Nz^*]/N$. We have

$$Z_t^N - z^* = z_N^* - z^* - \lambda \int_0^t Z_s^N(Z_s^N - z^*)ds + Y_t^N.$$

It follows that

$$Z_t^N - z^* = (z_N^* - z^*)e^{-\lambda\int_0^t Z_s^N ds} + Y_t^N - \lambda\int_0^t Z_s^N e^{-\lambda\int_s^t Z_r^N dr} Y_s^N ds.$$

Consequently

$$|Z_t^N - z^*| \le |z_N^* - z^*| + 2\sup_{0\le s\le t}|Y_s^N|. \tag{2}$$

We can also rewrite the above stochastic differential equation in the form

$$\begin{aligned}
Z_t^N - z^* &= z_N^* - z^* - (\lambda-\gamma)\int_0^t (Z_s^N - z^*)ds + \widetilde{Y}_t^N, \quad \text{where} \\
\widetilde{Y}_t^N &= Y_t^N - \lambda\int_0^t (Z_s^N - z^*)^2 ds.
\end{aligned} \tag{3}$$

A combination of (2) and (3) yields the existence of a constant C such that

$$\int_0^T (Z_t^N - z^*)^2 dt \leq |z_N^* - z^*| + C\left(\sup_{0\leq t\leq T}|Y_t^N|\right) \wedge \left(\sup_{0\leq t\leq T}|Y_t^N|^2\right). \tag{4}$$

For the bound by $\sup_{0\leq t\leq T}|Y_t^N|^2$, we first take the suare in (2).

We now define, for $0 < \alpha < 1/2$,

$$\widetilde{Y}_t^{N,\alpha} = N^\alpha[Y_t^N - \lambda\int_0^t (Z_s^N - z^*)^2 ds].$$

and deduce from (3)

$$N^\alpha(Z_t^N - z^*) = N^\alpha(z_N^* - z^*) - (\lambda-\gamma)\int_0^t N^\alpha(Z_s^N - z^*)ds + \widetilde{Y}_t^{N,\alpha}. \tag{5}$$

It follows from (5) that the map $\widetilde{Y}^{N,\alpha} \mapsto N^\alpha(Z^N - z^*)$ is continuous from $D([0,T])$ into itself. Here we equip $D([0,T])$ with the sup norm topology, which makes it a Hausdorff topologic vector space (equipped with the Skorohod topology, $D([0,T])$ is not a topologic vector space).

We are interested in the Large Deviations of $N^\alpha(Z^N - z^*)$, which means Moderate Deviations of $Z^N - z^*$. Note that the deviations of $N^\alpha(Z^N - z^*)$ in case $\alpha = 1/2$ are analyzed by the Central Limit Theorem, and in case $\alpha = 0$ by the Large Deviations. So with $0 < \alpha < 1/2$, we are clearly here in a regime which is intermediate between the CLT and LD, which is called the regime of Moderate Deviations.

We first note that the LD of $N^\alpha(Z^N - z^*)$ will be deduced from those of $\widetilde{Y}^{N,\alpha}$ thanks to the contraction principle, see e.g. Theorem 4.2.1 in [8]. So we essentially have to analyze the LD of $\widetilde{Y}^{N,\alpha}$. In fact, we will proceed in three steps. In the first step, we shall analyse the Large Deviations of

$$\overline{Y}_t^{N,\alpha} := N^{\alpha-1}\int_0^t\int_0^{\lambda N z^*(1-z^*)} \overline{\mathcal{M}}_1(ds,du) - N^{\alpha-1}\int_0^t\int_0^{\gamma N z^*} \overline{\mathcal{M}}_2(ds,du),$$

at the speed $N^{2\alpha-1}$, or in other words the Moderate Deviations of

$$\overline{Y}_t^{N} := \frac{1}{N}\int_0^t\int_0^{\lambda N z^*(1-z^*)} \overline{\mathcal{M}}_1(ds,du) - \frac{1}{N}\int_0^t\int_0^{\gamma N z^*} \overline{\mathcal{M}}_2(ds,du).$$

The second step will consist in showing that $\widetilde{Y}^{N,\alpha}$ and $\overline{Y}^{N,\alpha}$ have the same behavior as regards Large Deviations. Finally the third step will consist in applying the contraction principle, in order to deduce the LD of $N^\alpha(Z^N - z^*)$.

4.2.1 Step1 : Moderate Deviations of $\overline{Y}^N$

We shall use the notations $a_N = N^{2\alpha-1}$ and $\overline{Y}^{N,\alpha} := N^\alpha \overline{Y}^N$. Given a signed measure ν on $[0,T]$, we write

$$\Lambda_N(\nu) = \log \mathbb{E}\left[e^{\nu(\overline{Y}^{N,\alpha})}\right], \quad \text{where } \nu(\overline{Y}^{N,\alpha}) = \int_{[0,T]} \overline{Y}_t^{N,\alpha} \nu(dt)$$

for the logarithmic moment generating function of $\overline{Y}^{N,\alpha}$ at ν.

The crucial step of our derivation is the

Proposition 7. *For any signed measure ν on $[0,T]$, as $N \to \infty$,*

$$a_N \Lambda_N(a_N^{-1}\nu) \to \Lambda(\nu) := \frac{1}{2}\mathbb{E}\left[\nu(Y)^2\right],$$

where

$$Y_t := \int_0^t \int_0^{\lambda z^*(1-z^*)} \overline{\mathcal{M}}_1(ds, du) - \int_0^t \int_0^{\gamma z^*} \overline{\mathcal{M}}_2(ds, du).$$

Proof We first rewrite $a_N^{-1}\overline{Y}^{N,\alpha}$ in the form

$$a_N^{-1}\overline{Y}_t^{N,\alpha} = N^{-\alpha} \sum_{j=0}^{1} \sum_{\ell=1}^{N} (-1)^j M_{j,\ell}\left(\beta_j(z^*)t\right),$$

here $\beta_0(z) = \lambda z(1-z)$ and $\beta_1(z) = \gamma z$ and the processes $\{M_{j,\ell},\ j=0,1;\ 1 \le \ell \le N\}$ are i.i.d. compensated standard Poisson processes. We now have

$$\begin{aligned}
a_N \overline{\Lambda}_N(a_N^{-1}\nu) &= a_N \log \mathbb{E} \exp\left[\nu(a_N^{-1}\overline{Y}^{N,\alpha})\right] \\
&= N a_N \log \mathbb{E} \exp\left[N^{-\alpha}\nu\left(\sum_{j=0}^{1}(-1)^j M_{j,1}\left(\beta_j(z^*)\cdot\right)\right)\right] \\
&= N^{2\alpha} \log \mathbb{E}\Bigg[1 + N^{-\alpha}\nu\left(\sum_{j=0}^{1}(-1)^j M_{j,1}\left(\beta_j(z^*)\cdot\right)\right) \\
&\qquad + \frac{N^{-2\alpha}}{2}\nu\left(\sum_{j=0}^{1}(-1)^j M_{j,1}\left(\beta_j(z^*)\cdot\right)\right)^2 + O\left(N^{-3\alpha}\right)\Bigg] \\
&\to \frac{1}{2}\mathbb{E}\left[\nu\left(\sum_{j=0}^{1}(-1)^j M_{j,1}\left(\beta_j(z^*)\cdot\right)\right)^2\right] = \frac{\sigma^2}{2}\int_{[0,T]^2} t \wedge s\, \nu(dt)\nu(ds)
\end{aligned}$$

as $N \to \infty$, where we have used the notation $\sigma^2 := 2\frac{\gamma}{\lambda}(\lambda - \gamma)$. $\square$

The next step consists in establishing exponential tightness of the laws of $\overline{Y}^{N,\alpha}$, in the sense that

Proposition 8. *For any $R > 0$, there exists a compact set $K_R \subset\subset D([0,T]; \mathbb{R}^d)$ such that*

$$\limsup_N a_N \log \mathbb{P}(Y^{N,\alpha} \in (K_R)^c) \leq -R\,.$$

The proof of this Proposition follows essentially the lines of the proof of exponential tightness in section 4.2.4 of [5].

We now define the Fenchel–Legendre transform of Λ. Recall that we equip $D([0,T])$ with the supnorm topology. For each $\phi \in D([0,T]; \mathbb{R}^d)$,

$$\Lambda^*(\phi) = \sup_{\nu \in (D([0,T]))^*} \{\nu(\phi) - \Lambda(\nu)\}\,.$$

From Proposition 7 and Proposition 8 combined with an approximation of $\overline{Y}^{N,\alpha}$ by a piecewise linear continuous process (see [13] for the details), we deduce from Corollary 4.6.14 from [8] the following.

Theorem 9. *The sequence $\{\overline{Y}^{N,\alpha}, N \geq 1\}$ satisfies the Large Deviation Principle in $D([0,T]; \mathbb{R}^d)$ with the convex, good rate function Λ^* and with speed a_N, in the sense that for any Borel subset $\Gamma \subset D([0,T]; \mathbb{R}^d)$,*

$$\begin{aligned} -\inf_{\phi \in \mathring{\Gamma}} \Lambda^*(\phi) &\leq \liminf_N a_N \log \mathbb{P}(\overline{Y}^{N,\alpha} \in \Gamma) \\ &\leq \limsup_N a_N \log \mathbb{P}(\overline{Y}^{N,\alpha} \in \Gamma) \leq -\inf_{\phi \in \overline{\Gamma}} \Lambda^*(\phi)\,. \end{aligned}$$

Let us compute Λ^*. With the notation $s \wedge t := \inf(s,t)$,

$$\Lambda(\nu) = \frac{\sigma^2}{2} \int_{[0,T]^2} s \wedge t \; \nu(ds)\nu(dt)\,.$$

It is easily seen that $\Lambda^*(\phi) = +\infty$ if $\phi(0) \neq 0$. Let now $\phi \in C^2([0,T])$ such that $\phi(0) = 0$. The gradient of the map $\nu \mapsto \nu(\phi) - \Lambda(\nu)$ reads

$$\phi(t) - \sigma^2 \int_{[0,T]} s \wedge t \; \nu(ds).$$

We look for ν^* such that this gradient equals 0. This implies that

$$\begin{aligned}
&\phi'(t) = \sigma^2\nu^*((t,T]), \text{ hence}\\
&\phi'(T) = \sigma^2\nu^*(\{T\}), \quad \phi'(T) - \int_t^T \phi''(s)ds = \sigma^2\nu^*((t,T])\,,\\
&\nu^*(dt) = -\frac{1}{\sigma^2}\phi''(t)dt + \frac{1}{\sigma^2}\phi'(T)\delta_T(dt)\,.
\end{aligned}$$

From those identities, combined with $\phi(0) = 0$, we deduce that

$$\Lambda^*(\phi) = \frac{1}{2\sigma^2}\int_0^T |\phi'(t)|^2 dt\,.$$

4.2.2 Step2 : Moderate Deviations of $\widetilde{Y}^N$

What we want to show in this step is that $\widetilde{Y}^{N,\alpha}$ satisfies exactly the same Large Deviations result as $\overline{Y}^{N,\alpha}$. This will follow if we prove that $\widetilde{Y}^{N,\alpha}$ satisfies Proposition 7 (with the same expression in the limit) and Proposition 8. Let us state a property which allows us to conclude that $\widetilde{Y}^{N,\alpha}$ satisfies Proposition 7 with the correct limit.

Proposition 10. *For any $C > 0$, as $N \to \infty$,*

$$\begin{aligned}
&a_N \log \mathbb{E} \exp\left[Ca_N^{-1}\lambda(\widetilde{Y}^{N,\alpha} - \overline{Y}^{N,\alpha})\right] \to 0,\\
&a_N \log \mathbb{E} \exp\left[Ca_N^{-1}\lambda(\overline{Y}^{N,\alpha} - \widetilde{Y}^{N,\alpha})\right] \to 0\,.
\end{aligned} \tag{6}$$

We first prove

Corollary 11. *Given Proposition 7, if Proposition 10 holds true, then for any signed measure ν on $[0,T]$,*

$$a_N \log \mathbb{E}\left[e^{a_N^{-1}\nu(\widetilde{Y}^{N,\alpha})}\right] \;\to \frac{\sigma^2}{2}\int_{[0,T]^2} t\wedge s\,\nu(dt)\nu(ds)\,,$$

as $N \to \infty$.

Proof For any $\delta > 0$, we deduce from Hölder's inequality

$$\begin{aligned}
a_N \log \mathbb{E}\exp\{\nu(a_N^{-1}\widetilde{Y}^{N,\alpha})\} &= a_N \log \mathbb{E}\left[\exp\{\nu(a_N^{-1}\overline{Y}^{N,\alpha})\}\exp\{\{\nu(a_N^{-1}(\widetilde{Y}^{N,\alpha} - \overline{Y}^{N,\alpha}))\}\right]\\
&\le \frac{a_N}{1+\delta}\log \mathbb{E}\exp\{(1+\delta)a_N^{-1}\nu(\overline{Y}^{N,\alpha})\} + \frac{a_N\delta}{1+\delta}\log \mathbb{E}\exp\left\{\frac{1+\delta}{\delta a_N}\nu(\widetilde{Y}^{N,\alpha} - \overline{Y}^{N,\alpha})\right\},
\end{aligned}$$

so that, if we combine Proposition 7 and Proposition 10, we deduce that

$$\limsup_N a_N \log \mathbb{E} \exp\{\nu(a_N^{-1}\widetilde{Y}^{N,\alpha})\} \leq \frac{(1+\delta)\sigma^2}{2} \int_{[0,T]^2} t \wedge s \, \nu(dt)\nu(ds),$$

and letting $\delta \to 0$, we conclude that

$$\limsup_N a_N \log \mathbb{E} \exp\{\nu(a_N^{-1}\widetilde{Y}^{N,\alpha})\} \leq \frac{\sigma^2}{2} \int_{[0,T]^2} t \wedge s \, \nu(dt)\nu(ds).$$

For the inequality in the other direction, we note that, by similar arguments,

$$a_N \log \mathbb{E} \exp\left\{\frac{1}{a_N(1+\delta)}\nu(\overline{Y}^{N,\alpha})\right\} \leq \frac{a_N}{1+\delta} \log \mathbb{E} \exp\{a_N^{-1}\nu(\widetilde{Y}^{N,\alpha})\}$$
$$+\frac{a_N\delta}{1+\delta} \log \mathbb{E} \exp\{(\delta a_N)^{-1}\nu(\overline{Y}^{N,\alpha} - \widetilde{Y}^{N,\alpha})\},$$

so that

$$\liminf_N a_N \log \mathbb{E} \exp\{a_N^{-1}\nu(\widetilde{Y}^{N,\alpha})\} \geq \frac{\sigma^2}{2(1+\delta)} \int_{[0,T]^2} t \wedge s \, \nu(dt)\nu(ds),$$

hence, letting $\delta \to 0$ we conclude that

$$\liminf_N a_N \log \mathbb{E} \exp\{a_N^{-1}\nu(Y^{N,\alpha})\} \geq \frac{\sigma^2}{2} \int_{[0,T]^2} t \wedge s \, \nu(dt)\nu(ds).$$

□

Before we prove Proposition 10, we first need to establish a technical Lemma.

Lemma 12. *Let $\mathcal{M}$ be a standard Poisson random mesure on $\mathbb{R}_+^2$, and $\overline{\mathcal{M}}(dt,du) = \mathcal{M}(dt,du) - dt\,du$ the associated compensated measure. If φ is an $\mathbb{R}_+$–valued predictable process such that $\int_0^T \varphi_t dt$ has exponential moments of any order, and $a \in \mathbb{R}$, then there exists a constant C such that for any $0 \leq t \leq T$,*

$$\mathbb{E}\left[\sup_{0\leq s\leq t} \exp\left\{a \int_0^s \int_0^{\varphi_r} \overline{\mathcal{M}}(dr,du)\right\}\right] \leq C \left(\mathbb{E} \exp\left\{(e^{2a} - 1 - 2a) \int_0^t \varphi_s ds\right\}\right)^{1/2}.$$

Proof Consider with $b \geq 0$ the process

$$X_t = a \int_0^t \int_0^{\varphi_s} \overline{\mathcal{M}}(ds, du) - b \int_0^t \varphi_s ds \,. \tag{7}$$

It follows from Itô's formula that

$$\begin{aligned} e^{X_t} &= 1 - b \int_0^t e^{X_s} \varphi_s ds + a \int_0^t \int_0^{\varphi_s} e^{X_{s-}} \overline{\mathcal{M}}(ds, du) \\ &\quad + (e^a - 1 - a) \int_0^t \int_0^{\varphi_s} e^{X_{s-}} \mathcal{M}(ds, du) \,. \end{aligned}$$

It follows from Lemma 13 below that $M_t = \int_0^t \int_0^{\varphi_s} e^{X_{s-}} \overline{\mathcal{M}}(ds, du)$ is a martingale. Hence e^X is a martingale if $b = (e^a - 1 - a)$, a submartingale if we replace $=$ by $<$, and a supermartingale if we replace $=$ by $>$. Hence if $b \geq (e^a - 1 - a)$, $\mathbb{E} e^{X_t} \leq 1$. Now, using first Doob's L^2 inequality for submartingales, and later Cauchy's inequality, we have

$$\begin{aligned} &\mathbb{E}\left[\sup_{0 \leq s \leq t} \exp\left\{a \int_0^s \int_0^{\varphi_r} \overline{\mathcal{M}}(dr, du)\right\}\right] \\ &\lesssim \mathbb{E} \exp\left\{a \int_0^t \int_0^{\varphi_s} \overline{\mathcal{M}}(ds, du)\right\} \\ &\leq \left(\mathbb{E} \exp\left\{2a \int_0^t \int_0^{\varphi_s} \overline{\mathcal{M}}(ds, du) - 2b \int_0^t \varphi_s ds\right\}\right)^{1/2} \times \left(\mathbb{E} \exp\left\{2b \int_0^t \varphi_s ds\right\}\right)^{1/2} \end{aligned}$$

If $2b = e^{2a} - 1 - 2a$, the first factor on the last right hand side equals 1. $\square$

In order to complete the proof of Lemma 12, we still need to establish

Lemma 13. *The process φ satisfying the same assumptions as in Lemma 12, and X_t being given by (7), $M_t = \int_0^t \int_0^{\varphi_s} e^{X_{s-}} \overline{\mathcal{M}}(ds, du)$ is a martingale.*

Proof It is plain that M_t is a local martingale, whose predictable quadratic variation is given as

$$< M >_t = \int_0^t e^{2X_s} \varphi_s ds \begin{cases} \leq \exp\left\{2a \int_0^t \int_0^{\varphi_s} \mathcal{M}(ds, du)\right\} \int_0^t \varphi_s ds, & \text{if } a > 0\,; \\ \leq \exp\left\{-2(a + b) \int_0^t \varphi_s ds\right\} \int_0^t \varphi_s ds, & \text{if } a \leq 0\,. \end{cases}$$

All we need to show is that the above quantity is integrable. It is clearly a consequence of the assumption in case $a < 0$. In case $a > 0$, the second factor of the right hand side has finite exponential moments, so is square integrable, and all we need to show is that

$$\mathbb{E}\exp\left\{4a\int_0^t\int_0^{\varphi_s}\mathcal{M}(ds,du)\right\}<\infty. \tag{8}$$

Using Itô's formula we have

$$\begin{aligned}Y_t &= \exp\left\{8a\int_0^t\int_0^{\varphi_s}\mathcal{M}(ds,du)-(e^{8a}-1)\int_0^t\varphi_s ds\right\}\\ &= 1+(e^{8a}-1)\int_0^t\int_0^{\varphi_s}Y_{s-}\overline{\mathcal{M}}(ds,du).\end{aligned}$$

It is easy to conclude that $\mathbb{E}Y_t\le 1$. It follows from Cauchy–Schwartz that

$$\mathbb{E}\exp\left\{4a\int_0^t\int_0^{\varphi_s}\mathcal{M}(ds,du)\right\}\le\sqrt{\mathbb{E}Y_t}\sqrt{\mathbb{E}\exp\left\{(e^{8a}-1)\int_0^t\varphi_s ds\right\}},$$

and the result follows from our assumption on φ. □

We now turn to the

Proof of Proposition 10 We note that

$$\overline{Y}_t^{N,\alpha}-\widetilde{Y}_t^{N,\alpha}=\overline{Y}_t^{N,\alpha}-Y_t^{N,\alpha}+\lambda\int_0^t(Z_s^N-z^*)^2ds.$$

Proposition 10 will follow from the fact that for any $C>0$, as $N\to\infty$,

$$a_N\log\mathbb{E}\exp\left[Ca_N^{-1}\nu(Y^{N,\alpha}-\overline{Y}^{N,\alpha})\right]\to 0, \tag{9}$$

$$a_N\log\mathbb{E}\exp\left[Ca_N^{-1}\nu(\overline{Y}^{N,\alpha}-Y^{N,\alpha})\right]\to 0, \tag{10}$$

$$a_N\log\mathbb{E}\exp\left[Ca_N^{-1}N^\alpha\nu\left(\int_0^\cdot(Z_s^N-z^*)^2ds\right)\right]\to 0\,. \tag{11}$$

We shall prove (9) and (11). The proof of (10) is quite similar to that of (9).
Step 1 : Proof of (9) It suffices to consider one of the terms in the sum over j, and we suppress the index j for simplicity. We note that

$$\begin{aligned}a_N^{-1}(Y^{N,\alpha}-\overline{Y}^{N,\alpha}) &= N^{-\alpha}\int_0^t\int_{N\beta(z^*)}^{N[\beta(Z_s^N)\vee\beta(z^*)]}\overline{\mathcal{M}}(ds,du)\\ &\quad -N^{-\alpha}\int_0^t\int_{N\beta(Z_s^N)}^{N[\beta(Z_s^N)\vee\beta(z^*)]}\overline{\mathcal{M}}(ds,du)\end{aligned}$$

It is not hard to see that one can treat each of the two terms on the right separately, and we treat only the first term, the treatment of the second one being quite similar. We note that there exists a compensated Poisson process on $\mathbb{R}_+$ M such that this first term can be rewritten as

$$V_t^N := N^{-\alpha} M\left(N\int_0^t (\beta(Z_s^N)-\beta(z^*))^+ ds\right) .$$

We need to estimate $\mathbb{E}\exp[Ca_N^{-1}\nu(V^N)]$. If we decompose the signed measure ν as the difference of two measures as follows $\nu=\nu_+-\nu_-$, we again have two terms, and it suffices to treat one of them, say ν_+. Of course it suffices to treat the case where $\nu_+\neq 0$. Since the positive constant C is arbitrary, we can w.l.o.g. assume that ν_+ is a probability measure on $[0,T]$. It is then clear that

$$\exp\left[Ca_N^{-1}\int_0^T V_t^N \nu_+(dt)\right] \le \exp\left[Ca_N^{-1}\sup_{0\le t\le T} V_t^N\right] .$$

We choose a new parameter $0<\varrho<\alpha$, and we split the expression whose expectation needs to be estimated in two terms.

$$\begin{aligned}\exp\left\{CN^{-\alpha}\sup_{0\le t\le T} V_t^N\right\} &= \exp\left\{CN^{-\alpha}\sup_{0\le t\le T} V_t^N\right\}\mathbf{1}_{\sup_{0\le t\le T}|Z_t^N-z^*|\le N^{-\varrho}}\\ &\quad+\exp\left\{CN^{-\alpha}\sup_{0\le t\le T} V_t^N\right\}\mathbf{1}_{\sup_{0\le t\le T}|Z_t^N-z^*|> N^{-\varrho}}\end{aligned} \tag{12}$$

We now estimate the first term on the right hand side of (12). For that sake, we define the stopping time

$$\sigma_N=\inf\{0\le t\le T;\,|Z_t^N-z^*|>N^{-\varrho}\}$$

and note that

$$\begin{aligned}&\exp\left\{CN^{-\alpha}\sup_{0\le t\le T} M\left(N\int_0^t(\beta(Z_s^N)-\beta(z^*))^+ds\right)\right\}\mathbf{1}_{\sup_{0\le t\le T}|Z_t^N-z^*|\le N^{-\varrho}}\\ &\quad\le \exp\left\{CN^{-\alpha}\sup_{0\le t\le T} M\left(N\int_0^{t\wedge\sigma_N}(\beta(Z_s^N)-\beta(z^*))^+ds\right)\right\}\end{aligned}$$

Consequently the expectation of the first term on the right of (12) is bounded from above by

$$\begin{aligned}
&\mathbb{E}\exp\left\{CN^{-\alpha}\sup_{0\le t\le T} M\left(N\int_0^{t\wedge\sigma_N}(\beta(Z_s^N)-\beta(z^*))^+ds\right)\right\}\\
&\quad\le \mathbb{E}\exp\left\{(e^{2CN^{-\alpha}}-1-2CN^{-\alpha})N\int_0^{T\wedge\sigma_N}(\beta(Z_t^N)-\beta(z^*))^+dt\right\}\\
&\quad\le \exp\left\{cN^{1-2\alpha-\varrho}\right\}\ ,
\end{aligned}$$

where the first inequality follows from Lemma 12, and the second one exploits the Lipschitz property of β. Consider now the second term on the right hand side of (12).

$$\begin{aligned}
&\mathbb{E}\left(\exp\left\{N^{-\alpha}\sup_{0\le t\le T} M\left(N\int_0^{t}(\beta(Z_s^N)-\beta(z^*))^+ds\right)\right\}\mathbf{1}_{\sup_{0\le t\le T}|Z_t^N-z^*|>N^{-\varrho}}\right)\\
&\le\left(\mathbb{E}\exp\left\{2N^{-\alpha}\sup_{0\le t\le T} M\left(N\int_0^{t}(\beta(Z_s^N)-\beta(z^*))^+ds\right)\right\}\right)^{1/2}\\
&\quad\times\mathbb{P}\left(\sup_{0\le t\le T}|Z_t^N-z^*|>N^{-\varrho}\right)^{1/2}\\
&\le \exp\left\{cN^{1-2\alpha}\right\}\mathbb{P}\left(\sup_{0\le t\le T}\left|M\left(N\int_0^t\beta(Z_s^N)ds\right)\right|>N^{1-\varrho}\right)^{1/2}\ ,
\end{aligned}$$

where the second inequality follows from Lemma 12 and the boundedness of β. For the second factor in the last expression, we need to consider

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le T} M\left(N\int_0^t\beta(Z_s^N)ds\right)>N^{1-\varrho}\right)\quad\text{and}\\
&\mathbb{P}\left(\sup_{0\le t\le T}\left(-M\left(N\int_0^t\beta(Z_s^N)ds\right)\right)>N^{1-\varrho}\right)\ .
\end{aligned}$$

Both probabilities are estimated in a similar way. By an exponential estimate,

$$\mathbb{P}\left(\sup_{0\le s\le t} M\left(N\int_0^s\beta(Z_r^N)dr\right)>N^{1-\varrho}\right)\lesssim\exp\{-(16ct)^{-1}N^{1-2\varrho}\},\tag{13}$$

for N large enough. Finally the expectation of the second term of the right hand side of (12) is bounded by $\exp\{c_1N^{1-2\alpha}-c_2N^{1-2\varrho}\}$, with $c_1,c_2>0$, and

$$\mathbb{E}\exp\left\{N^{-\alpha}\sup_{0\le s\le t} M\left(N\int_0^t(\beta(Z_s^N)-\beta(z^*))^+ds\right)\right\}\le e^{cN^{1-2\alpha-\varrho}}+e^{c_1N^{1-2\alpha}-c_2N^{1-2\varrho}}.$$

From the inequality $\log(a+b) \leq \log(2) + \log(\sup(a,b))$, for N large enough,

$$a_N \log \mathbb{E} \exp \left\{ N^{-\alpha} \sup_{0\leq s\leq t} M \left(N \int_0^t (\beta(Z_s^N) - \beta(z^*))^+ ds \right) \right\} \leq a_N \log(2) + cN^{-\varrho},$$

which establishes (9).

Step 2 : Proof of (11) We in fact must prove that for any $C > 0$, as $N \to \infty$,

$$a_N \log \mathbb{E} \left[\exp \left\{ a_N^{-1} C N^\alpha \int_0^T (Z_t^N - z^*)^2 dt \right\} \right] \to 0\,.$$

In this proof, C will denote a constant whose value may change from line to line. We now introduce a new process, where $\bar{\beta} = \sup_{0\leq z\leq 1} \beta(z)$,

$$X_t^N := \frac{1}{N} \int_0^t \int_0^{N\bar{\beta}} \overline{\mathcal{M}}(ds, du)\,,$$

the event

$$A_b^N := \left\{ \sup_{0\leq t\leq T} |\overline{Y}_t^N| \leq b \right\} \cap \left\{ X_t^N \leq \bar{\beta} T \right\}\,,$$

and the stopping time

$$\bar{\tau}_b := \inf\{t > 0,\, |\overline{Y}_t^N| > b\} \wedge \inf\{t,\, X_t^N > \bar{\beta} T\}\,,$$

where the constant b will be chosen below. From (4), the fact that $|z^* - z_N^*| \leq N^{-1}$ and Cauchy–Schwartz,

$$a_N \log \mathbb{E} \left[\exp \left\{ a_N^{-1} C N^\alpha \int_0^T (Z_t^N - z^*)^2 dt \right\} \right]$$

$$\lesssim N^{\alpha-1} + a_N \log \mathbb{E} \left[\exp \left\{ a_N^{-1} C N^\alpha \sup_{0\leq t\leq T} |Y_t^N| \mathbf{1}_{(A_b^N)^c} \right\} \right] \tag{14}$$

$$+ a_N \log \mathbb{E} \left[\exp \left\{ a_N^{-1} C N^\alpha \sup_{0\leq t\leq T} |Y_s^N|^2 \mathbf{1}_{A_b^N} \right\} \right], \tag{15}$$

We take the limit successively in the two terms of the above right hand side.

Step 2a : Estimate of (14) We have

$$\mathbb{E} \left[\exp \left\{ a_N^{-1} C N^\alpha \sup_{0\leq t\leq T} |Y_s^N| \mathbf{1}_{(A_b^N)^c} \right\} \right]$$
$$\leq \mathbb{E} \left[\exp \left\{ C N^{-\alpha} \sup_{0\leq t\leq T} \left| \int_0^t \int_0^{N\beta(Z_s^N)} \overline{\mathcal{M}}(ds, du) \right| \right\} \mathbf{1}_{(A_b^N)^c} \right] + 1,$$

It remains to note that

$$\mathbb{E}\left[\exp\left\{CN^{-\alpha}\sup_{0\le t\le T}\left|\int_0^t\int_0^{N\beta(Z_s^N)}\overline{\mathcal{M}}(ds,du)\right|\right\}\right]\lesssim e^{CN^{1-2\alpha}},$$

and

$$\mathbb{P}((A_b^N)^c)\le\mathbb{P}\left(\sup_{0\le t\le T}|Y_t^N|>b\right)+\mathbb{P}\left(\sup_{0\le t\le T}X_t^N>\bar{\beta}T\right)\lesssim e^{-CN},$$

for some positive constant C, so that finally there exist two positive constants C_1 and C_2 such that, for N large enough,

$$\mathbb{E}\left[\exp\left\{a_N^{-1}CN^{\alpha}\sup_{0\le t\le T}|Y_s^N|\mathbf{1}_{(A_b^N)^c}\right\}\right]\le 1+\exp\{C_1N^{1-2\alpha}-C_2N\}\le 2.$$

Step 2b : Estimate of (15) Since Y^N is a martingale, it is clear that the process

$$\exp\left\{a_N^{-1}\frac{C}{2}N^{\alpha}|Y_t^N|^2\right\}$$

is a submartingale. Consequently, from Doob's L^2 submartingale inequality,

$$\begin{aligned}\mathbb{E}\left[\exp\left\{a_N^{-1}CN^{\alpha}\sup_{0\le t\le T}|Y_t^N|^2\mathbf{1}_{A_b^N}\right\}\right]&\le 4\mathbb{E}\left[\exp\left\{a_N^{-1}CN^{\alpha}|Y^N_{T\wedge\bar{\tau}_b}|^2\right\}\right]\\ &\le\sqrt{\mathbb{E}\left[\exp\left\{2CN^{1-\alpha}\left(|Y^N_{T\wedge\bar{\tau}_b}|^2-|\overline{Y}^N_{T\wedge\bar{\tau}_b}|^2\right)\right\}\right]}\\ &\quad\times\sqrt{\mathbb{E}\left[\exp\left\{2CN^{1-\alpha}|\overline{Y}^N_{T\wedge\bar{\tau}_b}|^2\right\}\right]}\end{aligned}\tag{16}$$

Consider first the first factor on the right hand side of (16). We have

$$\begin{aligned}|Y^N_{T\wedge\bar{\tau}_b}|^2-|\overline{Y}^N_{T\wedge\bar{\tau}_b}|^2&=\left(Y^N_{T\wedge\bar{\tau}_b}-\overline{Y}^N_{T\wedge\bar{\tau}_b}\right)\left(Y^N_{T\wedge\bar{\tau}_b}+\overline{Y}^N_{T\wedge\bar{\tau}_b}\right)\\ &\le(b+\bar{\beta}T)\left|Y^N_{T\wedge\bar{\tau}_b}-\overline{Y}^N_{T\wedge\bar{\tau}_b}\right|,\end{aligned}$$

and the result follows from (9) and (10).

We finally consider the second term in the right hand side of (16). We have

$$\left|\overline{Y}^N_{T\wedge\bar{\tau}_b}\right|^2\le\left|\overline{Y}^N_T\right|^2\mathbf{1}_{\{|Y_T^N|\le b\}}+(b+N^{-1})^2\mathbf{1}_{\{\bar{\tau}_b<T\}}.$$

Hence the second term on the right of (16) satisfies

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left\{CN^{1-\alpha}|\overline{Y}^N_{T\wedge\bar{\tau}_b}|^2\right\}\right] \\
(17)\qquad &\leq \sqrt{\mathbb{E}\left[\exp\left\{2CN^{1-\alpha}|\overline{Y}^N_T|^2\mathbf{1}_{\{|Y^N_T|\leq b\}}\right\}\right]\mathbb{E}\left[\exp\left\{2C'N^{1-\alpha}\mathbf{1}_{\{\bar{\tau}_b<T\}}\right\}\right]}
\end{aligned}
$$

We now write

$$
\overline{Y}^N_T=\frac{1}{N}\int_0^T\int_0^{N\beta(z^*)}\overline{\mathcal{M}}(ds,du)=\frac{\sqrt{T\beta(z^*)}}{\sqrt{N}}\xi_N,
$$

with

$$
\xi_N=\frac{\theta_N-aN}{\sqrt{aN}},\quad \text{where } \theta_N\sim \mathrm{Poi}(aN),\ a=T\beta(z^*).
$$

Clearly $\xi_N\Rightarrow\mathcal{N}(0,1)$. We choose $b=a/3$.

$$
\begin{aligned}
&\mathbb{E}\exp\left\{CN^{-\alpha}|\xi_N|^2\mathbf{1}_{\{|\xi_N|\leq\sqrt{aN}/3\}}\right\} \\
&\quad=\sum_{k=2aN/3}^{4aN/3}\exp\left\{CN^{-\alpha}\frac{(k-aN)^2}{aN}\right\}e^{-aN}\frac{(aN)^k}{k!} \\
&\quad\sim\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2\right\}e^{-aN}\frac{(aN)^{aN+x\sqrt{aN}}}{(aN+x\sqrt{aN})!}\sqrt{aN}dx \\
&\quad\lesssim\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{N}/3}^{\sqrt{N}/3}\exp\left\{CN^{-\alpha}x^2\right\}e^{x\sqrt{aN}}\left(1+\frac{x}{\sqrt{aN}}\right)^{-(aN+x\sqrt{aN})}dx \\
&\quad\leq\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2-\frac{x^2}{3}\right\}dx
\end{aligned}
$$

We have proved that the first factor on the right of (17) remains bounded, as $N\to\infty$. Next consider the second term on the right of (17). We have

$$
\exp\left\{4C'N^{1-\alpha}\mathbf{1}_{\{\tau_b<T\}}\right\}\leq 1+\exp\left\{4C'N^{1-\alpha}\right\}\mathbf{1}_{\{\tau_b<T\}},\ \text{and}
$$

$$
\mathbb{P}\left(\bar{\tau}_b<T\right)\leq\mathbb{P}\left(\sup_{0\leq t\leq T}|\overline{Y}^N_t|>b\right)+\mathbb{P}\left(\sup_{0\leq t\leq T}X^N_t>\overline{\beta}T\right)\lesssim e^{-cN}.
$$

It follows that the second factor in (16) is bounded from above by

$$
1+\exp\{C_1N^{1-\alpha}-C_2N\},
$$

where C_1 and C_2 are two positive constants. This last expression is bounded say by 2, as soon as N is large enough. □

4.2.3 Step3 : Moderate Deviations of $Z^N - z^*$

Recall that $\widetilde{Z}^{N,\alpha} := N^\alpha(Z^N - z^*)$ is the image of $\widetilde{Y}^{N,\alpha}$ by the mapping $x \mapsto y$ from $D([0,T])$ into itself, which is continuous if we equip $D([0,T])$ with the supnorm topology, defined by:

$$y(t) = -(\lambda - \gamma)\int_0^t y(s)ds + x(t),\ 0 \le t \le T.$$

Note also that the above mapping is a bijection. The following result is then a consequence of Corollary 11 and the contraction principle.

Theorem 14. *The collection of processes $\{N^\alpha(Z_t^N - z^*),\ 0 \le t \le T\}_{N\ge1}$ satisfies a large deviation principle with the good rate function*

$$I_T(\phi) = \begin{cases} \frac{1}{2\sigma^2}\int_0^T |\phi'(t) + (\lambda-\gamma)\phi(t)|^2 dt, & \textit{if } \phi \textit{ is absolutely continuous;} \\ +\infty, & \textit{otherwise.} \end{cases}$$

Remark 15. *While the LD rate function of a Poisson driven stochastic differential equation is very different from the rate function of LDs for its Brownian driven diffusion approximation, the rate function for moderate deviations of Poisson driven stochastic differential equations is identical to that of LDs for its Brownian driven diffusion approximation.*

4.2.4 Wentzell–Freidlin theory and extinction of an epidemic

We want to conclude from the Wentzell–Freidlin theory an estimate of the time needed for $N^\alpha(Z_t^N - z^*)$ to make a deviation of $-c$, i.e. to go from 0 to $-c$, which, for the value of N such that $z^* = cN^{-\alpha}$, means the time for Z_t^N to hit 0. For that sake, we first compute

$$\overline{V}_c = \min_{T>0}\ \min_{\phi,\ \phi(0)=0,\phi(T)=-c} I_T(\phi).$$

An application of Pontryaguin's maximum principle, see [15], yields

$$\overline{V}_c = \frac{\lambda}{2\gamma}c^2\,.$$

Using the same arguments as in [12] and [5], we then deduce from Theorem 14

Theorem 16. *Let* $T_c^{N,\alpha} := \inf\{t > 0,\ \widetilde{Z}_t^{N,\alpha} \le -c\}$*. For any* $\delta > 0$*,*

$$\lim_{N\to\infty} \mathbb{P}\left(\exp\{a_N^{-1}(\overline{V}_c - \delta)\} < T_c^{N,\alpha} < \exp\{a_N^{-1}(\overline{V}_c + \delta)\}\right) = 1\,.$$

Moreover

$$\lim_{N\to\infty} a_N \mathbb{E}(T_c^{N,\alpha}) = \overline{V}_c\,.$$

Recall that $a_N^{-1} = N^{1-2\alpha}$. In the CLT regime, $\alpha = 1/2$, $a_N^{-1} = 1$, while in the LD regime, $\alpha = 0$, $a_N^{-1} = N$.

Let us now compute the corresponding critical population size. $e^{N^{1-2\alpha}\overline{V}_c}$ is the order of magnitude of the time needed for $Z_t^N - z_t$ to make a deviation of size $cN^{-\alpha}$. This is sufficient to extinguish an epidemic, provided i^* is of the same order, so that the corresponding critical size is $N_\alpha \sim (1/i^*)^{1/\alpha}$, that is roughly the CLT critical population size raised to the power $1/2\alpha$. In the case of the SIR model with demography for measles, the CLT critical population size is of the order of a few millions, so e.g. with $\alpha = 1/3$, we go from 10^6 to 10^9, i.e. a few billions, which is the order of magnitude of of the biggest countries, China and India.

References

[1] Anderson, R. M. and May, R. M. *Infectious diseases of humans; dynamic and control*. Oxford University Press, 1991.

[2] Andersson, H. and Britton, T. *Stochastic epidemic models and their statistical analysis*. Springer Lecture Notes in Statistics. New York: Springer Verlag, 2000.

[3] Bailey, N.T.J. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, 1975.

[4] Bernoulli, D. (1760) Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir, Mém. Math. Phys. Acad. Roy. Sci., Paris 1–45, 1760.

[5] Britton, T. and Pardoux, E., Stochastic Epidemics in a Homogeneous Community, arXiv:1808.0535, to appear in [6].

[6] Britton, T. and Pardoux E. eds., *Stochastic epidemic models with inference*, Springer, 2019.

[7] Daley, D.J. and Gani, J. *Epidemic Modelling: an introduction*. Cambridge University Press, Cambdridge, 1999.

[8] Dembo, A. and Zeitouni, O., *Large Deviations, Techniques and Applications*, 2d ed., Applications of Mathematics **38**, Springer, New York, 1998.

[9] Diekmann, O., Heesterbeek, J.A.P. and Britton, T. *Mathematical tools for understanding infectious disease dynamics*. Princeton UP, 2013.

[10] Freidlin, M.I. and Wentzell, A.D., *Random perturbations of dynamical systems*, 3d ed. Grundlehren des Mathematischen Wissenschaften **260**, Springer, New York, 2012.

[11] Keeling, M.J. and Rohani, P. *Modeling infectious diseases in humans and animals*. Princeton UP, 2008.

[12] Kratz, P. and Pardoux, E., Large deviations for infectious diseases models, in *Séminaire de Probabilités XLIX*, C. Donati-Martin, A. Lejay, A. Rouault eds., Lecture Notes in Math. **2215**, pp. 221-327, 2018.

[13] Pardoux, E. , Moderate deviations and extinction of an endemic disease, submitted. arXiv:1905.08986

[14] Pardoux, E. and Samegni–Kepgnou, B., Large deviation principle for epidemic models, *Journal of Applied Probability* **54**, 905–920, 2017.

[15] Pontryagin, L. S.; Boltyanskii, V. G.; Gamkrelidze, R. V.; Mishchenko, E. F. (1962) *The mathematical theory of optimal processes*. Transl. by K. N. Trirogoff; ed. by L. W. Neustadt, John Wiley & Sons.