Electron. J. Probab. **0** (2012), no. 0, 1–28.
ISSN: 1083-6489 DOI: 10.1214/EJP.vVOL-PID

Moderate Deviations and Extinction of an Epidemic

Etienne Pardoux*

Abstract

Consider an epidemic model with a constant flux of susceptibles, in a situation where the corresponding deterministic epidemic model has a unique stable endemic equilibrium. For 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 paper, we discuss how the Central Limit Theorem, Moderate and Large Deviations allow us to give estimates of the extinction time of the epidemic.

Keywords: Extension of an epidemic, central limit theorem, large deviations, moderate deviations.
AMS MSC 2010: 60J75, 60F05, 60F10, 92D30.
Submitted to EJP on May 25, 2019, final version accepted on January 31, 2020.

1 Introduction

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

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.

*Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France. E-mail: etienne.pardoux@univ-amu.fr

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 law of large numbers and central limit theorems are rather old. They can be found e.g. in chapter 11 of Ethier and Kurtz [3]. There are also presented, in the framework of epidemic models, in Britton and Pardoux [1]. The Large Deviations results are close to those presented in Shwartz and Weiss [10], [11], although their assumptions are not quite satisfied in our models. Derivations adapted to our setup can be found in Kratz and Pardoux [5], Pardoux and Samegni–Kepgnou [7], and Britton and Pardoux [1]. The results concerning moderate deviations are new and constitute the core of this paper. Our derivation is essentially based upon a process version of the Gärtner–Ellis Theorem, Corollary 4.6.14 from Dembo and Zeitouni [2]. Our main results are Theorem 4.10 and Theorem 4.13. We also give the expression for the rate function in our three models of interest in section 4.5, and in case of the SIS model we give an explicit formula for the quasi–potential in section 4.7.2. Finally, again for the case of the SIS model, we compare in section 4.8 the upper bound of fluctuations given respectively by the central limit theorem, moderate deviations, and large deviations.

The results of this paper have been anounced in [6]

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. In section 4, we establish the moderate deviations result and explain how it can be used to predict the time taken for an epidemic to cease. Finally an Appendix establishes an estimate of exponential moments of the integral with respect to a compensated Poisson random measure. This estimate is used several times in our proofs.

In this paper, the same letter C denotes an arbitrary constant, whose value may change from line to line. The value at time t of a function ϕ will be denoted ϕ_t, but the value at time t of the first coordinate of ϕ will be denoted $\phi_1(t)$.

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 ODE. If we let $z(t) = i(t)$, we have $s(t) = 1 - z(t)$, and we obtain the ODE

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

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

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.

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

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

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

Note that $S_t^N = N^{-1}\mathcal{S}_t^N$ and $I_t^N = N^{-1}\mathcal{I}_t^N$, so that in particular all those processes live on the same time scale. The formulation of such a rate of infections 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 individuals at time t gives the above rate. Concerning recovery, it is assumed that each infectious individual recovers at rate γ, independently of the others.

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 a given rate multiplied by N, 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

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

Remark 2.1. One may think that it would be more natural to decide that births happen at rate μ times the total population. 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 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 N, 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.

$$\begin{aligned} Z_t^N &= z_N + \frac{1}{N} \sum_{j=1}^k h_j P_j \left(N \int_0^t \beta_j(Z_s^N) ds\right) \\ &= z_N + \int_0^t b(Z_s^N) ds + \frac{1}{N} \sum_{j=1}^k h_j M_j \left(N \int_0^t \beta_j(Z_s^N) ds\right), \end{aligned} \tag{3.1}$$

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 expressions has the advantage of being concise, we shall rather use the following equivalent formulation of (3.1). Let $\{\mathcal{M}_j,\, 1 \leq j \leq k\}$ 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 (3.1) in the form

$$\begin{aligned} Z_t^N &= z_N + \frac{1}{N}\sum_{j=1}^k h_j \int_0^t \int_0^{N\beta_j(Z_s^N)} \mathcal{M}_j(ds,du) \\ &= z_N + \int_0^t b(Z_s^N)ds + \frac{1}{N}\sum_{j=1}^k h_j \int_0^t \int_0^{N\beta_j(Z_s^N)} \overline{\mathcal{M}}_j(ds,du), \end{aligned} \tag{3.2}$$

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 (3.1) or (3.2) for its definition.

Let us state the assumptions which we will need in section 4 below. Those are more than necessary for the results of the present section to hold, see [1] for the proofs.

$(H.1)$ β_j is bounded , $1 \le j \le k$;

$(H.2)$ $b \in C^1(\mathbb{R}^d;\mathbb{R}^d)$, and $\nabla b : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$ is bounded and Lipshitz.

Remark 3.1. In practice, in our models, either the process Z_t^N takes its values in a compact subset of $\mathbb{R}^d$ (this is the case for all models with a constant population size), or else we restrict ourselves to such a situation, by stopping the process when the total population exceeds a given large value, see section 4.2.7 in [1]. As a consequence, the boundedness of β_j is not a severe assumption, while $(H.2)$ would follow from the fact that b is of class C^2. It is possible that our results could be extended to b being C^1, since it implies that ∇b is continuous, hence uniformly continuous since its argument remains in a compact set, but that would require a very careful check.

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 3.2. *Let Z_t^N denote the solution of the SDE* (3.1)*. Then $Z_t^N \to z_t$ a.s. locally uniformly in t, where $\{z_t,\, t \ge 0\}$ is the unique solution of the ODE*

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

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 3.3. *As $N \to \infty$, $\{U_t^N,\, t \ge 0\} \Rightarrow \{U_t,\, t \ge 0\}$ for the topology of locally uniform convergence, where $\{U_t,\, t \ge 0\}$ is a Gaussian process of the form*

$$U_t = \int_0^t \nabla_x b(s,z_s)U_s ds + \sum_{j=1}^k h_j \int_0^t \sqrt{\beta_j(s,z_s)}dB_j(s),\ t \ge 0\,, \tag{3.3}$$

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

The proof of this theorem can be found e.g. in [1]. It relies essentially upon the fact that if P is a rate 1 Poisson process, then $\sqrt{N}(P(Nt)/N - t)_{t\geq 0} \Rightarrow B$, where B is a standard Brownian motion.

Remark 3.4. One may wonder whether the above two theorems are compatible with the fact that for each N fixed, as we shall see below, extinction happens in finite time a.s. However, the extinction time tends to $+\infty$ as $N \to \infty$, so that the determinsitic limit of Z^N is a function whose infectious component remains > 0 for all t, and for any fixed t the fluctuations $\sqrt{N}(Z_t^N - z_t)$ tend to be Gaussian as $N \to \infty$.

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 $c_j(t) = 0\ dt$ a.e. on the set $\{t,\ \beta_j(\phi_t) = 0\}$ and

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

We define the rate function

$$I_{T,z}(\phi) := \begin{cases} \inf_{c\in\mathcal{A}_k(\phi)} I_T(\phi|c), & \text{if } \phi \in \mathcal{AC}_{T,d} \text{ and } \phi_0 = z; \\ \infty, & \text{otherwise.} \end{cases}$$

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 $0\log(0/0) = 0\log(0) = 0$. Given $z \in \mathbb{R}^d$, we shall denote by Z^{N,z_N} the solution of equation (3.1), starting from $Z_0^{N,z_N} = z_N = [Nz]/N$. The collection Z^{N,z_N} obeys a Large Deviations Principle, in the sense that

Theorem 3.5. *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 $F \subset D([0,T];\mathbb{R}^d)$,

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

A proof of this result can be found in [5] and in [1]. A slight reinforcement of this theorem allows us to conclude Theorem 3.6ă below. 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 ODE which appears in Theorem 3.2 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 3.6. *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,

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

We refer for the proof of this Theorem to [5] and [1].

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, see [9], and deduce in that case that $\overline{V} = \log\frac{\lambda}{\gamma} - 1 + \frac{\gamma}{\lambda}$. For other models, one can compute numerically a good approximation of the value of $\overline{V}$ for each given value of the parameters.

3.5 CLT and extinction of an epidemic

The discussion of this subsection, which motivates the moderate deviations approach of this paper, is taken from section 4.1 in [1]. Consider the SIR with demography.

$$i'(t) = \lambda i(t)s(t) - \gamma i(t) - \mu i(t),$$
$$s'(t) = -\lambda i(t)s(t) + \mu - \mu s(t).$$

We assume that $\lambda > \gamma + \mu$, in which case there is a unique stable endemic equilibrium, namely $z^* = (i^*, s^*) = (\frac{\mu}{\gamma+\mu} - \frac{\mu}{\lambda}, \frac{\gamma+\mu}{\lambda})$. We can study the extinction of an epidemic in the above model using the CLT. We note that the basic reproduction number R_0 and the expected relative time of a life an individual is infected, ε, are given by

$$R_0 = \frac{\lambda}{\gamma+\mu} \qquad\qquad \varepsilon = \frac{1/(\gamma+\mu)}{1/\mu} = \frac{\mu}{\gamma+\mu}. \tag{3.4}$$

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 we use the approximations $R_0 \approx \lambda/\gamma$ 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 central limit theorem tell us that $\sqrt{N}(I_t^N - i^*)$ converges to a Gaussian process, whose asymptotic variance can be shown to well approximated by $R_0^{-1} - R_0^{-2} \approx R_0^{-1}$, see Exercise 4.1.3 in [1], and its solution on pages 116–117. Note that $I_t^N = z_1^* + N^{-1/2}U_1^N(t) \approx z_t + N^{-1/2}U_1(t)$, and for large t the law of $U_1(t)$ is approximately $N(0, R_0^{-1})$. This suggests that for large t, the number of infectious individuals in the population is approximately Gaussian, with mean Ni^* and standard deviation $\sqrt{N/R_0}$. If Ni^* and $\sqrt{N/R_0}$ are of the same order, i.e. N is of the same order as $\frac{1}{(i^*)^2R_0}$, it is likely that the fluctuations described by the central limit theorem explain that the epidemic might cease in time of order one. This gives a critical population size roughly of the order of

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

in fact probably a bit larger than that.

Consider measles prior to vaccination. In that case it is known that $R_0 \approx 15$, and $\varepsilon \approx \frac{1/75}{1/(1/52)+1/75} \approx 1/3750$ we arrive at $N_c \sim (3750)^2/15$, which is almost 10^6. So, if the population is at most a million (or perhaps a couple of millions), we expect that the disease will go extinct quickly, whereas the disease will become endemic (for a rather long time) in a significantly larger population. 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).

4 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.

The assumptions $(H.1)$ and $(H.2)$ are assumed to hold throughout this section.

4.1 The set–up and preliminary estimates

We shall use the general model written in the form (3.2). We assume that the limiting law of large numbers ODE

$$z(t) = z + \int_0^t b(z(s))ds$$

has a unique stable equilibrium point z^* such that $z_1^* > 0$, called the endemic equilibrium, which is such that, provided $z_1(0) > 0$, $z(t) \to z^*$ as $t \to \infty$.

For the sake of simplifying many formulas below, we change our coordinates, and let $z^* = 0$. The reader should be aware of the fact that there is a price to pay for that translation of the origin. Indeed, since in the original coordinate system, the process Z_t^N was living on the set of vectors whose coordinates are integer multiples of N^{-1} (this is essential for the process to remain in the set where it makes sense, i.e. for proportions to remain between 0 and 1), the new origin generically does not belong to the set of point in $\mathbb{R}^d$ which our process Z_t^N may visit. The grid on which Z_t^N lives is translated by the vector $z^* - \{z^*\}_N$, where here and below $\{z\}_N := [Nz]/N$, $[Nz]$ denoting the vector whose i–th component is the integer part of the i–th component of Nz. However, this minor complexity will appear only in the formula for the initial condition of the SDE. Once the SDE starts on the correct grid, the solution remains there.

From now on 0 will be the endemic equilibrium (of course in the translated coordinate system), while $z^* \neq 0$ will denote that endemic equilibrium in the original coordinates (we shall need it for the formula of the initial condition of the SDE).

We want to study the moderate deviations at scale N^α of Z_t^N, where $0 < \alpha < 1/2$. Note that $\alpha = 0$ would correspond to the large deviations, and $\alpha = 1/2$ to the central limit theorem. We shall need below to consider the ODE starting from a point close to $z^* = 0$, namely we shall consider the function $\{z_N(t),\ 0 \le t \le T\}$, solution of the ODE

$$z_N(t) = N^{-\alpha} z + \int_0^t b(z_N(s))ds,$$

where $z \in \mathbb{R}^d$ is arbitrary. In fact, we shall be more interested in $\overline{z}_N(t) := N^\alpha z_N(t)$, which solves (below we exploit the fact that $b(0) = 0$)

$$\begin{aligned}\overline{z}_N(t) &= z + N^\alpha \int_0^t b(z_N(s))ds \\ &= z + \int_0^t \nabla b(0)\overline{z}_N(s)ds + \int_0^t \int_0^1 \left[\nabla b(\theta z_N(s)) - \nabla b(0)\right] d\theta\, \overline{z}_N(s)ds.\end{aligned}$$

It is not hard to prove that, under our standing assumption $(H.2)$ that b is of class C^1 and ∇b is bounded, as $N \to \infty$, $\overline{z}_N(t) \to \overline{z}(t)$ uniformly for $0 \le t \le T$, where $\overline{z}(t)$ solves the linearized ODE near the endemic equilibrium 0 :

$$\overline{z}(t) = z + \int_0^t \nabla b(0)\overline{z}(s)ds\,. \tag{4.1}$$

We want to study the moderate deviations of the process Z_t^N solution of the SDE (3.1) with the initial condition $z_N := \{z^* + N^{-\alpha} z\}_N - z^*$. This amounts to study the large deviations of $Z_t^{N,\alpha} := N^\alpha Z_t^N$ at speed $a_N = N^{2\alpha-1}$. We define

$$Y_t^N = \frac{1}{N} \sum_{j=1}^k h_j \int_0^t \int_0^{N\beta_j(Z_s^N)} \overline{\mathcal{M}}_j(ds, du), \text{ and } Y_t^{N,\alpha} = N^\alpha Y_t^N.$$

With these notations, the SDE for $Z_t^{N,\alpha}$ reads

$$\begin{aligned} Z_t^{N,\alpha} &= N^\alpha \left(\{z^* + N^{-\alpha} z\}_N - z^* \right) + \int_0^t N^\alpha b\left(N^{-\alpha} Z_s^{N,\alpha}\right) ds + Y_t^{N,\alpha} \\ &= N^\alpha \left(\{z^* + N^{-\alpha} z\}_N - z^* \right) + \int_0^t \nabla b(0) Z_s^{N,\alpha} ds + \int_0^t V_s^{N,\alpha} ds + Y_t^{N,\alpha}, \text{ where} \\ V_s^{N,\alpha} &= N^\alpha b\left(N^{-\alpha} Z_s^{N,\alpha}\right) - \nabla b(0) Z_s^{N,\alpha} = \left[\int_0^1 \left(\nabla b\left(\theta N^{-\alpha} Z_s^{N,\alpha}\right) - \nabla b(0) \right) d\theta \right] Z_s^{N,\alpha}. \end{aligned}$$

If we let $K := \sup_z \|\nabla b(z)\|$, we have

$$\|Z_t^{N,\alpha}\| \le \|z\| + \sqrt{d} N^{\alpha-1} + K \int_0^t \|Z_s^{N,\alpha}\| ds + \|Y_t^{N,\alpha}\|.$$

This combined with Gronwall's Lemma yields

$$\|Z_t^{N,\alpha}\| \le e^{Kt} \left(\|z\| + \sqrt{d} N^{\alpha-1} + \sup_{0\le s\le t} \|Y_s^{N,\alpha}\| \right). \tag{4.2}$$

From the boundedness and Lipschitz property of ∇b, and the formula for $V^{N,\alpha}$, we deduce te following (here and in the rest of the paper, $x_N \lesssim y_N$ means that there exists a constant C which is independent of N, such that $x_N \le C y_N$).

$$\|V_t^{N,\alpha}\| \lesssim \|Z_t^{N,\alpha}\|, \text{ and } \|V_t^{N,\alpha}\| \lesssim N^{-\alpha} \|Z_t^{N,\alpha}\|^2.$$

We deduce from the last three inequalities

$$\left\| V_t^{N,\alpha} \right\| \lesssim N^{\alpha-1} + \left(\|z\| + \sup_{0\le s\le t} \|Y_s^{N,\alpha}\| \right) \wedge N^{-\alpha} \left(\|z\| + \sup_{0\le s\le t} \|Y_s^{N,\alpha}\| \right)^2. \tag{4.3}$$

We now define

$$\widetilde{Y}_t^{N,\alpha} = \int_0^t V_s^{N,\alpha} ds + Y_t^{N,\alpha}, \ t \ge 0,$$

so that

$$Z_t^{N,\alpha} = N^\alpha \left(\{z^* + N^{-\alpha} z\}_N - z^* \right) + \int_0^t \nabla b(0) Z_s^{N,\alpha} ds + \widetilde{Y}_t^{N,\alpha}. \tag{4.4}$$

We will see below that the large deviations of $Z^{N,\alpha}$ will follow from those of $\widetilde{Y}^{N,\alpha}$ by a variant of the contraction principle. We first consider the simpler processes

$$\overline{Y}_t^N := \frac{1}{N} \sum_{j=1}^k h_j \int_0^t \int_0^{N\beta_j(0)} \overline{\mathcal{M}}_j(ds, du), \text{ and } \overline{Y}_t^{N,\alpha} = N^\alpha \overline{Y}_t^N \tag{4.5}$$

which are similar to Y^N and $Y^{N,\alpha}$, but with Z_s^N replaced by 0.

4.2 The limiting logarithmic moment generating function of $\overline{Y}^{N,\alpha}$

We note that writing the integral over $[0, N\beta_j(0)]$ as the sum from $\ell = 1$ to $\ell = N$ of integrals over $((\ell-1)\beta_j(0), \ell\beta_j(0)]$, we can rewrite $\overline{Y}^{N,\alpha}$ as follows.

$$\overline{Y}_t^{N,\alpha} = \frac{1}{N^{1-\alpha}} \sum_{\ell=1}^{N} Q_\ell(t), \quad \text{where}$$
$$Q_\ell(t) = \sum_{j=1}^{k} h_j \int_0^t \int_{(\ell-1)\beta_j(0)}^{\ell\beta_j(0)} \overline{\mathcal{M}}_j(ds, du).$$

The processes $Q_1, Q_2, \ldots, Q_N$ are i.i.d., and their law is that of

$$Q(t) = Q_1(t) = \sum_{j=1}^{k} h_j \int_0^t \int_0^{\beta_j(0)} \overline{\mathcal{M}}_j(ds, du). \tag{4.6}$$

Now let $\nu = (\nu_1, \ldots, \nu_d)$ be a vector of signed measures on $[0, T]$.

Lemma 4.1. *As $N \to \infty$,*

$$a_N \log \mathbb{E} \exp\left\{ a_N^{-1} \nu(\overline{Y}^{N,\alpha}) \right\} \to \frac{1}{2} \mathbb{E}\left(\nu(Q)^2\right).$$

Proof. We use in an essential way the above decomposition of $\overline{Y}^{N,\alpha}$.

$$\begin{aligned}
a_N \Lambda_N(a_N^{-1}\nu) &= N^{2\alpha-1} \log \mathbb{E} \exp\{N^{1-\alpha} \nu(\overline{Y}^N)\} \\
&= N^{2\alpha} \log \mathbb{E} \exp\{N^{-\alpha} \nu(Q)\} \\
&= N^{2\alpha} \log \mathbb{E}\{1 + N^{-\alpha}\nu(Q) + \frac{N^{-2\alpha}}{2}\nu(Q)^2 + N^{-3\alpha} R_3^N\} \\
&= N^{2\alpha} \log \left\{ 1 + \frac{N^{-2\alpha}}{2} \mathbb{E}[\nu(Q)^2] + N^{-3\alpha} \mathbb{E}[R_3^N] \right\} \\
&\to \frac{1}{2} \mathbb{E}[\nu(Q)^2],
\end{aligned}$$

provided

$$\sup_{N \geq 1} |\mathbb{E}[R_3^N]| < \infty, \tag{4.7}$$

which we will check below. From this it follows that the argument of the logarithm on the before last line is greater than or equal to 1, at least for N large enough, and the final conclusion follows easily from the fact that for any $x \geq 0$, $x - x^2/2 \leq \log(1+x) \leq x$. Let us now check (4.7). It follows from an exact Taylor formula that

$$|R_3^N| \leq \frac{|\nu(Q)|^3}{6} \exp(N^{-\alpha} |\nu(Q|).$$

But $\nu(Q)$ is an affine combination of mutually independent Poisson random variables, so that (4.7) follows easily by an explicit computation. □

4.3 The limiting logarithmic moment generating function of $\widetilde{Y}^{N,\alpha}$

We want to study the large deviations of $\widetilde{Y}^{N,\alpha}$. The main step will be to prove that Lemma 4.1 remains valid if we replace $\overline{Y}^{N,\alpha}$ by $\widetilde{Y}^{N,\alpha}$, which will follow from the next Proposition.

Proposition 4.2. *For any $C > 0$, $\nu = (\nu_1, \ldots, \nu_d)$ a vector of signed measures, as $N \to \infty$,*

$$a_N \log \mathbb{E} \exp\left[C a_N^{-1} \nu(\widetilde{Y}^{N,\alpha} - \overline{Y}^{N,\alpha}) \right] \to 0\,. \tag{4.8}$$

Before we establish that Proposition, let us first prove that it yields the wished result.

Proposition 4.3. *Given Lemma 4.1, if Proposition 4.2 holds true, then for any signed measure ν on $[0,T]$, as $N \to \infty$,*

$$a_N \log \mathbb{E} \exp\left\{ a_N^{-1} \nu(\widetilde{Y}^{N,\alpha}) \right\} \to \frac{1}{2} \mathbb{E}[\nu(Q)^2] \,.$$

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

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

so that, if we combine Lemma 4.1 and Proposition 4.2, we deduce that

$$\limsup_N a_N \log \mathbb{E} \exp\{ \nu(a_N^{-1} \widetilde{Y}^{N,\alpha}) \} \leq \frac{(1+\delta)}{2} \mathbb{E}[\nu(Q)^2],$$

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

$$\limsup_N a_N \log \mathbb{E} \exp\{ \nu(a_N^{-1} \widetilde{Y}^{N,\alpha}) \} \leq \frac{1}{2} \mathbb{E}[\nu(Q)^2].$$

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

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

with $\mu = -\nu$, which implies that

$$\liminf_N a_N \log \mathbb{E} \exp\{ a_N^{-1} \nu(\widetilde{Y}^{N,\alpha}) \} \geq \frac{1}{2(1+\delta)} \mathbb{E}[\nu(Q)^2],$$

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

$$\liminf_N a_N \log \mathbb{E} \exp\{ a_N^{-1} \nu(Y^{N,\alpha}) \} \geq \frac{1}{2} \mathbb{E}[\nu(Q)^2].$$

□

The remaining of this subsection will be devoted to the proof of Proposition 4.2.

We note that Proposition 4.2 is a consequence of the following two Propositions.

Proposition 4.4. *For any $C > 0$, as $N \to \infty$,*

$$a_N \log \mathbb{E} \exp\left[C a_N^{-1} \nu(Y^{N,\alpha} - \overline{Y}^{N,\alpha}) \right] \to 0 \,. \tag{4.9}$$

Proposition 4.5. *For any $C > 0$, as $N \to \infty$,*

$$a_N \log \mathbb{E} \exp\left[C a_N^{-1} \sup_{0 \leq t \leq T} \left\| \int_0^t V_s^{N,\alpha} ds \right\| \right] \to 0 \,. \tag{4.10}$$

We start with the

Proof of Proposition 4.4. The exponents in the expressions entering (4.9) are sums over the indices $1 \leq i \leq d$ and $1 \leq j \leq k$. Using repeatedly Schwartz's inequality, it is sufficient to prove the results with the sum replaced by each of the summands. Therefore in this proof we do as if $d = 1$, we fix $1 \leq j \leq k$ and for the sake of simplifying the notations, we drop the index j. We note that

$$a_N^{-1}(Y^{N,\alpha} - \overline{Y}^{N,\alpha}) = N^{-\alpha} \int_0^t \int_{N\beta(0)}^{N[\beta(Z_s^N)\vee\beta(0)]} \overline{\mathcal{M}}(ds, du) - N^{-\alpha} \int_0^t \int_{N\beta(Z_s^N)}^{N[\beta(Z_s^N)\vee\beta(0)]} \overline{\mathcal{M}}(ds, du)$$

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 standard Poisson process $M(t)$ on $\mathbb{R}_+$ such that the factor of $N^{-\alpha}$ in this first term can be rewritten as

$$W_t^N := M\left(N \int_0^t (\beta(Z_s^N) - \beta(0))^+ ds\right).$$

We need to estimate $\mathbb{E}\exp[CN^{-\alpha}\nu(W^N)]$. 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

$$\exp\left[CN^{-\alpha} \int_0^T W_t^N \nu_+(dt)\right] \leq \exp\left[CN^{-\alpha} \sup_{0\leq t\leq T} W_t^N\right].$$

We choose a new parameter $0 < \gamma < \alpha$, and we write the expression whose expectation needs to be estimated as a sum of two terms as follows.

$$\begin{aligned} &\exp\left\{CN^{-\alpha} \sup_{0\leq t\leq T} W_t^N\right\} \\ &\quad = \exp\left\{CN^{-\alpha} \sup_{0\leq t\leq T} W_t^N\right\} \mathbf{1}_{\sup_{0\leq t\leq T}\|Z_t^N\|\leq N^{-\gamma}} \\ &\qquad + \exp\left\{CN^{-\alpha} \sup_{0\leq t\leq T} W_t^N\right\} \mathbf{1}_{\sup_{0\leq t\leq T}\|Z_t^N\|> N^{-\gamma}}. \end{aligned} \tag{4.11}$$

We now estimate the first term on the right hand side of (4.11). For that sake, we define the stopping time

$$\sigma_N = \inf\{0 \leq t \leq T;\ \|Z_t^N\| > N^{-\gamma}\}$$

and note that

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

Consequently the expectation of the first term on the right of (4.11) is bounded from

above by

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

where the first inequality follows from Proposition 5.1 in the Appendix below, and the second one exploits the Lipschitz property of β. Consider now the second term on the right hand side of (4.11).

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

for some $c, C>0$, where the second inequality follows from Proposition 5.1 and the boundedness of β. Estimating the second factor in the last expression amounts to estimating the two probabilities (with another $c>0$)

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le T}M\left(N\int_0^t\beta(Z_s^N)ds\right)>cN^{1-\gamma}\right)\quad\text{and}\\
&\mathbb{P}\left(\sup_{0\le t\le T}\left(-M\left(N\int_0^t\beta(Z_s^N)ds\right)\right)>cN^{1-\gamma}\right).
\end{aligned}\tag{4.12}$$

We estimate the first probability. For any $a>0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le s\le t}M\left(N\int_0^s\beta(Z_r^N)dr\right)>cN^{1-\gamma}\right)\\
&\quad=\mathbb{P}\left(\sup_{0\le s\le t}\exp\left\{aM\left(N\int_0^s\beta(Z_r^N)dr\right)\right\}>\exp\{acN^{1-\gamma}\}\right)\\
&\quad\le e^{-acN^{1-\gamma}}\mathbb{E}\left(\sup_{0\le s\le t}\exp\left\{aM\left(N\int_0^s\beta(Z_r^N)dr\right)\right\}\right)\\
&\quad\lesssim e^{-acN^{1-\gamma}}\left(\mathbb{E}\exp\left\{(e^{2a}-1-2a)N\int_0^t\beta(Z_s^N)ds\right\}\right)^{1/2}\\
&\quad\le\exp\{-acN^{1-\gamma}+(e^{2a}-1-2a)NCt\}\\
&\quad\le\exp\left\{-\frac{c^2}{8Ct}N^{1-2\gamma}\right\},
\end{aligned}\tag{4.13}$$

where the second inequality follows from Proposition 5.1 and the last inequality by optimizing over $a>0$. One can easily convince oneself that a similar result holds for the second line of (4.12), making use of Proposition 5.1 with a negative a. Note also for further use that the same result also holds in case $\gamma=0$. In that case, the probability on

the second line of (4.12) is zero for large enough c, in which case the anounced estimate is of course true.

The expectation of the second term of the right hand side of (4.11) is thus dominated by (with c_1 and c_2 two positive constants)

$$\exp\{c_1 N^{1-2\alpha} - c_2 N^{1-2\gamma}\} \to 0, \quad \text{as } N \to \infty.$$

Finally

$$\begin{aligned}\mathbb{E}\exp\left\{N^{-\alpha}\sup_{0\le s\le t} M\left(N\int_0^t (\beta(Z_s^N)-\beta(z_s))^+ ds\right)\right\}\\ \le \exp\left\{CN^{1-2\alpha-\gamma}\right\} + \exp\{c_1 N^{1-2\alpha} - c_2 N^{1-2\gamma}\}\end{aligned}$$

It follows readily from the inequality $\log(a+b) \le \log(2) + \log(a\vee b)$ that for N large enough

$$a_N \log \mathbb{E}\exp\left\{N^{-\alpha}\sup_{0\le s\le t} M\left(N\int_0^t (\beta(Z_s^N)-\beta(0))^+ ds\right)\right\} \le a_N\log(2) + CN^{-\gamma},$$

which establishes (4.9). □

We now turn to the second proof.

Proof of Proposition 4.5. Recalling assumption $(H.1)$, we now define, with $\overline{\beta}_j := \sup_{z\in\mathbb{R}^d}\beta_j(z)$,

$$\xi_t^{N,j} := \frac{1}{N}\int_0^t\int_0^{N\overline{\beta}_j} \mathcal{M}_j(ds,du)\, 1\le j\le k,$$

the event

$$A_b^N := \left\{\sup_{0\le t\le T}\left\|\overline{Y}_t^N\right\| \le b\right\}\bigcap\bigcap_{j=1}^k\left\{\sup_{0\le t\le T}\xi_t^{N,j} \le (1+b')\overline{\beta}_j T\right\},$$

and the stopping time

$$\bar{\tau}_b := \inf\left\{t>0,\, \left\|\overline{Y}_t^N\right\| > b\right\}\bigwedge\bigwedge_{j=1}^k \inf\left\{t>0,\, \xi_t^{N,j} > (1+b')\overline{\beta}_j T\right\},$$

where the constant $b>0$ will be chosen below, and the constant $b'>0$ is arbitrary. From the estimate (4.3),

$$\begin{aligned}a_N\log\mathbb{E}\left[\exp\left\{a_N^{-1}C\sup_{0\le t\le T}\left\|\int_0^t V_s^{N,\alpha}ds\right\|\right\}\right]\\ \lesssim N^{\alpha-1} + a_N\log\mathbb{E}\left[\exp\left\{a_N^{-1}C\left(\|z\| + N^\alpha\sup_{0\le t\le T}\|Y_t^N\|\right)\mathbf{1}_{(A_b^N)^c}\right\}\right] \qquad (4.14)\\ + a_N\log\mathbb{E}\left[\exp\left\{a_N^{-1}C\left(N^{-\alpha}\|z\|^2 + N^\alpha\sup_{0\le t\le T}\|Y_s^N\|^2\right)\mathbf{1}_{A_b^N}\right\}\right], \qquad (4.15)\end{aligned}$$

We take the limit successively in the two terms of the above right hand side. Step 1 : Estimate of (4.14) We have

$$\begin{aligned}\mathbb{E}\left[\exp\left\{a_N^{-1}C\left(\|z\| + N^\alpha\sup_{0\le t\le T}\|Y_t^N\|\right)\mathbf{1}_{(A_b^N)^c}\right\}\right]\\ \lesssim \exp\left\{CN^{1-2\alpha}\|z\|\right\}\mathbb{P}\left((A_b^N)^c\right) + \mathbb{E}\left[\exp\left\{CN^{-\alpha}\sup_{0\le t\le T}\|NY_s^N\|\right\}\mathbf{1}_{(A_b^N)^c}\right] + 1,\end{aligned}$$

We first note that the arguments used in the proof of (4.13), in the particular case $\gamma = 0$, yield

$$\begin{aligned}\mathbb{P}((A_b^N)^c) &\leq \mathbb{P}\left(\sup_{0\leq t\leq T}\left\|\overline{Y}_t^N\right\| > b\right) + \sum_{j=1}^{k}\mathbb{P}\left(\sup_{0\leq t\leq T}(\xi_t^{N,j} - \bar{\beta}t) > b'\bar{\beta}T\right)\\ &\lesssim e^{-CN}\,, \end{aligned} \tag{4.16}$$

for some constant $C > 0$. We next estimate the product

$$\mathbb{E}\left[\exp\left\{CN^{-\alpha}\sup_{0\leq t\leq T}\|NY_s^N\|\right\}\right]\mathbb{P}\left((A_b^N)^c\right).$$

For the same reason as in the previous proof, we need only consider the case $d = k = 1$. It follows from Proposition 5.1 that the first factor satisfies

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

Finally there exist two positive constants C_1 and C_2 such that

$$\begin{aligned}\mathbb{E}\left[\exp\left\{a_N^{-1}C\left(\|z\| + N^{\alpha}\sup_{0\leq t\leq T}\|Y_t^N\|\right)\mathbf{1}_{(A_b^N)^c}\right\}\right] &\lesssim 1 + \exp\{C_1N^{1-2\alpha} - C_2N\}\\ &\leq 2\,,\end{aligned}$$

for N large enough. So $a_N \log$ of the above tends to 0, as $N \to \infty$.
Step 2 : Estimate of (4.15) We first note that

$$\begin{aligned}&a_N \log\mathbb{E}\left[\exp\left\{a_N^{-1}C\left(N^{-\alpha}\|z\|^2 + N^{\alpha}\sup_{0\leq t\leq T}\|Y_s^N\|^2\right)\mathbf{1}_{A_b^N}\right\}\right]\\ &\quad\leq CN^{-\alpha}\|z\|^2 + a_N\log\mathbb{E}\left[\exp\left\{a_N^{-1}CN^{\alpha}\sup_{0\leq t\leq T}\|Y_s^N\|^2\mathbf{1}_{A_b^N}\right\}\right].\end{aligned}$$

The first term on the right tends to 0 as $N \to \infty$. It remains to take care of the second term. Since Y_t^N is a martingale, it is clear that the process

$$\left\{\exp\left(a_N^{-1}\frac{C}{2}N^{\alpha}\|Y_t^N\|^2\right),\ t \geq 0\right\}$$

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

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq t\leq T}\exp\left\{a_N^{-1}CN^{\alpha}\|Y_t^N\|^2\mathbf{1}_{A_b^N}\right\}\right] &\leq \mathbb{E}\left[\sup_{0\leq t\leq T\wedge\bar{\tau}_b}\exp\left\{a_N^{-1}CN^{\alpha}\|Y_t^N\|^2\right\}\right]\\ &\leq 4\mathbb{E}\left[\exp\left\{a_N^{-1}CN^{\alpha}\|Y_{T\wedge\bar{\tau}_b}^N\|^2\right\}\right]\end{aligned}$$

Next

$$\begin{aligned}\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\|Y_{T\wedge\bar{\tau}_b}^N\|^2\right\}\right] &\leq \sqrt{\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\left(\|Y_{T\wedge\bar{\tau}_b}^N\|^2 - \|\overline{Y}_{T\wedge\bar{\tau}_b}^N\|^2\right)\right\}\right]}\\ &\quad\times\sqrt{\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\|\overline{Y}_{T\wedge\bar{\tau}_b}^N\|^2\right\}\right]}\end{aligned} \tag{4.17}$$

Consider first the first factor on the right hand side of (4.17). We deduce from the definition of $\bar{\tau}_b$ that

$$\begin{aligned}\|Y_{T\wedge\bar{\tau}_b}^N\|^2 - \|\overline{Y}_{T\wedge\bar{\tau}_b}^N\|^2 &= \left(Y_{T\wedge\bar{\tau}_b}^N + \overline{Y}_{T\wedge\bar{\tau}_b}^N, Y_{T\wedge\bar{\tau}_b}^N - \overline{Y}_{T\wedge\bar{\tau}_b}^N\right)\\ &\leq (c_T + b + 2N^{-1}\sup_j\|h_j\|)\left\|Y_{T\wedge\bar{\tau}_b}^N - \overline{Y}_{T\wedge\bar{\tau}_b}^N\right\|,\end{aligned}$$

with $c_T = \sum_{j=1}^k \|h_j\|(1+b')\overline{\beta}_j T$. Consequently the square of the first factor on the right of (4.17) is bounded from above by

$$\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\left\|Y^N_{T\wedge\bar\tau_b} - \overline{Y}^N_{T\wedge\bar\tau_b}\right\|\right\}\right] \le \mathbb{E}\left[\exp\left\{CN^{1-\alpha}\left\|Y^N_T - \overline{Y}^N_T\right\|\right\}\right],$$

where we have used Doob's optional sampling theorem for submartingales. From the same argument as above,we do as if $d=1$, note that

$$\begin{aligned}\exp\left\{CN^{1-\alpha}\left|Y^N_T - \overline{Y}^N_T\right|\right\} \le \exp\left\{CN^{1-\alpha}\left(Y^N_T - \overline{Y}^N_T\right)\right\}\\ + \exp\left\{CN^{1-\alpha}\left(\overline{Y}^N_T - Y^N_T\right)\right\}\end{aligned}$$

and exploit Proposition 4.4 in order to conclude concerning $a_N \log$ of the first factor on the right of (4.17).

We next note that

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

Hence the square of the second term on the right of (4.17) satisfies

$$\begin{aligned}\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\left\|\overline{Y}^N_{T\wedge\bar\tau_b}\right\|^2\right\}\right] \le \sqrt{\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\mathbf{1}_{\{\bar\tau_b<T\}}\right\}\right]}\\ \times\sqrt{\mathbb{E}\left[\exp\left\{CN^{1-\alpha}\left\|\overline{Y}^N_T\right\|^2\mathbf{1}_{\left\{\left\|\overline{Y}^N_T\right\|\le b\right\}}\right\}\right]}\end{aligned} \tag{4.18}$$

Consider first the second factor on the right of (4.18). We have

$$\left\|\overline{Y}^N_T\right\|^2 \le k\sum_{j=1}^k \frac{\|h_j\|^2}{N^2}\left|\int_0^T\int_0^{\beta_j(0)}\overline{\mathcal{M}}_j(ds,du)\right|^2.$$

Using the Cauchy–Schwartz inequality several times, it is clear that it is sufficient to do as if we had (dropping the index j)

$$\overline{Y}^N_T = \frac{1}{N}\int_0^T\int_0^{N\beta(0)}\overline{\mathcal{M}}(ds,du) = \sqrt{a/N}\,\xi_N,$$

with $a=\beta(0)\,T$ and $\xi_N = \frac{\theta_N - aN}{\sqrt{aN}}$, where $\theta_N \sim \text{Poi}(aN)$. We now choose $b=a/3$. We have

$$
\begin{aligned}
&\mathbb{E}\exp\left\{CN^{-\alpha}|\xi_N|^2\mathbf{1}_{\{|\xi_N|\le\sqrt{aN}/3\}}\right\}\\
&\quad=\sum_{k=\lceil 2aN/3\rceil}^{\lfloor 4aN/3\rfloor}\exp\left\{CN^{-\alpha}\frac{(k-aN)^2}{aN}\right\}e^{-aN}\frac{(aN)^k}{k!}\\
&\quad\lesssim\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2\right\}e^{-aN}\frac{(aN)^{aN+x\sqrt{aN}}}{(aN+x\sqrt{aN})!}\sqrt{aN}dx\\
&\quad\lesssim\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2\right\}e^{x\sqrt{aN}}\left(1+\frac{x}{\sqrt{aN}}\right)^{-(aN+x\sqrt{aN})}dx\\
&\quad\le\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2\right\}\\
&\qquad\times\exp\left\{x\sqrt{aN}-(aN+x\sqrt{aN})\left[\frac{x}{\sqrt{aN}}-\frac{x^2}{2aN}+\frac{x^3}{2(aN)^{3/2}}\mathbf{1}_{x<0}\right]\right\}dx\\
&\quad\le\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2-\frac{x^2}{2}+\frac{x^3}{2\sqrt{aN}}\mathbf{1}_{x>0}\right\}dx\\
&\quad\le\frac{1}{\sqrt{2\pi}}\int_{-\sqrt{aN}/3}^{\sqrt{aN}/3}\exp\left\{CN^{-\alpha}x^2-\frac{x^2}{3}\right\}dx
\end{aligned}
$$

We have proved that the second factor on the right of (4.18) remains bounded, as $N\to\infty$. We next consider the first factor on the right of (4.18). We first note that

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

But from (4.16), $\mathbb{P}\left(\bar{\tau}_b<T\right)\lesssim e^{-CN}$.

It follows that the left hand side of (4.18) is bounded from above by a constant times

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

where C_1 and C_2 are two positive constants. This last expression is bounded by 2, as soon as N is large enough. Finally $a_N\log$ of the left-hand side of (4.18) tends to 0, as $N\to\infty$. □

Remark 4.6. We note that the full strength of (4.3) is necessary for the proof of Proposition 4.5. Indeed, while $a_N\log\mathbb{E}\exp\{CN^{1-\alpha}\sup_{0\le t\le T}\|Y_t^N\|\}$ certainly does not converge to 0 as $N\to\infty$, clearly with high probability $\|Y_t^N\|^2$ is smaller than $\|Y_t^N\|$, but $\mathbb{E}\exp\{CN^{1-\alpha}\|Y_t^N\|^2\}=\infty$.

4.4 Large deviations of $\widetilde{Y}^{N,\alpha}$

We first define the Fenchel–Legendre transform of

$$
\begin{aligned}
\Lambda(\nu)&=\frac{1}{2}\mathbb{E}[\nu(Q)^2]\\
&=\sum_{j=1}^{k}\frac{\beta_j(0)}{2}\int_{[0,T]^2}s\wedge t\,\langle h_j,\nu\rangle(ds)\langle h_j,\nu\rangle(dt),
\end{aligned}
$$

where Q has been defined by (4.6), $\nu=(\nu_1,\dots,\nu_d)$ is a vector of signed measures and $\langle h_j,\nu\rangle(dt)=\sum_{i=1}^d h_j^i\nu_i(dt)$, h_j^i being the i–th coordinate of the vector h_j. We have

exploited the fact that $\nu(Q)$ is the sum over j of zero mean mutually independent random variables. For each $\phi \in D([0,T];\mathbb{R}^d)$, we define

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

The next step will consist in proving that the sequence of processes $\{\widetilde{Y}^{N,\alpha}\}_{N\geq 1}$ satisfies a Large Deviation Principle.

Theorem 4.7. *The sequence $\{\widetilde{Y}^{N,\alpha},\, N \geq 1\}$ satisfies the Large Deviation Principle in $D([0,T];\mathbb{R}^d)$ equipped with the supnorm topology, with the convex, good rate function Λ^* and with speed a_N, in the sense that for any Borel subset $\Gamma \subset D([0,T];\mathbb{R}^d)$,*

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

Since there is a difficulty with having a topology on $D([0,T];\mathbb{R}^d)$ which makes it a topological vector space, and allows for a simple characterization of the class of compact sets, we shall use a small detour for the proof of the above Theorem. Recall that

$$\widetilde{Y}_t^{N,\alpha} = Y_t^{N,\alpha} + \int_0^t V_s^{N,\alpha} ds,$$

where $Y_t^{N,\alpha}$ is piecewise constant, with jumps of size $h_j N^{\alpha-1}$. Let $Y_t^{N,\alpha,c}$ denote the continuous piecewise linear approximation of $Y_t^{N,\alpha}$, which is defined as follows. Let $0 = \tau_0^N < \tau_1^N < \tau_2^N < \cdots$ denote the successive jump times of the process $Y_t^{N,\alpha}$. For $i \geq 0$, on the interval $[\tau_i^N, \tau_{i+1}^N]$,

$$Y_t^{N,\alpha,c} = \frac{\tau_{i+1}^N - t}{\tau_{i+1}^N - \tau_i^N} Y_{\tau_i^N}^{N,\alpha} + \frac{t - \tau_i^N}{\tau_{i+1}^N - \tau_i^N} Y_{\tau_{i+1}^N}^{N,\alpha}\,.$$

Next we define

$$\widetilde{\widetilde{Y}}_t^{N,\alpha} = Y_t^{N,\alpha,c} + \int_0^t V_s^{N,\alpha} ds\,.$$

We note that

$$\sup_{0\leq t\leq T} \left\| \widetilde{Y}_t^{N,\alpha} - \widetilde{\widetilde{Y}}_t^{N,\alpha} \right\| \leq \sup_j \|h_j\| N^{\alpha-1}, \tag{4.19}$$

hence for any $\delta > 0$, for N large enough,

$$\mathbb{P}\left(\sup_{0\leq t\leq T} \left\| \widetilde{Y}_t^{N,\alpha} - \widetilde{\widetilde{Y}}_t^{N,\alpha} \right\| \geq \delta \right) = 0\,.$$

This implies clearly

Lemma 4.8. *The two sequences $\{\widetilde{Y}^{N,\alpha}\}_{N\geq 1}$ and $\left\{\widetilde{\widetilde{Y}}^{N,\alpha}\right\}_{N\geq 1}$ are exponentially equivalent in $D([0,T];\mathbb{R}^d)$, equipped with the supnorm topology, in the sense that for each $\delta > 0$,*

$$\limsup_{N\to\infty} a_N \log \mathbb{P}\left(\sup_{0\leq t\leq T} \left\| \widetilde{Y}_t^{N,\alpha} - \widetilde{\widetilde{Y}}_t^{N,\alpha} \right\| \geq \delta \right) = -\infty.$$

We shall prove below the following.

Proposition 4.9. *The sequence* $\left\{\widetilde{\widetilde{Y}}^{N,\alpha}\right\}_{N\geq 1}$ *is exponentially tight in* $C_0([0,T];\mathbb{R}^d)$*, the space of continuous functions from* $[0,T]$ *into* $\mathbb{R}^d$*, which start from* 0 *at* $t=0$*, in the sense that for any* $R>0$*, there exists a compact subset* $K_R \subset\subset C_0([0,T]:\mathbb{R}^d)$ *such that*

$$\limsup_{N\to\infty} a_N \log \mathbb{P}\left(\widetilde{\widetilde{Y}}^{N,\alpha} \notin K_R\right) \leq -R.$$

Let us now turn to the proof of the above Theorem.

Proof of Theorem 4.7. From (4.19), we deduce that

$$\begin{aligned} a_N \log \mathbb{E}\exp\left\{Ca_N^{-1}\nu\left(\widetilde{Y}_t^{N,\alpha} - \widetilde{\widetilde{Y}}_t^{N,\alpha}\right)\right\} \\ \leq CN^{\alpha-1} \to 0, \end{aligned}$$

as $N\to\infty$. Consequently, again by the argument of Proposition 4.3, we deduce from that same Proposition that for any signed measure ν on $[0,T]$, as $N\to\infty$,

$$a_N \log \mathbb{E}\exp\left\{a_N^{-1}\nu\left(\widetilde{\widetilde{Y}}^{N,\alpha}\right)\right\} \to \frac{1}{2}\mathbb{E}[\nu(Q)^2]\,.$$

This, together with Proposition 4.9, allows us to apply Corollary 4.6.14 from [2], to conclude that the sequence $\left\{\widetilde{\widetilde{Y}}^{N,\alpha}\right\}_{N\geq 1}$ satisfies a LDP in $C_0([0,T];\mathbb{R}^d)$ with the good rate function Λ^*, and speed a_N. Since $C_0([0,T];\mathbb{R}^d)$ is closed in $D([0,T];\mathbb{R}^d)$ equipped with the supnorm topology, it follows from Lemma 4.1.5 in [2] that the same LDP holds in the latter space, with the same rate function Λ^*, extended to that space by $\Lambda^*(\phi)=+\infty$ for $\phi\in D([0,T];\mathbb{R}^d)\backslash C_0([0,T];\mathbb{R}^d)$. The result now follows from Lemma 4.8, in view of Theorem 4.2.13 from [2]. □

We now turn to the

Proof of Proposition 4.9. Clearly it suffices to prove both that

$$\lim_{R\to\infty}\limsup_{N\to\infty} a_N \log \mathbb{P}\left(\sup_{0\leq t\leq T}\|V_t^{N,\alpha}\|\geq R\right) = -\infty\,, \tag{4.20}$$

and that the sequence $\{Y^{N,\alpha,c}\}_{N\geq 1}$ is exponentially tight in $C_0([0,T];\mathbb{R}^d)$. Let us first establish (4.20). It follows from (4.3) that

$$\|V_t^{N,\alpha}\| \leq C\left(\|z\| + 1 + \sup_{0\leq t\leq T}\|Y_t^{N,\alpha}\|\right).$$

Consequently, if $R > 2C(\|z\|+1)$, with $R' = (2C)^{-1}R$,

$$\begin{aligned} a_N \log \mathbb{P}(\sup_{0\leq t\leq T}\|V_t^{N,\alpha}\| > R) &\leq a_N \log \mathbb{P}(\sup_{0\leq t\leq T}\|Y_t^{N,\alpha}\| > R') \\ &\leq a_N \log\left(e^{-a_N^{-1}R'}\mathbb{E}\sup_{0\leq t\leq T}\exp\{a_N^{-1}\|Y_t^{N,\alpha}\|\}\right) \\ &\leq -R' + a_N \log \mathbb{E}\left(\sup_{0\leq t\leq T}\exp\{a_N^{-1}\|Y_t^{N,\alpha}\|\}\right). \end{aligned}$$

It follows from Doob's submartingale inequality and a combination of Lemma 4.1 and Proposition 4.4 that the $\limsup$ as $N\to\infty$ of the second term of the last right hand side is finite. (4.20) clearly follows.

It remains to consider $Y^{N,\alpha,c}$. Define the modulus of continuity of an element $x \in C_0([0,T];\mathbb{R}^d)$ as $w_x(\delta) = \sup_{0\le s,t\le T,|s-t|\le\delta} \|x(t)-x(s)\|$. It follows from Ascoli's theorem that for any sequence $\{\delta_\ell, \ell \ge 1\}$ of positive numbers, the following is a compact subset of $C_0([0,T];\mathbb{R}^d)$:

$$\bigcap_{\ell\ge 1}\{x:\ w_x(\delta_\ell) \le \ell^{-1}\}\,.$$

Suppose that for each $\ell \ge 1$, $R > 0$, we can find $\delta_{R,\ell} > 0$ such that for all $N \ge 1$,

$$\mathbb{P}(w_{Y^{N,\alpha,c}}(\delta_{R,\ell}) \ge \ell^{-1}) \le \exp\{-a_N^{-1}(R+\ell)\}\,. \tag{4.21}$$

From this we deduce that

$$\mathbb{P}(\cup_{\ell\ge1}\{w_{Y^{N,\alpha,c}}(\delta_{R,\ell}) \ge \ell^{-1}\}) \le \sum_{\ell\ge1} e^{-a_N^{-1}(R+\ell)} \le e^{-a_N^{-1}R},$$

so that

$$\limsup_{N\to\infty} a_N \log \mathbb{P}(\cup_{\ell\ge1}\{w_{Y^{N,\alpha,c}}(\delta_{R,\ell}) \ge \ell^{-1}\}) \le -R,$$

from which the result follows. A sufficient condition for (4.21) to be true is that for any $b > 0$,

$$\lim_{\delta\to0}\limsup_{N\to\infty} a_N \log \mathbb{P}\left(w_{Y^{N,\alpha,c}}(\delta) > b\right) = -\infty\,.$$

In turn a sufficient condition for this is that

$$\lim_{\delta\to0}\limsup_{N\to\infty} a_N \log \mathbb{P}\left(w_{Y^{N,\alpha}}(\delta) > b\right) = -\infty\,, \tag{4.22}$$

which we now prove. It is not hard to see that

$$\begin{aligned}
\mathbb{P}\left(w_{Y^{N,\alpha}}(\delta) > b\right) &\le 2\left(\frac{T}{\delta}+1\right)\sup_{0\le t\le T}\mathbb{P}\left(\sup_{t\le s\le t+2\delta}\|Y_s^{N,\alpha}-Y_t^{N,\alpha}\| \ge b/2\right)\\
&\le 2\left(\frac{T}{\delta}+1\right)\sup_{0\le t\le T}\mathbb{P}\left(\sup_{t\le s\le t+2\delta}\exp\{a_N^{-1}\delta^{-1/2}\|Y_s^{N,\alpha}-Y_t^{N,\alpha}\|\} \ge \exp\{ba_N^{-1}/2\sqrt{\delta}\}\right)\\
&\le 2\left(\frac{T}{\delta}+1\right)\exp\{-ba_N^{-1}/2\sqrt{\delta}\}\sup_{0\le t\le T}\mathbb{E}\exp\{a_N^{-1}\delta^{-1/2}\|Y_{t+2\delta}^{N,\alpha}-Y_t^{N,\alpha}\|\},
\end{aligned}$$

where we have used Doob's submartingale inequality at the last step. Clearly

$$\exp\{a_N^{-1}\delta^{-1/2}\|Y_{t+2\delta}^{N,\alpha}-Y_t^{N,\alpha}\|\} \le \prod_{j=1}^{k}\exp\left\{N^{-\alpha}\delta^{-1/2}\|h_j\|\left|\int_t^{t+2\delta}\int_0^{N\beta_j(Z_s^N)}\overline{\mathcal{M}}_j(ds,du)\right|\right\}$$

Using repeatedly Cauchy–Schwartz's inequality, we see that it suffices to estimate for each j

$$\begin{aligned}
&\mathbb{E}\exp\left\{CN^{-\alpha}\delta^{-1/2}\left|\int_t^{t+2\delta}\int_0^{N\beta_j(Z_s^N)}\overline{\mathcal{M}}_j(ds,du)\right|\right\}\\
&\quad\le \mathbb{E}\exp\left\{CN^{-\alpha}\delta^{-1/2}\int_t^{t+2\delta}\int_0^{N\beta_j(Z_s^N)}\overline{\mathcal{M}}_j(ds,du)\right\}\\
&\quad+\mathbb{E}\exp\left\{-CN^{-\alpha}\delta^{-1/2}\int_t^{t+2\delta}\int_0^{N\beta_j(Z_s^N)}\overline{\mathcal{M}}_j(ds,du)\right\}\\
&\quad\le 2\exp\{8C^2N^{1-2\alpha}\bar\beta_j\},
\end{aligned}$$

where $\bar{\beta}_j = \sup_z \beta_j(z)$, we have used Proposition 5.1 and the inequality $e^x - 1 - x \leq x^2$, valid for $x \leq \log(2)$, which we have applied with $x = 2CN^{-\alpha}\delta^{-1/2}$ and $x = -2CN^{-\alpha}\delta^{-1/2}$ (recall that we will first let $N \to \infty$). Putting together the last estimates yields

$$\limsup_{N\to\infty} a_N \mathbb{P}\left(w_{Y^{N,\alpha}}(\delta) > b\right) \leq -\frac{b}{2\sqrt{\delta}} + C.$$

(4.22) follows, and the Proposition is proved. □

4.5 Computation of the rate function Λ^*

Let us compute Λ^* in the three examples which we discussed above in section 2. Here we do not translate z^* to the origin.

4.5.1 Computation of Λ^* for the SIS model

Recall that in this case $d = 1$, $k = 2$, $h_1 = 1$, $\beta_1(z) = \lambda z(1-z)$, $h_2 = -1$, $\beta_2(z) = \gamma z$. If $\lambda > \gamma$, there is a unique stable endemic equilibrium $z^* = 1 - \gamma/\lambda$. We first compute

$$\Lambda(\nu) = \frac{1}{2}\mathbb{E}\int_{[0,T]\times[0,T]} Q(s)Q(t)\nu(ds)\nu(dt),$$

where

$$Q(t) = \int_0^t \int_0^{\beta_1(z^*)} \overline{\mathcal{M}}_1(ds, du) - \int_0^t \int_0^{\beta_2(z^*)} \overline{\mathcal{M}}_2(ds, du)\,.$$

It is easy to check that $\mathbb{E}[Q(t)Q(s)] = \sigma^2(z^*)\, s \wedge t$, where

$$\sigma^2(z^*) = \beta_1(z^*) + \beta_2(z^*) = 2\frac{\gamma}{\lambda}(\lambda - \gamma).$$

Consequently

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

We now need to compute $\Lambda^*(\phi)$ in case $\phi \in C^2([0,T])$. We should take the supremum over the signed measures ν on $[0,T]$ of the quantity

$$\int_{[0,T]} \phi(t)\nu(dt) - \frac{\sigma^2(z^*)}{2}\int_{[0,T]\times[0,T]} s \wedge t\, \nu(ds)\nu(dt)\,.$$

The supremum is achieved at the signed measure ν which makes the gradient with respect to ν of the above zero, if any. We first note that for such a ν to exist, we need that $\phi(0) = 0$, unless $\Lambda^*(\phi) = +\infty$. Now the optimal ν must satisfy

$$\begin{aligned}\phi(t) &= \sigma^2(z^*)\int_{[0,T]} s \wedge t\, \nu(ds) \\ &= \sigma^2(z^*)\int_{[0,t]} s\, \nu(ds) + \sigma^2(z^*)t\int_{(t,T]} \nu(dt).\end{aligned}$$

So necessarily

$$\nu(dt) = -\frac{\phi''(t)}{\sigma^2(z^*)}dt + \frac{\phi'(T)}{\sigma^2(z^*)}\delta_T(dt).$$

Substituting this signed measure ν in the above formula, we obtain that

$$\begin{aligned}\int_{[0,T]} \phi(t)\nu(dt) &= \frac{\phi'(T)}{\sigma^2(z^*)}\phi(T) - \int_0^T \frac{\phi''}{\sigma^2(z^*)}(t)\phi(t)dt \\ &= \frac{1}{\sigma^2(z^*)}\int_0^T |\phi'(t)|^2 dt\,.\end{aligned}$$

Consequently

$$\Lambda^*(\phi) = \begin{cases} \frac{1}{2\sigma^2(z^*)} \int_0^T |\phi'(t)|^2 dt, & \text{if } \phi(0) = 0 \text{ and } \phi \text{ is absolutely continuous;} \\ +\infty, & \text{otherwise.} \end{cases}$$

4.5.2 Computation of Λ^* for the SIRS model

In this model, $d = 2$ and $k = 3$. We have $h_1 = \binom{1}{-1}$, $\beta_1(z) = \lambda z_1 z_2$, $h_2 = \binom{-1}{0}$, $\beta_2(z) = \gamma z_1$, and $h_3 = \binom{0}{1}$, $\beta_3(z) = \rho(1 - z_1 - z_2)$. In the case $\lambda > \gamma$, there is a unique stable endemic equilibrium, namely $z^* = \binom{\frac{\rho}{\gamma+\rho}\left(1-\frac{\gamma}{\lambda}\right)}{\frac{\gamma}{\lambda}}$. In order to simplify the notations, we shall write $a = \beta_1(z^*)$, $b = \beta_2(z^*)$, $c = \beta_3(z^*)$ and $A = ab + ac + bc$. We have

$$\begin{aligned} \Lambda(\nu) = \frac{a}{2} \int_{[0,T]\times[0,T]} s \wedge t(\nu_1 - \nu_2)(ds)(\nu_1 - \nu_2)(dt) \\ + \frac{b}{2} \int_{[0,T]\times[0,T]} s \wedge t\nu_1(ds)\nu_1(dt) + \frac{c}{2} \int_{[0,T]\times[0,T]} s \wedge t\nu_2(ds)\nu_2(dt) \end{aligned}$$

The functional to be maximized with respect to ν is

$$\begin{aligned} \langle \nu_1, \phi_1 \rangle + \langle \nu_2, \phi_2 \rangle - \frac{a}{2} \int_{[0,T]\times[0,T]} s \wedge t(\nu_1 - \nu_2)(ds)(\nu_1 - \nu_2)(dt) \\ - \frac{b}{2} \int_{[0,T]\times[0,T]} s \wedge t\nu_1(ds)\nu_1(dt) - \frac{c}{2} \int_{[0,T]\times[0,T]} s \wedge t\nu_2(ds)\nu_2(dt) \end{aligned}$$

Writing that the gradient w.r.t. ν_1 and ν_2 of this functional is zero leads to the identities

$$\begin{aligned} \phi_1(t) &= a \int_{[0,T]} s \wedge t(\nu_1 - \nu_2)(ds) + b \int_{[0,T]} s \wedge t\nu_1(ds), \\ \phi_2(t) &= a \int_{[0,T]} s \wedge t(\nu_2 - \nu_1)(ds) + c \int_{[0,T]} s \wedge t\nu_2(ds). \end{aligned}$$

This implies the identities

$$\begin{aligned} \nu_1(dt) &= -\left(\frac{a+c}{A}\phi_1''(t) + \frac{a}{A}\phi_2''(t)\right) dt + \left(\frac{a+c}{A}\phi_1'(T) + \frac{a}{A}\phi_2'(T)\right) \delta_T, \\ \nu_2(dt) &= -\left(\frac{a}{A}\phi_1''(t) + \frac{a+b}{A}\phi_2''(t)\right) dt + \left(\frac{a}{A}\phi_1'(T) + \frac{a+b}{A}\phi_2'(T)\right) \delta_T. \end{aligned}$$

Finally we deduce that $\Lambda^*(\phi)$ is $+\infty$ unless ϕ is absolutely continuous and $\phi(0) = 0$, in which case

$$\Lambda^*(\phi) = \frac{a}{2A} \int_0^T |\phi_1'(t) + \phi_2'(t)|^2 dt + \frac{c}{2A} \int_0^T |\phi_1'(t)|^2 dt + \frac{b}{2A} \int_0^T |\phi_2'(t)|^2 dt \,.$$

4.5.3 Computation of Λ^* for the SIR model with demography

In this case, $d = 2$, $k = 4$, $h_1 = \binom{1}{-1}$, $\beta_1(z) = \lambda z_1 z_2$, $h_2 = \binom{-1}{0}$, $\beta_2(z) = (\gamma + \mu) z_1$, $h_3 = \binom{0}{1}$, $\beta_3(z) = \mu$ and $h_4 = \binom{0}{-1}$, $\beta_4(z) = \mu z_2$. In the case $\lambda > \gamma + \mu$, there is a unique stable endemic equilibrium, namely $z^* = \binom{\mu\left(\frac{1}{\gamma+\mu} - \frac{1}{\lambda}\right)}{\frac{\gamma+\mu}{\lambda}}$. We shall use the notations $a = \beta_1(z^*)$, $b = \beta_2(z^*)$, $c = \beta_3(z^*) + \beta_4(z^*)$ and $A = ab + ac + bc$ We have

$$\begin{aligned} \Lambda(\nu) = \frac{a}{2} \int_{[0,T]\times[0,T]} s \wedge t(\nu_1 - \nu_2)(ds)(\nu_1 - \nu_2)(dt) \\ + \frac{b}{2} \int_{[0,T]\times[0,T]} s \wedge t\nu_1(ds)\nu_1(dt) + \frac{c}{2} \int_{[0,T]\times[0,T]} s \wedge t\nu_2(ds)\nu_2(dt) \end{aligned}$$

Formally the functional $\Lambda(\nu)$ has exactly the same form as in the case of the SIRS model, only the constants have different values. The same computations as in the previous subsection lead to the same result, namely that $\Lambda^*(\phi)$ is $+\infty$ unless ϕ is absolutely continuous and $\phi(0) = 0$, in which case

$$\Lambda^*(\phi) = \frac{a}{2A}\int_0^T |\phi_1'(t) + \phi_2'(t)|^2 dt + \frac{c}{2A}\int_0^T |\phi_1'(t)|^2 dt + \frac{b}{2A}\int_0^T |\phi_2'(t)|^2 dt\,.$$

4.6 Moderate deviations of Z^N

We again equip $D([0,T];\mathbb{R}^d)$ with the supnorm topology. Let for $z \in \mathbb{R}^d$ $F_z : D([0,T];\mathbb{R}^d) \mapsto D([0,T];\mathbb{R}^d)$ be the continuous map which to x associates y solution of the ODE

$$y(t) = z + \int_0^t \nabla b(0) y(s) ds + x(t),$$

and for each $N \geq 1$ $F_{z,N} : D([0,T];\mathbb{R}^d) \mapsto D([0,T];\mathbb{R}^d)$ be the continuous map which to x associates y_N solution of the ODE

$$y_N(t) = N^\alpha(\{z^* + N^{-\alpha} z\}_N - z^*) + \int_0^t \nabla b(0) y_N(s) ds + x(t).$$

We have

$$F_{z,N}(x)(t) - F_z(x)(t) = \exp[\nabla b(0) t]\, \left(N^\alpha(\{z^* + N^{-\alpha} z\}_N - z^*) - z\right), \tag{4.23}$$

which converges to 0 as $N \to \infty$, uniformly in $t \in [0,T]$ and $x \in D([0,T];\mathbb{R}^d)$. We want to study the moderate deviations of Z^N, or in other words the large deviations of $Z^{N,\alpha} = N^\alpha Z^N$. In what follows, we shall denote by $Z_z^{N,\alpha}$ the process $Z^{N,\alpha}$ starting from $Z^{N,\alpha}(0) = z$. Since from (4.4), $Z_z^{N,\alpha}$, which from now on denotes the process $Z^{N,\alpha}$ starting from $Z^{N,\alpha}(0) = z$, is given as $Z_z^{N,\alpha} = F_{z,N}(\widetilde{Y}^{N,\alpha})$, the following statement is a consequence of Theorem 4.7, (4.23) and Corollary 4.2.21 from [2].

Theorem 4.10. *Assume that $(H.1)$ and $(H.2)$ hold. The collection of processes $\{Z_z^{N,\alpha}(t),\ 0 \leq t \leq T\}_{N\geq 1}$ satisfies a large deviations principle with speed a_N and the good rate function*

$$\begin{aligned} I_{z,T}(\phi) &= \Lambda^*(F_z^{-1}(\phi)) \\ &= \begin{cases} \Lambda^*\left(\phi(\cdot) - z - \nabla b(0)\int_0^\cdot \phi(s)ds\right) & \text{if } \phi(0) = z; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

More precisely, for any Borel subset $\Gamma \subset D([0,T];\mathbb{R}^d)$,

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

Since the mapping F_z has the nice property that $F_z(x)(t) - F_{z'}(x)(t) = \exp[\nabla b(0)t](z - z')$, it follows readily again from Corollary 4.2.21 in [2] that the above result can be extended to the following statement.

Theorem 4.11. *Assume that $(H.1)$ and $(H.2)$ hold. For any closed set $F \subset D([0,T];\mathbb{R}^d)$, for any sequence $z_N \to z$,*

$$\limsup_{N\to\infty} a_N \log \mathbb{P}(Z_{z_N}^{N,\alpha} \in F) \leq -\inf_{\phi\in F} I_{z,T}(\phi)\,.$$

For any open set $G \subset D([0,T];\mathbb{R}^d)$, for any sequence $z_N \to z$,

$$\liminf_{N\to\infty} a_N \log \mathbb{P}(Z_{z_N}^{N,\alpha} \in G) \geq -\inf_{\phi\in G} I_{z,T}(\phi)\,.$$

From this last Theorem, we can deduce, with the same proof as that of Corollary 5.6.15 in [2], the following Corollary.

Corollary 4.12. *Assume that $(H.1)$ and $(H.2)$ hold. Let K denote an arbitrary compact subset of $\mathbb{R}^d$.*

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

$$\limsup_{N\to\infty} a_N \log \sup_{z\in K} \mathbb{P}(Z_z^{N,\alpha} \in F) \leq - \inf_{\phi\in F, z\in K} I_{z,T}(\phi)\,.$$

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

$$\liminf_{N\to\infty} a_N \log \inf_{z\in K} \mathbb{P}(Z_z^{N,\alpha} \in G) \geq - \sup_{z\in K} \inf_{\phi\in G} I_{z,T}(\phi)\,.$$

4.7 Wentzell–Freidlin theory and extinction of an epidemic

We now define

$$\begin{aligned} V(z,z',t) &= \inf_{\phi,\, \phi(0)=z, \phi(t)=z'} I_{z,t}(\phi), \\ V(z,z') &= \inf_{t>0} V(z,z',t), \\ \overline{V}_a &= \inf_{z,\, z_1=-a} V(0,z), \end{aligned}$$

where $a>0$, and we recall that we have translated the endemic equilibrium z^* at the origin.

We can now state our main result.

Theorem 4.13. *Assume that $(H.1)$ and $(H.2)$ hold. For some $a>0$, let $T_{z,a}^N := \inf\{t > 0,\ Z_{z,1}^{N,\alpha}(t) \leq -a\}$, where $Z_{z,1}^{N,\alpha}(t)$ denotes the first coordinate of the process $Z_z^{N,\alpha}(t)$. The following hold.*

For any $z\in\mathbb{R}^d$ such that $z_1 > -a$, and any $\eta>0$,

$$\lim_{N\to\infty} \mathbb{P}\left(e^{a_N^{-1}(\overline{V}_a - \eta)} < T_{z,a}^N < e^{a_N^{-1}(\overline{V}_a+\eta)} \right) = 1\,,$$

and

$$\lim_{N\to\infty} a_N \log \mathbb{E}(T_{z,a}^N) = \overline{V}_a\,.$$

Given Corollary 4.12, the proof of the above result follows the exact same steps as that of Theorem 5.7.11 in [2], with some minor modifications, to adapt to the fact that our processes have discontinuous trajectories, see the proof of Theorem 7.14 in [5], or of Theorem 4.2.17 in [1].

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, $\alpha = 0$, $a_N^{-1} = N$.

4.7.1 Interpretation. The critical population size

Going back to the original coordinates, i.e. $z^* \neq 0$, we should interpret $Z^{N,\alpha}(t)$ as $Z^{N,\alpha}(t) = N^\alpha(Z^N(t) - z^*)$. So (dropping the index for the starting point in order to simplify our notations), T_a^N is the first time when $Z_1^N(t) \leq z_1^* - aN^{-\alpha}$. For T_a^N to be finite, we need to have $z_1^* - aN^{-\alpha} \geq 0$, since $Z_1^N(t)$ cannot become negative. This is of course no problem for the limit theorem, since $aN^{-\alpha} \to 0$ as $N\to\infty$, while z_1^* is fixed. However, a deviation of the order of $-aN^{-\alpha}$ is enough for $Z_1^N(t)$ to hit zero, if z_1^* is of the order of $N^{-\alpha}$, which means that N is of the order of $(z_1^*)^{-1/\alpha}$. $e^{N^{1-2\alpha}\overline{V}_a}$ is the order of magnitude of the time needed for $Z_t^N - z_t$ to make a deviation of size $aN^{-\alpha}$. This is sufficient to extinguish an epidemic, provided z_1^* is of the same order, so that the corresponding critical size is $N_\alpha \sim (1/z_1^*)^{1/\alpha}$, which is roughly the CLT critical population size raised to the power $1/2\alpha$.

4.7.2 The value of $\overline{V}_a$ in the SIS model

In the particular case of the SIS model, we can compute explicitly the value of the quasi–potential $\overline{V}_a$. In this case, $d = 1$, the linearized ODE around the endemic equilibrium translated at 0 reads

$$\dot{x} = -(\lambda - \gamma)x + u,$$

and the cost functional to minimize is

$$I_T(u) = \frac{\lambda}{4\gamma(\lambda-\gamma)} \int_0^T u(t)^2 dt\,.$$

We are looking for the minimal cost for driving x from 0 to $-a$. We now exploit the Pontryagin maximum principle, see [9]. The Hamiltonian reads

$$H(x,p,u) = -(\lambda-\gamma)px + pu - \frac{\lambda}{4\gamma(\lambda-\gamma)}u^2\,.$$

The optimal control $\hat{u}$ must maximize the Hamiltonian, so it satisfies $\hat{u} = \frac{2\gamma(\lambda-\gamma)}{\lambda}p$. Since the final time is free and the system is autonomous, the Hamiltonian vanishes along the optimal trajectory, so that along such a trajectory, either $p = 0$, in which case $\hat{u} = 0$, or else $x = \frac{\gamma}{\lambda}p$, hence $\hat{u} = 2(\lambda-\gamma)x$. Finally the pieces of optimal trajectory which move towards the origin correspond to $u \equiv 0$, those which move away from the origin (this is the case we are interested in) satisfy the time reversed ODE $\dot{x} = (\lambda-\gamma)x$. There is no optimal trajectory from $x = 0$ to $x = -a$. However, if we start from $x = -\varepsilon$, the optimal trajectory is $x(t) = -e^{(\lambda-\gamma)t}\varepsilon$, so $\hat{u}(t) = -2(\lambda-\gamma)e^{(\lambda-\gamma)t}\varepsilon$, the final state $-a$ is reached at time $(\lambda-\gamma)^{-1}\log(a/\varepsilon)$, and the optimal cost is $\frac{\lambda}{2\gamma}(a^2-\varepsilon^2)$. A possible sub–optimal control starting from 0 is as follows. Choose $u = -1$ for a time of order ε, until $x(t)$ reaches $-\varepsilon$, which costs or the order of ε, and then choose the optimal feedback, until $-a$ is reached. Letting $\varepsilon \to 0$, the cost converges to

$$\bar{V}_a = \frac{\lambda}{\gamma}\frac{a^2}{2}\,.$$

4.8 Comparison between the CLT, MD and LD

We do that comparison in case of the SIS model, for which we have explicit expressions for the rate functions and the quasi–potentials. We still translate z^* at the origin, and start our process at the origin : $Z_0^N = 0$. We fix $a > 0$ and want to compare (for t large) the upper bounds for $\mathbb{P}(N^\alpha Z_t^N \geq a)$ in the three cases $\alpha = 1/2$ (the central limit theorem), $0 < \alpha < 1/2$ (moderate deviations) and $\alpha = 0$ (large deviations).

We start with the **central limit theorem**. It is easy to see that $U_t = \lim_{N\to\infty}\sqrt{N}Z_t^N$ is an Ornstein–Uhlenbeck process (in particular it is a Gaussian process), which solves the SDE

$$U_t = -(\lambda-\gamma)\int_0^t U_s ds + \sqrt{2\gamma(\lambda-\gamma)/\lambda}B_t,$$

so that the asymptotic variance of U_t is γ/λ. Consequently for $a > 0$ fixed and any $\eta > 0$, there exist t and N large enough such that we have the following upper bound for the probability of a positive deviation of $\sqrt{N}Z_t^N$

$$\mathbb{P}(\sqrt{N}Z_t^N \geq a) \leq \exp\left\{-\frac{\lambda a^2}{2\gamma} + \eta\right\}.$$

Consider next the **moderate deviations**. Theorem 4.10 combined with the computation from the last subsection indicates that for $0 < \alpha < 1/2$, any $\eta > 0$, there exists t

and N large enough such that

$$\mathbb{P}(N^\alpha Z_t^N \geq a) \leq \exp\left\{-N^{1-2\alpha}\left(\frac{\lambda a^2}{2\gamma} - \eta\right)\right\}.$$

We finally consider the **large deviations**. Here we need to assume that $a < \gamma/\lambda$. We exploit the computations from sections 4.2.6 and A.6 in [1]. The optimal trajectory to go from ε to a is the original ODE, but time reversed, i.e. it follows the ODE $\dot{x}_t = \beta_2(x_t) - \beta_1(x_t)$. The running cost is $(\beta_2(x_t) - \beta_1(x_t)) \log\left(\frac{\beta_2(x_t)}{\beta_1(x_t)}\right)$, so the total cost is

$$\begin{aligned}\int_{T_\varepsilon}^{T_a} \log\left(\frac{\beta_2(x_t)}{\beta_1(x_t)}\right)(\beta_2(x_t) - \beta_1(x_t))dt &= \int_{T_\varepsilon}^{T_a} \log\left(\frac{\beta_2(x_t)}{\beta_1(x_t)}\right)\dot{x}_t dt \\ &= \int_\varepsilon^a \log\left(\frac{\beta_2(x)}{\beta_1(x)}\right)dx \\ &\to a + \left(\frac{\gamma}{\lambda} - a\right)\log\left(1 - a\frac{\lambda}{\gamma}\right), \quad \text{as } \varepsilon \to 0.\end{aligned}$$

Consequently, from Theorem 3.5, for any $\eta > 0$, there exists t and N large enough such that

$$\mathbb{P}(Z_t^N \geq a) \leq \exp\left\{-N\left[a + \left(\frac{\gamma}{\lambda} - a\right)\log\left(1 - a\frac{\lambda}{\gamma}\right) - \eta\right]\right\}.$$

We note that Moderate Deviations resembles much more the CLT than Large Deviations. The fact that the discontinuity in the form of the rate function is exactly at $\alpha = 0$ is typical of random variables with light tails. The situation would be quite different with heavy tails, see e. g. section VIII.4 in Petrov [8].

Note however that for small a,

$$a + \left(\frac{\gamma}{\lambda} - a\right)\log\left(1 - a\frac{\lambda}{\gamma}\right) \sim \frac{\lambda a^2}{2\gamma},$$

which is not too surprising, and in a sense reconcile Large Deviations and Moderate Deviations.

5 Appendix

In this Appendix, we establish the following technical result.

Proposition 5.1. *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 for any $0 \leq t \leq T$,*

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

Proof. Consider with $b \geq 0$ the process

$$X_t = a\int_0^t \int_0^{\varphi_s} \overline{\mathcal{M}}(ds, du) - b\int_0^t \varphi_s ds\,. \tag{5.1}$$

It follows from Itô's formula that

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

From Lemma 5.2 below, $M_t = \int_0^t \int_0^{\varphi_s} e^{X_{s-}} \overline{\mathcal{M}}(ds, du)$ 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 $>$. Consequently 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 Schwartz's inequality, we have

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

If $2b = e^{2a} - 1 - 2a$, it follows from the previous argument that the first factor on the second right hand side is less than or equal to 1, hence the result follows. □

In order to complete the proof of Proposition 5.1, we still need to establish

Lemma 5.2. *The process φ satisfying the same assumptions as in Proposition 5.1, and X_t being given by* (5.1)*, $M_t = \int_0^t \int_0^{\varphi_s} e^{X_{s-}} \overline{\mathcal{M}}(ds, du)$ is a martingale.*

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

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

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 it remains to show that

$$\mathbb{E}\exp\left\{4a\int_0^t \int_0^{\varphi_s} \mathcal{M}(ds, du)\right\} < \infty. \tag{5.2}$$

Using Itô's formula we have

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

The same computation with φ_s replaced by $\varphi_s^n = \varphi_s \wedge n$, and then Y_s replaced by Y_s^n would show that Y_t^n is a martingale satisfying $\mathbb{E}Y_t^n = 1$. But $0 \leq Y_t^n \to Y_t$ a.s., hence Fatou's Lemma implies that $\mathbb{E}Y_t \leq 1$. Since

$$4a\int_0^t \int_0^{\varphi_s} \mathcal{M}(ds, du) = 4a\int_0^t \int_0^{\varphi_s} \mathcal{M}(ds, du) - \frac{e^{8a} - 1}{2}\int_0^t \varphi_s ds + \frac{e^{8a} - 1}{2}\int_0^t \varphi_s ds,$$

it follows from Schwartz's inequality that

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

and the result follows from our assumption on φ. □

References

[1] Britton,T. and Pardoux, E.: Stochastic Epidemics in a Homogeneous Community, Part I of *Stochastic epidemic models with inference*, T. Britton & E. Pardoux eds., Lecture Notes in Math. **2255**, pp. 1–120, Springer 2019. arXiv:1808.05350

[2] Dembo, A. and Zeitouni, O.: *Large Deviations, Techniques and Applications*, 2d ed., Applications of Mathematics **38**, Springer, New York, 1998. MR-1619036

[3] Ethier, S. N. and Kurtz, T.G.: *Markov processes. Characterization and convergence*, J. Wiley 1986. MR-0838085

[4] Freidlin, M.I. and Wentzell, A.D.: *Random perturbations of dynamical systems*, 3d ed. Grundlehren des Mathematischen Wissenschaften **260**, Springer, New York, 2012. MR-2953753

[5] Kratz, P. and Pardoux, E.: Large deviations for infectious diseases models, in textitSéminaire de Probabilités XLIX, C. Donati-Martin, A. Lejay, A. Rouault eds., Lecture Notes in Math. **2215**, pp. 221-327, 2018. MR-3838866

[6] Pardoux, E.: Deviations from the law of large numbers, and extinction of an endemic disease, in *Mathematical modeling of random and deterministic phenomena*, Manou–Abi, S.M., Dabo–Niang, S., Salone, J.J. eds., pp. 3–30, ISTE and Wiley, 2020.

[7] Pardoux, E. and Samegni–Kepgnou, B.: Large deviation principle for epidemic models, *Journal of Applied Probability* **54**, 905–920, 2017. MR-3707836

[8] Petrov, V. V.: *Sums of Independent Random Variables*, Ergebnisse der Mathematik und ihrer Grenzgebiete **82**, Springer Verlag, 1975. MR-0388499

[9] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. G. and Mishchenko, E. F.: *The mathematical theory of optimal processes*. Transl. by K. N. Trirogoff; ed. by L. W. Neustadt, John Wiley & Sons, 1962. MR-0166037

[10] Shwartz, A. and Weiss, A.: textitLarge Deviations for Performance Analysis, Chapman Hall, London, 1995. MR-1335456

[11] Shwartz, A. and Weiss, A.: Large deviations with diminishing rates, textitMathematics of Operations Research **30** 281–310, 2005. MR-2142034

Acknowledgments. It is a pleasure to thank Pierre Petit for an inspiring discussion on moderate deviations, and two anonymous Referees for their careful reading of the paper.