Large deviation based design of an interacting particle method to compute moment generating functions

by

Mengjie Liu

B.Sc., Nankai University, Tianjin, P. R. China, 2018

M.Sc., Brown University, Providence, RI, 2019

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

February 2025

This dissertation by Mengjie Liu is accepted in its present form
by the Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date_____ _____

Paul Dupuis, Ph.D., Advisor

Recommended to the Graduate Council

Date_____ _____

Oanh Nguyen, Ph.D., Reader

Date_____ _____

Matthew Harrison, Ph.D., Reader

Approved by the Graduate Council

Date_____ _____

Thomas A. Lewis, Dean of the Graduate School

Vita

Mengjie Liu was born in 1996 in Shijiazhuang, Hebei Province, China. She went to Nankai University in September 2014 for her undergraduate study and obtained her B.Sc. degree in Mathematics in June 2018. In September 2018, she started the Ph.D. program in the Division of Applied Math at Brown University, and obtained her M.Sc. degree in Applied Math in May 2019. While at Brown, she has worked under the supervision of Prof. Paul Dupuis.

Acknowledgments

I would first like to express my heartfelt gratitude to my advisor, Professor Paul Dupuis, for his unwavering support and guidance throughout my Ph.D. journey at Brown University. As an exceptional advisor, his sharp intellect, tireless dedication, and genuine care for his students have profoundly shaped my academic and personal growth. I am deeply grateful for the privilege of learning from his expertise and passion for research. His enthusiasm for exploring new ideas and his readiness to address my questions have been invaluable to my work. Being a student under his mentorship has been an honor and a source of great pride.

I would also like to extend my sincere thanks to my dissertation committee for their time, thoughtful feedback, and invaluable suggestions during the review of my dissertation. In particular, I am deeply appreciative of Professor Matthew Harrison and Professor Oanh Nguyen for their insightful guidance, which significantly influenced the direction and development of my thesis.

I would like to express my gratitude to all the teachers and professors who have played an important role in my academic journey, including Professors Hongjie Dong, Chi-Wang Shu, Kavita Rammanan, and many others. Your guidance and support have been invaluable to me throughout my years as a student.

I would like to extend my heartfelt gratitude to my dear friends, both near and far. Yuqing Wang, Boxin Zhao and Yichen Bai in the United States, as well as

Jingyi Cui in China, hold a special place in my heart. I will always treasure the joyful gatherings with Enrui Zhang, Xiaoyu Xie, Sicheng Liu, Xin Lian, Qian Zhang, Yi Jiang, Tianming Yu, Xiaoding and Minglang Yin during the final year of my graduate studies. These moments of friendship and connection have left a lasting impression on my journey.

I am deeply grateful to my friend Sining Gong. From the moment we arrived on campus together, you have been more than just friends—you have been kind neighbors and steadfast companions, offering unwavering support and encouragement throughout our time at Brown. I would also like to extend my heartfelt thanks to my office mates, Sicheng and Xiaoding, as well as Xin and Tianmin. I will always treasure the memories we created together, from our lively office discussions about everything under the sun to our shared meals at local restaurants. These moments will remain a cherished part of my journey.

I want to express my deepest gratitude to my dear friend, Yaning. From the moment we met as undergraduate classmates and roommates, our friendship has only grown stronger over the years. Your constant support and companionship have been a cornerstone of my experience. Whether it was late-night study sessions during undergrad or our adventures exploring new things, your presence has brought so much joy and encouragement to my life. I am truly fortunate to have you as my friend. I couldn't have achieved this without you.

Last but certainly not least, I would like to express my deepest gratitude to my parents, Junguo Liu and Zhinan Liu. Their unwavering encouragement and trust have been the foundation of my determination and confidence throughout my doctoral studies. I am profoundly thankful for their unconditional love and steadfast support of my decisions, which have made this journey possible. Finally, I owe

my heartfelt thanks to my husband, Zhen Zhang. His unwavering companionship and support have been the driving force that enabled me to persevere through the challenges of this demanding Ph.D. journey.

Abstract of "Large deviation based design of an interacting particle method to compute moment generating functions", by Mengjie Liu, Ph.D., Brown University, February 2025

This thesis introduces a novel interacting particle scheme that combines large deviation analysis with the standard splitting method and rare event analysis to compute the moment generating function for ergodic Markov processes. The proposed scheme differs fundamentally from existing methods, as the thresholds that trigger splitting depend on both space and time.

We introduce a quasi-stationary distribution and the Feller property, both of which are essential for analyzing the scheme. We derive the interacting particle schemes based on space and timedependent thresholds, proving the unbiasedness of the estimator. Furthermore, we establish a theorem that bounds the second moment of the proposed unbiased estimator in terms of key parameters of the scheme.

The thesis not only establishes new theoretical results but also provides numerical evidence demonstrating the effectiveness of our method.

Contents

List of Tables

List of Figures

Chapter One

Introduction and Preliminaries

1.1 Monte Carlo Simulation

A common technique for the estimation of probabilities and functionals of probability measures is the standard Monte Carlo simulation. Let $\mathcal{P}(\mathbb{R})$ denote the space of probability measures on $\mathbb{R}$. Suppose X is a random variable with finite expectation taking values in $\mathbb{R}$ distributed according to $\mu \in \mathcal{P}(\mathbb{R})$. To estimate the expectation of X, denoted by $\mathbb{E}[X]$, here we assume $\mathbb{E}[X] < \infty$, we can generate N independently and identically distributed (iid) samples from μ, denoted as $X_0, X_1, \ldots, X_{N-1}$, and the estimate for $\mathbb{E}[X]$ is just the sample mean:

$$Q_N = \frac{1}{N} \sum_{i=0}^{N-1} X_i.$$

Law of Large Numbers (LLN) ensures that this estimate converges to the correct value as the number of samples N goes to infinity. The sample size N can be chosen if we have the rate of convergence.

In cases where the variance is large, and especially if it is large compared to $\mathbb{E}[X]$, it may take a large number of samples before the variance of the estimator is acceptably small relative to $\mathbb{E}[X]$.

We now turn to continuous time Markov chain moment generating functions. We want the numerical approximation of

$$E_x \exp\left\{\alpha \int_0^n c(X(s))ds\right\}, \tag{1.1.1}$$

where n can be large, E_x denotes conditioning on $X(0) = x$, $\{X(t), t \in [0, \infty)\}$ is an ergodic Markov process taking values in a compact Polish space $\mathcal{X}$, $c : \mathcal{X} \to \mathbb{R}$ is continuous, and $\alpha \in \mathbb{R}$.

Expected values of the form (1.1.1) arise whenever the growth rate of a system is stochastic, and occur for example in mathematical finance, climate modeling and epidemiology [22, 21]. They also occur in uncertainty quantification, where quantities such as (1.1.1), when computed under a nominal or design model, are used to bound costs under alternative models, with model distances defined in terms of relative entropy [10, 3].

The reason that the design of Monte Carlo numerical methods for this type of problem is difficult is that the expected value is often largely determined by rare events. This means that straightforward Monte Carlo cannot be expected to work well, and one will want to consider a more efficient method of numerical approximation.

1.2 Important Sampling Method and Splitting Particles Method

Among the more complicated but more efficient methods available for the Monte Carlo estimation, the most well known are importance sampling and particle splitting (or branching) methods, among which one could include interacting particle methods such as [7]. The schemes analyzed in this dissertation belong to the second category.

Importance Sampling Schemes

The first method is referred as importance sampling. Suppose that X has distribution θ, and consider an alternative sampling distribution τ. It is required that θ be absolutely continuous with respect to τ, so that the Radon-Nikodym derivative $f(x) \doteq (d\theta/d\tau)(x)$ exists. IID samples $\bar{X}_0, \bar{X}_1, \ldots$ with distribution τ are generated, and the estimate

$$\bar{Q}_N \doteq \frac{1}{N} \sum_{k=0}^{N-1} \bar{X}_k f\left(\bar{X}_k\right)$$

is considered in lieu of Q_N. Since

$$E\bar{X}_k f\left(\bar{X}_k\right) = \int_{\mathbb{R}} x f(x) \tau(dx) = \int_{\mathbb{R}} x\theta(dx) = EX,$$

$\bar{Q}_N$ is an unbiased estimate of EX, with a rate of convergence determined by

$$\operatorname{var}\left[\bar{X}_0 f\left(\bar{X}_0\right)\right] = \int_{\mathbb{R}} x^2 f(x)\theta(dx) - \left[\int_{\mathbb{R}} x\theta(dx)\right]^2.$$

The goal is to choose τ so that the individual samples $f(x)$ cluster tightly around θ, thereby reducing the variance. The optimization of this quantity over all possible τ is inappropriate. For example, suppose θ is supported on $[0, \infty)$, and $\theta(dx) = g(x)dx$. Let $m = \mathbb{E}[X] = \int_0^\infty x\theta(dx)$ and $\tau(dx) = \frac{xg(x)}{m}dx$. Then θ is absolutely continuous with respect to τ, with $f(x) = \frac{m}{x}$. Furthermore,

$$\operatorname{var}\left[\bar{X}_0 f\left(\bar{X}_0\right)\right] = \int x^2 f(x)\theta(dx) - m^2 = 0.$$

However, the distribution τ is not practically useful, as it requires knowledge of m, which is exactly the quantity we are trying to estimate. Rather than optimizing without constraints, it is more common to minimize over a family of parameterized

alternative sampling distributions. When the distribution of X_0 is linked to a large deviations problem, the form of the large deviations variational problem naturally suggests specific sampling distributions.

The characterization of optimal achievable performance of importance sampling schemes in the large deviation is discussed in [12, 13] in terms of a deterministic differential game. And [14] suggests a way to construct importance sampling schemes based on subsolutions of the Isaacs equation.

There is a challenge in applying importance sampling to simulate a process Y_t, where Y_t represents the integral of a continuous-time Markov chain, and x is the value of the Markov chain $X(t)$ at time t. This difficulty arises because it is challenging to select an appropriate reference distribution for the continuous process Y_t. In contrast, splitting-type schemes only require the ability to simulate trajectories under the original dynamical model.

Branching Methods

For branching methods, one of the most discussed applications is the first entrance probabilities. Consider a continuous time Markov process $X(t)$, and the probability of interest is given by:

$$p = P\left(\tau_A \leq \tau_B \mid X(0) = x\right),$$

where $\tau_C = \min\{t \geq 0 : X(t) \in C\}$ for sets A and B. In this context, we focus on the continuous time case. Assume that B is a "typical set" and that A is rare, meaning

that with high (perhaps very high) probability, the process will enter B before A.

In branching processes, the goal is to modify the process so that the a priori rare outcomes become more likely for at least some of the descendants (or particles), and then normalize to produce an unbiased estimate of p. Similar to ordinary Monte Carlo methods, one can generate multiple independent copies of the branched process and use the sample average to estimate p. Since the copies are independent, the performance of the scheme is determined by the variance of a single copy.

Key questions in constructing the branching process include determining when to branch (the triggers) and how much to branch (the rates). A natural and standard way to define triggers is through the use of thresholds, that is, nested sets $\{C_j, j = 0, 1, \ldots, m\}$ such that

$$C_0 \supset C_1 \supset \cdots \supset C_m \supset A,$$

with a branching rule that is triggered when these sets are entered. For example, branching could occur the first time a particular particle, or perhaps a selected set of particles containing it, enters a threshold set. An example figure of the triggers of branching scheme is Figure 1.1, here $C_0 = B^c$.

Within this framework where branching is triggered by thresholds, there are many possible variations. In what appears to be the oldest class of such methods, particles do not interact, and branching occurs for any given particle the first time it reaches a threshold that has not been reached before. The number of descendants could be random and may depend on factors such as the position of the particle upon entering the threshold, but it is independent of the state of all other particles at the time of branching. (If multiple thresholds are entered simultaneously, the distribution of

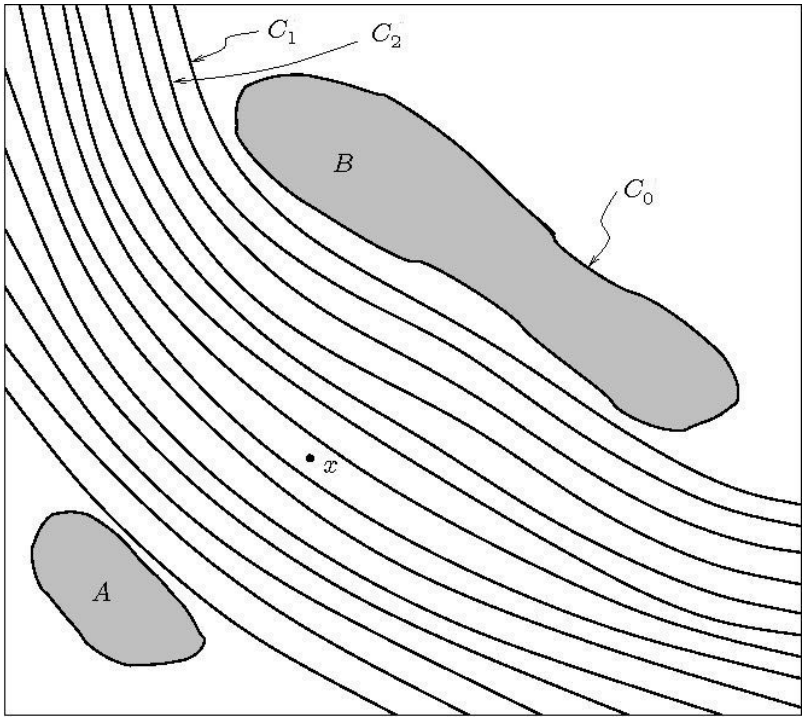

Figure 1.1: The Sets A and B and Trigger Sets C_j.

the number of descendants is the appropriate convolution of distributions.) The mean number of descendants is often assumed to be constant (which we will assume for discussion purposes), and thus these schemes are sometimes called *Fixed Rate Branching* schemes.

In addition to the basic fixed branching rate schemes [18], more efficient variants have been developed that eliminate less useful particles while still remaining unbiased [24, 25].

Designing such schemes presents several challenges. In large-scale problems, one must carefully ensure that the total number of particles whose distribution is inherently influenced by the placement of the thresholds does not become excessively large. At the same time, to significantly reduce the variance of the estimator, it is essential that enough particles successfully reach the rare set. The main challenge in the design is to strike the right balance between these competing requirements.

In a second class of methods, branching occurs only when all particles that started from the previous threshold (say, C_{j-1}) have either reached C_j or B. In these methods, the branching is not conditionally independent of the state of other particles. Instead, the process is designed to ensure that a fixed and deterministic number (denoted as N) of particles remain at each threshold. Specifically, suppose the process has a continuous trajectories and $\{X_{j-1,n}, n = 1, \ldots, N\}$ represent the positions of N particles after reaching threshold C_{j-1}. Starting from these positions, the particles evolve independently under the dynamics of X until they reach either C_j or B. Let N_j denote the number of particles that reach C_j before B, and let S_j represent the corresponding locations of these particles.

To maintain N particles, we set the positions of the first N_j particles, $X_{j,n}$ for $n = 1, \ldots, N_j$, to coincide with the locations in S_j (the order does not matter), and independently sample the remaining $N - N_j$ particles from S_j according to the uniform distribution. If $N_j = 0$, the simulation terminates, and zero is returned as the estimate. Repeating this process over, say, m thresholds that separate the initial point from A, we obtain an unbiased estimate

$$\hat{p} = \prod_{j=1}^{m} \frac{N_j}{N}$$

of $P(\tau_A \leq \tau_B)$.

Methods of this type are studied in [8, 17], and since the total computational effort (or at least the total number of particles) is predetermined, they are referred to as *Fixed Effort Branching*. Other name for these schemes includes *interacting particle systems (IPS)* methods.

There are known methods for incorporating large deviation information to assist in the design of splitting, including the RESTART algorithm [6]. The idea of the RESTART algorithm (Repetitive Simulation Trials After Reaching Thresholds) is to enhance the frequency of rare events by introducing nested sets of states and performing oversampling in high-importance regions. In order to ensure unbiasedness of the estimate, particles are split every time they cross into a splitting threshold. The proof of unbiasedness can be found in [6].

The RESTART algorithm is defined using an importance function that determines the thresholds C_j. This importance function can be either time-dependent or time-independent. For clarity, the algorithm is initially described using a time-independent importance function $V(y,t)$. The framework can be extended to handle the time-dependent case by treating time as an additional spatial variable, such that $x = (y,t) \in \mathbb{R}^d \times [0,T] \subset \mathbb{R}^{d+1}$. Regardless of whether the importance function is time-dependent or not, its design will be based on an associated Hamilton-Jacobi-Bellman (HJB) partial differential equation.

However, very little is known regarding how to use large deviation information in the design of interacting particle schemes. Current analyses have examined the variance by first allowing the number of particles to tend to infinity (thus enabling the exploitation of exchangeability and asymptotically negligible interactions between particles), followed by making the probability small. However, interacting particle schemes are designed in part to control computational effort in a predictable manner, so this approach does not provide the ideal measure of performance. It is preferable to keep the number of particles fixed while reducing the probability and, similar to fixed rate schemes, compare the standard deviation of the estimator with the probability. This type of analysis is challenging. In standard splitting, two particles that reach A have a relatively straightforward dependency: they share a common ancestor at

some point in the past. Before that, their histories are identical, and after that, their evolutions are independent. By partitioning the particles according to the time and location of their last common ancestor, this simple dependency can be exploited. Although more complex methods are harder to analyze, the dependency between particles remains constrained.

The situation with interacting particle schemes is very different. Here, dependence between particles is reintroduced with each resampling, which makes the analysis much more difficult. In the non-interacting case, one can analyze the variance of the estimator using large deviation estimates and techniques related to Freidlin-Wentsell type "small noise" asymptotics [1]. In the interacting case, the relevant techniques are related to those used to analyze the empirical or occupation measure of a Markov process, though there are numerous difficulties that make the analysis significantly more technical and complex. Among the difficult features are: (i) the natural time scale for the process is not the natural time scale for the algorithm, which is in fact indexed by thresholds; (ii) the transition kernels associated with threshold crossing typically do not satisfy the Feller property; (iii) the transition kernels also depend on the large deviation parameter; and (iv) the derived estimates should be valid all the way from a small number of particles (where large deviation effects dominate) to many particles (where law of large numbers effects dominate). A large deviation analysis of a different use of interacting particle methods in [5] suggests that they exhibit certain robustness properties not found with importance sampling or standard splitting.

The idea of this thesis is to design and analyze interacting particle schemes based exclusively on a large deviations and rare event analysis. As we will demonstrate, the use of large deviation ideas will suggest interacting particle schemes that differ significantly from existing approaches, and in particular the thresholds that are used

to trigger splitting through resampling depend on both space and time.

1.3 Preliminaries

In this section, we introduce some useful concepts and theorems which will be needed throughout this thesis.

We begin by giving the definition of rate function and large deviation principle. See [[7], Section 1 and 2] for detailed discussion. Let $\mathcal{X}$ be a complete separable metric space, i.e., a *Polish space*. And we will use $\mathcal{X}$ as a Polish space in this thesis.

Definition 1.3.1. (Rate Function). A function I on $\mathcal{X}$ is called a rate function on $\mathcal{X}$, or simply a rate function, if I maps $\mathcal{X}$ into $[0, \infty]$ and if for each $M < \infty$ the level set $\{x \in \mathcal{X} : I(x) \leq M\}$ is a compact subset of $\mathcal{X}$.

Let $\{X_m, m \in \mathbb{N}\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in a Polish space $\mathcal{X}$.

Definition 1.3.2. (Large Deviation Principle). Let I be a rate function on $\mathcal{X}$. The sequence $\{X_m\}$ is said to satisfy the large deviation principle on $\mathcal{X}$ with rate function I if the following two conditions hold.

(a) Large deviation upper bound. For each closed subset F of $\mathcal{X}$

$$\limsup_{m\to\infty} \frac{1}{m} \log \mathbb{P}\{X_m \in F\} \leq - \inf_{x\in F} I(x).$$

(b) Large deviation lower bound. For each open subset G of $\mathcal{X}$

$$\liminf_{m\to\infty} \frac{1}{m} \log \mathbb{P}\{X_m \in G\} \geq - \inf_{x\in G} I(x).$$

Another way of studying the large deviations is the following Laplace principle, which is equivalent to the large deviation principle.

Definition 1.3.3. (Laplace Principle). Let I be a rate function on $\mathcal{X}$. The sequence $\{X^m\}$ is said to satisfy the Laplace principle on $\mathcal{X}$ with rate function I if for all bounded continuous functions $h : \mathcal{X} \to \mathbb{R}$

$$\lim_{m\to\infty} \frac{1}{m} \log \mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\} = - \inf_{x\in\mathcal{X}} \{h(x) + I(x)\}$$

The term Laplace principle upper bound refers to the validity of

$$\limsup_{m\to\infty} \frac{1}{m} \log \mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\} \leq - \inf_{x\in\mathcal{X}} \{h(x) + I(x)\}$$

for all bounded continuous function h while the term Laplace principle lower bound refers to the validity of

$$\liminf_{m\to\infty} \frac{1}{m} \log \mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\} \geq - \inf_{x\in\mathcal{X}} \{h(x) + I(x)\}$$

for all bounded continuous functions h.

Next we introduce the relative entropy, which build the fundamental proof of weak convergence in large deviations through the variational formula which we introduce later in Proposition 1.3.5. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and $\mathcal{P}(\mathcal{X})$ the set of probability measures on $(\mathcal{X}, \mathcal{A})$.

Definition 1.3.4. (Relative Entropy). Given $\gamma, \theta \in \mathcal{P}(\mathcal{X})$, the relative entropy of γ with respect to θ is defined by

$$R(\gamma\|\theta) = \int_{\mathcal{X}} \log\left(\frac{d\gamma}{d\theta}(x)\right)\gamma(dx)$$

if γ is absolutely continuous with respect to θ, i.e. $\gamma \ll \theta$, and $\log([d\gamma/d\theta](x))$ is integrable with respect to γ, and by ∞ otherwise.

Next proposition states the variational formula result.

Proposition 1.3.5. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, k a bounded measurable function mapping $\mathcal{S}$ into $\mathbb{R}$, and θ a probability measure on $\mathcal{X}$. The following conclusions hold.

(a) We have the variational formula

$$-\log\int_{\mathcal{X}} e^{-k}d\theta = \inf_{\gamma\in\mathcal{P}(\mathcal{X})}\left\{R(\gamma\|\theta) + \int_{\mathcal{X}} kd\gamma\right\}$$

(b) Let γ_0 denote the probability measure on $\mathcal{X}$ which is absolutely continuous with respect to θ and satisfies

$$\frac{d\gamma_0}{d\theta}(x) \doteq e^{-k(x)}\cdot\frac{1}{\int_{\mathcal{X}} e^{-k}d\theta}$$

Then the infimum in the variational formula is attained uniquely at γ_0.

Proposition 1.3.6. Let $\mathcal{X}$ be a Polish space, k is a measurable function mapping $\mathcal{X}$ into $\mathbb{R}$ which is either bounded from below or bounded from above, and θ a

probability measure on $\mathcal{X}$. Then we have the variational formula

$$-\log \int_{\mathcal{X}} e^{-k} d\theta = \inf_{\gamma \in \Delta(\mathcal{X})} \left\{ R(\gamma \| \theta) + \int_{\mathcal{X}} k d\gamma \right\}$$

where $\Delta(\mathcal{X}) \doteq \{\gamma \in \mathcal{P}(\mathcal{X}) : R(\gamma\|\theta) < \infty\}$

[4] proves many properties of relative entropy, here we choose some of them we used in this thesis: convexity, lower semicontinuity, compactness, uniform integrability, and chain rule, etc.

Let $\mathcal{X}$ be a Polish space. We denote by $\mathcal{C}_b(\mathcal{X})$ the space of bounded continuous functions mapping $\mathcal{X}$ into $\mathbb{R}$ and by $\Psi_b(\mathcal{X})$ the space of bounded measurable functions mapping $\mathcal{X}$ into $\mathbb{R}$.

Lemma 1.3.7. Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. The relative entropy $R(\cdot\|\cdot)$ has the following properties.

(a) (Donsker-Varadhan variational formula.) For each γ and θ in $\mathcal{P}(\mathcal{X})$

$$\begin{aligned} R(\gamma\|\theta) &= \sup_{g \in \mathcal{C}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} g d\gamma - \log \int_{\mathcal{X}} e^g d\theta \right\} \\ &= \sup_{\psi \in \Psi_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} \psi d\gamma - \log \int_{\mathcal{X}} e^\psi d\theta \right\}. \end{aligned}$$

(b) $R(\gamma\|\theta)$ is a convex, lower semicontinuous function of $(\gamma, \theta) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$. In particular, $R(\gamma\|\theta)$ is a convex, lower semicontinuous function of each variable γ or θ separately. In addition, for fixed $\theta \in \mathcal{P}(\mathcal{X})$, $R(\cdot\|\theta)$ is strictly convex on the set $\{\gamma \in \mathcal{P}(\mathcal{X}) : R(\gamma\|\theta) < \infty\}$.

(c) For each $\theta \in \mathcal{P}(\mathcal{X})$, $R(\cdot\|\theta)$ has compact level sets. That is, for each $M < \infty$

the set $\{\gamma \in \mathcal{P}(\mathcal{X}) : R(\gamma\|\theta) \leq M\}$ is a compact subset of $\mathcal{P}(\mathcal{X})$.

(d) Let $\mathcal{X} = \mathbb{R}^d$ and let $\{\gamma_n, n \in \mathbb{N}\}$ and $\{\theta_n, n \in \mathbb{N}\}$ be sequences in $\mathcal{P}\left(\mathbb{R}^d\right)$. Assume that for each $\alpha \in \mathbb{R}^d$

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \exp\langle \alpha, y\rangle \theta_n(dy) < \infty \text{ and } \sup_{n\in\mathbb{N}} R\left(\gamma_n\|\theta_n\right) < \infty.$$

Then $\{\gamma_n\}$ is both tight and uniformly integrable in the sense that

$$\lim_{C\to\infty} \sup_{n\in\mathbb{N}} \int_{\left\{y\in\mathbb{R}^d : \|y\|>C\right\}} \|y\|\gamma_n(dy) = 0.$$

(f) Let $\sigma(dy|x)$ and $\tau(dy|x)$ be stochastic kernels on $\mathcal{Y}$ given $\mathcal{X}$ and θ a probability measure on $\mathcal{X}$. Then the function mapping $x \in \mathcal{X} \mapsto R(\sigma(\cdot\|x)\|\tau(\cdot\|x))$ is measurable and

$$\int_{\mathcal{X}} R(\sigma(\cdot\|x)\|\tau(\cdot\|x))\theta(dx) = R(\theta \otimes \sigma\|\theta \otimes \tau).$$

Theorem 1.3.8. (Chain Rule). Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and α and β probability measures on $\mathcal{X} \times \mathcal{Y}$. We denote by α_1 and β_1 the first marginals of α and β and by $\alpha(dy \mid x)$ and $\beta(dy \mid x)$ the stochastic kernels on $\mathcal{Y}$ given $\mathcal{X}$ for which we have the decompositions

$$\alpha(dx \times dy) = \alpha_1(dx) \otimes \alpha(dy \mid x) \text{ and } \beta(dx \times dy) = \beta_1(dx) \otimes \beta(dy \mid x).$$

Then the function mapping $x \in \mathcal{X} \mapsto R(\alpha(\cdot \mid x)\|\beta(\cdot \mid x))$ is measurable and

$$R(\alpha\|\beta) = R\left(\alpha_1\|\beta_1\right) + \int_{\mathcal{X}} R(\alpha(\cdot \mid x)\|\beta(\cdot \mid x))\alpha_1(dx)$$

Theorem 1.3.9. (Decomposition Theorem). Let $\sigma(dx \times dy \mid z)$ be a stochastic kernel on $\mathcal{X} \times \mathcal{Y}$ given $\mathcal{S}$ and θ a probability measure on $(\mathcal{S}, \mathcal{A})$. We denote by

$\sigma_1(dx \mid z)$ the first marginal of $\sigma(dx \times dy \mid z)$, which is the stochastic kernel on $\mathcal{X}$ given $\mathcal{S}$ defined by $\sigma_1(A \mid z) \doteq \sigma(A \times \mathcal{Y} \mid z)$ for Borel sets A and $z \in \mathcal{Y}$. Then there exists a stochastic kernel $\sigma(dy \mid x, z)$ on $\mathcal{Y}$ given $\mathcal{X} \times \mathcal{S}$ such that θ-a.s. for $z \in \mathcal{S}$

$$\sigma(A \times B \mid z) = \int_{A \times B} \sigma_1(dx \mid z)\sigma(dy \mid x, z) = \int_A \sigma(B \mid x, z)\sigma_1(dx \mid z)$$

for all Borel subesets A of $\mathcal{X}$ and B of $\mathcal{Y}$. This decomposition is summarized as $\sigma(dx \times dy \mid z) = \sigma_1(dx \mid z) \otimes \sigma(dy \mid x, z)$.

Theorem 1.3.10. (Varadhan's Lemma). Let $\{X_m\}_{m \geq 1}$ be a family of random variables that satisfy the large deviation principle on $\mathcal{X}$ with rate function $I(x)$, and let $f : \mathcal{X} \to \mathbb{R}$ be a continuous function. Then

$$\lim_{m \to \infty} \frac{1}{m} \log \mathbb{E}\left[e^{mf(X_m)}\right] = \sup_x \left(f(x) - I(x)\right),$$

where the supremum is taken over all x in the support of $I(x)$, which is the effective domain of the rate function $I(x)$.

We also show that a uniform Laplace principle as formulated here implies the uniform large deviation principle as defined by Freidlin and Wentzell in [16] (as the parameter varies over a compact set).

Definition 1.3.11. (compact level set) Let $\mathcal{X}$ be a Polish space and $\mathcal{Y}$ be a topological space. A family of rate functions I_y parametrized by $y \in \mathcal{Y}$ is said to have compact level sets on compacts if for each compact subset K of $\mathcal{Y}$ and each $M < \infty$,

$$\bigcup_{y \in K} \{x \in \mathcal{X} : I_y(x) \leq M\}$$

is a compact subset of $\mathcal{X}$.

Definition 1.3.12. (Uniform Laplace Function) Let I_y be a family of rate functions on $\mathcal{X}$ parametrized by y in a topological space $\mathcal{Y}$ and assume that this family has compact level sets on compacts. Let $\{X_n\}$ be a sequence of $\mathcal{X}$-valued random variables with distributions that depend on $y \in \mathcal{Y}$, and denote the corresponding expectation operator by $\mathbb{E}_y$. The sequence $\{X_n\}$ is said to satisfy the **Laplace principle on $\mathcal{X}$ with rate function I_y uniformly on compacts** if for all compact subsets K of $\mathcal{Y}$ and all bounded continuous functions $h : \mathcal{X} \to \mathbb{R}$,

$$\limsup_{n\to\infty} \sup_{y\in K} \left| \frac{1}{n} \log \mathbb{E}_y \exp\{-nh(X_n)\} - F(y,h) \right| = 0, \tag{1.6}$$

where $F(y,h) \doteq -\inf_{x\in\mathcal{X}} [h(x) + I_y(x)]$.

(a) The term **uniform Laplace principle upper bound** refers to the validity of

$$\limsup_{n\to\infty} \sup_{y\in K} \left(\frac{1}{n} \log \mathbb{E}_y \exp\{-nh(X_n)\} - F(y,h) \right) \leq 0$$

for all compact subsets K of $\mathcal{Y}$ and all bounded continuous functions h.

(b) The term **uniform Laplace principle lower bound** refers to the validity of

$$\liminf_{n\to\infty} \inf_{y\in K} \left(\frac{1}{n} \log \mathbb{E}_y \exp\{-nh(X_n)\} - F(y,h) \right) \geq 0$$

for all compact subsets K of $\mathcal{Y}$ and all bounded continuous functions h.

Last, we introduce the most important condition that we need in this thesis.

Assume $p(x, dy)$ a transition probability function defined on polish space $\mathcal{S}$. The Feller property is defined as follows:

Definition 1.3.13. (Feller Property). The function mapping $x \in \mathcal{S} \mapsto p(x,\cdot) \in$

$\mathcal{P}(\mathcal{S})$ is continuous in the topology of weak convergence on $\mathcal{P}(\mathcal{S})$, which means if $\{x_n, n \in \mathbb{N}\}$ is any sequence in $\mathcal{S}$ such that $x_n \to x \in \mathcal{S}$, then $p(x_n, \cdot) \Rightarrow p(x, \cdot)$.

1.4 Asymptotic Optimality

As we mentioned before, our interest is to calculate the moment generating function for a continuous time Markov chain. It may take a huge number of samples to get an accurate estimation. Since we have known the knowlege of large deviation principle, now we can show the 'optimal' convergent rate of the second moment for rare events simulation.

Let $X = X_0, X_1, \ldots$ be a sequence of iid $\mathbb{R}$-valued random variables with distribution μ, and assume the moment generating function $\mathbb{E}\left\{e^{\alpha X}\right\}$ is finite for all $\alpha \in \mathbb{R}$. Let $X^m = \frac{X_0 + X_1 + \cdots + X_{m-1}}{m}$. Suppose sequence $\{X^m/m\}$ satisfy the Laplace principle on $\mathbb{R}$ with rate function I, then for any fixed bounded continuous functions h, we have

$$\lim_{m\to\infty} -\frac{1}{m} \log \mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\} = \inf_{x\in\mathcal{X}}\{h(x) + I(x)\} = \sigma.$$

In order to estimate $\mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\}$, we generate iid samples $(X_0^m, X_1^m, ..., X_{K-1}^m)$ such that X_i^m has same distribution with X^m, then we can obtain the unbiased estimator of $\mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\}$ by using the average:

$$Q_K^m = \frac{1}{K} \sum_{i=0}^{K-1} \exp\left[-mh\left(X_i^m\right)\right].$$

And we know $Var(x) = E(x^2) - E^2(x)$ where x is a random variable, then the rate of convergence for the estimator Q_K^m is determined by the variance. Hence it suffices

to consider the second moment $\mathbb{E}\left\{\exp\left[-2mh\left(X_0^m\right)\right]\right\}$.

By Jensen's inequality, we know

$$\lim_{m\to\infty} -\frac{1}{m}\log\mathbb{E}\left\{\exp\left[-2mh\left(X^m\right)\right]\right\} \leq \lim_{m\to\infty} -\frac{1}{m}\log\left(\mathbb{E}\left\{\exp\left[-mh\left(X^m\right)\right]\right\}\right)^2 = 2\sigma.$$

We say the estimator Q_K^m is asymptotically optimal if this upper bound is achieved, i.e. if

$$\lim_{m\to\infty} -\frac{1}{m}\log\mathbb{E}\left\{\left[(Q_K^m)^2\right]\right\} = 2\sigma.$$

Definition 1.4.1. The decay rate of an unbiased estimator Q_N^m is given by

$$\liminf_{m\to\infty} -\frac{1}{m}\log E[(Q_N^m)^2].$$

Observe that Jensen's inequality places an upper bound on the decay rate of any unbiased estimator: it can be no better than twice the large deviation limit. When this upper bound is achieved, we say that the estimator is *asymptotically optimal*.

This thesis studies the performance of one of those more efficient schemes for rare event simulation, namely, interacting particle scheme.

The thesis is structured as follows. Chapter 2 formulates the problem and the algorithm whose idea combines the large deviations analysis and rare event estimation. It also defines a quasi-stationary distribution required for the analysis of the scheme and provides examples of process models. Chapter 3 derives the interacting particle schemes that are based on space and time dependent thresholds, and proves unbiasedness of the estimator. Chapter 4 reintroduces the large deviation parameter, and identifies the key quantities whose large deviation limits will determine the

second moment of the estimator. This chapter also presents the main result, characterizing the second moment in terms of the key design parameter and the number of particles. Chapter 5 states the large deviation properties of the quantities identified in Chapter 4, and gives the proof of main result. Proofs of the large deviation results for the constitutive elements are presented in Chapter 7, and numerical results are discussed in Chapter 6.

Chapter Two

Interacting Particle System Formulation

In this chapter we formulate the interacting particle method, which combines the large deviations analysis and rare event estimation. We begin by identifying the moment generating function for continuous Markov Chain and the related large deviation result in the next section. In the following section we introduce the space and time dependent thresholds and the Feller Property. In the last section we define a quasi-stationary distribution that will be used in the analysis of the algorithm estimator and large deviation analysis.

2.1 Notation and Problem Formulation

Our problem is to formulate the numerical approximation of

$$E_x \exp\left\{\alpha \int_0^n c(X(s))ds\right\}, \tag{2.1.1}$$

where n can be large, E_x denotes conditioning on $X(0) = x$, $\{X(t), t \in [0,\infty)\}$ is an ergodic Markov process taking values in a compact Polish space $\mathcal{X}$, $c : \mathcal{X} \to \mathbb{R}$ is continuous, and $\alpha \in \mathbb{R}$. To see the large deviation asymptotics, we rewrite the expected value in terms of the empirical measure of X, which is defined by

$$\mu^n(A) \doteq \frac{1}{n}\int_0^n 1_A(X(s))ds,$$

where $A \in \mathcal{B}(\mathcal{X})$ and 1_A is the indicator function. Then (2.1.1) can be written as

$$E_x \exp\left\{n\alpha \int_{\mathcal{X}} c(s)\mu^n(ds)\right\}.$$

Under suitable conditions which we will state in next part, $\{\mu^n, n \in \mathbb{N}\}$ will satisfy a large deviation principle on $\mathcal{P}(\mathcal{X})$ and have a rate function I, then we can use

Varadhan's lemma to get the following result:

$$\lim_{n\to\infty}\frac{1}{n}\log E_x \exp\left\{\alpha\int_0^n c(X(s))ds\right\} = \sup_{\gamma\in\mathcal{P}(\mathcal{X})}\left\{\alpha\int_{\mathcal{X}} cd\gamma - I(\gamma)\right\}.$$

2.2 Feller Property

In this section we introduce a common condition for a Markov process that assumes a large deviation principle.

Condition 2.2.1. Assume $\mathcal{X}$ is a compact Polish space, and that $\{X(t)\}_{t\in[0,\infty)}$ is an ergodic Feller Markov process (1.3.13) with state space $\mathcal{X}$ and stationary distribution π. Assume also that the empirical measure of X satisfies the LDP in the topology of weak convergence with rate function $I : \mathcal{P}(\mathcal{X}) \to [0,\infty]$, uniformly with respect to $X(0) = x \in \mathcal{X}$ [4, Section 1.2], that $I(\mu) = 0$ if and only if $\mu = \pi$, that $c : \mathcal{X} \to \mathbb{R}$ is continuous, and that $\int_{\mathcal{X}} c(x)\pi(dx) = 0$.

Basic examples of process models that satisfy Condition 2.2.1 and which will also be used to show the numerical simulations later on are as follows.

Example 2.2.1. Let $\mathcal{X} = \{s_1, s_2, \ldots, s_K\}$ and suppose that $\{X(t), t \in [0,\infty)\}$ is a continuous time, pure jump, irreducible Markov process with jump rates $r_{x,y} \in [0,\infty), x, y \in \mathcal{X}, x \neq y$.

Example 2.2.2. Let $\mathcal{X}$ be a compact subset of a finite dimensional Euclidean space $\mathbb{R}^k$, and assume that $\mathcal{X}$ is the closure of its interior, that the interior is connected, and that $\mathcal{X}$ has smooth boundary. We then consider as a process model the solution of a stochastic differential equation with reflection on the boundary of $\mathcal{X}$ [20], and assume that the process has smooth (e.g., C^1 on an open neighborhood of $\mathcal{X}$) drift

$b : \mathcal{X} \to \mathbb{R}^k$, smooth and uniformly nondegenerate covariance matrix a mapping $\mathcal{X}$ into the space of $k \times k$ symmetric matrices, and smooth direction of constraint $\gamma : \partial\mathcal{X} \to \mathbb{R}^k$ pointing uniformly into the interior of $\mathcal{X}$.

Here is an example of Brownian motion with reflection at the boundary of a disk. Let $X \subset \mathbb{R}^2$ be a closed disk with center at the origin and radius R, defined as

$$X = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R^2\}.$$

Consider a particle moving randomly within X, where its trajectory is governed by a stochastic differential equation with reflection at the boundary. Specifically, the motion of this particle can be defined by the following stochastic differential equation:

$$dZ_t = b(Z_t)\, dt + \sigma(Z_t)\, dW_t + dL_t,$$

where:

- Z_t represents the particle's position at time t;
- $b : X \to \mathbb{R}^2$ is a drift function that describes the particle's tendency or directional movement;
- $\sigma : X \to \mathbb{R}^{2\times 2}$ is a covariance matrix function that describes the intensity of random fluctuations;
- W_t is a two-dimensional Brownian motion;
- L_t is a reflection process at the boundary ∂X to ensure the particle remains within X.

The reflection direction is typically defined as the inward-pointing unit vector, meaning that on the boundary L_t points from the circumference toward the center of the disk.

The following special case of the first example will be used for certain exact calculations and to develop some explicit expressions to help with interpreting the results. The calculations will appear in the next section.

Example 2.2.3. Consider the special case where $\mathcal{X} = \{-1, 1\}$, $r_{-1,1} = r_{1,-1} = 1$ and $c(x) = x$. Here of course $\pi(-1) = \pi(1) = 1/2$, so that $c(-1)\pi(-1)+c(1)\pi(1) = 0$

2.3 Space and Time dependent Thresholds

We consider the approximation of

$$M^n(\alpha) = E_x e^{n\alpha Y^n(1)}, \tag{2.3.1}$$

where

$$Y^n(t) = \frac{1}{n}\int_0^{nt} c(X(s))ds,$$

and $\{X(t)\}$ is an ergodic Markov process with state space $\mathcal{X}$. Here $\{X(t)\}$ satisfies Condition (2.2.1), and we want use the large deviation results to design the thresholds of a splitting scheme. We will consider schemes defined in terms of thresholds of the general form

$$C_j^n \doteq \{(y,t) \in \mathbb{R} \times [0,1] : V_B(y,t) \leq V_B(0,0) - j\Delta/n\}, \quad j = 0, 1, \ldots, K^n, \tag{2.3.2}$$

where V_B is of the form

$$V_B(y,t) = -\alpha y + B[t-1]. \tag{2.3.3}$$

The form (2.3.3) is motivated by the fact that the solution to a Hamilton-Jacobi-Bellman equation characterizes the asymptotics of $M^n(\alpha)$ through large deviation theory, and to be specific the equation that characterizes the large deviation limit

$$H(\alpha) \doteq \lim_{n\to\infty} -\frac{1}{n}\log M^n(\alpha),$$

as well as intermediate quantities of the form $E_x[e^{n\alpha Y^n(1)}|Y^n(t)=y]$, takes this form when $B = H(\alpha)$.

Remark 2.3.1. To show that $V_B(y,t)$ is the solution to a Hamilton-Jacobi-Bellman equation, we can derive the equation using the standard formal argument. Let $E_{y,x,t}[e^{n\alpha Y^n(1)}]$ denote $E[e^{n\alpha Y^n(1)}|Y^n(t)=y,\ X(t)=x]$. If Y satisfies an uniform large deviation principle with rate function $J(\phi) = \int_0^1 L(\dot\phi(t))dt$, then by Varadhan's lemma,

$$\begin{aligned}
V(y,t) &\doteq \lim_{n\to\infty} -\frac{1}{n}\log(E_{y,x,t}[e^{n\alpha Y^n(1)}]) \\
&= \inf_\phi \left[-\alpha\phi(1) + \int_t^1 L(\dot\phi(s))ds \middle| \phi(t)=y\right] \\
&= \inf_\phi \left[-\alpha\phi(1) + \int_t^{t+\delta} L(\dot\phi(s))ds + \int_{t+\delta}^1 L(\dot\phi(s))ds \middle| \phi(t)=y\right] \\
&= \inf_\phi \left(\left[-\alpha\phi(1) + \int_{t+\delta}^1 L(\dot\phi(s))ds \middle| \phi(t)=y\right] + \left[\int_{t+\delta}^t L(\dot\phi(s))ds \middle| \phi(t)=y\right]\right)
\end{aligned}$$

The second equality comes from uniform LDP. We denote $\beta = \dot\phi(t)$, then if δ small enough we can have

$$\begin{aligned}
&\inf_\phi \left[-\alpha\phi(1) + \int_{t+\delta}^1 L(\dot\phi(s))ds \middle| \phi(t)=y\right] \\
\approx &\inf_\phi \left[-\alpha\phi(1) + \int_{t+\delta}^1 L(\dot\phi(s))ds \middle| \phi(t+\delta)=y+\delta\beta\right]
\end{aligned}$$

$$= V(y + \delta\beta, t + \delta).$$

Thus, we can get

$$V(y,t) \approx \inf_{\phi}[\delta L(\beta) + V(y + \delta\beta, t + \delta)].$$

Expanding $V(y + \delta\beta, t + \delta)$ using Taylor series, we get

$$V(y + \delta\beta, t + \delta) \approx V(y,t) + \delta\beta\frac{\partial V}{\partial y}(y,t) + \delta\frac{\partial V}{\partial t}(y,t).$$

Thus,

$$V(y,t) \approx \inf_{\beta}\left[\delta L(\beta) + V(y,t) + \delta\beta\frac{\partial V}{\partial y}(y,t) + \delta\frac{\partial V}{\partial t}(y,t)\right].$$

Moving $V(y,t)$ to the right side and dividing by δ, then taking the limit as $\delta \to 0$

$$\frac{\partial V}{\partial t}(y,t) + \inf_{\beta}\left\{L(\beta) + \beta\frac{\partial V}{\partial y}(y,t)\right\} = 0.$$

We also have the terminal condition $V(y,1) = -\alpha y$. We can use verification method to prove that $V(y,t) = -\alpha y + H(\alpha)(t-1)$ is the unique solution to the Bellman equation.

Example 2.3.1 (Example 2.2.3 cont'd)**.** Here we give the calculation for the simple case where $\mathcal{X} = \{-1, 1\}, r_{-1,1} = r_{1,-1} = 1$ and $c(x) = x$.

We define

$$Y^n(t) = \frac{1}{n}\int_0^{nt} c(X(s))ds,$$

where $c(x) = x$. Let

$$\mu^n(A) = \frac{1}{n}\int_0^{n} 1_A(X(s))ds.$$

Suppose γ is a probability on $\{-1, 1\}$, and let

$$\sigma(-1) = \gamma(-1)/(1/2), \quad \sigma(1) = \gamma(1)/(1/2),$$

so $\sigma = d\gamma/d\pi$. $\pi(1) = \pi(-1) = \frac{1}{2}$ is the stationary distribution for the Markov Process. Then the rate function for $\{\mu_n\}$ is [11]

$$\begin{aligned} I(\gamma) &= 1 - \frac{1}{2}\sigma^{1/2}(-1)\sigma^{1/2}(1) - \frac{1}{2}\sigma^{1/2}(1)\sigma^{1/2}(-1) \\ &= 1 - \sigma^{1/2}(-1)\sigma^{1/2}(1) \\ &= 1 - \sqrt{4\gamma(-1)\gamma(1)}. \end{aligned}$$

Using the contraction principle, the rate function for $Y^n(1) = \int c(y)\mu^n(dy)$ is

$$J(b) = \inf\left\{ I(\gamma) : \int c(y)\gamma(dy) = b \right\}.$$

Also, Varadhan's lemma (1.3.10) says that

$$\begin{aligned} \frac{1}{n}\log E e^{n\alpha Y^n(1)} &\to \sup_b \{\alpha b - J(b)\} \\ &= \sup_b \left\{ \alpha b - \inf_\gamma \left\{ I(\gamma) : \int c(y)\gamma(dy) = b \right\} \right\} \\ &= \sup_\gamma \left\{ \alpha \int c(y)\gamma(dy) - I(\gamma) \right\}. \end{aligned}$$

We can find β^* explicitly by solving

$$\sup_{\gamma(1)} \left\{ \alpha\left[c(-1)(1-\gamma(1)) + c(1)\gamma(1)\right] - 1 + \sqrt{4(1-\gamma(1))\gamma(1)} \right\}$$

for $\gamma^*(1)$, and then let

$$\beta^* = \left[c(-1)(1-\gamma^*(1)) + c(1)\gamma^*(1)\right]$$

Next we calculate the exact value of β^*. We define

$$F(\theta) = \alpha\left[c(-1)(1-\theta) + c(1)\theta\right] - 1 + \sqrt{4(1-\theta)\theta}$$

$$= \quad 2\alpha\theta - \alpha - 1 + \sqrt{4(1-\theta)\theta}.$$

Then take the derivative of $F(\theta)$

$$F^{'}(\theta) = 2\alpha + \frac{1-2\theta}{\sqrt{(1-\theta)\theta}}.$$

In order to find the maximal of $F(\alpha)$, we let $F^{'}(\theta) = 0$. By solving this equation, we know when $\theta = \frac{1}{2} + \frac{\alpha\sqrt{1+\alpha^2}}{2(1+\alpha^2)}$, $F(\alpha)$ reaches the maximal. In this situation,

$$\beta^* = \frac{\alpha\sqrt{1+\alpha^2}}{(1+\alpha^2)},$$

$$H(\alpha) = \sqrt{1+\alpha^2} - 1.$$

■

The schemes we propose are novel, in that the splitting thresholds depend on time and the y-coordinate. (For comparison, existing interacting particle type schemes for the approximation of (2.3.1) as in [23] use thresholds that depend only on the t coordinate.) The form (2.3.3) of V_B, and in particular the parameter B, will allow us to tie the performance of the interacting particle scheme to the large deviation properties of $\{Y^n(\cdot)\}$. The sign of B will depend on the sign of α and the law of large numbers behavior of $\{Y^n(1)\}$. To simplify the description of the scheme we assume that $Y^n(1) \to 0$ in probability (or equivalently that $\int c d\pi = 0$ where π is the stationary distribution of X) and that $\alpha > 0$, in which case the appropriate values for B will be positive. The geometry of the thresholds is then as indicated in Figure 2.1, with β^* and h quantities that will be explained later on.

We will use the following.

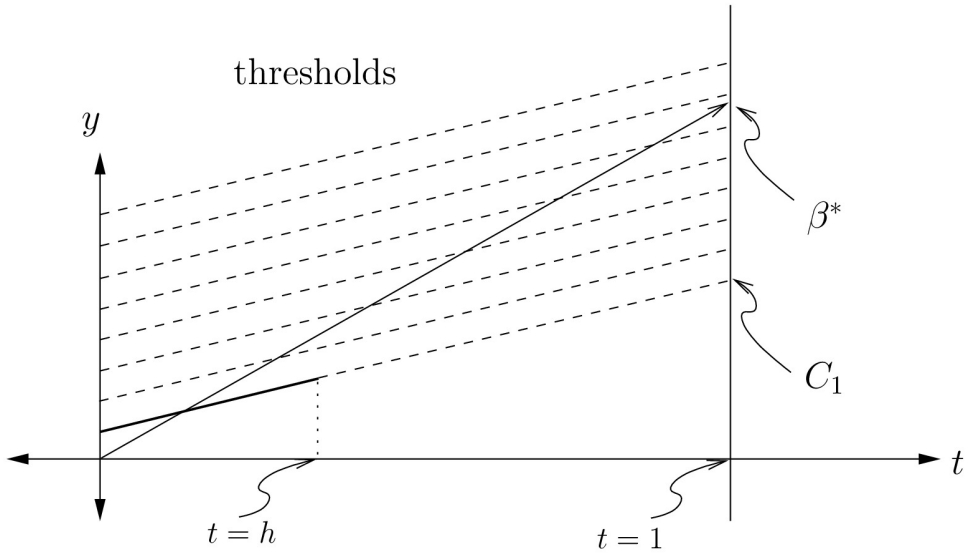

Figure 2.1: Resampling thresholds

Recall **Condition** 2.2.1: Assume $\mathcal{X}$ is a compact Polish space, and that $\{X(t)\}_{t\in[0,\infty)}$ is an ergodic Feller Markov process with state space $\mathcal{X}$ and stationary distribution π. Assume also that the empirical measure of X satisfies the large deviation principle (LDP) in the topology of weak convergence with rate function $I : \mathcal{P}(X) \to [0,\infty]$, uniformly with respect to $X(0) = x \in \mathcal{X}$, that $I(\nu) = 0$ if and only if $\nu = \pi$, that $c : \mathcal{X} \to \mathbb{R}$ is continuous, and that $\int_X c(x)\pi(dx) = 0$.

Remark 2.3.2. The assumption that c be continuous is motivated by the fact that we assume the LDP for the empirical measure holds in the topology of weak convergence. However, it is often the case that the LDP holds under the stronger τ-topology, in which case we expect that it would be sufficient for the results of this paper that c be bounded and measurable. Also, the restriction that $\mathcal{X}$ be compact can be removed, but doing so requires strong recurrence properties of $\{X(t)\}_{t\in[0,\infty)}$ that are typically formulated in terms of an appropriate Lyapunov function, and a great deal more care is needed in describing uniformity properties of the estimates.

Since it does not depend on rare events in a sensitive way, the estimation of $\int_{\mathcal{X}} c(x)\pi(dx)$ is expected to be much easier than that of $M^n(\alpha)$. Thus we assume

that c has been translated so that its mean under π is zero.

2.4 Quasi-stationary distribution

In this section we define a quasi-stationary distribution that will be used in the analysis of the algorithm and related quantities. The reader interested in getting to the formulation of the algorithm can go directly from the statement of Lemma 2.4.4 to Section 3. Following the logic of splitting methods in general, the idea behind the scheme is to decompose the original expected value into a collection of conditional expected values. When designed well, each such expectation can be accurately estimated with relatively small effort, and calculations for the entire group can be structured to be recursive in the threshold.

To simplify the presentation, we first present the algorithm for $\alpha = 1$ and without the large deviation scaling parameter (i.e., $n = 1$) and prove unbiasedness. We also first describe the algorithm when $X(0)$ is distributed according to a quasi-stationary distribution λ that will be defined shortly. This distribution plays a key role in understanding the large deviation properties of the scheme. The case of general initial conditions is analogous (see Remark 3.1.2) and owing to strong uniformity properties with respect to initial conditions (geometric ergodicity) also satisfies the conclusions of Theorem 4.4.1.

Recall $Y(1) = \int_0^1 c(X(s))ds$. Assuming there are K thresholds, we decompose according to

$$E_\lambda e^{Y(1)} = E_\lambda \left[1_{A_1^c} e^{Y(1)}\right] + \sum_{k=1}^{K-1} E_\lambda \left[1_{A_{k+1}^c} 1_{A_k} e^{Y(1)}\right] + E_\lambda \left[1_{A_K} e^{Y(1)}\right], \tag{2.4.1}$$

where by construction $A_{k+1} \subset A_k$, and A_{k+1}^c will denote the complement of A_{k+1}, but *relative to* A_k: $A_{k+1}^c = A_k \backslash A_{k+1}$. The sets A_k will be defined in terms of the thresholds and a parameter $h \in (0, 1]$ (which will scale with n in the large deviation setting). The first set is defined according to

$$A_1 \doteq \{Y(t) \text{ first reaches } C_1 \text{ at some } \tau_1 \in [0, h] \text{ and } 1 - \tau_1 \geq h\},$$

and therefore A_1^c is the set where $Y(t)$ fails to reach C_1 before time h, or there is less than h units of time remaining in $[0, 1]$ after it first reached C_1. See Figure 2.1 above.

Remark 2.4.1. Here h is a parameter that needs to be chosen to define a stationary distribution for decomposition.

We introduce some terminology that will be needed later. For a given initial condition $(Y(0), X(0))$, where $Y(0) = 0$ and $X(0)$ is distributed according to λ (and as usual independent of the future evolution of the process), we say that Y is a **success particle** if $\omega \in A_1$, and that it is a **failure particle** if $\omega \in A_1^c$. Note that success particles must reach the next threshold before time h has expired. Owing to the fact that $\int c d\mu = 0$ and since the geometry is as depicted in Figure 2.1, the probability of a success particle diminishes as C_1 is translated further from the origin. Given a deterministic initial condition $x \in \mathcal{X}$, we can define a subprobability measure $\theta(\cdot|x)$ on

$$Z \doteq \mathcal{X} \times W, \quad W \doteq \{(y, t) \in C_1 : t \in [0, h]\}$$

according to

$$\theta(A \times B|x) \doteq P_x\{X(\tau) \in A, (Y(\tau), \tau) \in B\},$$

where $\tau \doteq \inf\{t > 0 : (Y(t), t) \in C_1\}$.

Condition 2.4.2. Let $q(x, du) \doteq \theta(du \times W|x)$. Then there is a reference probability measure $\varrho \in \mathcal{P}(\mathcal{X})$ and $0 < \bar{c}_1 < \bar{c}_2 < \infty$ such that

$$\bar{c}_1 \varrho(du) \leq q(x, du) \leq \bar{c}_2 \varrho(du) \tag{2.4.2}$$

for any $x \in \mathcal{X}$. Assume also that there is $c_2 < 1$ such that $q(x, \mathcal{X}) \leq c_2$ for all $x \in \mathcal{X}$, and $q(x, du)$ indecomposable and Feller: for all bounded and continuous $f : \mathcal{X} \to \mathbb{R}$ the mapping

$$x \to \int_{\mathcal{X}} f(u) q(x, du)$$

is continuous. Lastly, we assume that for every $x \in \mathcal{X}$, $P_x\{c(X(\tau)) < B/\alpha, \tau \leq h\} = 0$.

The last part of the condition is mild, and requires that at the moment a success particle reaches the next threshold its velocity is at least as large as the slope of the threshold.

The special distribution λ that will be needed is characterized by

$$\lambda(A) = \frac{\int_{\mathcal{X}} q(x, A) \lambda(dx)}{\int_{\mathcal{X}} q(x, \mathcal{X}) \lambda(dx)} \tag{2.4.3}$$

for $A \in \mathcal{B}(\mathcal{X})$. If we append Λ to $\mathcal{X}$ and extend $q(x, du)$ to be a probability transition kernel by setting $q(x, \Lambda) \doteq 1 - q(x, \mathcal{X})$, then λ is a quasistationary distribution for the Markov process with this transition kernel and with respect to remaining in $\mathcal{X}$. In particular, if $X(0)$ has distribution λ and we condition on the event that the particle is a success particle, then $X(\tau)$ also has distribution λ. As we will see, λ must be concentrated on $\bar{\mathcal{X}} \doteq \{x \in \mathcal{X} : c(x) \geq B/\alpha\}$. We first discuss the existence and uniqueness as well as other properties of λ, and then show how one can verify

the conditions we assume for Examples 2.2.1 and 2.2.2.

Definition 2.4.3. (Successful State Space) $\bar{\mathcal{X}}$ is the subset of $\mathcal{X}$ that contains all x such that $c(x) \geq B/\alpha$,

$$\bar{\mathcal{X}} \doteq \{x \in \mathcal{X} : c(x) \geq B/\alpha\}. \tag{2.4.4}$$

Moreover, $\bar{\mathcal{X}}$ is compact.

Lemma 2.4.4. Assume Condition 2.4.2, and recall the definition $\bar{\mathcal{X}}$ (2.4.4). Then there is a unique probability distribution λ that satisfies (2.4.3), and this distribution is supported on $\bar{\mathcal{X}}$. Let $\mu_1 \in \mathcal{P}(\mathcal{X})$, and define $\mu_n \in \mathcal{P}(\mathcal{X})$ recursively by

$$\mu_{n+1} = g(\mu_n), \text{ where } g(\mu)(A) = \frac{\int_{\mathcal{X}} q(x, A)\mu(dx)}{\int_{\mathcal{X}} q(x, \mathcal{X})\mu(dx)} \tag{2.4.5}$$

for $A \in \mathcal{B}(\mathcal{X})$. Then $\mu_n(A) \to \lambda(A)$ for $A \in \mathcal{B}(\mathcal{X})$.

Proof. Using the condition $P_x\{c(X(\tau)) < B/\alpha, \tau \leq h\} = 0$, $\theta((\bar{\mathcal{X}})^c \times W|x) = 0$, and therefore $q(x, (\bar{\mathcal{X}})^c) = 0$. Although we define g for $\mu \in \mathcal{P}(\mathcal{X})$, it follows that λ and all $\mu_n, n > 1$ in the statement of the lemma must take values in $\mathcal{P}(\bar{\mathcal{X}})$, and without loss we can consider $q(x, du)$ as a transition kernel on $\bar{\mathcal{X}}$. Consider the eigenvalue problem

$$\alpha\lambda(A) = \int_{\bar{\mathcal{X}}} q(x, A)\lambda(dx), \quad A \in \mathcal{B}(\bar{\mathcal{X}}). \tag{2.4.6}$$

It follows from [19] that there is a positive eigenvalue α, larger in magnitude than any other eigenvalue, and a unique probability measure $\lambda \in \mathcal{P}(\bar{\mathcal{X}})$, with $\bar{v}(y) \doteq [d\lambda/d\varrho](y)$ uniformly bounded from above and from below away from zero on $\bar{\mathcal{X}}$, such that (2.4.6) holds. Letting $A = \bar{\mathcal{X}}$ we find that $\alpha = \int_{\bar{\mathcal{X}}} q(x, \bar{\mathcal{X}})\lambda(dx)$, and therefore

$$\lambda(A) = \frac{\int_{\bar{\mathcal{X}}} q(x, A)\lambda(dx)}{\int_{\bar{\mathcal{X}}} q(x, \bar{\mathcal{X}})\lambda(dx)}.$$

We extend λ to be a probability measure on $\mathcal{X}$ by assigning zero mass to $\mathcal{X}\backslash\bar{\mathcal{X}}$, in which case (2.4.3) follows.

Finally, we consider the last sentence of the lemma. Define $q^n(x, A)$ for $x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})$ recursively by

$$q^n(x, A) \doteq \int_{\mathcal{X}} q^{n-1}(x, dz)q(z, A).$$

By Theorem 10.1 in [19], q has a positive eigenvalue ρ that is larger in magnitude than any other eigenvalue, and unique (up to a multiplicative constant) bounded left and right eigenfunctions v and u, such that

$$q^n(x, dy) = \rho^n u(x)v(y)\varrho(dy)[1 + O(\Delta^n)], \tag{2.4.7}$$

where $0 < \Delta < 1$ is independent of x and y. Since (2.4.6) implies $\bar{v}(y) \doteq [d\lambda/d\varrho](y)$ is also a left eigenfunction, by uniqueness we have $\rho = \alpha$ and $\bar{v}(y) = v(y)/\int_{\mathcal{X}} v(z)\varrho(dz)$.

Suppose we start with an arbitrary distribution μ_1. Then by (2.4.2) we can find bounded $r(x)$ such that

$$\mu_2(dx) = g(\mu_1)(dx) = r(x)\varrho(dx).$$

Applying this recursively we find

$$\begin{aligned}
\mu_n(dy) = g(\mu_{n-1})(dy) &= \frac{\int_{\bar{\mathcal{X}}} q(x, dy)\mu_{n-1}(dx)}{\int_{\bar{\mathcal{X}}} q(x, \bar{\mathcal{X}})\mu_{n-1}(dx)} \\
&= \frac{\int_{\bar{\mathcal{X}}} q^{n-1}(x, dy)\mu_2(dx)}{\int_{\bar{\mathcal{X}}} q^{n-1}(x, \bar{\mathcal{X}})\mu_2(dx)} \\
&= \frac{\int_{\bar{\mathcal{X}}} q^{n-1}(x, dy)r(x)\varrho(dx)}{\int_{\bar{\mathcal{X}}} q^{n-1}(x, \bar{\mathcal{X}})r(x)\varrho(dx)}
\end{aligned}$$

$$= \frac{\rho^{n-1}\int_{\bar{\mathcal{X}}} u(x)v(y)r(x)\varrho(dx)\varrho(dy)[1+O(\Delta^{n-1})]}{\rho^{n-1}\int_{\bar{\mathcal{X}}}\int_{\bar{\mathcal{X}}} u(x)v(y)r(x)\varrho(dx)\varrho(dy)[1+O(\Delta^{n-1})]},$$

where the last equality comes from eqaution (2.4.7), so that

$$\mu_n(A) = \frac{\int_A v(y)\varrho(dy)\int_{\bar{\mathcal{X}}} u(x)r(x)\varrho(dx)[1+O(\Delta^{n-1})]}{\int_{\bar{\mathcal{X}}} v(y)\varrho(dy)\int_{\bar{\mathcal{X}}} u(x)r(x)\varrho(dx)[1+O(\Delta^{n-1})]}.$$

Since Δ is independent of x and y,

$$\mu_n(A) \to \frac{\int_A v(y)\varrho(dy)\int_{\bar{\mathcal{X}}} u(x)r(x)\varrho(dx)}{\int_{\bar{\mathcal{X}}} v(y)\varrho(dy)\int_{\bar{\mathcal{X}}} u(x)r(x)\varrho(dx)} = \frac{\int_A v(y)\varrho(dy)}{\int_{\bar{\mathcal{X}}} v(y)\varrho(dy)} = \int_{\mathcal{X}} \bar{v}(y)\varrho(dy) = \lambda(A).$$

□

Lemma 2.4.5. Suppose there are $z_i^* \in \mathcal{X}, i = 1, 2$, with $c(z_1^*) > (B + \Delta/h)/\alpha$ and $c(z_2^*) < (B + \Delta/h)/\alpha$. Then Condition 2.4.2 holds for Examples 2.2.1 and 2.2.2.

Remark 2.4.6. The existence of z_1^* means that there is a state such that the growth rate of integral increment (i.e. $c(z_1^*)$) is greater that the slope of the threshold, and the existence of z_2^* means that there is a state such that the growth rate of integral increment is lower that the slope of the threshold.

Proof. We give the proof only for the harder case of Example 2.2.2. The interested reader can easily fill in the details of the pure jump example. We start with the observation that owing to $\|c\|_\infty < \infty$, there is $\delta > 0$ such that no matter what the starting state $x \in \mathcal{X}$, $\tau \le \delta$ is impossible. We also note that, owing to the compactness of $\mathcal{X}$, nondegeneracy of the diffusion coefficient, smoothness of the problem data (drift, boundary, etc.) and standard properties of nondegenerate diffusions, X has a jointly continuous transition density with respect to Lebesgue measure, and there exist $0 < a < A < \infty$, $B < \infty$ such that if $p_\delta(x, z)dz = P(X(\delta) \in dz | X(0) = x)$,

then $p_\delta(x,z) \le B$ for all $x, z \in \mathcal{X}$, and for all $x_1, x_2 \in \mathcal{X}$

$$ap_\delta(x_1,z) \le p_\delta(x_2,z) \le Ap_\delta(x_1,z). \tag{2.4.8}$$

This shows that all initial conditions are in a strong sense comparable.

The existence of z_1^* means that there is a state such that, if the process stays in the state for $t \in [0,h]$, then $c(z_1^*) > (B+\Delta/h)/\alpha$ will lead to $\tau \le h$. Since c is continuous the same will be true for a ε-neighborhood of z_1^* for some $\varepsilon > 0$. Recall $\theta(\cdot|x)$ is a subprobability measure on $\mathcal{X} \times W$, where $W = \{(y,t) \in C_1 : t \in [0,h]\}$. Again using standard properties of diffusion processes, $\theta(\mathcal{X} \times W|x) \ge c_1 > 0$ for all $x \in \mathcal{X}$ follows. A similar argument using z_2^* shows that $\theta(\mathcal{X} \times W|x) \le c_2 < 1$ for all $x \in \mathcal{X}$. In order to *first* reach C_1 at time τ it must be true that

$$\int_{\tau-\sigma}^{\tau} c(X(s))ds > \frac{\sigma B}{\alpha}$$

for all $\sigma > 0$. This is impossible if $c(X(\tau))ds < B/\alpha$, and so q must be supported on $\bar{\mathcal{X}}$ (Definition 2.4.4).

Let $q_{h-\delta}(z,B) \doteq P\{X(\tau) \in B, \tau \le h | X(\delta) = z\}$. Then by the Markov property (recall that $\tau \le \delta$ is impossible)

$$q(x,du) = \int_{\mathcal{X}} q_{h-\delta}(z,du)p_\delta(x,z)dz.$$

If $f : \mathcal{X} \to \mathbb{R}$ is bounded and continuous then since $\int_{\mathcal{X}} f(u)q_{h-\delta}(z,du)$ is bounded and measurable, the continuity of

$$x \to \int_{\mathcal{X}} \int_{\mathcal{X}} f(u)q_{h-\delta}(z,du)p_\delta(x,z)dz$$

follows from dominated convergence. Thus $q(x, du)$ is Feller. Also, we note that owing to (2.4.8) $\varrho(du) = q(z, du)$ for any value of z serves as a reference measure of the form (2.4.2) for $q(x, du)$ for all $x \in \mathcal{X}$, and it follows from this that indecomposability holds. □

Remark 2.4.7. Owing to the scaling used later on, the fact that h will tend to ∞ means that the analogous condition for the n-dependent sequence regarding the existence of $z_i^*, i = 1, 2$, will be true if it is true when $h = \infty$.

Example 2.4.1 (Example 2.2.3 cont'd)**.** For Example 2.2.3 it must be the case that $X(\tau) = 1$ since when reaching a new threshold it must be that the derivative of Y is positive. Thus $q(-1, \{1\}) = q(1, \{1\}) = 1$, and therefore $(\lambda(\{-1\}), \lambda(\{1\})) = (0, 1)$.

We continue with the definition of the sets A_k that will be used in decomposition (2.4.1). Given A_k, let

$$A_{k+1} \doteq A_k \cap \{Y(t) \text{ first reaches } C_{k+1} \text{ at some } \tau_{k+1} \in [\tau_k, \tau_k + h] \text{ and } 1 - \tau_{k+1} \geq h\}.$$

We also define $A_{k+1}^c = A_k \backslash A_{k+1}$. This is done until we reach $k+1 = K$. If $\omega \in A_{k+1}^c$ then the following hold. (i) Y reached each threshold up to the kth within h units of time (after reaching the previous threshold), and before the terminal time 1. (ii) Subsequently, the particle either failed to reach threshold $k+1$ within h units of time after reaching C_k and before the terminal time 1, or after having reached C_{k+1} there were less than h units of time remaining. See Figure 2.2 above. We reiterate that, by construction, the X component of any success particle at any threshold is always distributed according to λ. The particles in A_{k+1}^c **failed** at threshold $k+1$, while the particles in A_{k+1} are **success** particles. By definition, $A_{K+1} = \emptyset$ (since there is no next threshold to reach), all particles in A_K will end up **failure** particles ($A_K = A_{K+1}^c$).

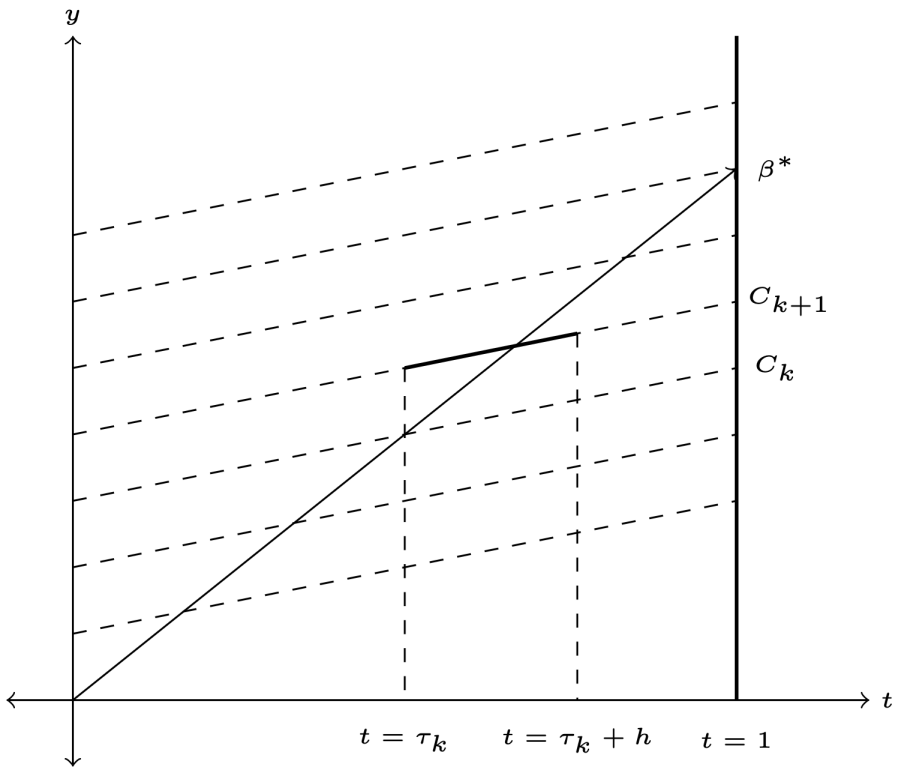

Figure 2.2: Success particles in A_{k+1}

We will need to describe how each of the terms in (2.4.1) can be approximated. While this will be done in detail in the next section, we make some remarks here. For some $k = 1, \ldots, K-1$ consider the term $E[1_{A_{k+1}} 1_{A_k^c} e^{Y(1)}]$. To estimate this expected value, we will need some random variables that estimate $E[1_{A_k^c} e^{Y(\tau_k)}]$, which involves the distribution of a particle that was successful for the first k thresholds. In particular, if say Q_k is a random variable used to estimate this quantity, then we will also want there to be random variables $(\bar{X}, \bar{Y}, \bar{\tau})$ such that the joint distribution of

$$\left(1_{A_k^c} e^{Y(\tau_k)}, X(\tau_k), Y(\tau_k), \tau_k\right)$$

is the same as that of

$$(Q_k, \bar{X}, \bar{Y}, \bar{\tau}).$$

To be precise, the variables $\bar{X}, \bar{Y}, \bar{\tau}$ will be used to provide a proper initial condition

for the algorithm to use for simulations that will take the algorithm to the next threshold, and the random variables will be constructed in the algorithm. As we will see, they can be used to construct unbiased estimators of $E[1_{A_{k+1}} 1_{A_k^c} e^{Y(1)}]$ and $E[1_{A_{k+1}^c} e^{Y(\tau_{k+1})}]$, and furthermore they can be implemented in a recursive way.

Chapter Three

The Detailed Description of Algorithm

Recall the decomposition presented in (2.4.1)

$$E_\lambda e^{Y(1)} = E_\lambda \left[1_{A_1^c} e^{Y(1)}\right] + \sum_{k=1}^{K-1} E_\lambda \left[1_{A_{k+1}^c} 1_{A_k} e^{Y(1)}\right] + E_\lambda \left[1_{A_K} e^{Y(1)}\right].$$

In this section, we will give a detailed description of the scheme. We will also discuss the initial distribution and resampling method to better understand the design of this estimator.

Notation Remark. To keep the two discrete indices clear, we use k as a *subscript* for threshold index, and j as a *superscript* for the particle index. The main

3.1 The Iterative Estimator

Step 1: Estimate $E_\lambda \left[1_{A_1^c} e^{Y(1)}\right]$

Recall A_1^c is the set where $Y(t)$ fails to reach C_1 before time h, or there is less than h units of time remaining in $[0,1]$ after it first reached C_1. We start with N particles (Y^j, X^j) with the desired initial distribution. To be precise, we have $Y^j = 0$ and X^j distributed according to λ for $j = 1, \ldots, N$. We will assume the X^j are independent, though this is not needed (and certainly the X components may be correlated at later thresholds). With an abuse of notation we denote this initialization for the N-particle system also by E_λ. We simulate independent trajectories according to the law of $X(\cdot)$, and define

$$\tau^j \doteq \inf\{t \geq 0 : Y^j(t) \in C_1\}.$$

Let

$$S_1 = \{j \in \{1, \ldots, N\} : \tau^j \leq h \text{ and } 1 - \tau^j \geq h\} \text{ and } F_1 = S_1^c.$$

Since A_1^c means either the particle was not successful in reaching the first threshold before h, or there was less than h units of time remaining, an unbiased estimate of $E_\lambda \left[1_{A_1} e^{Y(1)}\right]$ is

$$M_1 \doteq \frac{1}{N} \sum_{j \in F_1} e^{Y^j(1)}.$$

Let $N_1^S \doteq \#|S_1|$, the number of elements of S_1. We also record for future use the random variables

$$D_1 \doteq \left\{(R_1, X^j(\tau^j), Y^j(\tau^j), \tau^j), j \in S_1\right\},$$

where $R_1 \doteq N_1^S/N$. The distribution of $Y^j(\tau^j)$, given that it is a success particle, is independent of the total number of success particles. Hence

$$\begin{aligned} E_\lambda \left[R_1 e^{Y^j(\tau^j)}\right] &= E_\lambda \left[e^{Y^j(\tau^j)} \middle| j \in S_1\right] E_\lambda \left[N_1^S/N\right] \\ &= E_\lambda \left[e^{Y(\tau_1)} \middle| A_1\right] P_\lambda(A_1) = E_\lambda \left[1_{A_1} e^{Y(\tau_1)}\right] \end{aligned}$$

i.e., $R_1 e^{Y^j(\tau^j)}$ is an unbiased estimator of $E_\lambda[1_{A_1} e^{Y(\tau_1)}]$.

Step 2: Estimate $E_\lambda \left[1_{A_{k+1}^c} e^{Y(1)}\right]$

Recall A_{k+1}^c is the set where $Y(t)$ satisfied the following conditions: (i) Y reached each threshold up to the kth within h units of time (after reaching the previous threshold), and before the terminal time 1. (ii) Subsequently, the particle either failed to reach threshold $k+1$ within h units of time after reaching C_k and before the terminal time 1, or after having reached C_{k+1} there was less than h units of time remaining.

The input to stage $k+1$ of Step 2 will be random variables

$$D_k \doteq \left\{(R_k, X^i, Y^i, \tau^i), i \in \{1, \ldots N_k\}\right\},$$

with $N_k \leq N$. These correspond to success particles from the previous stage. (If $N_k = 0$ we can terminate the algorithm and define an estimator.) We resample these using any preferred method so long as it is independent of all past data to make a total of N, and call the *new* random variables (with some abuse of notation)

$$\bar{D}_k \doteq \left\{(R_k, X^j, Y^j, \sigma^j), j \in \{1, \ldots N\}\right\}.$$

The particular resampling used will not affect unbiasedness, though it can affect the variance of the estimate. For example, we could use $\lfloor N/N_k \rfloor$ copies of each element of D_k, and then use uniform iid sampling for the remaining $N - \lfloor N/N_k \rfloor N_k$ particles. We list a few resampling schemes in a remark at the end of this section. For each particle, with the data X^j, Y^j, σ^j being used to define the initial condition, we simulate trajectories

$$X^j(t), Y^j(t), \quad t \geq \sigma^j,$$

and define

$$\tau^j \doteq \inf\{t \geq \sigma^j : Y^j(t) \in C_{k+1}\}.$$

Let

$$S_{k+1} = \{j : \sigma^j \leq 1-h,\ \tau^j - \sigma^j \leq h, \text{ and } 1 - \tau^j \geq h\} \text{ and } F_{k+1} = S_{k+1}^c.$$

Thus to be in S_{k+1} the particle must start from ∂C_k before time $1-h$, reach C_{k+1} in no more than h units of time, and leave h units of time remaining. Then an unbiased

estimate of $E_\lambda[1_{A_{k+1}^c} e^{Y(1)}]$ is

$$M_{k+1} \doteq \frac{R_k}{N} \sum_{j \in F_k} e^{Y^j(1)}. \tag{3.1.1}$$

Let $R_{k+1} \doteq R_k N_{k+1}^S / N$, where now $N_{k+1}^S = \#\,|S_{k+1}|$. We record for future use the random variables

$$D_{k+1} \doteq \left\{ (R_{k+1}, X^j(\tau^j), Y^j(\tau^j), \tau^j), j \in S_{k+1} \right\},$$

and note that

$$R_{k+1} e^{Y^j(\tau^j)}$$

is an unbiased estimate of $E_\lambda[1_{A_{k+1}} e^{Y(\tau_{k+1})}]$. Note that we can also write $R_{k+1} e^{Y^j(\tau^j)}$ as

$$\left(\prod_{\ell=1}^{k+1} \frac{N_\ell^S}{N} \right) e^{Y^j(\tau^j)}.$$

Step 3: Estimate the $E\left[1_{A_{K+1}^c} e^{Y(1)}\right]$

A_{K+1}^c is the set where $Y(t)$ satisfied the following conditions: (i) Y reached each threshold up to the Kth within h units of time (after reaching the previous threshold), and before the terminal time 1. (ii) All particles will be failed particles and reach threshold at $t = 1$. In other words, all particles fall into F_{K+1}. We call this estimator M_{K+1}.

Overall estimator

The overall estimator is given by:

$$\sum_{k=1}^{K+1} M_k,$$

and as previously discussed, the performance of the estimator is determined by its second moment. When analyzing this second moment, we note that it can be decomposed into terms of the following form:

$$E_\lambda \left[\left(\prod\nolimits_{\ell=1}^{l} \frac{N_\ell^S}{N} \right) e^{Y^j(1)} \left(\prod\nolimits_{\ell=1}^{m} \frac{N_\ell^S}{N} \right) e^{Y^i(1)} 1_{\{j \in F_l, i \in F_m\}} \right]. \tag{3.1.2}$$

so that $Y^j(1)$ identifies the terminal location of a particle that became a failure particle when attempting to transition from threshold l to threshold $l+1$, and whose history prior to this threshold consisted of all success particles, and likewise $Y^i(1)$ identifies the terminal location of a particle that became a failure particle at threshold $m \leq l$.

Remark 3.1.1 (role of definition of success particles)**.** The requirement to reach the next threshold by time h, and also that there is at least h units of time remaining, is so there is a type of stationarity with respect to threshold index. This allows for the existence of the stationary distribution λ, which plays a key role in the analysis of the algorithm. One can write an unbiased estimator where these are not required. However, owing to the scaling used allowing the particle more time to reach the next threshold will not change the number of success particles greatly, since the overwhelming fraction of successes particles will end up crossing before h. On the other hand, dropping the requirement that there be at least h units of time before time 1 would likely not change the asymptotic theory since it will affect only the last few thresholds, and would likely lead to better performance. Lastly we note

the curious fact that, as α gets larger, in some sense the distribution λ gets simpler, since the set on which it is supported (i.e., $\bar{\mathcal{X}}$) gets smaller.

Remark 3.1.2 (general initial conditions)**.** The assumption that particles start in the distribution λ is also not needed. The only modification required to deal with an initial distribution $\vartheta \neq \lambda$ (including deterministic initial conditions) is that the collection of particle be started with the corresponding product measure on the $\mathcal{X}$ components, i.e., iid ϑ. With this modification the proof of unbiasedness is the same. If starting in another distribution the uniform ergodicity implied by Condition 4.1.1 stated below ensures that Theorem 4.4.1 is still valid.

Remark 3.1.3 (sampling schemes)**.** The main restriction of the resampling schemes we use is that they not depend on the states of the $\mathcal{X}$ components of success particles. Some choices are:

1. **IID Uniform Sampling**: In this scheme, we sample N particles from the N_k particles with replacement. Each particle has an equal chance of being selected, and the process repeats N times, ensuring that each particle is selected independently of others.

2. **Proportional Plus IID Uniform Sampling**: Here, we first use $\left\lfloor \frac{N}{N_k} \right\rfloor$ copies of each element from the set D_k. Then, for the remaining particles, we randomly sample $N - \left\lfloor \frac{N}{N_k} \right\rfloor N_k$ particles uniformly and independently from the remaining pool of particles. This strategy balances between proportional resampling and random selection, ensuring diversity in the particle set.

3. **Proportional Plus Uniform Without Replacement**: In this approach, we again use $\left\lfloor \frac{N}{N_k} \right\rfloor$ copies of each element from D_k. For the remaining particles, we choose them uniformly without replacement from all possible subsets of

exactly $N - \left\lfloor \frac{N}{N_k} \right\rfloor N_k$ particles within the N_k successful particles. This method avoids duplicates, ensuring that each of the remaining particles is distinct.

Chapter Four

Large Deviation Analysis

In this chapter we state our main result Theorem 4.4.1,which characterizes the second moment in terms of the key design parameter B and the number of particles. First we give the upper bound for the second moment of the estimator of (2.1.1). In the subsequent section we identify the key quantities whose large deviation limits make contribution to the estimator. Finally, we reintroduce the assumptions on the process models and discuss the choice of parameters.

4.1 Upper Bound for the Second Moment Estimator

We now reintroduce the large deviation parameter. Suppose we have K^n thresholds, where K^n scales proportionally with n (say κn), and Y is replaced by $n\alpha Y^n$. The parameter h is replaced by $\bar{h}/n^\eta$ with $\eta \in (0,1)$. We will want K^n to be large enough that the dynamics of "typical" success particles can reach time 1 before running out of thresholds, and not much larger. For example, as follows from the large deviations analysis presented below, the conditionally most likely velocity of success particles when we use $B = H(\alpha)$ to define the thresholds is $\beta^* = H'(\alpha)$, at least in the limit $n \to \infty$. Thus if we want the last threshold to just reach β^* at time 1 we will want

$$V_{H(\alpha)}(0,0) - V_{H(\alpha)}(\beta^*, 1) = \frac{\kappa n \Delta}{n} = \kappa\Delta,$$

where $V_{H(\alpha)}(y,t) = -\alpha y + H(\alpha)(t-1)$, and so, if only large n is considered one would choose κ close to

$$\frac{-H(\alpha) + \alpha\beta^*}{\Delta} = \frac{L(\beta^*)}{\Delta}.$$

However, when n is not too large, there can be a substantial contribution to the value being approximated due to trajectories for which $Y^n(1)$ is somewhat larger than β^*, and in this case one should choose κ to also be somewhat larger as well, say $\kappa = 1.2 \times (L(\beta^*)/\Delta)$. When $L(\beta^*)$ is not known a good value can be determined by performing a pilot run and examining the data to ensure typical success particles do not run out of thresholds before time $t = 1$.

The probability for the original particle to reach β^* at time 1 is approximately $e^{-nL(\beta^*)}$. Since we cross $nL(\beta^*)/\Delta$ thresholds to do this, the "typical" success probability should be about $e^{-\Delta}$. A more precise statement is that, if $\bar{p}^n$ is the probability of reaching the next threshold when the X component is in the stationary distribution λ^n [defined in (4.1.2)], then $\bar{p}^n = e^{-\Delta+o(1)}$ [see Lemma 7.3.4, and note that we assume here that $B = H(\alpha)$].

We note that the n that appears in the definition of the thresholds through (2.3.2) is to some degree canceled by the n scaling in the definition of Y^n. However, quantities such as $q(x, du)$ defined in (2.4.3) will depend on n owing to the replacement of h by $\bar{h}/n^\eta$. In other words, the time by which a particle must reach the next threshold to be declared a success particle depends on n. This leads to the introduction of the n-dependent versions of θ, q and λ from Section 3 that are defined below. However, when multiplied by n to offset the $1/n$ scaling used to define thresholds in the n-dependent sequence, we have $\bar{h}n^{1-\eta} \to \infty$, and hence these quantities will be essentially *independent* of n for large n.

The correct replacement for (2.4.3) in the asymptotic analysis is as follows. Let $C_\Delta \doteq \{(y,t) : V_B(y,t) \le V_B(0,0) - \Delta\}$, which is the same as C_1^1 when C_j^n is defined by (2.3.2). We let $W^n \doteq \{(y,t) \in C_\Delta : t \in [0, \bar{h}n^{1-\eta}]\}$ and $Z^n \doteq \mathcal{X} \times W^n$, and define

subprobability measures on Z^∞ according to

$$\theta(A_1 \times A_2|x) \doteq P_x\{X(\tau) \in A_1, (Y(\tau), \tau) \in A_2\}, \tag{4.1.1}$$

where $\tau \doteq \inf\{t > 0 : (Y(t), t) \in C_\Delta\}$. Let $q^n(x, du) \doteq \theta(du \times W^n|x)$. As in Section 3, the distribution λ^n is a probability measure that satisfies

$$\lambda^n(A) = \frac{\int_{\mathcal{X}} q^n(x, A)\lambda^n(dx)}{\int_{\mathcal{X}} q^n(x, \mathcal{X})\lambda^n(dx)}, \tag{4.1.2}$$

and under Condition 4.1.1 stated below it will be shown that this distribution is unique. Let $p^n(x) = q^n(x, \mathcal{X})$ be the probability Y reaches C_Δ before time $\bar{h}n^{1-\eta}$ when $X(0) = x$, and $\bar{p}^n = \int_{\mathcal{X}} p^n(x)\lambda^n(dx)$ be the averaged probability of success. The conditions we assume will impose $\bar{p}^n \in (0, 1)$. As noted previously, the n dependence of these quantities will disappear in the limit $n \to \infty$ owing to $\bar{h}n^{1-\eta} \to \infty$, and we let $W^\infty, Z^\infty, \lambda^\infty, q^\infty, p^\infty, \bar{p}^\infty$ denote the corresponding objects when $W^\infty \doteq \{(y, t) \in C_\Delta : t \in [0, \infty)\}$. Note that $\theta(A \times W^n|x) \uparrow \theta(A \times W^\infty|x)$ for each $x \in \mathcal{X}$ and $A \subset \mathcal{X}$ as $n \to \infty$. For convenience, we state the analogue of Condition 2.4.2 used in the asymptotic analysis. The same argument as that of Lemma 2.4.4 shows the condition is valid for interesting diffusion and pure jump models.

Condition 4.1.1. The following hold for all sufficiently large n. There is a reference probability measure $\varrho \in \mathcal{P}(\bar{\mathcal{X}})$ and $0 < \bar{c}_1 < \bar{c}_2 < \infty$ such that

$$\bar{c}_1\varrho(du) \le q^n(x, du) \le \bar{c}_2\varrho(du) \tag{4.1.3}$$

for $x \in \mathcal{X}$. There is $c_2 < 1$ such that $q^n(x, \mathcal{X}) \le c_2$ for all $x \in \mathcal{X}$, and $q^n(x, du)$ is indecomposable and Feller: for all bounded and continuous $f : \mathcal{X} \to \mathbb{R}$ the mapping

$$x \to \int_{\mathcal{X}} f(u)q^n(x, du)$$

is continuous. Lastly, we assume that for every $x \in \mathcal{X}$, $P_x\{c(X(\tau)) < B/\alpha, \tau \leq \bar{h}n^{1-\eta}\} = 0$.

With the algorithm as defined in the last section but with these adjustments, we designate the unbiased estimator of $M^n(\alpha) \doteq E_{\lambda^n} e^{n\alpha Y^n(1)}$ by

$$\mathcal{E}^n \doteq \sum_{k=1}^{K^n+1} M_k^n, \tag{4.1.4}$$

where we again abuse notation and let E_{λ^n} denote both that a single particle starts with distribution λ^n and also that the X component of each particle in an interacting particle system starts with distribution λ^n, independent from particle to particle.

4.1.1 Organization of the Large Deviations Analysis

There are certain scaling properties that it is convenient to note. With this in mind, we identify thresholds by quantities of the form $n\sigma_n$, where σ_n will be bounded, so that (at least along subsequences) σ_n converges to a limit $\bar{\sigma}$. Recall that a particle **fails** at threshold $n\sigma_n$ if it does not reach threshold $n\sigma_n + 1$ before time $\bar{h}n^{-\eta}$ or if it reaches the threshold after time $1 - \bar{h}n^{-\eta}$, and that $\eta \in (0, 1)$.

Note that it is only the location of failure particles that show up in the estimator $\mathcal{E}^n$, and since it is unbiased the rates of decay of the second moment characterize its accuracy (see the discussion below Condition 4.3.2). Regarding this decay rate, the next result shows that for upper bounds it is enough to obtain bounds for arbitrary pairs of failure particles, and for lower bounds the limiting (as $n \to \infty$) "worst case" pairs. The worst case pairs will be identified in Lemma 5.1.6 in terms of the explicit solution to a variational problem.

Lemma 4.1.2. Take as given arbitrary valid thresholds $n\sigma_n, n\tau_n \in \mathbb{Z}$, and let j (respectively l) denote a particle that fails at threshold $n\sigma_n$ (respectively $n\tau_n$). Let also $n\gamma_n \leq n\sigma_n \wedge n\tau_n$ denote the threshold at which the particles last had a common ancestor, with the convention that $n\gamma_n = 0$ if there is no common ancestor. Suppose there is $C \in (0, \infty)$ and for all $\delta > 0$ an integer $n_\delta < \infty$, such that for $n \geq n_\delta$ and for all such choices of thresholds, particles j, l, and $c_n \leq n\sigma_n \wedge n\tau_n$,

$$\frac{1}{n} \log E_{\lambda^n} \left[R^n_{n\sigma_n} e^{n\alpha Y^{j,n}(1)} R^n_{n\tau_n} e^{n\alpha Y^{l,n}(1)} 1_{\{j \in F^n_{n\sigma_n}, l \in F^n_{n\tau_n}, n\gamma_n = c_n\}} \right] \leq C + \delta \tag{4.1.5}$$

(see equation (3.1.2)). Then

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[\mathcal{E}^n\right]^2 \leq C. \tag{4.1.6}$$

Next suppose there exist a choice of thresholds and particles such that $j \in F^n_{n\sigma_n}, l \in F^n_{n\tau_n}, n\gamma_n = c_n$ and

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[R^n_{n\sigma_n} e^{n\alpha Y^{j,n}(1)} R^n_{n\tau_n} e^{n\alpha Y^{l,n}(1)} 1_{\{j \in F^n_{n\sigma_n}, l \in F^n_{n\tau_n}, n\gamma_n = c_n\}} \right] \geq C \tag{4.1.7}$$

for some $C \in (0, \infty)$. Then

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[\mathcal{E}^n\right]^2 \geq C. \tag{4.1.8}$$

Proof. Since the total number of thresholds is κn, $(\mathcal{E}^n)^2$ is no greater than N^{-2} times the sum of at most $\kappa^3 n^3$ terms of the form that appear in (4.1.5). Thus for $n \geq n_\delta$

$$E_{\lambda^n} \left[\mathcal{E}^n\right]^2 \leq \left(\frac{\kappa^3 n^3}{N^2} \right) e^{n(C+\delta)}.$$

Then (4.1.6) with C replaced by $C + \delta$ follows since $\frac{1}{n} \log n^3 \to 0$ as $n \to \infty$. Now send $\delta \to 0$.

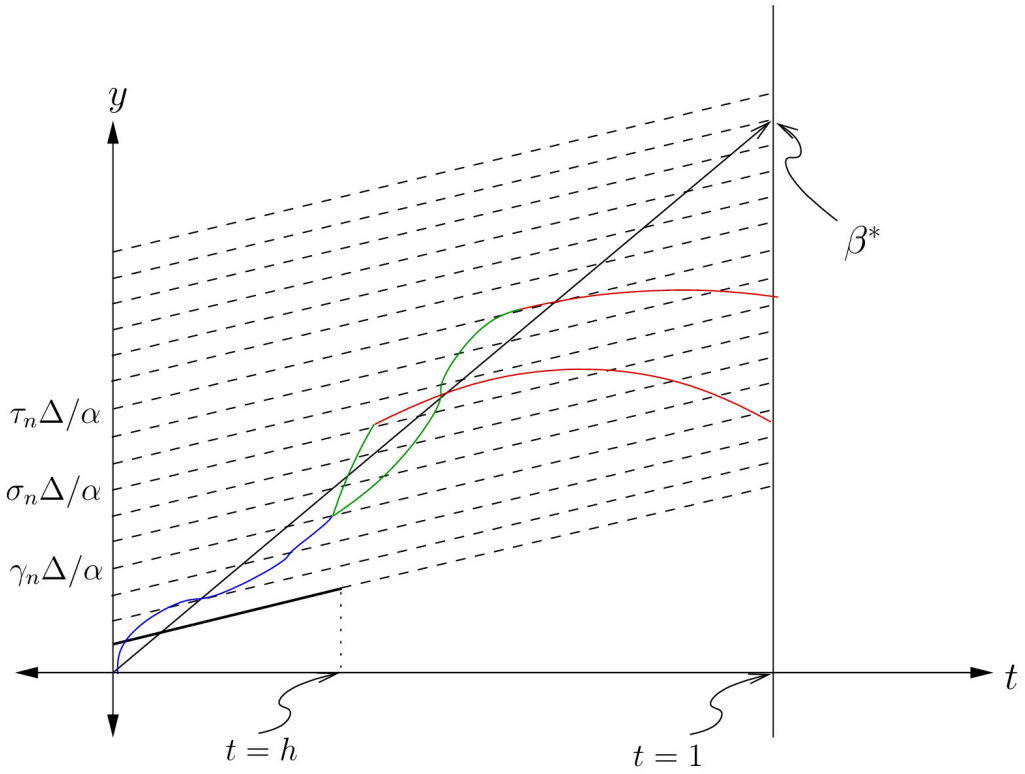

Figure 4.1: Evolution of particles

To prove (4.1.8) we use that $(\mathcal{E}^n)^2$ is bounded below by N^{-2} times the term

$$R^n_{n\sigma_n} e^{n\alpha Y^{j,n}(1)} R^n_{n\tau_n} e^{n\alpha Y^{l,n}(1)} 1_{\{j\in F^n_{n\sigma_n}, l\in F^n_{n\tau_n}, n\gamma_n = c_n\}}$$

appearing in (4.1.7). □

4.2 Decomposition of the Trajectories

Suppose $\partial C_a \doteq \{(y,t) \in \mathbb{R}\times[0,1] : V_B(y,t) - V_B(0,0) = -a\}$ and a is of of the form $\sigma_n\Delta$. Then the leftmost point $(y,0)$ of $\partial C_{\sigma_n\Delta}$ will satisfy $V_B(y,0) - V_B(0,0) = -\sigma_n\Delta$, and since $V_B(y,t) = -\alpha + B[t-1]$ we find $y = \sigma_n\Delta/\alpha$. Likewise the lefthand endpoints of thresholds numbered $n\gamma_n$ and $n\tau_2$ are $\gamma_n\Delta/\alpha$ and $\tau_n\Delta/\alpha$.

In Figure 4.1 we illustrate the various stages that lead to a pair of failure particles. The blue curve is the initial evolution of a particle. Owing to resampling it splits at threshold $n\gamma_n$ (which, as just noted, when rescaled has y-coordinate $\gamma_n\Delta/\alpha$ when $t = 0$), and the future independent evolution of two particles (so long as they are success particles) is illustrated by green curves. They eventually become failed particles (in the figure at different thresholds), and this portion of the evolution is depicted in red. In the figure both failure particles end at values lower than β^*. However, the goal for a well designed scheme is to produce at least some failure particles that end in a neighborhood of β^*, the location that contributes the most to $M^n(\alpha)$ when n is large.

As suggested by Lemma 4.1.2, typical terms that we will encounter in the analysis of the second moment are of the form

$$R^n_{n\sigma_n},\ (R^n_{n\sigma_n})^2 \text{ and } e^{n\alpha Y^{j,n}(1)},$$

where j is the index of a failure particle at threshold $n\sigma_n$. The last term can be simplified if one recognizes that prior to threshold $n\sigma_n$ this particle was always a success particle. As a consequence we can write this term as

$$e^{n\alpha Y^{j,n}(1)} = e^{n\alpha \bar{Y}^{j,n}(s^j_n)} e^{n\alpha (Y^{j,n}(1) - \bar{Y}^{j,n}(s^j_n))}.$$

where $(\bar{Y}^{j,n}(s^j_n), s^j_n)$ gives the location of the same particle when it first reached threshold $n\sigma_n$. As we will see, $(\bar{Y}^{j,n}(s^j_n), s^j_n)$ and $(Y^{j,n}(1) - \bar{Y}^{j,n}(s^j_n))$ have a well defined joint LDP that will asymptotically decouple from that of the quantities such as $R^n_{n\sigma_n}$ and $(R^n_{n\sigma_n})^2$, which will then allow for a characterization of the decay rate of the second moment. We next discuss this decoupling.

We want to obtain bounds on

$$\frac{1}{n}\log E_{\lambda^n}\left[R^n_{n\sigma_n}e^{n\alpha Y^{j,n}(1)}R^n_{n\tau_n}e^{n\alpha Y^{l,n}(1)}\right], \tag{4.2.1}$$

where as in Figure 4.1 the particles last had a common ancestor as threshold $n\gamma_n$, and j (resp., l) is the index of a failure particle at threshold $n\sigma_n$ (resp., $n\tau_n$), with $\gamma_n \le \sigma_n \le \tau_n \le K^n/n$. Recall also the definition

$$C^n_j \doteq \{(y,t) \in \mathbb{R}\times[0,1] : V_B(y,t) \le V_B(0,0) - j\Delta/n\}, \quad j = 0,1,\ldots,K^n.$$

We will use the Markov property and a conditioning argument to divide $E_{\lambda^n}[R^n_{n\sigma_n}e^{n\alpha Y^{j,n}(1)}R^n_{n\tau_n}e^{n\alpha Y^{l,n}(1)}]$ into three parts according to $(\gamma_n,\sigma_n,\tau_n)$, i.e., the times when (I) there is only one success particle, which is from threshold 1 to $n\gamma_n$, (II) there are two independent success particles (threshold $n\gamma_n$ to $n\sigma_n$), and (III) there is again only a single success particle (from threshold $n\sigma_n$ to $n\tau_n$). By symmetry and without loss of generality we can assume that index 1 corresponds to the particle which is successful all the way from threshold 1 to $n\tau_n$, and index 2 identifies the particle that last had a common ancestor with particle 1 at threshold $n\gamma_n$, and then failed at threshold $n\sigma_n$. As in Lemma 4.1.2, each of γ_n,σ_n and τ_n are considered as given. Some notation that will be needed is as follows. Define the sets

$$\mathcal{A}(\sigma_n,\tau_n) \doteq \left\{\cap_{m=n\sigma_n+1}^{n\tau_n}\{\text{particle 1 is successful at threshold } m\}\right\},$$

$$\mathcal{B}(\gamma_n,\sigma_n) \doteq \left\{\cap_{m=n\gamma_n+1}^{n\sigma_n}\{\text{particle 1 and particle 2 are successful at threshold } m\}\right\},$$

$$\mathcal{A}(\gamma_n) \doteq \{\cap_{m=1}^{n\gamma_n}\{\text{particle 1 is successful at threshold } m\}\}. \tag{4.2.2}$$

Note that $1 \in F^n_{n\tau_n}, 2 \in F^n_{n\sigma_n}$ imply $\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)\cap\mathcal{B}(\gamma_n,\sigma_n)\cap\mathcal{A}(\gamma_n)} = 1$ for some $\gamma_n \le \sigma_n$.

Then owing to the indicator function, the quantities that appear in Lemma 4.1.2 can be written in the form

$$E_{\lambda^n}\left[(R^n_{n\sigma_n})^2\ e^{n\alpha Y^{2,n}(1)}(R^n_{n\tau_n}/R^n_{n\sigma_n})e^{n\alpha Y^{1,n}(1)}\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)\cap\mathcal{B}(\gamma_n,\sigma_n)\cap\mathcal{A}(\gamma_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n},2\in F^n_{n\sigma_n}\}}\right].$$

As above we let $(\bar{Y}^{1,n}(t^1_n),t^1_n),(\bar{Y}^{1,n}(s^1_n),s^1_n),(\bar{Y}^{1,n}(g^1_n),g^1_n)$ and $(\bar{Y}^{2,n}(s^2_n),s^2_n)$ to denote the locations and times when particle 1 reaches thresholds $n\tau_n, n\sigma_n, n\gamma_n$ and particle 2 reaches threshold $n\sigma_n$, respectively [with (t,s,g) corresponding to (τ,σ,γ)]. Recall $R^n_k=\sum_{m=0}^k\log(N^S_m/N)$. Then we can write

$$\begin{aligned}
&E_{\lambda^n}\left[(R^n_{n\sigma_n})^2\ e^{n\alpha Y^{2,n}(1)}(R^n_{n\tau_n}/R^n_{n\sigma_n})e^{n\alpha Y^{1,n}(1)}\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)\cap\mathcal{B}(\gamma_n,\sigma_n)\cap\mathcal{A}(\gamma_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n},2\in F^n_{n\sigma_n}\}}\right]\\
=&E_{\lambda^n}\Bigg[e^{n\alpha[Y^{2,n}(1)-\bar{Y}^{2,n}(s^2_n)]+n\alpha[\bar{Y}^{2,n}(s^2_n)-\bar{Y}^{1,n}(g^1_n)]+n\alpha\bar{Y}^{1,n}(g^1_n)}\\
&\qquad\times e^{n\alpha[Y^{1,n}(1)-\bar{Y}^{1,n}(t^1_n)]+n\alpha[\bar{Y}^{1,n}(t^1_n)-\bar{Y}^{1,n}(s^1_n)]+n\alpha[\bar{Y}^{1,n}(s^1_n)-\bar{Y}^{1,n}(g^1_n)]+n\alpha\bar{Y}^{1,n}(g^1_n)}\\
&\qquad\times\prod_{m=0}^{n\sigma_n}\left(\frac{N^S_m}{N}\right)^2\prod_{m=n\sigma_n+1}^{n\tau_n}\left(\frac{N^S_m}{N}\right)\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)\cap\mathcal{B}(\gamma_n,\sigma_n)\cap\mathcal{A}(\gamma_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n},2\in F^n_{n\sigma_n}\}}\Bigg]\\
=&\ E_{\lambda^n}\left[\Phi^n_1\times\Phi^n_2\times\Phi^n_3\right], \qquad (4.2.3)
\end{aligned}$$

where

$$\begin{aligned}
\Phi^n_1 &\doteq e^{2n\alpha\bar{Y}^{1,n}(g^1_n)}e^{2\sum_{m=0}^{n\gamma_n}\log\frac{N^S_m}{N}}\mathbf{1}_{\mathcal{A}(\gamma_n)} \qquad (4.2.\\
\Phi^n_2 &\doteq e^{n\alpha[Y^{2,n}(1)-\bar{Y}^{2,n}(s^2_n)]+n\alpha[\bar{Y}^{2,n}(s^2_n)-\bar{Y}^{1,n}(g^1_n)]+n\alpha[\bar{Y}^{1,n}(s^1_n)-\bar{Y}^{1,n}(g^1_n)]}e^{2\sum_{m=n\gamma_n+1}^{n\sigma_n}\log\frac{N^S_m}{N}}\mathbf{1}_{\mathcal{B}(\gamma_n,\sigma_n)}\mathbf{1}_{\{2\in F^n_{n\sigma_n}\}}\\
\Phi^n_3 &\doteq e^{n\alpha[Y^{1,n}(1)-\bar{Y}^{1,n}(t^1_n)]+n\alpha[\bar{Y}^{1,n}(t^1_n)-\bar{Y}^{1,n}(s^1_n)]}e^{\sum_{m=n\sigma_n+1}^{n\tau_n}\log\frac{N^S_m}{N}}\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n}\}}.
\end{aligned}$$

Here we have grouped quantities according to the partition of thresholds $0\le n\gamma_n\le n\sigma_n\le n\tau_n$. The large deviation asymptotics of each of these terms can then be evaluated, so long as we know the joint large deviation asymptotics of each of the following quantities.

1. **Trajectories of failure particles.** We see that quantities such as $[Y^{2,n}(1) - \bar{Y}^{1,n}(g_n^1)]$ and $[Y^{1,n}(1) - \bar{Y}^{1,n}(t_n^1)]$ need to be considered in Φ_2^n and Φ_3^n. These are the increments of particles after the threshold at which they fail. As we will see, their large deviation properties are the same as that of the original process over the relevant time interval (e.g., $1 - g_n^1$ in Φ_2^n).

2. **Trajectories consisting only of success particles.** We also encounter quantities such as $[\bar{Y}^{1,n}(s_n^1) - \bar{Y}^{1,n}(g_n^1)]$ in Φ_2^n which, given $\mathbf{1}_{\mathcal{B}(\gamma_n,\sigma_n)} = 1$, is the sum of a certain number of increments of the original process, conditioned on each increment leading to a success particle. As we will show below, the large deviation properties of these can also be identified through the rate function on path space of the original process, though now it will be in terms of a conditional LDP.

3. **Empirical measures on the number of success particles.** The last quantity that appears is the empirical measure on the number of success particles. To be more precise, it is the empirical measure given that either particle 1 or particles 1 and 2 are known to be success particles between the key thresholds.

The large deviation rates for each of these quantities will decouple from the others. For example, the rate for the failure particle in Φ_3^n (the $[Y^{1,n}(1) - \bar{Y}^{1,n}(t_n^1)]$ term) decouples from that of the trajectory $[\bar{Y}^{1,n}(t_n^1) - \bar{Y}^{1,n}(s_n^1)]$ and the empirical measure on the number of success particles owing to the Markov property and the uniformity of the rate function for $[Y^{1,n}(1) - \bar{Y}^{1,n}(t_n^1)]$ with respect to the X component of particle 1 at threshold $n\tau_n$. As a consequence, the rate function for the three types of objects is the sum of the corresponding individual rate functions.

4.3 Some Important Assumptions

As far as the form of the rates functions is concerned, they will depend on the large deviation properties of the underlying process $\{X\}$. In addition, the resampling method used to maintain the number of particles at level N at each threshold will also impact the large deviation properties of the empirical measures for the number of success particles, but the effect will be shown to diminish as $N \to \infty$. To proceed further, we next state conditions that will be assumed on $\{X\}$ and the resulting large deviation rate function on path space for $\{Y^n\}$. Define the empirical measure for the ergodic Markov process $\{X(t), t \in [0, \infty)\}$ by

$$\mu^T(A) \doteq \frac{1}{T} \int_0^T 1_A(X(s))ds, \quad A \in \mathcal{B}(\mathcal{X}).$$

We view $\{\mu^T\}_{T\in(0,\infty)}$ as taking values in $\mathcal{P}(\mathcal{X})$, which is equipped with the topology of weak convergence.

It follows under Condition 2.2.1 that $\{Y^n(\cdot)\}_{n\in\mathbb{N}}$ satisfies an LDP on $C([0,1] : \mathbb{R})$ that is uniform with respect to $X(0)$. Here we consider $C([0,1] : \mathbb{R})$ with the usual sup norm, which makes it into a Polish space. To be precise, we have the following.

Lemma 4.3.1. Under Condition 2.2.1 $\{Y^n(\cdot)\}_{n\in\mathbb{N}}$ satisfies the LDP on $C([0,1] : \mathbb{R})$ with rate function

$$J(\phi) = \begin{cases} \int_0^1 L(\dot{\phi}(s))ds & \text{if } \phi \text{ is absolutely continuous, } \phi(0) = 0 \\ \infty & \text{else} \end{cases},$$

uniformly in $X(0)$, where

$$L(\beta) = \inf \left\{ I(\mu) : \int_{\mathcal{X}} c(x)\mu(dx) = \beta \right\}.$$

L is convex and lower semicontinuous, $L(\beta) = \infty$ if $|\beta| > \|c\|_\infty$, and $L(\beta) = 0$ if and only if $\beta = \int_{\mathcal{X}} c d\pi$.

Proof sketch. The claims follow from standard large deviation calculations. The convexity and lower semicontinuity of L follow from the corresponding properties assumed of I. For example, suppose that $\beta_i \to \beta$ as $i \to \infty$. Since $I(\mu)$ is lower semicontinuous and $\mathcal{P}(\mathcal{X})$ is compact, it follows that there is μ_i such that $I(\mu_i) = L(\beta_i)$ and $\int_{\mathcal{X}} c d\mu_i = \beta_i$. It is enough to show that given any subsequence i_k there is a further subsequence (again denoted i_k) such that $\liminf_k L(\beta_{i_k}) \geq L(\beta)$. By compactness of $\mathcal{P}(\mathcal{X})$, given i_k we can choose a subsequence such that μ_{i_k} converges to a limit μ. By the weak convergence $\mu_{i_k} \Rightarrow \mu$ and $\int_{\mathcal{X}} c d\mu_i = \beta_i$ we know $\int_{\mathcal{X}} c d\mu = \beta$, and therefore

$$L(\beta) \leq I(\mu) \leq \liminf_{k\to\infty} I(\mu_{i_k}) = \liminf_{k\to\infty} L(\beta_{i_k}).$$

Likewise $L(\beta) = 0$ if and only if $\beta = \int_{\mathcal{X}} c d\pi$ is true since $I(\mu) = 0$ if and only if $\mu = \pi$. Finally, $L(\beta) = \infty$ when $|\beta| > \|c\|_\infty$ since in this case it is impossible to satisfy the constraint in the definition of L.

The proof of the large deviation limits makes use of the following. Suppose we are given $\delta > 0$ with $1/\delta$ an integer, and a path $\phi : [0, 1] \to R$ with constant velocity on each interval of the form $(i\delta, i\delta + \delta)$. The uniform LDP for the empirical measure then immediately implies

$$\lim_{a\downarrow 0} \limsup_{n\to\infty} \sup_{x\in\mathcal{X}} \frac{1}{n} \log P_x \left\{ \sup_{t=k\delta, k=0,1,\ldots,1/\delta} |Y^n(t) - \phi(t)| \leq a \right\} \leq -\int_0^1 L(\dot{\phi}(s)) ds,$$

and therefore since $\|c\|_\infty < \infty$

$$\lim_{a\downarrow 0} \limsup_{n\to\infty} \sup_{x\in\mathcal{X}} \frac{1}{n} \log P_x \left\{ \sup_{0\leq t\leq 1} |Y^n(t) - \phi(t)| \leq a \right\} \leq -\int_0^1 L(\dot{\phi}(s)) ds.$$

Note that the collection of Lipschitz continuous paths that start at 0 is compact, and that the subset of paths with the indicated piecewise constant velocities form a dense subset. The uniform large deviation upper bound then follows from the last display and an open covering argument. To argue the lower bound we note that given any path ϕ with $\int_0^1 L(\dot{\phi}(s))ds < \infty$, it follows from convexity of $L(\cdot)$ that $\int_0^1 L(\dot{\phi}^\delta(s))ds \leq \int_0^1 L(\dot{\phi}(s))ds$, where ϕ^δ has piecewise constant velocities and the same average velocity over any interval of the form $(i\delta, i\delta+\delta)$ as ϕ. The lower bound then follows from $\phi^\delta \to \phi$ as $\delta \downarrow 0$, and that for $a > 0$ and sufficiently small $\delta > 0$,

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{x\in\mathcal{X}} \frac{1}{n} \log P_x \left\{ \sup_{0\leq t\leq 1} |Y^n(t) - \phi(t)| \leq a \right\} \\
&\quad \geq \liminf_{b\to 0} \liminf_{n\to\infty} \inf_{x\in\mathcal{X}} \frac{1}{n} \log P_x \left\{ \sup_{t=k\delta, k=0,1,\ldots,1/\delta} |Y^n(t) - \phi(t)| \leq b \right\} \\
&\quad \geq -\int_0^1 L(\dot{\phi}^\delta(s))ds \geq -\int_0^1 L(\dot{\phi}(s))ds.
\end{aligned}$$

□

Some convex duality relations that will be useful for both the statement of the main results and the proofs are as follows. If we define the Legendre transform

$$H(\alpha) \doteq \sup_{\beta\in\mathbb{R}} [\alpha\beta - L(\beta)],$$

then H is a convex function and, since automatically $L(\beta) = \infty$ if $|\beta| \geq \|c\|_\infty$, $H(\alpha) < \infty$ and therefore $H : \mathbb{R} \to \mathbb{R}$. Under the standing assumption $\int_{\mathcal{X}} c d\pi = 0$ (see Remark 2.3.2) we also have that $H(\alpha) \geq -L(0) = 0$ and since $L(\beta) \geq 0$ for all β that $H(0) = 0$. Finally we note that by Varadhan's lemma and the contraction

principle

$$\lim_{n\to\infty}\frac{1}{n}\log E_x e^{n\alpha Y^n(1)} = \sup_{\gamma\in\mathcal{P}(\mathcal{X})}\left\{\alpha\int_{\mathcal{X}} cd\gamma - I(\gamma)\right\} = \sup_{\beta\in\mathbb{R}}[\alpha\beta - L(\beta)] = H(\alpha)\in(0,\infty).$$

We make an additional very mild assumption on the regularity properties of L and H, which in particular holds for the pure jump and reflecting diffusion examples considered previously. We also rule out trivial cases with assumptions on c and π.

Condition 4.3.2. H is continuously differentiable on $\mathbb{R}$ and L is essentially smooth. Also, c is not a constant a.e. with respect to π.

Essentially smooth means that L is continuously differentiable on the interior of $D_L \doteq \{\beta : L(\beta) < \infty\}$ and, if $\beta_i \in D_L$ satisfies $\beta_i \to \beta \in \partial D_L$, then $L'(\beta_i) \to L'(\beta)$. For the role of this assumption in large deviation theory we refer the reader to [15]. These smoothness conditions are not essential, but simplify some of the statements and arguments, since they imply, among other things, that convex conjugate points (in the sense of convex duality) are unique, and since the domain of L has nonempty interior that H is strictly convex. Since L is a proper convex function, we note for later use that $L(\beta) = \sup_{\alpha\in\mathbb{R}}[\beta\alpha - H(\alpha)]$.

As is discussed in many places (e.g., [4, Section 14.1]), for any unbiased estimator with parameters that can be chosen to improve performance, minimizing the variance is equivalent to minimizing the second moment. Recall that E_{λ^n} indicates that particles start with iid distributions in the $\mathcal{X}$ component with marginal λ^n. Using Jensen's inequality, for any unbiased (single sample) estimator $\hat{M}^n(\alpha)$ we have

$$E_{\lambda^n}\left(\hat{M}^n(\alpha)^2\right) \geq \left(E_{\lambda^n}\hat{M}^n(\alpha)\right)^2 = \left(E_{\lambda^n}e^{n\alpha Y^n(1)}\right)^2.$$

Thus the optimal (smallest) rate of growth of the second moment is precisely twice the growth rate $H(\alpha)$:

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[\hat{M}^n(\alpha)\right]^2 \geq 2H(\alpha).$$

A scheme that replaces lim inf by lim and $\geq$ by $=$ is called *asymptotically optimal*. If a scheme is asymptotically optimal, then the number of independent copies needed so that the average of those samples has variance that is proportional to the quantity being estimated grows subexponentially.

Remark 4.3.3. The IPS schemes we construct are never asymptotically optimal, though when $B = H(\alpha)$ we will show that the gap can be made small by taking N large. One can construct asymptotically optimal importance sampling and (ordinary) splitting schemes, but these require a somewhat elaborate change of measure for importance sampling as well as knowing $H(\alpha)$. In the case of splitting one can consider a work-normalized analogue of the second moment, and then for asymptotic optimality one must again know $H(\alpha)$ and also choose the analogue of Δ so that the mean number of descendents that make it to the next threshold is precisely 1. In contrast, the IPS is somewhat robust with respect to errors in the selection of Δ, and as we discuss in Remark 4.4.3 lends itself to diagnostics which allow one to to learn $H(\alpha)$ based on numerical experimentation. For the models considered here, one might consider using the IPS scheme to find good values of B and Δ, and then a switch to the corresponding ordinary splitting scheme (or its RESTART variant).

In the analysis of the second moment, we will find it convenient when $B \neq H(\alpha)$ to interpret the interacting system as the system that would be best for a different value of α, and in particular to use that the thresholds defined in terms of B correspond to this alternate α_B. Let $\bar{C}$ be the supremum of $c(x)$ with respect to

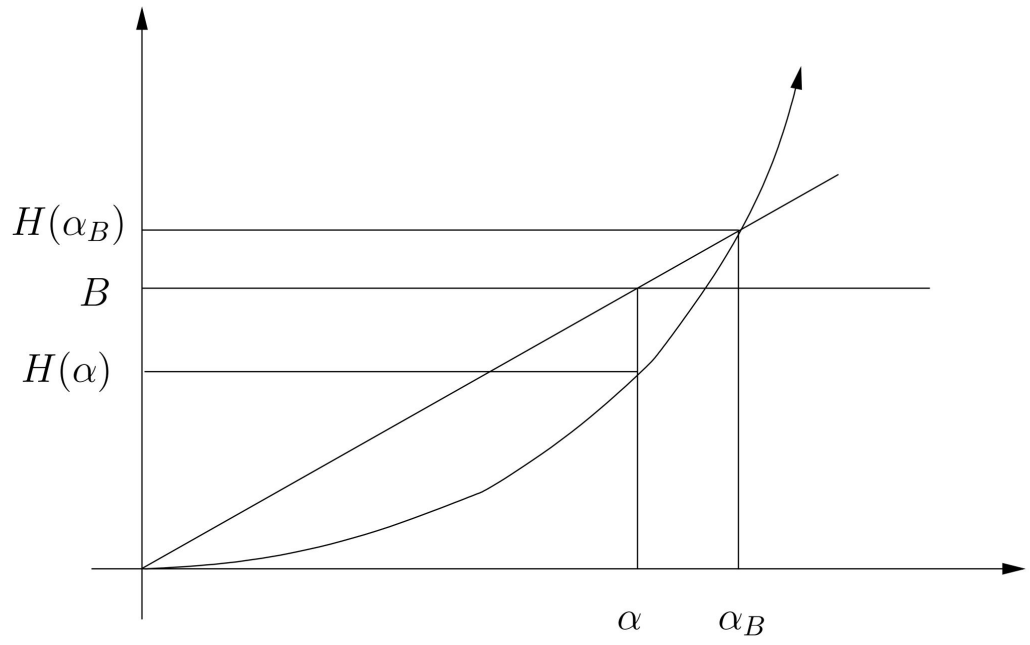

Figure 4.2: Relations between B and H(B)

the support of the stationary distribution π. One can easily show that $H'(z) \uparrow \bar{C}$ as $z \to \infty$, and as a consequence for any $0 < B < \alpha\bar{C}$ there is a unique α_B such that

$$\frac{H(\alpha_B)}{\alpha_B} = \frac{B}{\alpha}. \tag{4.3.1}$$

(here we use $\alpha > 0$ and $H'(0) = 0$, see Figure 4.3). Moreover, we have $\alpha_B < \alpha$ (resp., $\alpha_B > \alpha$) if $B < H(\alpha)$ (resp., $B > H(\alpha)$). Note also that by the strict convexity of H, whenever $B \neq H(\alpha)$ (and hence $\alpha_B \neq \alpha$)

$$H(\alpha_B) + H(2\alpha - \alpha_B) > 2H\left(\frac{\alpha_B}{2} + \frac{2\alpha - \alpha_B}{2}\right) = 2H(\alpha). \tag{4.3.2}$$

As we will see, the values $H'(\alpha_B)$ and $H'(2\alpha - \alpha_B)$ have important interpretations in connection with the computational scheme. Specifically, $H'(\alpha_B)$ is the mean velocity followed by success particles, while $H'(2\alpha - \alpha_B)$ is the velocity followed by those success particles that will contribute the greatest amount to the second moment of the estimator (see Remark 5.1.7). Figure 4.3 illustrates the relation between these and β^*

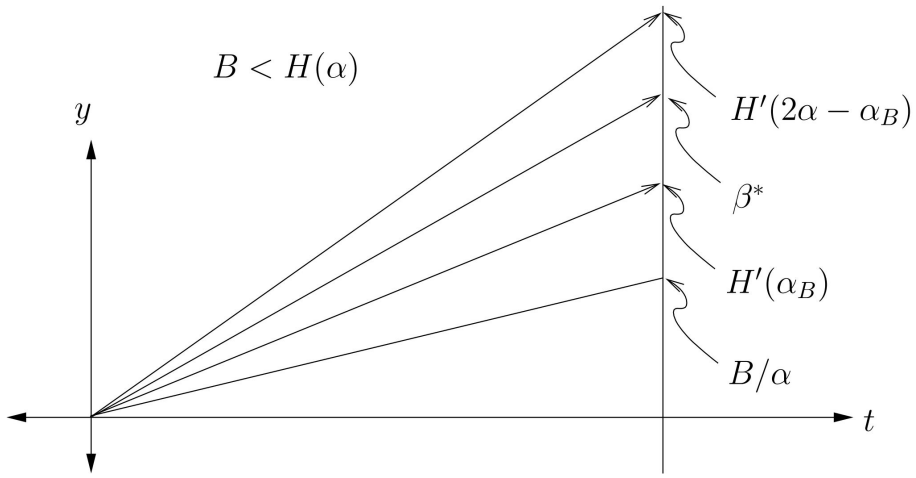

Figure 4.3: Critical velocities when $B \neq H(\alpha)$

and the threshold slope B/α when $B < H(\alpha)$. The inequalities depicted follow from $H(\alpha_B)/\alpha_B = B/\alpha$, which implies $H'(\alpha_B) > B/\alpha$, and the monotonicity of $H'(\cdot)$ and $B \to \alpha_B$. When $B > H(\alpha)$ we get the reversed situation $H'(\alpha_B) > \beta^* > H'(2\alpha - \alpha_B)$ but, unlike in the case $B < H(\alpha)$, $H'(2\alpha - \alpha_B) > B/\alpha$ can fail for large enough $B > H(\alpha)$. When $H'(2\alpha - \alpha_B) < B/\alpha$ happens $H'(2\alpha - \alpha_B)$ is no longer a possible velocity for a trajectory of success particles, and a different argument is required for the second moment. Let

$$\bar{B} \doteq \min\{B > 0 : H'(2\alpha - \alpha_B) \leq B/\alpha\} \tag{4.3.3}$$

Although one can present bounds on the second moment when $B \geq \bar{B}$ which show it to be not optimal, since this case is arguably less important we limit the analysis of the schemes to $B < \bar{B}$.

The following assumption, whose validity depends on the resampling method used, will play a crucial role in the analysis. For $\boldsymbol{x}^N \in S^N$ let $\Theta^N(\boldsymbol{x}^N) \doteq \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{x}^N(i)}$ denote the empirical measure.

Condition 4.3.4. If $\Theta^N(\boldsymbol{x}^N) \Rightarrow \mu$ as $N \to \infty$, then under $q^{\infty,N}(\boldsymbol{x}^N, d\boldsymbol{z}^N)$

$$\Theta^N(\boldsymbol{z}^N) \Rightarrow g(\mu) \in \mathcal{P}(\bar{\mathcal{X}})$$

in distribution, where $g(\mu)$ has the unique fixed point $\lambda^\infty \in \mathcal{P}(\bar{\mathcal{X}})$.

Lemma 4.3.5. Assume Conditions 2.2.1, 4.1.1 and 4.3.2, and that any of the resampling schemes discussed in Remark 3.1.3 are used. Then Condition 4.3.4 holds.

Proof. Suppose that $\mu^N(dx) = \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{x}_i^N}(dx) \Rightarrow \mu(dx)$. Let B_i^N be independent Bernoulli random variables that tell if particle i will reach the next threshold, which it does with probability $q^\infty(\boldsymbol{x}_i^N, \mathcal{X})$. If is does, it will have location Z_i^N, which has distribution

$$p(\boldsymbol{x}_i^N, d\boldsymbol{z}_i^N) = q^\infty(\boldsymbol{x}_i^N, d\boldsymbol{z}_i^N)/q^\infty(\boldsymbol{x}_i^N, \mathcal{X}).$$

Let M^N be the number of success particles. Then since $q^\infty(\boldsymbol{x}_i^N, \mathcal{X}) \geq \bar{c}_1 > 0$, M^N/N converges a.s. to

$$R = \int_{\mathcal{X}} q^\infty(x, \mathcal{X})\mu(dx) \in [\bar{c}_1, 1].$$

In addition, owing to the Feller property, the standard argument using

$$E\left[\frac{1}{N}\sum_{i=1}^N B_i^N \delta_{Z_i^N}(dz)f(z) - \frac{1}{N}\sum_{i=1}^N q^\infty(\boldsymbol{x}_i^N, dz)f(z)\right]^2 \to 0$$

shows that $\sum_{i=1}^N B_i^N \delta_{Z_i^N}(dz)/N$ converges in distribution and in the weak topology to the subprobability $\int_{\mathcal{X}} q^\infty(x, dz)\mu(dx)$.

There is also the resampling aspect to address. The first case in Remark 3.1.3 is

iid uniform resampling from

$$\gamma^N(dz) \doteq \frac{1}{M_N} \sum_{i=1}^{N} B_i^N \delta_{Y_i^N}(dz).$$

Suppose we generate N samples Z_j^N according to γ^N. Since $\gamma^N(\cdot) \Rightarrow \int_{\mathcal{X}} q^\infty(x, \cdot)\mu(dx)/R = g(\mu)(\cdot)$, a minor variant in the proof of the Glivenko-Cantelli lemma shows $\sum_{j=1}^{N} \delta_{Z_j^N}(dz)/N \Rightarrow g(\mu)(dz)$.

Next suppose that we do proportional sampling, and suppose that $\lfloor N/M_N \rfloor = m_N$. Then $N/M_N \geq \lfloor N/M_N \rfloor$ and so $1 = Nm_N/M_N \geq m_N \lfloor N/M_N \rfloor$. After the resampling we have

$$\frac{1}{N} m_N M_N \gamma^N(dz) + \frac{1}{N} \sum_{j=1}^{N - m_N M_N} \delta_{Z_j^N}(dz).$$

The same small change in the proof of the Glivenko-Cantelli lemma gives that

$$\frac{1}{N - m_N M_N} \sum_{j=1}^{N - m_N M_N} \delta_{Z_j^N}(dz) \Rightarrow g(\mu)(dz)$$

and so

$$\frac{1}{N} m_N M_N \gamma^N(dz) + \frac{N - m_N M_N}{N} \frac{1}{N - m_N M_N} \sum_{j=1}^{N - m_N M_N} \delta_{Z_j^N}(dz) \Rightarrow g(\mu)(dz).$$

The last sampling scheme is similar and omitted. □

4.4 Main Result

We now come to the main result of the paper. The proof is given at the end of Section 5.2 after large deviation estimates on the three quantities identified earlier in this section are proved. We recall $\bar{B}$ as defined in (4.3.3).

Theorem 4.4.1. Assume Conditions 2.2.1, 4.1.1, 4.3.2 and 4.3.4. Suppose we are given V_B of the form $V_B(y,t) = -\alpha y + B[t-1]$, where $B \in (0, \bar{B})$, and consider the n-dependent version of the interacting particle scheme as defined in Section 3 with thresholds given by (2.3.2). Define α_B as the unique solution to (4.3.1). Then there is $C_N \downarrow 0$ as $N \to \infty$, such that

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[\mathcal{E}^n\right]^2 \leq H(\alpha_B) + H(2\alpha - \alpha_B) + C_N. \tag{4.4.1}$$

There is also $\bar{C}_N \downarrow 0$ as $N \to \infty$, such that

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[\mathcal{E}^n\right]^2 \geq \max\{H(\alpha_B) + H(2\alpha - \alpha_B) - \bar{C}_N, 2H(\alpha)\}, \tag{4.4.2}$$

Remark 4.4.2. When $B = H(\alpha)$ we have $\alpha_B = \alpha$ and we see from (4.4.1) that, while the scheme may not be asymptotically optimal, the gap can be made as small as desired by choosing N large. When $B \neq H(\alpha)$ the lower bound allows for the possibility of better performance with fewer particles, but it is likely that this is due to the bound not being as tight as possible. To understand the tradeoff between lower variance and additional computational effort, it would be most useful to have a more quantitative description of C_N. See Remark 5.1.4.

Remark 4.4.3. The interacting scheme requires the choice of a small number of parameters, namely Δ, which controls the spacing between thresholds, B, which influences the conditional distribution of the velocity of success particles, N, the

number of particles, and η, which is used to define the n dependent time horizon used to produce a stationarity used in the analysis. One might expect that the last parameter is not so crucial, since one can show that the probability of reaching the next threshold by taking a longer time is very small, and so the additional successes would not be expected to reduce variance much.

Leaving aside the choice of N, one is left to determine good values for B and Δ. A natural adaptive approach is to start with a relatively small value of n, and based on simulations at that level of n gather data such as the fraction of particles that are success particles and estimates of $E_x e^{n\alpha Y^n(1)}$ as well as the estimated variance, and then use these to adjust the parameters B and Δ before raising the value of n. Suppose that a subsequence $\{n_k\}$ is to be considered. By starting with a small value of n_0 where one is confident that both the estimate of $E_x e^{n_0\alpha Y^{n_n}(1)}$ as well as its estimated variance are accurate, one might adjust according to a stochastic approximation type iteration of the form

$$\begin{aligned} B_{k+1} &= B_k + \delta_k \left(\frac{1}{n_k} \log \mathcal{E}^{n_k} - B_k \right) \\ \Delta_{k+1} &= \Delta_k + \bar{\delta}_k (b^k - p_0), \end{aligned}$$

where $\mathcal{E}^{n_k}$ is the estimate of the quantity of interest, b^k is the fraction of success particles, p_0 is a target fraction, and $\delta_k > 0$ and $\bar{\delta}_k > 0$ are gain sequences whose choice (and also possibly the choice of n_{k+1}) could be influenced by the ratio of the estimated variance of $\mathcal{E}^{n_k}$ to $\mathcal{E}^{n_k}$. The equation for Δ_k is straightforward, while that of B_k is motivated by the observation that $\frac{1}{n_k} \log \mathcal{E}^{n_k}$ would be presumably close to $\frac{1}{n_k} \log M^{n_k}(\alpha)$, and that $\frac{1}{n_k} \log M^{n_k}(\alpha) \to H(\alpha)$ as $k \to \infty$.

Chapter Five

Asymptotic Properties of Constitutive Elements

In this section we formulate the large deviation statements concerning each of the three elements identified in the last section, and then tie them together to obtain large deviation growth/decay rates for the second moment. These rates are expressed in terms of a variational problem. We use a dynamic programming type argument solve the variational problem, and then give the proof of Theorem 4.4.1.

We phrase the limits for the individual elements in terms of Laplace principles that are uniform in certain parameters. As is known [4, Proposition 1.12], this implies the corresponding large deviation principle that is uniform in the same parameters. What is important for our purposes is that the uniformity will allow for a joint LDP involving all the key elements identified in the last section. Recall the quantities Φ_n^r defined in and below (4.2.4). Proofs of all the lemmata in this section are given in the Appendix. Throughout this section we assume Conditions 2.2.1, 4.1.1, 4.3.2 and 4.3.4, though in fact only Lemmas 5.1.3 and 5.1.5 use Condition 4.3.4.

5.1 The Large Deviation Lower Bound

5.1.1 Trajectories of Failure Particles

The following lemma is concerned with the distribution of a particle when it fails. The estimate presented follows almost directly from the rate function for the process $\{Y^n\}$ on path space.

Lemma 5.1.1. Consider the random variables of the form $[Y^{1,n}(1) - \bar{Y}^{1,n}(t)]$ appearing in Φ_n^3, where $t \in [0,1]$. Let $E_{z,t}$ denote expected value given that the X component of particle 1 equals z at time t. For any bounded and continuous func-

tion $f : \mathbb{R} \to \mathbb{R}$ we have the limit

$$\lim_{n\to\infty} \sup_{z\in\mathcal{X},t\in[0,1]} \left| \frac{1}{n} \log E_{z,t} e^{-nf(Y^{1,n}(1)-\bar{Y}^{1,n}(t))} + \inf_{\beta\in\mathbb{R}} \{[1-t]L(\beta) + f(\beta[1-t])\} \right| = 0.$$

The analogous statement holds for the random variables $[Y^{2,n}(1) - \bar{Y}^{1,n}(g)], g \in [0,1]$ that appear in Φ_n^2.

5.1.2 Trajectories Consisting Only of Success Particles

The next lemma presents distributional properties of a particle that is successful for many thresholds in a row. The rate function is in fact the same as that of the underlying process conditioned on reaching the last threshold, and this turns out to be correct because the biggest contribution to this conditional distribution (i.e., the most typical way it occurs) comes from repeatedly making each of the many thresholds before the time constraint required to be a success particle is exceeded.

Lemma 5.1.2. Consider the random variables $(\bar{Y}^{1,n}(g_n^1), g_n^1)$ appearing in Φ_n^1. Let E_z denote expected value given that the X component of particle 1 equals z at time 0, and suppose that $\gamma_n \to \bar{\gamma}$ as $n \to \infty$ (recall that $n\gamma_n$ is the threshold at which the success particle being followed splits into two). For any bounded and continuous function $f : \mathbb{R} \times [0,1] \to \mathbb{R}$ we have the limit

$$\lim_{n\to\infty} \sup_{z\in\mathcal{X}} \left| \frac{1}{n} \log E_z \left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)} \middle| \mathbf{1}_{\mathcal{A}(\gamma_n)} = 1 \right] + \inf_{(\beta,s)\in\partial C_{\bar{\gamma}\Delta}} \left\{ sL(\beta/s) - \bar{\gamma}\Delta \frac{H(\alpha_B)}{B} + f(\beta,s) \right\} \right| = 0.$$

The analogous statement holds for the random variables $(\bar{Y}^{2,n}(s_n^2) - \bar{Y}^{1,n}(g_2^1), s_n^2 - g_2^1), (\bar{Y}^{1,n}(s_n^1) - \bar{Y}^{1,n}(g_2^1), s_n^1 - g_2^1)$ and $(\bar{Y}^{1,n}(t_n^1) - \bar{Y}^{1,n}(s_2^1), t_n^1 - s_2^1)$ that appear in Φ_n^2 and Φ_n^3, save that $\bar{\gamma}$ is replaced by $(\bar{\tau} - \bar{\sigma})$ and $(\bar{\sigma} - \bar{\gamma})$ in Φ_n^2 and Φ_n^3, respectively,

where $\sigma_n \to \bar{\sigma}$ and $\tau_n \to \bar{\tau}$ as $n \to \infty$.

5.1.3 Empirical Measures on the Number of Success Particles

The next result give the asymptotic behavior of the empirical number of success particles. We consider the double limit as $n \to \infty$ and $N \to \infty$. When $N \to \infty$ a type of law of large numbers behavior dominates, and as a consequence the second moment is nearly the square of the first moment, indicating that from the point of view of large deviations the distribution of the empirical measure is close to a constant.

Lemma 5.1.3. Suppose that $\mathbf{1}_{\mathcal{A}(\theta n)} = 1$ indicates that a particle (which without loss we take to be labeled as particle 1) is always successful up to threshold θn. Then there are $\bar{C}_N^1, \bar{C}_N^2 \in [0, \infty)$ with $\bar{C}_N^r \to 0$ as $N \to \infty$, such that for $r = 1, 2$

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[e^{r \sum_{m=0}^{\theta n - 1} \log \frac{N_m^S}{N}} \middle| \mathbf{1}_{\mathcal{A}(\theta n)} = 1 \right] \geq r\theta \left[-\Delta \frac{H(\alpha_B)}{B} - \bar{C}_N^r \right].$$

Moreover, there there are $C_N^1, C_N^2 \in [0, \infty)$ with $C_N^r \to 0$ as $N \to \infty$, $r = 1, 2$, such that

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[e^{r \sum_{m=0}^{\theta n - 1} \log \frac{N_m^S}{N}} \middle| \mathbf{1}_{\mathcal{A}(\theta n)} = 1 \right] \leq r\theta \left[-\Delta \frac{H(\alpha_B)}{B} + C_N^r \right].$$

In addition, results of the same form but will possibly different $\bar{C}_N^r, C_N^r$ hold when we condition on two distinct particles being repeatedly successful up to threshold θn.

Remark 5.1.4. The case of Example 2.2.3 is simple enough that one can explicitly

evaluate the constants in the lemma just stated, and we find

$$C_N^1 = \frac{H(\alpha_B)}{B} \log\left(1 + \frac{(1-p)}{pN}\right),$$

$$C_N^2 = \frac{H(\alpha_B)}{B} \log\left(1 + \frac{3p - 3p^2}{p^2 N} + \frac{3p - 3p^2 + 1}{p^2 N^2}\right).$$

By combining the estimates on these three types of quantities and using the stated uniformity properties, we obtain the following.

Lemma 5.1.5. There is $C_N \geq 0$ which satisfies $C_N \to 0$ as $N \to \infty$, such that given any $n\gamma_n \leq n\sigma_n \leq n\tau_n \leq K^n$ that satisfy $\gamma_n \to \bar{\gamma}, \sigma_n \to \bar{\sigma}, \tau_n \to \bar{\tau}$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log E_{\lambda^n}\left[R^n_{n\sigma_n} e^{n\alpha Y^{j,n}(1)} R^n_{n\tau_n} e^{n\alpha Y^{l,n}(1)} 1_{\{j\in F^n_{n\sigma_n}, l\in F^n_{n\tau_n}, n\gamma_n = c_n\}}\right] \\
&\quad \leq \sup\left\{\alpha(f_2 + f_1) - (1-s^2)L\left(\frac{f_2 - z_2}{1-s^2}\right) - (1-t^1)L\left(\frac{f_1 - z_3}{1-t^1}\right) - \bar{\gamma}\Delta\frac{H(\alpha_B)}{B}\right. \\
&\qquad \left. - g^1 L\left(\frac{z_1}{g^1}\right) - (s^2 - g^1)L\left(\frac{z_2 - z_1}{s^2 - g^1}\right) - (t^1 - g^1)L\left(\frac{z_3 - z_1}{t^1 - g^1}\right)\right\} + C_N,
\end{aligned}$$

where the supremum is over all $f_2, f_1 \in \mathbb{R}$, $(z_1, g^1) \in \partial C_{\bar{\gamma}\Delta}$, $(z_2, s^2) \in \partial C_{\bar{\sigma}\Delta}$, and $(z_3, t^1) \in \partial C_{\bar{\tau}\Delta}$. (Here f_1 and f_2 correspond to the final locations of particles 1 and 2, and $(z_1, g^1), (z_2, s^2)$ and (z_3, t^1) to the locations of branching, last success of particle 2, and last success of particle 1, respectively)

Finally we have the following simplification of the variational problem appearing in the last lemma. In the lemma statement the optimizers are identified, which through large deviation theory can then be used to approximately characterize the particle interactions that give the largest contribution to the second moment.

Lemma 5.1.6. Suppose that $B \in (0, \bar{B})$ with $\bar{B}$ as defined in (4.3.3). We have

$$H(\alpha_B) + H(2\alpha - \alpha_B) \tag{5.1.1}$$

$$= \sup \left\{ \alpha(f_2 + f_1) - (1 - s^2) L\left(\frac{f_2 - z_2}{1 - s^2}\right) - (1 - t^1) L\left(\frac{f_1 - z_3}{1 - t^1}\right) - \bar{\gamma}\Delta \frac{H(\alpha_B)}{B} \right.$$
$$\left. - g^1 L\left(\frac{z_1}{g^1}\right) - (s^2 - g^1) L\left(\frac{z_2 - z_1}{s^2 - g^1}\right) - (t^1 - g^1) L\left(\frac{z_3 - z_1}{t^1 - g^1}\right) \right\},$$

where the supremum is over all $0 \le \bar{\gamma} \le \bar{\sigma} \le \bar{\tau} \le \kappa$ and $f_2, f_1 \in \mathbb{R}$, $(z_1, g^1) \in \partial C_{\bar{\gamma}\Delta}$, $(z_2, s^2) \in \partial C_{\bar{\sigma}\Delta}$, and $(z_3, t^1) \in \partial C_{\bar{\tau}\Delta}$. When $B = H(\alpha)$ the supremum is for arbitrary $0 \le \bar{\gamma} \le \bar{\sigma} \le \bar{\tau} \le \kappa$ so long as $f_1 = f_2 = \beta^*$ and all the increments [e.g., $(f_2 - z_2)/(1 - s^2)$] appearing on the right hand side of (5.1.1) are also β^*. When $B \neq H(\alpha)$ the supremum only occurs when $\bar{\gamma} \le \bar{\sigma} \le \bar{\tau}$, so that $t^1 = s^2 = g^1 = 1$, and when $f_1 = f_2 = z_1 = \tilde{\beta} = H'(2\alpha - \alpha_B)$, i.e., the convex dual of $2\alpha - \alpha_B$.

Remark 5.1.7. One can interpret the form of the optimizers in the following way. When $B = H(\alpha)$ the growth of the second moment is the slowest. The geometry of the thresholds leads to success particles that have average velocity β^*, which is the location that provides the dominant contribution to $E_z e^{n\alpha Y^n(1)}$ in the limit $n \to \infty$. In this case the largest contribution to the second moment comes from pairs of particles which end near β^*, and the threshold of their last common descendant is not very important. If $B < H(\alpha)$ then the thresholds lead to success particles with average velocity $\beta = H'(\alpha_B) < \beta^*$. In this case the second moment gets its biggest contribution from relatively rare success particles, namely those that follow instead a velocity close to $\tilde{\beta} = H'(2\alpha - \alpha_B) > \beta^*$, and then split to form a pair of failure particles just before the terminal time. The picture reverses when $B > H(\alpha)$, with typical trajectories of success particles following a faster velocity $\beta = H'(\alpha_B) > \beta^*$ than that which gives the greatest contribution to $E_z e^{n\alpha Y^n(1)}$, and the second moment of the estimator getting its largest contribution from rare success particles that follow the slower velocity $\tilde{\beta} = H'(2\alpha - \alpha_B) < \beta^*$. Eventually we reach B large enough that $\tilde{\beta} = B/\alpha$, in which case it is no longer possible that $\tilde{\beta}$ is the velocity for a success particle, and this value of B coincides with $\bar{B}$. Intuitively, the second moment of estimator in the optimal case $B = H(\alpha)$ is due to many relatively uncorrelated

particles that follow the conditional LLN trajectory for success particles, while in the suboptimal cases simulations corresponding to relatively rare events (simulations needed to produce an unbiased estimator) end up getting much of their contribution from strongly correlated particles that follow a highly atypical trajectory for success particles.

5.2 Proof of Main Result

The proof of the main result now follows from Lemmas 4.1.2, 5.1.5 and 5.1.6. Let $n\sigma_n$ and $n\tau_n$ denote times at which particles fail, and let $n\gamma_n$ be the threshold of their last common ancestor. By compactness of $[0, \kappa]$ (recall that κ was defined at the beginning of Section 4) and a standard argument using subsequences, we can assume $\gamma_n \to \bar{\gamma}, \sigma_n \to \bar{\sigma}, \tau_n \to \bar{\tau}$ as in Lemma 5.1.5. Then by combining the first part of Lemma 4.1.2, the first part of Lemma 5.1.5 and Lemma 5.1.6, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{\lambda^n}\left[(\mathcal{E}^n)^2\right] \leq H(\alpha_B) + H(2\alpha - \alpha_B) + C_N,$$

where $C_N \to 0$ as $N \to \infty$. This gives the second (upper bound) part of Theorem 4.4.1.

The proof of (4.4.2) is trivial in the case $B = H(\alpha)$, since it follows from the fact that the estimator is unbiased and Jensen's inequality. When this is not the case we argue as follows. According to Lemma 5.1.6 the variational problem has a maximizer that corresponds to $(\bar{Y}^{1,n}(g_n^1), g_n^1)$ reaching a small neighborhood of the point $(\tilde{\beta}, 1)$, after which it splits and produces two failure particles that are also in a small neighborhood of $\tilde{\beta}$. Conditioning on the existence of such a sequence of

success particles, Lemmas 5.1.2 (which contributes $2\tilde{\beta}\alpha - L(\tilde{\beta}) + \bar{\gamma}\Delta H(\alpha_B)/B$) and 5.1.3 (which contributes $-2\bar{\gamma}\Delta H(\alpha_B)/B$) and the identity $H(\alpha_B)/\alpha_B = B/\alpha$ give the lower bound

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[(\mathcal{E}^n)^2 | \mathbf{1}_{\mathcal{A}(\gamma_n)} = 1 \right] \\
\geq & 2\tilde{\beta}\alpha - L(\tilde{\beta}) + \bar{\gamma}\Delta \frac{H(\alpha_B)}{B} - 2\bar{\gamma}\Delta \frac{H(\alpha_B)}{B} \\
= & 2\tilde{\beta}\alpha - L(\tilde{\beta}) - H(\alpha_B) - \bar{\gamma}\Delta \frac{H(\alpha_B)}{B} + H(\alpha_B) \\
= & 2\tilde{\beta}\alpha - L(\tilde{\beta}) - \tilde{\beta}\alpha_B + H(\alpha_B) \\
= & H(2\alpha - \alpha_B) + H(\alpha_B).
\end{aligned}$$

where we also use that $V_B(\tilde{\beta}, 1) = V_B(0, 0) - \bar{\gamma}\Delta$ implies $\bar{\gamma}\Delta = B - \tilde{\beta}\alpha$. Since Condition 4.1.1 implies

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{\lambda^n} \left[\mathbf{1}_{\mathcal{A}(\gamma_n)} = 1 \right] \geq \log \left(1 - (1 - \bar{c}_1)^N \right),$$

the lower bound follows by combining the last two displays.

■

Chapter Six

Numerical Results

In this chapter, we present some computational results for interacting particle scheme under a continuous time Markov Chain setting.

6.1 Problem Setup and Notations

In this section we will show some interesting numerical result for **Example 2.2.3**. Recall the setting for this example: $X = \{-1, 1\}$ and suppose that $\{X(t), t \in [0, \infty)\}$ is a continuous-time, pure jump, irreducible Markov process with jump rates $r_{-1,1} = r_{1,-1} = 1$, and $c(x) = x$. Then the quantity of interest defined in (2.1.1) becomes

$$e_n = E_x \exp\left\{\alpha \int_0^n X(s)ds\right\}.$$

The Monte Carlo method estimates of e_n for different values of n are given in Table (6.1) with sample size 500, 000, 000. For simplicity, we take these approximations as the true values of e_n and compare the results from IPS scheme to them.

n	5	10	20	30	40	50
e_n	8.35632	69.8142	4.73402e+3	2.92961e+5	2.750504e+7	1.07696e+9
$\log(e_n)$	2.12301	4.24583	8.46253	12.587797	17.12988	20.797416

Table 6.1: estimates of e^n for different n using Monte Carlo Method

Also from Example (2.3.1) we can calculate the explicit solution of Bellman equation and hence we can know the parameter choose for the IPS scheme, recall

$$H(\alpha) = \sqrt{1 + \alpha^2} - 1.$$

First we choose $B = H(\alpha)$ then we can know the form of the thresholds

$$V_B(y, t) = -\alpha y + B[t - 1] = -\alpha y + (\sqrt{1 + \alpha^2} - 1)[t - 1],$$

$$\bar{C}_j^n = \{(y,t) \in \mathbb{R} \times [0,1] : y = \frac{H(\alpha)}{\alpha} t + j\frac{\Delta}{\alpha n}\}, \quad j = 0, 1, \ldots, K^n.$$

Throughout this chapter, we use N as the number of particles of the IPS estimator at each splitting step and K^n as the number of threshold. Let M be the sample size, and $\{\hat{e}_n^i, i = 0, \ldots, M-1\}$ represents the collection of M estimates.

$$\hat{e}_n = \frac{1}{M} \sum_{i=0}^{M-1} \hat{e}_n^i$$

is the IPS estimate of e_n.

First we show the the figure 6.1 compares the performance of the Monte Carlo method and Importance Sampling (IPS, $N = 100$) across different sample sizes. In these experiments, we choose $\alpha = 1$, $K^n = 10$, $\Delta = 0.6$, $h = \frac{1}{n^{0.7}}$ and $N = 100$. The x-axis represents the number of samples, while the y-axis shows the estimator values. Blue circles represent the Monte Carlo method, and red squares represent the IPS method, with error bars indicating the standard error of the estimates. As the sample size increases, both methods converge towards similar estimates, and the standard error shrink. However, for smaller sample sizes, the IPS method exhibits significantly lower standard error compared to the Monte Carlo method, demonstrating greater efficiency.

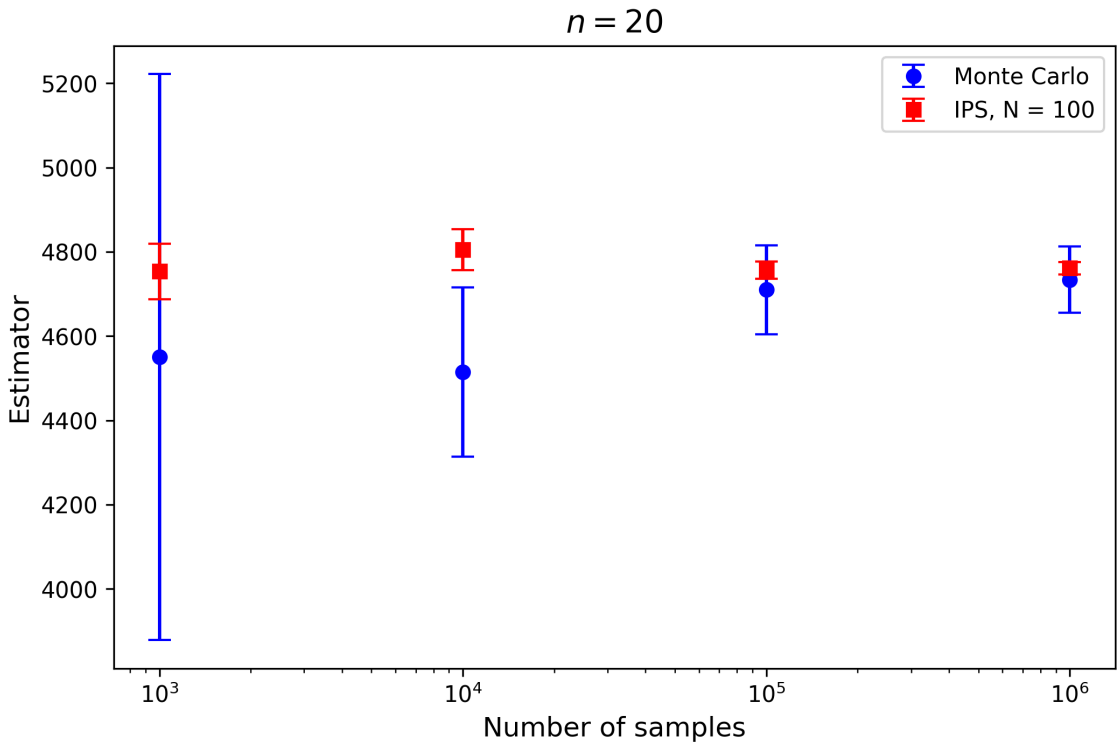

Figure 6.1: Comparison of Monte Carlo and IPS Estimators Across Sample Sizes

6.2 The Large Deviation Rate of the Second Moment

In Figure (6.2), we compare the large deviation rate of the second moment of the IPS estimators for different values of n and N. In each case, the sample size M is 10,000. The large deviation rate of the second moment is approximated by computing

$$-\log\left[\frac{1}{M}\sum_{i=0}^{M-1}\hat{e}_i^2\right].$$

The black dashed line represents the asymptotic optimal rate $-2\log e_n$, which is approximately a linear function of n. Here, e_n is calculated by using Monte Carlo method. The theoretical results from the previous chapter indicate that the IPS scheme is not asymptotically optimal since its large deviation rate of the second

moment is a function of N. However, as the number of particles N goes to ∞, the IPS scheme becomes asymptotically optimal.

As shown in Figure (6.2), the large deviation rate of IPS converges to the asymptotic optimal rate as N increases, which supports the theoretical results.

Another interesting question is what will the large deviation rate of the second moment of the IPS estimators change if we don't choose the optimal parameter for the threshold, i.e. $B \neq H(\alpha)$.

In Figure (6.3), we calculate the large deviation rate of the second moment of the IPS estimators for different values of B. In each case, the sample size M is 50,000, the particle number $N = 200$ and $n = 20$. From (4.3.1) we know that for each chosen B, there is a unique α_B such that

$$\frac{H(\alpha_B)}{\alpha_B} = \frac{B}{\alpha} = B.$$

In Figure (6.3), the x-axis represents different values of B, while the y-axis shows the corresponding large deviation rates. The dashed line represents the theoretical optimal rate $H(2 - \alpha_B) + H(\alpha_B)$, which is derived from the main result **Theorem** 4.4.1. The blue points represent the numerical results for various B values for B in $\{0.1 + 0.05j \mid \text{for } j = 0, 1, 2, ..., 10\} \cup \{\sqrt{2} - 1\}$. The large deviation rate of the second moment is approximated by computing

$$-\frac{1}{20} \log \left[\frac{1}{M} \sum_{i=0}^{M-1} \hat{e}_i^2 \right].$$

From the Figure (6.3), we observe that as B changes, the large deviation rate

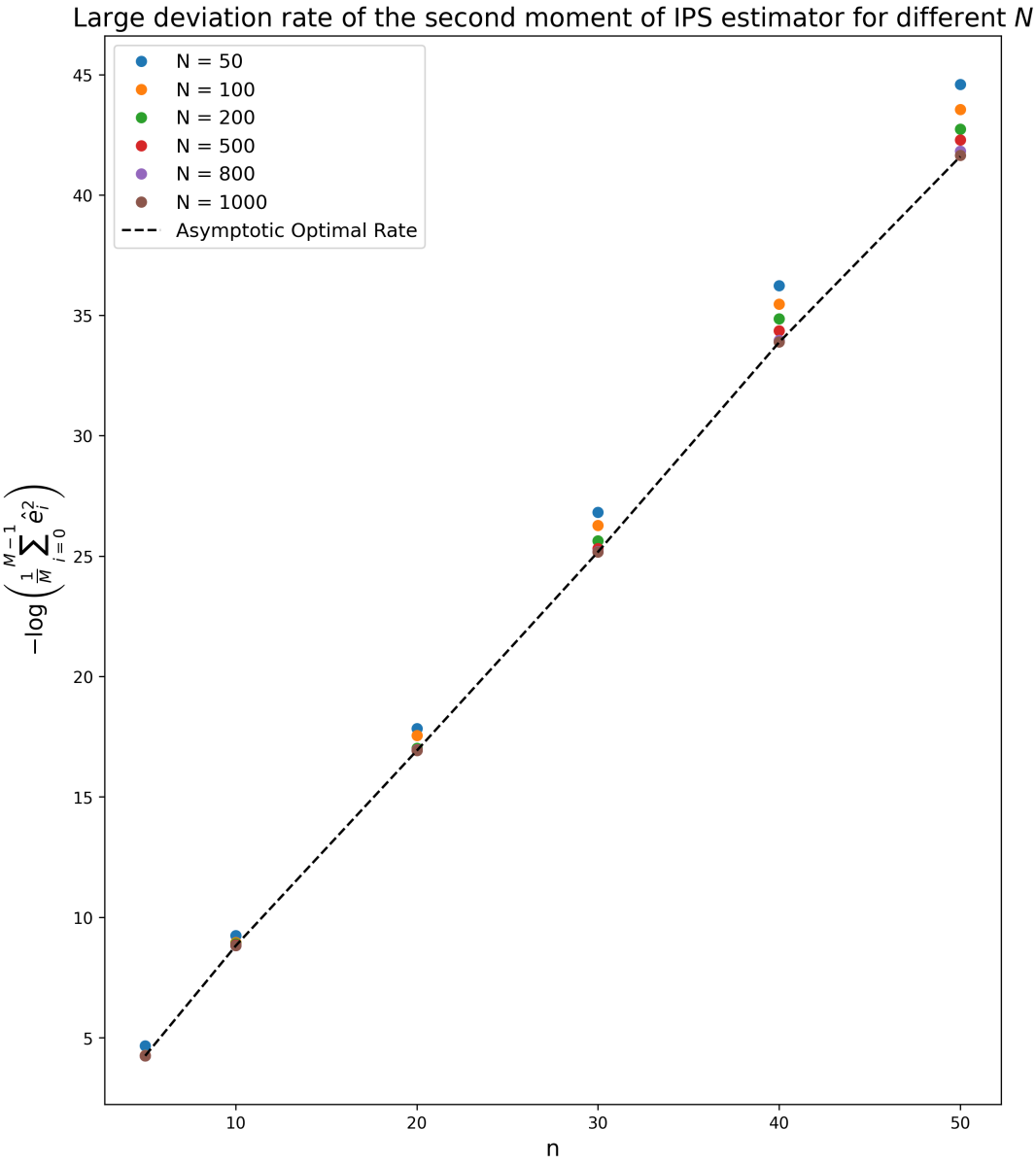

Figure 6.2: Large deviation rate of the second moment of the interacting particle system estimator for different n and N

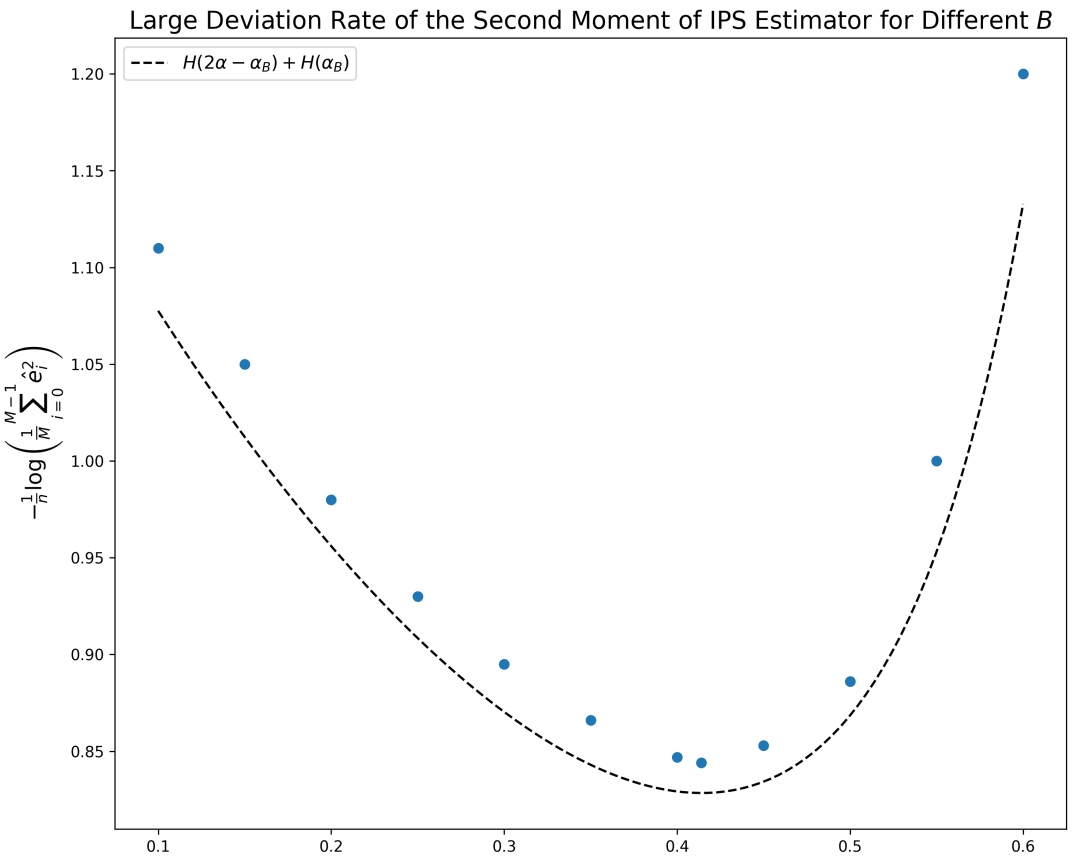

Figure 6.3: Large deviation rate of the second moment of the interacting particle system estimator for different B

follows a U-shaped curve. When we choose the optimal value of $B = H(\alpha) = \sqrt{2}-1$, the large deviation rate reach the minimum.

Therefore, this figure supports the theoretical findings, indicating that the choice of B is important for the IPS estimator's performance, and selecting the optimal B value can significantly reduce the estimation's deviation rate. Additionally, this suggests that if we can not calculate the exact value for $H(\alpha)$, we can use the approximate value for $H(\alpha)$ in the IPS scheme.

Chapter Seven

Appendix

7.1 Proof of Lemma 5.1.1

Using an argument by contradiction, it is enough to show that if $t_n \to t \in [0,1]$ then

$$\lim_{n\to\infty} \sup_{z\in\mathcal{X}} \left| \frac{1}{n} \log E_{z,t} e^{-nf(Y^{1,n}(1)-\bar{Y}^{1,n}(t_n))} + \inf_{\beta\in\mathbb{R}} \{(1-t)L(\beta) + f(\beta(1-t))\} \right| = 0. \tag{7.1.1}$$

The case $t = 1$ is straightforward and omitted. Except for the fact that being a failure particle constrains the trajectory of $Y^{1,n}$ for a vanishingly small time after t_n, the result would follow directly from Lemma 4.3.1, and the fact that this LDP implies the corresponding uniform Laplace principle for the quantities $\{(Y^{1,n}(1) - \bar{Y}^{1,n}(t_n))\}$ [2, Proposition 1.4]. For example, the uniformity with respect to z in Lemma 5.1.1 follows from the corresponding uniformity in Lemma 4.3.1. Since the paths of $Y^{1,n}$ are Lipschitz continuous with constant $\|c\|_\infty$, it is easy to check that

$$Y^{1,n}(1) - \bar{Y}^{1,n}(t_n) \text{ and } Y^{1,n}(1) - \bar{Y}^{1,n}(t)$$

satisfy a property called *super-exponential closeness*, and hence have the same large deviation asymptotics [2, Theorem 1.17]. Without the additional constraint that a failure particle cannot cross the next threshold in the next $\bar{h}/n^\eta$ units of time, this would imply (7.1.1), but with

$$\inf_{\beta\in\mathbb{R}} \{(1-t)L(\beta) + f(\beta(1-t))\}$$

replaced by

$$\inf_{\beta\in\mathbb{R}} \left\{ \inf_{\phi} \left\{ \int_t^1 L(\dot{\phi}(s))ds + f(\beta(1-t)) : \int_t^1 \dot{\phi}(s)ds = \beta(1-t) \right\} \right\},$$

where the infimum is over absolutely continuous ϕ, and we recover the previous expression using Jensen's inequality and the convexity of L.

Adding the constraint that the process not cross to the next threshold before $\bar{h}/n^{\eta}$ units of time have passed does not affect the upper bound, since it only reduces the probability of the event. It does however slightly complicate the proof of the lower bound, since one needs to know that with sufficiently high probability the process this will not happen. This is a type of *controllability* property. Although notationally detailed, it is in fact not difficult to prove such a result because of the uniform ergodicity and the fact that $\int c(x)\pi(dx) = 0$, where π is the stationary distribution of $\{X\}$. Suppose we shift so that the initial time t is 0. Then it is sufficient to argue that for any $\delta > 0$, we can keep $Y^n(s) \leq \Delta/n\alpha$ for all $s \leq \bar{h}/n^{\eta}$ with probability at least $e^{-n\delta}$, for all sufficiently large n and uniformly in z. This is the same as keeping $\int_0^s c(X(r))dr \leq \Delta/\alpha$ for $s \leq \bar{h}n^{1-\eta}$. Suppose we know that

$$\inf_{x\in\mathcal{X}} P_x \left\{ \sup_{0\leq s\leq 1} \int_0^s c(X(r))dr \leq \frac{\Delta}{\alpha} \text{ and } \int_0^1 c(X(r))dr \leq 0 \right\} \geq \sigma > 0. \tag{7.1.2}$$

Then by the Markov property $\int_0^s c(X(r))dr \leq \Delta/\alpha$ for $s \leq \bar{h}n^{1-\eta}$ occurs with probability at least $\sigma^{\lceil \bar{h}n^{1-\eta} \rceil}$ uniformly in x, and owing to $\eta > 0$ we have the further lower bound $\sigma^{\lceil \bar{h}n^{1-\eta} \rceil} \geq e^{-n\delta}$ for all sufficiently large n. The proof of (7.1.2) follows from that boundedness of c and the existence of a point $x^* \in \mathcal{X}$ with $c(x^*) < 0$. It uses many of the same arguments as Lemma 2.4.5 and is omitted. Since the lower bound also holds, (7.1.1) follows.

7.2 Proof of Lemma 5.1.2

We prove only the first statement of the lemma, since the proofs of the analogous statements are also analogous. Also, we prove the conditional Laplace principle that allows $(\bar{Y}^{1,n}(g_n^1), g_n^1)$ to be located anywhere on $\partial C_{\bar{\gamma}\Delta}$, although when applied later on we will be concerned only with events where $g_n^1 \leq 1$. Recall that

$$q^n(x, A) \doteq \theta(A \times W^n | x)$$

that for some probability measure ϱ and numbers $0 < c_1 < c_2 < \infty$

$$c_1 \varrho(A) \leq q^n(x, A) \leq c_2 \varrho(A), \tag{7.2.1}$$

and that for all Borel $A \subset \mathcal{X}$

$$q^n(x, A) \uparrow q^\infty(x, A).$$

Recall also that λ^n is uniquely defined by

$$\lambda^n(A) = \frac{\int_{\mathcal{X}} q^n(x, A)\lambda^n(dx)}{\int_{\mathcal{X}} q^n(x, \mathcal{X})\lambda^n(dx)}$$

and analogously for λ^∞. Let $\varkappa^n \doteq \int_{\mathcal{X}} q^n(x, \mathcal{X})\lambda^n(dx)$. Then we claim $\varkappa^n \to \varkappa^\infty$.

Although this is likely a known result, for completeness we provide a proof. It follows from (7.2.1) that for any $\mu \in \mathcal{P}(\mathcal{X})$

$$c_1 \varrho(A) \leq \int_{\mathcal{X}} q^n(x, A)\mu(dx) \leq c_2 \varrho(A),$$

and so from the definition of λ^n

$$\frac{c_1}{c_2}\varrho(A) \leq \lambda^n(A) \leq \frac{c_2}{c_1}\varrho(A),$$

and similarly for λ^∞. Thus these measures have densities f^n and f^∞ with respect to ϱ that are uniformly bounded below away from zero and from above. Hence we can consider them as elements of $L^2(d\varrho)$.

Now consider a subsequence (labeled by n) such that $f^n \to g$ weakly in $L^2(d\varrho)$ and $\varkappa_n \to \omega$. Then g must be a density, and we claim in fact that $\omega = \varkappa^\infty$ and $g = f^\infty$. We let $q^n(x, dy) = h^n(x, y)\varrho(dy)$ define $h^n(x, y)$ and use that, except possibly on a null set,

$$h^n(x, y) \uparrow h^\infty(x, y).$$

From the uniqueness of the solutions to the eigenvalue problem for λ^∞, we need only show that

$$\omega g(y)\varrho(dy) = \int_{\mathcal{X}} g(x)\varrho(dx)h^\infty(x, y)\varrho(dy). \tag{7.2.2}$$

We use that

$$\varkappa^n f^n(y)\varrho(dy) = \int_{\mathcal{X}} f^n(x)\varrho(dx)h^n(x, y)\varrho(dy).$$

We certainly have that $\varkappa^n f^n(y)\varrho(dy) \to \omega g(y)\varrho(dy)$ in the weak topology.[1] Also, if $m : \mathcal{X} \to \mathbb{R}$ is bounded and continuous

$$\begin{aligned}
&\int_{\mathcal{X}}\int_{\mathcal{X}} m(y)f^n(x)\varrho(dx)h^n(x, y)\varrho(dy) - \int_{\mathcal{X}}\int_{\mathcal{X}} m(y)g(x)\varrho(dx)h^\infty(x, y)\varrho(dy) \\
&\quad = \int_{\mathcal{X}}\int_{\mathcal{X}} m(y)f^n(x)\varrho(dx)[h^n(x, y) - h^\infty(x, y)]\varrho(dy) \\
&\qquad + \int_{\mathcal{X}}\int_{\mathcal{X}} m(y)[f^n(x) - g(x)]\varrho(dx)h^\infty(x, y)\varrho(dy).
\end{aligned}$$

[1] Or more precisely, the analogue of the weak topology for probability measures on $\mathcal{P}(\mathcal{X})$ for non-negative measures on $\mathcal{B}(\mathcal{X})$ with total mass less than c_2/c_1.

Then

$$\int_{\mathcal{X}}\int_{\mathcal{X}} m(y)f^n(x)\varrho(dx)[h^n(x,y)-h^\infty(x,y)]\varrho(dy)\to 0$$

by dominated convergence, and

$$\int_{\mathcal{X}}\int_{\mathcal{X}} m(y)[f^n(x)-g(x)]\varrho(dx)h^\infty(x,y)\varrho(dy)\to 0$$

by Fubini's theorem and by $f^n\to g$ weakly in $L^2(d\varrho)$. Thus for all bounded and continuous m

$$\int_{\mathcal{X}} m(y)\omega g(y)\varrho(dy)=\int_{\mathcal{X}}\int_{\mathcal{X}} m(y)g(x)\varrho(dx)h^\infty(x,y)\varrho(dy)$$

and (7.2.2) follows.

Since $\varkappa^n\to\varkappa^\infty$, given $\varepsilon>0$ we can choose $\bar{n}<\infty$ such that for $n\geq\bar{n}$

$$\int_{\mathcal{X}} q^n(x,\mathcal{X})\lambda^n(dx)\geq(1-\varepsilon)\int_{\mathcal{X}} q^\infty(x,\mathcal{X})\lambda^\infty(dx).$$

Let $\mathcal{A}^*(\gamma_n)$ denote the event that particle 1 makes it to the next threshold, without a restriction on the time it allowed for this to occur, $n\gamma_n$ times in a row. Then using the uniform ergodicity, uniformly in $z\in\mathcal{X}$

$$\liminf_{n\to\infty}\frac{1}{n}\log P_z\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}\geq\liminf_{n\to\infty}\frac{1}{n}\log P_z\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}+\bar{\gamma}\log(1-\varepsilon),$$

and by sending $\varepsilon\downarrow 0$

$$\liminf_{n\to\infty}\frac{1}{n}\log P_z\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}\geq\liminf_{n\to\infty}\frac{1}{n}\log P_z\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}. \tag{7.2.3}$$

Next consider

$$\frac{1}{n}\log E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}\middle|\,\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\right].$$

Since $\mathbf{1}_{\mathcal{A}(\gamma_n)}=1$ implies $g_n^1<\infty$, we can rewrite this as

$$E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}\middle|\,\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\right]=E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}1_{\{g_n^1<\infty\}}1_{\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}}\right]P_z\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}. \tag{7.2.4}$$

Let $g_n \doteq \inf\{t\geq 0:(Y^n(t),t)\in\partial C_{\gamma_n\Delta}\}$, where Y^n is the original process. Then the events $\{g_n<\infty\}$ and $\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}$ have the same probability. Thus by (7.2.3), $P_z\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}\leq P_z\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}$, and the large deviation properties of the original process to obtain the last equality (see Lemma 7.3.4) we obtain

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\log P_z\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}\\
=&\lim_{n\to\infty}\frac{1}{n}\log P_z\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}\\
=&\lim_{n\to\infty}\frac{1}{n}\log P_z\{g_n<\infty\}\\
=&-\Delta\gamma\frac{H(\alpha_B)}{B}.
\end{aligned} \tag{7.2.5}$$

Similarly

$$E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}1_{\{g_n^1<\infty\}}1_{\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}}\right]\leq E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}1_{\{g_n^1<\infty\}}1_{\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}}\right]$$

and

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{n}\log E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}1_{\{g_n^1<\infty\}}1_{\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}}\right]\\
\geq&\liminf_{n\to\infty}\frac{1}{n}\log E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}1_{\{g_n^1<\infty\}}1_{\{\mathbf{1}_{\mathcal{A}^*(\gamma_n)}=1\}}\right]+\|f\|_\infty\bar{\gamma}\log(1-\varepsilon),
\end{aligned}$$

and therefore by the argument analogous to that used for $P_z\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}$, uniformly

in z

$$\lim_{n\to\infty}\frac{1}{n}\log E_z\left[e^{-nf(\bar{Y}^{1,n}(g_n^1),g_n^1)}1_{\{g_n^1\le 1\}}1_{\{\mathbf{1}_{\mathcal{A}(\gamma_n)}=1\}}\right]=\lim_{n\to\infty}\frac{1}{n}\log E_z\left[e^{-nf(Y^n(g_n),g_n)}1_{\{g_n\le 1\}}\right] \tag{7.2.6}$$

$$=\inf_{(\beta,s)\in\partial C_{\bar{\gamma}\Delta}}\{sL(\beta/s)+f(\beta,s)\}\,.$$

The conclusion of the lemma follows by combining (7.2.4), (7.2.5) and (7.2.6).

7.3 Proof of Lemma 5.1.3

To simplify the notation we assume $\theta = 1$. Since we will consider limits when both $n \to \infty$ (the large deviation parameter) and $N \to \infty$ (the number of particles), in this section we make the dependence on N explicit in the notation. Consider the Markov process $\{\boldsymbol{X}_k^N, k = 1, \ldots, n\}$ on $\bar{\mathcal{X}}^N \subset \mathcal{X}^N$, with a generic point in $\mathcal{X}^N$ denoted by $\boldsymbol{x}^N = (x_1, \ldots, x_N)$, and with the ith component of $\boldsymbol{X}_k^N$ denoting the $\mathcal{X}$-component of particle i at the beginning of threshold k. We recall that these random variables take values in $\bar{\mathcal{X}} = \{x \in \mathcal{X} : c(x) \ge B/\alpha\}$. Owing to the fact that a success particle must reach the next threshold by time $\bar{h}/n^\eta$, there is also an n dependence that is not made explicit in the notation for $\boldsymbol{X}_k^N$. Also, throughout this section we construct $\{\boldsymbol{X}_k^N\}$ under the condition that particle 1 is a success particle, without making this explicit in the notation.

The transition kernel for $\{\boldsymbol{X}_k^N\}$ can be constructed as follows. Let

$$\left\{s_k^n(x,i), x \in \bar{\mathcal{X}}, i = 2, \ldots, N, k = 1, \ldots, n\right\}$$

be independent Bernoulli random fields, with

$$P\{s_k^n(x,i)=1\}=p^n(x)\in(0,1).$$

Given $\boldsymbol{X}_k^N=\boldsymbol{x}^N$, for $S\subset\{1,2,\ldots,N\}$ let

$$\rho^{n,N}(S,\boldsymbol{x}^N)=\prod_{i\in S\setminus\{1\}}p^n(x_i)\cdot\prod_{i\notin S}[1-p^n(x_i)].$$

S identifies success particles, and note that always $1\in S$ with $s_k^n(x,1)=1$.

Define the transition kernel $p^n(x,dy)\doteq q^n(x,dy)/p^n(x)$, and similarly $p^\infty(x,dy)\doteq q^\infty(x,dy)/p^\infty(x)$. For those i such that $s_k^n(x_i,i)=1$ let $\{X_{k+1}^N(i),i\in S\}$ have the joint distribution $\prod_{i\in S}p^n(x_i,d\bar{x}_i)$, and lastly apply the resampling rule to complete $\boldsymbol{X}_k^N$, i.e., to determine $\{X_{k+1}^N(i),i\notin S\}$. If $p_S^{n,N}(\boldsymbol{x}^N,d\boldsymbol{y}^N)$ denotes the resulting transition kernel, then for each S the Feller property of the single particle transition kernel (see Condition 4.1.1) $p^n(x,d\bar{x})$ implies that $p_S^{n,N}(\boldsymbol{x}^N,d\boldsymbol{y}^N)$ is also Feller. Since the overall transition kernel is the sum

$$q^{n,N}(\boldsymbol{x}^N,d\boldsymbol{y}^N)=\sum_{S:1\in S}\rho^{n,N}(S,\boldsymbol{x}^N)p_S^{n,N}(\boldsymbol{x}^N,d\boldsymbol{y}^N)$$

it is also Feller. Since also $\mathcal{X}^N$ is compact $q^{n,N}(\boldsymbol{x}^N,d\boldsymbol{y}^N)$ must have a unique stationary distribution, which we denote by $\lambda^{n,N}$. We note that each marginal of $\lambda^{n,N}$ is λ^n. To see this, one can start with an initial distribution $\theta_0^{n,N}(d\boldsymbol{x}^N)$ that is product measure with marginals λ^n. If $\theta_k^{n,N}$ is the distribution at time k of the chain with transitions $q^{n,N}$ and initial distribution $\theta_0^{n,N}$, then all marginals of $\theta_k^{n,N}$ are λ^n, and since $\theta_k^{n,N}\Rightarrow\lambda^{n,N}$ it is also true for $\lambda^{n,N}$.

The next result will allow us to deduce the large deviation properties of the

empirical measure for the number of success particles from those of the empirical measure for $\{\boldsymbol{X}_k^N\}$.

Lemma 7.3.1. Let $\{Z_i\}$ be N independent Bernoulli random variables with success parameters $p_i \in (0,1]$, and let

$$b \doteq \frac{1}{N} E \sum_{i=1}^{N} Z_i = \frac{1}{N} \sum_{i=1}^{N} p_i.$$

Then

$$b^2 \leq E\left(\frac{1}{N}\sum_{i=1}^{N} Z_i\right)^2 \leq b^2 + \frac{1}{4N}.$$

Proof. Using the independence

$$\begin{aligned} E\left(\frac{1}{N}\sum_{i=1}^{N} Z_i\right)^2 &= E\left(\frac{1}{N}\sum_{i=1}^{N} p_i + \frac{1}{N}\sum_{i=1}^{N}(Z_i - p_i)\right)^2 \\ &= E\left(\frac{1}{N}\sum_{i=1}^{N} p_i\right)^2 + E\left(\frac{1}{N}\sum_{i=1}^{N}(Z_i - p_i)\right)^2 \\ &= b^2 + \frac{1}{N}\cdot\frac{1}{N} E\sum_{i=1}^{N}(Z_i - p_i)^2 = b^2 + \frac{1}{N}\cdot\frac{1}{N}\sum_{i=1}^{N}(p_i - p_i^2). \end{aligned}$$

The result now follows since $p \to p - p^2$ has a minimum of zero and a maximum of $1/4$. □

The next result follows from the previous lemma and a straightforward conditioning argument.

Lemma 7.3.2.

$$E_{\lambda^{n,N}}\left[\prod_{k=1}^{n}\left(\frac{N_k^S}{N}\right)\middle|\, \mathbf{1}_{\mathcal{A}(n)} = 1\right] = \frac{1}{N^n} E_{\lambda^{n,N}}\left[\prod_{k=0}^{n-1}\left[1 + \sum_{i=2}^{N} p^n(\boldsymbol{X}_k^N(i))\right]\right]$$

and

$$\frac{1}{N^{2n}}E_{\lambda^{n,N}}\left[\prod_{k=0}^{n-1}\left[1+\sum_{i=2}^{N}p^n(\boldsymbol{X}_k^N(i))\right]^2\right] \le E_{\lambda^{n,N}}\left[\prod_{k=1}^{n}\left(\frac{N_k^S}{N}\right)^2\middle|\mathbf{1}_{\mathcal{A}(n)}=1\right]$$
$$\le \frac{1}{N^{2n}}E_{\lambda^{n,N}}\left[\prod_{k=0}^{n-1}\left(\left[1+\sum_{i=2}^{N}p^n(\boldsymbol{X}_k^N(i))\right]^2+\frac{1}{4}\right)\right].$$

We next present a result on large deviation asymptotics in n for fixed N.

Lemma 7.3.3. Let $f:\bar{\mathcal{X}}\to(0,\infty)$ be continuous. Then for $r=1,2$

$$\liminf_{n\to\infty}\frac{1}{n}\log E_{\lambda^{n,N}}e^{\sum_{k=0}^{n-1}r\log\left[\frac{1}{N}\left(\sum_{i=1}^{N}f(\boldsymbol{X}_k^N(i))\right)\right]}\ge r\log\left(\int_{\bar{\mathcal{X}}}f(x)\lambda^{\infty}(dx)\right).$$

Moreover, there there are $C_N^1, C_N^2\in[0,\infty)$ with $C_N^r\to 0$ as $N\to\infty$, $r=1,2$, such that

$$\limsup_{n\to\infty}\frac{1}{n}\log E_{\lambda^{n,N}}e^{\sum_{k=0}^{n-1}r\log\left[\frac{1}{N}\left(\sum_{i=1}^{N}f(\boldsymbol{X}_k^N(i))\right)\right]}\le r\left[\log\left(\int_{\bar{\mathcal{X}}}f(x)\lambda^{\infty}(dx)\right)+C_N^r\right].$$

In addition, results of the same form but with possibly different C_N^r hold when we condition on two distinct particles being repeatedly successful up to threshold n.

Proof. The proof is based on a representation formula. Let $\xi^{n,N}$ denote the empirical measure of $\{\boldsymbol{X}_k^N\}$ in the k variable:

$$\xi^{n,N}(d\boldsymbol{x}^N)\doteq\frac{1}{n}\sum_{k=0}^{n-1}\delta_{\boldsymbol{X}_k^N}(d\boldsymbol{x}^N).$$

Recall that $\Theta^N(\boldsymbol{x}^N)\doteq\frac{1}{N}\sum_{i=1}^{N}\delta_{\boldsymbol{x}^N(i)}$. We can rewrite

$$E_{\lambda^{n,N}}e^{\sum_{k=0}^{n-1}r\log\left[\frac{1}{N}\left(\sum_{i=1}^{N}f(\boldsymbol{X}_k^N(i))\right)\right]}=E_{\lambda^{n,N}}e^{nr\int_{\bar{\mathcal{X}}^N}\log\left[\int_{\bar{\mathcal{X}}}f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right]\xi^{n,N}(d\boldsymbol{x}^N)}.$$

We next state a control representation for the quantity in the last display. Relative entropy representations of this sort were first used in [9], and the proof of the representation is very similar to that used in the large deviation analysis of the empirical measure of a Markov chain in Chapter 6 of [4]. Recall that $q^{n,N}$ is the overall transition kernel for $\{\boldsymbol{X}_k^N\}$ given that particle 1 is always a success particle. Following the same line of proof as in [4], we have the formula

$$\begin{aligned}&\frac{1}{n}\log E_{\lambda^{n,N}} e^{nr\int_{\bar{\mathcal{X}}^N}\log\left[\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right]\xi^{n,N}(d\boldsymbol{x}^N)}\\ &\quad=\sup E_{\lambda^{n,N}}\left[r\int_{\bar{\mathcal{X}}^N}\log\left[\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right]\bar{\xi}^{n,N}(d\boldsymbol{x}^N)-R\left(\gamma^{n,N}\middle\|\,\bar{\xi}^{n,N}\otimes q^{n,N}\right)\right],\end{aligned}$$

where the new quantities $\gamma^{n,N}$ and $\bar{\xi}^{n,N}$ are defined as follows. Instead of using the original transition kernel $q^{n,N}$ as in the construction of $\{\boldsymbol{X}_k^N\}$, we consider a sequence of controls $\{\bar{q}_k^{n,N}\}$ which pick the conditional distribution of a controlled process $\{\bar{\boldsymbol{X}}_k^N\}$, in that $\bar{q}_k^{n,N}(d\boldsymbol{x}^N|\bar{\boldsymbol{X}}_0^N,\ldots,\bar{\boldsymbol{X}}_{k-1}^N)$ is the distribution of $\bar{\boldsymbol{X}}_k^N$ given the previous states of the process. Then $\bar{\xi}^{n,N}(d\boldsymbol{x}^N)$ is the empirical measure of $\{\bar{\boldsymbol{X}}_k^N\}$ up to threshold n, and we define

$$\bar{\xi}^{n,N}\otimes q^{n,N}(d\boldsymbol{x}^N,d\boldsymbol{y}^N)\doteq\frac{1}{n}\sum_{k=0}^{n-1}\delta_{\bar{\boldsymbol{X}}_k^N}(d\boldsymbol{x}^N)q^{n,N}(\bar{\boldsymbol{X}}_k^N,d\boldsymbol{y}^N),$$

and

$$\gamma^{n,N}(d\boldsymbol{x}^N,d\boldsymbol{y}^N)\doteq\frac{1}{n}\sum_{k=0}^{n-1}\delta_{\bar{\boldsymbol{X}}_k^N}(d\boldsymbol{x}^N)\bar{q}_k^{n,N}(d\boldsymbol{y}^N|\bar{\boldsymbol{X}}_0^N,\ldots,\bar{\boldsymbol{X}}_k^N).$$

Let $[\mu]_m$ indicate the mth marginal of a joint distribution μ on a product space. Key properties that we will need are that $\bar{\xi}^{n,N}=[\gamma^{n,N}]_1$, and that by using a standard conditioning argument as in [4, Lemma 6.12], for any bounded and continuous $g:\boldsymbol{\mathcal{X}}^N\to\mathbb{R}$

$$E_{\lambda^{n,N}}\left[\int_{\mathcal{X}^N}g(\boldsymbol{x}^N)[\gamma^{n,N}]_1(d\boldsymbol{x}^N)-\int_{\mathcal{X}^N}g(\boldsymbol{x}^N)[\gamma^{n,N}]_2(d\boldsymbol{x}^N)\right]^2\le 4\left\|g\right\|_\infty/n. \quad (7.3.1)$$

In particular, this is true for $g(\boldsymbol{x}^N) = \int_{\bar{\mathcal{X}}^N} h(x)\Theta^N(\boldsymbol{x}^N)(dx)$ when h is bounded and continuous. We think of $\bar{\xi}^{n,N}(d\boldsymbol{x}^N)$ as a controlled empirical measure, and since by the chain rule

$$R\left(\gamma^{n,N}\middle\|\, \bar{\xi}^{n,N}\otimes q^{n,N}\right) = \frac{1}{n}\sum_{k=0}^{n-1} R\left(\bar{q}_k^{n,N}(d\boldsymbol{y}^N|\bar{\boldsymbol{X}}_0^N,\ldots,\bar{\boldsymbol{X}}_k^N)\middle\|\, q^{n,N}(\bar{\boldsymbol{X}}_k^N, d\boldsymbol{y}^N)\right),$$

we pay relative entropy cost to perturb away from the original dynamics.

As noted previously, $q^{\infty,N}$ has a unique stationary distribution $\lambda^{\infty,N}(d\boldsymbol{x}^N)$, and this distribution has the property that each marginal is λ^∞. We first argue that for $r = 1, 2$,

$$\begin{aligned}&\liminf_{n\to\infty}\sup E_{\lambda^{n,N}}\left[r\int_{\bar{\mathcal{X}}^N}\log\left[\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right]\bar{\xi}^{n,N}(d\boldsymbol{x}^N) - R\left(\gamma^{n,N}\middle\|\,\bar{\xi}^{n,N}\otimes q^{n,N}\right)\right]\\ &\quad\geq r\int_{\bar{\mathcal{X}}^N}\log\left(\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right)\lambda^{\infty,N}(d\boldsymbol{x}^N).\end{aligned}$$

In the variational representation we choose

$$\bar{q}_k^{n,N}(d\boldsymbol{x}^N|\bar{\boldsymbol{X}}_0^N,\ldots,\bar{\boldsymbol{X}}_{k-1}^N) = q^{n,N}(\bar{\boldsymbol{X}}_{k-1}^N, d\boldsymbol{x}^N),$$

in which case the relative entropy term disappears. Suppose we show that given any subsequence of $\{n\}$, there is a further subsequence (again denoted by $\{n\}$) with

$$\begin{aligned}&\liminf_{n\to\infty} E_{\lambda^{n,N}}\left[r\int_{\bar{\mathcal{X}}^N}\log\left[\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right]\bar{\xi}^{n,N}(d\boldsymbol{x}^N)\right]\\ \geq& r\int_{\bar{\mathcal{X}}^N}\log\left(\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right)\lambda^{\infty,N}(d\boldsymbol{x}^N).\end{aligned}\tag{7.3.2}$$

Then the same statement holds for the original sequence, and also for the supremum over possible controls.

Since $p^n(x) \to p^\infty(x)$ and $q^n(x,dy) \uparrow q^\infty(x,dy)$, for any bounded and continuous f [in fact even for bounded and measurable f] $\int_{\bar{\mathcal{X}}} f(y)p^n(x,dy) \to \int_{\bar{\mathcal{X}}} f(y)p^\infty(x,dy)$, and the analogous statement is therefore true for $q^{n,N}(\boldsymbol{x}^N, d\boldsymbol{y}^N)$. Since $\bar{\xi}^{n,N}(d\boldsymbol{x}^N)$ takes values in a compact set, for any subsequence of $\{n\}$, there is a further subsequence (again denoted $\{n\}$) along which $\bar{\xi}^{n,N}(d\boldsymbol{x}^N)$ converges in distribution to a limit $\bar{\xi}^{\infty,N}(d\boldsymbol{x}^N)$. It then follows from (7.3.1) and Fatou's lemma that

$$\int_{\mathcal{X}^N} g(\boldsymbol{x}^N)\bar{\xi}^{\infty,N}(d\boldsymbol{x}^N) = \int_{\mathcal{X}^N}\int_{\mathcal{X}^N} g(\boldsymbol{y}^N)q^{\infty,N}(\boldsymbol{x}^N, d\boldsymbol{y}^N)\bar{\xi}^{\infty,N}(d\boldsymbol{x}^N)$$

a.s., and therefore $\bar{\xi}^{n,N}(d\boldsymbol{x}^N)$ converges in distribution to $\lambda^{\infty,N}(d\boldsymbol{x}^N)$. Hence (7.3.2) follows by dominated convergence.

To conclude the argument we need to show that as $N \to \infty$ the right hand side of (7.3.2) converges to $r\log\left(\int_{\bar{\mathcal{X}}} f(x)\lambda^\infty(dx)\right)$. We will do this as part of a more general result that will also be needed for the upper bound. We define a mapping from $\mathcal{P}((\bar{\mathcal{X}}^N)^2)$ to $\mathcal{P}(\mathcal{P}(\bar{\mathcal{X}})^2)$ by $\Upsilon^N : m(d\boldsymbol{x}^N, d\boldsymbol{y}^N) \to \sigma(d\mu, d\nu)$ if

$$\sigma(A \times B) = m(\{\boldsymbol{x}^N, \boldsymbol{y}^N\} : \Theta^N(\boldsymbol{x}^N) \in A, \Theta^N(\boldsymbol{y}^N) \in B)).$$

Thus Υ^N maps a distribution on sequences of the form $\{\boldsymbol{x}^N, \boldsymbol{y}^N\}$ to a distribution on pairs of probability measures. Consider any sequence $\{\gamma^{\infty,N}\}_{N\in\mathbb{N}}$ taking values in $\mathcal{P}((\bar{\mathcal{X}}^N)^2)$ such that

$$\limsup_{N\to\infty} E_{\lambda^{\infty,N}}\left[R\left(\gamma^{\infty,N}\middle\|\, [\gamma^{\infty,N}]_1 \otimes q^{\infty,N}\right)\right] < \infty \tag{7.3.3}$$

and

$$E_{\lambda^{\infty,N}}\left[\int_{\mathcal{X}^N} g(\boldsymbol{x}^N)[\gamma^{\infty,N}]_1(d\boldsymbol{x}^N) - \int_{\mathcal{X}^N} g(\boldsymbol{x}^N)[\gamma^{\infty,N}]_2(d\boldsymbol{x}^N)\right]^2 = 0 \tag{7.3.4}$$

hold. (Note that both these statements are true when $[\gamma^{\infty,N}] = \lambda^{\infty,N} \otimes q^{\infty,N}$.) By the information inequality for relative entropy [4, part (f) of Lemma 2.4]

$$R\left(\Upsilon^N[\gamma^{\infty,N}] \middle\| \Upsilon^N[[\gamma^{\infty,N}]_1 \otimes q^{\infty,N}]\right) \leq R\left(\gamma^{\infty,N} \middle\| [\gamma^{\infty,N}]_1 \otimes q^{\infty,N}\right). \tag{7.3.5}$$

Since $\{\Upsilon^N[\gamma^{\infty,N}]\}$ and $\{\Upsilon^N[[\gamma^{\infty,N}]_1 \otimes q^{\infty,N}]\}$ take values in the compact space $\mathcal{P}(\mathcal{P}(\bar{\mathcal{X}})^2)$, they are automatically tight as sequences of $\mathcal{P}(\mathcal{P}(\bar{\mathcal{X}})^2)$-valued random variables, and therefore converge in distribution, at least along a subsequence.

Let $\Upsilon^N[\gamma^{\infty,N}](d\mu, d\nu)$ converge in distribution to $\rho(d\mu, d\nu)$. By (7.3.4) with $g(\boldsymbol{x}^N) = \int_{\bar{\mathcal{X}}^N} h(x)\Theta^N(\boldsymbol{x}^N)(dx)$

$$E_{\lambda^{\infty,N}}\left[\int_{\mathcal{X}} h(x)[\Upsilon^N[\gamma^{\infty,N}]]_1(dx) - \int_{\mathcal{X}} h(x)[\Upsilon^N[\gamma^{\infty,N}]]_2(dx)\right]^2 \leq 0.$$

Since $\|g\|_\infty$ is uniformly bounded when $\|h\|_\infty$ is, $[\rho]_1 = [\rho]_2$ a.s. It also follows from Condition 4.3.4 that the limit in distribution of $\Upsilon^N[[\gamma^{\infty,N}]_1 \otimes q^{\infty,N}](d\mu, d\nu)$ is $[\rho]_1(d\mu)\delta_{g(\mu)}(d\nu)$.

By (7.3.3) and Fatou's lemma we have that $R\left(\rho(d\mu, d\nu) \| [\rho]_1(d\mu)\delta_{g(\mu)}(d\nu)\right) < \infty$ a.s. Then $[\rho]_1(d\mu) = [\rho]_2(d\mu)$ and $\rho(d\mu, d\nu) = [\rho]_1(d\mu)\delta_{g(\mu)}(d\nu)$ imply that $[\rho]_1$ is invariant under the deterministic dynamic $\mu \to g(\mu)$. We claim that since λ^∞ is the unique fixed point of g, this implies $[\rho]_1(d\mu) = \delta_{\lambda^\infty}(d\mu)$. Indeed, if $\bar{\mu}_1$ is a random probability measure with distribution $[\rho]_1$ and we define $\bar{\mu}_{n+1} = g(\bar{\mu}_n)$, then each $\bar{\mu}_n$ has distribution $[\rho]_1$. But then $\bar{\mu}_n \to \lambda^\infty$ implies $[\rho]_1(d\mu) = \delta_{\lambda^\infty}(d\mu)$. Therefore by dominated convergence

$$\begin{aligned} &E_{\lambda^{\infty,N}}\left[r\int_{\bar{\mathcal{X}}^N} \log\left[\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right][\gamma^{\infty,N}]_1(d\boldsymbol{x}^N)\right] \\ =&E_{\lambda^{\infty,N}}\left[r\log\left[\int_{\bar{\mathcal{X}}} f(x)[\Upsilon^N[\gamma^{\infty,N}]]_1(dx)\right]\right] \end{aligned} \tag{7.3.6}$$

$$\to r \log \left[\int_{\bar{\mathcal{X}}} f(x) \lambda^{\infty}(dx) \right].$$

Since as observed both (7.3.3) and (7.3.4) hold when $[\gamma^{\infty,N}] = \lambda^{\infty,N} \otimes q^{\infty,N}$, as a special case we obtain

$$\lim_{N\to\infty} r \int_{\bar{\mathcal{X}}^N} \log \left(\int_{\bar{\mathcal{X}}} f(x) \Theta^N(\boldsymbol{x}^N)(dx) \right) \lambda^{\infty,N}(d\boldsymbol{x}^N) = r \log \left[\int_{\bar{\mathcal{X}}} f(x) \lambda^{\infty}(dx) \right]$$

and complete the proof of the lower bound.

We next turn to the proof of the upper bound, and consider

$$B_N = \limsup_{n\to\infty} \sup_{\{\bar{q}_k^{n,N}\}} E_{\lambda^{n,N}} \left[r \int_{\bar{\mathcal{X}}^N} \log \left[\int_{\bar{\mathcal{X}}} f(x) \Theta^N(\boldsymbol{x}^N)(dx) \right] \bar{\xi}^{n,N}(d\boldsymbol{x}^N) - R\left(\gamma^{n,N} \middle\| \bar{\xi}^{n,N} \otimes q^{n,N} \right) \right]$$

Consider an arbitrary sequence $\{\bar{q}_k^{n,N}\}$. Now using that $\gamma^{n,N}$ also takes values in a compact set, it converges (along a subsequence) to a limit that we label $\gamma^{\infty,N}$. Then as for the lower bound Condition 4.1.1 implies $\bar{\xi}^{n,N} \otimes q^{n,N} \Rightarrow [\gamma^{\infty,N}]_1 \otimes q^{\infty,N}$ in distribution. By the dominated convergence theorem, Fatou's lemma and the lower semicontinuity of relative entropy,

$$\begin{aligned} &\limsup_{n\to\infty} E_{\lambda^{n,N}} \left[r \int_{\bar{\mathcal{X}}^N} \log \left[\int_{\bar{\mathcal{X}}} f(x) \Theta^N(\boldsymbol{x}^N)(dx) \right] \bar{\xi}^{n,N}(d\boldsymbol{x}^N) - R\left(\gamma^{n,N} \middle\| \bar{\xi}^{n,N} \otimes q^{n,N} \right) \right] \\ &\quad \le E_{\lambda^{\infty,N}} \left[r \int_{\bar{\mathcal{X}}^N} \log \left[\int_{\bar{\mathcal{X}}} f(x) \Theta^N(\boldsymbol{x}^N)(dx) \right] [\gamma^{\infty,N}]_1(d\boldsymbol{x}^N) - R\left(\gamma^{\infty,N} \middle\| [\gamma^{\infty,N}]_1 \otimes q^{\infty,N} \right) \right]. \end{aligned}$$

Since $\{\bar{q}_k^{n,N}\}$ is arbitrary we arrive at

$$B_N \le E_{\lambda^{\infty,N}} \left[r \int_{\bar{\mathcal{X}}^N} \log \left[\int_{\bar{\mathcal{X}}} f(x) \Theta^N(\boldsymbol{x}^N)(dx) \right] [\gamma^{\infty,N}]_1(d\boldsymbol{x}^N) - R\left(\gamma^{\infty,N} \middle\| [\gamma^{\infty,N}]_1 \otimes q^{\infty,N} \right) \right],$$

where it follows from (7.3.1) and Fatou's lemma that (7.3.4) holds.

We next argue that

$$\limsup_{N\to\infty} E_{\lambda^{\infty,N}}\left[r\int_{\bar{\mathcal{X}}^N}\log\left[\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)\right][\gamma^{\infty,N}]_1(d\boldsymbol{x}^N) - R\left(\gamma^{\infty,N}\middle\|\,[\gamma^{\infty,N}]_1\otimes q^{\infty,N}\right)\right]$$
$$- r\log\left(\int_{\bar{\mathcal{X}}} f(x)\lambda^{\infty}(dx)\right)\le 0, \tag{7.3.7}$$

Note that we can assume (7.3.3) holds since if the limit is ∞ then (7.3.7) is automatic. Since also (7.3.4) holds it follows from (7.3.6) that

$$\begin{aligned}
&E_{\lambda^{\infty,N}}\left[r\log\left[\int_{\bar{\mathcal{X}}^N}\int_{\bar{\mathcal{X}}} f(x)\Theta^N(\boldsymbol{x}^N)(dx)[\gamma^{\infty,N}]_1(d\boldsymbol{x}^N)\right]\right]\\
=&E_{\lambda^{\infty,N}}\left[r\log\left[\int_{\bar{\mathcal{X}}} f(x)[\Upsilon^N[\gamma^{\infty,N}]]_1(dx)\right]\right]\\
\to& r\log\left[\int_{\bar{\mathcal{X}}} f(x)\lambda^{\infty}(dx)\right].
\end{aligned}$$

Since $R(\cdot\|\cdot)\ge 0$, this implies (7.3.7), and therefore $B_N \le r\log\left(\int_{\bar{\mathcal{X}}} f(x)\lambda^{\infty}(dx)\right) + C_N$ with $C_N\to 0$, and completes the proof of the upper bound. □

Lemma 7.3.4. Suppose that Conditions 2.2.1, 4.1.1 and 4.3.2 hold, and that $\gamma_n\to\bar{\gamma}$. Then with Y^n the original single particle process, uniformly in $z\in\mathcal{X}$

$$\lim_{n\to\infty}\frac{1}{n}\log P_z\left(\inf\left\{t\ge 0:(Y^n(t),t)\in\partial C^n_{n\gamma_n}\right\}<\infty\right) = -\bar{\gamma}\Delta\frac{H(\alpha_B)}{B}.$$

If in addition Condition 4.3.4 holds, then with λ^{∞} and $p^{\infty}(x)$ as defined in Section 5 below (4.1.2)

$$\log\left(\int_{\bar{\mathcal{X}}} p^{\infty}(x)\lambda^{\infty}(dx)\right) = -\Delta\frac{H(\alpha_B)}{B}.$$

Proof. Using the large deviation properties of $\{Y^n\}$ and convexity of L we have

$$\lim_{n\to\infty}\frac{1}{n}\log P_z\left(\inf\left\{t\ge 0:(Y^n(t),t)\in\partial C^n_{n\gamma_n}\right\}<\infty\right) \tag{7.3.8}$$

$$= -\inf\{tL(\beta) : -\alpha t\beta + Bt = -\Delta, t \in [0,\infty)\} \tag{7.3.9}$$
$$= -\inf\left\{\frac{\Delta}{\alpha\beta - B}L(\beta) : \beta > B/\alpha\right\}.$$

We recall the relations $H(\alpha_B)/\alpha_B = B/\alpha$,

$$H(\alpha_B) + L(\beta_B) - \alpha_B\beta_B = 0 \text{ and } L'(\beta_B) = \alpha_B,$$

where $\beta_B = H'(\alpha_B)$ is the point dual to α_B. It is easy to check that the optimal β in (7.3.9) is β_B. Indeed, if there is a solution to

$$(B - \alpha\beta)L'(\beta) + \alpha L(\beta) = 0$$

that also satisfies the constraint $\beta > B/\alpha$ then it will be the minimizer. Inserting $\beta = \beta_B$ and using the identities above we find

$$(B-\alpha\beta_B)L'(\beta_B)+\alpha L(\beta_B) = B\alpha_B-\alpha\beta_B\alpha_B+\alpha\alpha_B\beta_B-\alpha H(\alpha_B) = \alpha\left[H(\alpha_B) - H(\alpha_B)\right] = 0.$$

Also,

$$\alpha\beta_B - B = \alpha\beta_B - \alpha\frac{H(\alpha_B)}{\alpha_B} = \frac{\alpha}{\alpha_N}\left[\alpha_B\beta_B - H(\alpha_B)\right] = \frac{\alpha}{\alpha_N}L(\beta_B) > 0$$

shows that the constraint is satisfied. We then find that the minimizing value is

$$-\Delta\left\{\frac{1}{\alpha\beta_B - B}L(\beta_B)\right\} = -\Delta\frac{H(\alpha_B)}{B}\left\{\frac{\alpha_B\beta_B/H(\alpha_B) - 1}{\alpha\beta_B/B - 1}\right\} = -\Delta\frac{H(\alpha_B)}{B}.$$

For the second part, we use $\gamma_n = 1$ and that the product of the fraction of success particles provides an unbiased estimate of the probability that the original

single particle process $\{Y^n\}$ reaches the nth threshold:

$$E_{\lambda^n}\prod_{m=0}^{n-1}\frac{N_m^S}{N}=P_{\lambda^n}\left(\inf\left\{t\geq 0:(Y^n(t),t)\in\partial C_n^n\right\}<\infty\right).$$

The second part now follows from the first part and Lemma 7.3.3. □

Lemma 5.1.3 now follows from Lemmas 7.3.3 and 7.3.4.

7.4 Proof of Lemma 5.1.5

We recall the definitions of Φ_j^n given in (4.2.4). Let $\Xi_k \doteq (\boldsymbol{X}_k, \boldsymbol{Y}_k, \boldsymbol{t}_k)$ denote the vector of (X, Y, t) values associated with the N particles at threshold k. Then by a conditioning argument and the Markov property we have

$$E_{\lambda^n}\left[\Phi_1^n\times\Phi_2^n\times\Phi_3^n\right]=E_{\lambda^n}\left[\Phi_1^n\times E_{\lambda^n}\left[\Phi_2^n\times E_{\lambda^n}\left[\Phi_3^n|\Xi_{\sigma_n}\right]||\Xi_{\gamma_n}\right]\right].$$

We now apply Lemmas 5.1.1, 5.1.2 and 5.1.3 to obtain upper bounds, starting with the innermost conditional expectation and working out. Thus, for example, we need an upper bound on $E_{\lambda^n}\left[\Phi_3^n|\Xi_{\sigma_n}\right]$ where

$$\Phi_3^n\doteq e^{n\alpha[Y^{1,n}(1)-\bar{Y}^{1,n}(t_n^1)]+n\alpha[\bar{Y}^{1,n}(t_n^1)-\bar{Y}^{1,n}(s_n^1)]}e^{\sum_{m=n\sigma_n+1}^{n\tau_n}\log\frac{N_m^S}{N}}\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n}\}}.$$

Note that since $E_{\lambda^n}[\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n}\}}=1|\Xi_{\sigma_n}]\leq 1$, an upper bound on $\frac{1}{n}\log E_{\lambda^n}\left[\Phi_3^n|\Xi_{\sigma_n}\right]$ is given by

$$\frac{1}{n}\log E_{\lambda^n}\left[e^{n\alpha[Y^{1,n}(1)-\bar{Y}^{1,n}(t_n^1)]+n\alpha[\bar{Y}^{1,n}(t_n^1)-\bar{Y}^{1,n}(s_n^1)]}e^{\sum_{m=n\sigma_n+1}^{n\tau_n}\log\frac{N_m^S}{N}}\middle|\,\Xi_{\sigma_n},\mathbf{1}_{\mathcal{A}(\sigma_n,\tau_n)}\mathbf{1}_{\{1\in F^n_{n\tau_n}\}}=1\right].$$

Using the lemmas previously mentioned, given $\delta > 0$, for all sufficiently large n and uniformly in $(\bar{Y}^{1,n}(s_n^1), s_n^1)$, we have the upper bound

$$\sup\left\{\alpha(f_1 - z_3) - (1-t^1)L\left(\frac{f_1 - z_3}{1-t^1}\right) + \alpha(z_3 - \bar{Y}^{1,n}(s_n^1))\right.$$
$$\left. -(t^1 - s_n^1)L\left(\frac{z_3 - \bar{Y}^{1,n}(s_n^1)}{t^1 - s_n^1}\right) + n(\tau_n - \sigma_n)\Delta\frac{H(\alpha_B)}{B} - n(\tau_n - \sigma_n)\left(\Delta\frac{H(\alpha_B)}{B} - C_N^1\right)\right\} + \delta.$$

The first two terms are due to $\alpha[Y^{1,n}(1) - \bar{Y}^{1,n}(t_n^1)]$ and use Lemma 5.1.1; the next three are due to $\alpha[\bar{Y}^{1,n}(t_n^1) - \bar{Y}^{1,n}(s_n^1)]$ and use Lemma 5.1.2; the final term is due to the empirical measure and uses Lemma 5.1.3. This returns a function of $(\bar{Y}^{1,n}(s_n^1), s_n^1)$, which is then used when bounding $E_{\lambda^n}\left[\Phi_2^n \times E_{\lambda^n}\left[\Phi_3^n | \Xi_{\sigma_n}\right] | | \Xi_{\gamma_n}\right]$. Arguments of the same sort are used to get upper bounds on Φ_1^n and Φ_2^n. We then use the convexity of L combine two terms involving L into the one term $(t^1 - g^1)L\left(z_3 - z_1/t^1 - g^1\right)$, canceling off terms of the form $\Delta H(\alpha_B)/B$, and are then left with an upper bound for $\frac{1}{n}\log E_{\lambda^n}\left[\Phi_1^n \times \Phi_2^n \times \Phi_3^n\right]$ for all sufficiently large n given as a supremum of the same form as that appearing in Lemma 5.1.5, save that $\bar{\gamma}$ is replaced by $n\gamma_n$, and C_N is replaced by $C_N^1 \vee C_N^2 + \delta$. The statement of Lemma 5.1.5 now follows by sending $n \to \infty$ and then $\delta \downarrow 0$. It is worth noting that the only term for which the $\Delta H(\alpha_B)/B$ does not cancel, and hence the term where the choice of B has implications, is that which corresponds to Φ_1^n.

7.4.1 Proof of Lemma 5.1.6

Lower bound

Since κ is large enough, the trajectory that starts at $(0,0)$ and uses constant velocity β^* will cross all the thresholds $\partial C_{\bar{\gamma}\Delta}$, $\partial C_{\bar{\sigma}\Delta}$ and $\partial C_{\bar{\tau}\Delta}$ before time $t = 1$. We first

claim that when $B = H(\alpha)$ we have the lower bound $2H(\alpha)$, which is achieved for arbitrary $0 \le \bar{\gamma} \le \bar{\sigma} \le \bar{\tau} \le \kappa$ so long as

$$\frac{f_2 - z_2}{1 - s^2} = \frac{f_1 - z_3}{1 - t^1} = \frac{z_1}{g^1} = \frac{z_2 - z_1}{s^2 - g^1} = \frac{z_3 - z_1}{t^1 - g^1} = \beta^*. \tag{7.4.1}$$

To see this, we use

$$f_2 = f_2 - z_2 + z_2 - z_1 + z_1 = \beta^*(1 - s^2) + \beta^*(s^2 - g^1) + \beta^* g^1 = \beta^*,$$

$$f_1 = f_1 - z_3 + z_3 - z_1 + z_1 = \beta^*(1 - t^1) + \beta^*(t^1 - g^1) + \beta^* g^1 = \beta^*,$$

and that $V_{H(\alpha)}(g_1\beta^*, g_1) = V_{H(\alpha)}(\bar{\gamma}\Delta/\alpha, 0)$ implies $\bar{\gamma}\Delta = g^1 L(\beta^*)$ (see (2.3.3)), so that

$$\begin{aligned}
&\left\{\alpha(f_2 + f_1) - (1 - s^2)L\left(\frac{f_2 - z_2}{1 - s^2}\right) - (1 - t^1)L\left(\frac{f_1 - z_3}{1 - t^1}\right) - \bar{\gamma}\Delta\frac{H(\alpha_B)}{B}\right. \\
&\qquad \left. - g^1 L\left(\frac{z_1}{g^1}\right) - (s^2 - g^1)L\left(\frac{z_2 - z_1}{s^2 - g^1}\right) - (t^1 - g^1)L\left(\frac{z_3 - z_1}{t^1 - g^1}\right)\right\} \\
&= 2\alpha\beta^* - 2L(\beta^*) - \bar{\gamma}\Delta + g^1 L(\beta^*) \\
&= 2H(\alpha).
\end{aligned}$$

When $B < H(\alpha)$, we want to show the lower bound $H(\alpha_B) + H(2\alpha - \alpha_B)$. Let $\tilde{\beta}$ be the convex dual of $2\alpha - \alpha_B$, i.e., $\tilde{\beta} = H'(2\alpha - \alpha_B)$. Note that since $B < H(\alpha)$ we have $\alpha_B < \alpha$ and thus $2\alpha - \alpha_B > \alpha$. It follows that

$$\tilde{\beta} = H'(2\alpha - \alpha_B) > H'(\alpha) = \beta^* > H(\alpha)/\alpha > B/\alpha.$$

We find that lower the bound is achieved only when $t^1 = s^2 = g^1 = 1$, and $f_1 = f_2 = z_1 = \tilde{\beta}$ (the values of z_2 and z_3 do not matter). The value of $\bar{\gamma}$ that allows

$(\tilde{\beta}, 1) \in \partial C_{\bar{\gamma}\Delta}$ can be found from

$$V_B(\tilde{\beta}, 1) = V_B(\bar{\gamma}\Delta/\alpha, 0)$$

which implies

$$\alpha\tilde{\beta} - B = \bar{\gamma}\Delta,$$

and we take $\bar{\gamma} = \bar{\sigma} = \bar{\tau} = (\alpha\tilde{\beta} - B)/\Delta$. Then using $\alpha H(\alpha_B) = B\alpha_B$ we compute

$$\begin{aligned}
&\left\{\alpha(f_2 + f_1) - (1 - s^2)L\left(\frac{f_2 - z_2}{1 - s^2}\right) - (1 - t^1)L\left(\frac{f_1 - z_3}{1 - t^1}\right) - \bar{\gamma}\Delta\frac{H(\alpha_B)}{B}\right. \\
&\qquad \left. -g^1 L\left(\frac{z_1}{g^1}\right) - (s^2 - g^1)L\left(\frac{z_2 - z_1}{s^2 - g^1}\right) - (t^1 - g^1)L\left(\frac{z_3 - z_1}{t^1 - g^1}\right)\right\} \\
&= 2\alpha\tilde{\beta} - L(\tilde{\beta}) - \bar{\gamma}\Delta\frac{H(\alpha_B)}{B} \\
&= 2\alpha\tilde{\beta} - L(\tilde{\beta}) - \alpha_B\tilde{\beta} + H(\alpha_B) \\
&= \tilde{\beta}(2\alpha - \alpha_B) - L(\tilde{\beta}) + H(\alpha_B) \\
&= H(\alpha_B) + H(2\alpha - \alpha_B).
\end{aligned}$$

When $B > H(\alpha)$ we need to use $B < \bar{B}$ [recall its definition (4.3.3)] to ensure that $\bar{\gamma}\Delta = \alpha\tilde{\beta} - B > 0$, and thus $\tilde{\beta}$ is a possible velocity for a trajectory consisting of success particles. The argument is then the same as for the case $B < H(\alpha)$, and we obtain the value $H(\alpha_B) + H(2\alpha - \alpha_B)$ by taking $t^1 = s^2 = g^1 = 1$ and $f_1 = f_2 = z_1 = \tilde{\beta}$.

7.4.1.1 Upper bound

The proof uses a dynamic programming argument. As for the lower bound the quantity to be considered is

$$\left\{\alpha(f_2+f_1)-(1-s^2)L\left(\frac{f_2-z_2}{1-s^2}\right)-(1-t^1)L\left(\frac{f_1-z_3}{1-t^1}\right)-\bar{\gamma}\Delta\frac{H(\alpha_B)}{B}\right.$$
$$\left.-g^1L\left(\frac{z_1}{g^1}\right)-(s^2-g^1)L\left(\frac{z_2-z_1}{s^2-g^1}\right)-(t^1-g^1)L\left(\frac{z_3-z_1}{t^1-g^1}\right)\right\},$$

where the supremum is over all $0\le\bar{\gamma}\le\bar{\sigma}\le\bar{\tau}\le\kappa$ and $f_2,f_1\in\mathbb{R}$, $(z_1,g^1)\in\partial C_{\bar{\gamma}\Delta}$, $(z_2,s^2)\in\partial C_{\bar{\sigma}\Delta}$, and $(z_3,t^1)\in\partial C_{\bar{\tau}\Delta}$.

Stage III. We define the terminal value

$$W_{\mathrm{III}}(f_1,f_2)\doteq\alpha(f_1+f_2).$$

Stage II. We define

$$W_{\mathrm{II}}((z_2,s^2),(z_3,t^1))\doteq\sup\left\{-(1-s^2)L\left(\frac{f_2-z_2}{1-s^2}\right)-(1-t^1)L\left(\frac{f_1-z_3}{1-t^1}\right)+W_{\mathrm{III}}(f_1,f_2)\right\}$$

where the sup is over f_1,f_2 and $(z_2,s^2)\in\partial C_{\bar{\sigma}\Delta}$, and $(z_3,t^1)\in\partial C_{\bar{\tau}\Delta}$. We use the inequality $\alpha\beta\le H(\alpha)+L(\beta)$, with equality if and only if $\beta=\beta^*$. We claim that

$$W_{\mathrm{II}}((z_2,s^2),(z_3,t^1))=-V_{H(\alpha)}(z_2,s^2)-V_{H(\alpha)}(z_3,t^1),$$

where $V_{H(\alpha)}(y,t)=-\alpha y+H(\alpha)[t-1]$ (which is the solution to the minimization problem with running cost L and terminal cost $-\alpha y$). To prove this it is enough to

show

$$\sup\left\{-(1-s^2)L\left(\frac{f_2-z_2}{1-s^2}\right)+\alpha f_2-(1-s^2)H(\alpha)-\alpha z_2\right\}=0$$

where the supremum is over f_2, and also the analogous equality for (z_3,t^1). However this follows by writing the last display as

$$\left\{-(1-s^2)L\left(\frac{f_2-z_2}{1-s^2}\right)-(1-s^2)H(\alpha)+(1-s^2)\alpha\left(\frac{f_2-z_2}{1-s^2}\right)\right\},$$

which is always less than or equal to zero, and equals zero when $(f_2-z_2)/(1-s^2)=\beta^*$.

Stage I. We define

$$W_{\mathrm{I}}((z_1,g^1))\doteq\sup\left\{-(s^2-g^1)L\left(\frac{z_2-z_1}{s^2-g^1}\right)-(t^1-g^1)L\left(\frac{z_3-z_1}{t^1-g^1}\right)+W_{\mathrm{II}}((z_2,s^2),(z_3,t^1))\right\}.$$

We claim that $W_{\mathrm{I}}((z_1,g^1))=-2V_{H(\alpha)}(z_1,g^1)$. Using the expression found for W_{II}, we can break this into two optimizations of the same form as in Stage II. The argument is in fact that same, where now for one case we use that

$$\left\{-(s^2-g^1)L\left(\frac{z_2-z_1}{s^2-g^1}\right)-(s^2-g^1)H(\alpha)+(s^2-g^1)\alpha\left(\frac{z_2-z_1}{s^2-g^1}\right)\right\}=0,$$

with value zero when $(z_2-z_q)/(s^2-g^1)=\beta^*$. Handling the other case in the same way, we find that $W_{\mathrm{I}}((z_1,g^1))=-2V_{H(\alpha)}(z_1,g^1)$.

Stage 0. For the final stage, we define

$$W_0((0,0))\doteq\sup\left\{-g^1L\left(\frac{z_1}{g^1}\right)-\bar{\gamma}\Delta\frac{H(\alpha_B)}{B}+W_{\mathrm{I}}((z_1,g^1))\right\}$$

Using $-2V_{H(\alpha)}(z_1, g^1) = -2\left[-\alpha z_1 + H(\alpha)(g^1 - 1)\right]$, this becomes

$$W_0((0,0)) = \sup\left\{-g^1 L\left(\frac{z_1}{g^1}\right) - \bar{\gamma}\Delta\frac{H(\alpha_B)}{B} + 2\alpha z_1 + 2H(\alpha)(1-g^1)\right\},$$

with the supremum over $(z_1, g^1) \in \partial C_{\bar{\gamma}\Delta}$ and $\bar{\gamma}\Delta = \alpha z_1 - g^1 B \geq 0$. When $B = H(\alpha)$ (and therefore $H(\alpha_B) = B$) the right hand side is

$$\sup g^1\left\{-L\left(\frac{z_1}{g^1}\right) + \alpha\frac{z_1}{g^1} - H(\alpha)\right\} + 2H(\alpha).$$

This equals $2H(\alpha)$ since $\alpha\beta \leq H(\alpha) + L(\beta)$ and we can obtain the equality for any value of g^1 by taking $z^1/g^1 = \beta^*$.

When $B \neq H(\alpha)$ we claim that $W_0((0,0)) = H(\alpha_B) + H(2\alpha - \alpha_B)$. Some simplification using $\bar{\gamma}\Delta = \alpha z_1 - g^1 B$ gives that the supremum is over

$$-g^1 L\left(\frac{z_1}{g^1}\right) - \alpha z_1\frac{H(\alpha_B)}{B} + g^1 H(\alpha_B) + 2\alpha z_1 + 2H(\alpha)(1-g^1).$$

After rearrangement it is enough to show that for $g^1 \in [0,1]$ and $z_1 \in \mathbb{R}$

$$\begin{aligned} &g^1\left\{-L\left(\frac{z_1}{g^1}\right) + \alpha\left(\frac{z_1}{g^1}\right) - H(\alpha)\right\} - H(\alpha_B) - H(2\alpha - \alpha_B) \\ &+ g^1\left\{\alpha\left(\frac{z_1}{g^1}\right)\left[1 - \frac{H(\alpha_B)}{B}\right] - H(\alpha)\left[1 - \frac{H(\alpha_B)}{H(\alpha)}\right]\right\} + 2H(\alpha) \leq 0 \end{aligned}$$

with equality for some choice. Using $H(\alpha_B)/B = \alpha_B/\alpha$ and some more rearrangement this becomes

$$\begin{aligned} &g^1\left\{-L\left(\frac{z_1}{g^1}\right) + \left(\frac{z_1}{g^1}\right)[2\alpha - \alpha_B] - H(2\alpha - \alpha_B)\right\} \\ &+ [1 - g^1]\left\{2H(\alpha) - H(\alpha_B) - H(2\alpha - \alpha_B)\right\} \leq 0. \end{aligned}$$

The inequality is true since by convex duality

$$-L\left(\frac{z_1}{g^1}\right) + \left(\frac{z_1}{g^1}\right)[2\alpha - \alpha_B] - H(2\alpha - \alpha_B) \leq 0$$

with equality when $z_1/g^1 = \tilde{\beta} = H'(2\alpha - \alpha_B)$. (Here we use $B < \bar{B}$, so that $\bar{\gamma}\Delta = \alpha z_1 - g^1 B > 0$.) Also by strict convexity

$$2H(\alpha) - H(\alpha_B) - H(2\alpha - \alpha_B) < 0.$$

Thus the combined quantity is strictly less than zero, with equality only when $g^1 = 1$. When combined with the lower bound this shows that the supremum is only achieved for this case when $g^1 = 1$ and $z_1/g^1 = \tilde{\beta}$. This completes the proof. ■

Bibliography

[1] A. Budhiraja and M. Conroy, *Empirical measure and small noise asymptotics under large deviation scaling for interacting diffusions*, 2020.

[2] A. Budhiraja and P. Dupuis, *Representations for functional of hilbert space valued diffusions*, (2001).

[3] ——, *A variational representation for positive functionals of infinite dimensional brownian motion*, Probab. Math. Statist., 20 (2001).

[4] ——, *Analysis and approximation of rare events*, Representations and Weak Convergence Methods. Series Prob. Theory and Stoch. Modelling, 94 (2019).

[5] Y. Cai and P. Dupuis, *Analysis of an interacting particle method for rare event estimation*, Queueing Systems, 73 (2013).

[6] T. Dean and P. Dupuis, *The design and analysis of a generalized restart/dpr algorithm for rare event simulation*, Annals OR, 189 (2011), pp. 63–102.

[7] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems With Applications*, vol. 100, 05 2004.

[8] P. Del Moral and P. Lezaud, *Branching and interacting particle interpretations of rare event probabilities*, in Stochastic Hybrid Systems: Theory and Safety Critical Applications, Springer, 2006, pp. 277–323.

[9] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, John Wiley & Sons, 2011.

[10] P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and P. Plechac, *Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics*, 2015.

[11] P. Dupuis and Y. Liu, *On the large deviation rate function for the empirical measures of reversible jump markov processes*, The Annals of Probability, 43 (2015).

[12] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stochastics: An International Journal of Probability and Stochastic Processes, 76 (2004), pp. 481–508.

[13] ——, *Dynamic importance sampling for uniformly recurrent markov chains*, The Annals of Applied Probability, 15 (2005).

[14] P. Dupuis and H. Wang, *Subsolutions of an isaacs equation and efficient schemes for importance sampling*, Math. Oper. Res., 32 (2005), pp. 723–757.

[15] R. S. Ellis, *Large deviations for a general class of random vectors*, The Annals of Probability, 12 (1984), pp. 1–12.

[16] M. Freidlin, J. Szucs, and A. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren der mathematischen Wissenschaften, Springer New York, 2012.

[17] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic, *Splitting for rare event simulation: analysis of simple cases*, in Proceedings of the 28th Conference on Winter Simulation, WSC '96, USA, 1996, IEEE Computer Society, p. 302–308.

[18] ——, *A large deviations perspective on the efficiency of multilevel splitting*, IEEE Transactions on Automatic Control, 43 (1998), pp. 1666–1679.

[19] T. E. Harris, *The theory of branching processes*, Die Grundlehren der Mathematischen Wissenschaften, Bd. 119, Springer-Verlag, Berlin, 1963.

[20] P.-L. Lions and A. S. Sznitman, *Stochastic differential equations with reflecting boundary conditions*, Communications on Pure and Applied Mathematics, 37 (1984), pp. 511–537.

[21] H. Pham, *Some applications and methods of large deviations in finance and insurance*, 2007.

[22] F. Ragone, J. Wouters, and F. Bouchet, *Computation of extreme heat waves in climate models using a large deviation algorithm*, Proceedings of the National Academy of Sciences, 115 (2017), p. 24–29.

[23] M. Rousset, *On the control of an interacting particle estimation of schrödinger ground states*, SIAM J. Math. Anal., 38 (2006), pp. 824–844.

[24] M. Villén-Altamirano and J. Villen-Altamirano, *Analysis of restart simulation: Theoretical basis and sensitivity study*, European Transactions on Telecommunications, 13 (2002), pp. 373–385.

[25] M. Villén-Altamirano, J. Villén-Altamirano, et al., *Restart: a method for accelerating rare event simulations*, Queueing, Performance and Control in ATM (ITC-13), (1991), pp. 71–76.