On the large deviation rate function for the empirical measures of reversible jump Markov processes*

Paul Dupuis† and Yufei Liu‡
Division of Applied Mathematics
Brown University
Providence, RI 02912

September 12, 2013

Abstract

The large deviations principle for the empirical measure for both continuous and discrete time Markov processes is well known. Various expressions are available for the rate function, but these expressions are usually as the solution to a variational problem, and in this sense not explicit. An interesting class of continuous time, reversible processes was identified in the original work of Donsker and Varadhan for which an explicit expression is possible. While this class includes many (reversible) processes of interest, it excludes the case of continuous time pure jump processes, such as a reversible finite state Markov chain. In this paper we study the large deviations principle for the empirical measure of pure jump Markov processes and provide an explicit formula of the rate function under reversibility.

1 Introduction

Let $X(t)$ be a time homogeneous Markov process with Polish state space S, and let $P(t,x,dy)$ be the transition function of $X(t)$. For $t \in [0,\infty)$, define T_t by

$$T_t f(x) \doteq \int_S f(y)\, P(t,x,dy).$$

Then T_t is a contraction semigroup on the Banach space of bounded, Borel measurable functions on S [6, Chapter 4.1]. We use $\mathcal{L}$ to denote the infinitesimal generator of T_t and $\mathcal{D}$ the domain of $\mathcal{L}$ (see [6, Chapter 1]). Hence for each bounded measurable function $f \in \mathcal{D}$,

$$\mathcal{L}f(x) = \lim_{t\downarrow 0} \frac{1}{t}\left[\int_S f(y)\, P(t,x,dy) - f(x)\right].$$

*The authors thank a referee for useful comments and a correction.

†Research supported in part by the Department of Energy (DE-SCOO02413), the National Science Foundation (DMS-1008331), and the Army Research Office (W911NF-09-1-0155,W911NF-12-1-0222).

‡Research supported in part by the Department of Energy (DE-SCOO02413).

The empirical measure (or normalized occupation measure) up to time T of the Markov process $X(t)$ is defined by

$$\eta_T(\cdot) \doteq \frac{1}{T}\int_0^T \delta_{X(t)}(\cdot)\,dt. \tag{1.1}$$

Let $\mathcal{P}(S)$ be the metric space of probability measures on S equipped with the Levy-Prohorov metric, which is compatible with the topology of weak convergence. For $\eta \in \mathcal{P}(S)$, define

$$I(\eta) \doteq -\inf_{\substack{u\in\mathcal{D}\\ u>0}} \int_S \frac{\mathcal{L}u}{u} d\eta. \tag{1.2}$$

It is easy to check that I thus defined is lower semicontinuous under the topology of weak convergence. Consider the following regularity assumption.

Condition 1.1 *There exists a probability measure λ on S such that for $t > 0$ the transition functions $P(t,x,dy)$ have densities with respect to λ, i.e.,*

$$P(t,x,dy) = p(t,x,y)\,\lambda(dy). \tag{1.3}$$

Under additional recurrence and transitivity conditions, Donsker and Varadhan [2, 3] prove the following. For any open set $O \subset \mathcal{P}(S)$

$$\liminf_{T\to\infty} \frac{1}{T}\log P(\eta_T(\cdot)\in O) \geq -\inf_{\eta\in O} I(\eta), \tag{1.4}$$

and for any closed set $C \subset \mathcal{P}(S)$

$$\limsup_{T\to\infty} \frac{1}{T}\log P(\eta_T(\cdot)\in C) \leq -\inf_{\eta\in C} I(\eta). \tag{1.5}$$

We refer to (1.4) as the large deviation lower bound and (1.5) as the large deviation upper bound. Under ergodicity, the empirical measure η_T converges to the invariant distribution of the Markov process $X(t)$. The large deviation principle characterizes this convergence through the associated rate function. While there are many situations where an explicit formula for (1.2) would be useful, it is in general difficult to solve the variational problem. The main existing results on this issue are for the self-adjoint case in the continuous time setting, see [2, 9, 11]. Specifically, suppose there is a σ-finite measure φ on S, and that the densities in (1.3) satisfy the following reversibility condition:

$$p(t,x,y) = p(t,y,x) \text{ almost everywhere } (\varphi\times\varphi). \tag{1.6}$$

Then T_t is self-adjoint. If we denote the closure of $\mathcal{L}$ by $\bar{\mathcal{L}}$ (see, e.g., [6, p. 16]) and the domain of $\bar{\mathcal{L}}$ by $\mathcal{D}(\bar{\mathcal{L}})$, then $\bar{\mathcal{L}}$ is self-adjoint and negative semidefinite (since T_t is a contraction). We denote by $(-\bar{\mathcal{L}})^{1/2}$ the canonical positive semidefinite square root of $-\bar{\mathcal{L}}$ [10, Chapter 12]. Let $\bar{\mathcal{D}}_{1/2}$ be the domain of $(-\bar{\mathcal{L}})^{1/2}$. Donsker and Varadhan [2, Theorem 5] show under certain conditions that I defined by (1.2) has

the following properties: $I(\mu) < \infty$ if and only if $\mu \ll \varphi$ and $(d\mu/d\varphi)^{1/2} \in \bar{\mathcal{D}}_{1/2}$, and with $f \doteq d\mu/d\varphi$ and $g \doteq f^{1/2}$,

$$I(\mu) = \left\| (-\bar{\mathcal{L}})^{1/2} g \right\|^2, \tag{1.7}$$

where $\|\cdot\|$ denotes the L^2 norm with respect to φ. Typically, φ is taken to be the invariant distribution of the process.

It should be noted that this explicit formula does not apply to one of the simplest Markov processes, namely, continuous time Markov jump processes with bounded infinitesimal generators. Let $\mathcal{B}(S)$ be the Borel σ-algebra on S and let $\alpha(x, \Gamma)$ be a transition kernel on $S \times \mathcal{B}(S)$. Let $B(S)$ denote the space of bounded Borel measurable functions on S and let $q \in B(S)$ be nonnegative. Then

$$\mathcal{L} f(x) \doteq q(x) \int_S (f(y) - f(x))\, \alpha(x, dy) \tag{1.8}$$

defines a bounded linear operator on $B(S)$ and $\mathcal{L}$ is the generator of a Markov process that can be constructed as follows. Let $\{X_n, n \in \mathbb{N}\}$ be a Markov chain in S with transition probability $\alpha(x, \Gamma)$, i.e.

$$P(X_{n+1} \in \Gamma | X_0, X_1, \ldots, X_n) = \alpha(X_n, \Gamma) \tag{1.9}$$

for all $\Gamma \in \mathcal{B}(S)$ and $n \in \mathbb{N}$. Let $\tau_1, \tau_2, \ldots$ be independent and exponentially distributed with mean 1, and independent of $\{X_n, n \in \mathbb{N}\}$. Define a sojourn time s_i for each $i = 1, 2, \ldots$ by

$$q(X_{i-1})\, s_i = \tau_i. \tag{1.10}$$

Then

$$X(t) = X_n \text{ for } \sum_{i=1}^{n} s_i \le t < \sum_{i=1}^{n+1} s_i$$

(with the convention $\sum_{i=1}^{0} s_i = 0$) defines a Markov process $\{X(t), t \in [0, \infty)\}$ with infinitesimal generator $\mathcal{L}$, and we call this process a Markov jump process.

A very simple special case is as follows. Using the notation above, assume $S = [0, 1]$, $q \equiv 1$ and for each $x \in [0, 1]$, $\alpha(x, \cdot)$ is the uniform distribution on $[0, 1]$. The infinitesimal generator $\mathcal{L}$ defined in (1.8) reduces to

$$\mathcal{L} f(x) = \int_0^1 f(y)\, dy - f(x),$$

which is clearly self-adjoint with respect to Lebesgue measure. If C is the collection of all Dirac measures on S, then C is closed under the topology of weak convergence on $\mathcal{P}(S)$. Hence a large deviation upper bound would imply

$$\limsup_{T \to \infty} \frac{1}{T} \log P(\eta_T \in C) \le - \inf_{\mu \in C} I(\mu). \tag{1.11}$$

However, the probability that the very first exponential holding time is bigger than T is exactly $\exp\{-T\}$, and when this happens, the empirical measure is a Dirac measure located at some point that is uniformly distributed on $[0,1]$. Hence

$$\liminf_{T\to\infty}\frac{1}{T}\log P\left(\eta_T(\cdot)\in C\right)\geq\liminf_{T\to\infty}\frac{1}{T}\log P\left(\tau_1>T\right)=-1.$$

In fact, we will prove later that the rate function for the empirical measure of this Markov jump process never exceeds 1. However, if the upper bound held with the function defined in (1.7), one would have $I(\delta_a)=\infty$ for $a\in[0,1]$, and by (1.11)

$$\limsup_{T\to\infty}\frac{1}{T}\log P\left(\eta_T(\cdot)\in D\right)=-\infty,$$

which is impossible.

This example shows that this type of Markov jump process is not covered by [2, 3]. In fact, the transition function $P(t,x,dy)$ takes the form

$$P(t,x,dy)=e^{-t}\delta_x(dy)+\left(1-e^{-t}\right)1_{[0,1]}(y)\,dy,$$

which means that we cannot find a reference probability measure λ on S such that $P(t,x,\cdot)$ has a density with respect to $\lambda(\cdot)$ for almost all $x\in S$ and $t>0$, which is a violation to Condition 1.1 used in [2, 3], and also violates the form of reversibility needed for (1.7).

A condition such as Condition 1.1 holds naturally for Markov processes that possess a "diffusive" term in the dynamics, which is not the case for Markov jump processes, and the form of the rate function given in (1.7) will not be valid for these processes either. The purpose of the current paper is to establish a large deviation principle for the empirical measures of reversible Markov jump processes, and to provide an explicit formula for the rate function like the one given in (1.7). We also show why the boundedness of the rate function results from the fact that tilting of the exponential holding times with bounded relative entropy cost can be used for target measures that are not absolutely continuous with respect to the invariant distribution.

Finally we mention that [1] evaluates (1.2) for certain classes of measures when $\mathcal{L}$ is the generator of a jump Markov process satisfying various conditions. However, it does not present an expression for an arbitrary measure, and indeed in appears that the authors are unaware that (1.7) is not the correct rate function for such processes, or that the large deviation principle had not been established.

The paper is organized as follows. Section 2 presents our assumptions on the process. In Section 3 we state the main result, Theorem 3.1. The proof of Theorem 3.1 is divided into two sections, Section 4 for the upper bound and Section 5 for the lower bound. In the final section, we discuss the special feature of Markov jump processes that leads to the boundedness of the rate function.

2 Assumptions

Our first assumption is that the Polish state space S is compact. While compactness is not needed, it lets us focus on the novel features of the problem. For standard techniques to deal with the non-compact case see, e.g., [3].

A construction of Markov jump processes was given in the Introduction, and we continue to use the notation introduced there. The jump intensity q in (1.8) is assumed to be continuous on S, and there exist $0 < K_1 \leq K_2 < \infty$ such that

$$K_1 \leq q(x) \leq K_2. \tag{2.1}$$

Reversibility seems necessary to obtain an explicit formula for the rate function, and we will make such an assumption. Recall that $\mathcal{D}$ is the domain of $\mathcal{L}$.

Condition 2.1 *$\mathcal{L}$ is self-adjoint (or reversible) under π in the following sense: for any $f, g \in \mathcal{D}$*

$$\int_S (\mathcal{L}f(x))\, g(x)\, \pi(dx) = \int_S (\mathcal{L}g(x))\, f(x)\, \pi(dx). \tag{2.2}$$

An equivalent condition for (2.2) to hold is the "detailed balance" condition, i.e., for π-a.e. $x, y \in S$

$$q(x)\, \alpha(x, dy)\, \pi(dx) = q(y)\, \alpha(y, dx)\, \pi(dy). \tag{2.3}$$

Note that (2.3) directly implies $\int_S (\mathcal{L}f(x))\, \pi(dx) = 0$ for all $f \in \mathcal{D}$.

To ensure ergodicity of $X(t)$, we need several conditions on the transition function α in (1.9). Recall that $\mathcal{P}(S)$ is the metric space of probability measures on S equipped with Levy-Prohorov metric, which is compatible with the topology of weak convergence.

Condition 2.2 *α satisfies the Feller property. That is, $\alpha(x, \cdot) : S \longmapsto \mathcal{P}(S)$ is continuous in x.*

Remark 2.3 *The Feller property and the compactness of S guarantee α has an invariant distribution [4, Proposition 8.3.4], which we denote by $\widetilde{\pi}$. The boundedness of q enables us to define a probability measure π according to*

$$\pi(A) \doteq \frac{\int_A \frac{1}{q(x)} \widetilde{\pi}(dx)}{\int_S \frac{1}{q(x)} \widetilde{\pi}(dx)}. \tag{2.4}$$

Since $\widetilde{\pi}$ is invariant under α, i.e., $\widetilde{\pi}(\cdot) = \int_S \alpha(x, \cdot)\, \widetilde{\pi}(dx)$, we have

$$\int_S (\mathcal{L}f(x))\, \pi(dx) = \frac{1}{\int_S \frac{1}{q(x)} \widetilde{\pi}(dx)} \int_S \int_S [f(y) - f(x)]\, \alpha(x, dy)\, \widetilde{\pi}(dx) = 0.$$

By Echeverria's Theorem [6, Theorem 4.9.17], π is an invariant distribution of $X(t)$.

Condition 2.4 *α satisfies the following transitivity condition. There exist positive integers l_0 and n_0 such that for all x and ζ in S*

$$\sum_{i=l_0}^{\infty} \frac{1}{2^i} \alpha^{(i)}(x, dy) \ll \sum_{j=n_0}^{\infty} \frac{1}{2^j} \alpha^{(j)}(\zeta, dy),$$

where $\alpha^{(k)}$ denotes the k-step transition probability.

Remark 2.5 *Under this condition, $\widetilde{\pi}$ is the unique invariant distribution of α [4, Lemma 8.6.2]. Thus π defined by (2.4) is the unique probability distribution that satisfies $\int_S (\mathcal{L}f(x))\, \pi(dx) = 0$, and hence by [6, Theorem 4.9.17] is the unique invariant distribution of $X(t)$.*

Condition 2.6 *There exists an integer N and a positive real number c such that*

$$\alpha^{(N)}(x, \cdot) \leq c\widetilde{\pi}(\cdot)$$

for all $x \in S$.

Remark 2.7 *This type of assumption is common in the large deviation analysis of empirical measures. See e.g., [5, Hypothesis 1.1].*

Condition 2.8 *The support of π is S.*

Remark 2.9 *This condition guarantees that any probability measure $\eta \in \mathcal{P}(S)$ can be approximated by measures that are absolutely continuous with respect to π. Indeed, given $\delta > 0$ let $\{x_j, N_j, j = 1, \ldots, J\}$ be such that $J < \infty$, $x_j \in N_j \in \mathcal{B}(S)$, the N_j are disjoint, $\cup_{j=1}^{J} N_j = S$, $\pi(N_j) > 0$ and $\sup\{d(x_i, y) : y \in N_j\} \leq \delta$ for $j = 1, \ldots, J$ (this can be done by an open covering argument). Given any $\eta \in \mathcal{P}(S)$ and $A \in \mathcal{B}(S)$ let*

$$\bar{\eta}^{\delta}(A) = \sum_{j=1}^{J} \frac{\pi(A \cap N_j)}{\pi(N_j)} \eta(N_j).$$

Then $\bar{\eta}^{\delta}$ is absolutely continuous with respect to π. Since $\bar{\eta}^{\delta}(N_j) = \eta(N_j)$ and $\sup\{d(x_i, y) : y \in N_j\} \leq \delta$, $\bar{\eta}^{\delta} \to \eta$ in the weak topology as $\delta \to 0$.

Remark 2.10 *Condition 2.8 excludes the existence of transient states. Although one can obtain an LDP for $X(t)$ that has transient states, one would end up with a rate function that depends on the initial state.*

3 A large deviation principle

3.1 Definition of the rate function

In this subsection, we define the rate function I. In later sections we prove that I thus defined is the correct form of the large deviation rate function for the empirical

measures of the Markov jump processes. All conditions stated in Section 2 will be assumed throughout the rest of the paper. We wish to study the large deviation principle for the empirical measures $\eta_T \in \mathcal{P}(S)$ defined by (1.1). Under compactness of S and Condition 2.2, η_T converges in distribution to an invariant distribution of $\mathcal{L}$. As pointed out in Remark 2.5, π is the unique invariant distribution of $\mathcal{L}$, and thus η_T converges in distribution to π. Let H be the collection of all distributions that are absolutely continuous with respect to π, i.e.

$$H \doteq \{\eta \in \mathcal{P}(S) : \eta \ll \pi\}. \tag{3.1}$$

For $\eta \in H$, and assuming that the integral is well defined, consider

$$-\int_S \theta^{1/2}(x)\,\mathcal{L}\left(\theta^{1/2}(x)\right)\pi(dx),$$

where $\theta = d\eta/d\pi$. This is a rewriting of $||(-\bar{\mathcal{L}})^{1/2} g||^2$ in (1.7). By inserting the form of $\mathcal{L}$ from (1.8), we obtain the candidate rate function

$$I(\eta) = \int_S q(x)\,\eta(dx) - \int_{S\times S} \theta^{1/2}(x)\,\theta^{1/2}(y)\,q(x)\,\alpha(x,dy)\,\pi(dx). \tag{3.2}$$

Note that by applying (2.3) and using the Cauchy-Schwartz inequality, one can prove that I defined by (3.2) is nonnegative. Recall that K_2 is the upper bound of q as in (2.1), and thus I is bounded above by K_2. In addition, it is straightforward to show that I is convex on H.

We want to extend the definition of I to all measures in $\mathcal{P}(S)$. As pointed out in Remark 2.9, H is dense in $\mathcal{P}(S)$ under the topology of weak convergence. Hence we can extend the definition of I to all of $\mathcal{P}(S)$ via lower semicontinuous regularization with respect to the topology of weak convergence. Thus if $\eta_n \to \eta$ weakly and $\{\eta_n\} \in H$, $\liminf_{n\to\infty} I(\eta_n) \geq I(\eta)$, and equality holds for at least one such sequence. This extension guarantees that the extended I is convex, lower semicontinuous and bounded above by K_2 on all of $\mathcal{P}(S)$. The compactness of S and the lower semicontinuity of I ensure that I has compact level sets. Being a nonnegative, lower semicontinuous function with compact level sets, I indeed is a valid large deviation rate function.

We have finished the definition of the rate function I, and are now ready to state the large deviation principle.

3.2 A large deviation principle

Our main result is the following:

Theorem 3.1 *Let $X(t)$ be a Markov jump process satisfying all the assumptions in Section 2. Let I be defined as in Section 3.1. Then the large deviation bounds (1.4) and (1.5) hold.*

To prove Theorem 3.1, it suffices to show the equivalent Laplace principle [4, Theorem 1.2.3]. Specifically, we establish that for any bounded continuous function $F : \mathcal{P}(S) \to \mathbb{R}$

$$\lim_{T\to\infty} -\frac{1}{T} \log E\left[\exp\left\{-TF\left(\eta_T\right)\right\}\right] = \inf_{\eta\in\mathcal{P}(S)} \left[F\left(\eta\right) + I\left(\eta\right)\right]. \tag{3.3}$$

By adding a constant to both sides of (3.3) we can assume $F \geq 0$ and do that for the rest of the paper. The proof is based on a weak convergence approach and is split into two parts: a Laplace upper bound and a Laplace lower bound.

Relative entropy plays a key role in the proof, and we hence state the definition and a few important properties. Details can be found in [4].

Definition 1 *Let $(\mathcal{V}, \mathcal{A})$ be a measurable space. For $\vartheta \in \mathcal{P}(\mathcal{V})$, the* ***relative entropy*** *$R(\cdot \|\vartheta)$ is a mapping from $\mathcal{P}(\mathcal{V})$ into the extended real numbers. It is defined by*

$$R\left(\gamma \|\vartheta\right) \doteq \int_{\mathcal{V}} \left(\log \frac{d\gamma}{d\vartheta}\right) d\gamma$$

when $\gamma \in \mathcal{P}(\mathcal{V})$ is absolutely continuous with respect to ϑ and $\log d\gamma/d\vartheta$ is integrable with respect to γ. Otherwise we set $R(\gamma \|\vartheta) \doteq \infty$.

If $\mathcal{V}$ is a Polish space and $\mathcal{A}$ the associated σ-algebra, then $R(\cdot \|\cdot)$ is nonnegative, jointly convex and jointly lower semicontinuous (with respect to the weak topology on $\mathcal{P}(\mathcal{V})^2$). We state the following two properties of relative entropy.

Lemma 3.2 *(Variational formula) Let $(\mathcal{V}, \mathcal{A})$ be a measurable space, k a bounded measurable function mapping $\mathcal{V}$ into $\mathbb{R}$, and ϑ a probability measure on $\mathcal{V}$. The following conclusions hold.*

(a) We have the variational formula

$$-\log \int_{\mathcal{V}} e^{-k} d\vartheta = \inf_{\gamma\in\mathcal{P}(\mathcal{V})} \left\{R\left(\gamma \|\vartheta\right) + \int_{\mathcal{V}} k d\gamma\right\}. \tag{3.4}$$

(b) The infimum in (3.4) is attained uniquely at γ_0 defined by

$$\frac{d\gamma_0}{d\vartheta}\left(x\right) \doteq e^{-k(x)} / \int_{\mathcal{V}} e^{-k} d\vartheta.$$

Theorem 3.3 *(Chain rule) Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and β and γ probability measures on $\mathcal{X} \times \mathcal{Y}$. We denote by $[\beta]_1$ and $[\gamma]_1$ the first marginals of β and γ and by $\beta(dy|x)$ and $\gamma(dy|x)$ the stochastic kernels on $\mathcal{Y}$ given $\mathcal{X}$ for which we have the decompositions*

$$\beta\left(dx \times dy\right) = [\beta]_1\left(dx\right) \otimes \beta\left(dy|dx\right) \text{ and } \gamma\left(dx \times dy\right) = [\gamma]_1\left(dx\right) \otimes \gamma\left(dy|dx\right).$$

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

$$R\left(\beta \|\gamma\right) = R\left([\beta]_1 \|[\gamma]_1\right) + \int_{\mathcal{X}} R\left(\beta\left(\cdot|x\right) \|\gamma\left(\cdot|x\right)\right) [\beta]_1\left(dx\right).$$

We devote the next two sections to proving the Laplace upper bound and the Laplace lower bound, respectively.

4 Proof of the Laplace upper bound

In this section, we prove the Laplace upper bound part of (3.3), i.e.

$$\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T\right)\right\}\right] \geq \inf_{\eta\in\mathcal{P}(S)}\left[F\left(\eta\right)+I\left(\eta\right)\right]. \tag{4.1}$$

Recalling the construction of $X(t)$ in the Introduction, we define a random integer R_T as the index when the total "waiting time" first exceeds T, i.e.

$$\sum_{i=1}^{R_T-1} s_i \leq T < \sum_{i=1}^{R_T} s_i. \tag{4.2}$$

Then the empirical measure η_T can be written as

$$\begin{aligned}
\eta_T\left(\cdot\right) &= \frac{1}{T}\int_0^T \delta_{X(t)}\left(\cdot\right)dt \\
&= \frac{1}{T}\left[\sum_{i=1}^{R_T-1}\delta_{X_{i-1}}\left(\cdot\right)s_i + \delta_{X_{R_T-1}}\left(\cdot\right)\left(T-\sum_{i=1}^{R_T-1}s_i\right)\right] \\
&= \frac{1}{T}\left[\sum_{i=1}^{R_T-1}\delta_{X_{i-1}}\left(\cdot\right)\frac{\tau_i}{q\left(X_{i-1}\right)} + \delta_{X_{R_T-1}}\left(\cdot\right)\left(T-\sum_{i=1}^{R_T-1}\frac{\tau_i}{q\left(X_{i-1}\right)}\right)\right].
\end{aligned} \tag{4.3}$$

The proof of (4.1) will be partitioned into two cases: $R_T/T > C$ and $0 \leq R_T/T \leq C$, where C will be sent to ∞ after sending $T\to\infty$.

4.1 The case $R_T/T > C$

Let $F:\mathcal{P}(S)\to\mathbb{R}$ be nonnegative and continuous. Then since $F\geq 0$,

$$\begin{aligned}
-\frac{1}{T}\log E\left[1_{\{(C,\infty)\}}(R_T/T)e^{-TF(\eta_T)}\right] &\geq -\frac{1}{T}\log P\left\{\sum_{i=1}^{\lfloor TC\rfloor+1} s_i \leq T\right\} \\
&= -\frac{1}{T}\log P\left\{\sum_{i=1}^{\lfloor TC\rfloor+1}\frac{\tau_i}{q\left(X_{i-1}\right)} \leq T\right\} \\
&\geq -\frac{1}{T}\log P\left\{\sum_{i=1}^{\lfloor TC\rfloor+1}\tau_i \leq K_2T\right\}.
\end{aligned}$$

Using Chebyshev's inequality, for any $\alpha\in(0,\infty)$

$$\begin{aligned}
P\left\{\sum_{i=1}^{\lfloor TC\rfloor+1}\tau_i \leq K_2T\right\} &= P\left\{e^{-\alpha\sum_{i=1}^{\lfloor TC\rfloor+1}\tau_i} \geq e^{-\alpha K_2T}\right\} \\
&\leq e^{\alpha K_2T}E\left[e^{-\alpha\sum_{i=1}^{\lfloor TC\rfloor+1}\tau_i}\right] \\
&= e^{\alpha K_2T}e^{(\lfloor TC\rfloor+1)\log\frac{1}{1+\alpha}}.
\end{aligned}$$

For the last equality we have used that if τ is exponentially distributed with mean 1 then $Ee^{a\tau} = 1/(1-a)$ for any $a \in (-\infty, 1)$. Combining the last two inequalities,

$$\begin{aligned}\liminf_{T\to\infty} &-\frac{1}{T}\log E\left[1_{\{(C,\infty)\}}(R_T/T)\cdot\exp\{-TF(\eta_T)\}\right]\\ &\geq \sup_{\alpha\in(0,\infty)}\left[-K_2\alpha + C\log(1+\alpha)\right]\\ &= -C + C\log C + K_2 - C\log K_2.\end{aligned}$$

Note that $-C + C\log C + K_2 - C\log K_2 \to \infty$ as $C \to \infty$.

4.2 The case $0 \leq R_T/T \leq C$

4.2.1 A stochastic control representation

In this case we adapt a standard weak convergence argument, see [4] for details. Specifically, we first establish a stochastic control representation for the left hand side of (3.3) and then obtain a lower bound for the limit as $T \to \infty$. In the representation, all distributions can be perturbed from their original form, but such a perturbation pays a relative entropy cost. We distinguish the new distributions and random variables by an overbar. In the following, the barred quantities are constructed analogously to their unbarred counterparts. Hence $\bar{\tau}_i$ and $\bar{X}_i$ are chosen recursively according to stochastic kernels $\bar{\sigma}_i(\cdot)$ and $\bar{\alpha}_i(\cdot)$, i.e., $\bar{\sigma}_i(\cdot)$ and $\bar{\alpha}_i(\cdot)$ are conditional distributions that can depend on the whole past. Specifically, $\bar{\sigma}_i(\cdot)$ depends on $\left\{\bar{X}_0, \bar{\tau}_1, \bar{X}_1, \bar{\tau}_2, \ldots, \bar{X}_{i-1}\right\}$ and $\bar{\alpha}_i(\cdot)$ depends on $\left\{\bar{X}_0, \bar{\tau}_1, \bar{X}_1, \bar{\tau}_2 \ldots, \bar{X}_{i-1}, \bar{\tau}_i\right\}$; $\bar{s}_i$ is defined by (1.10) using $\bar{X}_i$ and $\bar{\tau}_i$; $\bar{R}_T$ is defined by (4.2) using $\bar{s}_i$; and $\bar{\eta}_T$ is defined by (4.3) using $\bar{X}_i$, $\bar{\tau}_i$ and $\bar{R}_T$. It will be sufficient to consider any deterministic sequence $\{r_T\}$ such that $0 \leq r_T/T \leq C$, and $r_T/T \to A$ for some $A \in [0, C]$ as $T \to \infty$. We restrict consideration to controlled processes such that $\bar{R}_T = r_T$ by placing an infinite cost penalty on controls which lead to any other outcome with positive probability. Let $\mathbf{1}(A)$ denote the indicator function of a set A, and recall that our convention is $0 \cdot \infty = 0$. By applying [4, Proposition 4.5.1] and Theorem 3.3 the following is valid:

$$-\frac{1}{T}\log E\left[\exp\left\{-TF(\eta_T) - T\cdot\infty\cdot\mathbf{1}\left(r_T \neq R_T\right)\right\}\right] \tag{4.4}$$

$$\begin{aligned}&= -\frac{1}{T}\log E\left[\exp\left\{-TF(\eta_T) - T\cdot\infty\cdot\mathbf{1}\left(\left\{\sum_{i=1}^{r_T-1} s_i \leq T < \sum_{i=1}^{r_T} s_i\right\}^c\right)\right\}\right]\\ &= \inf E\left[F(\bar{\eta}_T) + \infty\cdot\mathbf{1}\left(\left\{\sum_{i=1}^{r_T-1}\bar{s}_i \leq T < \sum_{i=1}^{r_T}\bar{s}_i\right\}^c\right) + \frac{1}{T}\sum_{i=1}^{r_T}\left[R\left(\bar{\alpha}_{i-1}\,\|\alpha\right) + R\left(\bar{\sigma}_i\,\|\sigma\right)\right]\right]\end{aligned} \tag{4.5}$$

where the infimum is taken over all control measures $\{\bar{\alpha}_i, \bar{\sigma}_i\}$. Since in Section 5 we will prove a similar but more involved representation formula, Lemma 5.1, we omit the proof of this representation. Due to the restriction $\bar{R}_T = r_T$, one can write $\bar{\eta}_T$ as

$$\bar{\eta}_T(\cdot) = \frac{1}{T}\left[\sum_{i=1}^{r_T-1} \delta_{\bar{X}_{i-1}}(\cdot)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} + \delta_{\bar{X}_{r_T-1}}(\cdot)\left(T - \sum_{i=1}^{r_T-1}\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right]. \tag{4.6}$$

In the following proof, we repeatedly extract further subsequences of T. To keep the notation concise, we abuse notation and use T to denote all subsequences. Note also that in proving a lower bound for (4.4) it suffices to consider a subsequence of T such that

$$\sup_T -\frac{1}{T}\log E\left[\exp\left\{-TF(\eta_T) - T\cdot\infty\cdot\mathbf{1}\left(r_T \neq R_T\right)\right\}\right] < \infty. \tag{4.7}$$

We assume this condition for the rest of this subsection.

The relative entropy cost in (4.5) includes two parts, $RE_T^1 \doteq \frac{1}{T}\sum_{i=1}^{r_T} R\left(\bar{\alpha}_{i-1}\|\alpha\right)$ and $RE_T^2 \doteq \frac{1}{T}\sum_{i=1}^{r_T} R\left(\bar{\sigma}_i\|\sigma\right)$. We will prove that for any sequence of controls $\{\bar{\alpha}_i, \bar{\sigma}_i\}$ in (4.5)

$$\liminf_{T\to\infty} E\left[F\left(\bar{\eta}_T\right) + RE_T^1 + RE_T^2\right] \geq \inf_{\eta\in\mathcal{P}(S)}\left[F(\eta) + I(\eta)\right]. \tag{4.8}$$

Toward this end, it is enough to show that along any subsequence of T such that $r_T/T \to A$, we can extract a further subsequence along which (4.8) holds. In addition, it suffices to consider only functions F that besides being nonnegative, are also lower semicontinuous and convex. This restriction is valid since I is convex and lower semicontinuous, and follows a standard argument in the large deviation literature. The interested reader can find the details in [8].

In light of (4.5) and (4.7) we assume without loss of generality

$$\sup_T E\left[F\left(\bar{\eta}_T\right) + RE_T^1 + RE_T^2\right] < \infty. \tag{4.9}$$

Since the proof of (4.8) is lengthy, we analyze each term on the left hand side of (4.8) separately in the following subsections.

4.2.2 The term RE_T^1

The cost RE_T^1 comes from distorting the dynamics of the embedded Markov chain, and indeed the analysis gives a very similar conclusion to that of an ordinary Markov chain ([4, Chapter 8]). For any probability measure ν on $S\times S$ we will use notations $[\nu]_1$ and $[\nu]_2$ to denote the first and second marginals of ν. We have the following result for RE_T^1.

Lemma 4.1 *Consider any sequence of controls $\{\bar{\alpha}_i, \bar{\sigma}_i\}$ in (4.5) such that (4.9) holds. Along any subsequence of T satisfying $r_T/T \to A$, define a sequence of random probability measures on $S\times S$ via*

$$\mu_T(dx, dy) \doteq \frac{1}{r_T}\sum_{i=1}^{r_T}\delta_{\bar{X}_{i-1}}(dx)\,\bar{\alpha}_{i-1}(dy).$$

Then one can extract a further subsequence such that $E\mu_T$ converges in distribution to a probability measure $\tilde{\mu}$ on $S \times S$, and

$$\liminf_{T\to\infty} E\left[RE_T^1\right] \geq AR\left(\tilde{\mu}\,\|[\tilde{\mu}]_1 \otimes \alpha\right).$$

Furthermore, if $A > 0$ then $\tilde{\mu}$ satisfies

$$[\tilde{\mu}]_1 = [\tilde{\mu}]_2 . \tag{4.10}$$

Proof. By the chain rule (Theorem 3.3) and the joint convexity of relative entropy

$$\begin{aligned} E\left[RE_T^1\right] &= E\left[\frac{1}{T}\sum_{i=1}^{r_T} R\left(\bar{\alpha}_{i-1}\,\|\alpha\right)\right] \\ &= E\left[\frac{1}{T}\sum_{i=1}^{r_T} R\left(\delta_{\bar{X}_{i-1}} \otimes \bar{\alpha}_{i-1}\,\middle\|\,\delta_{\bar{X}_{i-1}} \otimes \alpha\right)\right] \\ &\geq E\left[\frac{r_T}{T} R\left(\mu_T\,\|[\mu_T]_1 \otimes \alpha\right)\right] \\ &\geq \frac{r_T}{T} R\left(E\mu_T\,\|[E\mu_T]_1 \otimes \alpha\right). \end{aligned}$$

Since $S \times S$ is compact, for any subsequence of T there exists a further subsequence along which $E\mu_T$ converges weakly to a probability measure $\tilde{\mu}$. Under the Feller property of α (Condition 2.2), $[E\mu_T]_1 \otimes \alpha$ converges weakly to $[\tilde{\mu}]_1 \otimes \alpha$. The lower semicontinuity of relative entropy then implies

$$\liminf_{T\to\infty} E\left[RE_T^1\right] \geq \liminf_{T\to\infty} \frac{r_T}{T} R\left(E\mu_T\,\|[E\mu_T]_1 \otimes \alpha\right) \geq AR\left(\tilde{\mu}\,\|[\tilde{\mu}]_1 \otimes \alpha\right).$$

This finishes the first part of Lemma 4.1. For the second part, we employ a standard martingale argument. Let $\mathcal{F}_i$ be the σ-algebra generated by the random variables $\left\{\left(\bar{X}_0, \ldots, \bar{X}_i\right), \left(\bar{\tau}_1, \ldots, \bar{\tau}_i\right)\right\}$. Thus $\mathcal{F}_i$ is a sequence of increasing σ-algebra's and, since $\bar{\alpha}_i$ selects the conditional distribution of $\bar{X}_i$, for any bounded continuous function f on S

$$E\left[\left(f\left(\bar{X}_i\right) - \int_S f\left(y\right)\bar{\alpha}_i\left(dy\right)\right)\middle|\,\mathcal{F}_{i-1}\right] = 0.$$

Hence for integers $0 \leq i < k \leq r_T - 1$

$$E\left[\left(f\left(\bar{X}_i\right) - \int_S f\left(y\right)\bar{\alpha}_i\left(dy\right)\right)\left(f\left(\bar{X}_k\right) - \int_S f\left(y\right)\bar{\alpha}_k\left(dy\right)\right)\right] = 0,$$

and thus for any bounded continuous function f on S

$$\begin{aligned}
&E\left[\int_{S\times S} f(x)\,\mu_T(dx,dy) - \int_{S\times S} f(y)\,\mu_T(dx,dy)\right]^2 \\
&= E\left[\frac{1}{r_T}\sum_{i=1}^{r_T} f\left(\bar{X}_{i-1}\right) - \frac{1}{r_T}\sum_{i=1}^{r_T}\int_S f(y)\,\bar{\alpha}_{i-1}(dy)\right]^2 \\
&\le \frac{1}{r_T^2}\sum_{i=1}^{r_T} E\left[f\left(\bar{X}_{i-1}\right) - \int_S f(y)\,\bar{\alpha}_{i-1}(dy)\right]^2 \\
&\le \frac{4}{r_T}\|f\|_\infty^2 .
\end{aligned}$$

Since $0 < A = \lim_{T\to\infty} r_T/T$, we have $r_T/T \ge A/2$ for all T large enough. Using Chebyshev's inequality and the last display we conclude that $[\mu_T]_1 - [\mu_T]_2$ converges to 0 in probability as $T \to \infty$, and therefore $[\tilde{\mu}]_1 = [\tilde{\mu}]_2$ with probability 1. This concludes the second part of Lemma 4.1. ■

4.2.3 The term RE_T^2

We now turn to the second cost RE_T^2. This cost comes from distorting the exponential sojourn times. We introduce a function ℓ which is closely related to the relative entropy of exponential distributions: $\ell(x) \doteq x\log x - x + 1$ for any $x \ge 0$.

Lemma 4.2 *Given any sequence of controls $\{\bar{\alpha}_i, \bar{\sigma}_i\}$, fix a subsequence of T for which the conclusions in Lemma 4.1 holds. Then we can extract a further subsequence along which*

$$\liminf_{T\to\infty} E\left[RE_T^2\right] \ge \int_{S\times\mathbb{R}_+} \ell(u)\,\tilde{\xi}(dx,du).$$

Here $\tilde{\xi}$ is a finite measure on $S\times\mathbb{R}_+$ and is related to $\tilde{\mu}$ in Lemma 4.1 by

$$\int_{\mathbb{R}_+} u\tilde{\xi}(dx,du) = A\,[\tilde{\mu}]_1(dx). \tag{4.11}$$

Before proving this lemma, we define $g : \mathbb{R}_+ \to \mathbb{R}$ by $g(b) \doteq -\log b + b - 1$. The functions g and ℓ are related by

$$g(x) = x\ell(1/x),$$

and g has the following property.

Lemma 4.3 *Let σ be an exponential distribution with mean 1. Then*

$$\inf\left\{R(\gamma\,\|\sigma) : \int_{\mathbb{R}_+} u\gamma(du) = b\right\} = g(b). \tag{4.12}$$

Proof. Let σ_b be the exponential distribution with mean b, i.e.,

$$\sigma_b(du) = \frac{1}{b} e^{-\frac{u}{b}} du.$$

Then $\frac{d\sigma_b}{d\sigma}(u) = \frac{1}{b} e^{\left(1-\frac{1}{b}\right)u}$ for $u > 0$. Picking any γ such that $R(\gamma \| \sigma) < \infty$ and $\int_{\mathbb{R}_+} u\gamma(du) = b$,

$$\begin{aligned}
R(\gamma \| \sigma) &= \int_{\mathbb{R}_+} \log\left(\frac{d\gamma}{d\sigma}\right) \gamma(du) \\
&= \int_{\mathbb{R}_+} \log\left(\frac{d\gamma}{d\sigma_b}\right) \gamma(du) + \int_{\mathbb{R}_+} \log\left(\frac{d\sigma_b}{d\sigma}\right) \gamma(du) \\
&= R(\gamma \| \sigma_b) + \int_{\mathbb{R}_+} \left[-\log b + \left(1 - \frac{1}{b}\right) u\right] \gamma(du) \\
&= R(\gamma \| \sigma_b) + g(b) \\
&\geq g(b)
\end{aligned}$$

and the infimum in (4.12) is achieved when $R(\gamma \| \sigma_b) = 0$, i.e., $\gamma = \sigma_b$. ■

Proof of Lemma 4.2. Lemma 4.3 guarantees that

$$RE_T^2 \geq \frac{1}{T} \sum_{i=1}^{r_T} g\left(\int u\bar{\sigma}_i(du)\right). \tag{4.13}$$

Recall the definition of $\mathcal{F}_i$ as the σ-algebra generated by the controlled process up to time i. Since $\bar{\sigma}_i$ selects the conditional distribution of $\bar{\tau}_i$,

$$E[\bar{\tau}_i | \mathcal{F}_{i-1}] = \int u\bar{\sigma}_i(du).$$

Define $\bar{m}_i \doteq \int u\bar{\sigma}_i(du)$, for $i = 1, \ldots, r_T - 1$. The definition of $\bar{m}_{r_T}$ requires more work. Recalling the definition of $\bar{R}_T$ by the equation analogous to (4.2) and the restriction that $\bar{R}_T = r_T$,

$$T - \sum_{i=1}^{r_T-1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} \leq \frac{\bar{\tau}_{r_T}}{q\left(\bar{X}_{r_T-1}\right)}.$$

Multiplying both sides by $q\left(\bar{X}_{r_T-1}\right)$ and taking expectation conditioned on $\mathcal{F}_{r_T-1}$,

$$q\left(\bar{X}_{r_T-1}\right)\left(T - \sum_{i=1}^{r_T-1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right) \leq E[\bar{\tau}_{r_T} | \mathcal{F}_{r_T-1}] = \int u\bar{\sigma}_{r_T}(du).$$

Define

$$\bar{\Delta}_T \doteq q\left(\bar{X}_{r_T-1}\right)\left(T - \sum_{i=1}^{r_T-1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right), \tag{4.14}$$

and define $\bar{m}_{r_T}$ by

$$\bar{m}_{r_T} \doteq \begin{cases} \int u\bar{\sigma}_{r_T}(du) & \text{if } \bar{\Delta}_T \leq \int u\bar{\sigma}_{r_T}(du) < 1 \\ 1 & \text{if } \bar{\Delta}_T \leq 1 \leq \int u\bar{\sigma}_{r_T}(du) \\ \bar{\Delta}_T & \text{if } 1 < \bar{\Delta}_T \leq \int u\bar{\sigma}_{r_T}(du) \end{cases} \tag{4.15}$$

i.e., $\bar{m}_{r_T}$ is the median of the triplet $\left(\bar{\Delta}_T, \int u\bar{\sigma}_{r_T}(du), 1\right)$. Since g is increasing on $(1,\infty)$, we have $g\left(\int u\bar{\sigma}_{r_T}(du)\right) \geq g(\bar{m}_{r_T})$ in all three cases. Thus by (4.13),

$$RE_T^2 \geq \frac{1}{T}\sum_{i=1}^{r_T} g\left(\int u\bar{\sigma}_i(du)\right) \geq \frac{1}{T}\sum_{i=1}^{r_T} g(\bar{m}_i). \tag{4.16}$$

Next consider the measure on $S \times \mathbb{R}_+$ defined by

$$\xi_T(dx, du) \doteq \frac{1}{T}\sum_{i=1}^{r_T} \delta_{\bar{X}_{i-1}}(dx)\,\delta_{(\bar{m}_i)^{-1}}(du)\,\bar{m}_i. \tag{4.17}$$

The total mass of $E\xi_T$ is

$$E\xi_T(S \times \mathbb{R}_+) = \frac{1}{T}\sum_{i=1}^{r_T} E[\bar{m}_i].$$

According to (4.9) and the assumption that $F \geq 0$, we have

$$\sup_T E\left[RE_T^2\right] < \infty. \tag{4.18}$$

By (4.16) $\sup_T E\left[\sum_{i=1}^{r_T} g(\bar{m}_i)/T\right] < \infty$. We also have by a straightforward calculation that $x \leq \max\{50, 10g(x)/9\}$. Using this and the fact that $r_T/T \leq C$ we have $\sup_T E\left[\sum_{i=1}^{r_T} \bar{m}_i/T\right] < \infty$, i.e., the total mass of $E\xi_T$ has a bound uniform in T. Thus when viewed as a sequence of measures on the compact space $S \times [0,\infty]$, $E\xi_T$ is tight due to the uniform boundedness of the total mass. We denote the weak limit by $\tilde{\xi}$, which is a finite measure. Since the function ℓ is nonnegative and continuous,

$$\begin{aligned} \liminf_{T\to\infty} E\left[RE_T^2\right] &\geq \liminf_{T\to\infty} E\left[\frac{1}{T}\sum_{i=1}^{r_T} g(\bar{m}_i)\right] \\ &= \liminf_{T\to\infty} E\left[\int_{S\times\mathbb{R}_+} \ell(u)\,\xi_T(dx, du)\right] \\ &= \liminf_{T\to\infty} \int_{S\times\mathbb{R}_+} \ell(u)\,E\xi_T(dx, du) \\ &\geq \int_{S\times\mathbb{R}_+} \ell(u)\,\tilde{\xi}(dx, du). \end{aligned} \tag{4.19}$$

We next explore the relation between $\tilde{\xi}$ and $\tilde{\mu}$. In order to establish (4.11), it suffices to show that for any bounded continuous function f on S

$$\int_{S\times\mathbb{R}_+} uf(x)\,\tilde{\xi}(dx, du) = A\int_S f(x)\,[\tilde{\mu}]_1(dx).$$

By the definitions of ξ_T and μ_T

$$\int_{S\times\mathbb{R}_+} uf(x)\, E\xi_T(dx,du) = \frac{r_T}{T}\int_S f(x)\,[E\mu_T]_1(dx)\,. \tag{4.20}$$

Then (4.18) and (4.19) imply there is a uniform upper bound on

$$\int_{\mathbb{R}_+} \ell(u)\int_S f(x)\, E\xi_T(dx,du)\,.$$

If we consider $\int_S f(x)\, E\xi_T(dx,du)$ as a sequence of measures on $\mathbb{R}_+$ with bounded total mass, then $\int_S f(x)\, E\xi_T(dx,du)$ converges weakly to $\int_S f(x)\,\tilde{\xi}(dx,du)$. Since ℓ is superlinear, [4, Theorem A.3.19] implies that

$$\lim_{T\to\infty}\int_{S\times\mathbb{R}_+} uf(x)\, E\xi_T(dx,du) = \int_{S\times\mathbb{R}_+} uf(x)\,\tilde{\xi}(dx,du)\,.$$

Using

$$\lim_{T\to\infty}\frac{r_T}{T}\int_S f(x)\,[E\mu_T]_1(dx) = A\int_S f(x)\,[\tilde{\mu}]_1(dx)$$

and (4.20) we arrive at (4.11). ■

4.2.4 The term $E\bar{\eta}_T$

Lemma 4.4 *Given any sequence of controls $\{\bar{\alpha}_i,\bar{\sigma}_i\}$, fix a subsequence of T for which the conclusions in Lemma 4.2 hold. Then we can extract a further subsequence along which*

$$\liminf_{T\to\infty} E\left[F(\bar{\eta}_T)\right] \geq F(\tilde{\eta})$$

for some probability measure $\tilde{\eta}$ on S, which is related to $\tilde{\xi}$ in Lemma 4.2 by

$$q(x)\,\tilde{\eta}(dx) = [\tilde{\xi}]_1(dx)\,. \tag{4.21}$$

Proof. As a sequence of probability measures on the compact space S, we can always extract a subsequence of T such that $E\bar{\eta}_T$ converges weakly to a probability measure on S which we denote by $\tilde{\eta}$. The convexity and lower semicontinuity of F imply that

$$\liminf_{T\to\infty} E\left[F(\bar{\eta}_T)\right] \geq \liminf_{T\to\infty} F(E\bar{\eta}_T) \geq F(\tilde{\eta})\,.$$

By the definitions of $\bar{\eta}_T$ in (4.6) and $\bar{\Delta}_T$ in (4.14)

$$\begin{aligned}
&q(x)\, E\bar{\eta}_T(dx) \\
&= \frac{q(x)}{T} E\left[\sum_{i=1}^{r_T-1} \delta_{\bar{X}_{i-1}}(dx)\, \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} + \delta_{\bar{X}_{r_T-1}}(dx)\left(T - \sum_{i=1}^{r_T-1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right] \\
&= \frac{1}{T} E\left[\sum_{i=1}^{r_T-1} \delta_{\bar{X}_{i-1}}(dx)\, \bar{\tau}_i + \delta_{\bar{X}_{r_T-1}}(dx)\, \bar{\Delta}_T\right] \\
&= \frac{1}{T}\left(\sum_{i=1}^{r_T-1} E\left[E\left[\delta_{\bar{X}_{i-1}}(dx)\, \bar{\tau}_i | \mathcal{F}_{i-1}\right]\right] + E\left[\delta_{\bar{X}_{r_T-1}}(dx)\, \bar{\Delta}_T\right]\right) \\
&= \frac{1}{T}\left(\sum_{i=1}^{r_T-1} E\left[\delta_{\bar{X}_{i-1}}(dx)\, \bar{m}_i\right] + E\left[\delta_{\bar{X}_{r_T-1}}(dx)\, \bar{\Delta}_T\right]\right).
\end{aligned}$$

Recalling the definition of ξ_T in (4.17), we have

$$[E\xi_T]_1(dx) = \frac{1}{T}\sum_{i=1}^{r_T} E\left[\delta_{\bar{X}_{i-1}}(dx)\, \bar{m}_i\right].$$

This implies the total variation bound

$$\|q(x)\, E\bar{\eta}_T(dx) - [E\xi_T]_1(dx)\|_{\mathrm{TV}} \leq \frac{1}{T} E\left|\bar{\Delta}_T - \bar{m}_{r_T}\right|.$$

Recalling the definition of $\bar{m}_{r_T}$ in (4.15) we conclude that

$$\|q(x)\, E\bar{\eta}_T(dx) - [E\xi_T]_1(dx)\|_{\mathrm{TV}} \leq \frac{1}{T}.$$

By taking limits we arrive at (4.21). ■

Lemma 4.1, Lemma 4.2 and Lemma 4.4 together imply for a sequence of controls $\{\bar{\alpha}_i, \bar{\sigma}_i\}$ satisfying (4.5), along any subsequence of T such that $r_T/T \to A$, we can extract a further subsequence along which

$$\liminf_{T\to\infty} E\left[F(\bar{\eta}_T) + RE_T^1 + RE_T^2\right] \geq F(\tilde{\eta}) + AR(\tilde{\mu}\,\|[\tilde{\mu}]_1 \otimes \alpha) + \int_{S\times\mathbb{R}_+} \ell(u)\, \tilde{\xi}(dx, du) \tag{4.22}$$

where $\tilde{\eta}$, $\tilde{\mu}$ and $\tilde{\xi}$ satisfy the constraints (4.11), (4.21), and (4.10) if $A > 0$.

Recall that our goal is to prove (4.8). Hence we need to establish the relationship between the right hand side of (4.22) and the rate function I defined in Section 3.1.

4.2.5 Properties of the rate function I

We prove the following lemma, for which we adopt the convention $0 \cdot \infty \doteq 0$. This is in fact the key link, showing that the rate function that is naturally obtained by the weak convergence analysis used to prove the upper bound in fact equals I for suitable measures, and also indicating how to construct controls to prove the lower bound for this same collection of measures. Note that the constraints appearing in the lemma hold for the subsequence appearing in (4.22) due to Lemmas 4.1, 4.2 and 4.4.

Lemma 4.5 *Let $I(\eta)$ be defined by (3.2). Suppose that $\eta \ll \pi$, that μ and ξ satisfy the constraints*

$$q(x)\,\eta(dx) = [\xi]_1(dx) \;\; \text{and} \;\; \int_{\mathbb{R}_+} u\xi(dx,du) = A\,[\mu]_1(dx), \tag{4.23}$$

and that when $A > 0$ the constraint $[\mu]_1 = [\mu]_2$ is also true. Then

$$I(\eta) \le AR(\mu \,\|\, [\mu]_1 \otimes \alpha) + \int_{S\times\mathbb{R}_+} \ell(u)\,\xi(dx,du). \tag{4.24}$$

Moreover,

$$I(\eta) = \inf\left[AR(\mu \,\|\, [\mu]_1 \otimes \alpha) + \int_{S\times\mathbb{R}_+} \ell(u)\,\xi(dx,du)\right]$$

where the infimum is over all possible choices of $A \ge 0$, μ and ξ satisfying these constraints.

The proof of this lemma is detailed. The reason we present it here instead of in an appendix is the previously mentioned fact that the construction of A, μ and ξ that minimize the right hand side of (4.24) indicates how to hit target measures η that are absolutely continuous with respect to the invariant measure in the proof of the Laplace lower bound.

Proof. We first prove the inequality (4.24). If the right hand side of (4.24) is ∞, there is nothing to prove. Hence we assume it is finite. First assume $A > 0$, in which case $R(\mu \,\|\, [\mu]_1 \otimes \alpha) < \infty$. Define

$$Q \doteq \int_S q(x)\,\pi(dx), \tag{4.25}$$

so that by (2.4)

$$\widetilde{\pi}(dx) = q(x)\,\pi(dx)/Q. \tag{4.26}$$

Since $\widetilde{\pi}$ is invariant under α, by [4, Lemma 8.6.2] $[\mu]_1 \ll \widetilde{\pi}$. By (2.1) q is bounded from below, and hence $[\mu]_1 \ll \pi$. Recall that the definition of I in (3.2) uses $\theta = d\eta/d\pi$. Define $\Theta \doteq \{x \in S : \theta(x) = 0\}$. By (4.23)

$$\begin{aligned}
[\mu]_1(dx) &= \frac{\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)}{A}\,[\xi]_1(dx) \\
&= \frac{\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)}{A}\,q(x)\,\eta(dx) \\
&= \frac{\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)}{A}\,q(x)\,\theta(x)\,\pi(dx)
\end{aligned} \tag{4.27}$$

where for a measure ν on $S \times \mathbb{R}_+$, $\nu_{2|1}$ denotes the regular conditional distribution on the second argument given the first. Thus $[\mu]_1(\Theta) = 0$. Now suppose that

$$\int_{S\times S} \theta^{1/2}(x)\,\theta^{1/2}(y)\,q(x)\,\alpha(x,dy)\,\pi(dx) = 0.$$

Then for π-a.e. $x \in S\backslash\Theta$, $\alpha(x,\Theta) = 1$, and hence $(\mu_1 \otimes \alpha)[(S\backslash\Theta) \times \Theta] = 1$. On the other hand, $\mu((S\backslash\Theta) \times \Theta) = 0$ due to $[\mu]_1 = [\mu]_2$. This violates the fact that $R(\mu \| [\mu]_1 \otimes \alpha) < \infty$. We conclude that

$$\int_{S\times S} \theta^{1/2}(x)\,\theta^{1/2}(y)\,q(x)\,\alpha(x,dy)\,\pi(dx) > 0.$$

Lemma 3.2 implies that

$$\begin{aligned}
&-\log \int_{S\times S} \theta^{1/2}(x)\,\theta^{1/2}(y)\,\alpha(x,dy)\,\widetilde{\pi}(dx) \\
&\quad = -\log \int_{S\times S} e^{\frac{1}{2}[\log\theta(x)+\log\theta(y)]}\alpha(x,dy)\,\widetilde{\pi}(dx) \\
&\quad \le R(\mu \| \widetilde{\pi} \otimes \alpha) - \frac{1}{2}\int_{S\times S} [\log\theta(x) + \log\theta(y)]\,\mu(dx,dy). \qquad (4.28)
\end{aligned}$$

Strictly speaking, the inequality above does not fall into the framework of Lemma 3.2 because $\log\theta$ is not bounded. However, if one goes through the proof of this lemma [4, Proposition 1.4.2], then the above inequality is true as long as the right hand side is not of the form $\infty - \infty$. Towards this end, it suffices to prove

$$\frac{1}{2}\int_{S\times S} [\log\theta(x) + \log\theta(y)]\,\mu(dx,dy) = \int_S \log\theta(x)\,[\mu]_1(dx) < \infty. \qquad (4.29)$$

In the appendix we will prove [this being the only place where Condition 2.6 is used] that

$$R([\mu]_1 \| \widetilde{\pi}) < \infty. \qquad (4.30)$$

For now, we assume this is true. Using (4.25), (4.26), and (4.27) to evaluate the relative entropy,

$$\infty > R([\mu]_1 \| \widetilde{\pi}) = \int_S \log\left(\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)\right)[\mu]_1(dx) + \int_S \log\theta(x)\,[\mu]_1(dx) + \log\frac{Q}{A}. \qquad (4.31)$$

We know from (2.1) that $Q \ge K_1$. Also, by (4.23) and the nonnegativity of ℓ

$$\begin{aligned}
&\int_S \log\left(\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)\right)[\mu]_1(dx) \\
&= \frac{1}{A}\int_S \left(\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)\right)\log\left(\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)\right)[\xi]_1(dx) \\
&= \frac{1}{A}\int_S \ell\left(\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)\right)[\xi]_1(dx) + \frac{1}{A}\int_{S\times\mathbb{R}_+} u\xi(dx,du) - \frac{1}{A}\int_S [\xi]_1(dx) \\
&= \frac{1}{A}\int_S \ell\left(\int_{\mathbb{R}_+} u\xi_{2|1}(du|x)\right)[\xi]_1(dx) + \int_S [\mu]_1(dx) - \frac{1}{A}\int_S q(x)\,\eta(dx) \qquad (4.32) \\
&\ge 1 - \frac{1}{A}K_2.
\end{aligned}$$

the second constraint in (4.23) is used for the first equality; the definition of ℓ gives the second equality; both parts of (4.23) assure the third equality; finally the non-negativity of ℓ is used. Thus rearranging (4.31) gives (4.29).

The chain rule of relative entropy gives

$$\begin{aligned}
&R\left(\mu \,\|\widetilde{\pi}\otimes\alpha\right)-\frac{1}{2}\int_{S\times S}\left[\log\theta\left(x\right)+\log\theta\left(y\right)\right]\mu\left(dx,dy\right)\\
&=R\left(\left[\mu\right]_1\|\widetilde{\pi}\right)+\int_S R\left(\mu_{2|1}\|\alpha\right)\left[\mu\right]_1\left(dx\right)-\int_S\log\theta\left(x\right)\left[\mu\right]_1\left(dx\right)\\
&=R\left(\left[\mu\right]_1\|\widetilde{\pi}\right)+R\left(\mu\,\|\left[\mu\right]_1\otimes\alpha\right)-\int_S\log\theta\left(x\right)\left[\mu\right]_1\left(dx\right).
\end{aligned}\tag{4.33}$$

By (4.31) and (4.32) and the convexity of ℓ

$$\begin{aligned}
&R\left(\left[\mu\right]_1\|\widetilde{\pi}\right)-\int_S\log\theta\left(x\right)\left[\mu\right]_1\left(dx\right)\\
&=\int_S\log\left(\int_{\mathbb{R}_+}u\xi_{2|1}\left(du|x\right)\right)\left[\mu\right]_1\left(dx\right)+\log\frac{Q}{A}\\
&=\frac{1}{A}\int_S\ell\left(\int_{\mathbb{R}_+}u\xi_{2|1}\left(du|x\right)\right)\left[\xi\right]_1\left(dx\right)+\int_S\left[\mu\right]_1\left(dx\right)-\frac{1}{A}\int_S q\left(x\right)\eta\left(dx\right)+\log\frac{Q}{A}\\
&\leq\frac{1}{A}\int_{S\times\mathbb{R}_+}\ell\left(u\right)\xi\left(dx,du\right)+1-\frac{1}{A}\int_S q\left(x\right)\eta\left(dx\right)+\log\frac{Q}{A}.
\end{aligned}\tag{4.34}$$

In summary (4.28), (4.33) and (4.34) imply

$$\begin{aligned}
&-\log\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)q\left(x\right)\alpha\left(x,dy\right)\pi\left(dx\right)\\
&\quad=-\log\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)\alpha\left(x,dy\right)\widetilde{\pi}\left(dx\right)-\log Q\\
&\quad\leq R\left(\mu\,\|\left[\mu\right]_1\otimes\alpha\right)+\frac{1}{A}\int_{S\times\mathbb{R}_+}\ell\left(u\right)\xi\left(dx,du\right)+1-\frac{1}{A}\int_S q\left(x\right)\eta\left(dx\right)+\log\frac{1}{A}.
\end{aligned}$$

Thus

$$\begin{aligned}
&-\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)q\left(x\right)\alpha\left(x,dy\right)\pi\left(dx\right)\\
&\leq-\exp\left\{-\left(R\left(\mu\,\|\left[\mu\right]_1\otimes\alpha\right)+\frac{1}{A}\int_{S\times\mathbb{R}_+}\ell\left(u\right)\xi\left(dx,du\right)+1-\frac{1}{A}\int_S q\left(x\right)\eta\left(dx\right)+\log\frac{1}{A}\right)\right\}.
\end{aligned}$$

(4.24) then follows from the fact that $-e^{-r}\leq ar+a\log a-a$ for any $r\in\mathbb{R}$ and $a\in\mathbb{R}_+$ by taking $a=A$ and

$$r=R\left(\mu\,\|\left[\mu\right]_1\otimes\alpha\right)+\frac{1}{A}\int_{S\times\mathbb{R}_+}\ell\left(u\right)\xi\left(dx,du\right)+1-\frac{1}{A}\int_S q\left(x\right)\eta\left(dx\right)+\log\frac{1}{A}.$$

For the case when $A=0$, (4.23) implies that $\int_{\mathbb{R}_+} u\xi\left(dx,du\right)=0$, which means that $\int_{\mathbb{R}_+} u\xi_{2|1}\left(du|x\right)=0$ $[\xi]_1$-a.e. Hence by the convexity of ℓ and $q\left(x\right)\eta\left(dx\right)=[\xi]_1\left(dx\right)$

$$\begin{aligned}\int_{S\times\mathbb{R}_+}\ell\left(u\right)\xi\left(dx,du\right) &\geq \int_S \ell\left(\int_{\mathbb{R}_+} u\xi_{2|1}\left(du|x\right)\right)[\xi]_1\left(dx\right)\\ &= \int_S [\xi]_1\left(dx\right)\\ &= \int_S q\left(x\right)\eta\left(dx\right)\\ &\geq \int_S q\left(x\right)\eta\left(dx\right)-\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)q\left(x\right)\alpha\left(x,dy\right)\pi\left(dx\right)\\ &= I\left(\eta\right).\end{aligned}$$

Thus (4.24) also holds in this case, and completes the proof of the first part of Lemma 4.5.

We now turn to the second part of Lemma 4.5. The definitions and constructions used here will also be used to construct what are essentially optimal controls to prove the reverse inequality in the next section, and indeed the particular forms of the definitions are suggested by that use. In particular, $A\kappa(x)$ will correspond to a dilation of the mean for the exponential random variables. In light of the second part of Lemma 3.2, we define μ by

$$\frac{d\mu}{d\left(\widetilde{\pi}\otimes\alpha\right)}\left(x,y\right)\doteq\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)\Big/\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)\left(\widetilde{\pi}\otimes\alpha\right)\left(dx,dy\right). \quad (4.35)$$

Note that by the Cauchy-Schwartz inequality, the detailed balance condition (2.3) and the relation between π and $\widetilde{\pi}$ (see (2.4)) imply

$$\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)\left(\widetilde{\pi}\otimes\alpha\right)\left(dx,dy\right)\leq\int_{S\times S}\theta\left(x\right)\alpha\left(x,dy\right)\widetilde{\pi}\left(dx\right)\leq\frac{K_2}{Q}.$$

Hence μ is well defined and $[\mu]_1=[\mu]_2$. Then Lemma 3.2 implies that

$$-\log\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)\alpha\left(x,dy\right)\widetilde{\pi}\left(dx\right)=R\left(\mu\,\|\widetilde{\pi}\otimes\alpha\right)-\int_S\log\theta\left(x\right)[\mu]_1\left(dx\right). \quad (4.36)$$

If $R\left(\mu\,\|\widetilde{\pi}\otimes\alpha\right)=\infty$ or $-\int_S\log\theta\left(x\right)[\mu]_1\left(dx\right)=\infty$, the last display implies

$$\int_{S\times S}\theta^{1/2}\left(x\right)\theta^{1/2}\left(y\right)q\left(x\right)\alpha\left(x,dy\right)\pi\left(dx\right)=0.$$

By letting $A\doteq 0$ and $\xi\left(dx,du\right)\doteq q\left(x\right)\eta\left(dx\right)\delta_0\left(du\right)$, then ξ and μ satisfy (4.23) and

$$AR\left(\mu\,\|[\mu]_1\otimes\alpha\right)+\int_{S\times\mathbb{R}_+}\ell\left(u\right)\xi\left(dx,du\right)=\int_S q\left(x\right)\eta\left(dx\right)=I\left(\eta\right).$$

Next assume $R(\mu \| \widetilde{\pi} \otimes \alpha) < \infty$ and $-\int_S \log \theta(x) [\mu]_1 (dx) < \infty$. Define A by

$$A \doteq \exp\left\{ -\left[R(\mu \| \widetilde{\pi} \otimes \alpha) - \int_S \log \theta(x) [\mu]_1 (dx) - \log Q \right] \right\}. \tag{4.37}$$

Define the measure

$$\rho(dx) \doteq q(x) \theta(x) \pi(dx), \tag{4.38}$$

and

$$\kappa \doteq d[\mu]_1 / d\rho. \tag{4.39}$$

Then for any $x \in S \backslash \Theta$ (recall $\Theta = \{x \in S : \theta(x) = 0\}$)

$$\kappa(x) = \frac{d[\mu]_1}{d\rho}(x) = \frac{1}{Q\theta(x)} \frac{d[\mu]_1}{d\widetilde{\pi}}(x).$$

In addition

$$\begin{aligned} \int_S \kappa(x) \log \kappa(x) \rho(dx) &= \int_S \log \kappa(x) [\mu]_1 (dx) \\ &= R([\mu]_1 \| \widetilde{\pi}) - \int_S \log \theta(x) [\mu]_1 (dx) - \log Q. \end{aligned} \tag{4.40}$$

Define

$$b(x) \doteq \begin{cases} 0 & \text{for } x \in \Theta \\ A\kappa(x) & \text{for } x \notin \Theta \end{cases} \tag{4.41}$$

and

$$\xi(dx, du) \doteq q(x) \eta(dx) \delta_{b(x)}(du). \tag{4.42}$$

Then ξ satisfies the first part of (4.23). To see that the second part of (4.23) is satisfied, note that

$$[\mu]_1 (\Theta) = 0 = \int_{\Theta \times \mathbb{R}_+} u\xi(dx, du)$$

and

$$\begin{aligned} \int_{\mathbb{R}_+} u\xi(dx, du) &= b(x) q(x) \eta(dx) \\ &= A\kappa(x) q(x) \theta(x) \pi(dx) \\ &= A[\mu]_1 (dx). \end{aligned}$$

By using the definitions we arrive at the following, each line of which is explained below:

$$
\begin{aligned}
& AR\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) + \int_{S\times\mathbb{R}_+} \ell\left(u\right) \xi\left(dx, du\right) \\
&= AR\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) + \int_S \ell\left(b\left(x\right)\right) q\left(x\right) \eta\left(dx\right) \\
&= AR\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) + \int_\Theta q\left(x\right) \eta\left(dx\right) + \int_{S\setminus\Theta} \ell\left(b\left(x\right)\right) \rho\left(dx\right) \\
&= AR\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) + \int_S q\left(x\right) \eta\left(dx\right) + A\log A - A + A\int_S \kappa\left(x\right) \log\kappa\left(x\right) \rho\left(dx\right) \\
&= \int_S q\left(x\right) \eta\left(dx\right) + A\log A - A + A\left(R\left(\mu \,\|\, \widetilde{\pi} \otimes \alpha\right) - \int_S \log\theta\left(x\right) [\mu]_1\left(dx\right) - \log Q\right) \\
&= \int_S q\left(x\right) \eta\left(dx\right) - A.
\end{aligned}
$$

The first equality uses (4.42) and the second uses (4.41). The third uses (4.41) again, expands ℓ, and uses $\kappa \doteq d\left[\mu\right]_1 / d\rho$ and $\eta\left(\Theta\right) = \rho\left(\Theta\right) = 0$. Equality four then uses (4.40) and the fifth follows from (4.37). Note that (4.36) and (2.4) imply

$$
A = \int_{S\times S} \theta^{1/2}\left(x\right) \theta^{1/2}\left(y\right) q\left(x\right) \alpha\left(x, dy\right) \pi\left(dx\right). \tag{4.43}
$$

Hence we obtain

$$
AR\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) + \int_S \ell\left(u\right) \xi\left(dx, du\right) = I\left(\eta\right).
$$

■

The representation formula (4.4), the lower bound (4.22) and Lemma 4.5 together give

$$
\begin{aligned}
& \liminf_{T\to\infty} -\frac{1}{T} \log E\left[\exp\left\{-TF(\eta_T) - T\cdot\infty\cdot\left(1_{\{r_T/T\}^c}\left(R_T/T\right)\right)\right\}\right] \\
& \qquad \geq \inf_{\eta\in\mathcal{P}(S)} \left[F\left(\eta\right) + I\left(\eta\right)\right].
\end{aligned} \tag{4.44}
$$

4.3 Combining the cases

In the last section, we showed that (4.44) is valid for any sequence $\{r_T\}$ such that $r_T/T \to A \in [0, C]$. An argument by contradiction shows that the bound is uniform in A. Thus

$$
\begin{aligned}
& \liminf_{T\to\infty} -\frac{1}{T} \log\left\{\sum_{r_T=1}^{\lfloor TC\rfloor} E\left[\exp\left\{-TF(\eta_T) - T\cdot\infty\cdot\left(1_{\{r_T/T\}^c}\left(R_T/T\right)\right)\right\}\right]\right\} \\
& \geq \liminf_{T\to\infty} -\frac{1}{T} \log\left\{TC\cdot\bigvee_{r_T=1}^{\lfloor TC\rfloor} E\left[\exp\left\{-TF(\eta_T) - T\cdot\infty\cdot\left(1_{\{r_T/T\}^c}\left(R_T/T\right)\right)\right\}\right]\right\} \\
& \geq \inf_{\eta\in\mathcal{P}(S)} \left[F\left(\eta\right) + I\left(\eta\right)\right].
\end{aligned}
$$

We now partition $E\left[\exp\left\{-TF(\eta_T)\right\}\right]$ according to the two cases to obtain the overall lower bound

$$\begin{aligned}\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\left\{-TF(\eta_T)\right\}\right] \\ \geq \min\left\{\inf_{\eta\in\mathcal{P}(S)}\left[F(\eta)+I(\eta)\right], \left[-C+C\log C+K_2-C\log K_2\right]\right\}\end{aligned}$$

Letting $C\to\infty$ we have the desired Laplace upper bound

$$\liminf_{T\to\infty} -\frac{1}{T}\log E\left[\exp\left\{-TF(\eta_T)\right\}\right] \geq \inf_{\eta\in\mathcal{P}(S)}\left[F(\eta)+I(\eta)\right]. \tag{4.45}$$

5 Proof of Laplace lower bound

We turn to the proof of the reverse inequality

$$\limsup_{T\to\infty} -\frac{1}{T}\log E\left[\exp\left\{-TF(\eta_T)\right\}\right] \leq \inf_{\eta\in\mathcal{P}(S)}\left[F(\eta)+I(\eta)\right]. \tag{5.1}$$

Let F be a nonnegative bounded and continuous function. Fix an arbitrary $\varepsilon>0$ and choose η such that

$$F(\eta)+I(\eta) \leq \inf_{\nu\in\mathcal{P}(S)}\left[F(\nu)+I(\nu)\right]+\varepsilon. \tag{5.2}$$

As pointed out in Remark 2.9, H defined in (3.1) is dense in $\mathcal{P}(S)$. Since I was extended from H to $\mathcal{P}(S)$ via lower semicontinuous regularization, we can assume without loss of generality that $\eta\ll\pi$. Define $\theta\doteq d\eta/d\pi$. We now argue we can further assume there exists $\delta>0$ such that

$$\delta\leq\theta(x)\leq\frac{1}{\delta} \tag{5.3}$$

for all $x\in S$. If $\eta^\delta\doteq(1-\delta)\eta+\delta\pi$ then $d\eta^\delta/d\pi\geq\delta$, and the continuity of F and the convexity of I imply that the difference between $F\left(\eta^\delta\right)+I\left(\eta^\delta\right)$ and $F(\eta)+I(\eta)$ can be made arbitrarily small.

Thus we can assume θ is uniformly bounded from below away from zero. Let $n\in\mathbb{N}$, and define

$$\eta^n(dx)\doteq\theta(x)\,1_{\{\theta(x)\leq n\}}\pi(dx)+\frac{\eta(\{x:\theta(x)>n\})}{\pi(\{x:\theta(x)>n\})}1_{\{\theta(x)>n\}}\pi(dx).$$

Then $d\eta^n/d\pi\leq\left[\eta(\{x:\theta(x)>n\})/\pi(\{x:\theta(x)>n\})\right]\vee n$, and since $\eta\ll\pi$ implies $\pi(\{x:\theta(x)>n\})\to0$, η^n converges weakly to η. It then follows from the continuity of F and the definition of I and convexity of $\theta\to-\theta^{1/2}$ that we can choose η satisfying (5.2) with 2ε replacing ε and also (5.3). Hence we assume η satisfies (5.2) and (5.3). Furthermore by Lusin's Theorem [7, Theorem 7.10], we can also assume that θ is continuous.

The proof of the lower bound will use the following representation. The infimum in the representation is taken over all control measures $\{\bar{\alpha}_i, \bar{\sigma}_i\}$, and the properties of such measures and how $\bar{\eta}_T$ and $\bar{R}_T$ are constructed from them were discussed immediately above the similar representation (4.4). The proof of the lemma is given in the appendix.

Lemma 5.1 *Let $F : \mathcal{P}(S) \to \mathbb{R}$ be bounded and continuous. Then*

$$-\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T\right)\right\}\right] = \inf E\left[F\left(\bar{\eta}_T\right) + \frac{1}{T}\sum_{i=1}^{\bar{R}_T}\left(R\left(\bar{\alpha}_{i-1}\left\|\alpha\right.\right) + R\left(\bar{\sigma}_i\left\|\sigma\right.\right)\right)\right]$$

where the infimum is taken over all control measures $\{\bar{\alpha}_i, \bar{\sigma}_i\}$.

Suppose that given any measure $\eta \in \mathcal{P}(S)$ satisfying (5.2) and (5.3), one can construct $\bar{\alpha}_i$ and $\bar{\sigma}_i$ such that given any subsequence of T, there is a further subsequence T_n such that

$$\lim_{T_n\to\infty} E\left[F\left(\bar{\eta}_{T_n}\right) + \frac{1}{T_n}\sum_{i=1}^{\bar{R}_{T_n}}\left(R\left(\bar{\alpha}_{i-1}\left\|\alpha\right.\right) + R\left(\bar{\sigma}_i\left\|\sigma\right.\right)\right)\right] = F\left(\eta\right) + I\left(\eta\right).$$

Then Lemma 5.1 implies the Laplace lower bound (5.1). The construction of suitable $\bar{\alpha}_i$ and $\bar{\sigma}_i$ turns on many of the same constructions as those used in the proof of the second part of Lemma 4.5. We first define $\mu \in \mathcal{P}(S \times S)$ as in (4.35). Then automatically $[\mu]_1 = [\mu]_2$, and hence if we define p as the regular conditional probability such that $\mu = [\mu]_1 \otimes p$, then $[\mu]_1$ is invariant under p [4, Lemma 8.5.1 (a)]. Define $\bar{\alpha}_i \doteq p$ for each i, and let $\left\{\bar{X}_i\right\}$ be the corresponding Markov chain. Next define $\rho(dx) \doteq q(x)\,\eta(dx)$ and

$$\kappa(x) \doteq \frac{d[\mu]_1}{d\rho}(x) = \frac{1}{Q\theta(x)}\frac{d[\mu]_1}{d\widetilde{\pi}}(x). \tag{5.4}$$

By (5.3), there is $M < \infty$ such that $1/M \le \kappa \le M$, and due to the continuity of θ, κ is also continuous. Notice that

$$\eta(dx) = \left(q(x)\,\kappa(x)\right)^{-1}[\mu]_1(dx). \tag{5.5}$$

Assumption (5.3) guarantees that

$$-\log\int_{S\times S}\theta^{1/2}(x)\,\theta^{1/2}(y)\left(\widetilde{\pi}\otimes\alpha\right)(dx,dy) < \infty \text{ and } -\int_S \log\theta(x)\,[\mu]_1(dx) < \infty,$$

and (4.36) then implies that $R\left(\mu\left\|\widetilde{\pi}\otimes\alpha\right.\right) < \infty$. Define A as in (4.43). Let $\bar{\sigma}_i$ be the exponential distribution with mean $\left[A\kappa\left(\bar{X}_{i-1}\right)\right]^{-1}$ for each i. Thus we can construct a Markov jump process $\bar{X}(t)$ using $\bar{\alpha}_i$ and $\bar{\sigma}_i$ instead of α and σ, and the infinitesimal $\bar{\mathcal{L}}$ generator will be bounded and continuous and takes the form:

$$\bar{\mathcal{L}}f(x) = A\kappa(x)\,q(x)\int_S\left[f(y) - f(x)\right]p(x,dy).$$

(5.5) and the fact that $[\mu]_1$ is invariant under p imply $\int_S \left(\bar{\mathcal{L}} f(x)\right) \eta(dx) = 0$, and η is an invariant distribution of the continuous time process $\bar{X}$. We claim that η is the unique invariant distribution of $\bar{X}$. Indeed, by [6, Proposition 4.9.2] any invariant distribution ν for $\bar{X}$ satisfies $\int_S \left(\bar{\mathcal{L}} f(x)\right) \nu(dx) = 0$. If we define

$$\widetilde{\nu}(dx) \doteq \frac{A\kappa(x)\, q(x)\, \nu(dx)}{\int_S A\kappa(x)\, q(x)\, \nu(dx)},$$

then $\widetilde{\nu}$ is invariant under p. However, by Condition 2.4 and [4, Lemma 8.6.3(c)] the invariant measure under p is unique, and hence the invariant measure of $\bar{X}$ is also unique. By the definition of $\bar{\eta}_T$ in (4.3),

$$\begin{aligned}\bar{\eta}_T(\cdot) &= \frac{1}{T}\int_0^T \delta_{\bar{X}(t)}(\cdot)\, dt \\ &= \frac{1}{T}\left[\sum_{i=1}^{\bar{R}_T-1} \delta_{\bar{X}_{i-1}}(dx)\, \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} + \delta_{\bar{X}_{\bar{R}_T-1}}(dx)\left(T - \sum_{i=1}^{\bar{R}_T-1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right]. \quad (5.6)\end{aligned}$$

Since S is compact we can extract a subsequence of T such that $\bar{\eta}_T$ converges weakly, and by [6, Theorem 4.9.3] this weak limit is η. We claim the following along the same subsequence.

Lemma 5.2 *$E\left[\bar{R}_T/T\right] \to A$, $E\left[\sum_{i=1}^{\bar{R}_T} R(\bar{\sigma}_i \| \sigma)/T\right] \to \int_S \ell(A\kappa(x))\, q(x)\, \eta(dx)$ and $E\left[\sum_{i=1}^{\bar{R}_T} \delta_{\bar{X}_{i-1}}(dx)/T\right] \to A[\mu]_1(dx)$.*

Proof. As in the proof of the upper bound, a minor nuisance is dealing with the residual time $T - \sum_{i=1}^{\bar{R}_T} \bar{\tau}_i$. However, this is more easily controlled here since it is bounded by an exponential with known mean. Since $\bar{\eta}_T \to \eta$ weakly, we have for any bounded and continuous function f on the space of subprobability measures on S that $\lim_{T\to\infty} E[f(\bar{\eta}_T)] = f(\eta)$. To prove the first part of the lemma, define f by

$$f(\nu) \doteq \int_S \kappa(x)\, q(x)\, \nu(dx).$$

Since both κ and q are bounded and continuous, f is also bounded and continuous. Using (5.5)

$$f(\eta) = \int_S \kappa(x)\, q(x)\, \eta(dx) = \int_S [\mu]_1(dx) = 1. \quad (5.7)$$

Thus $\lim_{T\to\infty} E\left[f\left(\bar{\eta}_T\right)\right] = 1$. Now by (5.6) and the definition of $\bar{R}_T$

$$\begin{aligned}
&E\left[\left|f\left(\bar{\eta}_T\right) - f\left(\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\delta_{\bar{X}_{i-1}}\left(dx\right)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right|\right] \\
&= \frac{1}{T}E\left[\kappa\left(\bar{X}_{\bar{R}_T-1}\right)q\left(\bar{X}_{\bar{R}_T-1}\right)\left(\sum_{i=1}^{\bar{R}_T}\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} - T\right)\right] \\
&\leq \frac{1}{T}E\left[\kappa\left(\bar{X}_{\bar{R}_T-1}\right)q\left(\bar{X}_{\bar{R}_T-1}\right)\frac{\bar{\tau}_{\bar{R}_T}}{q\left(\bar{X}_{\bar{R}_T-1}\right)}\right] \\
&\leq \frac{M}{T}E\left[\bar{\tau}_{\bar{R}_T}\right] \\
&\leq \frac{AM^2}{T} \to 0
\end{aligned}$$

as $T \to \infty$. Hence

$$\lim_{T\to\infty} E\left[f\left(\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\delta_{\bar{X}_{i-1}}\left(dx\right)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right] = 1.$$

Recall that $\mathcal{F}_i$ is the σ-algebra generated by $\left\{\left(\bar{X}_0,\ldots,\bar{X}_i\right),\left(\bar{\tau}_1,\ldots,\bar{\tau}_i\right)\right\}$. Then

$$
\begin{aligned}
&E\left[f\left(\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\delta_{\bar{X}_{i-1}}\left(dx\right)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right]\\
&=\frac{1}{T}E\left[\sum_{i=1}^{\bar{R}_T}\kappa\left(\bar{X}_{i-1}\right)q\left(\bar{X}_{i-1}\right)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right]\\
&=\frac{1}{T}E\left[\sum_{i=1}^{\infty}\kappa\left(\bar{X}_{i-1}\right)\bar{\tau}_i\mathbf{1}\left(\sum_{j=1}^{i-1}\frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)}\leq T\right)\right]\\
&=\frac{1}{T}\sum_{i=1}^{\infty}E\left[E\left[\kappa\left(\bar{X}_{i-1}\right)\bar{\tau}_i\mathbf{1}\left(\sum_{j=1}^{i-1}\frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)}\leq T\right)\middle|\mathcal{F}_{i-1}\right]\right]\\
&=\frac{1}{T}\sum_{i=1}^{\infty}E\left[\kappa\left(\bar{X}_{i-1}\right)\mathbf{1}\left(\sum_{j=1}^{i-1}\frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)}\leq T\right)E\left[\bar{\tau}_i|\mathcal{F}_{i-1}\right]\right]\\
&=\frac{1}{T}\sum_{i=1}^{\infty}E\left[\kappa\left(\bar{X}_{i-1}\right)\mathbf{1}\left(\sum_{j=1}^{i-1}\frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)}\leq T\right)\frac{1}{A\kappa\left(\bar{X}_{i-1}\right)}\right]\\
&=\frac{1}{AT}E\left[\sum_{i=1}^{\infty}\mathbf{1}\left(\sum_{j=1}^{i-1}\frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)}\leq T\right)\right]\\
&=\frac{1}{A}E\left[\frac{\bar{R}_T}{T}\right].
\end{aligned}
$$

This completes the proof of the first statement in the lemma.

The proof of the second statement is similar. Define f by

$$
f\left(\nu\right)\doteq\int_S\ell\left(A\kappa\left(x\right)\right)q\left(x\right)\nu\left(dx\right).
$$

Then as before,

$$
f\left(\eta\right)=\lim_{T\to\infty}E\left[f\left(\eta_T\right)\right]=\lim_{T\to\infty}E\left[f\left(\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\delta_{\bar{X}_{i-1}}\left(dx\right)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right].
$$

Using $g(x) = x\ell(1/x)$ and Lemma 4.3, we have

$$
\begin{aligned}
&E\left[f\left(\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\delta_{\bar{X}_{i-1}}(dx)\frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right] \\
&= E\left[\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\ell\left(A\kappa\left(\bar{X}_{i-1}\right)\right)\bar{\tau}_i\right] \\
&= \frac{1}{T}\sum_{i=1}^{\infty}E\left[\ell\left(A\kappa\left(\bar{X}_{i-1}\right)\right)\mathbf{1}\left(\sum_{j=1}^{i-1}\frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)}\leq T\right)E\left[\bar{\tau}_i|\mathcal{F}_{i-1}\right]\right] \\
&= E\left[\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\frac{1}{A\kappa\left(\bar{X}_{i-1}\right)}\ell\left(A\kappa\left(\bar{X}_{i-1}\right)\right)\right] \\
&= E\left[\frac{1}{T}\sum_{i=1}^{\bar{R}_T}g\left(\frac{1}{A\kappa\left(\bar{X}_{i-1}\right)}\right)\right] \\
&= E\left[\frac{1}{T}\sum_{i=1}^{\bar{R}_T}R\left(\bar{\sigma}_i\,\|\sigma\right)\right],
\end{aligned}
$$

and the second part of the lemma follows.

The proof of the third part follows very similar lines as the first two, and is omitted. ■

Now the Laplace lower bound is straightforward. The definition of μ in (4.35), the continuity of θ, and the bound (5.3) imply $x \to R\left(p\left(x,\cdot\right)\|\alpha\left(x,\cdot\right)\right)$ is bounded and continuous. By Lemma 5.2 and the chain rule for relative entropy,

$$
\begin{aligned}
&\lim_{T\to\infty}E\left[F\left(\bar{\eta}_T\right)+\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\left(R\left(\bar{\alpha}_{i-1}\,\|\alpha\right)+R\left(\bar{\sigma}_i\,\|\sigma\right)\right)\right] \\
&= \lim_{T\to\infty}E\left[F\left(\bar{\eta}_T\right)\right]+\lim_{T\to\infty}\int_S R\left(p\left(x,\cdot\right)\|\alpha\left(x,\cdot\right)\right)E\left[\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\delta_{\bar{X}_{i-1}}(dx)\right] \\
&\quad+\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{i=1}^{\bar{R}_T}R\left(\bar{\sigma}_i\,\|\sigma\right)\right] \\
&= F\left(\eta\right)+AR\left(\mu\,\|[\mu]_1\otimes\alpha\right)+\int_S\ell\left(A\kappa\left(x\right)\right)q\left(x\right)\eta\left(dx\right).
\end{aligned}
$$

Returning to the proof of the second part of Lemma 4.5, we find that with this choice of A, μ and κ, the rate function $I(\eta)$ coincides with $AR\left(\mu\,\|[\mu]_1\otimes\alpha\right)+\int_S\ell\left(A\kappa\left(x\right)\right)q\left(x\right)\eta\left(dx\right)$ (note that this η corresponds to a special of Lemma 4.5 where $\Theta \doteq \{x\in S:\theta\left(x\right)=0\}$ is empty). This completes the proof of the Laplace lower bound.

6 On the boundedness of rate function

As pointed out in the Introduction, continuous time jump Markov processes differ from the type of processes considered by Donsker and Varadhan in [2, 3], in that the dynamics do not have a "diffusive" component, and hence Condition 1.1 does not hold. For jump Markov models, the process only moves when a jump occurs, and there is no continuous change of position. For these processes the rate function is bounded, whereas for the processes of [2, 3] the rate function is infinity when the target measure is not absolutely continuous with respect to the reference measure. We now consider the source and implications of this distinction.

Consider a process satisfying all the conditions in Section 2 that has π as its invariant distribution. In order to hit a different probability measure $\eta \in \mathcal{P}(S)$, we need to perturb the original dynamics, which includes the distortion of the Markov chain transition probability α and the distortion of the exponential holding time σ. Each of these distortions must pay a relative entropy cost, and the minimum of the (suitably normalized) sum of these costs asymptotically approximates the rate function $I(\eta)$. When η is singular with respect to π, the relative entropy cost from the distortion of α can be made arbitrarily small, and the rate function is almost entirely due to contributions coming from the distortion of σ. We will illustrate this point via the following example.

Recall the model mentioned in the Introduction, where the state space S is $[0,1]$, the jump intensity is $q \equiv 1$, and for each $x \in [0,1]$, $\alpha(x,\cdot)$ is the uniform distribution on $[0,1]$. The invariant distribution π is just the uniform distribution on $[0,1]$. Now consider a Dirac measure $\eta \doteq \delta_{1/2}$ as a target measure. η is not absolutely continuous with respect to π. However, we can approximate η weakly via a sequence of probability measures that are absolutely continuous with respect to π. For each $n \in \mathbb{N}$ define a probability measure η^n by its Radon-Nikodym derivative θ^n with respect to π according to

$$\theta^n(x) \doteq \begin{cases} n-1 & \text{for } x \in \left(\frac{1}{2}-\frac{1}{2n}, \frac{1}{2}+\frac{1}{2n}\right) \\ \frac{1}{n-1} & \text{otherwise} \end{cases}.$$

Using the formula (3.2) for rate function, we have

$$I(\eta^n) = 1 - \left(\int_0^1 (\theta^n(x))^{1/2}\,dx\right)\left(\int_0^1 (\theta^n(y))^{1/2}\,dy\right) = 1 - \frac{4(n-1)}{n^2}.$$

According to the definition of rate function in Section 3.1, the rate function is bounded above by 1. However $I(\eta^n) \to 1$ as $n \to \infty$, and one can check that this is true for any sequence of absolutely continuous measures converging weakly to η. Thus $I(\eta) = 1$.

We now consider fixed $n \in \mathbb{N}$ and examine the perturbed dynamics that can hit the measure η^n. This is most easily understood by examining the minimizer in the variational formula for the rate function, whose form was suggested during the proof of the Laplace principle lower bound in Section 5. Recall that $\bar{\sigma}_i(\cdot)$ and $\bar{\alpha}_i(\cdot)$ are perturbed dynamics for the exponential holding time and the Markov chain, $\bar{\sigma}_i(\cdot)$ depends on $\left\{\bar{X}_0, \bar{\tau}_1, \bar{X}_1, \bar{\tau}_2, \ldots, \bar{X}_{i-1}\right\}$ and $\bar{\alpha}_i(\cdot)$ depends on $\left\{\bar{X}_0, \bar{\tau}_1, \bar{X}_1, \bar{\tau}_2 \ldots, \bar{X}_{i-1}, \bar{\tau}_i\right\}$. $\bar{\tau}_i$ and

$\bar{X}_i$ are chosen recursively according to stochastic kernels $\bar{\sigma}_i(\cdot)$ and $\bar{\alpha}_i(\cdot)$. Specifically, $\bar{s}_i$ is defined by (1.10) using $\bar{X}_i$ and $\bar{\tau}_i$; $\bar{R}_T$ is defined by (4.2) using $\bar{s}_i$; and $\bar{\eta}_T$ is defined by (4.3) using $\bar{X}_i$, $\bar{\tau}_i$ and $\bar{R}_T$. Following the procedure in Section 5, we first define $\mu \in \mathcal{P}(S \times S)$ as in (4.35). Thus μ is the product measure. As before, we use $[\mu]_1$ to denote the first marginal of μ and p to denote the regular conditional probability such that $\mu = [\mu]_1 \otimes p$. Since μ is a product measure defined by (4.35), $[\mu]_1$ and p are in fact the same measure and the density with respect to π can be calculated as

$$\frac{d[\mu]_1}{d\pi}(x) = \begin{cases} \frac{n}{2} & \text{for } x \in \left(\frac{1}{2} - \frac{1}{2n}, \frac{1}{2} + \frac{1}{2n}\right) \\ \frac{n}{2(n-1)} & \text{otherwise} \end{cases}. \tag{6.1}$$

As in Section 5 let $\bar{\alpha}_i \doteq p$ for each i. A direct calculation of A using formula (4.43) shows that $A = 4(n-1)/n^2$. Also, κ defined in (5.4) reduces to

$$\kappa(x) = \begin{cases} \frac{n}{2(n-1)} & \text{for } x \in \left(\frac{1}{2} - \frac{1}{2n}, \frac{1}{2} + \frac{1}{2n}\right) \\ \frac{n}{2} & \text{otherwise} \end{cases}.$$

As in Section 5, $\bar{\sigma}_i$ should be the exponential distribution with mean $\left[A\kappa\left(\bar{X}_{i-1}\right)\right]^{-1}$. Hence if $\bar{X}_{i-1}$ falls into $(1/2 - 1/(2n), 1/2 + 1/(2n))$, $\bar{\sigma}_i$ would be the exponential distribution with mean $n/2$, otherwise $\bar{\sigma}_i$ would be the exponential distribution with mean $n/[2(n-1)]$. Now the perturbed Markov jump process, denoted by $\bar{X}(t)$, is constructed using $\bar{\alpha}_i$ and $\bar{\sigma}_i$ defined as above. As proved in Lemma 5.2, the expected value of the relative entropy cost

$$\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\left(R\left(\bar{\alpha}_{i-1}\|\alpha\right) + R\left(\bar{\sigma}_i\|\sigma\right)\right)$$

convergences to

$$I(\eta^n) = AR\left(\mu\|[\mu]_1 \otimes \alpha\right) + \int_0^1 \ell\left(A\kappa(x)\right)\eta^n(dx)$$

as $T \to \infty$. We have noted that $p(x, dy) = [\mu]_1(dy)$ and $\alpha(x, dy) = \pi(dy)$, and by using (6.1)

$$\begin{aligned} & AR\left(\mu\|[\mu]_1 \otimes \alpha\right) \\ &= A\int_0^1 R\left(p(x,\cdot)\|\alpha(x,\cdot)\right)[\mu]_1(dx) \\ &= \frac{4(n-1)}{n^2}\left(\log n - \log 2 - \frac{\log(n-1)}{2}\right). \end{aligned}$$

This converges to 0 as $n \to \infty$. Hence the relative entropy cost that comes from the distortion of the Markov chain converges to 0. For the second term, we have

$$\int_0^1 \ell\left(A\kappa(x)\right)\eta^n(dx) = \frac{2(n-1)}{n^2}\left(\log(n-1) + 2\log 2 - \log n\right) - \frac{4(n-1)}{n^2} + 1$$

which converges to 1 as $n \to \infty$. Thus as η^n approaches the target distribution η, the relative entropy cost that comes from the distortion of Markov chain vanishes, and the rate function becomes solely determined by the relative entropy cost that comes from the distortion of exponential waiting times.

One can generalize the argument to more general discrete target measures, where one utilizes the original dynamics to make sure neighborhoods of the various points are visited, and then uses the time dilation to control their relative weight.

7 Appendix

7.1 Proof of inequality (4.30)

Proof. Recall that $R\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) < \infty$, where $[\mu]_1 = [\mu]_2$ and $\widetilde{\pi}$ is invariant under α. Additionally, we also have Condition 2.6, i.e., there exists an integer N and a real number $c \in (0, \infty)$ such that

$$\alpha^{(N)}\left(x, \cdot\right) \leq c\widetilde{\pi}\left(\cdot\right) \tag{7.1}$$

for all $x \in S$. Now let p be the regular conditional probability such that $\mu = [\mu]_1 \otimes p$. Then

$$R\left(\mu \,\|\, [\mu]_1 \otimes \alpha\right) = R\left([\mu]_1 \otimes p \,\|\, [\mu]_1 \otimes \alpha\right) < \infty.$$

The chain rule of relative entropy implies that

$$R\left([\mu]_1 \otimes p \underset{N}{\otimes \cdots} \otimes p \,\middle\|\, [\mu]_1 \otimes \alpha \underset{N}{\otimes \cdots} \otimes \alpha\right) = N \cdot R\left([\mu]_1 \otimes p \,\|\, [\mu]_1 \otimes \alpha\right) < \infty. \tag{7.2}$$

Indeed, since $[\mu]_1$ is invariant under p, for any integer n the n-th marginal of $\left[[\mu]_1 \otimes p \underset{n-1}{\otimes \cdots} \otimes p\right]$ is

$$\left[[\mu]_1 \otimes p \underset{n-1}{\otimes \cdots} \otimes p\right]_n = [\mu]_1 .$$

Hence (7.2) follows by induction:

$$\begin{aligned}
&R\left([\mu]_1 \otimes p \underset{n}{\otimes \cdots} \otimes p \,\middle\|\, [\mu]_1 \otimes \alpha \underset{n}{\otimes \cdots} \otimes \alpha\right) \\
&= R\left([\mu]_1 \otimes p \underset{n-1}{\otimes \cdots} \otimes p \,\middle\|\, [\mu]_1 \otimes \alpha \underset{n-1}{\otimes \cdots} \otimes \alpha\right) + \int_S R\left(p \,\|\, \alpha\right) d\left[[\mu]_1 \otimes p \underset{n-1}{\otimes \cdots} \otimes p\right]_n \\
&= (n-1) \cdot R\left([\mu]_1 \otimes p \,\|\, [\mu]_1 \otimes \alpha\right) + \int_S R\left(p \,\|\, \alpha\right) d\,[\mu]_1 \\
&= n \cdot R\left([\mu]_1 \otimes p \,\|\, [\mu]_1 \otimes \alpha\right).
\end{aligned}$$

Let $[\nu]_{k|j}$ denote the conditional probability of the k-th argument of ν given the j-th argument of ν. Note that one can define a mapping from $\mathcal{P}\left(S^{N+1}\right)$ to $\mathcal{P}\left(S^2\right)$ such

that each $\nu \in \mathcal{P}\left(S^{N+1}\right)$ is mapped to $[\nu]_1 \otimes [\nu]_{N+1|1}$. Since the relative entropy for induced measures is always smaller, (7.2) implies

$$R\left([\mu]_1 \otimes p^{(N)} \middle\| [\mu]_1 \otimes \alpha^{(N)}\right) < \infty.$$

Now since $[\mu]_1$ is invariant under p, it is also invariant under $p^{(N)}$, and therefore $\left[[\mu]_1 \otimes p^{(N)}\right]_2 = [\mu]_1$. Using the chain rule of relative entropy again gives

$$R\left([\mu]_1 \middle\| \left[[\mu]_1 \otimes \alpha^{(N)}\right]_2\right) < \infty.$$

This implies (4.30), since

$$\begin{aligned}
\infty &> R\left([\mu]_1 \middle\| \left[[\mu]_1 \otimes \alpha^{(N)}\right]_2\right) \\
&= R\left([\mu]_1 \| \widetilde{\pi}\right) - \log \int_S \frac{d\left([\mu]_1 \otimes \alpha^{(N)}\right)_2}{d\widetilde{\pi}} [\mu]_1 \\
&\geq R\left([\mu]_1 \| \widetilde{\pi}\right) - \log c,
\end{aligned}$$

where c is from (7.1). ■

7.2 Proof of Lemma 5.1

The proof of the representation is standard, save for the fact that R_T is random. We include a proof here for completeness.

Proof. Define for each $k \in \mathbb{N}_+$

$$\eta_T^k(\cdot) \doteq \frac{1}{T}\left[\sum_{i=1}^{R_T \wedge k-1} \delta_{X_{i-1}}(\cdot) \frac{\tau_i}{q\left(X_{i-1}\right)} + \delta_{X_{R_T \wedge k-1}}(\cdot)\left(T - \sum_{i=1}^{R_T \wedge k-1} \frac{\tau_i}{q\left(X_{i-1}\right)}\right)\right].$$

For any measure $\nu^k \in \mathcal{P}\left((S \times \mathbb{R}_+)^k\right)$, we can decompose ν^k as

$$\nu^k = \bar{\alpha}_0 \otimes \bar{\sigma}_1 \otimes \bar{\alpha}_1 \otimes \bar{\sigma}_2 \otimes \cdots \otimes \bar{\alpha}_{k-1} \otimes \bar{\sigma}_k. \tag{7.3}$$

Choose the barred random variables $\bar{X}_i$ and $\bar{\tau}_i$ according to $\bar{\alpha}_i$ and $\bar{\sigma}_i$ as before and define the corresponding $\bar{R}_T \wedge k$ the following way: if $\sum_{i=1}^k \bar{\tau}_i / q\left(\bar{X}_{i-1}\right) > T$, then $\bar{R}_T \wedge k \doteq \bar{R}_T$ where $\bar{R}_T$ is the integer that satisfies

$$\sum_{i=1}^{\bar{R}_T - 1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} \leq T < \sum_{i=1}^{\bar{R}_T} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)},$$

otherwise define $\bar{R}_T \wedge k \doteq k$. We also define

$$\bar{\eta}_T^k(\cdot) \doteq \frac{1}{T}\left[\sum_{i=1}^{\bar{R}_T \wedge k-1} \delta_{\bar{X}_{i-1}}(\cdot) \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)} + \delta_{\bar{X}_{\bar{R}_T \wedge k-1}}(\cdot)\left(T - \sum_{i=1}^{\bar{R}_T \wedge k-1} \frac{\bar{\tau}_i}{q\left(\bar{X}_{i-1}\right)}\right)\right]. \tag{7.4}$$

If we denote the multi-dimensional probability measure corresponding to the original dynamics by $\mu^k \in \mathcal{P}\left((S \times \mathbb{R}_+)^k\right)$, i.e.,

$$\mu^k \doteq \alpha^{(k)} \times \left(\prod_k \sigma\right),$$

then applying Lemma 3.2 gives

$$-\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T^k\right)\right\}\right] = \inf_{\nu^k \in \mathcal{P}\left((S\times\mathbb{R}_+)^k\right)} \left[\int_{(S\times\mathbb{R}_+)^k} F\left(\bar{\eta}_T^k\right) d\nu^k + \frac{1}{T} R\left(\nu^k \middle\| \mu^k\right)\right]. \tag{7.5}$$

By applying Theorem 3.3 repeatedly to $R\left(\nu^k \middle\| \mu^k\right)$ we obtain

$$R\left(\nu^k \middle\| \mu^k\right) = E\left[\sum_{i=1}^k \left(R\left(\bar{\alpha}_{i-1} \| \alpha\right) + R\left(\bar{\sigma}_i \| \sigma\right)\right)\right].$$

We can thus rewrite (7.5) as

$$-\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T^k\right)\right\}\right] = \inf_{\nu^k \in \mathcal{P}\left((S\times\mathbb{R}_+)^k\right)} E\left[F\left(\bar{\eta}_T^k\right) + \frac{1}{T}\sum_{i=1}^k \left(R\left(\bar{\alpha}_{i-1} \| \alpha\right) + R\left(\bar{\sigma}_i \| \sigma\right)\right)\right] \tag{7.6}$$

Now for each $\nu^k \in \mathcal{P}\left((S \times \mathbb{R}_+)^k\right)$, we construct another measure $\widehat{\nu}^k \in \mathcal{P}\left((S \times \mathbb{R}_+)^k\right)$ recursively as follows: define $\widehat{\alpha}_0 \doteq \bar{\alpha}_0$ and $\widehat{\sigma}_1 \doteq \bar{\sigma}_1$. For all $2 \le i \le k$, define $\widehat{\alpha}_{i-1}$ and $\widehat{\sigma}_i$ by

$$\left(\widehat{\alpha}_{i-1}, \widehat{\sigma}_i\right) = \begin{cases} \left(\bar{\alpha}_{i-1}, \bar{\sigma}_i\right) & \text{if } \sum_{j=1}^{i-1} \frac{\bar{\tau}_j}{q\left(\bar{X}_{j-1}\right)} \le T \\ (\alpha, \sigma) & \text{otherwise} \end{cases}.$$

Thus we return to the original dynamics with zero relative entropy cost after $\bar{R}_T$. Define $\widehat{\nu}^k$ using $\widehat{\alpha}_i$ and $\widehat{\sigma}_i$ by (7.3). From the definition (7.4) we have $E\left[F\left(\widehat{\eta}_T^k\right)\right] = E\left[F\left(\bar{\eta}_T^k\right)\right]$, and

$$\begin{aligned} E\left[\sum_{i=1}^k \left(R\left(\widehat{\alpha}_{i-1} \| \alpha\right) + R\left(\widehat{\sigma}_i \| \sigma\right)\right)\right] &= E\left[\sum_{i=1}^{\widehat{R}_T \wedge k} \left(R\left(\widehat{\alpha}_{i-1} \| \alpha\right) + R\left(\widehat{\sigma}_i \| \sigma\right)\right)\right] \\ &= E\left[\sum_{i=1}^{\bar{R}_T \wedge k} \left(R\left(\bar{\alpha}_{i-1} \| \alpha\right) + R\left(\bar{\sigma}_i \| \sigma\right)\right)\right] \\ &\le E\left[\sum_{i=1}^k \left(R\left(\bar{\alpha}_{i-1} \| \alpha\right) + R\left(\bar{\sigma}_i \| \sigma\right)\right)\right]. \end{aligned}$$

Hence we can rewrite (7.6) as

$$-\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T^k\right)\right\}\right] = \inf_{\nu^k \in \mathcal{P}\left((S\times\mathbb{R}_+)^k\right)} E\left[F\left(\bar{\eta}_T^k\right) + \frac{1}{T}\sum_{i=1}^{\bar{R}_T \wedge k} \left(R\left(\bar{\alpha}_{i-1} \| \alpha\right) + R\left(\bar{\sigma}_i \| \sigma\right)\right)\right]. \tag{7.7}$$

Using the pointwise convergence of both $R_T \wedge k \to R_T$ and $\bar{R}_T \wedge k \to \bar{R}_T$ as $k \to \infty$, by the dominated convergence theorem

$$\lim_{k\to\infty} -\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T^k\right)\right\}\right] = -\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T\right)\right\}\right],$$

$$\lim_{k\to\infty} E\left[F\left(\bar{\eta}_T^k\right)\right] = E\left[F\left(\bar{\eta}_T\right)\right].$$

Also, by the monotone convergence theorem

$$\lim_{k\to\infty} E\left[\sum_{i=1}^{\bar{R}_T\wedge k}\left(\left(R\left(\bar{\alpha}_{i-1}\,\|\alpha\right)+R\left(\bar{\sigma}_i\,\|\sigma\right)\right)\right)\right] = E\left[\sum_{i=1}^{\bar{R}_T}\left(\left(R\left(\bar{\alpha}_{i-1}\,\|\alpha\right)+R\left(\bar{\sigma}_i\,\|\sigma\right)\right)\right)\right].$$

Hence by taking limits on both sides of (7.7), we arrive at

$$-\frac{1}{T}\log E\left[\exp\left\{-TF\left(\eta_T\right)\right\}\right] = \inf \overline{E}\left[F\left(\bar{\eta}_T\right)+\frac{1}{T}\sum_{i=1}^{\bar{R}_T}\left(\left(R\left(\bar{\alpha}_{i-1}\,\|\alpha\right)+R\left(\bar{\sigma}_i\,\|\sigma\right)\right)\right)\right]$$

where the infimum is taken over all controlled measures $\{\bar{\alpha}_i, \bar{\sigma}_i\}$. This proves the lemma. ■

References

[1] M. Chen and Y. Lu. On evaluating the rate function of large deviations for jump processes. *Acta Mathematica Sinica, New Series*, 6:206–219, 1990.

[2] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.*, 28:1–47, 1975.

[3] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, III. *Comm. Pure Appl. Math.*, 29:389–461, 1976.

[4] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York, 1997.

[5] R.S. Ellis and A.D. Wyner. Uniform large deviation property of the empirical process of a Markov chain. *Ann. Probab.*, 17:1147–1151, 1989.

[6] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[7] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, New York, 1999.

[8] Y. Liu. *Large Deviations Rate Functions for the Empirical Measure: Explicit Formulas and an Application to Monte Carlo.* PhD thesis, Brown University, 2013.

[9] R. Pinsky. On evaluating the Donsker-Varadhan I-function. *Ann. Probab.*, 13:342–362, 1985.

[10] Walter Rudin. *Functional Analysis.* McGraw-Hill, New York, 1991.

[11] D.W. Stroock. *An Introduction to the Theory of Large Deviations.* Springer-Verlag, New York, 1984.