Lectures on the Large Deviation Principle

Fraydoun Rezakhanlou*
Department of Mathematics, UC Berkeley

March 25, 2017

*This work is supported in part by NSF grant DMS-1407723

1 Introduction

Many questions in probability theory can be formulated as a law of large numbers (LLN). Roughly a LLN describes the most frequently visited (or occurred) states in a large system. To go beyond LLN, we either examine the states that deviate from the most visited states by small amount, or those that deviate by large amount. The former is the subject of Central Limit Theorem (CLT). The latter may lead to a Large Deviation Principle (LDP) if the probability of visiting a non-typical state is exponentially small and we can come up with a precise formula for the exponential rate of convergence as the size of the system goes to infinity.

In this introduction we attempt to address four basic questions:

- 1. What does LDP mean?
- 2. What are some of our motivations to search for an LDP?
- 3. How ubiquitous is LDP?
- 4. What are the potential applications?

We use concrete examples to justify our answers to the above questions. To prepare for answering the first two questions, let us describe a scenario that is often encountered in *equilibrium statistical mechanics*: Imagine that we are studying a model of an evolving state and we have a candidate for the energy of each occurring state. When the system reaches equilibrium, states with lower energies are more likely to occur. A simple and natural model for such a system at equilibrium is given by a *Gibbs measure*. If the space of states E is finite, $\beta > 0$ and $H : E \to \mathbb{R}$ is the energy function, then a measure of the form

$$\nu_\beta(x) = \frac{1}{Z_\beta} e^{-\beta H(x)} = e^{-\beta H(x) - \log Z_\beta} =: e^{-\beta I(x)}, \tag{1.1}$$

assigns more weight to states of lower energy. (Here Z_β is the normalizing constant: $Z_\beta = \sum_{x \in E} e^{-\beta H(x)}$.) Note that the most probable state is the one at which H takes its smallest value. Roughly, an LDP means that we are dealing with probability measures that are approximately of the form (1.1) with a constant β that goes to ∞ as the size of the system increases without bound. We explain this by two models: Bernoulli trials and their continuum analog Brownian motion.

In a Bernoulli trial, experiments with exactly two possible outcomes, say 0 and 1 are repeated independently from each other. Let $p \in (0, 1)$ be the probability of the occurrence of 1 in each trial. Writing X_n for the value of the n-th trial and $S_n = X_1 + \cdots + X_n$, we certainly have that for $x \in [0, 1]$

$$\mathbb{P}(S_n = [nx]) = \binom{n}{[nx]} p^{[nx]} (1-p)^{n-[nx]},$$

where $[nx]$ denotes the integer part of nx. We write $\ell = [nx]$ and use Stirling's formula to assert that if $x \in (0,1)$, then $\mathbb{P}(S_n = [nx])$ is approximately equal to

$$\begin{aligned}
&\sqrt{\frac{n}{2\pi\ell(n-\ell)}} \exp\left[n\log n - \ell\log\ell - (n-\ell)\log(n-\ell) + \ell\log p + (n-\ell)\log(1-p)\right] \\
&= \sqrt{\frac{n}{2\pi\ell(n-\ell)}} \exp\left\{n\left[\frac{\ell}{n}\log p + \left(1-\frac{\ell}{n}\right)\log(1-p) - \frac{\ell}{n}\log\frac{\ell}{n} - \left(1-\frac{\ell}{n}\right)\log\left(1-\frac{\ell}{n}\right)\right]\right\} \\
&= \frac{1}{\sqrt{2\pi n x(1-x)}} \exp\left(-nI(x) + O(\log n)\right) = \exp\left(-nI(x) + O(\log n)\right),
\end{aligned}$$

where

$$I(x) = x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}.$$

Note that I is a convex function with the following properties,

$$I(p) = I'(p) = 0, \quad I''(p) = (p(1-p))^{-1}.$$

In particular, $I \geq 0$ and takes its minimum value at a unique point $x = p$. This is consistent with the fact that $n^{-1}S_n \to p$ as $n \to \infty$. Moreover

$$\lim_{n\to\infty} n^{-1}\log\mathbb{P}(S_n = [nx]) = -I(x), \tag{1.2}$$

for $x \in [0,1]$. In the language of Large Deviation Theory, the sequence $n^{-1}S_n$ satisfies an LDP with rate function $I(x)$. When $x \neq p$, the set $\{S_n = [nx]\}$ is regarded as an (exponentially) rare event and the limit (1.2) offers a precise exponential rate of the convergence of its probability to 0. Also note that if we consider a small deviation $x = p + y/\sqrt{n}$, then the Taylor expansion

$$nI(x) = \frac{1}{2}I''(p)y^2 + O(n^{-\frac{1}{2}}),$$

is compatible with the CLT

$$\mathbb{P}(S_n = [nx]) \simeq \frac{1}{\sqrt{2\pi p(1-p)}} e^{-\frac{1}{2}I''(p)y^2}.$$

Indeed the rate function I in (1.2) has very much the same flavor as our I in (1.1) and can be interpreted as the difference between an energy function and some kind of *entropy*. By this we mean that we may write $I = H - J$ where

$$H(x) = -x\log p - (1-x)\log(1-p), \quad J(x) = -x\log x - (1-x)\log(1-x).$$

This is a typical phenomenon and can be explained heuristically in the following way: To evaluate $\mathbb{P}(S_n = [nx])$, we need to do two things;

- Count those "micro-states" that correspond to the "macro-state" x. We have exactly

$$\binom{n}{[nx]} \simeq \exp(nJ(x)),$$

  many such micro-states.

- Calculate the probability of such micro-states which is approximately $\exp(-nH(x))$.

As we will see later on, most rate functions can be expressed as a difference of an energy function and entropy function. Yet another useful interpretation is that the rate function in many examples of interest (including the Bernoulli trial) can be written as some kind of *relative entropy.*

As our second example, we would like to examine the Wiener measure from Large Deviation point of view. We note that the standard d-dimensional Brownian measure may be regarded as a probability measure $\mathbb{P}$ (known as Wiener measure) on the space $\Omega = C([0,T];\mathbb{R}^d)$ of continuous trajectories $x : [0,T] \to \mathbb{R}^d$ such that for $t_0 = 0 < t_1 < \cdots < t_{k-1} < t_k = T$ and $y_0 = 0$, we have

$$\mathbb{P}(x(t_1) \in dy_1, \ldots, x(t_k) \in dy_k) = Z^{-1}_{t_1,\ldots,t_k} \exp\left\{-\frac{1}{2}\sum_{i=1}^{k} \frac{|y_i - y_{i-1}|^2}{t_i - t_{i-1}}\right\} dy_1 \ldots dy_k, \tag{1.3}$$

where

$$Z_{t_1,\ldots,t_k} = \prod_{i=1}^{k} (2\pi(t_i - t_{i-1}))^{d/2}.$$

Observe that the formula (1.3) resembles (1.1), and if we choose finer and finer grids $(t_0, t_1, \ldots, t_k)$ of the interval $[0,T]$, the energy function

$$\frac{1}{2}\sum_{i=1}^{k} \frac{|y_i - y_{i-1}|^2}{t_i - t_{i-1}} = \frac{1}{2}\sum_{i=1}^{k} (t_i - t_{i-1}) \left(\frac{|y_i - y_{i-1}|}{t_i - t_{i-1}}\right)^2,$$

approximates

$$\frac{1}{2}\int_0^T |\dot{x}(t)|^2 \, dt.$$

We are tempted to say that the probability of observing a path $x : [0,T] \to \mathbb{R}^d$ can be expressed as

$$\frac{1}{Z(T)} \exp\left(\frac{1}{2}\int_0^T |\dot{x}(t)|^2 \, dt\right). \tag{1.4}$$

To be more realistic, we wish to say that the Wiener measure is absolutely continuous with respect to a “Lebesgue-like” measure on Ω and its density is given by (1.4). To make sense of this, we encounter various challenges. For one thing, the normalizing constant $Z_{t_1,\dots,t_k}$ does not have a limit as we choose finer and finer grids for the interval $[0,T]$. Moreover, we do not have a good candidate for a “Lebesgue-like” (i.e. translation invariant) reference measure to compare $\mathbb{P}$ with. Nonetheless, the formula (1.4) can be utilized to predict two important properties of Brownian motion. For example we can use (1.4) to discover Cameron-Martin Theorem and Girsanov's formula. Here's how it goes: take a function $y \in \Omega$ and consider the translated Wiener measure $\mathbb{P}_y := \tau_y \mathbb{P}$ that is defined by

$$\int F(x(\cdot))\ \mathbb{P}_y(dx(\cdot)) = \int F(x(\cdot)+y(\cdot))\ \mathbb{P}(dx(\cdot)),$$

for every bounded continuous test function $F : \Omega \to \mathbb{R}$. A natural question is whether $\mathbb{P}_y \ll \mathbb{P}$ and if this is the case what is $\frac{d\mathbb{P}_y}{d\mathbb{P}}$? We can easily guess the answer if we use the formula (1.4). Indeed

$$\begin{aligned}\mathbb{P}_y(dx(\cdot)) \text{“}=\text{”} &Z^{-1}\ e^{-\frac{1}{2}\int_0^T |\dot{x}-\dot{y}|^2 dt}\ dx(\cdot)\\ &= Z^{-1}\ e^{-\frac{1}{2}\int_0^T |\dot{x}|^2 dt}\ e^{-\frac{1}{2}\int_0^T |\dot{y}|^2\ dt}\ e^{\int_0^T \dot{x}\cdot y\ dt}\ dx(\cdot)\\ &= e^{\int_0^T y\cdot dx - \frac{1}{2}\int_0^T |\dot{y}|^2 dt}\ \mathbb{P}(dx(\cdot)).\end{aligned}$$

Since the last line in the above formal calculation is all well-defined, we guess that $\mathbb{P}_y \ll \mathbb{P}$ if and only if $\int_0^T |\dot{y}|^2\ dt < \infty$ (Cameron-Martin) and that

$$\frac{d\mathbb{P}_y}{d\mathbb{P}} = \exp\left(\int_0^T y\cdot dx - \frac{1}{2}\int_0^T |\dot{y}|^2\ dt\right), \qquad (Girsanov). \tag{1.5}$$

Large Deviation Theory allows us to formulate a variant of (1.4) that is well-defined and can be established rigorously. The point is that if we take a small Brownian trajectory $\sqrt{\varepsilon}x(\cdot)$ and force it to be near a given $y \in \Omega$, then for $y \neq 0$ this is a rare event and the energy of such trajectory is so large that dominates the probability of its occurrence. That is, the normalizing constant Z can be safely ignored to assert

$$\mathbb{P}(\sqrt{\varepsilon}x(\cdot) \text{ is near } y(\cdot)) \approx \exp\left(-\frac{1}{2\varepsilon}\int_0^T |\dot{y}|^2\ dt\right). \tag{1.6}$$

This is Schilder's LDP and its generalization to general stochastic differential equations (SDE) is the cornerstone of the Wentzell-Freidlin Theory. Roughly, if x^ε solves

$$dx^\varepsilon = b(x^\varepsilon,t)\ dt + \sqrt{\varepsilon}\sigma(x^\varepsilon,t)\ dB,$$

for a standard Brownian motion B and an invertible matrix σ, then

$$\mathbb{P}(x^\varepsilon(\cdot) \text{ is near } y(\cdot)) \approx \exp\left(-\frac{1}{2\varepsilon}\int_0^T \left|\sigma(y,t)^{-1}\left(\dot{y}(t) - b(y(t),t)\right)\right|^2 \, dt\right). \tag{1.7}$$

The LDP (1.7) provides us with a powerful tool for examining how a small random perturbation of the ODE $\dot{x} = b(x,t)$ can affect its trajectories.

Our LDP (1.2) for the Bernoulli trial is an instance of the Cramér/Sanov LDP. More generally we may take a sequence of E-valued iid random variables $X_1, X_2, \dots, X_k, \dots$ with

$$\mathbb{P}(X_i \in A) = \mu(A),$$

and examine large deviations of the empirical measure $n^{-1}(\delta_{X_1} + \dots + \delta_{X_n})$ from μ, as $n \to \infty$. According to Sanov's theorem,

$$\mathbb{P}\left(n^{-1}(\delta_{X_1} + \dots + \delta_{X_n}) \text{ is near } \nu\right) \approx \exp\left(-n^{-1}H(\nu|\mu)\right), \tag{1.8}$$

where $H(\nu|\mu)$ is the entropy of ν relative to μ (aka KullbackLeibler divergence):

$$H(\nu|\mu) = \int \log \frac{d\nu}{d\mu} \, d\nu.$$

A sweeping generalization of Sanov's theorem was achieved by Donsker and Varadhan. To explain their result, let us set $\mathcal{E} = E^{\mathbb{Z}}$ to denote the space of sequences $\mathbf{x} = (x_i : i \in \mathbb{Z})$ and define $T : \mathcal{E} \to \mathcal{E}$ to be the shift operator:

$$(T\mathbf{x})_i = x_{i+1}.$$

Given a probability measure μ on E, we write $\mathbb{P}_\mu$ for the probability measure on Ω that we obtain by taking the products of μ's. Clearly the probability measure $\mathbb{P}_\mu$ is an invariant measure for the dynamical system associated with T. That is, for any bounded continuous $F : \mathcal{E} \to \mathbb{R}$, we have $\int F \, d\mathbb{P}_\mu = \int F \circ T \, d\mathbb{P}_\mu$. Given $\mathbf{x} \in \mathcal{E}$, we may define an empirical measure

$$\nu_n(\mathbf{x}) = n^{-1}\left(\delta_{\mathbf{x}} + \delta_{T(\mathbf{x})} + \dots + \delta_{T^{n-1}(\mathbf{x})}\right).$$

Note that $\nu_n(\mathbf{x})$ is a probability measure on $\mathcal{E}$ and by the celebrated Birkhoff Ergodic Theorem

$$\mathbb{P}_\mu\left\{\mathbf{x} : \lim_{n\to\infty} \nu_n(\mathbf{x}) = \mathbb{P}_\mu\right\} = 1.$$

By Donsker-Varadhan Theory, there is an LDP for the deviations of $\gamma_n(\omega)$ from $\mathbb{P}_\mu$: Given any T-invariant measure $\mathbb{Q}$,

$$\mathbb{P}_\mu\left(\nu_n(\mathbf{x}) \text{ is near } Q\right) \approx \exp\left(-n\mathcal{H}_\mu(Q)\right), \tag{1.9}$$

where $\mathcal{H}_\mu(Q)$ is closely related to *Kolmogorov-Sinai entropy.* This entropy is also known as the *metric entropy* and is a fundamental quantity in the theory of Dynamical Systems.

We now give a precise definition for LDP. Throughout, all probability measures are on a measure space $(E, \mathcal{B})$, where E is a *Polish* (separable complete metric space) and $\mathcal{B}$ is the corresponding σ-algebra of the Borel sets. To motivate the definition of LDP, let us recall two facts:

1. By definition, a sequence of probability measures $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ converges weakly to a probability measure $\mathbb{P}$ if and only if for every bounded continuous $F : E \to \mathbb{R}$,

$$\lim_{n\to\infty} \int F \, d\mathbb{P}_n = \int F \, d\mathbb{P}. \tag{1.10}$$

Equivalently,

$$\text{For every open set } U, \quad \liminf_{n\to\infty} \mathbb{P}_n(U) \geq \mathbb{P}(U), \quad \text{or} \tag{1.11}$$

$$\text{For every closed set } C, \quad \limsup_{n\to\infty} \mathbb{P}_n(C) \leq \mathbb{P}(C). \tag{1.12}$$

2. If $a_1, \ldots, a_k \in \mathbb{R}$, then

$$\lim_{n\to\infty} n^{-1} \log \left(\sum_{i=1}^{k} e^{-na_i} \right) = -\inf_i a_i. \tag{1.13}$$

Definition 1.1 Let $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ be a family of probability measures on a Polish space E and let $I : E \to [0, \infty]$ be a function.

1.We then say that the family $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ satisfies a *large deviation principle* (in short LDP) with rate function I, if the following conditions are satisfied:

(i) For every $a \geq 0$, the set $\{x : I(x) \leq a\}$ is compact.

(ii) For every open set U

$$\liminf_{n\to\infty} n^{-1} \log \mathbb{P}_n(U) \geq -\inf_{x\in U} I(x). \tag{1.14}$$

(iii) For every closed set C

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(C) \leq -\inf_{x\in C} I(x). \tag{1.15}$$

2. We say that the family $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ satisfies a *weak large deviation principle* (in short WLDP) with rate function I, if I is lower-semi continuous, (1.14) is valid for open sets, and (1.15) is true only for *compact* sets. $\square$

Remark 1.1 The statements (1.14) and (1.15) together is equivalent to saying that for every Borel set A,

$$-\inf_{A^o} I \le \liminf_{n\to\infty} n^{-1}\log \mathbb{P}_n(A) \le \limsup_{n\to\infty} n^{-1}\log \mathbb{P}_n(A) \le -\inf_{\bar{A}} I.$$

In particular, if for a set A we have that $\inf_{A^o} I = \inf_{\bar{A}} I$, then

$$\lim_{n\to\infty} n^{-1}\log \mathbb{P}_n(A) = -\inf_{A} I.$$

$\square$

We note that our requirements (1.14) and (1.15) are modeled after (1.11) and (1.12). Alternatively, we may establish an LDP by verifying this:

- For every bounded continuous $F : E \to \mathbb{R}$

$$\Lambda(F) := \lim_{n\to\infty} n^{-1}\log\int e^{nF}\, d\mathbb{P}_n = \sup_E (F - I). \tag{1.16}$$

Intuitively (1.16) is true because LDP roughly means

$$\mathbb{P}_n(y \text{ is near } x) \approx e^{-nI(x)}$$

and a sum exponentials is dominated by the largest exponential (see (1.13)).

Theorem 1.1 *(Varadhan) Given a rate function I that satisfies the condition* (**i**) *of Definition 1.1, the statements* (1.14) *and* (1.15) *are equivalent to the statement* (1.16). *Moreover the following statements are true:*

(i) *If F is lower semi-continuous and* (1.14) *is true for every open set, then*

$$\liminf_{n\to\infty} n^{-1}\log\int e^{nF}\, d\mathbb{P}_n \ge \sup_E (F - I). \tag{1.17}$$

(ii) *If F is upper semi-continuous and bounded above, and* (1.15) *is true for every closed set, then*

$$\limsup_{n\to\infty} n^{-1}\log\int e^{nF}\, d\mathbb{P}_n \le \sup_E (F - I). \tag{1.18}$$

(iii) *If* (1.17) *is true for every bounded continuous F, then* (1.14) *is valid.*

(iv) *If* (1.18) *is true for every bounded continuous F, then* (1.15) *is valid.*

Proof (i) To prove (1.17) take any $x \in E$ and $\varepsilon > 0$, and set $U_x = \{y : F(y) > F(x) - \varepsilon\}$, which is open by the lower semi-continuity assumption. We certainly have

$$\begin{aligned}\liminf_{n\to\infty} n^{-1} \log \int e^{nF}\ d\mathbb{P}_n &\geq \liminf_{n\to\infty} n^{-1} \log \int_{U_x} e^{nF}\ d\mathbb{P}_n \\ &\geq F(x) - \varepsilon - \inf_{U_x} I \geq F(x) - I(x) - \varepsilon.\end{aligned}$$

We then send $\varepsilon \to 0$ to deduce (1.17).

(ii) Fix some $\ell, \varepsilon > 0$ and set

$$W_x = \{y : \ F(y) < F(x) + \varepsilon,\ I(y) > I(x) - \varepsilon\}, \qquad K_\ell = \{x : I(x) \leq \ell\}.$$

By upper semi-continuity of F, and **(i)** of Definition 1.1, the set W_x is open and the set K_ℓ is compact. Choose an open set U_x such that

$$x \in U_x \subseteq \bar{U}_x \subseteq W_x.$$

By compactness of K_ℓ, there are $x_1, \dots, x_k \in K_\ell$ such that

$$K_\ell \subseteq U_{x_1} \cup \cdots \cup U_{x_k} := V.$$

Evidently,

$$\int e^{nF}\ d\mathbb{P}_n \leq \int_{E\setminus V} e^{nF}\ d\mathbb{P}_n + \sum_{i=1}^{k} \int_{U_{x_i}} e^{nF}\ d\mathbb{P}_n.$$

From this and (1.13) we deduce

$$\limsup_{n\to\infty} n^{-1} \log \int e^{nF}\ d\mathbb{P}_n \leq \max\{a, a_1, \dots, a_k\}, \tag{1.19}$$

where

$$\begin{aligned} a &= \limsup_{n\to\infty} n^{-1} \log \int_{E\setminus V} e^{nF}\ d\mathbb{P}_n, \quad \text{and} \\ a_i &= \limsup_{n\to\infty} n^{-1} \log \int_{U_{x_i}} e^{nF}\ d\mathbb{P}_n,\end{aligned}$$

for $i = 1, \dots, k$. Furthermore, by (1.15)

$$\begin{aligned} a &\leq \sup_E F + \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(E \setminus V) \\ &\leq \sup_E F - \inf_{E\setminus V} I \leq \sup_E F - \inf_{E\setminus K_\ell} I \leq \sup_E F - \ell, \\ a_i &\leq F(x_i) + \varepsilon + \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(U_{x_i}) \\ &\leq F(x_i) + \varepsilon + \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(\bar{U}_{x_i}) \\ &\leq F(x_i) + \varepsilon - \inf_{W_{x_i}} I \leq F(x_i) + 2\varepsilon - I(x_i) \leq \sup_E (F - I) + 2\varepsilon.\end{aligned}$$

From this and (1.19) we deduce

$$\limsup_{n\to\infty} n^{-1}\log\int e^{nF}\ d\mathbb{P}_n \le \max\left\{\sup_E F-\ell,\sup_E(F-I)+2\varepsilon\right\}.$$

We send $\varepsilon\to 0$ and $\ell\to\infty$ to deduce (1.18).

(iii) Take an open set U. Pick $x\in U$ and $\delta>0$ such that $B(x,\delta)\subseteq U$. Assume that $I(x)<\infty$, pick $\ell>0$, and set

$$F(y)=-\ell\ \min\left\{\delta^{-1}d(x,y),1\right\}.$$

By assumption

$$\liminf_{n\to\infty} n^{-1}\log\int e^{nF}\ d\mathbb{P}_n \ge \sup_E(F-I)\ge F(x)-I(x)=-I(x). \tag{1.20}$$

On the other hand,

$$\begin{aligned}\liminf_{n\to\infty} n^{-1}\log\int e^{nF}\ d\mathbb{P}_n &\le \liminf_{n\to\infty} n^{-1}\log\left[e^{-n\ell}+\mathbb{P}_n(B(x,\delta))\right]\\ &\le \max\left\{-\ell\ ,\ \liminf_{n\to\infty} n^{-1}\log\mathbb{P}_n(B(x,\delta))\right\}.\end{aligned}$$

From this and (1.20) we learn

$$-I(x)\le\max\left\{-\ell\ ,\ \liminf_{n\to\infty} n^{-1}\log\mathbb{P}_n(B(x,\delta))\right\}.$$

This is true for every $\ell>0$. By sending $\ell\to\infty$ we deduce (1.14).

(iv) Note that $\mathbb{P}_n(C)=\int e^{n\chi_C}\ dP_n$, where

$$\chi_C(x)=\begin{cases}0 & \text{if } x\in C,\\ -\infty & \text{if } x\notin C.\end{cases}$$

Given a closed set C, we approximate the upper semi-continuous function χ_C from above by a sequence bounded continuous functions given by

$$F_\ell(x)=-\ell\min\{d(x,C),1\}.$$

That is, $F_\ell\downarrow\chi_C$ as $\ell\to\infty$. By (1.18),

$$\limsup_{n\to\infty} n^{-1}\log\mathbb{P}_n(C)\le\limsup_{n\to\infty} n^{-1}\log\int e^{nF_\ell}\ d\mathbb{P}_n\le\sup(F_\ell-I).$$

We are done if we can show

$$\inf_\ell \sup_E (F_\ell - I) \le -\inf_C I. \tag{1.21}$$

We first show (1.21) assuming that $\inf_C I < \infty$. Suppose to the contrary (1.21) is not valid. If this is the case, then we can find $\varepsilon > 0$ such that

$$\inf_\ell \sup_{E\setminus C} (F_\ell - I) \ge -\inf_C +\varepsilon.$$

As a result, we can find a sequence x_ℓ such that

$$\ell \min\{d(x_\ell, C), 1\} + I(x_\ell) \le \inf_C I - \varepsilon.$$

From this we learn that $\ell d(x_\ell, C)$ is bounded, which in turn implies that $d(x_\ell, C) \to 0$ as $\ell \to \infty$. We also know that $I(x_\ell) \le \inf_C I - \varepsilon$. The compactness of the set

$$\{y : I(y) \le \inf_C I - \varepsilon\},$$

allows us to take a subsequence of $\{x_\ell\}_\ell$ that converges to a point $y \in C$. By the lower semi-continuity of I we deduce that $I(y) \le \inf_C I - \varepsilon$, which is absurd because $y \in C$. Thus, (1.21) must be true.

Finally, if $\inf_C I = \infty$ and (1.21) is not valid, then

$$\liminf_{\ell\to\infty} \sup_{E\setminus C} (F_\ell - I) \ge -A,$$

for some finite A. As a result, for a sequence x_ℓ,

$$\ell \min\{d(x_\ell, C), 1\} + I(x_\ell) \le A.$$

Again $d(x_\ell, C) \to 0$ as $\ell \to \infty$ and $I(x_\ell) \le A$. Hence we can take a subsequence of $\{x_\ell\}_\ell$ that converges to a point $y \in C$. By lower semi-continuity of I we deduce that $I(y) \le A$, which is absurd because $\inf_C I = \infty$. This completes the proof. □

As we will see later on, we establish LDP by verifying (1.16) for a suitable family of functions F. For this to work though, we need to learn how to recover I from Λ. Proposition 1.1 below gives a duality formula for expressing I in terms of Λ. Recall that $C_b(E)$ denotes the space of bounded continuous functions $F : E \to \mathbb{R}$.

Proposition 1.1 *Let I be a lower semi-continuous function that is bounded below and define $\Lambda : C_b(E) \to \mathbb{R}$ by*

$$\Lambda(F) = \sup_E (F - I).$$

Then

$$I(x) = \sup_{F\in C_b(E)} (F(x) - \Lambda(F)) = \sup_{F\in C_b(E)} \inf_{y\in E} (F(x) - F(y) + I(y)).$$

Proof The inequality

$$\sup_{F\in C_b(E)} \inf_{y\in E}(F(x)-F(y)+I(y)) \le I(x),$$

is immediate because we can choose $y=x$. For the reverse inequality, it suffices to show that if $I(x)<\infty$ and $\varepsilon>0$, then there exists $\bar F \in C_b(E)$ such that $\bar F(x)=I(x)$ and for every $y\in E$

$$\bar F(x)-\bar F(y)+I(y)\ge I(x)-\varepsilon, \quad \text{or equivalently} \quad \bar F(y)\le I(y)+\varepsilon.$$

It is well-known that lower semi-continuous functions can be approximated from below by functions in $C_b(E)$. More precisely, we can find a sequence of functions $F_k\in C_b(E)$ such that $F_k\uparrow I$. In fact the desired $\bar F$ can be selected as $\bar F = F_k - F_k(x)+I(x)$, where k is large enough so that $I(y)-F_k(y)\le\varepsilon$. □

Once an LDP is available, we can formulate several other LDPs that can often be established with ease. Two of such LDPs are stated in Theorem 1.2 below.

Theorem 1.2 *Assume that the family $\{\mathbb{P}_n\}$ satisfies LDP with rate function I.*

- **(i)** *(Contraction Principle) If $\Phi : E\to E'$ is a continuous function, then the family $\{\mathbb{P}'_n\}$, defined by*

$$\mathbb{P}'_n(A) := \mathbb{P}_n\left(\Phi^{-1}(A)\right),$$

satisfies an LDP with rate function $I'(x')=\inf\{I(x)\ :\ \Phi(x)=x'\}$.

- **(ii)** *If $G:E\to\mathbb{R}$ is a bounded continuous function, then the family*

$$d\mathbb{P}_n^G := \frac{1}{Z_n(G)} e^{nG} d\mathbb{P}_n, \ \ \textit{with} \quad Z_n(G)=\int e^{nG}\ d\mathbb{P}_n,$$

satisfies an LDP with the rate function $I^G(x)=I(x)-G(x)+\sup_E(G-I)$.

Proof (i) Given a bounded continuous function $F':E'\to\mathbb{R}$, we use Theorem 1.1 to assert

$$\begin{aligned}\lim_{n\to\infty} n^{-1}\log\int e^{nF'}\ d\mathbb{P}'_n &= \lim_{n\to\infty} n^{-1}\log\int e^{nF'\circ\Phi}\ d\mathbb{P}_n \\ &= \sup_E(F'\circ\Phi - I) = \sup_{x'\in E'}\ \sup_{x:\Phi(x)=x'} (F'\circ\Phi(x)-I(x)) \\ &= \sup_{x'\in E'}(F'(x')-I'(x')).\end{aligned}$$

This and another application of Theorem 1.1 yields the desired result provided that we can verify property **(i)** of Definition 1.1 for I'. This is an immediate consequence of the identity

$$\{x' \in E' \ : \ I'(x') \le \ell\} = \Phi\left(\{x \ : \ I(x) \le \ell\}\right). \tag{1.22}$$

To see (1.22), observe that if $I'(x') \le \ell$, then we can find a sequence $\{x_k\}$ in E such that $\Phi(x_k) = x'$ and $I(x_k) \le \ell + k^{-1}$. Since such a sequence is precompact, it has a convergent subsequence that converges to x with $\Phi(x) = x'$ and $I(x) \le \ell$. This shows that the left-hand side of (1.22) is a subset of the right-hand side. The other direction is obvious.

(ii) Again by Theorem 1.1,

$$\begin{aligned}\lim_{n\to\infty} n^{-1} \log \int e^{nF} \ d\mathbb{P}_n^G &= \lim_{n\to\infty} n^{-1} \log \int e^{n(F+G)} \ d\mathbb{P}_n - \lim_{n\to\infty} n^{-1} \log Z_n(G) \\ &= \sup_E (F + G - I) - \sup_E (G - I) = \sup_E (F - I^G),\end{aligned}$$

for every bounded continuous F. This and another application of Theorem 1.1 yields the desired result provided that we can verify property **(i)** of Definition 1.1 for I^G. Indeed since I is lower semi-continuous and G is continuous, the function I^G is lower semi-continuous. So I^G has closed level sets. On the other hand,

$$\{x \ : \ I^G(x) \le \ell\} \subseteq \{x \ : \ I(x) \le \ell'\},$$

for $\ell' = \ell + \sup_E G + \sup_E (G - I)$. Since the set on the right-hand side is compact, the set on the left-hand side is compact, completing the proof of part **(ii)**. □

Remark 1.1 It is also of interest to establish LDP for the family

$$\mathbb{P}_n^{x',\Phi} := \mathbb{P}_n(\ \cdot \ | \Phi = x').$$

When the family $\{\mathbb{P}_n^{x',\Phi}\}$ is sufficiently regular in x' variable, we expect to have an LDP with rate function

$$I_{x'}^{\Phi}(x) = \begin{cases} I(x) - I'(x'), & \text{if} \quad \Phi(x) = x', \\ \infty & \text{otherwise} \ . \end{cases}$$

(See for example [R].) This and Theorem 1.2**(ii)** may be applied to LDP (1.9) to obtain LDP for (grand canonical and micro canonical) Gibbs measures. Remarkably such LDPs have recently been used by Chaterjee [C] to study long time behavior of solutions of *nonlinear Schrödinger Equation.* □

Exercise 1.1 Suppose that the sequence $\{\mathbb{P}_n\}$ satisfies an LDP with rate function I and let U be an open neighborhood of the set $C_0 = \{x : \ I(x) = 0\}$. Show

$$\lim_{n\to\infty} \mathbb{P}_n(U^c) = 0.$$

□

2 A General Strategy

According to Theorem 1.1, an LDP is equivalent to (1.18) for every $F \in C_b(E)$. Moreover, Proposition 1.1 allows us to express I in terms of Λ. This suggests the following natural strategy for tackling an LDP:

- (i) Find a family of functions $\mathcal{V} \subseteq C_b(E)$ such that we can show

  (2.1) $$\Lambda(V) := \lim_{n\to\infty} n^{-1} \log Z_n(V),$$

  exists for every $V \in \mathcal{V}$, where $Z_n(V) = \int e^{nV}\, d\mathbb{P}_n$.

- (ii) If the family $\mathcal{V}$ is rich enough, then we have LDP with a rate function of the form

  (2.2) $$I(x) = \sup_{V\in\mathcal{V}} (V(x) - \Lambda(V)).$$

Here is how it works in practice:

Theorem 2.1 *Let $\mathcal{V} \subseteq C_b(E)$ be a family of functions such that the limit* (2.1) *exists for every $V \in \mathcal{V}$. Then the following statements are true:*

(i) *For every compact set C,*

(2.3) $$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(C) \le -\inf_C I,$$

where I is defined by (2.2).

(ii) *For every open set U,*

(2.4) $$\liminf_{n\to\infty} n^{-1} \log \mathbb{P}_n(U) \ge -\inf_{U\cap\rho(\mathcal{V})} I,$$

where

(2.5) $$\rho(\mathcal{V}) := \left\{ x : \lim_{n\to\infty} \mathbb{P}_n^V = \delta_x \quad \textit{for some } V \in \mathcal{V} \right\}.$$

(Recall $d\mathbb{P}_n^V = Z_n(V)^{-1}\, e^{nV} d\mathbb{P}_n$.). Moreover if $\lim_{n\to\infty} \mathbb{P}_n^V = \delta_x$ for some $V \in \mathcal{V}$, then $I(x) = V(x) - \Lambda(V)$.

(iii) *If for every $x \in E$ with $I(x) < \infty$, we can find a sequence $\{x_k\} \subseteq \rho(\mathcal{V})$ such that*

(2.6) $$\lim_{k\to\infty} x_k = x, \qquad \limsup_{k\to\infty} I(x_k) \ge I(x),$$

then $\{\mathbb{P}_n\}$ satisfies a weak LDP with rate function I.

Remark 2.1 Theorem 2.1 suggests a precise strategy for establishing a weak LDP principle that consists of two steps:

- A probabilistic step of finding a rich family $\mathcal{V}$ of functions F for which we can calculate the limit in (2.1)
- An analytical step of establishing certain regularity of I as was formulated in (2.6).

$\square$

For some examples, we can establish (2.1) for a family $\mathcal{V}$ that consists of continuous but unbounded functions (like linear functions). For such examples, we state a variant of Theorem 2.1.

Theorem 2.2 *Let $\mathcal{V}$ be a family of continuous functions such that the limit* (2.1) *exists for every $V \in \mathcal{V}$. Assume further*

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \int_{|V'|>\ell} V' \, d\mathbb{P}_n^V = 0, \tag{2.7}$$

for every $V, V' \in \mathcal{V}$. Then the statements **(i)-(iii)** *of Theorem 2.1 are true.*

The proofs of Theorems 2.1-2.2 will be given in Subsection 2.3. The main three steps of the proofs are as follows:

- (i) Use Chebyshev-type inequalities to obtain upper bound LDP (in short ULDP) for compact sets, namely (2.3) for a rate function I_u.
- (iii) Use entropy-type inequalities to establish lower bound LDP (in short LLDP) for open sets, namely (1.14) for a rate function I_l.
- (iv) Show that $I_u(x) = I_l(x)$ for $x \in \rho(\mathcal{V})$.

We explain these three steps in Subsections 2.1-2.3. The ULDP for closed sets will be achieved by showing that the family $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ is *exponentially tight*. That is, for an error that is exponentially small, we can replace a closed set with a (possibly large) compact set. This will be explained in Subsection 2.3.

2.1 Upper bound estimates

Note that if replace V with $G = V - \Lambda(V)$ in (2.1), then the right-hand side becomes 0. Also, for ULDP, we only need an inequality of the form

$$\limsup_{n\to\infty} n^{-1} \log \int e^{nG} \, d\mathbb{P}_n \le 0. \tag{2.8}$$

Theorems 2.1-2.2**(i)** are immediate consequences of the following more general fact:

Proposition 2.1 *Let $\mathcal{G}$ be a family of lower semi-continuous functions such that* (2.8) *holds for every $G \in \mathcal{G}$. Then for every compact set C,*

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(C) \le -\inf_C I_u, \tag{2.9}$$

where I_u is defined by

$$I_u(x) = \sup_{G\in\mathcal{G}} G(x). \tag{2.10}$$

Proof Given any Borel set A, we clearly have

$$e^{\inf_A G}\ \mathbb{P}_n(A) \le \int e^{nG}\ d\mathbb{P}_n.$$

From this, and our assumption (2.8) we deduce

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(A) \le -\inf_A G,$$

for every $G \in \mathcal{G}$. Optimizing this over G yields

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(A) \le -\sup_{G\in\mathcal{G}}\ \inf_{y\in A} G(y). \tag{2.11}$$

It remains to show that in the case of a compact set A, we can interchange sup with inf on the right-hand side of (2.11). To achieve this, take a compact set C and set $\hat{I} = \inf_C I_u$. Given $\varepsilon > 0$, and $x \in C$, we can find $G_x \in \mathcal{G}$ such that

$$\hat{I} - \varepsilon < G_x(x). \tag{2.12}$$

Since G_x is lower semi-continuous, we know that the set

$$U_x = \{y :\ G_x(y) > \hat{I} - \varepsilon\}$$

is open. Since C is compact and $x \in U_x$ by (2.12), we can find $x_1, \dots, x_k \in C$ such that

$$C \subseteq U_{x_1} \cup \dots \cup U_{x_k}.$$

Now we use this and apply (2.11) to $A = U_{x_i}$ to assert

$$\begin{aligned}
\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(C) &\le \limsup_{n\to\infty} n^{-1} \log \left[\mathbb{P}_n(U_{x_1}) + \dots + \mathbb{P}_n(U_{x_k})\right] \\
&\le -\min_{i\in\{1,\dots,k\}}\ \sup_{G\in\mathcal{G}}\ \inf_{y\in U_{x_i}} G(y) \\
&\le -\min_{i\in\{1,\dots,k\}}\ \inf_{y\in U_{x_i}} G_{x_i}(y) \le -\hat{I} + \varepsilon,
\end{aligned}$$

where we used (2.12) for the last inequality. We finally send $\varepsilon \to 0$ to conclude (2.9). $\square$

Remark 2.2 For several important examples (see Example 2.1 below or Sanov's theorem in Section 3) we can achieve this (even with equality in (2.8)) for linear continuous F and the space of such F is rich enough to give us the best possible rate function. More precisely, the space E is a closed subset of a separable Banach space X and $\Lambda(L)$ exists for every continuous linear $L \in X^*$. Using

$$\mathcal{G} := \{L(\cdot) - \Lambda(L) \ : \ L \in X^*\},$$

we deduce an ULDP with the rate function

$$I_u(x) = \Lambda^*(x) = \sup_{L \in X^*} (L(x) - \Lambda(L)),$$

which is the Legendre transform of Λ. We note that in this case I_u is convex and the family $\mathcal{G}$ as above would give us the optimal rate function when the rate function is convex. $\square$

Example 2.1 (Cramér's theorem-Upper bound) Let $X_1, X_2, \dots, X_n, \dots$ be a sequence of iid $\mathbb{R}^d$-valued random variables with distribution μ. Set

$$\mathbb{P}_n(A) = \mathbb{P}\left(n^{-1}(X_1 + \cdots + X_n) \in A\right).$$

Then for every $v \in \mathbb{R}^d$ and $n \geq 1$,

$$n^{-1} \log \int e^{nv\cdot x} \ \mathbb{P}_n(dx) = \log \int e^{v\cdot x} \ \mu(dx) =: \lambda(v)$$

and Theorem 2.1 yields an ULDP for $I_u = \lambda^*$. $\square$

2.2 Lower bound estimates

On the account of Proposition 2.1, we search for a family of functions for which (2.8) holds. The hope is that this family is rich enough so that our candidate I_u for the ULDP also serves as a rate function for the lower bound. To understand how this can be achieved, let us assume that for a given $\bar{x} \in E$, the supremum in (2.10) is achieved at a function $\bar{G} \in \mathcal{G}$. For $\bar{G}$, we better have equality in (2.10); otherwise by adding a small positive constant to $\bar{G}$, (2.10) is still true and such a modified $\bar{G}$ contradicts the optimality of $\bar{G}$. Let us also assume that in fact $\bar{G}$ is a bounded continuous function. Hence, if $I_u = I$ is indeed the large deviation rate function, by Theorem 1.1, we should have

$$\limsup_{n\to\infty} n^{-1} \log \int \ e^{n\bar{G}} \ d\mathbb{P}_n = \sup_E (\bar{G} - I) = 0.$$

Now by applying Theorem 1.2**(ii)**, we learn that the family $\{\mathbb{P}_n^{\bar{G}}\}$ satisfies an LDP with rate function $I^{\bar{G}} = I - \bar{G}$. Note that $I^{\bar{G}}(\bar{x}) = 0$. If $\bar{x}$ is the only point at which $I^{\bar{G}}$ vanishes, then by Exercise 1.1, we have a weak LLN for the family $\{\mathbb{P}_n^{\bar{G}}\}$ of the form

$$\mathbb{P}_n^{\bar{G}} \Rightarrow \delta_{\bar{x}}. \tag{2.13}$$

We may now recast $I_u(x)$ as the entropy cost for producing a probability measure $\mathbb{P}_n^{\bar{G}}$ that concentrates about $\bar{x}$ because

$$\lim_{n\to\infty} n^{-1} H\left(\mathbb{P}_n^{\bar{G}}|\mathbb{P}_n\right) = \lim_{n\to\infty}\left[\int \bar{G}\ d\mathbb{P}_n^{\bar{G}} - n^{-1}\log Z_n(\bar{G})\right] = \bar{G}(\bar{x}) = I_u(\bar{x}). \tag{2.14}$$

In summary a maximizer in the variational problem (2.10) is closely related to those measures $\mathbb{P}_n^{\bar{G}}$ for which a LLN holds. We now design a strategy for lower bounds LDP that is based on (2.13). Namely, we search for those G for which the corresponding $\mathbb{P}_n^G$ satisfies (2.8). Our candidate for $I_l(\bar{x})$ is the smallest "entropy cost" that is needed to achieve (2.6). Motivated by this strategy, let us make a definition.

Definition 2.1 Given $x \in E$, define $\mathcal{C}(x)$ to be the set of measurable functions $G : E \to \mathbb{R}$ such that $\sup_E G < \infty$, and $\mathbb{P}_n^G \Rightarrow \delta_x$ as $n \to \infty$. □

Proposition 2.2 *Let $\mathcal{G}'$ be a family of upper semi-continuous functions such that* $\sup_E G < \infty$*, and*

$$\liminf_{n\to\infty} n^{-1}\log\int e^{nG}\ d\mathbb{P}_n \geq 0, \tag{2.15}$$

holds for every $G \in \mathcal{G}'$. Then for every open set U,

$$\liminf_{n\to\infty} n^{-1}\log\mathbb{P}_n(U) \geq -\inf_U I_l, \tag{2.16}$$

where I_l is defined by

$$I_l(x) = \inf\left\{G(x) :\ G \in \mathcal{G}' \cap \mathcal{C}(x)\right\}. \tag{2.17}$$

As a preparation for the proof of Proposition 2.2, we establish a useful inequality. Given a probability measure $\mathbb{P}$ and a bounded measurable function G let us define

$$\Phi(G) := \log\int e^G\ d\mathbb{P}.$$

The function Φ is convex and its Gâteaux Derivative can be readily calculated. In fact the *subdifferential* of Φ at a function G is a probability measure that is given by

$$d\mathbb{P}_G := e^{G-\Phi(G)}\ d\mathbb{P}.$$

More precisely, for any pair of bounded measurable functions F and G,

$$\Phi(F) - \Phi(G) \geq \int (F - G)\, d\mathbb{P}_G. \tag{2.18}$$

This is an immediate consequence of the Jensen's inequality:

$$e^{\int (F-G)\, d\mathbb{P}_G} \leq \int e^{F-G}\, d\mathbb{P}_G = \int e^{F-\Phi(G)}\, d\mathbb{P} = e^{\Phi(F)-\Phi(G)}.$$

Proof of Proposition 2.2 The main ingredients for the proof are Theorem 1.2**(iii)** and the inequality (2.18). If we set

$$\Lambda_n(G) := n^{-1} \log \int e^{nG}\, d\mathbb{P}_n, \qquad d\mathbb{P}_n^G := e^{n(G-\Lambda_n(G))}\, d\mathbb{P}_n,$$

then (2.18) reads as

$$\Lambda_n(F) \geq \Lambda_n(G) + \int (F - G)\, d\mathbb{P}_n^G. \tag{2.19}$$

As a result,

$$\begin{aligned} \liminf_{n\to\infty} \Lambda_n(F) &\geq \liminf_{n\to\infty} \Lambda_n(G) + \liminf_{n\to\infty} \int (F - G)\, d\mathbb{P}_n^G \\ &\geq F(x) - G(x), \end{aligned} \tag{2.20}$$

for every $G \in \mathcal{G}' \cap \mathcal{C}(x)$. Optimizing this over G yields

$$\liminf_{n\to\infty} \Lambda_n(F) \geq F(x) - I_l(x).$$

From this and Theorem 1.1**(iii)** we deduce (2.16). □

For Theorem 2.2, we need to prove a variant of Proposition 2.2 that works for unbounded continuous functions:

Proposition 2.3 *Let $\mathcal{G}'$ be as in Proposition 2.2 except that instead of the requirement $\sup_E G < \infty$, we assume*

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \int_{|G|>\ell} G\, d\mathbb{P}_n^V = 0, \tag{2.21}$$

for every $G \in \mathcal{G}'$. Then the conclusion Proposition 2.2 holds true.

The proof is almost identical to the proof of Proposition 2.2. The only place where $\sup_E G < \infty$ was used was in the second inequality of (2.20). The condition (2.21) allows us to replace G with $G\mathbb{1}(|G| \leq \ell)$ that is upper semi-continuous and bounded for a small error.

2.3 Proofs of Theorems 2.1 and 2.2

Part **(i)** is an immediate consequence of Proposition 2.1. As for part **(ii)**, take an open set U and pick $x \in U \cap \rho(\mathcal{V})$. By Propositions 2.2 and 2.3

$$\liminf_{n\to\infty} n^{-1} \log \mathbb{P}_n(U) \le -(\bar{V}(x) - \Lambda(\bar{V})),$$

for a function $\bar{V} \in \mathcal{V}$ such that $\mathbb{P}_n^{\bar{V}} \Rightarrow \delta_x$. To deduce (2.4), we need to show that $I(x) = \bar{V}(x) - \Lambda(\bar{V})$, or equivalently,

$$\bar{V}(x) - \Lambda(\bar{V}) \ge V(x) - \Lambda(V), \tag{2.22}$$

for every $V \in \mathcal{V}$. By (2.19) and (2.1)

$$\liminf_{n\to\infty} \int (V - \bar{V})\ d\mathbb{P}_n^{\bar{V}} \le \liminf_{n\to\infty} [\Lambda_n(V) - \Lambda_n(\bar{V})] = \Lambda(V) - \Lambda(\bar{V}).$$

This implies (2.22) when both V and $\bar{V}$ are bounded because $\mathbb{P}_n^{\bar{V}} \Rightarrow \delta_x$. In the case of unbounded V and $\bar{V}$, use (2.7) to replace $V - \bar{V}$, with

$$(V - \bar{V})\mathbb{1}(|V|, |\bar{V}| \le \ell).$$

Then pass to the limit $n \to \infty$, to deduce (2.22).

Finally for part **(iii)**, if (2.6) is true for a sequence $\{x_k\}$ in $\rho(\mathcal{V})$, we also have

$$\lim_{k\to\infty} I(x_k) = I(x),$$

because I is lower semi-continuous. Now if $x \in U$ and (2.6) is true, then $x_k \in U$ for large k. For such k, part **(ii)** implies

$$\liminf_{n\to\infty} n^{-1} \log \mathbb{P}_n(U) \ge -I(x_k).$$

We then use $I(x_k) \to I(x)$ to deduce

$$\liminf_{n\to\infty} n^{-1} \log \mathbb{P}_n(U) \ge -I(x),$$

as desired. □

2.4 Exponential tightness

Recall that according Prohorov's theorem, a sequence of probability measures $\mathbb{P}_n$ has a convergent subsequence if it is *tight*. That is, for every $\ell > 0$, we can find a compact set K_ℓ such that

$$\sup_n \mathbb{P}_n(E \setminus K_\ell) \le \ell^{-1}.$$

This condition would allow us to restrict $\mathbb{P}_n$'s to a large compact set and use the fact that the space of probability measures on a compact metric space is compact.

In the same manner, we define exponential tightness so that off of a large compact set, probabilities are exponentially small.

Definition 2.1 We say that a sequence of probability measures $\mathbb{P}_n$ is *exponentially tight*, if for every $\ell > 0$, there exists a compact set K_ℓ such that

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(E \setminus K_\ell) \leq -\ell.$$

$\square$

Theorem 2.3 **(i)** *Suppose that the sequence of probability measures $\{\mathbb{P}_n\}$ is exponentially tight, and* (2.3) *is true for every compact set. Then* (2.3) *is also true for every closed set.*

(ii) *If* (2.3) *is true for every closed set for a function I_u with compact level sets, then the sequence of probability measures $\{\mathbb{P}_n\}$ is exponentially tight.*

(iii) *If the sequence of probability measures $\{\mathbb{P}_n\}$ is exponentially tight, and* (2.16) *is true for a lower semi-continuous function I_l, then I_l has compact level sets.*

Proof (i) Let K_ℓ be as in Definition 2.1. Given a closed set C,

$$\begin{aligned}\limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(C) &\leq \limsup_{n\to\infty} n^{-1} \log \left[\mathbb{P}_n(C \cap K_\ell) + \mathbb{P}_n(E \setminus K_\ell)\right] \\ &\leq \max\left\{-\inf_{C\cap K_\ell} I, -\ell\right\} \leq \max\left\{-\inf_{C} I, -\ell\right\}.\end{aligned}$$

We then send $\ell \to \infty$ to deduce (2.3) for C.

Proof (ii) Fix $\ell > 0$. Set $C_\ell = \{x : I(x) \leq \ell + 2\}$ and

$$U_k = \{x : d(x, C_\ell) < k^{-1}\}.$$

By ULDP,

$$\limsup_{n\to\infty} n^{-1}\mathbb{P}_n(U_k^c) \leq -\inf_{U_k^c} I \leq -\inf_{C_\ell^c} I \leq -(\ell + 2).$$

Hence, we can find n_k such that for $n > n_k$,

$$\mathbb{P}_n(U_k^c) \leq e^{-n(\ell+1)} = e^{-n}e^{-n\ell}.$$

Without loss of generality, we may assume that $n_k \geq k$, so that for $n > n_k \geq k$,

$$\mathbb{P}_n(U_k^c) \leq e^{-k}e^{-n\ell}.$$

To get rid of the restriction $n > n_k \geq k$, choose compact sets $C_{1,k}, C_{2,k}, \dots, C_{1,n_k}$ such that

$$\mathbb{P}_j \left(C^c_{j,k} \right) \leq e^{-k} e^{-j\ell},$$

for $j = 1, 2, \dots, n_k$. We now have

$$\mathbb{P}_n \left(U^c_k \cap C^c_{1,k} \cap C^c_{2,k} \cap \cdots \cap C^c_{1,n_k} \right) \leq e^{-k} e^{-n\ell},$$

for every n. As a result,

$$\mathbb{P}_n \left(K^c_\ell \right) \leq e^{-n\ell},$$

for the set

$$K_\ell = \cap_{k=1}^\infty \left[U_k \cup C_{1,k} \cup C_{2,k} \cup \cdots \cup C_{1,n_k} \right].$$

We are done if we can show that K_ℓ is compact. For this it suffices to show that K_ℓ is totally bounded. This is obvious, because for each k, the set

$$U_k \cup C_{1,k} \cup C_{2,k} \cup \cdots \cup C_{1,n_k},$$

can be covered by finitely many balls of radius $1/k$.

Proof (iii) If we apply (2.16) to the open set $U_\ell = K^c_\ell$ with K_ℓ as in the Definition 2.1, we learn

$$\{x : \ I(x) \leq \ell\} \subseteq K_\ell.$$

This implies the compactness of the level sets of I because by lower semi-continuity, these level sets are closed. □

Next result gives us a practical way of verifying exponential tightness.

Lemma 2.1 *The sequence $\{\mathbb{P}_n\}$ is exponentially tight if there exists a function $F : \mathbb{E} \to \mathbb{R}$ such that the set $\{x : F(x) \leq \ell\}$ is compact for every ℓ, and*

$$a := \limsup_{n \to \infty} n^{-1} \log \int e^{nF} \ d\mathbb{P}_n < \infty.$$

Proof By Chebyshev's inequality,

$$\mathbb{P}_n(F > \ell) \leq e^{-n\ell} \int e^{nF} \ d\mathbb{P}_n.$$

Hence

$$\limsup_{n \to \infty} n^{-1} \log \mathbb{P}_n(F > \ell) \leq a - \ell.$$

This implies the desired result because the set $\{F \leq \ell\}$ is compact. □

3 Cramér and Sanov Large Deviation Principles

3.1 Cramér's theorem

In the case of Cramér's LDP, we have a sequence $X_1, X_2, \dots$ of $\mathbb{R}^d$-valued iid random variables with distribution μ. Recall

$$\mathbb{P}_n(A) = \mathbb{P}\left(n^{-1}(X_1 + \cdots + X_n) \in A\right). \tag{3.1}$$

Theorem 3.1 *Assume that $\int e^{x\cdot v}\mu(dx) < \infty$, for every $v \in \mathbb{R}^d$. Then the sequence $\{\mathbb{P}_n\}$ (defined by (3.1)) satisfies LDP with the rate function*

$$I(x) = \lambda^*(x) = \sup_{v\in\mathbb{R}^d}(x\cdot v - \lambda(v)), \tag{3.2}$$

where $\lambda(v) = \log \int e^{x\cdot v}\mu(dx)$.

Proof *Step 1.* To apply Theorem 2.2, we need to come up with a family $\mathcal{V}$ such that $\Lambda(V)$ can be calculated for every $V \in \mathcal{V}$. As we mentioned in Example 2.1, we choose $\mathcal{V} = \{L_v : v \in \mathbb{R}^d\}$ with $L_v(x) = v \cdot x$. It is not hard to see that $\lambda(v) := \Lambda(L_v)$ for every $v \in \mathbb{R}^d$. To complete the proof we need to verify three things: the exponential tightness, (2.6), and (2.7). As a preparation, first observe that since

$$e^{r|x|} \le \sum_{i=1}^{d}\left(e^{drx_i} + e^{-drx_i}\right),$$

we have

$$\int e^{r|x|}\, d\mu < \infty, \tag{3.3}$$

for every $r > 0$. This in particular implies

$$n^{-1}\log\int e^{n|x|}\, d\mathbb{P}_n \le \log\int e^{|x|}\, \mu(dx) < \infty.$$

We then use this to apply Lemma 2.1 in the case of $F(x) = |x|$ to deduce exponential tightness.

Step 2. We now turn to the proofs of (2.6) and (2.7). For this, we need to identify the measures $\mathbb{P}_n^{L_v}$, $v \in \mathbb{R}^d$, and the set $\rho(\mathcal{V})$. Observe

$$\mathbb{P}_n^{L_v}(A) = \mathbb{P}^v\left(n^{-1}(X_1 + \cdots + X_n) \in A\right),$$

where $X_1, X_2, \ldots$ are now iid with law

$$\mu^v(dx) = e^{v\cdot x-\lambda(v)}\ \mu(dx).$$

From this and (3.3), we can readily verify (2.7). As for (2.6), note that by the weak LLN,

$$\mathbb{P}_n^{L_v} \Rightarrow \delta_{m(v)}.$$

where $m(v) = \int x\ \mu^v(dx)$. Hence

$$\rho(\mathcal{V}) = \{m(v) :\ v \in \mathbb{R}^d\}.$$

*Step 3.*To establish (2.6), we need to understand the function λ better. For one thing, we may use (3.3) to show that the function λ is smooth with $\nabla\lambda(v) = m(v)$. Now it is clear that if for a given x, the supremum in (3.2) is achieved at some v, then $x = \nabla\lambda(v)$ and as a result $x \in \rho(\mathcal{V})$. Also, by Theorem 2.1**(ii)**, we have

$$I(m(v)) = v \cdot m(v) - \lambda(v). \tag{3.4}$$

As we will see in Exercise 3.1 below, in some cases the supremum is not achieved for a finite v. To deal with such a possibility, let us consider a restricted supremum of the form

$$I_k(x) = \sup_{|v|\le k}(x \cdot v - \lambda(v)). \tag{3.5}$$

Fix $x \notin \rho(\mathcal{V})$ with $I(x) < \infty$ and choose v_k such that

$$I_k(x) = x \cdot v_k - \lambda(v_k), \qquad |v_k| \le k.$$

We then set $x_k = m(v_k)$. We claim that the sequence $\{x_k\}$ satisfies the requirements stated in (2.6). To see this, observe that since v_k is a maximizer in the variational problem (3.4), we must have $|v_k| = k$ for each $k > 0$; otherwise $x = \nabla\lambda(v_k)$ which contradicts $x \notin \rho(\mathcal{V})$. as a result, $|v_k| = k$, and if $w \cdot v_k \ge 0$, then $(x - \nabla\lambda(v_k)) \cdot w \ge 0$. So, there exists some scalar $t_k \ge 0$ such that

$$x - \nabla\lambda(v_k) = t_k v_k.$$

So,

$$x = x_k + t_k v_k, \qquad I_k(x) = (\nabla\lambda(v_k) + t_k v_k) \cdot v_k - \lambda(v_k) = I(m(v_k)) + t_k|v_k|^2 \to I(x).$$

From this we learn that $\{t_k|v_k|^2\}$ is bounded because $I(x) < \infty$. This in turn implies $\lim_{k\to\infty} x_k = x$. This completes the proof of (2.6) because

$$I(m(v_k)) \le I_k(x) \to I(x),$$

as $k \to \infty$. □

Exercise 3.1

(i) Show that $\lambda(v)$ grows at most linearly as $|v| \to \infty$ if the support of μ is bounded.

(ii) Show

$$\lim_{|v|\to\infty} \frac{\lambda(v)}{|v|} = \infty,$$

if the support of μ is $\mathbb{R}^d$ (i.e. $\mu(U) > 0$ for every nonempty open set U). (Hint: Use

$$\lambda(v) \geq \log \int_{A_R} e^{x\cdot v} \mu(dx),$$

where the set $A_R = \{x : 2x \cdot v \geq |v||x|,\ |x| > R\}$.)

(iii) Show

$$D^2\lambda(v) = \int (x - m(v)) \otimes (x - m(v))\ \mu^v(dx).$$

Use this identity to show that λ is strictly convex unless the measure μ is concentrated on a linear subset of $\mathbb{R}^d$ of codimension one. □

3.2 Sanov's theorem

For our next LDP, let μ be a probability measure on a Polish space E and assume that $X_1, X_2, \dots$ is a sequence of E-values iid random variables with $\mathbb{P}(X_i \in A) = \mu(A)$ for every Borel set A. Write $\mathcal{M} = \mathcal{M}(E)$ for the space of Radon probability measures on E and $\|f\| = \sup_E |f|$ for the uniform norm. Equip the space $\mathcal{M}$ with the topology of weak convergence. As it is well-known we may use Wasserstein distance $\mathcal{D}$ to metrize $\mathcal{M}$. Moreover the metric space $(\mathcal{M}, \mathcal{D})$ is a Polish space. We now define a family of probability measures $\mathcal{P}_n$ on $\mathcal{M}$ by

$$\mathbb{P}_n(A) = \mathbb{P}\left(n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n}) \in A\right) \tag{3.6}$$

for every Borel set $A \subseteq \mathcal{M}$. Recall that the relative entropy is defined by

$$H(\alpha|\beta) = \begin{cases} \int \log \frac{d\alpha}{d\beta}\ d\alpha & \text{if } \alpha \ll \beta, \\ \infty & \text{otherwise,} \end{cases}$$

for α and $\beta \in \mathcal{M}$. We are now ready to state and prove Sanov's theorem.

Theorem 3.2 *The family $\{\mathbb{P}_n\}$, defined by* (3.6) *satisfies LDP with a rate function $I(\nu) = H(\nu|\mu)$.*

As a preparation for the proof of Theorem 3.2, we establish the following variational expression of Donsker and Varadhan that in essence identifies the Legendre transform of the entropy.

Theorem 3.3 *We have*

$$H(\nu|\mu) = \sup_{f\in C_b(E)} \left(\int f \, d\nu - \lambda(f) \right) = \sup_{f\in B_b(E)} \left(\int f \, d\nu - \lambda(f) \right), \tag{3.7}$$

where $B_b(E)$ denotes the space of bounded Borel-measuable functions $f : E \to \mathbb{R}$.

Proof *Step 1.* Let us write

$$I(\nu) = \sup_{f\in C_b(E)} \left(\int f \, d\nu - \lambda(f) \right), \quad I'(\nu) = \sup_{f\in B_b(E)} \left(\int f \, d\nu - \lambda(f) \right). \tag{3.8}$$

We first to show that $H(\nu|\mu) \geq I'(\nu)$. Indeed, if $H(\nu|\mu) < \infty$, then for some Borel function $h \geq 0$ we can write $d\nu = h \, d\mu$, or in short $\nu = h\mu$. We then use Young's inequality to assert that for any $f \in B_b(E)$,

$$\begin{aligned}\int f \, d\nu - H(\nu|\mu) &= \int h(f - \log h) \, d\mu = \log \exp \left[\int h \left(f - \mathbb{1}(h > 0) \log h\right) \, d\mu \right] \\ &\leq \log \left[\int h \exp \left(f - \mathbb{1}(h > 0) \log h\right) \, d\mu \right] = \log \int e^f \, d\mu,\end{aligned}$$

as desired.

Step 2. We now turn to the proof of $I = I'$. To prove this, assume that $I(\nu) < \infty$ and pick any $f \in B_b(E)$. We wish to show

$$\int f \, d\nu \leq \lambda(f) + I(\nu). \tag{3.9}$$

Recall that by Lusin's theorem, for any $\varepsilon > 0$, there exists a continuous function f_ε such that

$$\|f_\varepsilon\| \leq \|f\|, \quad (\mu + \nu)(A_\varepsilon) := (\mu + \nu) \left(\{x : f(x) \neq f_\varepsilon(x)\}\right) \leq \varepsilon.$$

From this and

$$\int f_\varepsilon \, d\nu \leq \lambda(f_\varepsilon) + I(\nu),$$

we deduce

$$\int f \, d\nu - 2\varepsilon\|f\| \leq \log \left[\int e^f \, d\mu + 2\varepsilon e^{\|f\|} \right] + I(\nu),$$

for every $\varepsilon > 0$. We then send ε to 0 to deduce (3.9), which in turn implies $I' \le I$, and hence $I = I'$.

Step 3. We next turn to the proof of

$$H(\nu|\mu) \le I'(\nu). \tag{3.10}$$

For this, we first show that if $I'(\nu) < \infty$, then $\nu \ll \mu$. Indeed if $\mu(A) = 0$ for a Borel set A, then choose $f(x) := \ell 1\!\!1(x \in A)$ in (3.9) to assert that $\ell\nu(A) \le I'(\nu)$. Since $\ell > 0$ is arbitrary, we deduce that $\nu(A) = 0$.

We note that if the supremum in the definition of $I'(\nu)$ is achieved for $\bar f \in B_b(E)$, then we must have $\nu = e^{\bar f - \lambda(\bar f)}\mu$. Equivalently, if $\nu = h\mu$, then $\bar f = \log h$. However, in general, $\log h$ is not bounded. Because of this, let us pick some ℓ and ε with $0 < \varepsilon < \ell$ and define $h_\ell = \min\{h, \ell\}$, and

$$h_{\ell,\varepsilon} = \begin{cases} h & \text{if } \ h \in (\varepsilon, \ell), \\ \ell & \text{if } \ h \ge \ell, \\ \varepsilon & \text{if } \ h \le \varepsilon. \end{cases}$$

We then choose $f = \log h_{\ell,\varepsilon}$ in (3.9) to assert that for $\nu = h\mu$,

$$\int (\log h_{\ell,\varepsilon}) \ d\nu \le I(\nu) + \log \int h_{\ell,\varepsilon} \ d\mu. \tag{3.11}$$

Since $h_{\ell,\varepsilon} \downarrow h_\ell$ and $\log h_\ell \le \log \ell$, we may use Monotone Convergence Theorem to send $\varepsilon \to 0$ in (3.11) to deduce

$$\int (\log h_\ell) \ d\nu \le I(\nu) + \log \int h_\ell \ d\mu \le I(\nu).$$

we now send $\ell \to \infty$ to conclude (3.10). □

Exercise 3.2 Show

$$\lambda(f) = \log \int e^f \ d\mu = \sup_{\nu \in \mathcal{M}} \left(\int f \ d\nu - H(\nu|\mu) \right).$$

□

Proof of Theorem 3.2 Given $f \in C_b(E)$, define $L_f(\nu) = \int f \ d\nu$ and set $\mathcal{V} = \{L_f : \ f \in C_b(E)\}$. We have

$$n^{-1} \log \int e^{nL_f} \ dP_n = \log \int e^f \ d\mu =: \lambda(f),$$

which implies that $\Lambda(L_f) = \lambda(f)$. Theorem 2.1 is applicable to the family $\mathcal{V}$ with the associated rate function given by

$$I(\nu) = \sup_{f \in C_b(E)} \left(\int f \ d\nu - \lambda(f) \right). \tag{3.12}$$

By Theorem 3.3, we know that $I(\nu) = H(\nu|\mu)$. We need to take care of three things: identify $\rho(\mathcal{V})$, verify (2.6), and the exponential tightness of the family $\{\mathbb{P}_n\}$.

Step 1. As for $\rho(\mathcal{V})$, let us write

$$\nu_n := n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n}), \quad \mathbb{Q}_n^f := \mathbb{P}_n^{L_f}, \quad \text{and} \quad d\mu^f := e^{f-\lambda(f)}\ d\mu.$$

Observe that for any $J \in C_b(\mathcal{M})$,

$$\begin{aligned}\int J\ d\mathbb{Q}_n^f &= \mathbb{E}\ J(\nu_n)\ e^{f(X_1)+\cdots+f(X_n)-n\lambda(f)} \\ &= \int J\left(n^{-1}(\delta_{x_1} + \cdots + \delta_{x_n})\right)\ \mu^f(dx_1)\ldots\mu^f(dx_n),\end{aligned}$$

which means that the probability measure $\mathbb{Q}_n^f$ is the law of $n^{-1}(\delta_{X_1^f} + \cdots + \delta_{X_n^f})$ where now $X_1^f, X_2^f, \ldots$ are iid with $\mathbb{P}(X_i^f \in A) = \mu^f(A)$. By LLN,

$$\mathbb{P}\left(\left\{\omega : \lim_{n\to\infty} n^{-1}(\delta_{X_1^f(\omega)} + \cdots + \delta_{X_n^f(\omega)}) = \mu^f\right\}\right) = 1.$$

From this we can readily deduce

$$\lim_{n\to\infty} \int J\ d\mathbb{Q}_n^f = J(\mu^f),$$

for every $f \in C_b(\mathcal{M})$. Equivalently

$$\lim_{n\to\infty} \mathbb{Q}_n^f = \delta_{\mu^f}.$$

Hence

$$\rho(\mathcal{V}) = \left\{\mu^f :\ f \in C_b(E)\right\} = \{h\mu \in \mathcal{M} :\ h, h^{-1} \in C_b(E)\}. \tag{3.13}$$

Here by a measure $\nu = h\mu \in \mathcal{M}$, we really mean that the measure $\nu \ll \mu$ and $d\nu/d\mu = h$. So, when $\nu = h\mu \in \rho(\mathcal{V})$, in really means that $h = e^f$ for some $f \in C_b(E)$. By Theorem 2.1**(iii)**,

$$I(h\mu) = \int L_{\log h}\ d\mathbb{Q}^{\log h} - \lambda(\log h) = \int h\log h\ d\mu = H(h\mu|\mu), \tag{3.14}$$

for every function $h > 0$ such that $h, h^{-1} \in C_b(E)$.

Step 2. We now verify (2.6). First observe that if $\phi(h) = h\log h - h + 1$, then $\phi \geq 0$ and $H(h\mu|\mu) = \int \phi(h)\ d\mu$. We now set $h^\ell = h\mathbb{1}(\ell^{-1} \leq h \leq \ell)$ and observe

$$H(h\mu|\mu) = \lim_{\ell\to\infty} \int \phi(h^\ell)\ d\mu.$$

We then pick $\varepsilon > 0$ and use Lusin's theorem to find a continuous function $h^{\varepsilon,\ell}$ such that $h^{\ell,\varepsilon} : E \to [\ell^{-1}, \ell]$, and

$$\mu\left(\left\{x : \ h^{\ell,\varepsilon} \neq h^{\ell}\right\}\right) \leq \varepsilon.$$

We now have

$$\left|\int \phi\left(h^{\ell,\varepsilon}\right) \ d\mu - \int \phi\left(h^{\ell}\right) \ d\mu\right| \leq \varepsilon \ell \log \ell. \tag{3.15}$$

Note that the right-hand side is small for large ℓ if we choose $\varepsilon = \ell^{-2}$. To turn $h^{\ell,\varepsilon}$ into a probability density, set $z^{\ell,\varepsilon} = \int h^{\ell,\varepsilon} \ d\mu$ and $\hat{h}^{\ell,\varepsilon} = h^{\ell,\varepsilon}/z^{\ell,\varepsilon}$. We note

$$\left|z^{\ell,\varepsilon} - 1\right| \leq \varepsilon \log \ell + \nu\left(\{x : \ h(x) \notin (\ell^{-1}, \ell)\}\right). \tag{3.16}$$

We also have

$$\left|\int \phi\left(\hat{h}^{\ell,\varepsilon}\right) \ d\mu - \int \phi\left(h^{\ell,\varepsilon}\right) \ d\mu\right| = \left(1 - \frac{1}{z^{\ell,\varepsilon}}\right) \int \phi\left(h^{\ell,\varepsilon}\right) \ d\mu + \frac{1}{z^{\ell,\varepsilon}} + \log z^{\ell,\varepsilon} - 1.$$

From this, (3.16) and (3.15), we can readily deduce that if $h_\ell = \hat{h}^{\ell,\ell^{-2}}$, then

$$\lim_{\ell \to \infty} H(h_\ell \mu | \mu) = H(h\mu | \mu).$$

Since h_ℓ is continuous and $h_\ell \in [\ell^{-1}, \ell]$, we have established (2.6).

Step 3. For the exponential tightness, pick a sequence of compact sets $K_\ell \subseteq E$ such that

$$\mu(K_\ell^c) \leq e^{-\ell^2}, \tag{3.17}$$

for $\ell = 1, 2, \dots$, and form $\mathcal{K}_\ell = \cap_{k=\ell}^{\infty} \mathcal{K}^k$, where

$$\mathcal{K}^k := \{\nu \in \mathcal{M} : \ \nu(K_k) \geq 1 - k^{-1}\}.$$

Evidently each $\mathcal{K}^k$ is closed and by Prohorov's theorem, each $\mathcal{K}_\ell$ is compact in $\mathcal{M}$. We then apply Chebyshev's inequality to assert

$$\begin{aligned}
\mathbb{P}_n(\mathcal{K}_\ell^c) &\leq \sum_{k=\ell}^{\infty} \mathbb{P}_n((\mathcal{K}^k)^c) = \sum_{k=\ell}^{\infty} \mathbb{P}(\nu_n \notin \mathcal{K}^k) = \sum_{k=\ell}^{\infty} \mathbb{P}(\nu_n(K_k^c) \geq k^{-1}) \\
&= \sum_{k=\ell}^{\infty} \mathbb{P}(nk^2 \nu_n(K_k^c) \geq k) \leq \sum_{k=\ell}^{\infty} e^{-kn} \int e^{nk^2 \nu_n(K_k^c)} \ d\mathbb{P} \\
&= \sum_{k=\ell}^{\infty} e^{-kn} \left(\mu(K_k) + e^{k^2} \mu(K_k^c)\right)^n \leq 2^n \sum_{k=\ell}^{\infty} e^{-kn} \leq 2^{n+1} e^{-n\ell},
\end{aligned}$$

where we used (3.17) for the second inequality. This readily implies the exponential tightness. $\square$

Remark 3.1 In view of Theorem 2.2, the exponential tightness and ULDP imply that the large deviation rate function has compact level sets. This however can be readily verified for $H(\nu|\mu)$ because the function $h \mapsto h \log h$ grows faster than linearly at infinity. Indeed the set

$$\mathcal{K}_\ell = \{h : \ H(h\mu|\mu) \le \ell\} = \{h : \ \int h \log h \ d\mu \le \ell\},$$

is a weakly compact subset of $L^1(\mu)$. This means that the family $\mathcal{K}$ is uniformly integrable. This can be established directly and is an immediate consequence of the following bound: For any Borel set A,

$$\nu(A) \le \frac{\log 2}{\log\left(\frac{1}{\mu(A)} + 1\right)} + \frac{H(\nu|\mu)}{\log \frac{1}{\mu(A)}}. \tag{3.18}$$

Here how the proof goes; for $k \ge 1$,

$$\int_{h>k} h \ d\mu \le (\log k)^{-1} \int h \log h \ d\mu.$$

So, for any set Borel set A, and $\nu = h\mu$,

$$\nu(A) \le (\log k)^{-1} \int h \log h \ d\mu + \int_A h \mathbb{1}(h \le k) \ d\mu \le (\log k)^{-1} H(\nu|\mu) + k\mu(A).$$

This implies (3.18) by choosing

$$k = \frac{\log 2}{\mu(A) \log\left(\frac{1}{\mu(A)} + 1\right)}.$$

$\square$

Exercise 3.3 Use (3.7) to deduce

$$\nu(A) \le \frac{H(\nu|\mu) + \log 2}{\log\left(\frac{1}{\mu(A)} + 1\right)}. \tag{3.19}$$

Exercise 3.4 (i) Given a Polish space E and $\mu \in \mathcal{M}(E)$, define $\hat{I} : \mathcal{M}([0,1] \times E) \to [0,\infty]$ by

$$\hat{I}(\nu) = \sup_{f \in C_b([0,1]\times E)} \left(\int f \ d\nu - \hat{\Lambda}(f) \right),$$

where

$$\hat{\Lambda}(f) = \int_0^1 \lambda(f(\theta,\cdot))\ d\theta = \int_0^1 \log\left[\int e^{f(\theta,x)}\ \mu(dx)\right]\ d\theta.$$

Show that if $\hat{I}(\nu) < \infty$, then $\nu \ll \hat{\mu}$, where $d\hat{\mu} = d\theta \times d\mu$. Moreover, in the definition of $\hat{I}$, we can take the supremum over $B_b([0,1] \times E)$.

(ii) Show that if $\hat{I}(\nu) < \infty$, then we can write $\nu(d\theta, dx) = h(\theta, x)\ d\theta\mu(dx)$ with $\int h(\theta,x)\mu(dx) = 1$ for every $\theta \in [0,1]$ and that

$$\hat{I}(\nu) = \int h \log h\ d\hat{\mu}.$$

(iii) Given a sequence of E-valued iid random variables $X = (X_1, X_2, \dots)$ with the distribution μ, define

$$\hat{\nu}_n(d\theta, dx; X) = n^{-1}\sum_{i=1}^n \delta_{(i/n, X_i)}.$$

The map $X \mapsto \hat{\nu}_n(\cdot; X)$ induces a probability measure $\hat{\mathbb{P}}_n$ on $\hat{\mathcal{M}} = \mathcal{M}([0,1]\times E)$. Show that the family $\{\hat{\mathbb{P}}_n\}$ satisfies LDP principle with the rate function $\hat{I}$. □

3.3 Sanov's theorem implies Cramér's theorem

We may apply Contraction Principle (Theorem 1.2**(i)**) to Sanov's theorem to establish a Cramér's theorem. For simplicity, first we assume that measure μ has a bounded support $\{x :\ |x| \le k\}$. To deduce Theorem 3.1 from Theorem 3.2 in this case, we choose $E = \{x :\ |x| \le k\}$ and consider $\Phi : \mathcal{M}(E) \to \mathbb{R}^d$, defined by $\Phi(\nu) = \int x\ d\nu$. Note that Φ is a continuous function and that if $\mathbb{P}_n$ is defined as in Theorem 3.2, then

$$\mathbb{P}'_n(A) := \mathbb{P}_n\left(\{\nu :\ \Phi(\nu) \in A\}\right) = \mathbb{P}\left(n^{-1}(X_1 + \dots + X_n) \in A\right).$$

As a result, we may apply Theorem 3.2 and Theorem 1.2**(i)** to assert that the family $\{\mathbb{P}'_n\}$ satisfies an LDP with the rate

$$I'(m) = \inf_{\nu\in\mathcal{M}}\{H(\nu|\mu) :\ \Phi(\nu) = m\}. \tag{3.20}$$

This does not immediately imply Cramér's theorem for the sequence $\{\mathbb{P}'_n\}$ because Theorem 3.1 suggests a large deviation rate function of the form

$$\hat{I}(m) = \sup_{v\in\mathbb{R}^d}(m\cdot v - \hat{\lambda}(v)), \tag{3.21}$$

where $\hat{\lambda}(v) = \log\int e^{x\cdot v}\ \mu(dx)$. To prove Theorem 3.1 when the support of μ is bounded, we need to show that $I' = \hat{I}$. We offer two different proofs for this.

The first proof is based on *Minimax Principle*. Indeed

$$\begin{aligned} I'(m) &= \inf_{\nu\in\mathcal{M}} \{H(\nu|\mu) :\ \Phi(\nu) = m\} \\ &= \inf_{\nu\in\mathcal{M}} \sup_{v\in\mathbb{R}^d} \left\{ H(\nu|\mu) - v\cdot\left[\int x\ \nu(dx) - m\right]\right\} \\ &= \inf_{\nu\in\mathcal{M}} \sup_{v\in\mathbb{R}^d} \left\{ H(\nu|\mu) - \int v\cdot x\ \nu(dx) + v\cdot m\right\}, \end{aligned} \tag{3.22}$$

simply because if $\Phi(\nu) - m \neq 0$, then the supremum over v is $+\infty$. Since E is compact in our case, the space $\mathcal{M} = \mathcal{M}(E)$ is also compact. As a result, the Minimax Principle is applicable; the supremum and the infimum in (3.22) can be interchanged. This would yield the equality of I' with $\hat{I}$:

$$\begin{aligned} I'(m) &= \sup_{v\in\mathbb{R}^d} \inf_{\nu\in\mathcal{M}} \left\{ H(\nu|\mu) - \int v\cdot x\ \nu(dx) + v\cdot m\right\} \\ &= \sup_{v\in\mathbb{R}^d} \left\{ -\log\int e^{v\cdot x}\ \mu(dx) + v\cdot m\right\} = \hat{I}(m), \end{aligned}$$

where we used Exercise 3.2 for the second equality.

Our second proof is more direct, though we need to borrow Step 3 from the proof of Theorem 3.1. First observe that the proof of $I' \geq \hat{I}$ is straight forward: for any measure $\nu = h\mu$ satisfying $\Phi(\nu) = \int \Psi\ d\nu = m$,

$$\begin{aligned} H(\nu|\mu) &= \sup_{f\in C_b(E)} \left\{ \int f\ d\nu - \log\int e^f\ d\nu\right\} \\ &\geq \sup_{v\in\mathbb{R}^d} \left\{ \int v\cdot\Psi\ d\nu - \log\int e^{v\cdot\Psi}\ d\nu\right\} \\ &= \sup_{v\in\mathbb{R}^d} (m\cdot v - \hat{\lambda}(v)) = \hat{I}(m). \end{aligned}$$

This yields an ULDP with rate $\hat{I}$. (This not surprising, even in the Minimax Principle, the inequality $\inf\sup \geq \sup\inf$ is trivial.) As for the reverse inequality, note that for I' we are minimizing the entropy of ν (relative to μ) with the constraints

$$\int h\ d\mu = 1, \qquad \int xh(x)\ \nu(dx) = m.$$

Since $\partial H(h\mu|\mu) = \log h$, we may use the method of Lagrange multipliers to assert that for a minimizing $\bar{h}$ of the optimization problem (3.20), we can find scalar $\bar{\lambda}$ and vector $\bar{v}$ such that

$$\log\bar{h}(x) = \bar{v}\cdot x - \bar{\lambda} \quad \text{or} \quad \bar{h}(x) = e^{v\cdot x - \bar{\lambda}}. \tag{3.23}$$

In fact since $\int \bar{h}\ d\mu = 1$, we must have $\bar{\lambda} = \hat{\lambda}(\bar{v}) = \log \int e^{\bar{v}\cdot x}\ \mu(dx)$. This suggests that the minimizing $\bar{h}$ should be of the form (3.23). Though when such minimizing $\bar{h}$ exists, we must have

$$m = \int x\ \nu(dx) = \int x e^{v\cdot x - \hat{\lambda}(v)}\ \nu(dx) = \nabla \hat{\lambda}(v).$$

As a result,

$$I'(\nabla \hat{\lambda}(v)) \le H\left(e^{v\cdot \Psi - \hat{\lambda}(v)}\mu|\mu\right) = \int [v\cdot \Psi - \hat{\lambda}(v)]\ d\mu = m\cdot v - \hat{\lambda}(v).$$

This means that $I'(m) \le \hat{I}(m)$ provided that $m = \nabla\hat{\lambda}(v)$ for some $v \in \mathbb{R}^d$. We may deduce an LDP for $\{\mathbb{P}'_n\}$ with rate function $\hat{I}$ if we can show that whenever $\hat{I}(m) < \infty$, we can find a sequence of $m_k = \nabla\hat{\lambda}(v_k)$ such that $m_k \to m$ and $\hat{\lambda}(m_k) \to \hat{\lambda}(m)$ in large k limit. This is exactly the property (2.6) and was discussed in Step 3 of the proof of Theorem 3.1.

Note that when the support of μ is unbounded, we cannot apply Contraction Principle because when $E = \mathbb{R}^d$, the transformation $\Phi : \mathcal{M} \to \mathbb{R}^d$, defined by $\Phi(\mu) = \int x\ d\nu$ is not continuous. Though this issue can be taken care of with some additional work. We leave the details to Exercise 3.4 below.

Exercise 3.5 (i) Suppose that E and E' are two Polish spaces and $\Phi : E \to E'$ is a Borel function. Assume that there are compact subsets $K_\ell \subset E$ such that the restriction of Φ to each K_ℓ is continuous. Let $\{\mathbb{P}_n\}$ be a sequence of probability measures on E that satisfies LDP with rate function I. If

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(K_\ell^c) = -\infty,$$

then the sequence $\{\mathbb{P}_n \circ \Phi^{-1}\}$ satisfies LDP with rate $I'(x') = \inf_{\Phi(x)=x'} I(x)$.

(ii) Show that if $E = \mathbb{R}^d$, and μ satisfies $\int e^{x\cdot v}\ d\mu < \infty$ for every $v \in \mathbb{R}^d$, then there exists an increasing function $\tau : [0,\infty) \to [0,\infty)$ such that $\tau(0) = 0$, $r^{-1}\tau(r) \to \infty$ in large r limit, and $\int e^{\tau(|x|)}\ \mu(dx) < \infty$.

(iii) Apply part **(i)** and **(ii)** when $\{\mathbb{P}_n\}$ is as in Theorem 3.2, $E = \mathbb{R}^d$, μ satisfies $\int e^{x\cdot v}\ d\mu < \infty$ for every $v \in \mathbb{R}^d$, $\Phi(\mu) = \int x\ d\mu$, and

$$K_\ell = \left\{\nu \in \mathcal{M}(\mathbb{R}^d) :\ \int \tau(|x|)\ d\nu \le \ell\right\}.$$

□

4 Donsker-Varadhan Theory

Let $X_1, X_2, \dots$ be a sequence of iid random variables. By Sanov's theorem, we have an LDP for the empirical measure ν_n associated with this sequence. As a consequence we have a LDP for the law of the sequence $n^{-1}(\Psi(X_1) + \dots + \Psi(X_n))$ provided that $\Psi \in C_b(E)$. However, Sanov's theorem is not strong enough to yield an LDP for the sequence

$$Z_n(\Psi) := n^{-1} \left(\Psi(X_1, X_2) + \Psi(X_2, X_3) + \dots + \Psi(X_n, X_{n+1}) \right),$$

where now $\Psi : E \times E \to \mathbb{R}$ is a bounded continuous function. We note that by LLN

$$\lim_{n\to\infty} Z_n = \int \Psi(x, y)\ \mu(dx)\mu(dy),$$

because if $X_i' = (X_i, X_{i+1})$, then $X_1', X_3', \dots$ and $X_2', X_4', \dots$ are two sequences of iid random variables. In the same fashion we can show that for any bounded continuous $\Psi : E^k \to \mathbb{R}$,

$$\lim_{n\to\infty} Z_n(\Psi) = \int \Psi(x_1, \dots, x_k)\ \mu(dx_1) \dots \mu(dx_k), \tag{4.1}$$

where

$$Z_n(\Psi) = n^{-1} \left(\Psi(X_1, \dots, X_k) + \Psi(X_2, \dots, X_{k+1}) + \dots + \Psi(X_n, \dots, X_{n+k}) \right).$$

Equivalently

$$\lim_{n\to\infty} n^{-1} \left(\delta_{(X_1,\dots,X_k)} + \delta_{(X_2,\dots,X_{k+1})} + \dots + \delta_{(X_n,\dots,X_{n+k})} \right) = \prod_{i=1}^{k} \mu, \tag{4.2}$$

in the weak topology. (Throughout this section all convergence of measures are in weak topology.) It is the best to rephrase (4.2) as an Ergodic Theorem. More precisely, let us consider

$$\mathcal{E} := E^{\mathbb{Z}} = \{\mathbf{x} = (x_n : n \in \mathbb{Z}) :\ x_n \in E \text{ for each } \ n \in \mathbb{Z}\},$$

and define $T(\mathbf{x})_n = x_{n+1}$ to be the shift operator. We equip $\mathcal{E}$ with the product topology so that $\mathcal{E}$ is again a Polish space and T is a continuous function. Writing $\mathbb{P}_\mu$ for the product measure $\prod_{n\in\mathbb{Z}} \mu$, we may apply LLN to assert that for any $\Psi \in C_b(\mathcal{E})$,

$$\lim_{n\to\infty} n^{-1} \left(\Psi(\mathbf{x}) + \Psi(T(\mathbf{x})) + \dots + \Psi(T^{n-1}(\mathbf{x})) \right) = \int \Psi\ d\mathbb{P}_\mu. \tag{4.3}$$

This is exactly (4.1) if

$$\Psi \in C_b^{loc}(E) := \{\Psi \in C_b(E) :\ \Psi \text{ depends on finitely many coordinates}\}.$$

The proof of (4.3) for general $\Psi \in C_b(\mathcal{E})$ follows from the denseness of $C_b^{loc}(E)$ in $C_b(E)$. Furthermore, by taking a countable dense set of functions $\Psi \in C_b(\mathcal{E})$, we learn

$$\mathbb{P}_\mu \left\{ \mathbf{x} \in \mathcal{E} : \lim_{n \to \infty} \nu_n(\mathbf{x}) = \mathbb{P}_\mu \right\} = 1, \tag{4.4}$$

where

$$\nu_n(\mathbf{x}) = n^{-1} \left(\delta_{\mathbf{x}} + \delta_{T(\mathbf{x})} + \cdots + \delta_{T^{n-1}(\mathbf{x})} \right).$$

The statement (4.4) is an instance of an Ergodic Theorem for the dynamical system $(\mathcal{E}, T)$. More generally, we may set

$$\mathcal{M}_S = \mathcal{M}_S(\mathcal{E}) := \left\{ Q \in \mathcal{M}(\mathcal{E}) : \int F \circ T \, dQ = \int F \, dQ, \text{ for every } F \in C_b(\mathcal{E}) \right\}.$$

In the theory of Dynamical Systems, the set $\mathcal{M}_S(\mathcal{E})$ consists of *invariant measures* of the dynamical system $(\mathcal{E}, T)$. In probabilistic language, $\mathcal{M}_S$ consists of the laws of E-valued stationary processes. We can readily show that $\mathcal{M}_S(\mathcal{E})$ is a closed convex subspace of $\mathcal{M}(\mathcal{E})$. Hence $\mathcal{M}_S(\mathcal{E})$ is a Polish space. Here are some examples of stationary processes.

Example 4.1
(i) (*iid sequences*) The product measure $\mathbb{P}_\nu \in \mathcal{M}(E)$, for every $\nu \in \mathcal{M}(E)$.

(ii) (*Markovian sequences*) Consider a measurable family $\{p(x, \cdot) : x \in E\}$ of probability measures on E and regard $p(x, dy)$ as a Markov kernel. Suppose that $\pi \in \mathcal{M}(E)$ is an invariant measure for p. That is,

$$\int p(x, A) \, \pi(dx) = \pi(A),$$

for every $A \in \mathcal{B}$. We now build a stationary process Q that is the law of a Markov process with marginal π and kernel p. The measure Q is uniquely determined by identifying its finite dimensional marginals; the law of $(x_k, x_{k+1}, \ldots, x_{k+\ell-1}, x_{k+\ell})$ is given by

$$\pi(dx_k) p(x_k, dx_{k+1}) \ldots p(x_{k+\ell-1}, dx_{k+\ell}).$$

(iii) (*Periodic sequences*) Given $\mathbf{x} \in \mathcal{E}$, define an n-periodic sequence $\Pi_n(\mathbf{x}) = (x_i^n : i \in \mathbb{Z})$ such that $x_{i+kn}^n = x_i$ for $i \in \{1, 2 \ldots, n\}$ and $k \in \mathbb{Z}$. We then define a modified empirical measure as follows:

$$\hat{\nu}_n(\mathbf{x}) = n^{-1} \left(\delta_{\Pi_n(\mathbf{x})} + \delta_{T \circ \Pi_n(\mathbf{x})} + \cdots + \delta_{T^{n-1} \circ \Pi_n(\mathbf{x})} \right). \tag{4.5}$$

Evidently $\hat{\nu}_n(\mathbf{x}) \in \mathcal{M}_S(\mathcal{E})$. □

According to the Birkhoff Ergodic Theorem,

$$Q\left(\left\{\mathbf{x}\in\mathcal{E}:\ \lim_{n\to\infty}\nu_n(\mathbf{x})=Q_{\mathbf{x}}^T\right\}\right)=1, \tag{4.6}$$

where $Q_{\mathbf{x}}^T(d\mathbf{y})$ is a conditional measure of Q with respect to the σ-algebra $\mathcal{I}^T$ of T-invariant Borel sets:

$$\mathcal{F}=\{A: A \text{ is a Borel set and } T(A)=A\}.$$

We also write $\mathcal{M}_{er}$ for the space of ergodic invariant measures. More precisely, $\mathcal{M}_{er}$ consists of those invariant measures Q such that for every $A\in\mathcal{I}^T$, we have that either $Q(A)=1$ or 0. In particular if $Q\in\mathcal{M}_{er}$, then

$$Q\left(\left\{\mathbf{x}\in\mathcal{E}:\ \lim_{n\to\infty}\nu_n(\mathbf{x})=Q\right\}\right)=1. \tag{4.7}$$

Donsker-Varadhan Theory establishes an LDP for the LLN stated in (4.4). To describe this LDP let us recall that any limit point of the empirical measure $\nu_n(\mathbf{x})$ is necessarily in $\mathcal{M}_S(\mathcal{E})$ even though $\nu_n(\mathbf{x})\notin\mathcal{M}_S(\mathcal{E})$ in general (except when $\mathbf{x}$ is n-periodic). This suggests to formulate an LDP for probability measures defined on $\mathcal{M}_S(\mathcal{E})$. To achieve this goal, let us modify our LLN slightly to assert that in fact for $Q\in\mathcal{M}_{er}(\mathcal{E})$,

$$Q\left(\left\{\mathbf{x}\in\mathcal{E}:\ \lim_{n\to\infty}\hat\nu_n(\mathbf{x})=Q\right\}\right)=1, \tag{4.8}$$

where $\hat\nu_n$ was defined in (4.5). This is because for a local function $\Psi=\Psi_k$, with $\Psi_k\in C_b(E^k)$,

$$\left|\int\Psi\ d\nu_n(\mathbf{x})-\int\Psi\ d\hat\nu_n(\mathbf{x})\right|\le 2kn^{-1}\|\Psi\|. \tag{4.9}$$

Since $\hat\nu_n\in\mathcal{M}_S(\mathcal{E})$, let us define a sequence of probability measures $\{\mathcal{P}_n^\mu\}$ on $\mathcal{M}_S(\mathcal{E})$ by

$$\mathcal{P}_n^\mu(A)=\mathbb{P}_\mu\left(\{\mathbf{x}:\ \hat\nu_n(\mathbf{x})\in A\}\right).$$

Donsker-Varadhan [DV] establishes an LDP for the family $\{\mathcal{P}_n^\mu\}$. To facilitate the statement of this LDP and its proof, let us make some useful conventions.

Definition 4.1 (i) Given a σ-algebra $\mathcal{F}$, we abuse the notation and write $\mathcal{F}$ for the space of $\mathcal{F}$-measurable functions. We also write $b\mathcal{F}$ for the space of bounded $\mathcal{F}$-measurable functions.

(ii) Define

$$\mathcal{E}^i=\{\mathbf{x}^i=(x_j:\ j\le i):\ x_j\in E \text{ for } j\le i\},\qquad \mathcal{E}_i=\{\mathbf{x}_i=(x_j:\ j\ge i):\ x_j\in E \text{ for } j\le i\},$$

so that $\mathcal{E}=\mathcal{E}^i\times\mathcal{E}_{i+1}$ and any $\mathbf{x}\in\mathcal{E}$ can be written as $\mathbf{x}=(\mathbf{x}^i,\mathbf{x}_{i+1})$.

(iii) The σ-algebra of Borel sets depending on coordinates $(x_j : \ j \leq i)$ (resp. $(x_j : \ j \geq i)$) is denoted by $\mathcal{B}^i$ (resp. $\mathcal{B}_i$). The σ-algebra of Borel sets depending on coordinates $(x_j : i \leq j \leq k))$ is denoted by $\mathcal{B}_i^k$).

(iv) Given $Q \in \mathcal{M}(\mathcal{E})$, we write $Q_i'(d\mathbf{x}^i)$ for the restriction of Q to the σ-algebra $\mathcal{B}^i$. Further, the Q-conditional measure with respect to $\mathcal{B}^i$ is denoted by $Q_i(\mathbf{x}^i, d\mathbf{y})$. This measure is supported on the set $\{\mathbf{x}^i\} \times \mathcal{E}_{i+1}$. Identifying this support with $\mathcal{E}_{i+1}$, we also write $Q_i(\mathbf{x}^i, d\mathbf{y}_{i+1})$ for the Q-conditional measure but now as a probability measure on $\mathcal{E}_{i+1}$. Given $k > 0$, we also write $Q_i(\mathbf{x}^i, dy_{i+1}, \ldots, dy_{i+k})$ for the restriction of $Q_i(\mathbf{x}^i, d\mathbf{y}_{i+1})$ to the σ-algebra $\mathcal{B}_{i+1}^{i+k}$. Hence for $\Psi \in b\mathcal{B}^{i+k}$,

$$\int \Psi \ dQ = \int \left[\int \Psi(\mathbf{x}^i, y_{i+1}, \ldots, y_{i+k}) \ Q_i(\mathbf{x}^i, d\mathbf{y}_{i+1}) \right] \ Q_i'(d\mathbf{x}^i).$$

(v) We write $\mu^{\otimes k} \in \mathcal{M}(E^k)$ for the product of k copies of μ. When $Q \in \mathcal{M}$ and

$$Q(\mathbf{x}^i, dy_{i+1}, \ldots, dy_{i+k}) \ll \mu^{\otimes k},$$

the corresponding Radon-Nikodym derivative is denoted by $q_i(y_{i+1}, \ldots, y_{i+k} | \mathbf{x}^i)$.

(vi) Set

$$\mathcal{W}^i = \left\{ G \in \mathcal{B}^{i+1} : \ \int e^{G(\mathbf{x}^i, x_{i+1})} \ \mu(dx_{i+1}) = 1 \text{ for every } \mathbf{x}^i \in \mathcal{E}^i \right\}.$$

We also write

$$\mathcal{W}_b^i = \mathcal{W}^i \cap b\mathcal{B}^{i+1}, \qquad \hat{\mathcal{W}}^i = \mathcal{W}^i \cap C_b^{loc}(\mathcal{E}).$$

Note that for any $F \in \mathcal{B}^{i+1}$, the function

$$\hat{F}(\mathbf{x}^i, x_{i+1}) = F(\mathbf{x}^i, x_{i+1}) - \log \int e^{F(\mathbf{x}^i, y_{i+1})} \ \mu(dy_{i+1}),$$

belongs to $\mathcal{W}^i$. Also, if $Q_i(\mathbf{x}^i, \cdot) \ll \mu$ for Q-almost all $\mathbf{x}^i$, then $G(\mathbf{x}^i, x_{i+1}) = \log q_i(x_{i+1} | \mathbf{x}^i) \in \mathcal{W}^i$ after a modification on a Q-null set. $\square$

We are now ready to state Donsker-Varadhan LDP. Recall that $H(\alpha|\beta)$ denotes the relative entropy and was defined write after (3.6).

Theorem 4.1 *The family $\{\mathcal{P}_n^\mu\}$ satisfies an LDP with rate function $\mathcal{H}_\mu : \mathcal{M}_S(\mathcal{E}) \to [0, \infty]$, that is defined by*

$$\mathcal{H}_\mu(Q) = \int H(\hat{Q}_{\mathbf{x}}^1 | \mu) \ Q(d\mathbf{x}), \tag{4.10}$$

where $\hat{Q}_{\mathbf{x}}^1(dy_1) = Q(\mathbf{x}^0, dy_1)$ is the Q-conditional measure of x_1 given $\mathcal{B}^0$.

As a preparation for the proof Theorem 4.1, we find alternative formulas for the rate $\mathcal{H}_\mu(Q)$.

Theorem 4.2 **(i)** *We have*

$$\begin{aligned} \mathcal{H}_\mu(Q) &= \sup_{F\in b\mathcal{B}^{i+1}} \int \left[F(\mathbf{x}^i,x_{i+1}) - \log\int e^{F(\mathbf{x}^i,y_{i+1})}\ \mu(dy_{i+1})\right]\ dQ \\ &= \sup_{G\in\mathcal{W}_b^i}\int G\ dQ = \sup_{G\in\hat{\mathcal{W}}_b^i}\int G\ dQ = H(Q'_{i+1}|Q'_i\times\mu). \end{aligned} \tag{4.11}$$

(ii) *For every $Q\in\mathcal{M}_S(\mathcal{E})$, write $\hat{Q}^k_{\mathbf{x}}$ for $Q_0(\mathbf{x}^0,dy_1,\dots,dy_k)$. Then*

$$\mathcal{H}_\mu(Q) = k^{-1}\int H(\hat{Q}^k_{\mathbf{x}^0}|\mu^{\otimes k})\ Q'_0(d\mathbf{x}^0).$$

(iii) *Let $Q\in\mathcal{M}_{er}(\mathcal{E})$ with $\mathcal{H}_\mu(Q)<\infty$. Then*

$$Q\left(\left\{\mathbf{x}:\ \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \log q_i(x_i|\mathbf{x}^{i-1}) = \mathcal{H}_\mu(Q)\right\}\right) = 1. \tag{4.12}$$

(iv) *There exists a set $\bar{\mathcal{E}}\in\mathcal{B}^0\cap\mathcal{I}^T$ and a measurable map $R:\bar{\mathcal{E}}\to\mathcal{M}(\mathcal{E})$ such that the following properties are true for every $Q\in\mathcal{M}_S(\mathcal{E})$:*

- $Q(\bar{\mathcal{E}})=1$.
- $$Q(d\mathbf{x}) = \int_{\bar{\mathcal{E}}} R(\mathbf{y})(d\mathbf{x})\ Q(d\mathbf{y}), \tag{4.13}$$
- *If we write $\hat{R}(\mathbf{x})(dy_1)$ for the law of the coordinate y_1 with respect to $R(\mathbf{x})$, then*
$$\mathcal{H}_\mu(Q) = \int H\left(\hat{R}(\mathbf{x})|\mu\right)\ Q(d\mathbf{x}). \tag{4.14}$$

In particular, $\mathcal{H}_\mu$ is an affine function.

Proof **(i)** The proof of the second equality in (4.11) is obvious because

$$G(\mathbf{x}^i,x_{i+1}) = F(\mathbf{x}^i,x_{i+1}) - \log\int e^{F(\mathbf{x}^i,y_{i+1})}\ \mu(dy_{i+1}),$$

is always in $\mathcal{W}^i$. On the other hand,

$$\int \left[F(\mathbf{x}^i, x_{i+1}) - \log \int e^{F(\mathbf{x}^i, y_{i+1})} \; \mu(dy_{i+1}) \right] \; dQ,$$

equals

$$\int \left[\int F(\mathbf{x}^i, x_{i+1}) \; Q_i(\mathbf{x}^i, dy_{i+1}) - \log \int e^{F(\mathbf{x}^i, y_{i+1})} \; \mu(dy_{i+1}) \right] \; Q_i'(d\mathbf{x}^i),$$

and this is bounded above by

$$\int H\left(Q_i(\mathbf{x}^i, \cdot)|\mu\right) \; Q_i'(d\mathbf{x}^i) = \mathcal{H}_\mu(Q),$$

by Theorem 3.3. For the reverse inequality, use the concavity of the log-function to assert that (4.11) is bounded below by

$$\begin{aligned}
&\sup_{F \in b\mathcal{B}^{i+1}} \int \left[\int F(\mathbf{x}^i, x_{i+1}) \; Q_i(\mathbf{x}^i, dy_{i+1}) Q_i'(d\mathbf{x}^i) - \log \int e^{F(\mathbf{x}^i, y_{i+1})} \; \mu(dy_{i+1}) Q_i'(d\mathbf{x}^i) \right] \\
&\quad = \sup_{F \in b\mathcal{B}^{i+1}} \int \left[\int F(\mathbf{x}^i, x_{i+1}) \; Q_{i+1}'(d\mathbf{x}^{i+1}) - \log \int e^{F(\mathbf{x}^i, y_{i+1})} \; \mu(dy_{i+1}) Q_i'(d\mathbf{x}^i) \right] \\
&\quad = H(Q_{i+1}'|Q_i' \times \mu),
\end{aligned}$$

where we used Theorem 3.3 for the second equality. We now argue that $H(Q_{i+1}'|Q_i' \times \mu) \geq \mathcal{H}_\mu(Q)$. To prove this, we may assume that the left-hand side is finite which in turn implies that $Q_{i+1}' \ll Q_i' \times \mu$, or there exists a Borel function $h(\mathbf{x}^i, x_{i+1}) \geq 0$ such that

$$Q_{i+1}'(d\mathbf{x}^i, dx_{i+1}) = Q_i(\mathbf{x}^i, dx_{i+1}) Q_i'(d\mathbf{x}^i) = h(\mathbf{x}^i, x_{i+1}) Q_i'(d\mathbf{x}^i) \mu(dx_{i+1}).$$

Using this, we assert that indeed $h(\mathbf{x}^i, \cdot)$ may be regarded as $dQ_i(\mathbf{x}^i, \cdot)/d\mu$. Hence

$$H(Q_{i+1}'|Q_i' \times \mu) = \int \log h(\mathbf{x}^i, x_{i+1}) Q_i'(d\mathbf{x}^i) \mu(dx_{i+1}) = \int H(Q_i(\mathbf{x}^i, \cdot)|\mu) \; Q_i'(d\mathbf{x}^i) = \mathcal{H}_\mu(Q),$$

as desired.

It remains to check that the G-supremum in (4.11) can be restricted to $\hat{\mathcal{W}}_b^i$. We first show that we may restrict the supremum to $C_b(\mathcal{E})$. As in the proof of Theorem 3.3, pick $\varepsilon > 0$ and $G \in \mathcal{W}_b^i$, and use Lusin's theorem to find a function $F \in \mathcal{B}^{i+1} \cap C_b(\mathcal{E})$ such that if

$$A = \left\{ \mathbf{x}^{i+1} : \; G(\mathbf{x}^{i+1}) \neq F(\mathbf{x}^{i+1}) \right\},$$

then

$$\left(Q_{i+1}' + Q_i' \times \mu\right)(A) \leq \varepsilon.$$

We can readily show the expression

$$(4.15)\qquad \left|\int G\ dQ - \int \left[F(\mathbf{x}^i, x_{i+1}) - \log \int e^{F(\mathbf{x}^i, y_{i+1})}\ \mu(dy_{i+1})\right]\ Q_i'(d\mathbf{x}^i)\right|,$$

is bounded above by

$$c_1\varepsilon + \int \log\left(1 + c_1\mu(A(\mathbf{x}^i)\right) Q_i'(d\mathbf{x}^i),$$

where $A(\mathbf{x}^i) = \{x_{i+1} :\ (\mathbf{x}^i, x_{i+1}) \in A\}$, and $c_1 > 0$ is a constant that depends on $\|G\| + \|F\|$ only. Using the elementary bound $\log(1 + a) \le a$, we deduce that the expression (4.11) is bounded by

$$c_1\varepsilon + \int c_1\mu(A(\mathbf{x}^i)Q_i'(d\mathbf{x}^i) \le 2c_1\varepsilon,$$

as desired.

We can readily restrict the $F \in C_b(\mathcal{E})$-supremum in (4.11) to $F \in C_b^{loc}(\mathcal{E})$ by approximation in uniform norm.

(ii) Assume that $\mathcal{H}_\mu(Q) < \infty$. Then $Q_0(\mathbf{x}^0, dx_1) \ll \mu(dx_1)$ and $Q_0(\mathbf{x}^0, dx_1) = q_0(x_1|\mathbf{x}^0)\mu(dy_1)$. Let us write

$$f(\mathbf{x}) := q_0(x_1|\mathbf{x}^0),$$

and regard it as a function in $\mathcal{B}^1$. By stationarity of Q, we have that $Q_i(\mathbf{x}^i, dx_{i+1}) \ll \mu(dx_{i+1})$ and

$$Q_i(\mathbf{x}^i, dx_{i+1}) = q_i(x_{i+1}|\mathbf{x}^i)\ \mu(dx_{i+1}) = (f \circ T^i)(\mathbf{x}),$$

for every positive integer i. By the definition of the conditional measure,

$$\begin{aligned} Q_k'(d\mathbf{x}^k) &= Q_0'(d\mathbf{x}^0)Q_0(\mathbf{x}^0, dx_1)\dots Q_{k-1}(\mathbf{x}^{k-1}, dx_k) \\ &= q_0(x_1|\mathbf{x}^0)\dots q_{k-1}(x_k|\mathbf{x}^{k-1})\ Q_0'(d\mathbf{x}^0)\mu(dx_1)\dots\mu(dx_k) \\ &= \prod_{i=1}^{k} f \circ T^i(\mathbf{x})\ Q_0'(d\mathbf{x}^0)\mu(dx_1)\dots\mu(dx_k). \end{aligned}$$

Hence $\hat{Q}^k_{\mathbf{x}^0} \ll \mu^{\otimes k}$ for Q-almost all $\mathbf{x}$, and

$$\int H\left(\hat{Q}^k_{\mathbf{x}^0}|\mu^{\otimes k}\right) Q_0'(d\mathbf{x}^0) = k\int \log f(\mathbf{x})\ \mu(dx_1)Q_0'(d\mathbf{x}^0) = k\mathcal{H}_\mu(Q),$$

as desired.

(iii) Let f be as in part **(ii)** and set $g = \log f$. If $\mathcal{H}_\mu(Q) < \infty$, then

$$\int |g|\ dQ = \int |g(\mathbf{x})|\ Q(\mathbf{x}^0, dx_1)Q_0'(d\mathbf{x}^0) = \int f(\mathbf{x})|\log f(\mathbf{x})|\ \mu(dx_1)Q_0'(d\mathbf{x}^0) \le 1+\mathcal{H}_\mu(Q) < \infty.$$

Hence, we may apply Ergodic Theorem to assert

$$Q\left(\left\{\mathbf{x}: \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} g(T^i(\mathbf{x})) = \mathcal{H}_\mu(Q)\right\}\right) = 1,$$

proving (4.12).

(iv) *Step 1* By Theorem C.2 of the Appendix, we can find a set $\mathcal{E}_0 \in \mathcal{I}^T$ such that the map

$$R^T(\mathbf{x}) := \lim_{n\to\infty} \nu_n(\mathbf{x}),$$

exists, belongs to $\mathcal{M}_{er}(\mathcal{E})$, and $R^T(\mathbf{x}) = Q$ for Q-almost all $\mathbf{x}$ whenever $Q \in \mathcal{M}_{er}(\mathcal{E})$. In fact we may use (4.9) to replace ν_n with $\hat{\nu}_n$ in the definition of Q^T. Moreover, (4.9) allows us to replace ν_n with $\hat{\nu}_n$ in the definition of $\mathcal{E}_0$ that appears in the proof of Theorem C.2. (The set $\mathcal{E}_0$ is defined to be the set of $\mathbf{x}$ for which the limits in (C.6) and (C.9) below exist for a countable dense set of functions in $U_b(\mathcal{E})$, that can be chosen to consist of local functions.) Since $\hat{\nu}(\mathbf{x})$ depends on $\mathbf{x}_1$ only, the set $\mathcal{E}_1 \in \mathcal{B}_1$. We now replace T with T^{-1} and denote the corresponding $\mathcal{E}_0$ by $\bar{\mathcal{E}}$ that now belongs to $\mathcal{I}^T$ and $\mathcal{B}^0$.

Step 2 We may define a measurable map $\mathcal{F} : \mathcal{M}(\mathcal{E}) \times \mathcal{E} \to \mathcal{M}(\mathcal{E})$ such that $\mathcal{F}(Q, \mathbf{x}) = Q_{\mathbf{x}}$ is the Q-conditional measure, given $\mathcal{B}^0$. We then define R by

$$R(\mathbf{x}) = \mathcal{F}\left(R^T(\mathbf{x}), \mathbf{x}\right).$$

Now if $Q \in \mathcal{M}_{er}(\mathcal{E})$, then using the fact that $Q\big(\{R^T(\mathbf{x}) = Q\}\big) = 1$, we can assert

$$\begin{aligned}\int_{\bar{E}} R(\mathbf{y})(d\mathbf{x})\ Q(d\mathbf{y}) &= \int_{\bar{\mathcal{E}}} \mathcal{F}\left(R^T(\mathbf{y}), \mathbf{y}\right)(d\mathbf{x})\ Q(d\mathbf{y}) = \int_{\bar{\mathcal{E}}} \mathcal{F}\left(Q, \mathbf{y}\right)(d\mathbf{x})\ Q(d\mathbf{y}) \\ &= \int Q_{\mathbf{y}}(d\mathbf{x})\ Q(d\mathbf{y}) = Q(d\mathbf{x}).\end{aligned}$$

Hence (4.15) is true for every ergodic Q. Since R is independent of Q, both sides of (4.15) are linear in Q. Since any $Q \in \mathcal{M}_S(\mathcal{E})$ can be expressed as convex combination of ergodic measures, we have (4.15) for every $Q \in \mathcal{M}_S(\mathcal{E})$.

Step 3 By the definition of R, we know that $R(\mathbf{x})$ is $\mathcal{B}^0$ measurable. From this and (4.15) we learn that $R(\mathbf{x}) = Q_{\mathbf{x}}$ is the Q-conditional measure with respect to $\mathcal{B}^0$ for Q-almost all $\mathbf{x}$. This immediately implies (4.14) by the very definition of $\mathcal{H}_\mu$. □

Proof of Theorem 4.1 *Step 1 (Upper Bound)* Given $G \in C_b(\mathcal{E})$, define $L_G : \mathcal{M} \to \mathbb{R}$ by $L_G(Q) = \int G\ dQ$. The restriction of L_G to the set $\mathcal{M}_S$ is also denoted by L_G. We now set $\mathcal{V} = \{L_G :\ G \in \hat{\mathcal{W}}^1\}$, where $\hat{\mathcal{W}}^1$ was defined in Definition 4.1. We now argue that indeed $\Lambda(L_G) = 0$ for every $G \in \hat{\mathcal{W}}^1$. To see this, first observe that if $G \in \mathcal{W}^1$, then write $G(\mathbf{x}) = G(\mathbf{x}^1)$, and

$$\int e^{n \int G \, d\nu_n} \, d\mathbb{P}_\mu = \int e^{G(\mathbf{x}^0, x_1) + \cdots + G(\mathbf{x}^{n-1}, x_n)} \, d\mathbb{P}_\mu = 1,$$

because $G \in \mathcal{W}$. From this and (4.9) we deduce that if $G \in \hat{\mathcal{W}}^1$ and depends on k many coordinates $x_r, \ldots, x_{r+k-1}$, then

$$e^{-2k\|G\|} \le \int e^{n \int G \, d\hat{\nu}_n} \, d\mathbb{P}_n \le e^{2k\|G\|}.$$

From this, we can readily deduce that

$$\bar{\Lambda}(G) := \Lambda(L_G) = 0,$$

for every $G \in \hat{\mathcal{W}}^1$. In view of Theorem 2.1 and 4.1**(i)**, we have an ULDP for compact sets with the rate function

$$I_u(Q) = \sup_{G \in \hat{\mathcal{W}}^1} \int G \, dQ = \mathcal{H}_\mu(Q).$$

Step 2 (Exponential Tightness) Let us write

$$\alpha_n(\mathbf{x}) = n^{-1} \left(\delta_{x_1} + \cdots + \delta_{x_n} \right).$$

We note that $\alpha_n(\mathbf{x})$ is simply the one-dimensional marginal of $\hat{\nu}_n(\mathbf{x})$; for any $f \in C_b(E)$,

$$\int f(y_1) \; \hat{\nu}_n(\mathbf{x})(d\mathbf{y}) = \int f \; d\alpha_n(\mathbf{x}).$$

The idea is that we already have exponential tightness for the marginals of $\hat{\nu}_n$ by Sanov's theorem. This and stationarity would yield exponential tightness for $\hat{\nu}_n$. To see this, we use the proof of the exponential tightness of the sequence $\{\mathbb{P}_n\}$ of Sanov's theorem (see Step 3 of the proof of Theorem 3.2) to find compact subsets A_k of E such that

$$\mathbb{P}_\mu \left(\left\{ \mathbf{x} : \; \alpha_n(\mathbf{x})(A_k^c) > k^{-1} \text{ for some } k \ge \ell \right\} \right) \le e^{-\ell n}.$$

Let us write $\tilde{Q}$ for the one dimensional marginal of a stationary measure Q. If we set

$$\mathcal{A}_\ell = \left\{ Q \in \mathcal{M}_S(\mathcal{E}) : \; \tilde{Q}(A_k^c) \le k^{-1} \; \text{ for all } k \ge \ell \right\},$$

then

$$\mathcal{P}_n(\mathcal{A}_\ell^c) = \mathbb{P}_\mu \left(\left\{ \mathbf{x} : \; \alpha_n(\mathbf{x})(A_k^c) > k^{-1} \text{ for some } k \ge \ell \right\} \right) \le e^{-\ell n}.$$

We are done if we can show that $\mathcal{A}_\ell$ is a tight subset of $\mathcal{M}_S(\mathcal{E})$. Note that if we set

$$B_\ell = \prod_{i \in \mathbb{Z}} A_{\ell 2^{|i|}},$$

then by Tychonoff's theorem, each B_ℓ is compact in $\mathcal{E}$. Moreover, if we set

$$\mathcal{A}'_\ell = \left\{ Q \in \mathcal{M}_S(\mathcal{E}) : \ Q(B_\ell^c) \leq 3\ell^{-1} \ \text{ for all } \ell \in \mathbb{N} \right\},$$

then by Prohorov's theorem the set $\mathcal{A}'_\ell$ is tight. We are done if we can show that $\mathcal{A}_\ell \subseteq \mathcal{A}'_\ell$. This is straightforward because if $Q \in \mathcal{A}_\ell$, then

$$\begin{aligned} Q(B_\ell^c) &= Q\left(\{\mathbf{x} : \ x_i \notin A_{\ell 2^{|i|}} \text{ for some } i \in \mathbb{Z}\}\right) \\ &\leq \sum_{i \in \mathbb{Z}} Q\left(\{\mathbf{x} : \ x_i \notin A_{\ell 2^{|i|}}\}\right) \leq \ell^{-1} \sum_{i \in \mathbb{Z}} 2^{-|i|} = 3\ell^{-1}. \end{aligned}$$

Step 3 (Lower Bound for Ergodic Measures) We wish to show that if U is an open set in $\mathcal{M}_S(\mathcal{E})$ and $Q \in U$ with $\mathcal{H}_\mu(Q) < \infty$, then

$$\liminf_{n \to \infty} n^{-1} \log \mathcal{P}_n^\mu(U) \geq -\mathcal{H}_\mu(Q). \tag{4.16}$$

We first use (4.12) to establish (4.16) for $Q \in \mathcal{M}_{er}(\mathcal{E})$. Note that $\hat{\nu}_n(\mathbf{x})$ depends on $\mathbf{x}_1^n = (x_1, \dots, x_n)$ only. As a result, for any $\delta > 0$,

$$\begin{aligned} \mathcal{P}_n^\mu(U) &= \mathbb{P}_n^\mu\left(\{\mathbf{x} : \ \hat{\nu}_n(\mathbf{x}) \in U\}\right) \\ &= \int \mathbb{1}\left(\hat{\nu}_n(\mathbf{x}) \in U\right) \ \mu^{\otimes n}(dx_1, \dots, dx_n) \ Q'_0(d\mathbf{x}^0) \\ &\geq \int \mathbb{1}\left(\hat{\nu}_n(\mathbf{x}) \in U, \ q_0(x_1, \dots, x_n|\mathbf{x}^0) > 0\right) \ \mu^{\otimes n}(dx_1, \dots, dx_n) \ Q'_0(d\mathbf{x}^0) \\ &= \int \mathbb{1}\left(\hat{\nu}_n(\mathbf{x}) \in U, \ q_0(x_1, \dots, x_n|\mathbf{x}^0) > 0\right) e^{-\log q_0(x_1, \dots, x_n|\mathbf{x}^0)} \ Q_0(dx_1, \dots, dx_n|\mathbf{x}^0) \ Q'_0(d\mathbf{x}^0) \\ &= \int \mathbb{1}\left(\hat{\nu}_n(\mathbf{x}) \in U, \ q_0(x_1, \dots, x_n|\mathbf{x}^0) > 0\right) e^{-\log q_0(x_1, \dots, x_n|\mathbf{x}^0)} \ Q(d\mathbf{x}) \\ &\geq e^{-n\mathcal{H}_\mu(Q) - n\delta} \int \mathbb{1}\left(\hat{\nu}_n(\mathbf{x}) \in U, \ q_0(x_1, \dots, x_n|\mathbf{x}^0) > 0\right) \\ &\qquad\qquad \mathbb{1}\left(n^{-1} \log q_0(x_1, \dots, x_n|\mathbf{x}^0) \leq \mathcal{H}_\mu(Q) + \delta\right) \ Q(d\mathbf{x}). \end{aligned}$$

From this, (4.12), and (4.10) we deduce that for every $\delta > 0$,

$$\liminf_{n \to \infty} n^{-1} \log \mathcal{P}_n^\mu(U) \geq -\mathcal{H}_\mu(Q) - \delta.$$

This immediately implies (4.16) when $Q \in \mathcal{M}_{er}(\mathcal{E})$.

Step 3 (Lower Bound for General Stationary Measures) We now establish (4.16) for a measure Q that can be expressed as

$$Q = \sum_{i=1}^{r} \alpha_i Q_i, \tag{4.17}$$

with $\alpha_i \geq 0$ and $Q_i \in \mathcal{M}_{er}(\mathcal{E})$ for $i \in \{1, \ldots, r\}$, and $\sum_i \alpha_i = 1$. We first try to replace $\hat{\nu}_n$ with an expression that is similar to (4.17). Let us examine this issue before periodization; observe that if we set $n_i = [n\alpha_i]$, then

$$\nu_n(\mathbf{x}) = \alpha_1 \hat{\nu}_{n_1}(\mathbf{x}) + \alpha_2 \hat{\nu}_{n_2}(T^{n_1}(\mathbf{x})) + \cdots + \alpha_r \hat{\nu}_{n_r}(T^{n_1+\cdots+n_{r-1}}(\mathbf{x})) + O(2rn^{-1}).$$

Motivated by this, let us define

$$\bar{\nu}_n(\mathbf{x}) = \alpha_1 \hat{\nu}_{n_1}(\mathbf{x}) + \alpha_2 \hat{\nu}_{n_2}(T^{n_1}(\mathbf{x})) + \cdots + \alpha_r \hat{\nu}_{n_r}(T^{n_1+\cdots+n_{r-1}}(\mathbf{x})).$$

Choose a countable dense set $\{\Psi_m \;\; m \in \mathbb{N}\}$ of uniformly continuous local functions, and define the metric

$$D(Q, Q') = \sum_{m=1}^{\infty} 2^{-m} \min \left\{ \int \Psi_m \, dQ - \int \Psi_m \, dQ', \ 1 \right\},$$

on $\mathcal{M}(\mathcal{E})$ that induces the weak topology. If the local function Ψ_m depends on $k(m)$ many consecutive coordinates, then

$$Err(n) := \sup_{\mathbf{x}} D(\hat{\nu}_n(\mathbf{x}), \bar{\nu}_n(\mathbf{x})) \leq \sum_{m=1}^{\infty} 2^{-m} \min \left\{ 2rk(m)\|\Psi_m\| n^{-1}, \ 1 \right\},$$

which goes to 0 as $n \to \infty$. Since $Q \in U$, we may find $\delta > 0$ such that

$$B_\delta(Q) = \{Q' \in \mathcal{M}_S(\mathcal{E}) : \ D(Q', Q) < \delta\}.$$

We then set $U' = B_{\delta/2}(Q)$ and assert

$$\mathbb{P}_\mu \left(\{\mathbf{x} : \ \hat{\nu}_n(\mathbf{x}) \in U\} \right) \geq \mathbb{P}_\mu \left(\{\mathbf{x} : \ \bar{\nu}_n(\mathbf{x}) \in U'\} \right), \tag{4.18}$$

for sufficiently large n (i.e. those with $Err(n) < \delta/2$). We now find open sets $U_i \subseteq \mathcal{M}_S(\mathcal{E})$, such that $Q_i \in U_i$ for $i = 1, \ldots, r$, and if we choose any $Q'_i \in U_i$ for $i = 1, \ldots, r$, then we always have

$$\sum_{i=1}^{r} \alpha_i Q'_i \in U'.$$

From this and (4.18) we learn that for large n,

$$\begin{aligned}\mathcal{P}_n^{\mu}(U) &\geq \mathbb{P}_\mu\left(\{\mathbf{x}:\ \bar{\nu}_n(\mathbf{x}) \in U'\}\right)\\ &\geq \mathbb{P}_\mu\left(\{\mathbf{x}:\ \hat{\nu}_{n_1}(\mathbf{x}) \in U_1\}\right) \cdots \mathbb{P}_\mu\left(\left\{\mathbf{x}:\ \hat{\nu}_{n_r}\left(T^{n_1+\cdots+n_{r-1}}(\mathbf{x})\right) \in U_r\right\}\right)\\ &= \mathbb{P}_\mu\left(\{\mathbf{x}:\ \hat{\nu}_{n_1}(\mathbf{x}) \in U_1\}\right) \cdots \mathbb{P}_\mu\left(\{\mathbf{x}:\ \hat{\nu}_{n_r}(\mathbf{x}) \in U_r\}\right).\end{aligned}$$

From this, Step 2, and linearity of $\mathcal{H}_\mu$, we deduce

$$\begin{aligned}\liminf_{n\to\infty} n^{-1} \log \mathcal{P}_n^{\mu}(U) &\geq \sum_{i=1}^{r} \liminf_{n\to\infty} n^{-1} \log \mathbb{P}_\mu\left(\{\mathbf{x}:\ \hat{\nu}_{n_i}(\mathbf{x}) \in U_i\}\right)\\ &\geq -\left[\alpha_1 \mathcal{H}_\mu(Q_1) + \cdots + \alpha_r \mathcal{H}_\mu(Q_r)\right] = -\mathcal{H}_\mu(Q).\end{aligned}$$

This completes the proof of (4.16) for Q of the form (4.17).

Final Step It remains to prove that if $\mathcal{H}_\mu(Q) < \infty$, then we can find a sequence of stationary measures $\{Q^n\}$ such that

(i) $\lim_{n\to\infty} Q^n = Q$;

(ii) For each Q^n there exist $\alpha_1^n, \ldots, \alpha_{r(n)}^n \geq 0$ and $Q_1^n, \ldots, Q_{r(n)}^n \in \mathcal{M}_{er}(\mathcal{E})$ such that $\sum_i \alpha_i^n = 1$, and $Q^n = \sum_i \alpha_i^n Q_i^n$.

(iii) $\lim_{n\to\infty} \mathcal{H}_\mu(Q^n) = \mathcal{H}_\mu(Q)$.

This is left as an exercise. □

Exercise 4.1 (i) Construct the sequence $\{Q^n\}$ as in the Final Step of the proof of Theorem 4.1.

(ii) Use Contraction Principle to show that Theorem 4.1 implies Sanov's theorem. □

5 Large Deviation Principles for Markov Chains

We learned from Exercise 4.1**(i)** that the Donsker-Varadhan LDP implies Sanov's LDP. Needless to say that the latter is a much stronger LDP. To put Theorem 4.1 to a good use, let us utilize Theorem 1.2 to establish two new LDPs. Perhaps the most celebrated consequence of Theorem 4.1 is an LDP for the empirical measure of a Markov chain. This is an immediate consequence of Theorem 4.1 and Theorem 1.2**(i)-(ii)**.

Definition 5.1 **(i)** By a *regular kernel* we mean a uniformly positive bounded local continuous function $\pi(\mathbf{x}^1) = \pi(x_1|\mathbf{x}^0) \in \mathcal{B}^1$ such that $\int \pi(x_1|\mathbf{x}^0)\ \mu(dx_1) = 1$ for every $\mathbf{x}^0 \in \mathcal{E}^0$.

(ii) Given a regular kernel π, we write $\mathbb{P}^{\mathbf{x}^0} = \mathbb{P}^{\mathbf{x}^0}_{\pi,\mu}$ for the probability measure on $\mathcal{E}_1$ for which the law of $(x_1, \dots, x_k)$ is given by

$$\prod_{i=1}^{k} \pi(x_i|\mathbf{x}^{i-1})\mu(dx_i).$$

Note that if $\pi(x_1|\mathbf{x}^0) = \pi(x_1|x_0)$ depends only on (x_0, x_1), then $\mathbb{P}^{x_0} = \mathbb{P}^{x_0}_{\pi,\mu} := \mathbb{P}^{\mathbf{x}^0}_{\pi,\mu}$ is the law of a Markov chain that starts from x_0 and has a Markov kernel $\pi(x_1|x_0)\ \mu(dx_1)$.

(iii) Given a regular kernel π, we define $\mathcal{P}^{\mathbf{x}^0}_n = \mathcal{P}^{\mathbf{x}^0;\pi,\mu}_n \in \mathcal{M}(\mathcal{M}_S(\mathcal{E}))$ by

$$\mathcal{P}^{\mathbf{x}^0}_n(A) = \mathbb{P}^{\mathbf{x}^0}_{\pi,\mu}\left(\{\mathbf{x}_1 :\ \hat{\nu}_n(\mathbf{x}_1) \in A\}\right).$$

Here we are simply writing $\hat{\nu}_n(\mathbf{x}_1)$ for $\hat{\nu}_n(\mathbf{x}^0, \mathbf{x}_1)$ because $\hat{\nu}_n(\mathbf{x})$ depends only on $(x_1, \dots, x_n)$.

(iv) We write

$$\alpha_n(\mathbf{x}_1) = n^{-1}\left(\delta_{x_1} + \dots + \delta_{x_n}\right),$$

for the empirical measures associated with $\mathbf{x}_1$. The law of $\alpha_n(\mathbf{x}_1)$ with respect to $\mathbb{P}^{\mathbf{x}^0} = \mathbb{P}^{\mathbf{x}^0}_{\pi,\mu}$ is denoted by $\mathbb{P}^{\mathbf{x}^0}_n = \mathbb{P}^{\mathbf{x}^0;\pi,\mu}_n \in \mathcal{M}(\mathcal{M}(E))$:

$$\mathbb{P}^{\mathbf{x}^0}_n(A) = \mathbb{P}^{\mathbf{x}^0}_{\pi,\mu}\left(\{\mathbf{x}_1 :\ \alpha_n(\mathbf{x}_1) \in A\}\right).$$

□

Our goal is to use Theorem 4.1 to establish LDP for both $\{\mathcal{P}^{\mathbf{x}^0;\pi,\mu}_n\}$ and $\{\mathbb{P}^{\mathbf{x}^0;\pi,\mu}_n\}$.

Theorem 5.1 **(i)** *The sequence* $\{\mathcal{P}^{\mathbf{x}^0;\pi,\mu}_n\}$ *satisfies an LDP with rate function*

$$\mathcal{H}_{\mu,\pi}(Q) = \int H\left(\hat{Q}_{\mathbf{x}}|\hat{\pi}_{\mathbf{x}}\right)\ Q(d\mathbf{x}),$$

where $\hat{Q}_{\mathbf{x}}(dy_1) = Q(\mathbf{x}^0, dy_1)$ *is as in Theorem 4.1, and* $\hat{\pi}_{\mathbf{x}}(dy_1) = \pi(y_1|\mathbf{x}^0)\ \mu(dy_1)$. *This LDP is uniform in* $\mathbf{x}^0$. *Moreover,*

$$\mathcal{H}_{\mu,\pi}(Q) = \sup_{F \in \mathcal{B}^1} \int \left[F(\mathbf{x}^0, x_1) - \log \int e^{F(\mathbf{x}^0, y_1)} \pi(y_1|\mathbf{x}^0)\ \mu(dy_1)\right] Q(d\mathbf{x}). \tag{5.1}$$

(ii) *The sequence* $\{\mathbb{P}_n^{\mathbf{x}^0;\pi,\mu}\}$ *satisfies an LDP with rate function*

$$I(\alpha) = \inf\{\mathcal{H}_{\mu,\pi}(Q) : \ \tau(Q) = \alpha\}, \tag{5.2}$$

where $\tau : \mathcal{M}_S(\mathcal{E}) \to \mathcal{M}(E)$ *and* $\tau(Q)$ *denotes the one-dimensional marginal of* Q. *This LDP is uniform in* $\mathbf{x}^0$.

Proof (i) Set $G(\mathbf{x}) = \log \pi(x_1|\mathbf{x}^0)$ and observe that $G \in C_b^{loc}(\mathcal{E})$. Note that for any $F \in C_b(\mathcal{M}_S(\mathcal{E}))$,

$$\begin{aligned} e^{-2k\|G\|} \int e^{nF(\hat\nu_n(\mathbf{x}_1)) + \int G \ d\hat\nu_n(\mathbf{x}_1)} \ \mathbb{P}_\mu(d\mathbf{x}) \le \int e^{nF} \ d\mathcal{P}_n^{\mathbf{x}^0} &= \int e^{nF(\hat\nu_n(\mathbf{x}_1)) + \int G \ d\nu_n(\mathbf{x}_1)} \ \mathbb{P}_\mu(d\mathbf{x}) \\ &\le e^{2k\|G\|} \int e^{nF(\hat\nu_n(\mathbf{x}_1)) + \int G \ d\hat\nu_n(\mathbf{x}_1)} \ \mathbb{P}_\mu(d\mathbf{x}), \end{aligned}$$

where k is chosen so that $G(\mathbf{x})$ depends on $(x_{-k+1}, \dots, x_{-1}, x_0)$ only. From this it is clear that

$$\lim_{n\to\infty} n^{-1} \log \int e^{nF} \ d\mathcal{P}_n^{\mathbf{x}^0} = \lim_{n\to\infty} n^{-1} \log \int e^{nF + nL_G} \ d\mathcal{P}_n^\mu,$$

where $L_G(Q) = \int G \ dQ$. From this, Theorem 4.1 and Theorem 1.1 we deduce that the sequence $\{\mathcal{P}_n^{\mathbf{x}^0;\pi,\mu}\}$ satisfies an LDP with the rate function

$$\mathcal{H}_\mu(Q) - \int G \ dQ = \mathcal{H}_\mu(Q) - \int \left[\int \log \pi(y_1|\mathbf{x}^0) \ \hat{Q}_{\mathbf{x}}(dy_1) \right] \ Q(d\mathbf{x}) = \mathcal{H}_{\mu,\pi}(Q).$$

As for (5.1), choose $F(\mathbf{x}^0, x_1) = F'(\mathbf{x}^0, x_1) - \log \pi(x_1|x_0)$ in (4.11) to assert

$$\mathcal{H}_{\mu,\pi}(Q) = \sup_{F' \in b\mathcal{B}^1(\mathcal{E})} \int \left[F'(\mathbf{x}^0, x_1) - \log \int e^{F'(\mathbf{x}^0, x_1)} \pi(x_1|x_0) \ \alpha(dx_1) \right] \ dQ.$$

(ii) Since $\tau(\hat\nu_n(\mathbf{x}_1)) = \alpha_n(\mathbf{x}_1)$, we learn

$$\mathbb{P}_n^{\mathbf{x}^0;\pi,\mu}(A) = \mathcal{P}_n^{\mathbf{x}^0;\pi,\mu}(\tau^{-1}(A)).$$

This, Part **(i)** and Theorem 1.2**(ii)** imply Part **(ii)**. $\square$

We now try to find a simpler expression for the rate function I when $\mathbb{P}^{\mathbf{x}^0} = \mathbb{P}^{x_0}$ is Markovian. Recall that by (5.1) we can express $\mathcal{H}_\mu$ as a supremum. Using this and the form of I, we can think of two possible expressions for the rate function. In the first expression, we try to find out what type of stationary measures could serve as minimizers in (5.2). Since $\mathbb{P}^{x_0}$ is a Markov chain, we guess that the minimizer is a stationary Markov chain. Such a

Markov chain is completely determined by its one-dimensional marginal α and its Markovian kernel $q(dx_1|x_0) = q(x_0, dx_1)$. Based on this, we guess that I is equal to

$$I_1(\alpha) := \inf \left\{ \int H\big(q(\cdot|x_0)|\pi(\cdot|x_0)\mu(\cdot)\big)\ \alpha(dx_0) :\ \alpha \in \mathcal{I}^q \right\}, \tag{5.3}$$

where $\mathcal{I}^q$ denotes the space of invariant measures for the Markovian kernel q:

$$\int q(dx_1|x_0)\ \alpha(dx_0) = \alpha(dx_1).$$

Alternatively, we may try to find out what type of functions $F(\mathbf{x}^0, x_1)$ in expression (5.1) would be a maximizer for a Markovian Q. As we will see below, such F would be a function of x_1 only. Motivated by this, we define

$$I_2(\alpha) := \sup_{f \in b\mathcal{B}(E)} \int (f - \lambda(f))\ d\alpha, \tag{5.4}$$

where

$$\lambda(f)(x_0) = \log \int e^{f(x_1)} \pi(x_1|x_0)\ \mu(dx_1). \tag{5.5}$$

Theorem 5.2 *If $\pi(x_1|\mathbf{x}^0) = \pi(x_1|x_0)$ is Markovian, then $I = I_1 = I_2$, where I_1 and I_2 are defined by* (5.3) *and* (5.4).

Proof *Step 1* By restricting the supremum in (5.1) to functions of the form $F(\mathbf{x}^0, x_1) = f(x_1)$, we learn that if $\tau_1(Q) = \alpha$, then

$$I(\alpha) \geq I_2(\alpha). \tag{5.6}$$

We certainly have $I \leq I_3$, where

$$I_3(\alpha) = \inf \left\{ \mathcal{H}_{\mu,\pi}(Q) :\ \tau_1(Q) = \alpha,\quad Q \in \mathcal{M}_S(\mathcal{E}) \text{ is Markovian} \right\}.$$

We can easily verify that $I_3 = I_1$. In view of (5.6), it remains to show that $I_1 \leq I_2$.

Observe that if $I_1(\alpha) < \infty$, then $q(dx_1|x_0) \ll \mu(dx_1)$ for μ-almost all x_0. If we write

$$q(dx_1|x_0) = q(x_1|x_0)\ \mu(dx_1),$$

then

$$\int H\big(q(\cdot|x_0)|\pi(\cdot|x_0)\mu(\cdot)\big)\ \alpha(dx_0) = \int q(x_1|x_0) \log \frac{q(x_1|x_0)}{\pi(x_1|x_0)}\ \mu(dx_1)\alpha(dx_0) = H(\gamma|p^\alpha),$$

where

$$\gamma(dx_0, dx_1) = q(dx_1|x_0)\alpha(dx_0) = q(x_1|x_0)\mu(dx_1)\alpha(dx_0),$$
$$p^\alpha(dx_0, dx_1) = \pi(x_1|x_0)\mu(dx_1)\alpha(dx_0).$$

Here γ is the 2-dimensional marginal of a Markov $Q \in \mathcal{M}_S(\mathcal{E})$ and could be any probability measure on $E^2 = E \times E$ with identical marginals. Writing $\tau_1(\gamma)$ and $\tau_2(\gamma)$, for the marginals of γ, we may use Donsker-Varadhan variational formula (3.7) for the entropy, to assert

$$\begin{aligned} I_1(\alpha) &= \inf\left\{H(\gamma|p^\alpha):\ \gamma \in \mathcal{M}(E \times E),\quad \tau_1(\gamma) = \tau_2(\gamma) = \alpha\right\} \\ &= \inf_{\gamma\in\Gamma(\alpha)} \sup_{g\in b\mathcal{B}(E^2)} \left[\int g\ d\gamma - \log \int e^g\ dp^\alpha\right], \end{aligned} \tag{5.7}$$

where

$$\Gamma(\alpha) := \left\{\gamma \in \mathcal{M}(E \times E):\ \tau_1(\gamma) = \tau_2(\gamma) = \alpha\right\}.$$

Step 2 It remains to show that $I_1 \le I_2$ with I_1 and I_2 given by (5.7) and (5.4). We wish to interchange the supremum with infimum in (5.7) with the aid of the Minimax Theorem D.1. We note that if

$$J(g, \gamma) = \int g\ d\gamma - \log \int e^g\ dp^\alpha,$$

then J is linear in γ and concave in g. Moreover the set $\Gamma(\alpha)$ is convex and compact. The latter is an immediate consequence of the bound

$$\gamma\left((K \times K)^c\right) \le 2\alpha(K^c),$$

for $\gamma \in \Gamma(\alpha)$. From this and Minimax Theorem we deduce

$$\begin{aligned} I_1(\alpha) &= \sup_{g\in b\mathcal{B}(E^2)} \inf_{\gamma\in\Gamma(\alpha)} \left[\int g\ d\gamma - \log \int e^g\ dp^\alpha\right] \\ &= \sup_{g\in b\mathcal{B}(E^2)} \left[\inf_{\gamma\in\Gamma(\alpha)} \int g\ d\gamma - \log \int e^g\ dp^\alpha\right]. \end{aligned} \tag{5.8}$$

By Kantorovich's duality formula (see Theorem E.1 below), we have

$$\inf_{\gamma\in\Gamma(\alpha)} \int g\ d\gamma = \sup_{(f,h)\in\Gamma'(g)} \int (f + h)\ d\alpha,$$

where $\Gamma'(g)$ denotes the set of pairs $(f, h) \in C_b(E)^2$ such that

$$f(x_1) + h(x_0) \le g(x_0, x_1),$$

for all $x_0, x_1 \in E$. From this and (5.8) we deduce

$$\begin{aligned}
I_1(\alpha) &= \sup_{g\in b\mathcal{B}(E^2)} \sup_{(f,h)\in\Gamma'(g)} \left[\int (f+h)\ d\alpha - \log \int e^g\ dp^\alpha\right] \\
&\le \sup_{g\in b\mathcal{B}(E^2)} \sup_{(f,h)\in\Gamma'(g)} \left[\int (f+h)\ d\alpha - \log \int e^{f(x_1)+h(x_0)}\ p^\alpha(dx_0,dx_1)\right] \\
&= \sup_{f,h\in C_b(E)} \left[\int (f+h)\ d\alpha - \log \int e^{f(x_1)+h(x_0)}\ p^\alpha(dx_0,dx_1)\right] =: I_4(\alpha).
\end{aligned}$$

We are done if we can show that $I_4 = I_2$.

Final Step To verify the equality of I_4 and I_2, let us define

$$\begin{aligned}
\eta(f) &= \log \int_{E^2} e^{f(x_1)}\ p^\alpha(dx_0,dx_1), \\
\beta(dx_0) &= e^{-\eta(f)} \int_E e^{f(x_1)}\ p^\alpha(dx_0,dx_1) = e^{\lambda(f)(x_0)-\eta(f)}\ \alpha(dx_0),
\end{aligned}$$

so that $\beta \in \mathcal{M}(E)$. We can now write,

$$\begin{aligned}
I_4(\alpha) &= \sup_{f\in C_b(E)} \left\{\int f\ d\alpha + \sup_{h\in C_b(E)} \left[\int h\ d\alpha - \log \int e^{f(x_1)+h(x_0)}\ p^\alpha(dx_0,dx_1)\right]\right\} \\
&= \sup_{f\in C_b(E)} \left\{\int f\ d\alpha - \eta(f) + \sup_{h\in C_b(E)} \left[\int h\ d\alpha - \log \int e^h\ d\beta\right]\right\} \\
&= \sup_{f\in C_b(E)} \left\{\int f\ d\alpha - \eta(f) + H(\alpha|\beta)\right\} \\
&= \sup_{f\in C_b(E)} \left\{\int f\ d\alpha - \eta(f) - \int \big(\lambda(f) - \eta(f)\big)\ d\alpha\right\} \\
&= \sup_{f\in C_b(E)} \left\{\int f\ d\alpha - \int \lambda(f)\ d\alpha\right\} = I_2(\alpha).
\end{aligned}$$

We are done. □

Remark 5.1 The expression I_1 in (5.3) was formulated based on our correct prediction that when π is Markovian, the minimizing Q in (5.1) should also be Markovian. This and $\tau(Q) = \alpha$ does not determine Q uniquely. To figure out what the exact form of the minimizer Q is, let us first find out what the maximizing f in (5.4) is. Indeed, if the maximizing f is denoted by $\bar f$, then we would have $\partial\mathcal{K}(\bar f) = 0$, where $\mathcal{K}(f) = \int (f - \lambda(f))\ d\alpha$ and $\partial\mathcal{K}$ denotes the (Gâteau) derivative of $\mathcal{K}$. More precisely,

$$0 = \partial\mathcal{K}(\bar f)h = \frac{d}{d\varepsilon}\mathcal{K}(\bar f + \varepsilon h)|_{\varepsilon=0} = \int h\ d\alpha - \int e^{\bar f(x_1)-\lambda(\bar f)(x_0)} h(x_1)\pi(x_1|x_0)\ \mu(dx_1)\alpha(dx_0),$$

for every $h \in C_b(E)$. This means that if

$$\pi^{\bar{f}}(dx_1|x_0) = e^{\bar{f}(x_1)-\lambda(\bar{f})(x_0)}\pi(x_1|x_0)\ \mu(dx_1),$$

then $\pi^{\bar{f}}$ is a Markovian kernel for which α is an invariant measure. In fact it is straightforward to show if $Q^{\bar{f}}$ denotes the stationary Markov chain with $\tau(Q) = \alpha$ and Markov kernel $\pi^{\bar{f}}$, then

$$\mathcal{H}_{\mu,\pi}(Q^{\bar{f}}) = \mathcal{K}(\bar{f}) = I(\alpha).$$

□

Exercise 5.1 Define $T : \mathcal{M}(E) \to \mathcal{M}(E)$ by

$$T\alpha(dx_1) = \left[\int \pi(x_1|x_0)\ \alpha(dx_0)\right]\ \mu(dx_1).$$

Show

$$\|T(\alpha) - \alpha\| \leq R(I(\alpha)),$$

where

$$R(a) = \inf_{b>0}\left[\frac{a + b - \log(b+1)}{b}\right], \qquad \|\nu\| = \sup_{A\in\mathcal{B}(E)} |\nu(A)|.$$

Hint: Use $I(\alpha) \geq \int (f - \lambda(f))\ d\alpha$ for $f = \log(b+1)1\!\!1_A$. □

6 Wentzell-Freidlin Problem

Throughout this section, E denotes the space of continuous functions $x : [0, T] \to \mathbb{R}^d$ with $x(0) = 0$, and $\|x\| = \sup_t |x(t)|$ denotes the uniform norm. We also write $\mathbb{P} \in \mathcal{M}(E)$ for the Wiener measure. More precisely, the process $x(\cdot)$ is a Standard Brownian motion with respect to $\mathbb{P}$. Moreover the law of the process $n^{-1/2}x(\cdot)$ is denoted by $\mathbb{P}_n$: For every $f \in C_b(E)$,

$$\int f \; d\mathbb{P}_n = \int f\left(n^{-1/2}x\right) \mathbb{P}(dx).$$

Theorem 6.1 *(Schilder) The family $\{\mathbb{P}_n : \varepsilon > 0\}$ satisfies an LDP with the rate function*

$$I(x) = \begin{cases} \frac{1}{2}\int_0^T |\dot{x}(t)|^2 \; dt & x \in H^1, \\ \infty & \textit{otherwise}, \end{cases}$$

where H^1 denotes the space of weakly differentiable continuous functions.

Proof *Step 1 (weak ULDP)* Write BV_T for the space of vector-valued left-continuous functions $F = (F_1, \dots, F_d) : [0, T] \to \mathbb{R}^d$ such that each F_i is of bounded variation. We then set

$$\mathcal{V} = \{L_F : \; F \in BV_T\},$$

where $L_F(x) = \int F \cdot dx$. This integral may be understood in the Riemann-Steiltjes sense after an integration by parts:

$$L_F(x) = \int_0^T F \cdot dx = F(T) \cdot x(T) - \int_0^T x \cdot dF.$$

Alternatively ,

$$L_F(x) = \int_0^T F \cdot dx = \lim_{k\to\infty} \sum_{j=1}^k F(t_{j-1}) \cdot (x(t_j) - x_(t_{j-1})),$$

where $t_j = Tj/k$. From this we can readily deduce that the random variable $L_F(x)$ is a normal random variable with mean 0 and variance $\frac{1}{2}\int_0^T |F|^2 \; dt$: Indeed

$$\begin{aligned} \int e^{iL_F(x)} \; \mathbb{P}(dx) &= \lim_{k\to\infty} \int e^{i\sum_{j=1}^k F(t_{j-1})\cdot(x(t_j)-x(t_{j-1}))} \; \mathbb{P}(dx) \\ &= \lim_{k\to\infty} e^{2^{-1}\sum_{j=1}^k |F(t_{j-1})|^2(t_j-t_{j-1})} = e^{\frac{1}{2}\int_0^T |F|^2 \; dt}, \end{aligned}$$

for every $F \in BV_T$.

To apply Theorem 2.1, we wish to evaluate $\Lambda(L_F)$ for every $F \in BV_T$. Clearly

$$\int e^{nL_F}\ d\mathbb{P}_n = \int e^{n^{1/2}L_F}\ d\mathbb{P} = e^{\frac{n}{2}\int_0^T |F|^2\ dt}.$$

As a result,

$$\Lambda(L_F) = \frac{1}{2}\int_0^T |F|^2\ dt.$$

From this and Theorem 2.1**(i)** we deduce an ULDP for compact sets for the rate function

$$I_u(x) = \sup_{F\in BV_T}\left[\int F\cdot dx - \frac{1}{2}\int_0^T |F|^2\ dt\right].$$

Step 2 ($I = I_u$) We now verify $I = I_u$. The proof of $I_u \le I$ is straightforward because if $F \in BV_T$ and x is absolutely continuous with $\dot{x} \in L^2([0,T])$, then

$$\int F\cdot dx = \int F\cdot \dot{x}\ dt \le \frac{1}{2}\int_0^T |F|^2\ dt + \frac{1}{2}\int_0^T |\dot{x}|^2\ dt.$$

We first note that for a smooth x, we can readily establish the reverse inequality by selecting $F(t) = \dot{x}(t)$. Motivated by this, we set $t_j = t_j^k = Tj/k$, $x_j = x(t_j)$, and choose

$$F = \frac{k}{T}\sum_{j=1}^k (x_j - x_{j-1})1\!\!1_{[t_{j-1},t_j)}.$$

We certaily have,

$$\int F\cdot dx - \frac{1}{2}\int_0^T |F|^2\ dt = \frac{k}{T}\sum_{j=1}^k |x_j - x_{j-1}|^2 - \frac{k}{2T}\sum_{j=1}^k |x_j - x_{j-1}|^2 = \frac{k}{2T}\sum_{j=1}^k |x_j - x_{j-1}|^2.$$

From this we deduce

$$I_u(x) \ge \sup_{k\in\mathbb{N}}\ \frac{k}{2T}\sum_{j=1}^k \left|x(t_j^k) - x(t_j^k)\right|^2. \tag{6.1}$$

Note that if $x^k \in E$ denotes the linear interpolation of x between $(t_j^k : j = 0,\dots,k)$, then $x^k \in H^1$ and

$$\frac{1}{2}\int_0^T \left|\dot{x}^k\right|^2\ dt = \frac{k}{2T}\sum_{j=1}^k \left|x(t_j^k) - x(t_j^k)\right|^2.$$

Now if $I_u(x) < \infty$, then by (6.1), the sequence $\{\dot{x}^k : \ k \in \mathbb{N}\}$ is bounded in L^2. Hence this sequence has a subsequence that converges weakly to $y \in L^2([0,T])$ with

$$\frac{1}{2}\int_0^T |\dot{y}|^2 \ dt \le I_u(x). \tag{6.2}$$

Since $x^k \to x$ pointwise in large k limit, we have

$$\int \dot{\zeta} \cdot x \ dt = \lim_{k\to\infty} \int \dot{\zeta} \cdot x^k \ dt = -\lim_{k\to\infty} \int \zeta \cdot \dot{x}^k \ dt = \int \zeta \cdot y \ dt,$$

for every smooth ζ with support in the interval $(0,T)$. Hence x is weakly differentiable with $\dot{x} = y$. This and (6.2) imply that $I(x) \le I_u(x)$, as desired.

Step 3 (Exponential Tightness) Pick $\alpha \in (0,1/2)$, set $\delta_k = T/k$, and define

$$G(x) = \sup_{k\in\mathbb{N}} \sup_{\substack{|t-s|\le\delta_k \\ s,t\in[0,T]}} |x(t) - x(s)|\delta_k^{-\alpha}.$$

For the exponential tightness, it suffices to show

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} n^{-1} \log \mathbb{P}_n(G \ge \ell) = \infty. \tag{6.3}$$

Writing $t_i^k = iTk^{-1}$,

$$G(x) \le 3 \sup_{k\in\mathbb{N}} \sup_{1\le i\le k} \sup_{t\in[t_{i-1}^k, t_i^k]} \left| x(t) - x(t_{i-1}^k) \right| \delta_k^{-\alpha}.$$

As a result,

$$\begin{aligned}
\mathbb{P}_n(G \ge \ell) &\le \sum_{k=1}^\infty k\mathbb{P}_n\left(\sup_{t\in[0,\delta_k]} |x(t)| \ge 3^{-1}\delta_k^\alpha \ell \right) \le \sum_{k=1}^\infty k\mathbb{P}\left(\sup_{t\in[0,\delta_k]} |x(t)| \ge 3^{-1}\delta_k^\alpha \ell n^{\frac{1}{2}} \right) \\
&\le \sum_{k=1}^\infty 2k\mathbb{P}\left(|x(\delta_k)| \ge 3^{-1}\delta_k^\alpha \ell n^{\frac{1}{2}} \right) \le \sum_{k=1}^\infty 2k\mathbb{P}\left(|x(1)| \ge 3^{-1}\delta_k^{\alpha-\frac{1}{2}} \ell n^{\frac{1}{2}} \right) \\
&\le c_0 \sum_{k=1}^\infty k e^{-c_1 k^{2\alpha-1}\ell^2 n} \le c_0 \sum_{k=1}^\infty k e^{-c_2 k^{2\alpha-1}} e^{-\ell^2 n} \le c_3 e^{-\ell^2 n}.
\end{aligned}$$

This certainly implies (6.3).

Step 4 (LLDP) To apply Theorem 2.1, we need to determine $\rho(\mathcal{V})$ of (2.5). Note that if

$$d\mathbb{P}_n^F := d\mathbb{P}_n^{L_F} = e^{n\int_0^T F\cdot dx - \frac{n}{2}\int_0^T |F|^2 \ dt} \ d\mathbb{P}_n,$$

then for any $g \in C_b(E)$,

$$\int g \, d\mathbb{P}_n^F = \int g\left(n^{-1/2}x\right) \; e^{n^{1/2}\int_0^T F\cdot dx - \frac{n}{2}\int_0^T |F|^2 \; dt} \; \mathbb{P}(dx)$$
$$= \int g\left(n^{-1/2}(x + n^{1/2}f\right) \; \mathbb{P}(dx) = \int g\left(n^{-1/2}x + f\right) \; \mathbb{P}(dx),$$

where $f(t) = \int_0^t F(s) \; ds$, and we used Cameron-Martin's formula (Theorem 6.2 below) for the second equality. In other words, $\mathbb{P}_n^F$ is the law of $n^{-1/2}x + F$ with x a standard Brownian motion. From this it is clear that $\mathbb{P}_n^F \to \delta_f$ in large n limit. Hence

$$\rho(\mathcal{V}) = \left\{ f(t) = \int_0^t F(s) \; ds : \; F \in BV \right\}.$$

To complete the proof of lower bound, we still need to verify (2.6), namely if $I(x) < \infty$, then we can find $X_n \in \rho(\mathcal{V})$ such that $I(x_n) \to I(x)$ in large n limit. This is straightforward and follows from the fact that the space of smooth (hence BV) is dense in $L^2([0,T])$. □

As our next model, we consider a dynamical system that is perturbed by a small white noise. Given a bounded continuous vector field b, consider the equation

$$dy^n = b(y^n, t)dt + n^{-1/2}dB, \quad y^n(0) = 0, \tag{6.4}$$

where B is a standard Brownian motion. By this we mean that $y^n \in E$ satisfies implicitly $\Psi(y^n) = n^{-1/2}B$, where $\Psi : E \to E$ is defined by

$$\Psi(y)(t) = y(t) - \int_0^t b(y(s), s) \; ds.$$

Evidently the map Ψ is a continuous function.

Lemma 6.1 *The map Ψ is a homeomorphism.*

This lemma allows us to apply Contraction Principle (Theorem 1.2**(i)**) to the LDP of Theorem 6.1 with $\Phi = \Psi^{-1}$. Recall that $\mathbb{P}_n$ is the law of $n^{-1/2}B$ with B a standard Brownian motion. Let us write $\mathbb{Q}_n$ for the law of the process y^n of (6.4).

Corollary 6.1 *The sequence $\{\mathbb{Q}_n : \; n \in \mathbb{N}\}$ satisfies LDP with the rate function*

$$I'(y) = \begin{cases} \frac{1}{2}\int |\dot{y}(t) - b(y(t), t)|^2 \; dt & \text{if } y \text{ weakly differentiable;} \\ \infty & \text{otherwise.} \end{cases}$$

More generally, we may consider the stochastic differential equation

$$dy^n = b(y^n, t)dt + n^{-1/2}\sigma(y^n, t)dB, \quad y^n(0) = 0, \tag{6.5}$$

where both b and σ are Lipschitz continuous in x, uniformly in $t \in [0, T]$, and σ is a $d \times d$-invertible matrix for each (x, t). Let us continuous to write $\mathbb{Q}_n$ for the law of the process y^n that satisfies (6.5). We are tempted to define $\Phi(x) = y$ implicitly by the equation

$$y(t) = \int_0^t b(y(s), s)\ ds + \int_0^t \sigma(y(s), s)\ dx(s),$$

and use

$$\mathbb{Q}_n(A) = \mathbb{P}_n\left(\Phi^{-1}(A)\right),$$

to assert the following generalization of Corollary 5.1:

Theorem 6.2 *The sequence $\{\mathbb{Q}_n :\ n \in \mathbb{N}\}$ satisfies LDP with the rate function*

$$I''(y) = \begin{cases} \frac{1}{2}\int \left|\sigma(y(t), t)^{-1}\left(\dot{y}(t) - b(y(t), t)\right)\right|^2\ dt & \textit{if } y \textit{ weakly differentiable;} \\ \infty & \textit{otherwise.} \end{cases}$$

The difficulty is that the transformation Φ is no longer continuous and we need to show that Φ can be approximated by continuous functions for a price that is super exponentially small.

7 Stochastic Calculus and Martingale Problem

Let E be a Polish space and write $C_0(E)$ for the space of continuous functions $f : E \to \mathbb{R}$ that vanish at infinity. As before the space $C_b(E)$ is equipped with the uniform norm $\|\cdot\|$. Also $C_0(E)$ is a closed subset of $C_b(E)$. A *Feller Markov process* in E is specified with its transition probabilities $\{p_t(x,\cdot) : \ t \geq 0, x \in E\} \subseteq \mathcal{M}(E)$: If

$$T_t f(x) = \int f(y)\ p_t(x, dy),$$

then $T : C_0(E) \to C_0(E)$, $T_0(f) = f$, $T_t \circ T_s = T_{t+s}$ for every $s, t \geq 0$, and for every $f \in C_0(E)$, we have that $T_t f \to f$ as $t \to 0$. We note that $\|T_t f\| \leq \|f\|$, and the map $t \mapsto T_t f$ is continuous for every $f \in C_0(E)$.

The family $\{T_t : \ t \geq 0\}$ is an example of a *strongly continuous semigroup* and its infinitesimal generator is defined by

$$\mathcal{A}f = \lim_{t \to \infty} t^{-1}(T_t f - f).$$

The set of functions $f \in C_0(E)$ for which this limit exits is denoted by $Dom(\mathcal{A})$.

8 Miscellaneous

8.1 Random matrices and Erdős-Rényi graphs

We begin with formulating some natural combinatorial questions about graphs. As we will see in this section, we will be able to answer these questions with the aid of LD techniques and some sophisticated combinatorial and analytical tools. To begin, let us write $\mathcal{G}_n$ for the space of simple unordered graphs with n vertices and $\mathcal{G} = \cup_{n=1}^{\infty} \mathcal{G}_n$ for the space of all such graphs. More precisely, $\mathcal{G}_n$ is simply the space of edge sets G of unordered pairs $e = \{a, b\} \subseteq \{1, \dots, n\}$ with $a \neq b$. Since there are $n(n-1)/2$ many unordered edges, we have $\sharp \mathcal{G}_n = 2^{n(n-1)/2}$. We define two functions $\mathbf{v}, \mathbf{e} : \mathcal{G} \to \mathbb{N}$ to represent the number of vertices and edges of a graph: $\mathbf{v}(G) = n$ for $G \in \mathcal{G}_n$, and $\mathbf{e}(G) = \sharp G$. We may ask the following combinatorial question: How many graphs $G \in \mathcal{G}_n$ has exactly tn^3 many triangles? By a triangle in a graph G, we mean an unordered triplet $\{a, b, c\}$ such that $\{a, b\}, \{a, c\}, \{c, b\} \in G$. Note that generically we would have $O(n^3)$ many triangles in a graph of n vertices and since there are $2^{n(n-1)/2}$ such graphs, we may wonder whether or not we can calculate

$$T_0 := \lim_{n \to \infty} n^{-2} \log \sharp \{G \in \mathcal{G}_n : \boldsymbol{\tau}(G) \geq tn^3\}, \tag{8.1}$$

where $\boldsymbol{\tau}(G)$ is the total number of triangles in G. To give a probabilistic flavor to this problem, we may use the uniform probability measure $\mathbb{U}_n$ on $\mathcal{G}_n$ and wonder what would be the probability of having at least tn^3 many triangles in a randomly sampled graph of n vertices:

$$T := \lim_{n \to \infty} n^{-2} \log \mathbb{U}_n \left(\{G \in \mathcal{G}_n : \boldsymbol{\tau}(G) \geq tn^3\} \right).$$

This is clearly equivalent to (8.1) and $T = T_0 - \log 2/2$. More generally, we may pick $p \in (0, 1)$ and select edges independently with probability p. The outcome is known as the Erdős-Rényi $G(n, p)$ model and is a probability measure $\mathbb{U}_n^p$ on $\mathcal{G}_n$ so that any graph $G \in \mathcal{G}_n$ with m edges occurs with probability

$$p^m (1-p)^{\frac{n(n-1)}{2} - m}.$$

Evidently $\mathbb{U}_n^{1/2} = \mathbb{U}_n$. We are now interested in

$$T(p) := \lim_{n \to \infty} n^{-2} \log \mathbb{U}_n^p \left(\{G \in \mathcal{G}_n : \boldsymbol{\tau}(G) \geq tn^3\} \right). \tag{8.2}$$

To turn (8.2) to a more familiar LD problem, observe that we may regard a graph $G \in \mathcal{G}_n$ as a symmetric $n \times n$ matrix $X_n(G) = [x_{ij}(G)]_{i,j=1}^n$ such that

$$x_{ij}(G) = \mathbb{1}(\{i, j\} \in G).$$

With this notation, $(x_{ij}(G) : i > j)$ is a collection of $n(n-1)/2$ iid Bernoulli random variables under $\mathbb{U}_n^p$ with $\mathbb{U}_n^p(x_{ij} = 1) = p$ for every $i > j$. More generally, we may pick a probability measure $\mu \in \mathcal{M}(\mathbb{R})$, and let $X_n(\omega) = [x_{ij}]_{i,j=1}^n$ be a random symmetric $n \times n$ matrix where the entries for $i > j$ are iid with law $\mu \in \mathcal{M}(\mathbb{R})$, and either assume that $x_{ii} = 0$ for each i, or choose diagonal entries iid with law μ and independently from the off-diagonal entries. The law of such a matrix is denoted by $\mathbb{U}_n^\mu$. Writing $(\lambda_i(X_n) : i = 1, \dots, n)$ for the eigenvalues of the random matrix X_n, we may wonder whether or not we have a LDP for empirical measure of these eigenvalues. As it turns out, the random variable

$$\frac{1}{6} \sum_{i=1}^n \lambda_i(X_n)^3,$$

is indeed the total number of triangles in a graph G, when $X_n = X_n(G)$.

The primary purpose of this section is the statement and proof of a LDP for the family $\{X_n\}$ that has recently been obtained by Chaterjee and Varadhan [CV]. This LDP allows us to evaluate $T(p)$ and analogous quantities for large symmetric random matrices.

Before stating the main result of this section, let us go back to our Cramér-Sanov LDP and discuss some possible refinements. We then use these refinements to motivate Chaterjee-Varadhan's work.

Given a sequence $\mathbf{x} = (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}}$, define

$$\gamma_n(\theta; \mathbf{x}) = \sum_{i=1}^n x_i \ \mathbb{1}_{[(i-1)/n, i/n]}(\theta), \qquad \gamma_n'(d\theta; \mathbf{x}) = n^{-1} \sum_{i=1}^n x_i \ \delta_i(d\theta). \tag{8.3}$$

Writing $E = L^1([0,1])$ for the space of Lebesgue integrable functions and $E' = \mathcal{M}_{sn}([0,1])$ for the space of signed measures of finite variation on $[0,1]$, we certainly have that $\gamma_n(\cdot; \mathbf{x}) \in E$ and $\gamma_n'(\cdot; \mathbf{x}) \in E'$. We equip E' with the topology of weak convergence and regard E as the set of signed measures that are absolutely continuous with respect to the Lebesgue measure. If $\mathbf{x}$ is a sequence of iid random variables with law $\mu \in \mathcal{M}(\mathbb{R})$, then the transformations $\mathbf{x} \mapsto \gamma_n(\cdot; \mathbf{x}) \ d\theta$ and $\mathbf{x} \mapsto \gamma_n'(\cdot; \mathbf{x})$ induce two probability measures $\mathbb{P}_n$ and $\mathbb{P}_n'$ on E and E' respectively. We wish to establish a LDP for the families $\{\mathbb{P}_n\}$ and $\{\mathbb{P}_n'\}$. It is not hard to see that the families $\{\mathbb{P}_n\}$ and $\{\mathbb{P}_n'\}$ satisfy the same LDP. Moreover, either by modifying the proof of Theorem 3.1, or by applying the contraction principle to the LDP of Exercise 3.4 we can show that the family $\{\mathbb{P}_n\}$ satisfies LDP with a rate function $I : E \to [0, \infty]$ such that if $I(\gamma) < \infty$, then γ is absolutely continuous with respect to the Lebesgue measure, and

$$I(\gamma) = \int_0^1 h\left(\frac{d\gamma}{d\theta}\right) \ d\theta,$$

with

$$h(\rho) = \sup_{v \in \mathbb{R}} (\rho v - \lambda(v)), \tag{8.4}$$

where λ as before is given by $\lambda(v) = \log \int e^{xv}\ \mu(dx)$.

We now take a matrix X_n that is distributed according to $\mathbb{U}_n^{\mu}$. To preserve the matrix structure, and in analogy with (8.3), define

$$k_n(\theta_1, \theta_2; X_n) = \sum_{i,j=1}^{n} x_{ij}\ \mathbb{1}((\theta_1, \theta_2) \in J_i \times J_j),$$

$$k'_n(d\theta_1, d\theta_2; X_n) = n^{-2} \sum_{i,j=1}^{n} x_{ij}\ \delta_{(i/n, j/n)}(d\theta_1, d\theta_2),$$

where $J_i = [(i-1)/n, i/n]$. The transformation $X_n \mapsto k_n(\cdot, \cdot; X_n)\ d\theta_1 d\theta_2$ and $X_n \mapsto k'_n(\cdot, \cdot; X_n)$ push forward the probability measure $\mathbb{U}_n^{\mu}$ to the probability measures $\mathbb{Q}_n$ and $\mathbb{Q}'_n$ on $\mathcal{E} = \mathcal{M}_{sn}([0,1]^2)$. As before, we equip $\mathcal{E}$ with the weak topology and examine the question of LDP for the families $\{\mathbb{Q}_n\}$ and $\{\mathbb{Q}'_n\}$. Again these two families enjoy the same LDP and in just the same way we treated the families $\{\mathbb{P}_n\}$ or $\{\mathbb{P}'_n\}$, we can show the following LDP for the family $\{\mathbb{Q}_n\}$. (See also Theorem 8.2 below.)

Theorem 8.1 *The family $\{\mathbb{Q}_n\}$ satisfies LDP with the rate function $I : \mathcal{E} \to [0, \infty]$ such that if $I(\gamma) < \infty$, then γ is absolutely continuous with respect to the Lebesgue measure with a symmetric Radon-Nikodym derivative, and when $d\gamma = g\ d\theta_1 d\theta_2$,*

$$I(\gamma) = \frac{1}{2} \int_0^1 \int_0^1 h(g(\theta_1, \theta_2))\ d\theta_1 d\theta_2, \tag{8.5}$$

with h as in (8.4).

To explain the appearance of $1/2$ in (8.5), write C_{sym} for the symmetric continuous functions on $[0,1]^2$. Observe that if $f \in C_{sym}$, and $\Delta = \{(s,t) \in [0,1]^2 : s \le t\}$, then

$$\begin{aligned}
\lim_{n\to\infty} n^{-2} \log \int e^{n^2 \int f d\gamma}\ \mathbb{Q}_n(d\gamma) &= \lim_{n\to\infty} n^{-2} \log \int e^{2n^2 \int_\Delta f\ d\gamma}\ \mathbb{Q}_n(d\gamma) \\
&= \lim_{n\to\infty} n^{-2} \sum_{i,j=1}^{n} \lambda \left(2n^2 \int_{J_i \times J_j} f\ d\theta_1 d\theta_2 \right) \mathbb{1}(j \ge i) \\
&= \frac{1}{2} \int_0^1 \int_0^1 \lambda(2f)\ d\theta_1 d\theta_2 =: \Lambda(f).
\end{aligned} \tag{8.6}$$

On the other hand, the LD rate function is given by the Legendre transform of Λ:

$$\begin{aligned}
I(\gamma) &= \sup_{f \in C_{sym}} \left(\int f\ d\gamma - \frac{1}{2} \int_0^1 \int_0^1 \lambda(2f)\ d\theta_1 d\theta_2 \right) \\
&= \frac{1}{2} \sup_{f \in C_{sym}} \left(\int f\ d\gamma - \int_0^1 \int_0^1 \lambda(f)\ d\theta_1 d\theta_2 \right).
\end{aligned}$$

This equals the right-hand side of (8.5) when $\gamma \ll d\theta_1 d\theta_2$.

To avoid some technical issues and simplify our presentation, let us assume that the measure μ is concentrated on the set $[-\ell, \ell]$. This in particular implies that $|k_n|$ is bounded by ℓ. As a result, we may switch to the smaller state space $\mathcal{K}_\ell$ that consists of bounded measurable functions k with $|k| \le \ell$. Note that $\mathcal{K}_\ell$ may be identified with a closed subset of $\mathcal{E}$ and is a compact metric space with respect to the weak topology. In the case of an Erdős-Rényi graph, we can even choose the space $\mathcal{K}^1$ of bounded measurable functions k, taking values in the interval $[0, 1]$ for the state space. With a slight abuse of notation, let us regard $\mathbb{Q}_n$ as a probability measure on $\mathcal{K}_\ell$ ($\mathcal{K}^1$ in the case of Erdős-Rényi graph), and write $\mathcal{I} : \mathcal{K}_\ell \to [0, \infty)$ for the rate function:

$$\mathcal{I}(k) = \frac{1}{2} \int_0^1 \int_0^1 h(k(\theta_1, \theta_2)) \; d\theta_1 d\theta_2.$$

Unfortunately the LDP of Theorem 8.1 is not strong enough to allow us to evaluate $T(p)$ of (8.2). This is because, if we attempt to express $\boldsymbol{\tau}(G)$ in terms of $\hat{k}_n = k_n(\cdot, \cdot; X_n(G))$, we learn that for $G \in \mathcal{G}_n$, $\boldsymbol{\tau}(G) = \hat{\boldsymbol{\tau}}(\hat{k}_n)$, where

$$\hat{\boldsymbol{\tau}}(k) = \frac{1}{6} \int_{[0,1]^3} k(\theta_1, \theta_2) k(\theta_2, \theta_3) k(\theta_3, \theta_1) \; d\theta_1 d\theta_2 d\theta_3,$$

and the function $\hat{\boldsymbol{\tau}} : \mathcal{K} \to \mathbb{R}$ is not continuous with respect to the weak topology. Certainly $\hat{\boldsymbol{\tau}}$ is continuous with respect to the topology of (strong) L^1 convergence, but this is too strong for establishing an LDP. A natural question is whether or not we can strengthen the weak topology to guarantee the continuity of $\hat{\boldsymbol{\tau}}$ without spoiling our LDP.

As it turns out, the *cut metric* $d_\square$ of Frieze and Kannan would do the job as the LDP result of Chaterjee and Varadhan demonstrate. This metric comes from the cut norm

$$\|k\|_\square := \sup \left\{ \int_0^1 \int_0^1 k(\theta_1, \theta_2) f(\theta_1) g(\theta_2) \; d\theta_1 d\theta_2 : \; f, g \in \mathcal{B}([0,1]), |f|, |g| \le 1 \right\}.$$

We note that for $k \in \mathcal{K}_\ell$,

$$\int_{[0,1]^2} k(\theta_1, \theta_2) k(\theta_2, \theta_3) k(\theta_3, \theta_1) \; d\theta_1 d\theta_2 \le \ell^2 \|k\|_\square,$$

for each θ_3. This can be readily used to show the continuity of $\hat{\boldsymbol{\tau}}$ with respect to the cut metric.

More generally we can take any finite simple graph $H \in \mathcal{G}_m$ and define $\boldsymbol{\tau}^H : \mathcal{K}_\ell \to \mathbb{R}$ by

$$\boldsymbol{\tau}^H(k) = \int_{[0,1]^{\# H}} \prod_{\{i,j\} \in H} k(\theta_i, \theta_j) \prod_{r=1}^m d\theta_r.$$

Again $\boldsymbol{\tau}^H$ is continuous with respect to the cut metric. For example, if we take the cyclic graph

$$C_m = \big\{\{1,2\},\{2,3\},\dots,\{m-1,m\},\{m,1\}\big\},$$

then

$$\boldsymbol{\tau}^{C_m}(k) = \int_{[0,1]^m} k(\theta_1,\theta_2)k(\theta_2,\theta_3)\dots k(\theta_{m-1},\theta_m)k(\theta_m,\theta_1)\prod_{r=1}^m d\theta_r = \sum_{\lambda\in\sigma(k)} \lambda^m,$$

where $\sigma(k)$ denotes the spectrum of the Hilbert-Schmidt operator

$$\bar{k}(f)(\theta) = \int_0^1 k(\theta,\theta')f(\theta')\ d\theta'.$$

Note that in the case of the Erdős-Rényi graph, the expression

$$\binom{n}{m}\boldsymbol{\tau}^H\big(k_n(\cdot,\cdot;X_n(G))\big), \qquad G\in\mathcal{G}_n,$$

counts the number of subgraphs of G that are isomorphic to H.

On account of the continuity of the function $\boldsymbol{\tau}^H$, it is desirable to establish a LDP for the sequence $\{\mathbb{Q}_n\}$ of probability measures that are now defined on the metric space $(\mathcal{K}_\ell, d_\square)$.

Theorem 8.2 *The family $\{\mathbb{Q}_n\}$ satisfies a weak LDP with the rate function $\mathcal{I}:\mathcal{K}_\ell\to[0,\infty]$.*

Proof Let us write $\mathcal{B}_{sym}$ for the space of bounded Borel symmetric functions $f:[0,1]^2\to\mathbb{R}$. For each $f\in\mathcal{B}_{sym}$, define $L_f(k)=\int_{[0,1]^2} fk\ d\theta$, where $d\theta = d\theta_1 d\theta_2$. Note that L_f is bounded and continuous with respect to the metric $d_\square$. (In fact L_f is even weakly continuous.) Set $\mathcal{V}=\{L_f:\ f\in\mathcal{B}_{sym}\}$. By (8.6),

$$\Lambda(f) = \lim_{n\to\infty} n^{-2}\log\int e^{n^2 L_f(k)}\ \mathbb{Q}_n(dk) = \frac{1}{2}\int_{[0,1]^2}\lambda(2f)\ d\theta,$$

which implies the ULDP for compact sets by Theorem 2.1**(i)**.

In view of Theorem 2.1**(ii)-(iii)**, we would like to establish a LLN for the measures

$$d\mathbb{Q}_n^f = e^{L_f-\Lambda(f)}\ d\mathbb{Q}_n.$$

Write $\mathcal{S}_m$ for the set of functions that are constants on each interval of the form

$$J_{ij}^m := [(i-1)/m, i/m)\times[(j-1)/m, j/m),$$

for $i, j = 1, \dots, m$. For lower bounds, it suffices to verify

$$\lim_{n\to\infty} \mathbb{Q}_n^f = \delta_{\lambda'(f)}, \tag{8.7}$$

for every $f \in \mathcal{S}_m$ and $m \in \mathbb{N}$. This is because for an arbitrary f, we may take a finite partition of boxes and replace the value of f by its average on each box and get a sequence of simple $\{f_n\}$ such that

$$\lim_{n\to\infty} f_n = f, \qquad \limsup_{n\to\infty} \mathcal{I}(f_n) \geq \mathcal{I}(f).$$

This would allow us complete the proof of LLDP by applying Theorem 2.1**(iii)**.

It remains to verify (8.7). Note that we are using the $d_\square$ topology in (8.7); for the weak topology (8.7) is an immediate consequence of Theorem 8.1. Let us write $B_r^\square(g)$ for the set of k such that $d_\square(k, g) \leq r$. For (8.7), we need to show

$$\lim_{n\to\infty} \mathbb{Q}_n^f \left(\{k : \ k \notin B_r^\square(\lambda'(f))\}\right) = 0, \tag{8.8}$$

for every $f \in \mathcal{S}_m$. If we write D^J for the restriction of the metric $d_\square$ to the space of functions that are defined on J, we have the inequality

$$d_\square(k, g) \leq \sum_{i,j=1}^m D^{J_{ij}^m}\left(k_{J_{ij}^m}, g_{J_{ij}^m}\right),$$

where k_J denotes the restriction of k to J. As a result, $d_\square(k, g) \geq r$ implies that for some $J = J_{ij}^m$, we have $D^J(k_J, g_J) \geq rm^{-2}$. From this we learn that practically we may assume that $m = 1$ and f is constant in (8.8). From now on, we assume that f is constant that can be assume to be 0 without loss of generality. In summary, it suffices to establish a LLD for the sequence $\{\mathbb{Q}_n\}$ with respect the cut metric.

Let us write $m = \int x \ \mu(dx)$. We wish to show

$$\lim_{n\to\infty} \mathbb{U}_n^\mu \left(\{X : \ k_n(X) \notin B_r^\square(m)\}\right) = 0, \tag{8.9}$$

Note that we always have

$$d_\square(k, k') \leq 4 \sup_{A,B\in\mathcal{B}} \left| \int_{A\times B} (k - k') \ d\theta \right|, \tag{8.10}$$

where $\mathcal{B}$ denotes the set of Borel subsets of $[0, 1]$. On the other hand, if both k and k' are constants, the supremum in (8.9) can be restricted to sets $A, B \in \mathcal{B}_n$, where $\mathcal{B}_n$ is the σ-algebra generated by the of intervals $[(i-1)/n, i/n]$, $i = 1, \dots, n$, with $i < j$ and $i, j \in \{0, 1, \dots, n\}$. As a result,

$$\mathbb{U}_n^\mu \left(\{X : \ k_n(X) \notin B_r^\square(m)\}\right) \leq 2^{2n} \sup_{A,B\in\mathcal{B}_n} \mathbb{U}_n^\mu \left(\left\{X : \ \left| \int_{A\times B} (k_n(X) - m) \ d\theta \right| \geq 4r \right\}\right).$$

From this we learn that for (8.9), we only need to show

$$\limsup_{n\to\infty} n^{-2}\log \sup_{A,B\in\mathcal{B}_n} \mathbb{U}_n^{\mu}\left(\left\{X:\ \left|\int_{A\times B}(k_n(X)-m)\ d\theta\right|\geq 4r\right\}\right)<0. \tag{8.11}$$

This is a ULDP with the weak topology that is uniform on the sets $A\times B\in\mathcal{B}_n$. We can readily verify (8.11) by Chebyshev's inequality, in just the same way we prove Cramér's ULDP. □

In the case of the weak topology, a weak LDP implies a strong LDP because $\mathcal{K}_\ell$ is compact with respect to the weak topology. This is no longer the case if we use the cut metric $d_\square$ (see Exercise 8.1 below). We now employ a trick that would allow us to regain the compactness that in turn would facilitate a LDP for a suitable quotient of the metric space $(\mathcal{K}_\ell, d_\square)$.

Even though the labeling of vertices plays no role in our combinatorial questions, it does play a role in the very definition of k_n. For this reason, we only need a LDP that is insensitive to a relabeling of vertices. In large n limit, a relabeling becomes a measure preserving change of coordinates. More precisely, if $\gamma:[0,1]\to[0,1]$ is a Lebesgue measure preserving bijection, then we want to identify k with k^γ that is defined by

$$k^\gamma(\theta_1,\theta_2)=k(\gamma(\theta_1),\gamma(\theta_2)).$$

Let us write Γ for the set of such η and define equivalence classes

$$[k]:=\big\{k^\gamma:\ \gamma\in\Gamma\big\},\qquad k\in\mathcal{K}_\ell.$$

The set of all equivalence classes is denoted by $\widetilde{\mathcal{K}}_\ell$. Naturally the cut norm $d_\square$ induces a (pseudo)metric

$$\delta_\square([k],[k'])=\inf_{\gamma\in\Gamma} d_\square(k^\gamma,k'),$$

on $\widetilde{\mathcal{K}}_\ell$. Since $\|k\|_\square=\|k^\gamma\|_\square$, this metric is well defined. According to a fundamental theorem of Lovász and Szegedy [LS], the metric space $(\widetilde{\mathcal{K}}_\ell,\delta_\square)$ is compact. This compactness is a consequence of a deep regularity lemma of Szemerédi that practically allows us to verify total boundedness of this metric space. Before stating this lemma, let us make a couple more comments so that we can state the LDP of [CV].

We note that the LD rate function $\mathcal{I}$ can be easily defined on $\widetilde{\mathcal{K}}_\ell$ because $\mathcal{I}(k)=\mathcal{I}(k^\gamma)$ for every $\gamma\in\Gamma$. We abuse the notation and write $\mathcal{I}$ for the resulting function on $\widetilde{\mathcal{K}}_\ell$. Also, the map $k\mapsto[k]$ pushes forward $\mathbb{Q}_n$ into a probability measure $\widetilde{\mathbb{Q}}_n$ on the space $\widetilde{\mathcal{K}}_\ell$. We are now ready to state our LDP:

Theorem 8.3 *The family $\{\widetilde{\mathbb{Q}}_n\}$ satisfies a LDP with the rate function $\mathcal{I}:\widetilde{\mathcal{K}}_\ell\to[0,\infty]$.*

The idea behind the LDP of Theorem 8.3 is that if $\widetilde{C}$ is a closed set in $\widetilde{\mathcal{K}}_\ell$, then it is compact. If we lift $\widetilde{C}$ to a closed subset C of $\mathcal{K}_\ell$, even though C is not necessarily compact, it has a compact nucleus that produces other members of C by measure preserving transformations. These transformations are simply permutations at the n-th level. Since there are $n!$ many such transformations, and $n^{-2}\log n! \to 0$ in large n-limit, the set C is as good as a compact set for which we already have LDP.

As we mentioned earlier, the main ingredient for the proof of Theorem 8.3 is a powerful regularity lemma of Szemerédi. For our purposes we state a variant of this lemma that would give us a practical way of proving the total boundedness of $\widetilde{\mathcal{K}}_\ell$. As a preparation, let us recall that $\mathcal{B}_n$ is the σ-algebra generated by the intervals $J_i^n : i = 1, \dots, n$. We also write $\mathcal{B}_n^2$ for the σ-algebra generated by the boxes $J_{ij}^n : i, j = 1, \dots, n$. The space of permutations π of $\{1, \dots, n\}$ is denoted by $\Pi(n)$. Given $k = \sum_{i,j} x_{ij} 1\!\!1_{J_{ij}^n} \in \mathcal{B}_n^2$, we write k^π for $\sum_{i,j} x_{\pi(i)\pi(j)} 1\!\!1_{J_{ij}^n}$. We are now ready to state our main lemma.

Lemma 8.1 *For each $\varepsilon > 0$, there exists a compact subset $\mathcal{K}_\ell^\varepsilon$ of $\mathcal{K}_\ell$ and $n_0 = n_0(\varepsilon)$ such that if $n \geq n_0$, then*

$$\mathcal{K}_{n,\ell} := \mathcal{K}_\ell \cap \mathcal{B}_n^2 \subseteq \bigcup_{\pi \in \Pi(n)} \left\{ k \in \mathcal{B}_n^2 : \ \delta_\square \left(k^\pi, \mathcal{K}_\ell^\varepsilon\right) \leq \varepsilon \right\} =: \bigcup_{\pi \in \Pi(n)} \mathcal{K}_{n,\ell}^{\varepsilon,\pi}.$$

Armed with this lemma, we can readily prove Theorem 8.3.

Proof of Theorem 8.3 The lower bound LDP follows from the lower bound of Theorem 8.2. As for the upper bound, let $\widetilde{C}$ be a closed subset of $\widetilde{\mathcal{K}}_\ell$ and set

$$C = \cup\{[k] : \ [k] \in \widetilde{C}\}, \qquad C_n = C \cap \mathcal{B}_n^2 \subseteq \mathcal{K}_{n,\ell}.$$

To bound

$$\widetilde{\mathbb{Q}}_n(\widetilde{C}) = \mathbb{Q}_n(C_n),$$

pick $\varepsilon > 0$ and use Lemma 8.1 to assert

$$C_n \subseteq \cup_{\pi \in \Pi(n)} \left(\mathcal{K}_{n,\ell}^{\varepsilon,\pi} \cap C_n\right).$$

Since $\mathcal{K}_{n,\ell}^{\varepsilon,\pi}$ is the ε-neighborhood of the compact set $\mathcal{K}_\ell^\varepsilon$, we can find a finite subset $A \subseteq \mathcal{K}_\ell^\varepsilon$ such that

$$\mathcal{K}_\ell^\varepsilon \subseteq \bigcup_{f \in A} B_\varepsilon^\square(f).$$

Note that the set A is independent of n. As a result, we can find a subset $A' \subset C_n$ such that $\#A = \#A'$ and

$$C_n \subseteq \bigcup_{\pi \in \Pi(n)} \left(\mathcal{K}_{n,\ell}^{\varepsilon,\pi} \cap C_n\right) \subseteq \bigcup_{\pi \in \Pi(n)} \bigcup_{f \in A'} B_{2\varepsilon}^\square(f^\pi).$$

In fact $A' = \{f' : \ f \in A\}$, where f' is chosen so that $f' \in C_n \cap B_\varepsilon^\square(f)$, whenever this intersection is nonempty. Hence

$$\widetilde{\mathbb{Q}}_n(\widetilde{C}) = \mathbb{Q}_n(C_n) \le n!(\sharp A) \sup_{\pi \in \Pi(n)} \sup_{f \in A'} \mathbb{Q}_n\left(B_\varepsilon^\square(f^\pi)\right) = n!(\sharp A) \sup_{f \in A'} \mathbb{Q}_n\left(B_\varepsilon^\square(f)\right). \tag{8.12}$$

On the other hand, since the set $B_\varepsilon^\square(f)$ is weakly closed, we may apply Theorem 8.1 to assert

$$\lim_{\varepsilon \to \infty} \limsup_{n \to \infty} n^{-2} \log \mathbb{Q}_n(B_\varepsilon^\square(f)) \le -\mathcal{I}(f).$$

From this and (8.12) we learn

$$\lim_{\varepsilon \to \infty} \limsup_{n \to \infty} n^{-2} \log \mathbb{Q}_n(C_n) \le -\inf_{f \in A'} \mathcal{I}(f) \le -\inf_{C} \mathcal{I}.$$

This completes the proof of the upper bound. □

Exercise 8.1

(i) Let $k_n(\theta_1, \theta_2) = k_n(\theta) = \sum_{m \in \mathbb{Z}^2} a_n(m) e^{i\theta \cdot m}$ be the Fourier expansion of k_n. Show that $k_n \to k$ weakly if and only if

$$\lim a_{n \to \infty}(m) = a(m),$$

exists for each $m \in \mathbb{Z}^2$. However, if $\|k_n - k\|_\square \to 0$, then

$$\lim_{n \to \infty} \sup_{m \in \mathbb{Z}^2} |a_n(m) - a(m)| = 0.$$

(ii) When $k(\theta_1, \theta_2) = \bar{k}(\theta_1 - \theta_2)$ for an even function $\bar{k}$, find an expression for $\hat{\boldsymbol{\tau}}(k)$. Use this expression to show that $\hat{\boldsymbol{\tau}}$ is not continuous in the weak topology.

(iii) Show that if $k(\theta_1, \theta_2) = A(\theta_1)A(\theta_2)$, then

$$\|k\|_\square = \left(\int_0^1 |A(\theta_1)| \, d\theta_1 \right)^2.$$

Show that the sequence $k_n(\theta_1, \theta_2) = \cos(2\pi n \theta_1)\cos(2\pi n \theta_2)$ converges to 0 weakly, but has no $d_\square$-convergent subsequence.

(iv) Verify (8.10) and (8.11). □

A Probability Measures on Polish Spaces

By a *Polish space* (E, d), we mean that d is a metric on E and E is separable and complete with respect to d. The corresponding σ-algebra of the Borel subsets of E is denoted by $\mathcal{B}(E)$ or simply $\mathcal{B}$. We write $B_r(x)$ for the open ball of center x and radius r in E. The space of bounded continuous functions $f : E \to \mathbb{R}$ is denoted by $C_b(E)$. This space is equipped with the norm $\|f\| = \sup_{x\in E} |f(x)|$. We also write $U_b(E)$ for the space of bounded uniformly continuous functions. The space of Radon probability measures on E is denoted by $\mathcal{M} = \mathcal{M}(E)$. Given a sequence of probability measures, we say μ_n converges (weakly) to μ, or simply $\mu_n \Rightarrow \mu$ if

$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu, \tag{A.1}$$

for every $f \in C_b(E)$. Here are some equivalent definitions for weak convergences of measures:

Theorem A.1 *Let $\{\mu_n\}$ be a sequence in $\mathcal{M}(E)$ with E a Polish space. Then the following statements are equivalent:*

(i) $\mu_n \Rightarrow \mu$.

(ii) (A.1) *holds for every* $f \in U_b(E)$.

(iii) $\limsup_{n\to\infty} \mu_n(C) \le \mu(C)$ *for every closed set* C.

(iv) $\liminf_{n\to\infty} \mu_n(U) \ge \mu(U)$ *for every open set* U.

(v) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for every set* $A \in \mathcal{B}$ *such that* $\mu(\partial A) = 0$.

Proof The only non-trivial parts to show are that **(ii)** implies **(iii)** and that **(v)** implies **(i)**. For the former observe that if $d(x, C) = \inf\{d(x, y) : \ y \in C\}$, then

$$f_k(x) = (1 + d(x, C))^{-k},$$

is a sequence of uniformly (even Lipschitz) bounded continuous functions such that $f_k \downarrow 1\!\!1_C$. Hence

$$\mu(C) = \lim_{k\to\infty} \int f_k \, d\mu = \lim_{k\to\infty} \lim_{n\to\infty} \int f_k \, d\mu_n \ge \limsup_{n\to\infty} \mu_n(C).$$

To deduce **(i)** from **(v)**, approximate $f \in C_b(E)$ with

$$s_k = \sum_{i=-\infty}^{\infty} a_i \; 1\!\!1 \left(f \in [a_{i-1}, a_i)\right),$$

where $(a_i : i \in \mathbb{Z})$ are selected so that $0 < a_i - a_{i-1} \leq k^{-1}$, and $\mu(\{f = a_i\}) = 0$ for every $i \in \mathbb{Z}$. From **(v)**, we know that $\lim_{n\to\infty} \int s_k \, d\mu_n = \int s_k \, d\mu$. This implies **(i)** because

$$\int |s_k - f| \, d\mu_n, \int |s_k - f| \, d\mu \leq k^{-1}.$$

□

Theorem A.2 *If (E, d) is a Polish space, then there exists a metric d' on E that induces the same topology on E as the metric d, and the completion of E with respect to d' is compact. In particular, the space $U_b(E)$ for the metric d' is separable.*

Proof First assume that E is compact and choose a countable dense subset $\{x_n\}$ of E. Then choose a continuous function $g_{n,m}$ that has support in $B_{2/m}(x_n)$ and is 1 in $B_{1/m}(x_n)$ for each $n, m \in \mathbb{N}$. Let $\mathcal{A} = \{g_{n,m} : \ n, m \in \mathbb{N}\}$ and write $\mathcal{A}'$ for the set of polynomials of rational coefficients and variables in $\mathcal{A}$. Then apply Stone-Weierstrass Theorem to show that $\mathcal{A}'$ is dense in $C(E)$.

When E is a Polish space, by Urysohn-Tychonov type embeddings, we can embed E into $E' = [0, 1]^{\mathbb{N}}$, where E' is equipped with the product topology. More precisely, since E is a separable metric space, we may choose a countable base of open sets $\{U_n : \ n \in \mathbb{N}\}$ and continuous functions $f_n : E \to [0, 1]$, $n \in \mathbb{N}$ such that $U_n^c = \{x : \ f_n(x) = 0\}$, for every $n \in \mathbb{N}$. Then $\Phi : E \to [0, 1]^{\mathbb{N}}$, defined by $\Phi(x) = (f_n(x) : n \in \mathbb{N})$ is the desired embedding. Clearly Φ is an injective continuous function. On the other hand, for any open set $U \subseteq E$, the set $\Phi(U)$ is open in E': If $\bar{y} \in \Phi(U)$, choose $\bar{x}$ and $n \in \mathbb{N}$ such that $\Phi(\bar{x}) = \bar{y}$, and $\bar{x} \in U_n \subseteq U$, so that $\bar{y} \in V \cap \Phi(E) \subseteq \Phi(U)$ for open set $V = \{y = (y_i : i \in \mathbb{N}) : \ y_n > 0\}$.

From the embedding Φ, we learn that E is homeomorphic to a subspace $E'' = \Phi(E)$ of the compact space E'. Since the space E' is metrizable and compact; the closure $\hat{E} = \bar{E}''$ of E'' in E' is a compact metric space. In other words, equip E' with the product metric $\bar{d}$ so that (E', d') is a compact metric space. We then take the completion of E'' with respect to $\bar{d}$ to obtain the compact metric space $(\hat{E}, \bar{d})$. Clearly $U_b(\hat{E})$ is homeomorphic to $C_b(\hat{E}) = C(\hat{E})$ because each $f \in U_b(E)$ has a continuous extension to $\hat{E}$. Since $C(\hat{E})$ is separable, we deduce that $U_b(E)$ is separable. □

We learn from the proof of Theorem A.2 that any Polish space E is homeomorphic with a subset of the compact metric space $[0, 1]^{\mathbb{N}}$. On the other hand, the measure space $(E, \mathcal{B})$ is homeomorphic to a subspace of $\{0, 1\}^{\mathbb{N}}$.

Theorem A.3 *Equip $\bar{E} = \{0, 1\}^{\mathbb{N}}$ with the product topology and write $\bar{\mathcal{B}}$ for its Borel σ–algebra. Let $\mathcal{F} = \sigma\big(\{A_n : n \in \mathbb{N}\}\big)$ be a countably generated σ-algebra.*

(i) *The map*

$$F(x) := (1\!\!1_{A_n} : n \in \mathbb{N}),$$

satisfies $F^{-1}\big(\bar{\mathcal{B}}\big) = \mathcal{F}$.

(ii) *We have* $A_{\mathcal{F}}(x) = A'_{\mathcal{F}}(x) \in \mathcal{F}$, *where*

$$A_{\mathcal{F}}(x) = \cap\{A: \ x \in A \in \mathcal{F}\}, \qquad A'_{\mathcal{F}}(x) = \cap_{n=1}^{\infty}\{A_n: \ x \in A_n\}.$$

Moreover

$$\{A_{\mathcal{F}}(x): \ x \in E\} = \left\{F^{-1}(\bar{x}): \ \bar{x} \in F(E)\right\}.$$

(iii) *In the case* $\mathcal{F} = \mathcal{B} = \mathcal{B}(E)$, *the map* F *is injective.*

Proof (i) Write $D_n \subset \bar{E}$, for the set of $(x_n : n \in \mathbb{N}) \in \bar{E}$ such that $x_m = 1$. We can readily show that $\bar{\mathcal{B}} = \sigma\big(\{D_n : n \in \mathbb{N}\}\big)$. Since $F^{-1}(D_n) = A_n$, we deduce that $F^{-1}\big(\bar{\mathcal{B}}\big) = \mathcal{F}$.

(ii) By definition $A'_{\mathcal{F}}(x) \in \mathcal{F}$, and

$$F(x) = F(y) \Leftrightarrow A'_{\mathcal{F}}(x) = A'_{\mathcal{F}}(y).$$

On the other hand, if $x \in A \in \mathcal{F}$, then $A = F^{-1}(\bar{A})$ for some $\bar{A} \in \mathcal{E}$ by Part **(i)**. As a result, $F(x) \in \bar{A}$ and $F^{-1}\big(\{F(x)\}\big) \subseteq A$. This completes the proof of $A_{\mathcal{F}}(x) = A'_{\mathcal{F}}(x) \in \mathcal{F}$.

(iii) In the case of $\mathcal{F} = \mathcal{B}$, choose a countable dense set $\{x_n : n \in \mathbb{N}\}$ and select $\{A_n: \ n \in \mathbb{N}\} = \{B_{1/m}(1/n): \ m, n \in \mathbb{N}\}$. Since whenever $x \neq y$, we can find $n \neq m$ such that

$$x \in A_n \setminus A_m, \qquad y \in A_m \setminus A_m,$$

we learn that the corresponding F is injective. □

Remark A.1 We note that in general the map F is not surjective. In fact the set $F(E)$ may not be a Borel set. Nonetheless, it is true that the measure space $(E, \mathcal{B})$ is isomorphic to $(\bar{E}, \bar{\mathcal{B}})$. See for example [P]. □

Definition A.1 The sets $\big(A_{\mathcal{F}}(x): \ x \in E\big)$ of Theorem A.3 are called the *atoms* of $\mathcal{F}$. Note that the atoms are the "smallest" sets in $\mathcal{F}$. □

Example A.1 (i) When $\mathcal{F} = \mathcal{B}$, we have $A_{\mathcal{F}}(x) = \{x\}$.

(ii) Let $\mathcal{E} = E^{\mathbb{Z}} = \{\mathbf{x} = (x_i: \ i \in \mathbb{Z}): \ x_i \in E \text{ for } i \in \mathbb{Z}\}$, and write $\mathcal{B}^j$ for the σ-algebra generated by $(x_i: \ i \leq j)$. In other words, a Borel function F is $\mathcal{B}^j$-measurable iff it depends on coordinates $(x_i: \ i \leq j)$ only. Then

$$A_{\mathcal{B}^j}(\mathbf{x}) = \{\mathbf{y} = (y_i: \ i \in \mathbb{Z}) \in \mathcal{E}: \ y_i = x_i \text{ for } i \leq j\}.$$

□

B Conditional Measures

Let $(E, \mathcal{B})$ be a Polish measure space i.e. E is a Polish space and $\mathcal{B}$ its Borel σ-field. Given $Q \in \mathcal{M}(E)$ and $p \geq 1$, we write $L^p(Q)$ for the space of $\mathcal{B}$-measurable functions $f : E \to \mathbb{R}$ such that $\int |f|^p \, dQ < \infty$. The space of Q-essentially bounded functions is denoted by $L^\infty(Q)$. Given $\mathcal{F}$ a sub σ-algebra of $\mathcal{B}$, and $p \in [1, \infty]$, we write $L^p(Q; \mathcal{F})$ for the set of $f \in L^p(Q)$ such that f is also $\mathcal{F}$-measurable. The conditional expectation of $f \in L^1(Q)$ with respect to $\mathcal{F}$ is denoted by $g = \mathbb{E}^Q(f|\mathcal{F})$ and is the unique $g \in L^1(Q; \mathcal{F})$ such that for any $h \in L^\infty(Q; \mathcal{F})$,

$$\int hf \, dQ = \int hg \, dQ.$$

When $f = \mathbb{1}_A$, we simply write $Q(A|\mathcal{F})$ for $\mathbb{E}^Q(f|\mathcal{F})$.

There are two ways to prove the existence of a conditional expectation. For example, for square integrable functions, we may regard $L^2(Q; \mathcal{F})$ as a closed linear subspace of $L^2(Q)$ and $\mathbb{E}^Q(f|\mathcal{F})$, is simply the orthogonal projection of f onto $L^2(Q; \mathcal{F})$. For $f \in L^1(Q)$, set $Q' = fQ$ and regard Q' as a probability measure on $\mathcal{F}$. Since $Q' \ll Q$, by Radon-Nikodym Theorem we can find $g \in L^1(Q; \mathcal{F})$ such that for every $A \in \mathcal{F}$, we have $Q'(A) = \int g \, dQ$.

It is straightforward to show that $\mathbb{E}^Q(f + f'|\mathcal{F}) = \mathbb{E}^Q(f|\mathcal{F}) + \mathbb{E}^Q(f'|\mathcal{F})$, Q-almost surely. Equivalently, if A and $A' \in \mathcal{B}$ are disjoint,

$$Q(A \cup A'|\mathcal{F}) = Q(A|\mathcal{F}) + Q(A'|\mathcal{F}),$$

Q-almost surely. This suggests that perhaps we can construct a nice version of $\{Q(A|\mathcal{F}) : A \in \mathcal{B})\}$ such that $Q(\cdot|\mathcal{F})(x)$ is a probability measure for Q-almost all $x \in E$. If such a version can be constructed, will be called a conditional probability measure. We state and prove a theorem that would guarantee the existence of such conditional measures.

Theorem B.1 *Let $(E, \mathcal{B})$ be a Polish measure space, $\mathcal{F} \subseteq \mathcal{B}$ is a σ-algebra, and $Q \in \mathcal{M}(E)$.*

(i) *Then there exists a family $\{Q_x : x \in E\} \subseteq \mathcal{M}(E)$ such that*

- *For every $A \in \mathcal{B}$, the function $x \mapsto Q_x(A)$ is $\mathcal{F}$-measurable;*
- *For every $A \in \mathcal{B}$ and $B \in \mathcal{F}$, we have $Q(A \cap B) = \int_B Q_x(A) \, Q(dx)$.*

In other words, $Q_x(A) = Q(A|\mathcal{F})(x)$ for Q-almost all x.

(ii) *If we write $Q^{\mathcal{F}}$ for $Q_{|\mathcal{F}}$, then*

$$Q(dy) = \int Q_x(dy) \, Q^{\mathcal{F}}(dx).$$

(iii) *If $\mathcal{F}$ is countably generated, then $Q\big(x, A_{\mathcal{F}}(x)\big) = 1$, where $(A_{\mathcal{F}}(x) : x \in E)$ are the atoms of $\mathcal{F}$ (see Definition A.1).*

Proof (ii) For any $A \in \mathcal{F}$, we know that $Q(x, A) = Q(A|\mathcal{F})(x) = 1\!\!1_A(x)$, Q-almost surely. Now if $\mathcal{F} = \sigma(\{A_n : n \in \mathbb{N}\})$, then from Q-almost sure equality $Q(x, A_n) = 1\!\!1_{A_n}(x)$, for each $n \in \mathbb{N}$ we deduce

$$Q\left(x, A'_{\mathcal{F}}(x)\right) = 1,$$

where $A'_{\mathcal{F}}(x) = \cap_{n=1}^{\infty}\{A_n : x \in A_n\}$. From this and Theorem A.3**(ii)** we conclude that $Q\big(x, A_{\mathcal{F}}(x)\big) = 1$.

Example B.1 (i) If $\mathcal{B} = \mathcal{F}$, then we may choose $Q_x = \delta_x$.

(ii) If $Q(A) = 0$ or 1 for every $A \in \mathcal{F}$, then $Q_x = Q$.

(iii) Let $\mathcal{E} = E^{\mathbb{Z}}$, and write $\mathcal{B}^j$ for the σ-algebra generated by $(x_i : i \le j)$. Since E is a Polish space, the σ-algebra $\mathcal{F}$ is coubtably generated. Then by Example A.1 and Theorem B.1**(ii)**, $Q^j(\mathbf{x}, E^j(\mathbf{x})) = 1$, where $(Q^j(\mathbf{x}, \cdot) : \mathbf{x} \in \mathcal{E})$ denote the conditional measures with respect to $\mathcal{B}^j$, and

$$E^j(\mathbf{x}) := \{\mathbf{y} = (y_i : i \in \mathbb{Z}) \in \mathcal{E} : y_i = x_i \text{ for } i \le j\}.$$

Let us write $\mathcal{E}^j := \{\mathbf{x}^j = (x_i : i \le j) : x_i \in E \text{ for } i \le j\}$, and for $\mathbf{x}^j \in \mathcal{E}^j$, we set

$$\mathcal{E}^j(\mathbf{x}^j) := \{(\mathbf{x}^j, \mathbf{x}_j) : \mathbf{x}_j = (x_i : i > j) : x_i \in E \text{ for } i > j\} = \{\mathbf{x}^j\} \times \mathcal{E}^j.$$

Since $\mathbf{x} \mapsto Q_{\mathbf{x}}$ is $\mathcal{B}^j$-measurable, we may regard it as a kernel of the form $Q(\mathbf{x}^j, d\mathbf{x}_j)$ that is defined on Borel σ-algebra of the Polish space $\mathcal{E}^j(\mathbf{x}^j)$. Moreover, for any Borel $F(\mathbf{x}) = F(\mathbf{x}^j, \mathbf{x}_j)$, we have

$$\int F \, dQ = \int \left[\int_{\mathcal{E}^j} F(\mathbf{x}^j, \mathbf{x}_j) \; Q(\mathbf{x}^j, d\mathbf{x}_j)\right] \; Q^j(d\mathbf{x}^j),$$

where $Q^j(d\mathbf{x}^j)$ denotes the the restriction of Q to the σ-algebra generated by $(x_i : i > j)$.

□

C Ergodic Theorem

Let $\mathcal{E}$ be a Polish space and $T : \mathcal{E} \to \mathcal{E}$ a homeomorphism. The pair $(\mathcal{E}, T)$ is an example of a discrete dynamical system. A central question for the system $(\mathcal{E}, T)$ is the long term behavior of the orbits $(T^n(\mathbf{x}) : n \in \mathbb{N})$ for $\mathbf{x} \in \mathcal{E}$.

We write $\mathcal{M}_S = \mathcal{M}_S^T(\mathcal{E})$ for the space of invariant measure of the system $(\mathcal{E}, T)$:

$$\text{(C.1)} \quad \mathcal{M}_S = \mathcal{M}_S(\mathcal{E}) := \left\{Q \in \mathcal{M}(\mathcal{E}) : \int F \circ T \, dQ = \int F \, dQ, \;\; \text{for every } F \in C_b(\mathcal{E})\right\}.$$

We can readily show that $\mathcal{M}_S = \mathcal{M}_S^T(\mathcal{E})$ is a convex closed subset of $\mathcal{M}(\mathcal{E})$. Hence, the space $\mathcal{M}_S$ is also a Polish space.

As before, the σ-algebra of Borel subsets of $\mathcal{E}$ is denoted by $\mathcal{B}$. the σ-algebra of T-invariant Borel sets is denoted by $\mathcal{I}^T$:

$$\mathcal{I}^T = \{A \in \mathcal{B} : A \text{ is a Borel set and } T(A) = A\}.$$

We also write $\mathcal{M}_{er}$ for the space of ergodic invariant measures. That is, $Q \in \mathcal{M}_{er}$ if $Q \in \mathcal{M}_S$, and if $A \in \mathcal{I}^T$, then $Q(A) = 1$ or 0. One can show that $\mathcal{M}_{er}$ is exactly the set of extreme points of the convex set $\mathcal{M}_S$. That is, if $Q \in \mathcal{M}$, then $Q \in \mathcal{M}_{er}$ iff whenever $Q = tQ_1 + (1-t)Q_2$ for some $Q_1, Q_2 \in \mathcal{M}_S$, then $Q = Q_1 = Q_2$.

According to the Boltzmann Ergodic Hypothesis, the time average of an observable over an orbit is approximately equal to its ensemble average. A rigorous formulation of the Boltzmann Hypothesis is given by the Birkhoff Ergodic Theorem. By time averages, we are referring to the integration with respect to the empirical measures

$$\nu_n(\mathbf{x}) = n^{-1}\left(\delta_{\mathbf{x}} + \delta_{T(\mathbf{x})} + \cdots + \delta_{T^{n-1}(\mathbf{x})}\right).$$

For a compact notation we also regard $\nu_n(\mathbf{x})$ as an operator that is acting on $C_b(\mathcal{E})$:

$$\nu_n(\mathbf{x})(F) = \int F \, d\nu_n(\mathbf{x}) = n^{-1}\left(F(\mathbf{x}) + F(T(\mathbf{x})) + \cdots + F(T^{n-1}(\mathbf{x}))\right).$$

It is straightforward to show that any limit point of $\{\nu_n(\mathbf{x})\}$ belong to $\mathcal{M}_S$.

Theorem C.1 *(Birkhoff) For any $F \in C_b(\mathcal{E})$, and $Q \in \mathcal{M}_S$,*

$$Q\left(\left\{\mathbf{x} : \lim_{n\to\infty} \nu_n(\mathbf{x})(F) = P_T F\right\}\right), \tag{C.2}$$

where $P_T F = \mathbb{E}^Q(F|\mathcal{I}^T)$ is the conditional expectation of F, given the σ-algebra $\mathcal{I}^T$. Moreover if $F \in L^q(Q)$ for some $q \in [1,\infty)$, then

$$\lim_{n\to\infty} \int |\nu_n(\mathbf{x})(F) - P_T F|^q \, dQ = 0. \tag{C.3}$$

Using a countable dense set of bounded uniformly continuous functions, we use Ergodic Theorem to assert that for any $Q \in \mathcal{M}_S$,

$$Q\left(\left\{\mathbf{x} \in \mathcal{E} : \lim_{n\to\infty} \nu_n(\mathbf{x}) = Q^T_{\mathbf{x}}\right\}\right) = 1, \tag{C.4}$$

where $Q^T_{\mathbf{x}}(d\mathbf{y})$ is a conditional measure of Q with respect to the σ-algebra $\mathcal{I}^T$. In particular if $Q \in \mathcal{M}_{er}$, then

$$Q\left(\left\{\mathbf{x} \in \mathcal{E} : \lim_{n\to\infty} \nu_n(\mathbf{x}) = Q\right\}\right) = 1, \tag{C.5}$$

As the following result of Oxtoby [Ox] indicates, we can construct a universal set on which the limit of the empirical measures exists.

Theorem C.2 *There exists a Borel set $\mathcal{E}_0 \in \mathcal{I}^T$ such that $Q(\mathcal{E}_0) = 1$ for every $Q \in \mathcal{M}_S$, and if $\mathbf{x} \in \mathcal{E}_0$, then*

$$R^T(\mathbf{x}) := \lim_{n\to\infty} \nu_n(\mathbf{x}) \quad \textit{exists and is ergodic.} \tag{C.6}$$

Moreover, for $Q \in \mathcal{M}_{er}$,

$$Q\left(\left\{\mathbf{x} \in \mathcal{E}_0 : \ R^T(\mathbf{x}) = Q^T_{\mathbf{x}}\right\}\right) = 1, \tag{C.7}$$

where $Q^T_{\mathbf{x}}$ is the Q-conditional measure, given $\mathcal{I}^T$. In particular, for $Q \in \mathcal{M}_{er}$,

$$Q\left(\left\{\mathbf{x} \in \mathcal{E}_0 : \ R^T(\mathbf{x}) = Q\right\}\right) = 1. \tag{C.8}$$

Proof Set $\mathcal{E}_1$ to be the set of $\mathbf{x} \in \mathcal{E}$ for which that the limit in (C.6) exists. Since each ν_n is a continuous function, the set $\mathcal{E}_1$ is a Borel (even $F_{\sigma\delta}$) set and belongs to $\mathcal{I}^T$. We then define $R^T : \mathcal{E}_1 \to \mathcal{M}_S$ as in (C.6). Evidently the map R^T is measurable. By (C.4), we have $Q(\mathcal{E}_1) = 1$ for every $Q \in \mathcal{M}_S$. To figure out how to define $\mathcal{E}_0$, let us examine the ergodicity of $R^T(\mathbf{x})$. Define $P_T F : \mathcal{E}_1 \to \mathbb{R}$ by $P_T F(x) = \int F \ dR^T(\mathbf{x})$. We now take countable dense set $\mathcal{C}$ of functions $F \in U_b(\mathcal{E})$ and define $\mathcal{E}_0$ to be the set of $x \in \mathcal{E}_1$ such that

$$R_T(F)(x) := \int_{\mathcal{E}_1} \left[P_T F(\mathbf{y}) - P_T F(\mathbf{x})\right]^2 \ R^T(\mathbf{x})(d\mathbf{y}) = 0,$$

for every $F \in \mathcal{C}$. We claim that if $x \in \mathcal{E}_1$, then (C.6) holds true. This is because

$$R^T(\mathbf{x}) \left(\{\mathbf{y} : \ P_T F(\mathbf{y}) = P_T F(\mathbf{x})\}\right) = 1,$$

for $\mathbf{x} \in \mathcal{E}_0$ and $F \in \mathcal{C}$. This means that in the support of $R^T(\mathbf{x})$, the function $P_T F$ is constant for each $F \in \mathcal{C}$; so the measure $R^T(\mathbf{x})$ is ergodic.

We now check that $\mathcal{E}_1 \in \mathcal{I}^T$, and $Q(\mathcal{E}_1) = 1$ for every $Q \in \mathcal{M}_S(\mathcal{E})$. For this it suffices to show

$$\bar{R}(F) := \int R_T(F)(\mathbf{x}) \ Q(d\mathbf{x}) = 0,$$

for every $F \in C_b(\mathcal{E})$ and $Q \in \mathcal{M}_S(\mathcal{E})$. Indeed

$$\begin{aligned}
\bar{R}(F) &= \int \lim_{m\to\infty} \left[\int_{\mathcal{E}_1} \left[\int F\ d\nu_m(\mathbf{y}) - P_T F(\mathbf{x}) \right]^2 R^T(\mathbf{x})(d\mathbf{y}) \right] Q(d\mathbf{x}) \\
&= \int \lim_{m\to\infty} \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \left(\int F\ d\nu_m(T^i(\mathbf{x})) - P_T F(\mathbf{x}) \right)^2 Q(d\mathbf{x}) \\
&= \int \lim_{m\to\infty} \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \left(\int F\ d\nu_m - P_T F \right)^2 \circ T^i(\mathbf{x})\ Q(d\mathbf{x}) \\
&= \int \lim_{m\to\infty} P_T \left(\int F\ d\nu_m - P_T F \right)^2 Q(d\mathbf{x}) \\
&= \lim_{m\to\infty} \int P_T \left(\int F\ d\nu_m - P_T F \right)^2 Q(d\mathbf{x}) \\
&= \lim_{m\to\infty} \int \left(\int F\ d\nu_m - P_T F \right)^2 Q(d\mathbf{x}) \\
&= \int \lim_{m\to\infty} \left(\int F\ d\nu_m - P_T F \right)^2 Q(d\mathbf{x}) = 0,
\end{aligned}$$

completing the proof of (C.9). Here, we used

(i) the Bounded Convergence Theorem, for the first, fifth and seventh equalities;

(ii) the fact that $\int F\ d\nu_m \to P_T F$ on $\mathcal{E}_1$ as $m \to \infty$, for the first equality;

(iii) the definition of $R^T(\mathbf{x})$ for the second equality;

(iv) the invariance $R^T(\mathbf{x}) = Q^T(T(\mathbf{x}))$ for the third equality;

(v) Ergodic Theorem (C.2) for the forth and last equalities.

Finally (C.7) is an immediate consequence of (C.5) and the definition Q^T. $\square$

Corollary C.1 *(Choquet Theorem) For every $Q \in \mathcal{M}_S(\mathcal{E})$, we can find $\Theta \in \mathcal{M}\big(\mathcal{M}_{er}(\mathcal{E})\big)$ such that*

(C.9) $$Q = \int_{\mathcal{M}(\mathcal{M}_{er}(\mathcal{E}))} \alpha\ \Theta(d\alpha).$$

Proof By the definition of conditional measure, we always have

$$Q = \int Q^T_{\mathbf{x}}\ Q(d\mathbf{x}).$$

From (C.7),

$$Q = \int_{\mathcal{E}_0} R^T(\mathbf{x})\ Q(d\mathbf{x}).$$

We are done because $R^T(\mathbf{x}) \in \mathcal{M}_{er}(\mathcal{E})$. $\square$

D Minimax Principle

In many examples of interest, we use the Contraction Principle (Theorem 1.2**(i)**) to obtain new LDP. In view of (2.2), the expression we get for the rate function I' involves a supremum and an infimum. Sometimes we can simplify this expression by interchanging the infimum with the supremum. In this Appendix we state and prove a minimax principle (known as *Sion's minimax theorem*) that provides us with sufficient conditions under which we can perform such an interchange.

Let X and Y be a two topological vector spaces and $J : X' \times Y' \to \mathbb{R}$ with $X' \subseteq X$, $Y' \subseteq Y$.

Definition D.1 We say J satisfies the *minimax conditions* if the following statements are true:

(i) The set $X_y(a) = \{x :\ J(x,y) \le a\}$ is closed and convex for each $y \in Y$ and $a \in \mathbb{R}$.

(ii) The set $Y_x(a) = \{y :\ J(x,y) \ge a\}$ is closed and convex for each $x \in X$ and $a \in \mathbb{R}$. In other words, the function J is quasi-convex and lower semi-continuous (resp. concave and upper semi-continuous) in x-variable (resp. y-variable). The former means for $t \in [0,1]$,

$$\begin{aligned} &J(tx_1 + (1-t)x_2, y) \le \max\{J(x_1,y), J(x_2,y)\}, \\ \text{(resp.}\quad &J(x, ty_1 + (1-t)y_2) \ge \min\{J(x,y_1), J(x,y_2)\}.) \end{aligned} \tag{D.1}$$

$\square$

Theorem D.1 below is due to Sion. Its proof is adopted from [K].

Theorem D.1 *Assume that X' is compact and convex, Y' is convex and that $J : X' \times Y' \to \mathbb{R}$ satisfies the minimax conditions. Then*

$$\inf_{x\in X'} \sup_{y\in Y'} J(x,y) = \sup_{y\in Y'} \inf_{x\in X'} J(x,y). \tag{D.2}$$

Proof The proof of $\inf\sup \ge \sup\inf$ is trivial. As for the reverse inequality, let a be any number with

$$a < \inf_{x\in X'} \sup_{y\in Y'} J(x,y).$$

This means that $\cap_{y\in Y'}X_y(a)=\emptyset$. From this and the compactness of the sets $\{X_y : y\in Y'\}$, we learn that there are finitely many $y_1,\dots,y_k\in Y'$ such that $\cap_{i=1}^k X_{y_i}(a)=\emptyset$. Hence

$$a<\inf_{x\in X'}\sup_{1\le i\le k} J(x,y_i). \tag{D.3}$$

It remains to show that (D.3) implies that for some $\bar y\in Y'$, we have

$$a<\inf_{x\in X'} J(x,\bar y). \tag{D.4}$$

We prove this by induction on k. It is obvious when $k=1$. We now verify this for $k=2$. We hope to find $\bar y\in[y_1,y_2]:=\{ty_1+(1-t)y_2:\ t\in[0,1]\}$ such that (D.4) is true. Suppose to the contrary (D.4) fails to be true for all $\bar y\in[y_1,y_2]$ and we would like to arrive at a contradiction. To achieve this, pick a number b such that

$$a<b<\inf_{x\in X'}\max\{J(x,y_1),J(x,y_2)\}, \tag{D.5}$$

so that we have

$$X_{y_1}(b)\cap X_{y_2}(b)=\emptyset. \tag{D.6}$$

On the other hand, by (D.1),

$$X_{\bar y}(b)\subseteq X_{y_1}(b)\cup X_{y_2}(b), \tag{D.7}$$

for every $\bar y\in[y_1,y_2]$. Since we are assuming for now that $\inf_{x\in X'}J(x,\bar y)\le a$ for all $\bar y\in[y_1,y_2]$, we know that the closed sets $X_{\bar y}(b),X_{y_1}(b)$ and $X_{y_2}(b)$ are nonempty. So (D.7) and convexity (or even connectedness) of $X_{\bar y}(b)$ implies

$$X_{\bar y}(a)\subseteq X_{\bar y}(b)\subseteq X_{y_1}(b),\qquad \text{or}\qquad X_{\bar y}(a)\subseteq X_{\bar y}(b)\subseteq X_{y_2}(b). \tag{D.8}$$

Define

$$T_i=\{z\in[y_1,y_2]:\ X_z(a)\subseteq X_{y_i}(b)\},$$

for $i=1,2$. Certainly $y_1\in T_1$, $y_2\in T_2$, $T_1\cap T_2=\emptyset$ by (D.6), and $T_1\cup T_2=[y_1,y_2]$ by (D.8). We get a contradiction if we can show that both T_1 and T_2 are closed. For this, fix $i\in\{1,2\}$ and take a sequence $\{z_k\}$ in T_i with $z_k\to z$ in large k limit. By definition,

$$X_{z_k}(a)\subseteq X_{y_i}(b),$$

for every k. We wish to show that $X_z(a)\subseteq X_{y_i}(b)$. Pick $x_0\in X_z(a)$, so that $J(x_0,z)\le a<b$. By upper semi-continuity of $J(\cdot,y_i)$, we also have $J(x_0,z_k)<b$ for large k. That is, $x_0\in X_{z_k}(a)\subseteq X_{y_i}(b)$, as desired. This implies the closeness of both T_1 and T_2, and the contradiction we were seeking for. In summary when $k=2$, the inequality (D.3) implies the

existence of $\bar{y} \in Y'$ for which (D.4). For larger k, we argue by induction. If we already know how to deduce (D.4) from (D.3) for some $k \geq 2$, then in the case of $k+1$, set

$$X'' = \{x : \ J(x, y_{k+1}) \leq a\}.$$

and restrict J to $X'' \times Y$. If

$$a < \inf_{x \in X'} \sup_{1 \leq i \leq k+1} J(x, y_i). \tag{D.9}$$

and $X'' = \emptyset$, then (D.4) is true for $\bar{y} = y_{k+1}$ and we are done. Otherwise, apply the induction hypothesis to the function J, restricted to $X'' \times Y$: Since

$$a < \inf_{x \in X''} \sup_{1 \leq i \leq k} J(x, y_i),$$

by (D.9), we can find y' such that

$$a < \inf_{x \in X''} J(x, y').$$

This means

$$a < \inf_{x \in X'} \max\{J(x, y'), J(x, y_{k+1})\}.$$

We finally use our result for the case $k = 2$ to deduce (D.4) for some $\bar{y} \in Y'$. □

E Optimal Transport Problem

In the Monge transport problem, we search for a plan that minimizes the cost of transporting mass from a set of locations to another set of locations. In a general setting, we have two Polish spaces E_0 and E_1, and a Borel *cost function* $g : E_0 \times E_1 \to \mathbb{R}$. Given $\alpha_0 \in \mathcal{M}(E_0)$ and $\alpha_1 \in \mathcal{M}(E_1)$, we wish to minimize the total cost

$$\int g(x_0, T(x_0)) \ \alpha_0(dx_0), \tag{E.1}$$

over Borel functions $T : E_0 \to E_1$ such that $T^\sharp \alpha_0 = \alpha_1$. Here $T^\sharp : \mathcal{M}(E_0) \to \mathcal{M}(E_1)$ is defined by

$$T^\sharp \alpha_0(A_1) = \alpha_0 \left(T^{-1}(A_1)\right),$$

for every $A_1 \in \mathcal{B}(E_1)$. We may rewrite (E.1) as

$$\mathcal{C}(\gamma) := \int g(x_0, x_1) \ \gamma(dx_0, dx_1), \tag{E.2}$$

where $\gamma(dx_0, dx_1) = \delta_{T(x_0)}(dx_1)\alpha_0(dx_0)$. The type of measures that appear in the Monge optimization problem are characterize by two properties:

- If we write $\tau_0(\gamma)$ and $\tau_1(\gamma)$ for x_0 and x_1 marginals of γ respectively, we have that $\tau_0(\gamma) = \alpha_0$ and $\tau_1(\gamma) = \alpha_1$.
- The measure γ is supported on the graph of some Borel function $T : E_0 \to E_1$.

If we relax this optimization problem by dropping the second requirement, we obtain the Monge-Kantorovich transport problem:

$$\mathcal{D}(\alpha_0, \alpha_1) := \inf \left\{ \mathcal{C}(\gamma) : \ \gamma \in \Gamma(\alpha_0, \alpha_1) \right\}. \tag{E.3}$$

where

$$\Gamma(\alpha_0, \alpha_1) := \left\{ \mathcal{M}(E_0 \times E_1) : \ \tau_0(g) = \alpha_0, \tau_1(\gamma) = \alpha_1 \right\}.$$

To guarantee that $\mathcal{D}(\alpha_0, \alpha_1) > -\infty$, we assume that there are functions $a_0 \in L^1(\alpha_0)$ and $a_1 \in L^1(\alpha_1)$ such that

$$g(x_0, x_1) \geq a_0(x_0) + a_1(x_1). \tag{E.4}$$

According to Kantorovich, the variational problem (E.3) has dual formulations of the forms

$$\begin{aligned} \mathcal{D}^*(\alpha_0, \alpha_1) : &= \sup \left\{ \int f_0 \ d\alpha_0 + \int f_1 \ d\alpha_1 : \ (f_0, f_1) \in \mathcal{A}^*(\alpha_0, \alpha_1) \right\} \\ \mathcal{D}'(\alpha_0, \alpha_1) : &= \sup \left\{ \int f_0 \ d\alpha_0 + \int f_1 \ d\alpha_1 : \ (f_0, f_1) \in \mathcal{A}'(\alpha_0, \alpha_1) \right\}, \end{aligned} \tag{E.5}$$

where

$$\begin{aligned} \mathcal{A}^*(\alpha_0, \alpha_1) = \big\{ (f_0, f_1) \in C_b(E_0) \times C_b(E_1) : \ & g(x_0, x_1) \geq f_0(x_0) + f_1(x_1), \\ & \text{for all } (x_0, x_1) \in E_0 \times E_1 \big\}, \\ \mathcal{A}'(\alpha_0, \alpha_1) = \big\{ (f_0, f_1) \in L^1(\alpha_0) \times L^1(\alpha_1) : \ & g(x_0, x_1) \geq f_0(x_0) + f_1(x_1), \\ & \text{for } \alpha_0 \times \alpha_1 - \text{almost all } (x_0, x_1) \big\}. \end{aligned}$$

Theorem E.1 *Assume that* (E.4) *is true. Then* $\mathcal{D}(\alpha_0, \alpha_1) = \mathcal{D}'(\alpha_0, \alpha_1) = \mathcal{D}^*(\alpha_0, \alpha_1)$.

Proof *Step 1* We first assume that E_0 and E_1 are compact. We use Lagrange multipliers to drop the condition $\gamma \in \Gamma(\alpha_0, \alpha_1)$; we may write

$$\mathcal{D}(\alpha_0, \alpha_1) = \inf_{\gamma \in \mathcal{M}(E_0 \times E_1)} \ \sup_{f_0 \in C_b(E_0)} \ \sup_{f_1 \in C_b(E_1)} \left[\Lambda(\gamma, f_0, f_1) + \int f_0 \ d\alpha_0 + \int f_1 \ d\alpha_1 \right],$$

where

$$\Lambda(\gamma, f_0, f_1) = \int \big(g(x_0, x_1) - f_0(x_0) - f_1(x_1)\big)\ \gamma(dx_0, dx_1).$$

Since $E_0 \times E_1$ is compact, the space $\mathcal{M}(E_0 \times E_1)$ is compact. This allows us to apply Minimax Theorem to assert

$$\begin{aligned}\mathcal{D}(\alpha_0, \alpha_1) &= \sup_{f_0 \in C(E_0)} \sup_{f_1 \in C(E_1)} \inf_{\gamma \in \mathcal{M}(E_0 \times E_1)} \left[\Lambda(\gamma, f_0, f_1) + \int f_0\ d\alpha_0 + \int f_1\ d\alpha_1\right] \\ &= \sup_{f_0 \in C(E_0)} \sup_{f_1 \in C(E_1)} \left[\inf_{\gamma \in \mathcal{M}(E_0 \times E_1)} \Lambda(\gamma, f_0, f_1) + \int f_0\ d\alpha_0 + \int f_1\ d\alpha_1\right] \\ &= \sup_{f_0 \in C(E_0)} \sup_{f_1 \in C(E_1)} \left[R(\gamma, f_0, f_1) + \int f_0\ d\alpha_0 + \int f_1\ d\alpha_1\right],\end{aligned}$$

where

$$R(\gamma, f_0, f_1) = \inf_{(x_0, x_1) \in E_0 \times E_1} \big(g(x_0, x_1) - f_0(x_0) - f_1(x_1)\big).$$

Note that if $a = R(\gamma, f_0, f_1)$, then

$$\begin{aligned}&g(x_0, x_1) \geq f_0(x_0) + f_1(x_1) + a, \\ &a + \int f_0\ d\alpha_0 + \int f_1\ d\alpha_1 = \int f_0\ d\alpha_0 + \int (f_1 + a)\ d\alpha_1.\end{aligned}$$

Hence we can drop the R-term by modifying (for example) f_1. This implies that $\mathcal{D}(\alpha_0, \alpha_1) = \mathcal{D}^*(\alpha_0, \alpha_1)$. Slight modifications of the above argument would yield $\mathcal{D}(\alpha_0, \alpha_1) = \mathcal{D}'(\alpha_0, \alpha_1)$.

Step 2 We now drop the compactness assumption and instead we assume that $g \in C_b(E_0 \times E_1)$. Note that the space $\Gamma(\alpha_0, \alpha_1)$ is tight because for any pair of compact sets $K_0 \subseteq E_0$, $K_1 \subseteq E_1$, we have

$$\gamma\big((K_0 \times K_1)^c\big) \leq \alpha_0(K_0^c) + \alpha_1(K_1^c).$$

Since the function $\gamma \mapsto \int g\ d\gamma$ is continuous, we know that there exists $\bar{\gamma} \in \Gamma(\alpha_0, \alpha_1)$ such that

$$\mathcal{D}(\alpha_0, \alpha_1) = \int g\ d\bar{\gamma}.$$

Given $\delta > 0$, choose compact sets $K_0 \subseteq E_0$ and $K_1 \subseteq E_1$ such that

$$\alpha_0(K_0^c) + \alpha_1(K_1^c) \leq \delta.$$

This implies that $\bar{\gamma}(K^c) \leq \delta$ for $K = K_0 \times K_1$. We then define $\hat{\gamma}$ by

$$\hat{\gamma} = \bar{\gamma}(K)^{-1} 1\!\!1_K\ \bar{\gamma},$$

so that $\hat{\gamma} \in \mathcal{M}(K)$. We denote the marginals of $\bar{\gamma}$ by $\hat{\alpha}_0$ and $\bar{\alpha}_1$. We also find $\tilde{\gamma} \in \mathcal{M}(K)$ such that

$$\mathcal{D}(\hat{\alpha}_0, \hat{\alpha}_1) = \int g \ d\tilde{\gamma}.$$

Given $\delta \in (0,1)$, use Step 1 to find $\hat{f}_i \in C(K_i)$, for $i = 0, 1$ such that for every $(x_0, x_1) \in K$,

$$\hat{f}_0(x_0) + \hat{f}_1(x_1) \leq g(x_0, x_1),$$

and

$$\mathcal{D}(\hat{\alpha}_0, \hat{\alpha}_1) = \int g \ d\tilde{\gamma} \leq \int \hat{f}_0 \ d\hat{\alpha}_0 + \int \hat{f}_1 \ d\hat{\alpha}_1 + \delta. \tag{E.6}$$

We are hoping to use the pair $(\hat{f}_0, \hat{f}_1)$ to build a pair (f_0, f_1) so that $\int g \ d\bar{\gamma} \leq \int f_0 \ d\alpha_1 + \int f_1 \ d\alpha_1 + \varepsilon$, for some small ε. Note that f_i is only defined on K_i. Let us set

$$f_0(x_0) = \inf_{y_1 \in K_1} (g(x_0, y_1) - \hat{f}_1(y_1)), \qquad f_1(x_1) = \inf_{y_0 \in E_0} (g(y_0, x_1) - f_0(y_0)).$$

Evidently $f_0(x_0) + f_1(x_1) \leq g(x_0, x_1)$ for all $(x_0, x_1) \in E_0 \times E_1$, and $\hat{f}_i \leq f_i$ on K_i. As a result,

$$\begin{aligned} \int g \ d\tilde{\gamma} &\leq \int \hat{f}_0 \ d\hat{\alpha}_0 + \int \hat{f}_1 \ d\hat{\alpha}_1 + \delta \leq \int f_0 \ d\hat{\alpha}_0 + \int f_1 \ d\hat{\alpha}_1 + \delta \\ &=: \int f_0 \ d\alpha_0 + \int f_1 \ d\alpha_1 + \delta' + \delta, \end{aligned} \tag{E.7}$$

where $\delta' = \delta(f_0, f_1)$ represents the error term we get as we by replace $\hat{\alpha}_i$ with α_i for $i = 0, 1$. To show that δ' is small, we need some bounds on f_0 and f_1.

Step 3 To have pointwise bounds on f_0 and f_1, we first find such bounds for $\hat{f}_0$ and $\hat{f}_1$. We have an obvious upper bound

$$\hat{f}_0(x_0) + \hat{f}_1(x_1) \leq g(x_0, x_1) \leq \|g\|,$$

on the set K. As for the lower bound, observe that we almost have equality of g with $\hat{f}_0(x_0) + \hat{f}_1(x_1)$ on the support of $\tilde{\gamma}$. This is because by (E.6),

$$\int [g(x_0, x_1) - \hat{f}_0(x_0) - \hat{f}_1(x_1)] \ \tilde{\gamma}(dx_0, dx_1) \leq \delta.$$

This in particular implies that for some point $(\bar{x}_0, \bar{x}_1) \in K$,

$$\hat{f}_0(x_0) + \hat{f}_1(x_1) \geq g(x_0, x_1) - \delta \geq -\|g\| - \delta \geq \|g\| - 1 =: -2a_0 \tag{E.8}$$

Note that if we replace $(\hat{f}_0(x_0), \hat{f}_1(x_1))$ with $(\hat{f}_0(x_0)-c_0, \hat{f}_1(x_1)+c_0)$ for a constant c_0, nothing in the above argument would change. We may choose c_0 so that $\hat{f}_0(\bar{x}_0) = \hat{f}_1(\bar{x}_1)$. Assuming this, we may use (E.8) to assert

$$\hat{f}_0(\bar{x}_0) = \hat{f}_1(\bar{x}_1) \geq -a_0.$$

This in turn implies that for $(x_0, x_1) \in K$,

$$\hat{f}_0(x_0) \leq g(x_0, \bar{x}_1) - \hat{f}_1(\bar{x}_1) \leq \|g\| + a_0, \qquad \hat{f}_1(x_1) \leq g(\bar{x}_0, x_1) - \hat{f}_0(\bar{x}_0) \leq \|g\| + a_0.$$

Using these lower bounds we use the definition of f_0 and f_1 to deduce that for $x_0 \in E_0$ and $x_1 \in E_1$,

$$\begin{aligned} -2\|g\| - a_0 &\leq f_0(x_0) = \inf_{K_1}[g(x_0, \cdot) - \hat{f}_1] \leq g(x_0, \bar{x}_1) - \hat{f}_1(\bar{x}_1) \leq \|g\| + a_0, \\ -2\|g\| - a_0 &\leq f_1(x_1) = \inf_{E_0}[g(\cdot, x_1) - f_0] \leq 3\|g\| + a_0. \end{aligned}$$

To summarize, we set $a_1 = 3\|g\| + a_0$, so that

$$|f_0|, |f_1| \leq a_1. \tag{E.9}$$

We are now ready to bound the error δ' that appeared in (E.7). We can use (E.9) to show

$$\left| \int f_i \; d\hat{\alpha}_i - \int_{K_i} f_i \; d\alpha_i \right| = \left| \bar{\gamma}(K)^{-1} \int_K f_i(x_i) \; \bar{\gamma}(dx_0, dx_1) - \int f_i \; d\alpha_i \right| \leq \frac{2\delta a_1}{1-\delta}.$$

From this and (E.7) we deduce

$$\int g \; d\tilde{\gamma} \leq \int f_0 \; d\alpha_0 + \int f_1 \; d\alpha_1 + 4\delta a_1 + \delta, \tag{E.10}$$

for every $\delta \in (0, 1/2]$. We are almost done except that $\tilde{\gamma}$ does not α_0 and α_1 for its marginals.

Step 4 The support of $\tilde{\gamma}$ is contained in the set K. To extend it to the whole $E_0 \times E_1$, we define

$$\gamma^* = \bar{\gamma}(K)\tilde{\gamma} + 1\!\!1_{K^c} \; \bar{\gamma}.$$

We have

$$\begin{aligned} \gamma^*(A \times E_1) &= \bar{\gamma}(K) \; \tilde{\gamma}(A \times E_1) + \bar{\gamma}\big(K^c \cap (A \times E_1)\big) \\ &= \bar{\gamma}(K) \; \hat{\alpha}(A) + \bar{\gamma}\big(K^c \cap (A \times E_1)\big) \\ &= \bar{\gamma}\big(K \cap (A \times E_1)\big) + \bar{\gamma}\big(K^c \cap (A \times E_1)\big) = \alpha_0(A). \end{aligned}$$

In the same we show that x_1-marginal of γ^* is α_1. On the other hand,

$$\left|\int g\ d\tilde{\gamma} - \int g\ d\gamma^*\right| \le \|g\|\left(1 - \bar{\gamma}(K) + \bar{\gamma}(K^c)\right) \le 2\delta\|g\|.$$

From this and (E.10) we deduce

$$\int g\ d\gamma^* \le \int f_0\ d\alpha_0 + \int f_1\ d\alpha_1 + 4\delta a_1 + \delta + 2\delta\|g\|.$$

This completes the proof assuming that $g \in C_b(E_0 \times E_1)$.

Final Step For the general case, choose a sequence $\{g_n :\ n \in \mathbb{N}\} \subseteq C_b(E_0 \times E_1)$ such that $g_n \le g_{n+1}$ for every $n \in \mathbb{N}$ and $\sup_n g_n = g$. All we need to show is

$$\mathcal{D}(\alpha_0, \alpha_1) \le \sup_n \inf_{\gamma\in\Gamma(\alpha_0,\alpha_1)} \int g_n\ d\gamma. \tag{E.11}$$

Indeed once (E.11) is established, then we can choose a large n such that

$$\mathcal{D}(\alpha_0, \alpha_1) - \delta \le \inf_{\gamma\in\Gamma(\alpha_0,\alpha_1)} \int g_n\ d\gamma,$$

for a given $\delta > 0$, and choose bounded continuous functions f_0 and f_1 such that

$$\begin{aligned} &f_0(x_1) + f_1(x_1) \le g_n(x_0, x_1) \le g(x_0, x_1), \\ &\inf_{\gamma\in\Gamma(\alpha_0,\alpha_1)} \int g_n\ d\gamma - \delta \le \int f_0\ d\alpha_0 + \int f_1\ d\alpha_1, \end{aligned}$$

by what we established in Step 3.

To establish (E.11), choose $\gamma_n \in \Gamma(\alpha_0, \alpha_1)$ such that

$$\inf_{\gamma\in\Gamma(\alpha_0,\alpha_1)} \int g_n\ d\gamma = \int g_n\ d\gamma_n,$$

and let $\bar{\gamma}$ be any limit point of the sequence $\{\gamma_n\}$. We have

$$\begin{aligned} \mathcal{D}(\alpha_0, \alpha_1) &\le \int g\ d\bar{\gamma} = \lim_{m\to\infty} \int g_m\ d\bar{\gamma} = \lim_{m\to\infty} \lim_{n\to\infty} \int g_m\ d\gamma_n \\ &\le \lim_{m\to\infty} \lim_{n\to\infty} \int g_n\ d\gamma_n = \lim_{n\to\infty} \int g_n\ d\gamma_n = \sup_n \inf_{\gamma\in\Gamma(\alpha_0,\alpha_1)} \int g_n\ d\gamma, \end{aligned}$$

as desired. □

Acknowledgments. This work is supported in part by NSF grant DMS-1106526.

References

[B] P. Billingsley, Convergence of probability measures. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1999.

[DV] M. D. Donsker; S. R. S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. I-IV. Comm. Pure Appl. Math. **28** (1975), 1–47; 28 (1975), 279–301; **29** (1976), no. 4, 389–461 ; **36** (1983), no. 2, 183–212.

[K] H. Komiya, Elementary proof for Sion's minimax theorem. Kodai Math. J. **11** (1988), no. 1, 57.

[Ol] S. Olla, Large deviations for Gibbs random fields. Probab. Theory Related Fields **77** (1988), no. 3, 343–357.

[Ox] J. C. Oxtoby, Ergodic sets. Bull. Amer. Math. Soc. **58**, (1952). 116–136.

[P] K. R. Parthasarathy, Probability measures on metric spaces. Reprint of the 1967 original. AMS Chelsea Publishing, Providence, RI, 2005.

[R1] F. Rezakhanlou, Hydrodynamic limit for a system with finite range interactions. Comm. Math. Phys. **129** (1990), no. 3, 445–480.

[R2] F. Rezakhanlou, Lectures on Dynamical Systems. http://math.berkeley.edu/r̃ezakhan/

[SV1] S. Chatterjee; S. R. S. Varadhan, Large deviations for random matrices. Commun. Stoch. Anal. **6** (2012), 1–13.

[SV2] S. Chatterjee; S. R. S. Varadhan, The large deviation principle for the Erdős-Rènyi random graph. European J. Combin. **32** (2011), no. 7, 1000–1017.

[V1] C. Villani, Topics in Optimal Transportaion. Graduate Studies in Mathematics, AMS, 2003.

[V2] C. Villani, Optimal Transport: Old and New. Springer Verlag (Grundlehren der mathematischen Wissenschaften), 2008.