Large Deviations

Richard Sowers

Department of Mathematics, University of Illinois at Urbana-Champaign
E-mail address: r-sowers@illinois.edu
Contents

Chapter 1. Some Simple Calculations and Laplace Asymptotics 5
  1. The Main Ideas and Examples 7
  2. The Contraction Principle 11
      Exercises 12

Chapter 2. The Gärtner-Ellis Theorem 15
      Exercises 21

Chapter 3. Schilder’s Theorem 23
      Exercises 27

Chapter 4. Freidlin-Wentzell-Theory 29
      Exercises 34

Chapter 5. Large Deviations for Markov Chains 35
  1. Pair Empirical Measure 42
      Exercises 43

Chapter 6. Refined Asymptotics 45
  1. Reference 48

Bibliography 49
CHAPTER 1

Some Simple Calculations and Laplace Asymptotics

The basic idea which underlines all of large deviations theory is exemplified in the following calculation.

Set

\[ I_\varepsilon \overset{\text{def}}{=} e^{-3/\varepsilon} + e^{-7/\varepsilon} \]

for all \( \varepsilon \in (0, 1) \). Then

\[ I_\varepsilon = e^{-3/\varepsilon} \left\{ 1 + e^{-4/\varepsilon} \right\} \]

for all \( \varepsilon \in (0, 1) \), and \( \lim_{\varepsilon \to 0} \{ 1 + e^{-4/\varepsilon} \} = 1 \). Thus

\[ \lim_{\varepsilon \to 0} \varepsilon \ln I_\varepsilon = -3 + \lim_{\varepsilon \to 0} \varepsilon \ln \{1 + e^{-4/\varepsilon}\} = -3. \]

**Definition 0.1.** If \( \{A_\varepsilon\}_{\varepsilon \in (0,1)} \) and \( \{B_\varepsilon\}_{\varepsilon \in (0,1)} \) are two real-valued sequences, we say that

\( A_\varepsilon \asymp B_\varepsilon \)

if

\[ \lim_{\varepsilon \to 0} \varepsilon \ln A_\varepsilon = \lim_{\varepsilon \to 0} \varepsilon \ln B_\varepsilon. \]

The generalization of (1) is thus that

\[ \sum_{n=1}^{N} a_n \varepsilon^{-c_j/\varepsilon} \asymp \exp \left[ -\frac{1}{\varepsilon} \min_{1 \leq n \leq N} c_n \right], \]

for generic \( a_n \)'s and \( c_n \)'s (we clearly have to disallow the situations like \( 7e^{-8/\varepsilon} + 7e^{-8/\varepsilon} \)).

**0.1. Laplace Asymptotics.** Let's extend these calculations to integrals. Define

\[ I_\varepsilon \overset{\text{def}}{=} \int_{x=0}^{1} \exp \left[ -\frac{1}{\varepsilon} f(x) \right] dx, \quad \varepsilon \in (0,1) \]

where \( f : [0,1] \to \mathbb{R} \) is smooth and has a unique minimum at some \( x^* \in (0,1) \). Let’s also assume that \( \dot{f}(x^*) > 0 \). Informally, we have that \( \dot{f}(x^*) > 0 \). Formally, we have that

\[ I_\varepsilon \approx \sum_{j=1}^{J-1} \exp \left[ -\frac{1}{\varepsilon} f(x_j) \right] \{x_{j+1} - x_j\} \asymp \exp \left[ -\frac{1}{\varepsilon} \inf_{1 \leq j \leq J-1} f(x_j) \right] \times \exp \left[ -\frac{1}{\varepsilon} \inf_{0 \leq x \leq 1} f(x) \right] \]

where \( 0 = x_1 < x_2 \ldots x_{J-1} < x_J = 1 \) is a partition of \([0,1]\) which is “sufficiently” fine. Let’s see if we can make this rigorous. We can write \( I_\varepsilon = \exp \left[ -\frac{1}{\varepsilon} \int_{x=x^*} f(x) \right] \tilde{I}_\varepsilon \) where

\[ \tilde{I}_\varepsilon = \int_{x=0}^{1} \exp \left[ -\frac{1}{\varepsilon} \{f(x) - f(x^*)\} \right] dx. \]

Our goal is then to show that \( \lim_{\varepsilon \to 0} \varepsilon \ln \tilde{I}_\varepsilon = 0 \).

Let’s break \( \tilde{I}_\varepsilon \) into two parts;

\[ \tilde{I}_\varepsilon = \tilde{I}_\varepsilon^A + \tilde{I}_\varepsilon^B \]

where

\[ \tilde{I}_\varepsilon^A \overset{\text{def}}{=} \int_{x \in [0,1]} \exp \left[ -\frac{1}{\varepsilon} \{f(x) - f(x^*)\} \right] dx \]

and

\[ \tilde{I}_\varepsilon^B \overset{\text{def}}{=} \int_{|x-x^*| > \varepsilon} \exp \left[ -\frac{1}{\varepsilon} \{f(x) - f(x^*)\} \right] dx. \]
\[ \tilde{I}_\varepsilon \overset{\text{def}}{=} \int_{x \in [0, 1]} \exp \left[ -\frac{1}{\varepsilon} \{ f(x) - f(x^*) \} \right] dx \]

where \( \gamma > 0 \) is some parameter to be determined.

To proceed, we note that

\[ c = \lim_{\delta \to 0} \inf_{x \in [0, 1]} \frac{f(x) - f(x^*)}{\delta^2} > 0. \]  

Thus for \( \varepsilon \in (0, 1) \) sufficiently small,

\[ \inf_{x \in [0, 1], |x - x^*| \geq \varepsilon} \frac{1}{\varepsilon} \{ f(x) - f(x^*) \} \geq \frac{c}{2} \varepsilon^{2\gamma - 1}. \]

Thus

\[ \tilde{I}_\varepsilon \leq \exp \left[ -\frac{c}{2} \varepsilon^{2\gamma - 1} \right]. \]

If \( \gamma < \frac{1}{2} \), then \( \lim_{\varepsilon \to 0} \varepsilon^{2\gamma - 1} = \infty \), so \( \lim_{\varepsilon \to 0} \varepsilon \ln \tilde{I}_\varepsilon = 0 \).

Let’s next consider \( \tilde{I}_\varepsilon^B \). We first expand \( f \) around \( x^* \). We have that

\[ f(x) = f(x^*) + \frac{1}{2} \tilde{f}(x^*)(x - x^*)^2 + \mathcal{E}_1(x) \]

for \( x \in [0, 1] \), where

\[ \mathcal{E}_1(x) = \frac{1}{2} (x - x^*)^3 \int_{s=0}^{s=1} (1 - s)^2 f^{(3)}(x^*) + s(x - x^*))ds \]

for all \( x \in [0, 1] \). Let’s next rescale the integral representation for \( \tilde{I}_\varepsilon^B \) by taking \( z = (x - x^*)/\sqrt{\varepsilon} \). If \( \varepsilon > 0 \) is small enough that \( (x^* - \varepsilon^\gamma, x^* + \varepsilon^\gamma) \subset (0, 1) \), we get that

\[ \tilde{I}_\varepsilon^B = \sqrt{\varepsilon} \int_{z=-\varepsilon^{\gamma - 1/2}}^{\varepsilon^{\gamma - 1/2}} \exp \left[ -\frac{1}{2} \tilde{f}(x^*)z^2 + \frac{1}{\varepsilon} \mathcal{E}_1 (x^* + \sqrt{\varepsilon}z) \right] dz \]

If \( |z| \leq \varepsilon^{\gamma - 1/2} \), then

\[ \left| \frac{1}{\varepsilon} \mathcal{E}_1 (x^* + \sqrt{\varepsilon}z) \right| \leq ||f^{(3)}||_{C[0, 1]} \frac{1}{\varepsilon} \sqrt{\varepsilon} |z|^3 \leq \varepsilon^{3\gamma - 1}. \]

This suggests that we take \( \gamma > \frac{1}{3} \) so that \( \lim_{\varepsilon \to 0} \varepsilon^{3\gamma - 1} = 0 \); note also that \( \lim_{\varepsilon \to 0} \varepsilon^{\gamma - 1/2} = \infty \). Doing so, we get that

\[ \lim_{\varepsilon \to 0} \frac{1}{\sqrt{\varepsilon}} \tilde{I}_\varepsilon^B = \int_{z \in \mathbb{R}} \exp \left[ -\frac{1}{2} \tilde{f}(x^*)z^2 \right] dz = \sqrt{\frac{2\pi}{\tilde{f}(x^*)}}. \]

In other words,

\[ \tilde{I}_\varepsilon^B = \sqrt{\varepsilon} \left\{ \sqrt{\frac{2\pi}{\tilde{f}(x^*)}} + \mathcal{E}_2(\varepsilon) \right\} \]

\[ \tilde{I}_\varepsilon = \sqrt{\varepsilon} \left\{ \sqrt{\frac{2\pi}{\tilde{f}(x^*)}} + \mathcal{E}_2(\varepsilon) \right\} + \tilde{I}_\varepsilon^A \]

where \( \lim_{\varepsilon \to 0} \mathcal{E}_2(\varepsilon) = 0 \) and thus

\[ \lim_{\varepsilon \to 0} \varepsilon \ln \tilde{I}_\varepsilon = 0. \]
0.2. I.I.D. Coin Tosses. Let’s now toss a large collection of independent and identically distributed (i.i.d.) coins. Let \( \{\xi_n\}_{n \in \mathbb{N}} \) be i.i.d. Bernoulli random variables; i.e.,

\[
\mathbb{P}\{\xi_n = 1\} = p \quad \text{and} \quad \mathbb{P}\{\xi_n = 0\} = 1 - p
\]

for all \( n \in \mathbb{N} \), where \( p \in (0, 1) \) is some fixed parameter. Define

\[
X_N \overset{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \xi_n
\]

for all \( N \in \mathbb{N} \). Thus \( NX_N \) is a binomial distribution;

\[
\mathbb{P}\{NX_N = k\} = \binom{N}{k} p^k (1 - p)^{N-k}
\]

for all \( k \in \{0, 1, \ldots, N\} \). The second key realization is that we can use Stirling’s formula to approximate \( n! \) for large \( n \); recall that \( n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \) for large \( n \) (more precisely, the difference between the two is negligible for large \( n \)). Thus we have the following string of approximate equalities: for any \( \alpha \in (0, 1) \),

\[
P\{X_N \approx \alpha\} \approx \frac{N!}{(N\alpha)!(N-N\alpha)!} p^{N\alpha} (1 - p)^{N-N\alpha}
\approx \frac{(N/e)^N}{(N\alpha/e)^N (N(1 - \alpha)/e)^{N(1-\alpha)} \sqrt{2\pi N \alpha} \sqrt{2\pi N (1 - \alpha)}} \times \frac{1}{\sqrt{2\pi N \alpha(1 - \alpha)}}
\times \frac{1}{\sqrt{2\pi N \alpha(1 - \alpha)}} \exp \left[N \left\{ \alpha \ln \frac{p}{\alpha} + (1 - \alpha) \ln \frac{1 - p}{1 - \alpha} \right\} \right]
\approx \frac{1}{\sqrt{2\pi N \alpha(1 - \alpha)}} \exp \left[-NH(\alpha, p)\right]
\]

where

\[
H(\alpha, p) \overset{\text{def}}{=} \alpha \ln \frac{\alpha}{p} + (1 - \alpha) \ln \frac{1 - \alpha}{1 - p}.
\]

This is of course the famous relative entropy. Specifically, it is the relative entropy of a coin with bias \( \alpha \) with respect to the true coin. Thus for any \( A \in [0, 1] \), we should heuristically have that

\[
P\{X_N \in A\} \approx \sum_{\alpha \in A} P\{X_N \approx \alpha\} \approx \sum_{\alpha \in A} \frac{1}{\sqrt{2\pi N \alpha(1 - \alpha)}} \exp \left[-NH(\alpha, p)\right] \approx \exp \left[-N \inf_{\alpha \in A} H(\alpha, p)\right].
\]

The \( 1/\sqrt{N} \) terms in (5) should of course negligible at the exponential rate of interest.

1. The Main Ideas and Examples

Large Deviations theory gives a background behind both Laplace asymptotics and the calculations of i.i.d. coin flips. Note that Laplace asymptotics primarily applies to continuous random variables, and the coin flips are discrete. We can also handle infinite-dimensional random variables.

There are three basic questions we would like to address.

(1) What is a large deviations principle? What is the “correct” way to formulate a study of rare events?
(2) Is there a unifying theory for studying the origination of rare events?
(3) How do rare events transform?
We will give the definition of a large deviations principle here. We will also motivate the answer to the second question; the formal result is called the Gätzner-Ellis result and will, due to its technicalities, be addressed in the next chapter. The last issue is essentially the “contraction principle” of Varadhan, which we will also address here.

1.1. Definition of an LDP. Our basic setup is as follows. We have a collection \( \{X_\varepsilon\}_{\varepsilon \in (0,1)} \) of random variables which take values in some Polish space \( X \). We will see that the following is the “correct” way to think about exponentially rare events.

**Definition 1.1.** We say that \( \{X_\varepsilon\}_{\varepsilon \in (0,1)} \) has a large deviations principle with rate function \( I : X \to [0, \infty] \) if
1. For each \( s \geq 0 \), \( \Phi(s) \defeq \{ x \in X : I(x) \leq s \} \) is compact.
2. For each closed subset \( F \) of \( X \),
   \[ \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}(X_\varepsilon \in F) \leq -\inf_{x \in F} I(x) \]
3. For each open subset \( G \) of \( X \),
   \[ \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}(X_\varepsilon \in G) \geq -\inf_{x \in G} I(x) \]

We will see why this is correct in a moment. It turns out, however, that there is an equivalent definition for a large deviations principle. We here let \( d \) be the metric on \( X \).

**Definition 1.2.** We say that \( \{X_\varepsilon\}_{\varepsilon \in (0,1)} \) has a large deviations principle with rate function \( I : X \to [0, \infty] \) if
1. For each \( s \geq 0 \), \( \Phi(s) \defeq \{ x \in X : I(x) \leq s \} \) is compact.
2. For each \( s \geq 0 \) and each \( \delta > 0 \),
   \[ \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}(d(X_\varepsilon, \Phi(s)) \geq \delta) \geq -s \]
3. For every \( x^* \in X \) and every \( \delta > 0 \),
   \[ \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}(d(X_\varepsilon, x^*) < \delta) \leq -I(x^*) \]

1.2. How to get at an LDP. Let’s return to the framework of Laplace asymptotics. Suppose that \( \{X_\varepsilon\}_{\varepsilon \in (0,1)} \) is a collection of \( \mathbb{R} \)-valued random variables such that, for some \( f \in C^2[0,1] \),

\[ \mathbb{P}(X_\varepsilon \in A) = c_\varepsilon \int_{x \in A} \exp \left[ -\frac{1}{\varepsilon} f(x) \right] dx \]

for all \( A \in \mathcal{B}[0,1] \), where
\[ c_\varepsilon \defeq \left\{ \int_{x \in A} \exp \left[ -\frac{1}{\varepsilon} f(x) \right] dx \right\}^{-1} \]

with \( \lim_{\varepsilon \to 0} \varepsilon \ln c_\varepsilon = 0 \).

**Lemma 1.1.** For any \( A \in \mathcal{B}[0,1] \),
\[ \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}(X_\varepsilon \in A) \leq -\inf_{x \in A} f(x) \]
\[ \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}(X_\varepsilon \in A) \geq -\inf_{x \in A^*} f(x). \]

**Proof.** Let \( \mathcal{L} \) be Lebesgue measure on \( ([0,1], \mathcal{B}[0,1]) \). The upper bound is obvious;

\[ \mathbb{P}(X_\varepsilon \in A) \leq c_\varepsilon \mathcal{L}(A) \exp \left[ -\frac{1}{\varepsilon} \inf_{x \in A} f(x) \right] \leq c_\varepsilon \exp \left[ -\frac{1}{\varepsilon} \inf_{x \in A} f(x) \right]. \]

To get the lower bound, fix \( x^* \in A^c \) and \( \delta > 0 \). Define \( \mathcal{O} \defeq \{ x \in [0,1] : f(x) > f(x^*) - \delta \} \cap A^c \). Since \( f \) is continuous, \( \mathcal{O} \) is open (in the topology \([0,1]\) inherits from \( \mathbb{R} \)) and thus \( \mathcal{L}(\mathcal{O}) > 0 \). Hence

\[ \mathbb{P}(X_\varepsilon \in A) \geq c_\varepsilon \mathcal{L}(\mathcal{O}) \exp \left[ -\frac{1}{\varepsilon} \{ f(x^*) + \delta \} \right]. \]
This gives the lower bound by taking $\varepsilon \to 0$ and then $\delta \to 0$, and they varying $x^*$ over $A^\circ$. □

The upper bound was easy and global; the lower bound required a localization (and the fact that $\mathcal{L}(O) > 0$). While the two bounds may agree (if $\bar{A} = \overline{A^\circ}$), they will not agree in general. For example, if $A$ is a single point or a countable dense set, they will disagree. These “technicalities” become even more pronounced in infinite-dimensional spaces.

Is there a robust way to recover $f$ even if we don’t start from Laplace asymptotics and where we may not originally know the exact distribution of $X_\varepsilon$ (recall that in (4), we only knew asymptotics of the distribution of $X_N$)? Our interest is to come up with a Laplace-type asymptotics when we don’t have (6) as a starting point.

To see how things work, let’s compute asymptotics of the moment generating functions of $X_\varepsilon$. Laplace asymptotics tells us that

$$
\lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{E}
\left[
\exp\left[\frac{\theta}{\varepsilon} X^*\right]
\right] = \lim_{\varepsilon \to 0} \varepsilon \ln \left\{ c_\varepsilon \int_{x=0}^{1} \exp\left[\frac{1}{\varepsilon} \{\theta x - f(x)\}\right] dx \right\} = \Lambda(\theta)
$$

where

$$
\Lambda(\theta) \overset{\text{def}}{=} \sup_{x \in [0,1]} \{\theta x - f(x)\}.
$$

The key observation we wish to exploit is almost the $\Lambda$ is the Legendre-Fenchel transform of $f$. More exactly, it is exactly the Legendre-Fenchel transform of the function

$$
f^\infty(x) \overset{\text{def}}{=} \begin{cases} f(x) & \text{if } x \in [0,1] \\ \infty & \text{else.} \end{cases}
$$

**Definition 1.3.** Let $X$ be a vector space with dual $X^\ast$. Fix $f : X \to [\infty, \infty]$. The Legendre-Fenchel transform of $f$ is the function

$$f^\ast(x^\ast) \overset{\text{def}}{=} \sup_{x \in X} \{(x, x^\ast)_X - f(x)\} \quad x^\ast \in X^\ast
$$

The bidual of $f$ is the function

$$f^{**}(x) \overset{\text{def}}{=} \sup_{x^\ast \in X^\ast} \{(x, x^\ast)_X - f^\ast(x^\ast)\} \quad x \in X
$$

The deus ex machina which we shall exploit (and in a moment prove) is that if $f$ is “nice”, $f = f^{**}$. The definition of nice involves some more definitions.

**Definition 1.4.** Fix a space $X$ and a map $f : X \to [\infty, \infty]$. Define

$$\text{epi}(f) \overset{\text{def}}{=} \{(x, \alpha) \in X \times \mathbb{R} : \alpha \geq f(x)\} \subset X \times \mathbb{R};$$

this is the epigraph of $f$.

- If $X$ is a vector space, we say that $f$ is convex if $\text{epi}(f)$ is a convex subset of $X \times \mathbb{R}$.
- If $X$ is a topological space, we say that $f$ is lower semicontinuous if $\text{epi}(f)$ is a closed subset of $X \times \mathbb{R}$.

We then have

**Lemma 1.2.** Suppose that $X$ is a topological vector space. If $f : X \to (\infty, \infty]$ is convex and lower semicontinuous, then $f = f^{**}$.

We will prove this in the following subsection. In the world of (6), if we know that $f$ is convex and lower semicontinuous, we can reconstruct it from the asymptotics of (7) via the formula

$$f(x) = \sup_{\theta \in \mathbb{R}} \langle \theta x - \Lambda(\theta) \rangle.$$
1.3. Some Convex Analysis. Let’s first understand what the Legendre-Fenchel transform is, at least for a function $f : \mathbb{R} \to \mathbb{R}$. Fix $\theta \in \mathbb{R}$. By definition, $f^*(\theta)$ is the smallest constant $c$ such that $c \geq \theta x - f(x)$ for all $x \in \mathbb{R}$; i.e., the smallest constant $c$ such that $\theta x - c \leq f(x)$. If we define $\ell(x) \overset{\text{def}}{=} \theta x - c$ for all $x \in \mathbb{R}$, then $\ell$ is a linear minorant of $f$, and $-c = \ell(0)$ (i.e., the $y$-intercept of $\ell$) and thus $c = -\ell(0)$. Hence

$$f^*(\theta) = \inf \{-\ell(0) : \ell \text{ is a linear minorant of } f\} = -\sup \{-\ell(0) : \ell \text{ is a linear minorant of } f\}.$$ 

Suppose now that $X$ is a topological vector space. Suppose that $f : X \to (-\infty, \infty]$ is convex. Define

$$\text{dom}(f) \overset{\text{def}}{=} \{x \in X : f(x) < \infty\}.$$ 

If a point is not in $\text{epi}(f)$, then we can use the Hahn-Banach theorem to separate it from $\text{epi}(f)$. This leads to the following.

**Lemma 1.3.** Suppose that $X$ is a topological vector space and that $f : X \to (-\infty, \infty]$ is convex and $f \not= \infty$. Fix $(x, \alpha) \not\in \text{epi}(f)$. Then there is an $x^* \in X^*$, $c \in \mathbb{R}$, and a $\lambda \geq 0$ such that

$$\langle x, x^* \rangle_X - \lambda \alpha > c \quad \langle z, x^* \rangle_X - \lambda f(z) < c \quad z \in \text{dom}(f)$$

(8)

If $f(x) < \infty$, then $\lambda > 0$.

**Proof.** By the Hahn-Banach theorem, there is a $(x^*, \lambda) \in X^* \times \mathbb{R}$ and a $c \in \mathbb{R}$ such that

$$\langle x, x^* \rangle_X - \lambda \alpha > c \quad \langle y, x^* \rangle_X - \lambda \beta < c \quad (y, \beta) \in \text{epi}(f)$$

(9)

Since $f \not= \infty$, there is a $\tilde{x} \in V$ such that $f(\tilde{x}) < \infty$. From the second inequality, we thus get that $\lambda \beta < c - \langle \tilde{x}, x^* \rangle_X$ for all $\beta > f(x)$. Letting $\beta \nearrow \infty$, we arrive at a contradiction if $\lambda < 0$; thus indeed $\lambda \geq 0$. Since $(x, f(x)) \in \text{epi}(f)$ for any $x \in \text{dom}(f)$, we get the (8). If $\lambda = 0$ and $x \in \text{dom}(f)$, then

$$\langle x, x^* \rangle_X > c > \langle x, x^* \rangle_X$$

which is a contradiction. Thus if $f(x) < \infty$, we must have $\lambda > 0$. \qed

We can in fact do a bit better.

**Corollary 1.4.** Under the same assumptions as Lemma 1.3, we can take $\lambda > 0$.

**Proof.** We start with (8). If $\lambda > 0$, we are done. We thus henceforth assume that $\lambda = 0$. By assumption there is an $\tilde{x} \in V$ with $f(\tilde{x}) \in \mathbb{R}$. Taking $\tilde{\alpha} < f(\tilde{x})$, we can apply Lemma 1.3 to $(\tilde{x}, \tilde{\alpha})$. We get an $\tilde{x}^* \in X^*$, $\tilde{c} \in \mathbb{R}$, and $\tilde{\lambda} > 0$ such that

$$\langle \tilde{x}, \tilde{x}^* \rangle_X - \tilde{\lambda} \tilde{\alpha} > \tilde{c} \quad \langle z, \tilde{x}^* \rangle_X - \tilde{\lambda} f(\tilde{x}) < \tilde{c}.$$ 

$z \in \text{dom}(f)$

Let’s now take a linear combination of this and (8). Fix $\eta \in (0, 1)$. Then

$$\eta \langle x, x^* \rangle_X + (1 - \eta) \tilde{\lambda} \tilde{\alpha} > \eta c + (1 - \eta) \tilde{c} \quad \langle z, \eta x^* + (1 - \eta) \tilde{x}^* \rangle_X - (1 - \eta) \tilde{\lambda} f(\tilde{x}) < \eta c + (1 - \eta) \tilde{c}, \quad z \in \text{dom}(f)$$

The claim will follow if we can find $\eta \in (0, 1)$ such that

$$\langle x, \eta x^* + (1 - \eta) \tilde{x}^* \rangle_X + (1 - \eta) \tilde{\lambda} \tilde{\alpha} > \eta c + (1 - \eta) \tilde{c}$$

But this is obvious;

$$\lim_{\eta \nearrow 1} \left( \left\{ \langle x, \eta x^* + (1 - \eta) \tilde{x}^* \rangle_X + (1 - \eta) \tilde{\lambda} \tilde{\alpha} \right\} - \{\eta c + (1 - \eta) \tilde{c}\} \right) = \langle x, x^* \rangle_X - c > 0.$$ 

\qed

We finally the separation theorem we will use.

**Corollary 1.5.** Under the same assumptions as Lemma 1.3, we can take $\lambda = 1$. 

10
PROOF. Using Corollary, we can assume that $\lambda > 0$ in (8) by $\lambda$; thus we can divide by $\lambda$. \hfill $\Box$

PROOF OF LEMMA 1.2. First, if $f \equiv \infty$, then $f^* \equiv -\infty$ and thus $f^{**} = \infty = f$. We thus now assume that $f$ is not identically $\infty$.

Fix $x \in X$. Note that

$$f^{**}(x) = \sup_{x^* \in X^*} \inf_{x^* \in X} \{(x - x', x^*)_X + f(x')\} \leq \sup_{x^* \in X^*} \{(x - x, x^*)_X + f(x)\} = f(x).$$

The reverse inequality requires a bit more work. Fix $x \in X$ and $\alpha < f(x)$. Apply Corollary 1.5; there is an $x^* \in X^*$ and $c \in \mathbb{R}$ be such that

$$\langle x, x^* \rangle_X - \alpha > c$$

$$\langle z, x^* \rangle_X - f(z) < c \quad z \in \text{dom}(f)$$

Then $f^*(x^*) \leq c$ and thus $\langle x, x^* \rangle_X - \alpha > f^*(x^*)$. Thus

$$\alpha < \langle x, x^* \rangle_X - f^*(x^*) \leq f^{**}(x).$$

\hfill $\Box$

2. The Contraction Principle

We next identify how rare events transform.

THEOREM 2.1 (Contraction Theorem). Suppose that $\{X_\varepsilon\}_{\varepsilon > 0}$ is a collection of random variables which take values in a Polish space $X$ and which has a large deviations principle with rate function $I$. Let $Y$ be a second Polish space and let $f : X \rightarrow Y$ be continuous. Define $Y_\varepsilon \overset{\text{def}}{=} f(X_\varepsilon)$ for all $\varepsilon > 0$. Then $\{Y_\varepsilon\}_{\varepsilon > 0}$ has a large deviations principle with rate function $I_{\varepsilon} \overset{\text{def}}{=} \inf_{x \in X} I(x)$.

PROOF. Note that

$$\{y \in Y : I_{\varepsilon}(y) \leq s\} = f \left( \{x \in X : I(x) \leq s\} \right);$$

forward images of compacts sets are compact. To get the rest of the proof, note that for any measurable subset $A$ of $Y$,

$$\mathbb{P}\{Y_\varepsilon \in A\} = \mathbb{P}\{X_\varepsilon \in f^{-1}(A)\},$$

and that

$$\inf_{x \in f^{-1}(A)} I(x) = \inf_{y \in A} I_{\varepsilon}(y).$$

\hfill $\Box$

The point of this is the following. Often one tries to find a large deviations principle for a “fundamental” problem and then pass it through a system. Another use is that in some cases a large deviations principles can be related to another large deviations principle “upstairs”.

EXAMPLE 2.2. Let $\{\xi_n\}_{n \in \mathbb{N}}$ be an i.i.d. collection of random variables with common law $\mu$ and which take values in some Polish space $X$. For each $N \in \mathbb{N}$, define

$$\mu_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{\xi_n};$$

this is a random element of $\mathcal{P}(X)$. Later on, we will see that $\{\mu_N\}_{N \in \mathbb{N}}$ has a large deviations principle with rate function

$$f^{(1)}(\nu) = \begin{cases} \int_{x \in X} \ln \frac{d\nu}{d\mu}(x) \nu(dx) & \text{if } \nu \ll \mu \\ \infty & \text{else} \end{cases}$$

(this is entropy). Assume now that $X = \mathbb{R}^d$ and furthermore that

$$\int_{x \in \mathbb{R}^d} \exp \{\langle \theta, x \rangle_{\mathbb{R}^d}\} \mu(dx) < \infty$$

11
for \( \theta \) in a small neighborhood of the origin. Define now

\[
S_N = \frac{1}{N} \sum_{n=1}^{N} \xi_n = \int_{x \in \mathbb{R}^d} x \mu_N(dx).
\]

these are random elements of \( \mathbb{R}^d \). We will also see later that \( S_N \) has a large deviations principle with rate function

\[
I^{(2)}(x) = \sup_{\theta \in \mathbb{R}^d} \left\{ \langle \theta, x \rangle - \ln \int_{x' \in \mathbb{R}^d} \exp \left[ \langle \theta, x' \rangle \right] \mu(dx') \right\}.
\]

Clearly these two should be related.

**Example 2.3.** Let \( W \) be a Brownian motion. For each \( \varepsilon \in (0, 1) \), define \( X_\varepsilon(t) = \sqrt{\varepsilon}W_t \). Then \( X_\varepsilon \) is a \( C[0,1] \)-valued random variable. We will later see that \( X_\varepsilon \) has a large deviations principle with rate function

\[
I(\phi) = \begin{cases} 
\frac{1}{2} \int_0^1 (\dot{\phi}(s))^2 ds & \text{if } \dot{\phi} \in L^2 \\
\text{else} &
\end{cases}
\]

Fix now \( b: \mathbb{R} \rightarrow \mathbb{R} \) which is Lipshitz-continuous and a \( y_o \in \mathbb{R} \). Then we can solve the SDE

\[
dY_\varepsilon^t = b(Y_\varepsilon^t) dt + \sqrt{\varepsilon}dW_t \quad t \in [0,1] \\
Y_\varepsilon^0 = y_o
\]

It turns out that we can write \( Y_\varepsilon \) as a nonlinear transformation of \( X_\varepsilon \), so we can immediately get a large deviations principle for \( Y_\varepsilon \).

**Exercises**

1. Let \( I_\varepsilon \equiv \varepsilon^{-5}e^{-3/\varepsilon} + \varepsilon^4e^{-7/\varepsilon} \). Compute \( \lim_{\varepsilon \rightarrow 0} \varepsilon \ln I_\varepsilon \).
2. Let \( X \) be geometric with parameter \( p \in (0, 1) \); i.e., \( P\{X = k\} = (1-p)^{k-1}p \) for \( k \in \mathbb{N} \), \( 0 \) else.
3. Compute \( \lim_{N \rightarrow \infty} \frac{1}{N} \ln P\{X \geq N\} \).
4. Let \( X \) be negative binomial \((2,p)\) (for some \( p \in (0, 1) \)); i.e., \( X \) is the position of the second heads in an i.i.d sequence of coin flips where each coin has probability \( p \) of being heads. More precisely,

\[
P\{X = k\} = \begin{cases} 
(k-1)p^2(1-p)^{k-2} & \text{for } k \in \{2,3\ldots\}\\
0 & \text{else}
\end{cases}
\]

Compute \( \lim_{N \rightarrow \infty} \frac{1}{N} \ln P\{X \geq N\} \).
5. Define \( I_N \equiv \frac{3}{N^2} + \frac{7}{N^4} \) for all \( N \in \mathbb{N} \). Find a nondecreasing function \( a: \mathbb{N} \rightarrow (0,\infty) \) such that \( \lim_{N \rightarrow \infty} a(N) = \infty \) and such that

\[
\lim_{N \rightarrow \infty} \frac{1}{a(N)} \ln I_N
\]

exists and is nontrivial. In other words, polynomials can also fit into our framework.
6. Prove (3)
7. Prove Laplace asymptotics if \( x^* = 0 \) and \( \dot{f}(x^*) > 0 \); i.e., the minimizer is at an endpoint.
8. Prove Laplace asymptotics if \( x^* \in (0,1) \) but \( \dot{f}(x^*) = 0 \) and \( f^{(4)}(x^*) > 0 \).
9. Show that

\[
\int_{x=x}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \leq \frac{1}{2} \exp^{-x^2/2}
\]

for all \( x > 0 \).
(9) Suppose that $\phi : \mathbb{R} \to \mathbb{R}$ is a measurable function such that
\[
\int_{x \in \mathbb{R}} e^{-\phi(x)} dx < \infty.
\]
Suppose further that there is a $\phi > 0$ such that $\phi \geq \frac{1}{\varepsilon}$ Lebesgue-almost surely. Show that
\[
\lim_{N \to \infty} \varepsilon \ln \int_{x \in \mathbb{R}} \exp \left\{ -\frac{1}{\varepsilon} \phi(x) \right\} dx \leq -\frac{1}{\varepsilon}.
\]
This calculation is relevant for localizing the region of integration to a compact set.

(10) Suppose that $\{X_n\}_{n \in (0,1)}$ is a collection of random variables such that, for some $f \in C^2[0,1]$,
\[
P\{X_n \in A\} = c_n \int_{x \in A} \exp \left\{ -\frac{1}{\varepsilon} f(x) \right\} dx
\]
for all $A \in \mathcal{B}[0,1]$, where
\[
c_n \overset{\text{def}}{=} \left\{ \int_{x \in A} \exp \left\{ -\frac{1}{\varepsilon} f(x) \right\} dx \right\}^{-1}
\]
Assume that $\lim_{N \to \infty} \varepsilon \ln c_n = 0$.
(a) Show that $\inf_{x \in [0,1]} f(x) = 0$.
(b) Show that $\lim_{\varepsilon \to 0} X_n = x^*$ in probability.

(11) Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a Lebesgue-integrable function such that $\lim_{r \to \infty} r^\alpha g(r) = c$ for some $\alpha > 0$ and some positive constant $c > 0$. Define
\[
f(x) = \int_{\sigma = 0}^{\infty} \exp \left\{ -\frac{1}{\varepsilon} f(r) \right\} dr \quad x > 0
\]
find the behavior of $f(x)$ for $x$ large. The idea is to get “fat tails” via a “stochastic volatility”.

(12) Suppose that $X_n$ is Gaussian with mean 0 and variance $\varepsilon^2 \sigma^2$, where $\sigma$ is some fixed nonzero parameter.
(a) Write the density of $X_n$.
(b) Compute $E[\exp(\theta X_n)]$ for all $\theta \in \mathbb{R}$.
(c) For each $\theta \in \mathbb{R}$, compute $\Lambda(\theta) \overset{\text{def}}{=} \lim_{\varepsilon \to 0} \varepsilon^2 \ln E[\exp(\frac{1}{\varepsilon} \theta X_n)]$.
(d) Compute $\Lambda^*(x) \overset{\text{def}}{=} \sup_{\theta \in \mathbb{R}} \{\theta x - \Lambda(\theta)\}$.

(13) Fix $f : \mathbb{R}^n \to \mathbb{R}$. Define
\[
f^*(\theta) \overset{\text{def}}{=} \sup_{x \in \mathbb{R}^n} \{\langle \theta, x \rangle_{\mathbb{R}^n} - f(x)\}
\]
for all $\theta \in \mathbb{R}^n$. Fix scalars $a$ and $b$ in $\mathbb{R}$ and a vector $c \in \mathbb{R}^n$. Define
\[
\tilde{f}(x) = af(b(x - c)) \quad x \in \mathbb{R}^n
\]
Compute $\tilde{f}$ in terms of $f^*$.

(14) Let $X_N$ be exponential with parameter $N\lambda$; i.e., $P\{X_N \geq t\} = e^{-\lambda N t}$ for all $t \geq 0$.
(a) Show that $X_N \to 0$ in probability.
(b) Compute $\Lambda(\theta) \overset{\text{def}}{=} \lim_{N \to \infty} \frac{1}{N} \ln E[\exp(\theta X_N)]$ for all $\theta \in \mathbb{R}$ (hint: the answer is very degenerate).
(c) Compute $I(x) = \sup_{\theta \in \mathbb{R}} \{\theta x - \Lambda(\theta)\}$.

(15) Fix constants $a$, $b$, and $c$ in $\mathbb{R}$. Compute $f^*$ for the following functions:
(a) when $a < b$,
\[
f(x) = \begin{cases} c & \text{if } a < x < b \\ \infty & \text{else} \end{cases}
\]
(b) when $a$ and $b$ are nonnegative,
\[
f(x) = \begin{cases} ax & \text{if } x > 0 \\ -bx & \text{if } x < 0 \end{cases}
\]

(16) Let $X$ be uniformly on $(0, 1)$.
(a) For each \( \theta \in \mathbb{R} \), compute
\[
\Lambda(\theta) \stackrel{\text{def}}{=} \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{E} \left[ \exp \left[ N \theta X \right] \right]
\]
(b) Compute \( I(x) \stackrel{\text{def}}{=} \sup_{\theta \in \mathbb{R}} \{ x\theta - \Lambda(\theta) \} \).

(17) Consider the coin flips of (4).
(a) Show that \( \lim_{N \to \infty} X_N = p \).
(b) Compute \( \Lambda(\theta) \stackrel{\text{def}}{=} \ln \mathbb{E} \left[ \exp \left[ \theta \xi_n \right] \right] \) for all \( \theta \in \mathbb{R} \).
(c) Compute \( I(x) \stackrel{\text{def}}{=} \sup_{\theta \in \mathbb{R}} \{ \theta x - \Lambda(\theta) \} \) for all \( x \in \mathbb{R} \).
(d) Show that \( p \mapsto H(p, \alpha) \) is increasing on \([0,1]\) and decreasing on \((0,\alpha)\).

(18) Show that Definitions 1.1 and 1.2 are equivalent.

(19) Fix a space \( V \) and a map \( f : V \to (-\infty, \infty) \).
(a) Assume that \( V \) is a vector space. Show that \( f \) is convex if and only if \( f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \) for all \( x \) and \( y \) in \( V \) and \( \lambda \in [0,1] \).
(b) Assume that \( V \) is topological. Show that \( f \) is lower semicontinuous if and only if \( \{ v \in V : f(v) \leq s \} \) is closed for all \( s \in \mathbb{R} \).
(c) Assume that \( V \) is topological. Show that \( f \) is lower semicontinuous if and only if \( f(x) \leq \lim_{n \to \infty} f(x_n) \) if \( x = \lim_{n \to \infty} x_n \).

(20) Suppose that \( f : V \to (-\infty, \infty) \), where \( V \) is a topological vector space. Show that
\[
f^\ast(x) = \sup \left\{ f(x) : f : V \to (-\infty, \infty) \text{ is convex, lower semicontinuous, and } \tilde{f} \leq f \text{ everywhere} \right\}.
\]

(21) Fix an \( \mathbb{R}^d \)-valued random variable \( X \). Show that
\[
\Lambda(\theta) \stackrel{\text{def}}{=} \ln \mathbb{E} \left[ \exp \left[ \langle \theta, X \rangle_{\mathbb{R}^d} \right] \right] \quad \theta \in \mathbb{R}^d
\]
is a convex function.

(22) Let \( f \) be a function on a vector space \( V \).
(a) Show that \( f^\ast \) is convex.
(b) Show that \( f^\ast \) is lower semicontinuous (hint: write \( \{ v^\ast \in V^\ast : f^\ast(v^\ast) > s \} \) as a union of open half-spaces for each \( s \in \mathbb{R} \).

(23) Suppose that \( \{ f_n \}_{n \in \mathbb{N}} \) is a collection of convex functions on a vector space \( V \). Suppose that \( f \stackrel{\text{def}}{=} \lim_{n \to \infty} f_n \) exists pointwise. Show that \( f \) is convex.

(24) Prove (10) and (11). One direction of (10) is very easy. The other direction requires some compactness.

(25) Suppose that \( \{ X^1_\varepsilon \}_{\varepsilon > 0} \) and \( \{ X^2_\varepsilon \}_{\varepsilon > 0} \) are, respectively, collections of \( \mathbb{R}^d_1 \)-valued and \( \mathbb{R}^d_2 \)-valued random variables with large deviations principles with respective action functionals \( I_1 \) and \( I_2 \). Assume that \( X^1_\varepsilon \) and \( X^2_\varepsilon \) are independent for each \( \varepsilon > 0 \). Define \( Y_\varepsilon \stackrel{\text{def}}{=} (X^1_\varepsilon, X^2_\varepsilon) \) (i.e., \( Y_\varepsilon \) is \( \mathbb{R}^{d_1+d_2} \)-valued). Show that \( \{ Y_\varepsilon \}_{\varepsilon > 0} \) has a large deviations principle and find its rate function.

(26) For Example 2.2, show that
\[
I^{(2)}(x) = \inf \left\{ I^{(1)}(\nu) : \int_{x' \in \mathbb{R}^d} x' \nu(dx') = x \right\}.
\]

(27) Consider Example 2.3. For each \( \varphi \in C[0,1] \), let \( T(\varphi) \in C[0,1] \) be the solution of the integral equation
\[
(T(\varphi))(t) = y_0 + \int_{s=0}^{t} b((T(\varphi)(s)) \, ds + \varphi(t) \quad t \in [0,1]
\]
(a) Show that \( T \) is a Lipschitz-continuous map from \( C[0,1] \) to itself.
(b) Find a stochastic differential equation for \( T(X^\varepsilon) \).
(c) Find a large deviations principle for \( T(X^\varepsilon) \). Be as explicit as possible.
CHAPTER 2

The Gärtner-Ellis Theorem

Our goal here is to formalize the calculations of Subsection 1.2. This will involve a number of important but disparate thoughts, so we will only connect things together at the end.

For reasons which will be clear a bit later, let’s set things up in a canonical way and have the measures depend on \( \varepsilon \) rather than the random variables. Our measurable space will be \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\), and we will define \( X(\omega) \equiv \omega \) for all \( \omega \in \mathbb{R}^d \).

Fix \( \{\mathbb{P}_\varepsilon\}_{\varepsilon > 0} \subset \mathcal{P}(\mathbb{R}^d) \), the collection of probability measures on \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\). For each \( \varepsilon > 0 \), let \( \mathbb{E}_\varepsilon \) be the associated expectation operator. For each \( \theta \in \mathbb{R}^d \) and \( \varepsilon > 0 \), define
\[
(13) \quad \Lambda_\varepsilon(\theta) \equiv \varepsilon \ln \mathbb{E}_\varepsilon \left[ \exp \left( \frac{1}{\varepsilon} \langle \theta, X \rangle_{\mathbb{R}^d} \right) \right].
\]

Suppose that \( \Lambda(\theta) \equiv \lim_{\varepsilon \downarrow 0} \Lambda_\varepsilon(\theta) \) exists for all \( \theta \in \mathbb{R}^d \). We define
\[
(14) \quad I(x) \equiv \sup_{\theta \in \mathbb{R}^d} \left\{ \langle \theta, x \rangle_{\mathbb{R}^d} - \Lambda(\theta) \right\}, \quad x \in \mathbb{R}^d.
\]

0.1. Structure of \( I \) and upper bound. Let’s start with deterministic calculations.

**Lemma 0.4.** We have that \( I \) is lower semicontinuous.

**Proof.** Assume that \( x = \lim_{n \to \infty} x_n \). For any \( \theta \in \mathbb{R} \),
\[
\lim_{n \to \infty} I(x_n) \geq \lim_{n \to \infty} \left\{ \langle \theta, x_n \rangle_{\mathbb{R}^d} - \Lambda(\theta) \right\} = \langle \theta, x \rangle_{\mathbb{R}^d} - \Lambda(\theta).
\]

Thus \( I(x) \leq \liminf_{n \to \infty} I(x_n) \), giving us the claimed semicontinuity. \( \square \)

Let’s next prove a bound which will be central to our calculations.

**Lemma 0.5 (Exponential Chebychev).** Fix \( \theta \in \mathbb{R}^d \) and \( L \in \mathbb{R} \). Then
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \left\{ \langle \theta, X \rangle_{\mathbb{R}^d} \geq L \right\} \leq -(L - \Lambda(\theta)).
\]

**Proof.** We have that
\[
\mathbb{P}_\varepsilon \left\{ \langle \theta, X \rangle_{\mathbb{R}^d} \geq L \right\} = \mathbb{P}_\varepsilon \left\{ \left[ \frac{1}{\varepsilon} \langle \theta, X \rangle_{\mathbb{R}^d} \right] \geq \frac{L}{\varepsilon} \right\} \leq e^{-L/\varepsilon} \mathbb{E} \left[ \exp \left( \frac{1}{\varepsilon} \langle \theta, X \rangle_{\mathbb{R}^d} \right) \right] = \exp \left( - \frac{1}{\varepsilon} (L - \Lambda_\varepsilon(\theta)) \right).
\]

The claim follows. \( \square \)

Let’s next start to work on the upper large deviations bound.

**Lemma 0.6.** For every \( K \subset \subset \mathbb{R}^d \),
\[
(15) \quad \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in K \} \leq - \inf_{x \in K} I(x).
\]

**Proof.** First fix \( s < \inf_{x \in K} I(x) \), and note that
\[
K \subset \{ x \in \mathbb{R}^d : I(x) > s \} = \bigcup_{\theta \in \mathbb{R}} \{ x \in \mathbb{R} : \langle \theta, x \rangle_{\mathbb{R}^d} > s + \Lambda(\theta) \}
\]

Since we thus cover \( K \) by a collection of open sets, we can extract a finite subcover; there is a finite subset \( \Theta \) of \( \mathbb{R}^d \) such that
\[
K \subset \bigcup_{\theta \in \Theta} \{ x \in \mathbb{R} : \langle \theta, x \rangle_{\mathbb{R}^d} > s + \Lambda(\theta) \}
\]
Thus
\[ \mathbb{P}_\varepsilon \{ X \in K \} \leq \sum_{\theta \in \Theta} \mathbb{P}_\varepsilon \{ (\theta, x)_{\mathbb{R}^d} > s + \Lambda(\theta) \} \]
and thus
\[ \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in K \} \leq \max_{\theta \in \Theta} \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ (\theta, x)_{\mathbb{R}^d} > s + \Lambda(\theta) \} \leq -s \]
where we have used Lemma 0.5.

Let’s now get some regularity at infinity. For each \( L > 0 \), define
\[ \bar{B}(L) \overset{\text{def}}{=} \{ x \in \mathbb{R}^d : \| x \| \leq L \} . \]

**Lemma 0.7.** Assume that
\[ \mathbf{0} \in \text{int}(\text{dom}(\Lambda)) \]
(i.e., the origin is in the interior of dom(\( \Lambda \))). Then \( \{ x \in \mathbb{R}^d : I(x) \leq s \} \) is bounded for all \( s \geq 0 \). Furthermore,
\[ \lim_{L \to \infty} \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \not\in \bar{B}(L) \} = -\infty . \]

**Proof.** To start, let \( \{ e_1, e_2, \ldots, e_d \} \) be the standard basis vectors in \( \mathbb{R}^d \). Define \( \mathcal{E}^\pm \overset{\text{def}}{=} \{ \pm e_1, \pm e_2, \ldots, \pm e_d \} . \)
By (16), there is a \( \delta > 0 \) such that \( K \overset{\text{def}}{=} \sup_{\mathbf{e} \in \mathcal{E}^\pm} \Lambda(\delta \mathbf{e}) \) is finite.
If \( I(x) \leq s \), then \( \langle \delta \mathbf{e}, x \rangle_{\mathbb{R}^d} - \Lambda(\delta \mathbf{e}) \leq s \) for all \( \mathbf{e} \in \mathcal{E}^\pm \); hence
\[ \sup_{\mathbf{e} \in \mathcal{E}^\pm} \langle \delta \mathbf{e}, x \rangle_{\mathbb{R}^d} \leq \frac{s + K}{\delta} . \]

Similarly,
\[ \mathbb{P}_\varepsilon \{ \| X \|_\infty \geq L \} = \mathbb{P}_\varepsilon \{ \langle X, \mathbf{e} \rangle_{\mathbb{R}^d} \geq L \text{ for some } \mathbf{e} \in \mathcal{E}^\pm \} \leq \sum_{\mathbf{e} \in \mathcal{E}^\pm} \mathbb{P}_\varepsilon \{ \langle X, \mathbf{e} \rangle_{\mathbb{R}^d} \geq L \} \]
\[ = \sum_{\mathbf{e} \in \mathcal{E}^\pm} \mathbb{P}_\varepsilon \{ \langle X, \delta \mathbf{e} \rangle_{\mathbb{R}^d} \geq \delta L \} . \]
Thus
\[ \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ \| X \|_\infty \geq L \} \leq \max_{\mathbf{e} \in \mathcal{E}^\pm} \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ \langle X, \delta \mathbf{e} \rangle_{\mathbb{R}^d} \geq \delta L \} \leq \max_{\mathbf{e} \in \mathcal{E}^\pm} \{ L \delta - \Lambda(\delta \mathbf{e}) \} . \]
This leads to (17). \( \square \)

The property (17) is called **exponential tightness**; here it is framed in terms of balls, which are compact in finite-dimensional spaces.

**Corollary 0.8.** If (16) holds, then
- For each \( s \geq 0, \{ x \in \mathbb{R}^d : I(x) \leq s \} \subset \subset \mathbb{R}^d \)
- For \( F \subset \mathbb{R}^d \) closed,
\[ \lim_{\varepsilon \searrow 0} \epsilon \ln \mathbb{P}_\varepsilon \{ X \in F \} \leq -\inf_{x \in F} I(x) . \]

**Proof.** Lemma 0.4 ensures that the level sets of \( I \) are closed; Corollary 0.8 ensures that they are also bounded. Fix next \( F \subset \mathbb{R}^d \) closed. Fix \( s < \inf_{x \in F} I(x) \). By Lemma 0.7, we can then find an \( L > 0 \) such that
\[ \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \not\in \bar{B}(L) \} \leq -s . \]
Also note that \( \inf_{x \in F \cap \bar{B}(L)} I(x) > s \). We have that
\[ \mathbb{P}_\varepsilon \{ X \in F \} \leq \mathbb{P}_\varepsilon \{ X \in F \cap \bar{B}(L) \} + \mathbb{P}_\varepsilon \{ X \not\in \bar{B}(L) \} . \]
Thus
\[ \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in F \} \leq \max \left( \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in F \cap \bar{B}(L) \}, \lim_{\varepsilon \searrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \not\in \bar{B}(L) \} \right) \leq -s . \]
0.2. Some More Convex Analysis. We now want to prove the lower bound. This will take a lot of work and require a number of concepts from convex analysis; our treatment is taken from [DZ98]. Since the \( \Lambda_v \)'s are convex, so is \( \Lambda \). Furthermore, \( I \) is convex. We define

\[
\text{ri}(\text{dom}(I)) = \{ x \in \text{dom}(I) : \text{if } y \in \text{dom}(I) \text{ then } x + \varepsilon(x - y) \in \text{dom}(I) \text{ for some } \varepsilon > 0 \}
\]

this is the \emph{relative interior} of \( \text{dom}(I) \).

Before proceeding, let's prove a regularity result.

**Lemma 0.9.** Let \( f : V \to (-\infty, \infty] \) be convex, where \( V \) is a finite-dimensional vector space. Then \( f \) is locally bounded on \( \text{int dom}(f) \).

**Proof.** Fix \( x^* \in \text{int dom}(f) \). Define \( \tilde{f}(x) \defeq f(x^* + x) - f(x^*) \). Then \( \tilde{f} \) is convex, \( \tilde{f}(0) = 0 \), and \( 0 \in \text{int dom}(\tilde{f}) \). We want to show that \( \tilde{f} \) is locally bounded near 0.

Let \( V \) be a basis of \( V \); since \( V \) is finite-dimensional, \( V \) is a finite set. Define \( V^\pm = V \cup (-V) \cup \{0\} \); i.e., \( V \) is the collection of basis vectors and their negatives, and also 0. Since \( 0 \in \text{int dom}(\tilde{f}) \), there is a \( \delta > 0 \) such that \( \delta v \in \text{dom}(\tilde{f}) \) for all \( v \in V^\pm \). Define \( M \defeq \max_{v \in V^\pm} |\tilde{f}(\delta v)| \); then \( M < \infty \).

Define now

\[
O \defeq \left\{ \sum_{v \in V} \lambda_v v : \sum_{v \in V} |\lambda_v| < \delta \right\}.
\]

Then \( O \) is an open neighborhood of 0. Fix \( x = \sum_{v \in V} \lambda_v v \in O \). For each \( v \in V \), define \( s_v \defeq \text{sgn}(\lambda_v) \); i.e., \( s_v = \lambda_v / |\lambda_v| \) if \( \lambda_v \neq 0 \), and \( s_v = 0 \) if \( \lambda_v = 0 \). Note that \( s_v v \in V^\pm \) for all \( v \in V \). We then write

\[
x = \left( \sum_{v \in V} |\lambda_v| \right) \sum_{v \in V} |\lambda_v| s_v v + \left( 1 - \frac{\sum_{v \in V} |\lambda_v|}{\delta} \right) 0.
\]

Convexity then implies that

\[
\tilde{f}(x) \leq \frac{\sum_{v \in V} |\lambda_v|}{\delta} \tilde{f}(\delta s_v v) \leq M.
\]

Thus \( \sup_{x \in O} \tilde{f}(x) \leq M \); i.e., \( \tilde{f} \) is bounded from above on \( O \).

Fix next \( x \in O \). Since \( O = -O \), \( -x \in O \), so \( \tilde{f}(-x) \leq M \). Suppose that \( \tilde{f}(x) < -M \); then there is an \( s < -M \) such that \( (x, s) \in \text{epi}(\tilde{f}) \). Clearly \( (-x, -M) \in \text{epi}(\tilde{f}) \) so by definition of convexity, \( (0, \frac{1}{2}(s + M)) \in \text{epi}(\tilde{f}) \); thus \( \tilde{f}(x) \leq \frac{1}{2}(s + M) < 0 \). But this contradicts the fact that \( \tilde{f}(0) = 0 \). Thus \( \tilde{f}(x) \geq -M \), so \( \inf_{x \in O} \tilde{f}(x) \geq -M \). \( \square \)

In other words, we can't have \( \lim_{y \to x} |f(y)| = \infty \) if \( x \in \text{int dom}(f) \).

Our first technical result is that if \( x \in \text{ri}(\text{dom}(I)) \), then the sup is actually a max in the definition of \( I \).

**Lemma 0.10.** Suppose that \( \Lambda \) is lower semicontinuous and that \( x \in \text{ri}(\text{dom}(I)) \). Then there is a \( \Theta_x \in \text{dom}(\Lambda) \) such that

\[
I(x) = \langle x, \Theta_x \rangle_{R^d} - \Lambda(\Theta_x).
\]

**Proof.** The basic idea is to find a supporting hyperplane for the subderivative of \( I \) at \( x \). The fact that the subderivative is homogeneous then plays an important role.

For \( v \in R^d \), define

\[
g(v) \defeq \inf_{\delta > 0} \frac{I(x + \delta v) - I(x)}{\delta}.
\]

Since \( g \) is the inf of a collection of convex functions, it is convex. Fix \( 0 < \delta_1 < \delta_2 \) and \( v \in R^d \). Then

\[
\frac{I(x + \delta_1 v) - I(x)}{\delta_1} = \frac{\frac{\delta_1}{\delta_2} I(x + \delta_2 v) + \left( 1 - \frac{\delta_1}{\delta_2} \right) x - I(x)}{\delta_1} \leq \frac{\frac{\delta_1}{\delta_2} I(x + \delta_2 v) + \left( 1 - \frac{\delta_1}{\delta_2} \right) I(x) - I(x)}{\delta_1} \leq \frac{I(x + \delta_2 v) - I(x)}{\delta_2}
\]
(this is a standard calculation) so in fact

\[ g(v) = \lim_{\delta \searrow 0} \frac{I(x + \delta v) - I(x)}{\delta}. \]

Note that \( g(\alpha v) = \alpha g(v) \) for all \( \alpha > 0 \); thus \( g(v) = \|v\|g\left(\frac{v}{\|v\|}\right) \) for all \( v \in \mathbb{R}^d \setminus \{0\} \). We would like to use the Lemma 1.3 to find the required \( \Theta_x \).

We next claim that \( \text{dom}(g) \) is in fact a vector space. Clearly it is convex and is a cone. Next, note that if \( g(v) < \infty \), then \( x + \delta v \in \text{dom}(I) \) for some \( \delta > 0 \). Since \( x \in \text{ri(dom}(I)) \), there is an \( \varepsilon > 0 \) such that \( x - \varepsilon \delta v = x + \varepsilon(x - (x + \delta v)) \in \text{dom}(I) \). Thus \( -v \in \text{dom}(g) \); thus \( -\text{dom}(g) \subset \text{dom}(g) \), so in fact \( \text{dom}(g) = -\text{dom}(g) \). Fix \( v \in \text{dom}(g) \) and \( \alpha \in \mathbb{R} \). Since \( \text{dom}(g) \) is a cone, \( |\alpha|v \in \text{dom}(g) \). Since \( \text{dom}(g) = -\text{dom}(g) \), \( \alpha v \in \text{dom}(g) \); i.e., \( \text{dom}(g) \) is a symmetric cone. Next fix \( v_1 \) and \( v_2 \) in \( \text{dom}(g) \) and \( \alpha_1 \) and \( \alpha_2 \) in \( \mathbb{R} \). Since \( \text{dom}(g) \) is a symmetric cone, \( 2\alpha_1 v_1 \) and \( 2\alpha_2 v_2 \) are both in \( \text{dom}(g) \). Since \( \text{dom}(g) \) is convex, we finally have that \( \alpha_1 v_1 + \alpha_2 v_2 = \frac{1}{2}(2\alpha_1 v_1 + 2\alpha_2 v_2) \in \text{dom}(g) \). Thus \( \text{dom}(g) \) is indeed a vector space, which we denote as \( V \); since \( V \subset \mathbb{R}^d \), \( V \) is finite-dimensional. By Lemma 0.9, \( g \) is bounded on compact sets.

The only thing we are missing is lower semicontinuity. We can modify things so that this happens. For each \( v \in \mathbb{R}^d \), define \( \bar{g}(v) \stackrel{\text{def}}{=} \lim_{\nu \searrow v} g(\nu) \); this is the lower-semicontinuous regularization of \( g \); it is indeed lower semicontinuous and it retains the convexity of \( g \).

We next apply Lemma 1.3 to separate \((0, -1)\) from \( \text{epi}(\bar{g}) \). Note that since \( g \) is bounded on compact sets, it is bounded on the unit sphere; by homogeneity, we must indeed have that \( \bar{g}(0) = 0 \). Thus by Lemma 1.3, there is a \( \Theta_x \in \mathbb{R}^d \) and \( c \in \mathbb{R} \) such that \( \langle z, \Theta_x \rangle_{\mathbb{R}^d} - \bar{g}(z) < c \) for all \( z \in \text{dom}(\bar{g}) \). Since \( g \geq \bar{g} \), for any \( z \in \text{dom}(g) \) and \( \alpha > 0 \), we have that

\[ g(\alpha z) > \alpha \langle z, \Theta_x \rangle_{\mathbb{R}^d} - c \]

and thus

\[ g(z) > \langle z, \Theta_x \rangle_{\mathbb{R}^d} - \frac{c}{\alpha}. \]

Take now \( \alpha \not\to \infty \). Unwinding things, we have that

\[ I(x + v) \geq I(x) + \langle \Theta_x, v \rangle_{\mathbb{R}^d} \]

for all \( v \in \mathbb{R}^d \). We now can use Lemma 1.2 to write \( \Lambda(\Theta_x) \) in terms of \( I \) (this is where we use the lower semicontinuity of \( \Lambda \)). We clearly have that

\[ \Lambda(\Theta_x) = \langle x, \Theta_x \rangle_{\mathbb{R}^d} - I(x). \]

However, we also have that for any \( v \in \mathbb{R}^d \),

\[ \langle x + v, \Theta_x \rangle_{\mathbb{R}^d} - I(x + v) \leq \langle x, \Theta_x \rangle_{\mathbb{R}^d} + I(x) \]

so in fact \( \Lambda(\Theta_x) = \langle x, \Theta_x \rangle_{\mathbb{R}^d} - I(x) \) which gives us the claim (and ensures that \( \Theta_x \in \text{dom}(\Lambda) \)).

We use the fact that \( x \in \text{ri(dom}(I)) \) to bound the subgradient from below.

Let’s next identify some assumptions under which \( \Lambda \) is differentiable at any solution of (18).

**Definition 0.1.** We say that \( \Lambda \) is essentially smooth if

1. \( \text{int(dom}(\Lambda) \neq \emptyset \)
2. \( \Lambda \) is differentiable on \( \text{int(dom}(\Lambda) \)
3. If \( \theta \in \partial \text{int(dom}(\Lambda) \), then \( \lim_{\theta' \to \theta} \|\nabla \Lambda(\theta')\| = \infty \).

Note that in the last requirement, we impose no condition on \( \lim_{\theta' \to \text{int(dom}(\Lambda) \) \( \|\theta'\| \to \infty \) \( \|\nabla \Lambda(\theta')\| \). Then we have

**Lemma 0.11.** Assume that \( \Lambda \) is essentially smooth and that \( \text{dom}(I) \neq \emptyset \). If \( \Theta_x \in \mathbb{R}^d \) satisfies (18), then \( \Theta_x \in \text{int(dom}(\Lambda) \).
PROOF. By the first requirement of essential smoothness, we can find \( \Theta^* \in \text{int dom}(\Lambda) \). By Lemma 0.9, for \( \delta > 0 \) sufficiently small, \( \Lambda \) is bounded on \( \Theta^* + B(\delta) \); i.e., \( M_{\delta} \overset{\text{def}}{=} \sup_{v \in B(\delta)} \Lambda(\Theta^* + v) \) is finite.

Let’s now use the fact that \( \Theta_x \) and \( \Theta^* + B(\delta) \) are both in \( \text{dom}(\Lambda) \) and that \( \text{dom}(\Lambda) \) is convex. For each \( \lambda \in [0, 1] \), define \( O_\lambda \overset{\text{def}}{=} (1 - \lambda)\Theta_x + \lambda(\Theta^* + B(\delta)) \) and set \( \theta_\lambda \overset{\text{def}}{=} (1 - \lambda)\Theta_x + \lambda\Theta^* \). Then \( O_\lambda \in \text{dom}(\Lambda) \) for each \( \lambda \in [0, 1] \), and since \( O_\lambda \) is open for each \( \lambda \in (0, 1] \), in fact \( O_\lambda \in \text{int dom}(\Lambda) \) for each \( \lambda \in (0, 1] \). By the second requirement of essential smoothness, we know that \( \Lambda \) is thus differentiable at \( \theta_\lambda \) for each \( \lambda \in (0, 1] \).

Let’s bound \( \nabla \Lambda(\theta_\lambda) \). Fix \( v \in B(\delta) \) and \( s \in (0, 1] \). Then

\[
\theta_\lambda + s\lambda v = (1 - \lambda)\Theta_x + \lambda(\Theta^* + sv) \in O_\lambda.
\]

\[
\Lambda(\theta_\lambda + sv) - \Lambda(\theta_\lambda) = \Lambda(s(\theta_\lambda + \lambda v) + (1 - s)\theta_\lambda) - \Lambda(\theta_\lambda) \leq s\Lambda(\theta_\lambda + v) + (1 - s)\Lambda(\theta_\lambda) - \Lambda(\theta_\lambda) = s(\Lambda(\theta_\lambda + v) - \Lambda(\theta_\lambda)).
\]

Thus

\[
\langle \nabla \Lambda(\theta_\lambda), \lambda v \rangle_{\mathbb{R}^d} \leq \lim_{s \downarrow 0} \frac{\Lambda(\theta_\lambda + sv) - \Lambda(\theta_\lambda)}{s} \leq \Lambda(\theta_\lambda + v) - \Lambda(\theta_\lambda).
\]

Using (19) with \( s = 1 \), we have that

\[
\Lambda(\theta_\lambda + 1) \leq (1 - \lambda)\Lambda(\Theta_x) + \lambda \Lambda(\Theta^* + v) = \Lambda(\Theta_x) + \lambda(\Theta^* + v) - \Lambda(\Theta_x).
\]

We can use (18) to bound \( \Lambda(\theta_\lambda) \) from below; we have that

\[
\Lambda(\theta_\lambda) \geq \langle \theta_\lambda, x \rangle_{\mathbb{R}^d} - I(x) = (1 - \lambda) \langle \Theta_x, x \rangle_{\mathbb{R}^d} + \lambda \langle \Theta^* + v, x \rangle_{\mathbb{R}^d} - I(x)
\]

\[
= \langle \Theta_x, x \rangle_{\mathbb{R}^d} - I(x) + \lambda \langle \Theta^* - \Theta_x, x \rangle_{\mathbb{R}^d} = \Lambda(\theta_\lambda) + \lambda(\Theta^* - \Theta_x, x)_{\mathbb{R}^d}.
\]

Setting \( M' \overset{\text{def}}{=} M + \|\Theta^* - \Theta_x\|\|x\| \), we thus have that

\[
\|\Lambda(\theta_\lambda + v) - \Lambda(\theta_\lambda)\| \leq M'\lambda.
\]

Using this in (20), we get that

\[
\langle \nabla \Lambda(\theta_\lambda), \lambda v \rangle_{\mathbb{R}^d} \leq M'\lambda.
\]

Letting \( v \) vary over \( B(\delta) \), we get that \( \|\nabla \Lambda(\theta_\lambda)\| \). Finally let \( \lambda \searrow 0 \); then \( \lim_{\lambda \searrow 0} \theta_\lambda = \Theta_x \). By the last requirement of essential smoothness, we thus have that \( \Theta_x \in \text{int dom}(\Lambda) \setminus \partial \text{int dom}(\Lambda) = \text{int dom}(\Lambda) \).

Finally, we prove that if \( \Lambda \) is differentiable at \( \Theta_x \) solving (18), then we have several other properties.

Lemmas 0.12. Suppose that \( \Theta_x \) solves (18). If \( \Lambda \) is differentiable at \( \Theta_x \), then

- \( x = \nabla \Lambda(\Theta_x) \)
- \( I(x + v) \geq I(x) + \langle \Theta_x, v \rangle_{\mathbb{R}^d} \) for all \( v \in \mathbb{R}^d \).

PROOF. Suppose that (18) holds. Then for any \( v \in \mathbb{R}^d \) and \( \delta > 0 \),

\[
\Lambda(\Theta_x + \delta v) \geq \langle x + \delta v, \Theta_x \rangle_{\mathbb{R}^d} - I(x) = \langle \delta v, \Theta_x \rangle_{\mathbb{R}^d} - \Lambda(\Theta_x).
\]

Thus

\[
\langle v, x \rangle_{\mathbb{R}^d} \leq \lim_{\delta \downarrow 0} \frac{\Lambda(\Theta_x + \delta v) - \Lambda(\Theta_x)}{\delta} = \langle v, \nabla \Lambda(\Theta_x) \rangle_{\mathbb{R}^d},
\]

This holding for any \( v \in \mathbb{R}^d \), we indeed have the first claim.

Suppose now that the second claim is not true. Then there is a \( v \in \mathbb{R}^d \) such that

\[
I(x + v) \leq I(x) + \langle v, \nabla \Lambda(\Theta_x) \rangle_{\mathbb{R}^d}.
\]

Then for any \( \psi \in \mathbb{R}^d \) and any \( \delta > 0 \),

\[
\langle x + \delta v, \Theta_x + \delta \psi \rangle_{\mathbb{R}^d} - \Lambda(\Theta_x + \delta \psi) \leq \langle x, \nabla \Lambda(\Theta_x) \rangle_{\mathbb{R}^d} + \langle v, \nabla \Lambda(\Theta_x) \rangle_{\mathbb{R}^d}
\]

Thus

\[
\langle x + v, \psi \rangle_{\mathbb{R}^d} \leq \lim_{\delta \downarrow 0} \frac{\Lambda(\Theta_x + \delta v) - \Lambda(\Theta_x)}{\delta} = \langle \psi, \nabla \Lambda(\Theta_x) \rangle_{\mathbb{R}^d}.
\]

Since this holds for all \( \psi \in \mathbb{R}^d \), we must have that \( x + v = \nabla \Lambda(\Theta_x) = x \); i.e., \( v = 0 \).
0.3. Lower Bound. We can now get back to proving the lower large deviations bound.

**Lemma 0.13.** Assume that $\Lambda$ is essentially smooth. Fix $G \subset \mathbb{R}^d$ open. Then

$$\lim_{\epsilon \searrow 0} \epsilon \ln \mathbb{P}_\epsilon \{X \in G\} \geq - \inf \{I(x) : x \in G \cap \text{ri dom}(I)\}.$$ 

**Proof.** Fix $x \in G \cap \text{ri dom}(I)$. Since $x \in \text{ri dom}(I)$, Lemma 0.10 ensures that there is a $\Theta_x \in \text{dom}(\Lambda)$ which satisfies (18). Note that

$$\mathbb{E}_\epsilon \left[ \exp \left[ \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right] \right] = \exp \left[ \frac{1}{\epsilon} \{ \Lambda_x(\Theta_x) - \langle \Theta_x, x \rangle_{\mathbb{R}^d} \} \right] < \infty$$

We write things in a fairly clever way.

$$\mathbb{P}_\epsilon \{X \in G\} = \mathbb{E}_\epsilon \left[ \chi_G(X) \mathbb{E}_\epsilon \left[ \exp \left[ \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right] \right] \right]$$

where $\mathbb{E}_\epsilon \in \mathcal{P}(\mathbb{R}^d)$ is defined as

$$\mathbb{E}_\epsilon[A] \overset{\text{def}}{=} \frac{\mathbb{E}_\epsilon[\chi_A \exp \left[ \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right]]}{\mathbb{E}_\epsilon[\exp \left[ \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right]]} \quad A \in \mathcal{B}(\mathbb{R}^d)$$

Note that

$$\lim_{\epsilon \searrow 0} \epsilon \ln \mathbb{E}_\epsilon \left[ \chi_A \exp \left[ \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right] \right] = - \chi_G \left( \epsilon \{ \langle \Theta_x, x \rangle_{\mathbb{R}^d} - \Lambda(\Theta_x) \} = -I(x) \right)$$

Fix now $\delta > 0$ such that $x + B(\delta) \in G$. Then

$$\mathbb{E}_\epsilon \left[ \chi_G(X) \exp \left[ - \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right] \right] \geq \mathbb{E}_\epsilon \left[ \chi_{x + B(\delta)}(X) \exp \left[ - \frac{1}{\epsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right] \right]$$

$$\geq \mathbb{E}_\epsilon \left[ \chi_{x + B(\delta)}(X) \exp \left[ - \frac{\delta}{\epsilon} \right] \right]$$

We claim that

$$\lim_{\epsilon \searrow 0} \mathbb{P}_\epsilon \{X \in x + B(\delta)\} = 1.$$ 

We will do this via the upper bound. Define

$$\tilde{\Lambda}_\epsilon(\theta) \overset{\text{def}}{=} \epsilon \ln \mathbb{E}_\epsilon \left[ \exp \left[ \frac{1}{\epsilon} \langle \theta, X \rangle_{\mathbb{R}^d} \right] \right] = \Lambda_x(\Theta_x + \theta) - \Lambda_x(\Theta_x).$$

Thus

$$\tilde{\Lambda}(\theta) \overset{\text{def}}{=} \lim_{\epsilon \searrow 0} \left\{ \Lambda_x(\Theta_x + \theta) - \Lambda_x(\Theta_x) \right\} = \Lambda(\Theta_x + \theta) - \Lambda(\Theta_x)$$

for all $\theta \in \mathbb{R}^d$. Define also

$$\tilde{I}(x') = \sup_{\theta \in \mathbb{R}^d} \left\{ \langle \theta, x \rangle_{\mathbb{R}^d} - \tilde{\Lambda}(\theta) \right\} = I(x') - I(x) - \langle \Theta_x, x' - x \rangle_{\mathbb{R}^d}.$$

By Lemma 0.10 we knew that $\Theta_x \in \text{dom}(\Lambda)$; by using Lemma 0.11 we actually know that $\Theta_x \in \text{int dom}(\Lambda)$. Thus $0 \in \text{int dom}(\tilde{\Lambda})$. Thus we can use the upper large deviations bound Corollary 0.8. Note that $(x + B(\delta))^c$ is closed.

We thus have that

$$\lim_{\epsilon \searrow 0} \epsilon \ln \mathbb{P}_\epsilon \{X \notin x + B(\delta)\} \leq - \inf_{x' \notin x + B(\delta)} I(x') = - \min_{x' - x \notin B(\delta)} \{I(x') - I(x)\}.$$ 

By now using the strict inequality in the second claim of Lemma 0.12, we get that

$$\min_{x' - x \notin B(\delta)} \{I(x') - I(x)\} > 0.$$ 

Thus

$$\lim_{\epsilon \searrow 0} \epsilon \ln \mathbb{P}_\epsilon \{X \notin x + B(\delta)\} < 0.$$
so
\[
\lim_{\varepsilon \downarrow 0} \hat{\mathbb{P}}_\varepsilon \{ X \in x + B(\delta) \} = 1.
\]
Thus
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \left[ \chi_G(X) \exp \left[ -\frac{1}{\varepsilon} \langle X - x, \Theta_x \rangle_{\mathbb{R}^d} \right] \right] \geq -\|\Theta_x\| \delta.
\]
Let \( \delta \downarrow 0 \) and combine things to get the claimed lower bound. \( \square \)

We finally have

**Corollary 0.14.** Assume that \( \Lambda \) is essentially smooth. Fix \( G \subset \mathbb{R}^d \) open. Then
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in G \} \geq -\inf \{ I(x) : x \in G \}.
\]

**Proof.** Fix \( x \in G \). We claim that
\[
(21) \quad \lim_{\varepsilon \downarrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in G \} \geq -I(x).
\]
This is clearly true if \( I(x) = \infty \). Assume next that \( x \in \text{dom}(I) \). Since \( \text{ri dom}(I) \neq \emptyset \), fix \( z \in \text{ri dom}(I) \).
Consider the point \( z_\lambda \overset{\text{def}}{=} (1 - \lambda)x + \lambda z \) for all \( \lambda \in [0, 1] \). We want to show that \( z_\lambda \in \text{ri dom}(I) \) for all \( \lambda \in (0, 1] \).
Indeed, fix \( y \in \text{dom}(I) \); then since \( z \in \text{ri dom}(I) \), there is an \( \varepsilon > 0 \) such that \( z + \varepsilon(z - y) \in \text{dom}(I) \). Define
\[
\hat{\varepsilon} = \frac{\varepsilon \lambda}{1 + \varepsilon(1 - \lambda)} \quad \text{and} \quad \hat{\lambda} = \frac{\lambda}{1 + \varepsilon(1 - \lambda)}
\]
We calculate that
\[
\begin{align*}
(1 - \hat{\lambda})x + \hat{\lambda}(z + \varepsilon(z - y)) &= (1 - \lambda)\hat{x} + \hat{\lambda}\hat{y} = (1 + \hat{\varepsilon})(1 - \lambda)x + (1 + \hat{\varepsilon})\lambda z - \hat{\varepsilon}y \\
&= (1 + \varepsilon)(1 - \lambda)x + (1 + \varepsilon)\lambda z - \varepsilon y \\
&= (1 + \varepsilon)(1 - \lambda)x + (1 + \varepsilon)\lambda z - \varepsilon y/1 + \varepsilon(1 - \lambda)
\end{align*}
\]
The import of these calculations is that since \( \text{dom}(I) \) is convex and \( x \) and \( z + \varepsilon(z - y) \), we thus know that \( z_\lambda + \hat{\varepsilon}(z_\lambda - y) \in \text{dom}(I) \). Since \( y \) was an arbitrary element of \( \text{dom}(I) \), we thus have that indeed \( z_\lambda \in \text{ri dom}(I) \) (note: it is much easier to deduce the correct formulae for \( \hat{\varepsilon} \) and \( \hat{\lambda} \) by taking \( z = 0 \); i.e., by considering \( \text{dom}(I) - x \)).

Thus \( z_\lambda \in \text{ri dom}(I) \) for all \( \lambda \in (0, 1] \). Since \( G \) is open, \( z_\lambda \in G \) for \( \lambda \in (0, 1] \) sufficiently small. Thus
\[
\lim_{\varepsilon \downarrow 0} \varepsilon \ln \mathbb{P}_\varepsilon \{ X \in G \} \geq -\inf \{ I(x) : x \in G \cap \text{ri dom}(I) \} \geq -I(z_\lambda)
\]
By convexity, \( I(z_\lambda) \leq (1 - \lambda)I(x) + \lambda I(z) \) so \( \lim_{\varepsilon \downarrow 0} I(z_\lambda) \leq I(x) \). Thus we again have (21). \( \square \)

**Theorem 0.15 (Gärtner-Ellis).** Assume that (16) holds and that \( \Lambda \) is essentially smooth. Then \( X^\varepsilon \) has large deviations principle with rate function (14)

**Exercises**

1. If \( f \) is convex, show that \( \text{dom}(f) \) is convex.
2. We here investigate the natural properties of the \( \Lambda_\varepsilon \)'s of (13).
   (a) Show that the \( \Lambda_\varepsilon \)'s are convex.
   (b) Show that the limit of convex functions is convex.
   (c) Show that the Legendre-Fenchel transform of a function is convex.
3. Show that if \( \text{dom}(I) \neq \emptyset \), then \( \text{ri dom}(I) \) for (1)
(4) Let $f : \mathbb{R}^2 \to (0, \infty)$ be
\[
    f(x, y) = \begin{cases} 
        \frac{1}{y} & \text{if } y \geq 1 \text{ and } x = 0 \\
        \infty & \text{else}
    \end{cases}
\]
Fix $\text{ri}(\text{dom}(f))$.

(5) Let $f : \mathbb{R}^d \to [-\infty, \infty]$ be a convex function and fix $a > 0$ and $b \in \mathbb{R}$, $c \in \mathbb{R}^d$, and $d$ in $\mathbb{R}$. Define
\[
    \tilde{f}(x) = af(bx + c) + d \text{ for all } x \in \mathbb{R}^d.
\]
Show that $\tilde{f}$ is convex.

(6) If $\mathcal{F}$ is a collection of convex functions from a vector space $V$ to $[-\infty, \infty]$, show that $f(x) \overset{\text{def}}{=} \inf_{g \in \mathcal{F}} g(x)$ is convex.

(7) Fix $f : V \to [-\infty, \infty]$ where $V$ is a topological For each $x \in V$, define $\tilde{f}(x) = \lim_{x' \to x} f(x')$.
(a) Show that $\tilde{f}$ is lower semicontinuous.
(b) Assume that $V$ is a topological vector space and $f$ is convex. Show that $\tilde{f}$ is also convex.

(8) Show that if a lower semicontinuous function $f$ has compact level sets, then $f$ attains its inf over a closed set.

(9) One of the basic results in large deviations is Cramer’s Theorem. Let $\{\xi_n\}_{n \in \mathbb{N}}$ be an i.i.d. collection of $\mathbb{R}^d$-valued random variables with common law $\mu$. Define
\[
    S_N \overset{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \xi_n
\]
(a) For $\theta \in \mathbb{R}^d$, compute $\Lambda(\theta) = \lim_{N \to -\infty} \frac{1}{N} \ln \mathbb{E} [\exp \{N \langle \theta, X_N \rangle \}]$ in terms of $\mu$.
(b) Show that $\Lambda$ is differentiable on $\text{int} \text{ dom}(\Lambda)$.
(c) Fix $x \in \mathbb{R}^d$ and suppose that $\Theta_x \in \mathbb{R}^d$ satisfies (18). Find the statistics of $\{\xi_n\}_{1 \leq n \leq N}$ under
\[
    \tilde{P}_N(A) = \frac{\mathbb{E}_A \exp [N \langle S_N, \Theta_x \rangle]}{\mathbb{E} \exp [N \langle S_N, \Theta_x \rangle]} \quad A \in \mathcal{F}
\]
In this case, the proof of the lower bound can be simplified.

(10) Suppose that $\{X_{1 \varepsilon}^1\}_{\varepsilon > 0}$ and $\{X_{2 \varepsilon}^2\}_{\varepsilon > 0}$ are, respectively, collections of $\mathbb{R}^{d_1}$-valued and $\mathbb{R}^{d_2}$-valued random variables. Assume that $X_{1 \varepsilon}^1$ and $X_{2 \varepsilon}^2$ are independent for each $\varepsilon > 0$. Assume that
\[
    \Lambda_1(\theta) = \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{E} \left[ \exp \left\{ \varepsilon^{-1} \langle \theta, X_{1 \varepsilon}^1 \rangle \right\} \right]
\]
\[
    \Lambda_2(\theta) = \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{E} \left[ \exp \left\{ \varepsilon^{-1} \langle \theta, X_{2 \varepsilon}^2 \rangle \right\} \right]
\]
both exist, satisfy (16) and are essentially smooth. Define $Y_{\varepsilon} \overset{\text{def}}{=} (X_{\varepsilon}^1, X_{\varepsilon}^2)$ (i.e., $Y_{\varepsilon}$ is $\mathbb{R}^{d_1 + d_2}$-valued).
(a) Compute
\[
    \Lambda(\theta) = \lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{E} \left[ \exp \left\{ \varepsilon^{-1} \langle \theta, Y_{\varepsilon} \rangle \right\} \right]
\]
for all $\theta = (\theta_1, \theta_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$
(b) Show that (16) holds for $\Lambda$
(c) Show that $\Lambda$ is essentially smooth.
(d) Compute the Legendre-Fenchel transform of $\Lambda$ in terms of the Legendre-Fenchel transforms of $\Lambda_1$ and $\Lambda_2$.

(11) Fix $\mu \in \mathcal{P}[0, \infty)$. Define
\[
    \Lambda(\theta) \overset{\text{def}}{=} \ln \int_{z \in [0, \infty)} e^{z \theta} \mu(dz) \quad \theta \in \mathbb{R}
\]
Let $I$ denote the Fenchel transform of $\Lambda$.
(a) Compute $I(x)$ for $x < 0$.
(b) Compute $\lim_{\theta \to \infty} \Lambda(\theta)$. (Hint: note that $e^{-z} = \int_{z \in [0, \infty)} \chi_{(r \geq z)} e^{-r} dr$).
(c) Compute $I(0)$. 

22
CHAPTER 3

Schilder’s Theorem

We here develop large deviations for path-valued random variables. Namely, let \( W \) be a Brownian motion. We want to consider the \( C[0,1] \)-valued random variable \( \varepsilon W \) as \( \varepsilon > 0 \).

Let’s first approximate. For each \( n \in \mathbb{N} \), define

\[
\hat{W}_t^n \stackrel{\text{def}}{=} W_{\lfloor tn \rfloor/n} \{ [tn] + 1 - tn \} + W_{\lfloor (tn+1) \rfloor/n} \{ tn - [tn] \}.
\]

Thus \( \hat{W}^n \) agrees with \( W \) at each \( k/n \), and is piecewise linear on each \([k/n,(k+1)/n]\). We can thus uniquely characterize \( X^{n,\varepsilon} \) as follows.

- \( \hat{W}_0^n = 0 \)
- It is piecewise linear on each \([k/n,(k+1)/n]\).
- \( \{ n^{1/2} (\hat{W}_{(k+1)/n}^{n} - \hat{W}_{k/n}^{n} ) \mid 0 \leq k \leq n-1 \} \) is an independent collection of standard normal random variables.

We will see that \( (\varepsilon \hat{W}^n)_{\varepsilon \in (0,1)} \) has a large deviations principle for each \( n \in \mathbb{N} \). Since \( \hat{W}^n \) is in fact totally determined by its values at the \( k/n \)’s, it is in fact finite-dimensional; wish to exploit this fact.

Define \( T_n : \mathbb{R}^n \rightarrow C[0,1] \) as

\[
T_n(\bar{x}) (t) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \left\{ \sum_{1 \leq j \leq \lfloor tn \rfloor} x_j \right\} + (tn - \lfloor tn \rfloor) \bar{x}_{\lfloor tn \rfloor + 1} \quad t \in [0,1]
\]

for \( \bar{x} = (x_1, x_2 \ldots x_n) \in \mathbb{R}^n \). Note that

\begin{equation}
T_n(\bar{x}) \left( \frac{k+1}{n} \right) - T_n(\bar{x}) \left( \frac{k}{n} \right) = \frac{1}{\sqrt{n}} x_k
\end{equation}

for all \( k \in \{0, 1 \ldots n-1\} \). Fix next an independent collection \( \{ \eta_i \}_{1 \leq i \leq n} \) of standard normal random variables and define \( \hat{W}_t^n \stackrel{\text{def}}{=} (T_n(\bar{\eta})) (t) \) for all \( t \in [0,1] \), where \( \bar{\eta} \stackrel{\text{def}}{=} (\eta_1, \eta_2 \ldots \eta_n) \). Clearly \( \hat{W}_0^n = 0 \), \( \hat{W}^n \) is piecewise linear on each \([k/n,(k+1)/n]\) and from (22) we have that \( \{ n^{1/2} (\hat{W}_{(k+1)/n}^{n} - \hat{W}_{k/n}^{n} ) \}_{k=0}^{n-1} \) is an independent collection of standard normals.

**Lemma 0.16.** We have that \( (\varepsilon \hat{W}^n)_{\varepsilon \in (0,1)} \) has large deviations principle in \( C[0,1] \) with action functional

\[
I_n(\varphi) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{2} \sum_{k=0}^{n-1} \left| \frac{\varphi \left( \frac{k+1}{n} \right) - \varphi \left( \frac{k}{n} \right) }{\sqrt{n}} \right|^2 & \text{if } \varphi \text{ is piecewise linear on each } [k,(k+1)/n] \text{ and } \varphi(0) = 0 \\ \infty & \text{else} \end{cases}
\]

**Proof.** By the above comments, the law of \( \hat{W}^n \) and the law of \( \hat{W}^n \) agree. Note that \( T_n(\mathbb{R}^n) \) is the collection of functions which vanish at 0 and which are piecewise-linear on each \([k,(k+1)/n]\). We further note that if \( T_n(\psi_1, \psi_2 \ldots \psi_n) = \psi \), then

\[
\frac{1}{2} \sum_{k=1}^n \psi_k^2 = \frac{1}{2} \sum_{k=1}^n \left| \frac{\psi \left( \frac{k+1}{n} \right) - \psi \left( \frac{k}{n} \right) }{\sqrt{n}} \right|^2.
\]

\[ \square \]
Letting $n$ become large, let’s define

\[
I(\varphi) \overset{\text{def}}{=} \left\{ \begin{array}{ll}
\frac{1}{2} \int_{s=0}^{1} (\dot{\varphi}(s))^2 \, ds & \text{if } \varphi \text{ is absolutely continuous and } \varphi(0) = 0 \\
\infty & \text{else;}
\end{array} \right.
\]

see [FW98]. Define $C_0[0,1] = \{ \varphi \in C[0,1] : \varphi(0) = 0 \}$, and for $\varphi \in C_0[0,1]$, set $\|\varphi\|_C \overset{\text{def}}{=} \sup_{0 \leq t \leq 1} |\varphi(t)|$.

Our goal is to show that $\varepsilon W$ has a large deviations principle in $C_0[0,1]$ with rate functional given by (23); this is Theorem 0.22 below.

**Lemma 0.17.** For each $s \geq 0$, the level sets $\{ \varphi \in C_0[0,1] : I(\varphi) \leq s \}$ of $I$ are compact in $C_0[0,1]$.

**Proof.** If $I(\varphi) \leq s$, then for any $0 \leq t_1 < t_2 \leq 1$,

\[
|\varphi(t_2) - \varphi(t_1)| = \left| \int_{s=t_1}^{t_2} \dot{\varphi}(s) \, ds \right| = t_2 - t_1 \left| \frac{1}{t_2 - t_1} \int_{s=t_1}^{t_2} \dot{\varphi}(s) \, ds \right|
\]

\[
\leq t_2 - t_1 \sqrt{ \frac{1}{t_2 - t_1} \int_{s=t_1}^{t_2} |\dot{\varphi}(s)| \, ds } \leq \sqrt{t_2 - t_1} \sqrt{2s}.
\]

Thus $\overline{\Phi}(s)$ is compact. We thus need to show that $\Phi(s)$ is closed. This follows from the fact that

\[
I(\varphi) = \sup \left\{ \frac{1}{2} \sum_{j=0}^{n} \frac{|\varphi(t_{j+1}) - \varphi(t_j)|^2}{t_{j+1} - t_j} \mid 0 = t_0 < t_1 \cdots < t_n = 1 \right\}.
\]

\[\square\]

Let’s next prove the lower bound. Again we use a measure change.

**Lemma 0.18.** For any $\varphi \in C_0[0,1]$ and $\delta > 0$,

\[
\lim_{\varepsilon \to 0} \varepsilon^2 \ln \mathbb{P}\left\{ \|X^\varepsilon - \varphi\|_C < \delta \right\} \geq -I(\varphi)
\]

**Proof.** The result is obvious if $I(\varphi) = \infty$. Assume that $I(\varphi) < \infty$; thus $\dot{\varphi}$ is well-defined, and we set

\[\dot{W}^\varepsilon = W_t - \frac{1}{\varepsilon} \varphi(t) = W_t - \frac{1}{\varepsilon} \int_{s=0}^{t} \dot{\varphi}(s) \, ds;\]

then $X^\varepsilon = \varepsilon \dot{W}^\varepsilon$. Define

\[\tilde{\mathbb{P}}^\varepsilon(A) \overset{\text{def}}{=} \mathbb{E} \left[ \chi_A \exp \left[ \frac{1}{\varepsilon} \int_{s=0}^{1} \dot{\varphi}(s) \, ds - \frac{1}{2\varepsilon^2} \int_{s=0}^{1} (\dot{\varphi}(s))^2 \, ds \right] \right] \quad A \in \mathcal{F} \]

By Girsanov’s theorem, $\dot{W}^\varepsilon$ is a Brownian motion under $\tilde{\mathbb{P}}^\varepsilon$. We thus have that

\[\mathbb{P}\left\{ \|X^\varepsilon - \varphi\|_C < \delta \right\} = \tilde{\mathbb{E}}^\varepsilon \left[ \chi_{\|X^\varepsilon\|_C < \delta} \exp \left[ -\frac{1}{\varepsilon} \int_{s=0}^{1} \dot{\varphi}(s) \, dW_s - \frac{1}{\varepsilon^2} I(\varphi) \right] \right] \geq \tilde{\mathbb{E}}^\varepsilon \left[ \chi_{\|X^\varepsilon\|_C < \delta} \exp \left[ -\frac{1}{\varepsilon} \int_{s=0}^{1} \dot{\varphi}(s) \, dW_s \right] \right] \exp \left[ -\frac{1}{\varepsilon^2} I(\varphi) \right] \]

Girsanov’s theorem tells us that

\[
\lim_{\varepsilon \to 0} \tilde{\mathbb{E}}^\varepsilon \{ \|X^\varepsilon\|_C < \delta \} = 1,
\]

so for $\varepsilon > 0$ sufficiently small, $\tilde{\mathbb{E}}^\varepsilon \{ \|X^\varepsilon\|_C < \delta \}$ is positive. Thus we can use Jensen’s inequality to calculate that

\[\tilde{\mathbb{E}}^\varepsilon \left[ \chi_{\|X^\varepsilon\|_C < \delta} \exp \left[ -\frac{1}{\varepsilon} \int_{s=0}^{1} \dot{\varphi}(s) \, dW_s \right] \right] = \tilde{\mathbb{E}}^\varepsilon \left[ \chi_{\|X^\varepsilon\|_C < \delta} \exp \left[ -\frac{1}{\varepsilon} \int_{s=0}^{1} \dot{\varphi}(s) \, dW_s \right] \right] \frac{\mathbb{P}^\varepsilon \{ \|X^\varepsilon\|_C < \delta \}}{\mathbb{P}^\varepsilon \{ \|X^\varepsilon\|_C < \delta \}} \]

\[24\]
We next have that
\[ \tau \]
and
\[ \frac{\mathbb{E}^\varepsilon \chi_{\{\|X^\varepsilon\|_C < \delta\}} \left( \int_{s=0}^{1} \phi(s) dW_s \right)}{\mathbb{P}^\varepsilon \{\|X^\varepsilon\|_C < \delta\}} \leq \sqrt{\frac{\mathbb{E}^\varepsilon \chi_{\{\|X^\varepsilon\|_C < \delta\}} \left( \int_{s=0}^{1} \phi(s) dW_s \right)}{\mathbb{P}^\varepsilon \{\|X^\varepsilon\|_C < \delta\}}} \mathbb{E}^\varepsilon \left[ \left( \int_{s=0}^{1} \phi(s) dW_s \right)^2 \right] = \frac{1}{\sqrt{\mathbb{P}^\varepsilon \{\|X^\varepsilon\|_C < \delta\}}} \sqrt{2I(\varphi)}.

Combining things together, we get that
\[ \mathbb{P} \{\|X^\varepsilon - \varphi\|_C < \delta\} \geq \mathbb{P}^\varepsilon \{\|X^\varepsilon\|_C < \delta\} \exp \left[ \frac{1}{\varepsilon} \sqrt{\frac{2I(\varphi)}{\mathbb{P}^\varepsilon \{\|X^\varepsilon\|_C < \delta\}}} - \frac{1}{\varepsilon^2} I(\varphi) \right] \]
This gives us the stated claim. □

**Lemma 0.19.** We have that
\[ \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} |W_t| \geq L \right\} \leq \sqrt{\frac{8}{\pi}} \exp \left[ -\frac{L^2}{2} \right] \]
for all \( L > 0 \).

**Proof.** We can prove this either by the reflection principle or the Feynman-Kac formula. We shall use the latter. Define
\[ u(t, x) = 2 \int_{z=L}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp \left[ -\frac{(z-x)^2}{2t} \right] dz \]
for all \( t > 0 \) and \( x \leq L \). Clearly \( u \in C^\infty((0, \infty) \times (-\infty, L]) \) and
\[ \frac{\partial u}{\partial t} (t, x) = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} (t, x) \]
We also have that \( u(t, L) = 1 \) and \( \lim_{t \downarrow 0} u(t, \cdot) = 0 \) uniformly on compact subsets of \((-\infty, L)\).

The PDE for \( u \) and its regularity imply that for each \( \delta > 0 \), \( Z_0^\delta \overset{\text{def}}{=} u(1 + \delta - t, W_t) \) is a martingale. We have that
\[ Z_0^\delta = u(1 + \delta, 0) = 2 \int_{r=L/\sqrt{1+\delta}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{z^2}{2} \right] dz. \]
Define next
\[ \tau \overset{\text{def}}{=} \inf \{ t \geq 0 : W_t \geq L \}. \]
We next have that
\[ \{ \tau \leq 1 \} = \left\{ \sup_{0 \leq t \leq 1} W_t \geq L \right\} \]
If \( \tau \leq 1 \), then \( \tau = 1 = \tau \) and \( W_{\tau \wedge 1} = L \), so \( Z_0^\delta = u(1 + \delta - \tau, L) = 0 \). On the other hand, if \( \tau > 1 \), then \( \tau \wedge 1 = 1 \) and \( W_{\tau \wedge 1} < L \), in which case \( Z_{\tau \wedge 1} = u(\delta, W_1) \). Combining things together, we get that
\[ u(1 + \delta, 0) = Z_0^\delta \overset{\mathbb{E}}{=} E \left[ Z_{\tau \wedge 1} \right] = \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} W_t \geq L \right\} + E \left[ u(\delta, W_1) \chi_{\{\tau > 1\}} \right]. \]
Let \( \delta \downarrow 0 \) and use standard convergence results to see that
\[ \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} W_t \geq L \right\} = 2 \int_{r=L}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{z^2}{2} \right] dz. \]
Thus
\[ \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} |W_t| \geq L \right\} \leq \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} W_t \geq L \right\} + \mathbb{P} \left\{ \inf_{0 \leq t \leq 1} W_t \leq -L \right\} = 4 \int_{r=L}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{z^2}{2} \right] dz. \]
A standard bound on the Gaussian error function completes the proof.

From this we can show that $\tilde{X}^{n,\varepsilon}$ is a good approximation of $X^\varepsilon$.

**Corollary 0.20.** For all $n \in \mathbb{N}$, $\varepsilon > 0$, and $\delta > 0$,

$$\mathbb{P}\left\{\|\tilde{X}^{n,\varepsilon} - X^\varepsilon\|_C \geq \delta\right\} \leq \sqrt{\frac{2}{\pi}} n \exp\left[-\frac{\delta^2 n^2}{8\varepsilon^2}\right]$$

**Proof.** For every $t \in [0, 1]$,

$$|\tilde{X}_t^{n,\varepsilon} - X^\varepsilon|^2 \leq \varepsilon \left|\tilde{W}_t^n - W_{\lfloor tn \rfloor/n}\right|^2 + \varepsilon \left|W_{\lfloor tn \rfloor/n} - W_t\right| \leq 2\varepsilon \left|W_t - W_{\lfloor tn \rfloor/n}\right|.$$

Thus

$$\mathbb{P}\left\{\|\tilde{X}^{n,\varepsilon} - X^\varepsilon\|_C \geq \delta\right\} \leq \mathbb{P}\left\{\sup_{0 \leq k \leq n-1} \sup_{\frac{k}{n} \leq t \leq \frac{k+1}{n}} |W_t - W_{\frac{k}{n}}| \geq \frac{\delta}{2\varepsilon}\right\} \leq \sum_{k=0}^{n-1} \mathbb{P}\left\{\sup_{\frac{k}{n} \leq t \leq \frac{k+1}{n}} |W_t - W_{\frac{k}{n}}| \geq \frac{\delta}{2\varepsilon}\right\} \leq n \mathbb{P}\left\{\sup_{0 \leq t \leq 1} |W_t| \geq \frac{\delta \sqrt{n}}{2\varepsilon}\right\} \leq n \sqrt{\frac{8}{\pi}} \exp\left[-\frac{\delta^2 n}{32\varepsilon^2}\right].$$

This completes the proof.

We can now prove the upper bound.

**Lemma 0.21.** For $s \geq 0$ and $\delta > 0$,

$$\lim_{\varepsilon \to 0^+} \varepsilon \ln \mathbb{P}\{\text{dist}(X^\varepsilon, \Phi(s)) \geq \delta\} \leq -s.$$

**Proof.** We have that

$$\mathbb{P}\{\text{dist}(X^\varepsilon, \Phi(s)) \geq \delta\} \leq \mathbb{P}\left\{\|X^\varepsilon - \tilde{X}^{n,\varepsilon}\| \geq \frac{\delta}{2}\right\} + \mathbb{P}\left\{\text{dist}(\tilde{X}^{n,\varepsilon}, \Phi(s)) \geq \frac{\delta}{2}\right\} \leq n \sqrt{\frac{16}{\pi}} \exp\left[-\frac{\delta^2 n}{32\varepsilon^2}\right] + \mathbb{P}\left\{I\left(\tilde{X}^{n,\varepsilon}\right) \geq s\right\}. $$

We note that

$$I\left(\tilde{X}^{n,\varepsilon}\right) = \frac{n\varepsilon^2}{2} \sum_{k=1}^{n-1} |W_{(k+1)/n} - W_{k/n}|^2.$$  

We thus have that for $\alpha > 0$

$$\mathbb{P}\left\{I\left(\tilde{X}^{n,\varepsilon}\right) \geq s\right\} \leq \mathbb{P}\left\{\frac{1-\alpha}{\varepsilon^2} I\left(\tilde{X}^{n,\varepsilon}\right) \geq \frac{s(1-\alpha)}{\varepsilon^2}\right\} \leq \exp\left[-\frac{s(1-\alpha)}{\varepsilon^2}\right] \mathbb{E}\left[\exp\left[\frac{(1-\alpha)}{\varepsilon^2} I\left(\tilde{X}^{n,\varepsilon}\right)\right]\right]$$

$$= \exp\left[-\frac{s(1-\alpha)}{\varepsilon^2}\right] \mathbb{E}\left[\prod_{k=0}^{n-1} \exp\left[\frac{n(1-\alpha)}{2} (W_{(k+1)/n} - W_{k/n})^2\right]\right]$$

$$= \exp\left[-\frac{s(1-\alpha)}{\varepsilon^2}\right] \left(\sqrt{\frac{n}{2\pi}} \int_{z \in \mathbb{R}} \exp\left[-\frac{\alpha z^2}{2}\right] dz\right)^n = \exp\left[-\frac{s(1-\alpha)}{\varepsilon^2}\right] \alpha^{-n/2}.$$

Collect things together.

**Theorem 0.22.** We have that $\{X_\varepsilon\}_{\varepsilon > 0}$ has an LDP in $C_0[0,1]$ with rate functional $I$ as in (23), which is equivalently written as

$$I(\varphi) = \left\{\begin{array}{ll} \frac{1}{2} \int_0^1 (\dot{\varphi}(s))^2 ds & \text{if } \varphi \in L^2 \\ \infty & \text{else.} \end{array}\right.$$

**Proof.** Collect the above calculations together.
Exercises

(1) If $W$ is a Brownian motion, show that for each $\delta > 0$ \( \lim_{\varepsilon \to 0} P\{\sup_{0 \leq t \leq T} |\varepsilon W_t| \geq \delta \} = 0. \)

(2) We here study (24).

(a) Suppose that $0 = t_0 < t_1 \cdots < t_n = 1$. Show that
\[
\sum_{j=0}^{n} \frac{t_{j+1} - t_j}{|\varphi(t_{j+1}) - \varphi(t_j)|^2} \leq \int_{t=0}^{1} (\dot{\varphi}(t))^2 dt.
\]

(b) Fix an open subset $O$ of $[0, 1]$. Show that there is a \( \{\varphi_n\} \subset C([0, 1]; [0, 1]) \) such that \( \varphi_n \not\to \chi_O \).

(c) Since Lebesgue measure on $([0, 1], \mathcal{B}[0, 1])$ is regular, show that for any $A \in \mathcal{B}[0, 1]$ and any $\delta > 0$, there is a $\varphi \in C([0, 1]; [0, 1])$ such that
\[
\int_{t=0}^{1} |\chi_A(t) - \varphi(t)|^2 dt < \delta.
\]

(d) Assume that $\psi \in L^2[0, 1]$. Show that for $\delta > 0$, there is a $\psi^* \in C[0, 1]$ such that
\[
\int_{t=0}^{1} |\psi(t) - \psi^*(t)|^2 dt < \delta.
\]

(e) Show that if $\varphi \in C^1[0, 1]$, then
\[
I(\varphi) \leq \sup \left\{ \frac{1}{2} \sum_{j=0}^{n} \frac{|\varphi(t_{j+1}) - \varphi(t_j)|^2}{t_{j+1} - t_j} \mid 0 = t_0 \leq t_1 \cdots \leq t_n = 1 \right\}.
\]

(f) Show that if $I(\varphi) < \infty$, then there is a $\psi^* \in C^1[0, 1]$ such that
\[
\int_{t=0}^{1} |\dot{\varphi}(t) - \dot{\psi}^*(t)|^2 dt < \delta.
\]

(g) Show that if $I(\varphi) < \infty$, then
\[
I(\varphi) \leq \sup \left\{ \frac{1}{2} \sum_{j=0}^{n} \frac{|\varphi(t_{j+1}) - \varphi(t_j)|^2}{t_{j+1} - t_j} \mid 0 = t_0 \leq t_1 \cdots \leq t_n = 1 \right\}.
\]

(3) Show that since (24) holds, the level sets of $I$ are closed in $C[0, 1]$.

(4) Use the contraction principle (on Brownian scaling) and the large deviations principle in $C[0, 1]$ to get the large deviations principle for $\varepsilon W$ as an element of $C[0, T]$ for any $T > 0$. 

27
CHAPTER 4

Freidlin-Wentzell-Theory

Let’s next extend Schilder’s theorem to a nonlinear setting. Fix bounded and Lipshitz-continuous maps $b$ and $\sigma$ from $\mathbb{R}$ to $\mathbb{R}$. Define

$$
\bar{b} \stackrel{\text{def}}{=} \sup_{x \in \mathbb{R}} |b(x)| \quad \sigma \stackrel{\text{def}}{=} \sup_{x \in \mathbb{R}} |\sigma(x)| \quad \underline{\sigma} \stackrel{\text{def}}{=} \inf_{x \in \mathbb{R}} |\sigma(x)|
$$

$$
L_b \stackrel{\text{def}}{=} \sup_{x,y \in \mathbb{R} \atop x \neq y} \frac{|b(x) - b(y)|}{|x - y|} \quad L_\sigma \stackrel{\text{def}}{=} \sup_{x,y \in \mathbb{R} \atop x \neq y} \frac{|\sigma(x) - \sigma(y)|}{|x - y|}.
$$

Then $\bar{b}$, $\sigma$, $L_b$ and $L_\sigma$ are finite. We also assume that $\underline{\sigma} > 0$. Let $W$ be a Brownian motion and fix an initial point $x_0 \in \mathbb{R}$. For each $\varepsilon \in [0,1)$, define

$$
\begin{align*}
X^\varepsilon_t &= b(X^\varepsilon_t)dt + \varepsilon \sigma(X^\varepsilon_t)dW_t \quad t > 0 \\
X^\varepsilon_0 &= x_0
\end{align*}
$$

Note that $X^0$ is deterministic.

**Lemma 0.23.** For each $T > 0$, we have that $\lim_{\varepsilon \to 0} \sup_{0 \leq t \leq T} \mathbb{E}\left[\left|X^\varepsilon_t - X^0_t\right|^{2}\right] = 0$.

The proof is one of the problems. We want to find a large deviations principle for $X^\varepsilon$. To make things fit into the large deviations framework, let’s fix $T > 0$. We will treat $\{X^\varepsilon_t \mid 0 \leq t \leq T\}$ as an element of $C[0,T]$. Note that we already concluded a related result in Exercise 27 of Chapter 1; that will be central in our current calculations. Motivated by that exercise, let’s define

$$
I(\varphi) \stackrel{\text{def}}{=} \begin{cases} 
\frac{1}{2} \int_{s=0}^{1} \left( \frac{\varphi(s) - b(\varphi(s))}{\sigma(\varphi(s))} \right)^2 ds & \text{if } \varphi \text{ is absolutely continuous and } \varphi(0) = x_0 \\
\infty & \text{else};
\end{cases}
$$

(26)

Let’s first prove the lower bound; the arguments are very similar to the proof of the lower bound in Schilder’s theorem.

**Lemma 0.24.** For $G \subset C[0,T]$ open,

$$
\lim_{\varepsilon \to 0} \varepsilon^2 \ln \mathbb{P}\{X^\varepsilon \in G\} \geq - \inf_{\varphi \in G} I(\varphi)
$$

**Proof.** It is sufficient to fix $\varphi \in C[0,T]$ such that $I(\varphi) < \infty$ and show that

$$
\lim_{\delta \to 0} \lim_{\varepsilon \to 0} \varepsilon^2 \ln \mathbb{P}\{\|X^\varepsilon - \varphi\|_{C[0,T]} < \delta\} \geq -I(\varphi).
$$

Since $I(\varphi) < \infty$, $\varphi$ is absolutely continuous and $\varphi(0) = x_0$. Define

$$
\dot{X}^\varepsilon_t \stackrel{\text{def}}{=} X^\varepsilon_t - \varphi_t \quad t \in [0,T]
$$

$$
\dot{\xi}^\varepsilon_t \stackrel{\text{def}}{=} \frac{\dot{\varphi}_t - b(\psi_s + \dot{X}^\varepsilon_t)}{\sigma(\psi_s + \dot{X}^\varepsilon_t)} \quad t \in [0,T]
$$

$$
\dot{\tilde{W}}^\varepsilon_t \stackrel{\text{def}}{=} W_t - \frac{1}{\varepsilon} \int_{s=0}^{t} \dot{\xi}^\varepsilon_s ds \quad t \in [0,T]
$$

$$
\tilde{\mu}^\varepsilon(A) \stackrel{\text{def}}{=} \mathbb{E} \left[ \chi_A \exp \left( \frac{1}{\varepsilon} \int_{s=0}^{T} \dot{\xi}^\varepsilon_s dW_s - \frac{1}{2\varepsilon^2} \int_{s=0}^{T} (\dot{\xi}^\varepsilon_s)^2 ds \right) \right] \quad A \in \mathcal{F}
$$
Then $\tilde{W}^\varepsilon$ is a $\tilde{P}^\varepsilon$-Brownian motion and

$$d\hat{X}^\varepsilon_t = \varepsilon \sigma (\hat{X}^\varepsilon_t + \varphi_t) d\tilde{W}^\varepsilon_t \quad t \in [0, T]$$

$\hat{X}^\varepsilon_0 = 0$.

We then have that

$$\mathbb{P} \left\{ \|X^\varepsilon - \varphi\|_{C[0,T]} < \delta \right\} = \tilde{E}^\varepsilon \left\{ \chi_{\{\|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \}} \exp \left[ -\frac{1}{\varepsilon} \int_{s=0}^{T} \xi_s d\tilde{W}^\varepsilon_s - \frac{1}{2\varepsilon^2} \int_{s=0}^{T} (\xi_s^2)^2 ds \right] \right\}.$$

To proceed, we first note that for $x$ and $v$ in $\mathbb{R}$,

$$|(x + v)^2 - x^2| = |2xv + v^2| \leq 2|x||v| + v^2 \leq 2 \left( |x|\sqrt{|v|} \right) \sqrt{|v|} + v^2 \leq x^2|v| + |v| + v^2 = |v| (x^2 + 1 + |v|).$$

With this calculation in hand, we see that

We now compute that

$$\left| \left( \xi_s^2 - \left( \frac{\psi_t - b(\psi_t)}{\sigma(\psi_t)} \right) \right)^2 \right| \leq \frac{1}{\sigma^2(\psi_t + X_s^\varepsilon)} \left\{ \left( \frac{\psi_t - b(\psi_t)}{\sigma(\psi_t)} \right) \left( \frac{b(\psi_t) - b(\psi_t + X_s^\varepsilon)}{\sigma(\psi_t)} \right) \right\}^2 - \left( \frac{\psi_t - b(\psi_t)}{\sigma(\psi_t)} \right)^2$$

$$+ \frac{1}{\sigma^2(\psi_t + X_s^\varepsilon)} \left \{ \left( \frac{\psi_t - b(\psi_t)}{\sigma(\psi_t)} \right) - \left( \frac{b(\psi_t) - b(\psi_t + X_s^\varepsilon)}{\sigma(\psi_t)} \right) \right \} \sigma^2(\psi_t - \sigma(\psi_t + X_s^\varepsilon))$$

Combining these together, we get that if $\|\hat{X}^\varepsilon\|_{C[0,T]} \leq \delta$, then

$$\left| \frac{1}{2} \int_{s=0}^{T} (\xi_s^2)^2 ds - I(\varphi) \right| \leq \mathcal{E}(\delta)$$

where

$$\mathcal{E}(\delta) \stackrel{\text{def}}{=} \frac{L_\delta}{\sigma^2} \left\{ 2\sigma^2 I(\varphi) + 1 + L_\delta \right\} + \frac{\sigma^2}{\sigma^2} I(\varphi)L_\sigma \delta.$$

Note that $\lim_{\varepsilon \to 0} \tilde{P}_\varepsilon \{ \|\hat{X}^\varepsilon\|_{C[0,T]} \geq \delta \} = 0$; i.e. $\lim_{\varepsilon \to 0} \tilde{P}_\varepsilon \{ \|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \} = 1$. By Jensen’s inequality, as in the proof of Lemma 0.18. We have that

$$\tilde{P}_\varepsilon \left\{ \|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \right\} \exp \left[ -\frac{1}{\varepsilon} \int_{s=0}^{T} \xi_s^2 d\tilde{W}^\varepsilon_s \right]$$

$$\geq \tilde{P}_\varepsilon \left\{ \|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \right\} \exp \left[ -\frac{1}{\varepsilon} \tilde{P}_\varepsilon \left\{ \|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \right\} \right]$$

$$\geq \tilde{P}_\varepsilon \left\{ \|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \right\} \exp \left[ -\frac{1}{\varepsilon} \tilde{P}_\varepsilon \left\{ \|\hat{X}^\varepsilon\|_{C[0,T]} < \delta \right\} \right].$$

To finish, we need to bound $\tilde{P}_\varepsilon \left[ \int_{s=0}^{T} (\xi_s^2)^2 ds \right]$. We calculate that
Thus
\[
\int_{s=0}^{T} |\xi_t|^2 \, dt \leq 4\frac{\sigma^2}{\alpha^2} I(\varphi) + \frac{8T\tilde{b}^2}{\alpha^2}.
\]

Combine things together as in the proof of Lemma 0.18.

\[\square\]

To prove the compactness of the level sets and the upper bound, we will use the approximation of Exercise 27 of Chapter 1. For \(n \in \mathbb{N}\), define \(\tau_n(t) \overset{\text{def}}{=} \lfloor tn \rfloor / n\). For \(\varepsilon \in (0,1)\), let’s then define
\[
dx^n = b(X^n_t) \, dt + \varepsilon \sigma \left( X^n_{\tau_n(t)} \right) \, dW_t \quad t \in [0,T]
\]
and let’s define
\[
I_n(\varphi) = \begin{cases} \frac{1}{2} \int_{t=0}^{T} \left| \frac{d\varphi(t) - b(\varphi(t))}{\sigma(\varphi(t))} \right|^2 \, dt & \text{if } \varphi \text{ is absolutely continuous and } \varphi(0) = x_0 \\ \infty & \text{else} \end{cases}
\]

**Lemma 0.25.** We have that \(\{X_{n,\varepsilon}\}_{\varepsilon \in (0,1)}\) has large deviations principle in \(C[0,T]\) with action functional \(I_n\).

This is in the exercises.

We begin by showing that \(I_n\) is a good approximation of \(I\). Define
\[
\nu(s, \delta) \overset{\text{def}}{=} \sqrt{2\delta}^{1/2} + \tilde{b}\delta
\]
for all \(s > 0\) and \(\delta > 0\).

**Lemma 0.26.** Fix \(\varphi \in C[0,T]\) and \(n \in \mathbb{N}\). We have that
\[
|\varphi(t_2) - \varphi(t_1)| \leq \min \{\nu(I(\varphi), |t_2 - t_1|), \nu(I_n(\varphi), |t_2 - t_1|)\}
\]

\[
I(\varphi) \leq I_n(\varphi) \left( 1 + \frac{L_\sigma \nu(I_n(\varphi), 1/n)}{\alpha} \right)^2 \leq I_n(\varphi) \left( 1 + \frac{L_\sigma \nu(I(\varphi), 1/n)}{\alpha} \right)^2.
\]

**Proof.** We can assume that \(\varphi\) is absolutely continuous and \(\varphi(0) = x_0\); otherwise the claims are trivially true. Fix next \(\psi \in C[0,T]\), which we will take as either \(\varphi\) itself or \(\varphi \circ \tau_n\). If \(0 \leq t_1 \leq t_1 \leq T\), then
\[
|\varphi(t_2) - \varphi(t_1)| = \left| \int_{t_1}^{t_2} \dot{\varphi}(r) \, dr \right| = \left| \int_{t_1}^{t_2} \left\{ \left( \frac{\dot{\varphi}(r) - b(\varphi(r))}{\sigma(\varphi(r))} \right) \sigma(\varphi(r)) + b(\varphi(r)) \right\} \, dr \right| \leq \left| \int_{t_1}^{t_2} \left( \frac{\dot{\varphi}(r) - b(\varphi(r))}{\sigma(\varphi(r))} \right) \sigma(\varphi(r)) \, dr \right| + \left| \int_{t_1}^{t_2} b(\varphi(r)) \, dr \right| \leq \sqrt{\int_{t_1}^{t_2} \left( \frac{\dot{\varphi}(r) - b(\varphi(r))}{\sigma(\varphi(r))} \right)^2 \, dr} \sqrt{\int_{t_1}^{t_2} \sigma^2(\varphi(r)) \, dr} + \left| \int_{t_1}^{t_2} b(\varphi(r)) \, dr \right|.
\]

This proves the first two claims.

To see the last two claims, fix \(\psi_1\) and \(\psi_2\) in \(C[0,T]\). We will either take \(\psi_1 = \varphi\) and \(\psi_2 = \varphi \circ \tau_n\), or we will take \(\psi_1 = \varphi \circ \tau_n\) and \(\psi_2 = \varphi\). In either case, we have that
\[
|\psi_1(s) - \psi_2(s)| \leq |\varphi(s) - \varphi(\tau_n(s))| \leq \min \{\nu(I(\varphi), |t_2 - t_1|), \nu(I_n(\varphi), |t_2 - t_1|)\}.
\]

(28)
We proceed by calculating that

$$
\frac{1}{2} \int_{s=0}^{T} \left( \frac{\dot{\varphi}(s) - b(\varphi(s))}{\sigma(\psi_1(s))} \right)^2 ds = \frac{1}{2} \int_{s=0}^{T} \left( \frac{\dot{\varphi}(s) - b(\varphi(s))}{\sigma(\psi_2(s))} \right)^2 \left( \frac{\sigma(\psi_2(s))}{\sigma(\psi_1(s))} \right)^2 ds \\
= \frac{1}{2} \int_{s=0}^{T} \left( \frac{\dot{\varphi}(s) - b(\varphi(s))}{\sigma(\psi_2(s))} \right)^2 \left( 1 + \frac{\sigma(\psi_2(s)) - \sigma(\psi_1(s))}{\sigma(\psi_1(s))} \right)^2 ds \\
\leq \frac{1}{2} \int_{s=0}^{T} \left( \frac{\dot{\varphi}(s) - b(\varphi(s))}{\sigma(\psi_2(s))} \right)^2 \left( 1 + \frac{L_\sigma |\psi_1(s) - \psi_2(s)|}{\sigma} \right)^2 ds.
$$

We can then use (28) to finish the proof of the last two claims. □

We can now prove that the level sets of $I$ are compact.

**Lemma 0.27.** For each $s \geq 0$, the set

$$
\Phi(s) \overset{\text{def}}{=} \{ \varphi \in C[0, T] : I(\varphi) \leq s \}
$$

is a compact subset of $C[0, T]$.

**Proof.** The first part of Lemma 0.26 implies that $\Phi(s)$ is equicontinuous. To see that it is closed, fix a sequence $(\varphi_m)_{m \in \mathbb{N}}$ in $\Phi(s)$ and assume that $\varphi \overset{\text{def}}{=} \lim_{m \to \infty} \varphi_m$ is well-defined (in $C[0, T]$). For each $m \in \mathbb{N}$,

$$
I_n(\varphi_m) \leq s \left( 1 + \frac{L_\sigma v(s, 1/n)}{\sigma} \right)^2.
$$

Since the $I_n$’s are lower semicontinuous, we thus have that

$$
I_n(\varphi) \leq s \left( 1 + \frac{L_\sigma v(s, 1/n)}{\sigma} \right)^2.
$$

For $n$ large enough, we thus have that $I_n(\varphi) \leq 2s$, and for such $n$,

$$
I(\varphi) \leq I_n(\varphi) \left( 1 + \frac{L_\sigma v(2s, 1/n)}{\sigma} \right)^2 \leq s \left( 1 + \frac{L_\sigma v(2s, 1/n)}{\sigma} \right)^4.
$$

Let $n \to \infty$ and we see that $\Phi(s)$ is indeed closed. □

Let’s next show that $X^{n, \epsilon}$ is an exponentially good approximation of $X^\epsilon$. This is found in [Var84].

**Lemma 0.28.** For each $\delta > 0$, we have that

$$
\lim_{n \to \infty} \lim_{\epsilon \to 0} \epsilon^2 \ln \mathbb{P} \left\{ \|X^\epsilon - X^{n, \epsilon}\|_{C[0, T]} \geq \delta \right\} = -\infty.
$$

This will take some work to prove. We start by writing that

$$
X_t^\epsilon - X_t^{n, \epsilon} = \int_{s=0}^{t} \{ b(X_s^\epsilon) - b(X_s^{n, \epsilon}) \} ds + \epsilon \int_{s=0}^{t} \{ \sigma(X_s^\epsilon) - \sigma(X_s^{n, \epsilon}) \} dW_s \\
+ \epsilon \int_{s=0}^{t} \{ \sigma(X_s^{n, \epsilon}) - \sigma(X_s^{n, \epsilon}) \} dW_s
$$

**Lemma 0.29.** For $\nu > 0$, we have that

$$
\lim_{n \to \infty} \lim_{\epsilon \to 0} \epsilon^2 \ln \mathbb{P} \left\{ \sup_{0 \leq t \leq T} |X_t^{n, \epsilon} - X_t^{n, \epsilon} | \geq \nu \right\} = -\infty.
$$

**Proof.** If $j/n \leq t < (j+1)/n$ and $n \geq 2 \bar{b}/\nu$, we have that

$$
|X_t^{n, \epsilon} - X_t^{n, \epsilon} | = \epsilon \sigma |W_t - W_{j/n} + b_1/n \leq \epsilon \sigma \sup_{j/n \leq t < (j+1)/n} |W_t - W_{j/n}| + \nu/2.
$$

Thus
\[
\mathbb{P} \left\{ \sup_{0 \leq t \leq T} |X_{\tau(t)}^{n,\varepsilon} - X_{\tau(t)}^n| \geq \varkappa \right\} \leq \mathbb{P} \left\{ \max_{0 \leq j \leq \lfloor Tn \rfloor} \sup_{j/n \leq t < (j+1)/n} |W_{t} - W_{\tau(t)}| \geq \varkappa \right\} \\
\leq Tn \mathbb{P} \left\{ \sup_{0 \leq t \leq 1/n} |W_{t}| \geq \frac{\varkappa}{2\varepsilon\sigma} \right\} \leq Tn \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} |W_{t}| \geq \frac{\varkappa\sqrt{n}}{2\varepsilon\sigma} \right\} \\
\leq Tn \sqrt{\frac{8}{\pi}} \exp \left[-\frac{\varkappa^2 n}{8\varepsilon^2 \sigma^2}\right]
\]

where we have used Lemma 0.19. The claim follows. \(\square\)

We can now bound the error between \(X^\varepsilon\) and \(X^{n,\varepsilon}\).

**Proof of Lemma 0.28.** Let's now rewrite things a bit. For \(n \in \mathbb{N}\) and \(\varepsilon, \delta, \text{ and } \varkappa\) in \((0,1)\). Define
\[
\begin{align*}
\varrho_{\delta}^{1,n,\varepsilon} &\overset{\text{def}}{=} \inf \left\{ t \in [0,T] : |X_t^\varepsilon - X_t^{n,\varepsilon}| \geq \delta \right\} \quad \text{inf}() \overset{\text{def}}{=} T \\
\varrho_{\varkappa}^{2,n,\varepsilon} &\overset{\text{def}}{=} \inf \left\{ t \in [0,T] : |X_t^{n,\varepsilon} - X_t^{n,\varepsilon}_{\tau(t)}| \geq \varkappa \right\} \quad \text{inf}() \overset{\text{def}}{=} T \\
\varrho_{\delta,\varkappa}^{3,n,\varepsilon} &\overset{\text{def}}{=} \varrho_{\delta}^{1,n,\varepsilon} \wedge \varrho_{\varkappa}^{2,n,\varepsilon}
\end{align*}
\]

Lemma 0.29 is that
\[
\lim_{n \to \infty} \lim_{\varepsilon \to 0} \mathbb{E}^2 \ln \mathbb{P} \left\{ \varrho_{\varkappa}^{2,n,\varepsilon} < T \right\} = -\infty
\]

for \(\varkappa \in (0,1)\).

Fix for a moment \(n \in \mathbb{N}\) and \(\varepsilon\) and \(\varkappa\) in \((0,1)\). We write that
\[
\left\{ \sup_{0 \leq t \leq T} |X_t^{\varepsilon} - X_t^{n,\varepsilon}| \geq \delta \right\} = \left\{ \varrho_{\delta}^{1,n,\varepsilon} < T \right\} \subset \left\{ \varrho_{\delta}^{1,n,\varepsilon} < T, \varrho_{\varkappa}^{2,n,\varepsilon} = T \right\} \cup \left\{ \varrho_{\varkappa}^{2,n,\varepsilon} < T \right\}.
\]

Define now
\[
Z_t^{n,\varepsilon,\delta} \overset{\text{def}}{=} \left( |X_t^\varepsilon - X_t^{n,\varepsilon}|^2 + \varkappa^2 \right)^{1/2} \quad t \in [0,T]
\]
for all \(t \in [0,T]\). Then
\[
\left\{ \varrho_{\delta}^{1,n,\varepsilon} < T, \varrho_{\varkappa}^{2,n,\varepsilon} = T \right\} \subset \left\{ Z_t^{n,\varepsilon,\delta} \geq (\delta^2 + \varkappa^2)^{1/2} \right\}
\]

We next study the evolution of \(Z^{n,\varepsilon,\delta}\). Set
\[
\tilde{f}_{\varepsilon,\varkappa}(x) \overset{\text{def}}{=} (x^2 + \varkappa^2)^{1/2} \quad x \in \mathbb{R}
\]
Then
\[
\begin{align*}
\tilde{f}_{\varepsilon,\varkappa}(x) &= \frac{1}{\varepsilon^2} \left( x^2 + \varkappa^2 \right)^{1/2 - 1} (2x) \\
\tilde{f}_{\varepsilon,\varkappa}(x) &\overset{\text{def}}{=} \frac{1}{\varepsilon^2} \left( \frac{1}{\varepsilon^2} - 1 \right) \left( x^2 + \varkappa^2 \right)^{1/2 - 2} 4x^2 + \frac{2}{\varepsilon^2} \left( x^2 + \varkappa^2 \right)^{1/2 - 1}
\end{align*}
\]
for all \(x \in \mathbb{R}\). Note that
\[
\left| \tilde{f}_{\varepsilon,\varkappa}(x) \right| \leq \frac{2}{\varepsilon} \left( x^2 + \varkappa^2 \right)^{1/2 - 1/2} \quad \text{and} \quad \left| \tilde{f}_{\varepsilon,\varkappa}(x) \right| \leq \frac{6}{\varepsilon^2} \left( x^2 + \varkappa^2 \right)^{1/2 - 1}
\]
for all \(x \in \mathbb{R}\). Then
\[
d\tilde{Z}_t^{n,\varepsilon,\delta} = \tilde{f}_{\varepsilon,\varkappa}(X_t^\varepsilon - X_t^{n,\varepsilon}) \{ b(X_t^\varepsilon) - b(X_t^{n,\varepsilon}) \} dW_t + \varepsilon \tilde{f}_{\varepsilon,\varkappa}(X_t^\varepsilon - X_t^{n,\varepsilon}) \left\{ \sigma(X_t^\varepsilon) - \sigma(X_t^{n,\varepsilon}) \right\} dW_t \\
\quad + \frac{\varepsilon^2}{2} \tilde{f}_{\varepsilon,\varkappa}(X_t^\varepsilon - X_t^{n,\varepsilon}) \left\{ \sigma(X_t^\varepsilon) - \sigma(X_t^{n,\varepsilon}) \right\}^2 dt
\]
We have that
\[
\left| \tilde{f}_{\varepsilon,\varkappa}(X_t^\varepsilon - X_t^{n,\varepsilon}) \{ b(X_t^\varepsilon) - b(X_t^{n,\varepsilon}) \} \right| \leq \frac{2Lb}{\varepsilon^2} \tilde{Z}_t^{n,\varepsilon,\delta}
\]
If \(0 \leq t \leq \theta_{n,\varepsilon}^2\), then
\[
\left| X_t^\varepsilon - X_{\tau_n(t)}^n \right|^2 \leq \left( |X_t^\varepsilon - X_{\tau_n(t)}^n| + \varepsilon \right)^2 \leq 2 \left( |X_t^\varepsilon - X_{\tau_n(t)}^n|^2 + \varepsilon^2 \right)
\]
and thus
\[
\left\| X_t^\varepsilon - X_{\tau_n(t)}^n \right\|^2 \leq \frac{6L \sigma}{\varepsilon^2} Z_t^\varepsilon
\]
Note also that \(\tilde{Z}_0^\varepsilon = \varepsilon^{2/\varepsilon^2}\). Set now \(\lambda \equiv 2L + 6L \sigma^2\); then
\[
Z_t^{\varepsilon,\delta} \leq \varepsilon^{2/\varepsilon^2} + \frac{\lambda}{\varepsilon^2} \int_0^t Z_s^{\varepsilon,\delta} ds + M_t
\]
where \(M_t\) is a martingale. Setting
\[
\tilde{Z}_t^{\varepsilon,\delta} \equiv \exp \left[ -\frac{\lambda}{\varepsilon^2} (t \wedge \theta_{3,n,\varepsilon}^2) \right] Z_t^{\varepsilon,\delta}, \quad t \in [0,T]
\]
we have that \(\tilde{Z}_t^{\varepsilon,\delta}\) is a submartingale. Thus
\[
\mathbb{P}\left\{ \tilde{Z}_t^{\varepsilon,\delta} \geq (\delta^2 + \varepsilon^2)^{1/\varepsilon^2} \right\} \leq \mathbb{P}\left\{ \tilde{Z}_t^{\varepsilon,\delta} \geq e^{-\lambda T/\varepsilon^2} (\delta^2 + \varepsilon^2)^{1/\varepsilon^2} \right\}
\]
\[
\leq \frac{\mathbb{E}\left[ \tilde{Z}_t^{\varepsilon,\delta} \right]}{e^{-\lambda T/\varepsilon^2} (\delta^2 + \varepsilon^2)^{1/\varepsilon^2}} \leq \left( \frac{\varepsilon^2 e^{\lambda T}}{\delta^2 + \varepsilon^2} \right)^{1/\varepsilon^2}.
\]
Thus
\[
\lim_{\varepsilon \searrow 0} \lim_{\varepsilon \searrow 0} \varepsilon^2 \ln \mathbb{P}\left\{ \theta_3^{\varepsilon,\delta} < T, \theta_2^{\varepsilon,\delta} = T \right\} \leq \lim_{\varepsilon \searrow 0} \lim_{\varepsilon \searrow 0} \varepsilon^2 \ln \left( \frac{\varepsilon^2 e^{\lambda T}}{\delta^2 + \varepsilon^2} \right)^{1/\varepsilon^2} = -\infty.
\]
Collecting things together, we get the desired result. \(\square\)

**Exercises**

1. Prove Lemma 0.23. Hint: use something like Gronwall’s inequality.
2. We want to prove Lemma 0.25. For \(\varphi \in C[0,T]\), define \(T\varphi \in C[0,T]\) as the solution of
\[
(T\varphi)(t) = x_0 + \int_0^t b(\varphi(s)) ds + \sigma(\varphi(\tau_n(t)) \varphi(t) - \varphi(\varphi(\tau_n(t)))) + \sum_{0 \leq j \leq \lfloor tn \rfloor - 1} \sigma(j/n) (\varphi(j/n) - \varphi(j/n)) \quad t \in [0,T]
\]
(a) Show that \(T\) is well-defined and continuous.
(b) Show that \(X_{\tau_n}^\varepsilon = T_n(\varepsilon W)\).
(c) Show that (27) is the action functional for \(X_{\tau_n}^\varepsilon\).
Large Deviations for Markov Chains

Large deviations studies 'atypical' behavior. The Law of Large Numbers gives us one framework of 'typical' behavior. Let's see that we also have a notion of 'typical' behavior for a Markov chain.

To begin, the state space of our Markov chain will be a finite set $S$. The paths of the Markov chain will thus be elements of $\Omega \overset{\text{def}}{=} S^\mathbb{N}$. For each $n \in \mathbb{N} \cup \{0\}$, and $\omega = (\omega_0, \omega_1, \ldots)$, define the coordinate random variable $X_n(\omega) \overset{\text{def}}{=} \omega_n$. Fix a transition matrix $P = (p(s, s') \mid s, s' \in S)$ and an initial distribution $\pi = (\pi(s) \mid s \in S)$; we then consider the probability measure $P$ on $(\Omega, \mathcal{B}(\Omega))$ defined by requiring that

$$
P\left(\bigcap_{n=0}^{N} \{X_n = s_n\}\right) = \pi(s_0) \prod_{n=0}^{N-1} p(s_n, s_{n+1})$$

for all $n \in \mathbb{N}$ and all $(s_0, s_1, \ldots s_n) \in S^{N+1}$.

We are interested in the asymptotic behavior of the occupation measure of $X$.

To set up our notation, define $P_{\circ}(S) \overset{\text{def}}{=} \left\{ \pi \in \mathbb{R}_+^S : \sum_{s \in S} \pi_s = 1 \right\}$; this is of course equivalent to the collection of probability measures on $S$. For each $s \in S$ and $N \in \mathbb{N}$, define

$$L_N(s) \overset{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \chi_{\{s\}}(X_n).$$

clearly $L_N$ is a random element of $P_{\circ}(S)$. We topologize $P_{\circ}(S)$ as a subset of $L^1(\mathbb{R}^S)$.

The "typical" behavior of $L_N$ for $N$ large is given by the stationary distribution of the Markov chain. To avoid a number of details, recall a standard notion from the theory of Markov chains.

**Definition 0.2.** The Markov chain $P$ is irreducible if for every distinct $s$ and $s'$ in $S$ there is an $N \in \mathbb{N}$ and an $(s_0, s_1, s_2 \ldots s_N)$ in $S^N$ with $s_0 = s$ and $s_N = s'$ such that $\prod_{n=0}^{N-1} p_{s_n, s_{n+1}} > 0$.

We then have a standard result from the theory of Markov chains.

**Lemma 0.30.** If $P$ is irreducible, it has a unique stationary distribution $\mu \in P_{\circ}(S)$; i.e., a unique $\mu \in \mathcal{P}(S)$ such that

$$\mu(s) = \sum_{s' \in S} \mu(s') p(s, s') \quad s \in S$$

Furthermore,

$$\lim_{N \to \infty} L_N = \mu$$

in probability.

We will prove this in Subsection 0.4.

Define $T_P : B(S) \to B(S)$ as

$$(T_P f)(s) \overset{\text{def}}{=} \sum_{s' \in S} p(s, s') f(s')$$

for all $f \in B(S)$ and $s \in S$. This is the transition operator associated with $P$. 

35
Let’s now see if we can understand “unlikely” events. For \( q \in \mathcal{P}_x(S) \), define

\[
I(q) \overset{\text{def}}{=} \sup_{\Phi \in B(S)} \sum_{s \in S} \left\{ \ln \frac{e^{\phi(s)}}{(Te^\Phi)(s)} q(s) \right\}
\]

**Lemma 0.31.** We have that \( I : \mathcal{P}_x(S) \rightarrow [0, \infty] \). Furthermore, \( \{q \in \mathcal{P}_x(S) : I(q) \leq L\} \subset \mathcal{P}_x(S) \) for all \( L \geq 0 \).

**Proof.** First note that by taking \( \Phi \equiv 1 \), we get that \( I \geq 0 \). As in our earlier calculations, for each \( L \geq 0 \) we have that

\[
\{q \in \mathcal{P}_x(S) \leq L\} = \bigcap_{\Phi \in B(S)} \left\{ q \in \mathcal{P}_x(S) : \sum_{s \in S} \ln \frac{e^{\phi(s)}}{(Te^\Phi)(s)} q(s) \leq L \right\}.
\]

Since the map \( q \mapsto \sum_{s \in S} \ln \frac{e^{\phi(s)}}{(Te^\Phi)(s)} q(s) \) is continuous, \( \{q \in \mathcal{P}_x(S) \leq L\} \) is an intersection of closed sets and is thus closed. Since \( \mathcal{P}_x(S) \) is compact, we can conclude the proof.

**Lemma 0.32.** Fix \( q \in \mathcal{P}_x(S) \). If \( I(q) < \infty \), we have that

\[
\lim_{\delta \searrow 0} \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{P}\left\{ \|L_N - q\| < \delta \right\} + I(q) = 0.
\]

If \( I(q) = \infty \), then

\[
\lim_{\delta \searrow 0} \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{P}\left\{ \|L_N - q\| < \delta \right\} = -\infty.
\]

**Proof.** Fix \( \Phi \in B(S) \) and define

\[
\tilde{\Phi}(s) \overset{\text{def}}{=} \ln \frac{e^{\phi(s)}}{(Te^\Phi)(s)} = \Phi(s) - \log (Te^\Phi)(s) \quad s \in S.
\]

We now write that

\[
\mathbb{P}\left\{ \|L_N - q\| < \delta \right\} = \mathbb{E}\left[ \chi_{\{\|L_N - q\| < \delta\}} \exp \left[ -N \sum_{s \in S} \tilde{\Phi}(s)L_N(s) \right] \exp \left[ N \sum_{s \in S} \tilde{\Phi}(s)L_N(s) \right] \right].
\]

First note that if \( \|L_N - q\| < \delta \), then

\[
\left| \sum_{s \in S} \tilde{\Phi}(s)L_N(s) - \sum_{s \in S} \tilde{\Phi}(s)q(s) \right| = \left| \sum_{s \in S} \Phi(s) (L_N(s) - q(s)) \right| \leq \|\Phi\|_\infty \delta.
\]

Secondly,

\[
\exp \left[ N \sum_{s \in S} \tilde{\Phi}(s)L_N(s) \right] = \exp \left[ \sum_{n=1}^{N} \tilde{\Phi}(X_n) \right] = \exp \left[ \sum_{n=1}^{N} \Phi(X_n) - \sum_{n=1}^{N} \ln (Te^\Phi)(X_n) \right] = \exp \left[ \sum_{n=1}^{N} \Phi(X_n) - \sum_{n=0}^{N-1} \ln (Te^\Phi)(X_n) \right] \left( \frac{Te^\Phi}(X_0) \right)^{N-1} = \left( \prod_{n=1}^N \frac{e^{\Phi(X_n)}}{(Te^\Phi)(X_{n-1})} \right) \left( \frac{Te^\Phi}(X_0) \right)^{N-1}
\]

For \( N \in \mathbb{N} \), define the measure \( \hat{\mathbb{P}}_N \) on \( (\Omega, \mathcal{B}(\Omega)) \) as

\[
\hat{\mathbb{P}}_N(A) \overset{\text{def}}{=} \mathbb{E}\left[ \chi_A \left\{ \prod_{n=1}^N \frac{e^{\Phi(X_n)}}{(Te^\Phi)(X_{n-1})} \right\} \right]. \quad A \in \mathcal{B}(\Omega)
\]

To understand the measure \( \hat{\mathbb{P}}_N \), define the matrix \( P^\Phi = (p^\Phi(s,s'); s, s' \in S) \) as

\[
p^\Phi(s, s') \overset{\text{def}}{=} \frac{p(s, s') e^{\Phi(s')}}{(Te^\Phi)(s)}. \quad s, s' \in S
\]
Clearly \( \hat{P}\Phi \) is a stochastic matrix. For any \((s_0, s_1, \ldots s_N) \in S^N\),

\[
\hat{P}_N \{ \cap_{n=0}^N X_n = s_n \} = \mathbb{E} \left[ \prod_{n=0}^N \chi(X_n=s_n) \prod_{n=1}^N \frac{\exp(\Phi(s_n))}{(T_p e^\Phi)(s_{n-1})} \right] \\
= \left\{ \prod_{n=1}^N \frac{\exp(\Phi(s_n))}{(T_p e^\Phi)(s_{n-1})} \right\} \pi(s_0) \prod_{n=1}^N \hat{p}(s_n) = \pi(s_0) \prod_{n=1}^N \hat{p}(s_n).
\]

Thus \( \hat{P}_N \) is a probability measure and \( \{ X_n \}_{n=0}^N \) is Markovian with transition matrix \( \hat{P}\Phi \) under \( \hat{P}_N \). Collecting things together, we now have that

\[
P \{ \| L_N - q \| < \delta \} = \mathbb{E}_N \left[ \chi(\| L_N - q \| < \delta) \exp \left[ -N \sum_{s \in S} \hat{\Phi}(s) L_N(s) \right] \exp \left[ N \sum_{s \in S} \hat{\Phi}(s) L_N(s) \right] \right].
\]

Suppose now that \( I(q) < \infty \). Then

\[
I(q) = \sum_{s \in S} q(s) \ln \frac{\exp(\Phi(s))}{(T_p e^\Phi)(s)} = \sup_{\Phi \in B(S)} \sum_{s \in S} \{ \Phi(s) - \ln (T_p e^\Phi)(s) \} q(s)
\]

If the supremum is achieved, then the first-order conditions of optimality are thus that

\[
\sum_{s \in S} \left\{ \eta(s) \left( \frac{T_p(e^\Phi \eta)(s)}{(T_p e^\Phi)(s)} \right) \right\} q(s) = 0
\]

for all \( \eta \in B(S) \). We see that

\[
\frac{(T_p(e^\Phi \eta))(s)}{(T_p e^\Phi)(s)} = (\hat{T}_\Phi \eta)(s)
\]

for all \( s \in S \). Thus (31) means that

\[
q(s) = \sum_{s' \in S} \hat{P}(s, s')
\]

for all \( s \in S \), so \( q \) a stationary distribution for \( \hat{P}\Phi \). In fact, by Lemmas 0.35 and (0.36), we know that \( \lim_{N \to \infty} \hat{P}_N \{ \| L_N - q \| < \delta \} = 1 \). Since \( \Phi \in B(S) \),

\[
e^{-2\|\Phi\|_\infty} \leq \frac{(T_p e^\Phi)(X_0)}{(T_p e^\Phi)(X_N)} \leq e^{2\|\Phi\|_\infty}.
\]

Combining things we get the claim when \( I(q) < \infty \).

Suppose next that \( I(q) = \infty \). Clearly

\[
\lim_{\delta \downarrow 0} \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{P} \{ \| L_N - q \| < \delta \} = -\infty.
\]

Fix \( \Phi \in B(S) \); from (30) we have that

\[
\lim_{\delta \downarrow 0} \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{P} \{ \| L_N - q \| < \delta \} \leq - \sum_{s \in S} \left\{ \ln \frac{\exp(\Phi(s))}{(T_p e^\Phi)(s)} q(s) \right\}.
\]

Since \( I(q) = \infty \), upon letting \( \Phi \) vary, we get the desired claim. \( \square \)
0.4. Proofs. For each \( s \in \mathcal{S} \), define
\[
\tau_s \overset{\text{def}}{=} \inf\{ n \geq 1 : X_n = s \}
\]
and let \( \mathbb{P}_s \) be the probability measure on \( (\Omega, \mathcal{F}(\Omega)) \) defined by requiring that
\[
\mathbb{P}_s \left( \bigcap_{n=0}^{N} \{ X_n = s_n \} \right) = \delta_s(s_0) \prod_{n=0}^{N-1} p(s_n, s_{n+1})
\]
for all \( n \in \mathbb{N} \) and all \( (s_0, s_1, \ldots, s_n) \in \mathcal{S}^{N+1} \). Thus of course
\[
\mathbb{P}(A) = \sum_{s \in \mathcal{S}} \mathbb{P}_s(A) \pi_s \quad A \in \mathcal{F}(\Omega)
\]
For \( s \in \mathcal{S} \), define
\[
\tau_s^1 \overset{\text{def}}{=} \min\{ n > 0 : X_n = s \} \\
\tau_s^k \overset{\text{def}}{=} \min\{ n > \tau_s^{k-1} : X_n = s \} \quad k \in \mathbb{N} \quad (\inf \emptyset = \infty)
\]
For convenience, set \( \tau_s \overset{\text{def}}{=} \tau_s^0 \). For each \( k \in \mathbb{N} \), let’s also define
\[
S_s^k \overset{\text{def}}{=} \begin{cases} 
\tau_s^{k+1} - \tau_s^k & \text{if } \tau_s^k < \infty \\
0 & \text{else}
\end{cases}
\]
Also, for each \( s \in \mathcal{S} \), define
\[
V_s^n \overset{\text{def}}{=} \sum_{j=1}^{n} \chi_{\{ X_j = s \}} \\
V_s^\infty \overset{\text{def}}{=} \sum_{j=1}^{\infty} \chi_{\{ X_j = s \}} = \lim_{n \to \infty} V_s^n.
\]
Thus \( V_s^n \) is the number of times \( X \) visits \( s \) during the set of times \( \{1, 2 \ldots n\} \). Note that
\[
L_n(s) = \frac{1}{n} V_s^n.
\]
\( V_s^n \) is the number of times that \( X \) visits the state \( i \) in the times \( \{1, 2 \ldots n\} \).

We start by understanding the structure of the \( S_s^k \)'s.

**Lemma 0.33.** Fix \( s \) and \( s' \) in \( \mathcal{S} \), \( A \in \mathcal{F}_{s'} \) and \( k \) and \( n \) in \( \mathbb{N} \). Then
\[
\mathbb{P}_{s'} (\{ S_s^k = n \} \cap A \cap \{ \tau_s^k < \infty \}) = \mathbb{P}_s (\tau = n) \mathbb{P}_{s'} (A \cap \{ \tau_s^k < \infty \})
\]
Thus, conditionally on \( \tau_s^k < \infty \), \( S_s^k \) is independent of \( \mathcal{F}_{s'} \) and has the \( \mathbb{P}_s \)-law of \( \tau \).

**Proof.** We have that
\[
\mathbb{P}_{s'} (\{ S_s^k = n \} \cap A \cap \{ \tau_s^k < \infty \}) = \sum_{m=0}^{\infty} \mathbb{P}_{s'} (\{ S_s^k = n \} \cap A \cap \{ \tau_s^k = m \}).
\]
Since \( \tau_s^k \) is a stopping time,
\[
A \cap \{ \tau_s^k = m \} = (A \cap \{ \tau_s^k \leq m \}) \setminus \{ \tau_s^k \leq m - 1 \} \in \mathcal{F}_m
\]
if \( m \in \mathbb{N} \) is positive; of course \( A \cap \{ \tau_s^k = 0 \} = A \cap \{ \tau_s^k \leq 0 \} \in \mathcal{F}_0 \). Thus for any \( m \in \{0, 1 \ldots \},
\]
\[
\mathbb{P}_{s'} (\{ S_s^k = n \} \cap A \cap \{ \tau_s^k = m \}) = \mathbb{P}_{s'} \left( \bigcap_{j=1}^{n-1} \{ X_{m+j} \neq s \} \cap \{ X_{m+n} = s \} \cap A \cap \{ \tau_s^k = m \} \right)
\]
\[
= \mathbb{P}_s \left( \bigcap_{j=1}^{n-1} \{ X_j \neq s \} \cap \{ X_n = s \} \right) \mathbb{P}_{s'} (A \cap \{ \tau_s^k = m \}).
\]
Thus there is an\[\text{so in fact}\]

\[P_n\] test for series). Hence we indeed have recurrence. □

Fix now \(s \in S\) for all \(s \in S\), so \(P\) is recurrent.

**Proof.** First note that for any \(s\) and \(s'\) in \(S\) and any \(n \in \mathbb{N}\)

\[P_n\{V_s^\infty \geq n + 1\} = P_n\{\tau_{s'}^{n+1} < \infty\} = P_n\{\tau_{s'}^{n+1} < \infty, \tau_s^n < \infty\} = P_n\{S_s^n < \infty, \tau_{s'}^n < \infty\}\]

Observe now that

\[\sum_{s',s'' \in S} E_{s'}[V_{s''}^\infty] \pi(s') = \sum_{s'' \in S} E[V_{s''}^\infty] = \infty.\]

Thus there is an \(s'\) and \(s''\) in \(S\) such that \(E_{s'}[V_{s''}^\infty] = \infty\). Explicitly,

\[\sum_{n=1}^{\infty} p^{(n)}(s', s'') = \infty.\]

Fix now \(s \in S\). Since \(P\) is irreducible, there is an \(n'\) and \(n''\) in \(\mathbb{N}\) such that \(p^{(n'')}\{s, s'\} > 0\) and \(p^{(n')}\{s', s\} > 0\). Thus for any \(n\),

\[p^{(n'+n+n'')}\{s, s\} > p^{(n'')}\{s, s'\} p^{(n')}\{s', s''\} p^{(n')}\{s'', s\}\]

so in fact

\[E_s[V_{s'}^\infty] = \sum_{n=1}^{\infty} p^{(n)}(s, s) = \infty.\]

Rewriting this, we see that

\[\sum_{n=1}^{\infty} P_n\{V_s \geq n\} = E_s[V_{s'}^\infty] = \infty.\]

If \(P_n\{\tau_s < \infty\} < 1\), then by Lemma 0.34 we have that \(E_s[V_{s'}^\infty] < \infty\) which contradicts (34) (use the ratio test for series). Hence we indeed have recurrence.

Next fix \(s\) and \(s'\) in \(S\). By recurrence, \(P_{s'}\{\tau_s < \infty\} = 1\). Using Lemma 0.34, we have that

\[P_{s'}\{V_{s'}^\infty \geq n\} = (P_{s'}\{\tau_s < \infty\})^{n-1} P_{s'}\{V_{s'}^\infty \geq 1\} = (P_{s'}\{\tau_s < \infty\})^n = 1\]

so in fact \(P_{s'}\{V_{s'}^\infty = \infty\} = 1\). By irreducibility, there is an \(n \in \mathbb{N}\) such that \(p^{(n)}(s, s') > 0\). Thus

\[1 = P_{s'}\{V_{s'}^\infty = \infty\} = P_s\left\{\sum_{j=n}^{\infty} \chi_{X_j = s} = \infty\right\} = \sum_{s'' \in S} P_s\left\{\sum_{j=n}^{\infty} \chi_{X_j = s''} = \infty, X_n = s''\right\} = \sum_{s'' \in S} p^{(n)}(s', s'') P_{s''}\{V_{s''}^\infty = \infty\}.\]

Since \(\sum_{s'' \in S} p^{(n)}(s', s'') = 1\), we must in fact have that \(P_s\{V_{s'}^\infty = \infty\} = 1\). Furthermore,

\[1 = P_s\{V_{s'}^\infty = \infty\} \leq P_s\{\tau_{s'} < \infty\} = 1.\]

The proof is complete. □

For each \(s \in S\), define

\[\mu(s) \overset{\text{def}}{=} \begin{cases} \frac{1}{\pi(s, s)} & \text{if } \text{E}_s[\tau_s] < \infty \\ 0 & \text{else} \end{cases}\]

Then we have

**Lemma 0.35.** Since \(P\) is irreducible, \(\lim_{N \to \infty} L_N(s) = \mu(s)\) in probability for each \(s \in S\).
PROOF. Fix $s \in S$. Lemmas 0.33 and 0.34 imply that $\{S^n_k\}_{k \in \mathbb{N}}$ is an i.i.d. collection of random variables with common law being the $\mathbb{P}_s$-law of $\tau_s$ and by Lemma 0.34 the $S^n_k$’s are and $\tau_s^1$ are $\mathbb{P}$-a.s. finite. If $\mathbb{E}_s[\tau_s] < \infty$, then the strong law of large numbers implies that

$$\mathbb{P}_s\left\{\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} S^n_s = \mathbb{E}_s[\tau_s]\right\} = 1.$$ 

If $\mathbb{E}_s[\tau_s] = \infty$, then by virtue of the fact that the $S^n_s$’s are nonnegative,

$$\mathbb{P}_s\left\{\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} S^n_s \wedge M = \mathbb{E}_s[\tau_s \wedge M]\right\} = 1$$

for all $M \in \mathbb{N}$. Thus $\mathbb{P}_s$-a.s.

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} S^n_s \geq \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} S^n_s \wedge M = \mathbb{E}_s[\tau_s \wedge M].$$

Letting $M \not\to \infty$, we get that again

$$\mathbb{P}_s\left\{\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} S^n_s = \mathbb{E}_s[\tau_s]\right\} = 1.$$ 

By Lemma 0.34, we have that

$$\mathbb{P}\{\lim_{n \to \infty} V^n_s = \infty\} = \mathbb{P}\{V^\infty_s = \infty\} = 1.$$ 

If $V^n_s = k \in \mathbb{N}$, then $X$ visits the state $s$ exactly $k$ times during the times $\{1, 2 \ldots n\}$. This means that

$$\tau_s^1 + \sum_{1 \leq k' \leq k-1} S^k_s \leq n < \tau_s^1 + \sum_{1 \leq k' \leq k} S^k_s.$$ 

In other words,

$$\tau_s^1 + \sum_{1 \leq k' \leq V^n_s - 1} S^k_s \leq n < \tau_s^1 + \sum_{1 \leq k' \leq V^n_s} S^k_s.$$ 

Hence

$$\frac{\tau_s^1 + \sum_{1 \leq k' \leq V^n_s - 1} S^k_s}{V^n_s} \leq \frac{n}{V^n_s} < \frac{\tau_s^1 + \sum_{1 \leq k' \leq V^n_s} S^k_s}{V^n_s}.$$ 

Consequently

$$\mathbb{P}\left\{\lim_{n \to \infty} \frac{n}{V^n_s} = \mathbb{E}_s[\tau_s]\right\} = 1.$$ 

This implies the claim. □

This gives us the sought-after stationary distribution.

**Lemma 0.36.** Since $P$ is irreducible, $\mu$ is the unique stationary distribution.

**Proof.** Note that $0 \leq L_N(s) \leq 1$, so the convergence in probability given by Lemma 0.35 implies that

$$\mu(s) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N-1} \mathbb{P}\{X_n = s\}. $$

Let’s first see that $\mu \in \mathcal{P}_c(S)$. Clearly $\mu \geq 0$. Secondly, since $|S| < \infty$,

$$\sum_{s \in S} \mu(s) = \sum_{s \in S} \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{P}\{X_n = s\} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left\{\sum_{s \in S} \mathbb{P}\{X_n = s\}\right\} = 1.$$ 

Let’s next see that $\mu$ is indeed a stationary distribution. For any $s \in S$,

$$\mu(s) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{P}\{X_n = s\} = \lim_{N \to \infty} \frac{1}{N} \left(\pi(s) + \sum_{n=1}^{N-1} \left\{\sum_{s' \in S} \mathbb{P}\{X_{n-1} = s'\} p(s', s)\right\}\right)$$

40
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-2} \left\{ \sum_{s' \in S} \mathbb{P}\{X_n = s'\} p(s', s) \right\} = \sum_{s' \in S} p(s', s) \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-2} \mathbb{P}\{X_n = s'\} = \sum_{s' \in S} \mu(s') p(s', s).
\]

Fix next any stationary distribution \( \tilde{\mu} \in \mathcal{P}_c(S) \). Since \( \mu \in \mathcal{P}_c(S) \), there is an \( s^* \in S \) such that \( \mu(s^*) > 0 \); i.e., \( \mathbb{E}_s[\tau_s] < \infty \). Fix \( s \in S \). Then

\[
\tilde{\mu}(s) = \sum_{s_1 \in S} \tilde{\mu}(s_1) p(s_1, s) = \sum_{s_1 \in S, s_1 \neq s} \tilde{\mu}(s_1) p(s_1, s) + \tilde{\mu}(s^*) p(s^*, s)
\]

\[
= \sum_{s_1, s_2 \in S, s_1 \neq s} \tilde{\mu}(s_2) p(s_2, s_1) p(s_1, s) + \sum_{s_1 \in S, s_1 \neq s} \tilde{\mu}(s^*) p(s^*, s_1) p(s_1, s) + \tilde{\mu}(s^*) p(s^*, s)
\]

\[
= \sum_{s_1, s_2 \in S, s_1 \neq s} \tilde{\mu}(s_2) p(s_2, s_1) p(s_1, s) + \tilde{\mu}(s^*) \left\{ \sum_{s_1 \in S, s_1 \neq s} p(s^*, s_1) p(s_1, s) + p(s^*, s) \right\}
\]

Note that

\[
p(s^*, s) = \mathbb{P}_{s^*} \{X_1 = s\} = \mathbb{P}_{s^*} \{X_1 = s, \tau_{s^*} > 0\}
\]

\[
\sum_{s_1 \neq s} p(s^*, s_1) p(s_1, s) = \mathbb{P}_{s^*} \{X_1 \neq s^*, X_2 = s\} = \mathbb{P}_{s^*} \{X_1 = s, \tau_{s^*} > 1\}
\]

Extending (35) and using the fact that all the terms are nonnegative, we get that \( \tilde{\mu}(s) \geq \tilde{\mu}(s^*) \gamma(s) \) where

\[
\gamma(s) \overset{\text{def}}{=} \sum_{n=0}^{\infty} \mathbb{P}_{s^*} \{X_n = s, \tau_{s^*} > n\} = \mathbb{E}_{s^*} \left[ \sum_{n=0}^{\tau_{s^*} - 1} \chi(X_n = s) \right].
\]

Since this holds for all \( s \in S \), we have that \( \tilde{\mu} \geq \tilde{\mu}(s^*) \gamma \). Note that since \( \mathbb{E}_s[\tau_s] < \infty \), \( X_0 = X_{\tau_s} = s^* \) \( \mathbb{P}_{s^*}\)-a.s. Thus

\[
\gamma(s) = \mathbb{E}_{s^*} \left[ \sum_{n=1}^{\tau_{s^*}} \chi(X_n = s) \right] = \sum_{n=1}^{\infty} \mathbb{P}_{s^*} \{X_n = s, n \leq \tau_{s^*}\} = \sum_{n=1}^{\infty} \mathbb{P}_{s^*} \{X_n = s, n-1 < \tau_{s^*}\}
\]

\[
= \sum_{s' \in S} \sum_{n=1}^{\infty} \mathbb{P}_{s^*} \{X_n = s, X_{n-1} = s', n-1 < \tau_{s^*}\}
\]

\[
= \sum_{s' \in S} \sum_{n=1}^{\infty} \mathbb{P}_{s^*} \left( \{X_n = s\} \cap \{X_{n-1} = s'\} \cap \bigcap_{j=1}^{n-1} \{X_j \neq s^*\} \right)
\]

\[
= \sum_{s' \in S} \sum_{n=1}^{\infty} p(s', s) \mathbb{P}_{s^*} \left( \{X_{n-1} = s'\} \cap \bigcap_{j=1}^{n-1} \{X_j \neq s^*\} \right) = \sum_{s' \in S} \sum_{n=0}^{\infty} p(s', s) \mathbb{P}_{s^*} \left( \{X_n = s'\} \cap \bigcap_{j=1}^{n} \{X_j \neq s^*\} \right)
\]

\[
= \sum_{s' \in S} \sum_{n=0}^{\infty} \mathbb{P}_{s} \{X_n = s', \tau_{s^*} \geq n\} = \sum_{s' \in S} p(s', s) \gamma(s').
\]
Fix \( s \in S \). Since \( P \) is irreducible, there is an \( n \in \mathbb{N} \) such that \( p^{(n)}(s, s^*) > 0 \). Thus, noting that \( \gamma(s^*) = 1 \), we have that

\[
0 = \tilde{\mu}(s^*) - \tilde{\mu}(s^*)\gamma(s^*) = \sum_{s' \in S} p^{(n)}(s', s^*) \{ \tilde{\mu}(s') - \tilde{\mu}(s^*)\gamma(s') \}.
\]

Since \( \tilde{\mu} \geq \tilde{\mu}(s^*)\gamma \), all of the terms on the right are nonnegative. Thus

\[
0 \geq p^{(n)}(s, s^*)\{ \tilde{\mu}(s^*) - \mu(s)\gamma(s^*) \}.
\]

and thus in fact \( \tilde{\mu}(s) = \tilde{\mu}(s^*)\gamma(s) \). Thus all stationary distributions must coincide. \( \square \)

**Proof of Lemma 0.30.** Combine things. \( \square \)

### 1. Pair Empirical Measure

Let’s use the theory for Markov chains to look at *pairs*. Namely, consider

\[
L^{(2)}(s, s') \overset{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \chi((s, s'))(X_{n-1}, X_n) \quad (s, s') \in S^2
\]

This counts the number of times that \( X \) makes a transition from \( s \) to \( s' \). We note that \( X^{(2)}_n \overset{\text{def}}{=} (X_n, X_{n+1}) \) is Markovian with state space \( S^2 \) and transition matrix \( P^{(2)} = (p^{(2)}((s_1, s_2), (s'_1, s'_2)) : (s_1, s_2), (s'_1, s'_2) \in S^2) \) defined by

\[
p^{(2)}((s_1, s_2), (s'_1, s'_2)) = \begin{cases} p(s'_1, s'_2) & \text{if } s_2 = s'_1 \\ 0 & \text{else} \end{cases}
\]

We also define \( T : B(S^2) \to B(S^2) \) as

\[
(T_{p^{(2)}} f)(s_1, s_2) \overset{\text{def}}{=} \sum_{(s'_1, s'_2) \in S^2} p^{(2)}((s_1, s_2), (s'_1, s'_2)) f(s'_1, s'_2) = \sum_{s'_2} f(s_2, s'_2) p(s_2, s'_2)
\]

Define

\[
I^{(2)}(q) \overset{\text{def}}{=} \sup_{\Phi \in B(S^2)} \sum_{(s, s') \in S^2} q(s, s') \ln \frac{e^{\Phi(s_1, s_2)}}{(Te^\Phi)(s_1, s_2)}
\]

for all \( q \in \mathcal{P}_e(S^2) \). We are interested in a simpler formulation of this expression. For \( q \in \mathcal{P}_e(S^2) \), define

\[
q_1(s) \overset{\text{def}}{=} \sum_{s' \in S} q(s, s') \quad \text{and} \quad q_2(s) \overset{\text{def}}{=} \sum_{s' \in S} q(s', s)
\]

for all \( s \in S \). If \( q_1(s) > 0 \), define

\[
q_{2|1}(s'|s) \overset{\text{def}}{=} \frac{q(s, s')}{q_1(s)} \quad s' \in S
\]

**Lemma 1.1.** We have that

\[
I^{(2)}(q) = \begin{cases} \sum_{s \in S} q_1(s) H \left( q_{2|1} | s \right) p(s, \cdot) & \text{if } q_1 = q_2 \\ \infty & \text{else} \end{cases}
\]

for all \( q \in \mathcal{P}_e(S^2) \).
We use here the fact that the right-hand side of (36) doesn’t depend on \( q \).

Then

\[
\tau(1) = \sup \left\{ q(s) \Phi(s_1, s_2) - \sum_{s', s'' \in S} q(s) \ln \sum_{s'' \in S} e^{\Phi(s', s'')} p(s', s'') \right\}
\]

(38)

\[
= \sup_{\Phi \in B(S^2)} \left\{ \sum_{(s, s') \in S^2} q(s, s') \Phi(s_1, s_2) - \sum_{(s', s'') \in S^2} q(s, s') \ln \sum_{s'' \in S} e^{\Phi(s', s'')} p(s', s'') \right\}
\]

\[
= \sup_{\Phi \in B(S^2)} \left\{ \sum_{(s, s') \in S^2} q(s, s') \Phi(s_1, s_2) - \sum_{s' \in S} q_2(s') \ln \sum_{s'' \in S} e^{\Phi(s', s'')} p(s', s'') \right\}
\]

We use here the fact that the right-hand side of (36) doesn’t depend on \( s_1 \).

If \( q_1 = q_2 \), then we can continue (38) as

\[
I^{(2)}(q) = \sup_{\Phi \in B(S^2)} \left\{ \sum_{(s, s') \in S^2} q(s, s') \Phi(s_1, s_2) - \sum_{s \in S} q_1(s) \ln \sum_{s' \in S} e^{\Phi(s, s')} p(s, s') \right\}
\]

Next suppose that \( q_1 \neq q_2 \). Then there is an \( s^* \in S \) such that \( q_1(s^*) < q_2(s^*) \) (we use here the fact that both \( q_1 \) and \( q_2 \) sum up to 1). For each \( \alpha \in \mathbb{R} \), define

\[
\Phi_\alpha(s, s') = \begin{cases} 
\alpha & \text{if } s = s^* \\
0 & \text{else}
\end{cases}
\]

Then

\[
I(q) \geq \sup_{\alpha \in \mathbb{R}} \left\{ \sum_{(s, s') \in S^2} q(s, s') \Phi_\alpha(s_1, s_2) - \sum_{s' \in S} q_2(s') \ln \sum_{s'' \in S} e^{\Phi_\alpha(s', s'')} p(s', s'') \right\}
\]

\[
= \sup_{\alpha \in \mathbb{R}} \left\{ \alpha q_1(s^*) - q_2(s^*) \alpha - \sum_{s' \neq s^*} q_2(s) \right\} = \sup_{\alpha \in \mathbb{R}} \{ \alpha (q_1(s^*) - q_2(s^*)) - (1 - q_2(s^*)) \} = \infty.
\]

This completes the proof. \( \square \)

When \( q_1 = q_2 \), we say that \( q \) is shift-invariant.

Exercises

1. Show that \( \tau_n \) of (32) is a stopping time; i.e., for each \( n \in \mathbb{N} \),

\[
\{ \tau_n \leq n \} \in \sigma\{ X_0, X_1, \ldots X_n \}.
\]

2. Show that thanks to Lemmas 0.33 and 0.34

\[
\mathbb{P}_s \left( \cap_{k=1}^K \{ S^k \equiv n_k \} \right) = \prod_{k=1}^K \mathbb{P}_s \{ S^k_s \equiv n_k \}
\]

for all \( K \in \mathbb{N} \) and \( (n_1, n_2, \ldots n_K) \in \mathbb{N}^K \).
(3) Show that $I$ of (29) can be rewritten as

$$I(q) = \sup_{\Phi \in B(S)} \sum_{s \in S} \left\{ \ln \frac{\Phi(s)}{(T\Phi)(s)} q(s) \right\} = \sup_{0 < \Phi \leq 1} \sum_{s \in S} \left\{ \ln \frac{\Phi(s)}{(T\Phi)(s)} q(s) \right\}.$$ 

(4) Assume that $S = \{1, 2\}$. Compute $I$ of (29) for the matrices

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \alpha & 1 - \alpha \end{pmatrix}$$

where for the third matrix $\alpha$ is a fixed element of $(0, 1)$. Note that for the first and second matrices, there is no randomness in the dynamics. For the first, every distribution is stationary, but for the second there is only one stationary distribution. The third matrix is irreducible.
CHAPTER 6

Refined Asymptotics

Let’s now discuss some issues of refined asymptotics. A general theory would be much less than instructive than a full analysis of a special case, so we return to the case of Subsection 0.2. We again define $X_N$ as in (4). Fix $\alpha \in (p, 1)$. Then we should have that

$$\lim_{N \to \infty} \frac{1}{N} \ln \mathbb{P}\{X_N \geq \alpha\} \leq - \inf_{\alpha' \geq \alpha} H(\alpha', p) < 0$$

so $\{X_N \geq \alpha\}$ is a rare event.

Fix now a measurable map $\Psi : \mathbb{R} \to \mathbb{R}_+$. We want to compute

$$I_N = \mathbb{E}[\Psi(X_N - \alpha) \chi_{[\alpha, \infty)}(X_N)] = \mathbb{E}[\Psi(X_N - \alpha) \chi_{[0, \infty)}(X_N - \alpha)].$$

If $\Psi$ is nice enough (e.g., bounded), we should have that $\lim_{N \to \infty} I_N = 0$, but how quickly?

To start, let’s write down the measure change suggested by the Gartner-Ellis theorem. We have that

$$H(\alpha, p) = \sup_{\theta \in \mathbb{R}} \{\theta \alpha - \ln(pe^{\theta} + 1 - p)\}.$$

The supremum is attained at $\theta^*$, where

$$\alpha = \frac{pe^{\theta^*}}{pe^{\theta^*} + 1 - p}.$$

We then write that

$$I_N = \tilde{I}_N \exp[-NH(\alpha, p)]$$

where

$$\tilde{P}_N(A) \overset{\text{def}}{=} \frac{\mathbb{E}[\chi_A[N\theta^*(X_N - \alpha)]]}{\mathbb{E}[(N\theta^*(X_N - \alpha))]}.$$

$$\tilde{I}_N = \tilde{\mathbb{E}}_N[\Phi(X_N - \alpha) \exp[-N\theta^*(X_N - \alpha)]]$$

We recall that under $\tilde{P}_N$, $\{\xi_1, \xi_2 \ldots \xi_N\}$ are i.i.d. Bernoulli random variables with

$$\tilde{P}_N(\xi_n = 1) = \frac{pe^{\theta^*}}{pe^{\theta^*} + 1 - p} = \alpha.$$

Thus, as expected, $\tilde{\mathbb{E}}_N[X_N] = \alpha$.

To make headway with this specific problem, let’s define

$$\tilde{X}_N \overset{\text{def}}{=} N(X_N - \alpha) = \sum_{n=1}^{N} \xi_n - N\alpha$$

$$S_N \overset{\text{def}}{=} \{n - N\alpha : n \in \{0, 1 \ldots N\}\}$$

Then

$$\tilde{I}_N = \tilde{\mathbb{E}}_N[\Psi\left(N^{-1}\tilde{X}_N\right) \exp[-\theta^*\tilde{X}_N]]$$

and $\tilde{X}_N$ takes values in $S_N$. We will later choose some specific $\Phi$’s. For the moment, we know that $\tilde{\mathbb{E}}_N[\tilde{X}_N] = 0$. It is easy to see that $\tilde{\mathbb{E}}_N\left[(\tilde{X}_N)^2\right] = N\alpha(1 - \alpha)$, so in fact the variance of $\tilde{X}_N$ tends
to infinity. Note further by the central limit theorem that \( \frac{1}{\sqrt{N}} \tilde{X}_N \) should approximately be Gaussian as \( N \to \infty \).

**Remark 0.2.** As an example of the purpose of things, let’s assume that \( \Psi(z) = z^\alpha \) for some \( \alpha > 0 \). Admitting some prescience, let’s replace \( \tilde{X}_N \) by a Gaussian with mean 0 and variance \( N\alpha(1 - \alpha) \). We can then compute that

\[
\int_0^\infty \frac{1}{\sqrt{2\pi N\alpha(1 - \alpha)}} \left( \frac{z}{N} \right)^\alpha \exp\left[-\frac{z^2}{2N\alpha(1 - \alpha)}\right] dz \\
\approx \frac{1}{N^{\alpha+1/2}/2\pi\alpha(1 - \alpha)} \int_0^\infty z^\alpha e^{-\theta^* z} dz = \frac{1}{N^{\alpha+1/2}/2\pi\alpha(1 - \alpha)} \Gamma(\alpha + 1) (\theta^*)^{\alpha+1}
\]

This is the type of asymptotic result we want.

To mimic this calculation, let’s assume some regularity of \( \Psi \).

**Assumption 0.3.** We assume that

\[
\sup_{z \geq 0} \left| \Psi(z) \right| < \infty
\]

for some \( \beta > 0 \) and that \( C \) \( \text{def} \lim_{N \to \infty} \frac{\Psi(z)}{z^\beta} \) is well-defined, finite, and positive for some \( \alpha > 0 \).

The negative exponential in \( \tilde{I}_N \) should mean that most of the mass of the integral in \( \tilde{I}_N \) is centered where \( \tilde{X}_N \) is \( O(1) \). Thus, we want to find the behavior of \( \tilde{P}_N \{ \tilde{X}_N = s \} \) for \( s \in \mathcal{S}_N \) of order 1 as \( N \to \infty \).

We proceed by using characteristic functions. Fix \( N \in \mathbb{N} \) and \( \psi \in \mathbb{R} \) and define

\[
\mathcal{P}_N(\psi) = \mathbb{E}_N \left[ \exp[i\psi \tilde{X}_N] \right] \quad \psi \in \mathbb{R}
\]

We then have that

\[
\mathcal{P}_N(\psi) = \sum_{n=0}^{N} \exp[i\psi(n - N\alpha)] \tilde{P}_N \left\{ \tilde{X}_N = n - N\alpha \right\}
\]

(39)

\[
\mathcal{P}_N(\psi) = \mathbb{E}_N \left[ \exp \left[ i\psi \left\{ \sum_{n=1}^{N} \xi_n - N\alpha \right\} \right] \right] = \exp[-i\psi N\alpha] (e^{i\psi} + 1 - \alpha)^N
\]

where

\[
\hat{\mathcal{P}}(\psi) \overset{\text{def}}{=} \left\{ 1 + \alpha (e^{i\psi} - 1) \right\} e^{-i\psi}
\]

Firstly, we can rearrange the first equation of (39) get that

\[
\mathcal{P}_N(\psi)e^{iN\alpha\psi} = \sum_{n=0}^{N} e^{i\psi n} \tilde{P}_N \left\{ \tilde{X}_N = n - N\alpha \right\}.
\]

Integrating against \( e^{i\psi m} \), we get that

\[
\tilde{P}_N \left\{ \tilde{X}_N = n - N\alpha \right\} = \int_{\psi \in (-\pi, \pi)} \mathcal{P}_N(\psi) \exp[-i(n - N\alpha)\psi] d\psi
\]

so that in fact

\[
\hat{\mathcal{P}}_N \left\{ \tilde{X}_N = s \right\} = \int_{\psi \in (-\pi, \pi)} \mathcal{P}_N(\psi) \exp[-i\psi] d\psi
\]

for all \( s \in \mathcal{S}_N \).

Let’s next look at the second equation of (39). Since \( \mathcal{P}_M \) is a characteristic function, |\( \mathcal{P}_N \)| \( \leq 1 \). We actually have that |\( \mathcal{P}_N(\psi) \)| \( = |\hat{\mathcal{P}}(\psi)|^N \). Note that \( \hat{\mathcal{P}}_N(0) = 1 \). We can actually get a more precise estimate.
Lemma 0.4. There is a $K_1 > 0$ such that

$$|\tilde{P}(\psi)| \leq \exp[-K_1\psi^2] \quad \psi \in (-\pi, \pi)$$

Secondly, we have that

$$\tilde{P}(\psi) = \exp\left[-\frac{\alpha(1-\alpha)}{2}\psi^2 + E(\psi)\right] \quad \psi \in (-\pi, \pi)$$

where there is a $K_2 > 0$ such that $|E(\psi)| \leq K_2|\psi|^3$ for all $|\psi| \leq 1$.

Proof. We first calculate that

$$|\tilde{P}(\psi)|^2 = |1 - \alpha + \alpha \cos \psi + i\alpha \sin \psi|^2 = (1 - \alpha + \alpha \cos \psi)^2 + \alpha^2 \sin^2 \psi$$

$$= (1 - \alpha)^2 + \alpha^2 + 2\alpha(1 - \alpha) \cos \psi = (1 - \alpha)^2 + 2\alpha(1 - \alpha) \cos \psi + \alpha^2 - 2\alpha(1 - \alpha)(1 - \cos \psi)$$

$$= 1 - 2\alpha(1 - \alpha)(1 - \cos \psi).$$

Since $-1 < \cos \psi < 1$ for all nonzero $\psi \in (-\pi, \pi)$, we have that $0 < 1 - \cos \psi < 1$ for all nonzero $\psi \in (-\pi, \pi)$. We also have that $0 < \alpha(1 - \alpha) \leq \frac{1}{4}$. Thus for all nonzero $\psi \in (-\pi, \pi)$, we have that

$$0 < 2\alpha(1 - \alpha)(1 - \cos \psi) < 4\alpha(1 - \alpha) \leq 1$$

and thus $0 < |\tilde{P}(\psi)| < 1$ for all nonzero $\psi \in (-\pi, \pi)$. Hence

$$K_1 \overset{\text{def}}{=} \sup_{\psi \in (-\pi, \pi) \setminus \{0\}} \frac{\ln \{1 - 2\alpha(1 - \alpha)(1 - \cos \psi)\}}{\psi^2}$$

is well-defined and $KK_1 \leq 0$. We have that

$$\lim_{|\psi| \to \pi} \frac{\ln \{1 - 2\alpha(1 - \alpha)(1 - \cos \psi)\}}{\psi^2} = \lim_{\psi \to 0} \frac{-2\alpha(1 - \alpha) \sin \psi}{2\psi} = -\alpha(1 - \alpha) < 0.$$ 

We can use l'Hôpital's rule to also see that

$$\lim_{|\psi| \to \pi} \frac{\ln \{1 - 2\alpha(1 - \alpha)(1 - \cos \psi)\}}{\psi^2} = \lim_{\psi \to 0} \frac{-2\alpha(1 - \alpha) \sin \psi}{2\psi} = -\alpha(1 - \alpha) < 0.$$ 

Note of course that $|\tilde{P}(0)| = 1 \leq \exp[-K_10^2]$. This proves the first claim.

To see the second claim, we want to use the principal branch of the logarithm. Note that if $\tilde{P}(\psi) \in \mathbb{R}_-$, then $\sin \psi = 0$. Thus $\tilde{P}(\psi) \not\in \mathbb{R}_-$ for all $\psi \in (-\pi, \pi)$. We can thus define

$$f(\psi) \overset{\text{def}}{=} \ln \tilde{P}(\psi) = \ln \{1 + \alpha (e^{i\psi} - 1)\} - i\psi \alpha$$

for all $\psi \in (-\pi, \pi)$, and calculate that

$$f'(\psi) = \frac{iae^{i\psi}}{1 + \alpha(e^{i\psi} - 1)} - i\alpha$$

$$f''(\psi) = \frac{-\alpha e^{i\psi}}{1 + \alpha(e^{i\psi} - 1)} = \frac{\alpha^2 e^{2i\psi}}{(1 + \alpha(e^{i\psi} - 1))^2}$$

Thus $f(0) = 0$, $f'(0) = 0$, and $f''(0) = -\alpha + \alpha^2 = -\alpha(1 - \alpha)$. This gives the second claim.

We can next connect things back together. For any $s \in S_N$, we have that

$$\hat{P}_N \left\{ \hat{X}_N = s \right\} = \frac{\sqrt{N}}{2\pi} \int_{\psi \in (-\pi \sqrt{N}, \pi \sqrt{N})} \mathcal{P}_N \left( \frac{\psi}{\sqrt{N}} \right) \exp \left[ -\frac{i}{\sqrt{N}} \psi \right] d\psi$$

$$= \frac{1}{2\pi \sqrt{N}} \int_{\psi \in (-\pi \sqrt{N}, \pi \sqrt{N})} \exp \left[ -\frac{\alpha(1 - \alpha)}{2} \psi^2 + N\mathcal{E} \left( \frac{\psi}{\sqrt{N}} \right) \right] \exp \left[ -\frac{i}{\sqrt{N}} \psi \right] d\psi.$$
1. Reference

Two common references for large deviations are [DZ98] and [FW98]. See also the references therein. Another nice source is [Var84].
Bibliography

