Large Deviation Principles
Spring 2010
Scott Robertson[*]

Chris Almost[†]

The course website contains links to the assignments and further reading.

Contents

[*] scottrob@andrew.cmu.edu
[†] cdalmost@cmu.edu

1 Introduction to large deviations principles

1.1 Motivation

Large deviation theory is concerned with the study of probabilities of rare events. We will be asymptotically computing these probabilities on an exponential scale. When we say "rare" we do not mean $\mathbb{P}[A] \approx 0$, rather we mean events for which $\frac{1}{n}\log\mathbb{P}_n[A]$ is moderately sized for large n, where the family of probability measures $\{\mathbb{P}_n\}_{n\geq 1}$ converges (perhaps to a point mass).

Suppose we have n stocks $S_1,\dots,S_n$ and we own \$1 of each at the beginning of the day. Assume the one day rates of return r_i are i.i.d. with mean zero. What is the probability your portfolio loss exceeds some given percentage, say 50%? Our wealth at the end of the day is $W_1 = \sum_{i=1}^n \frac{1}{S_i^0}S_i^1$, so the portfolio rate of return is $\frac{1}{n}\sum_i r_i$ (recall $W_0 = \$n$). We wish to compute $\mathbb{P}[\frac{1}{n}\sum_i r_i \leq 0.5]$. This is an example of a ruin probability.

Why $\frac{1}{n}\log\mathbb{P}_n[A]$ and not, say, $\frac{1}{n^2}\log\mathbb{P}_n[A]$ or $n^2\log\mathbb{P}_n[A]$? The answer is because it works for empirical averages of i.i.d. $N(0,1)$ random variables. It is well know that $\frac{1}{n}\sum_{i=1}^n Z_i \sim N(0,\frac{1}{n})$. It can be shown (see K&S p. 112)

$$\frac{a\sqrt{2\pi n}}{1+na^2}e^{-na^2/2} \leq \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n Z_i \geq a\right] = \int_a^\infty \sqrt{\frac{n}{2\pi}}e^{-nx^2/2}dx \leq \sqrt{\frac{2\pi}{na}}e^{-na^2/2}.$$

so $\frac{1}{n}\log\mathbb{P}_n[(a,\infty)] \to -a^2/2$. Since $\frac{1}{n}\log\mathbb{P}_n[A]$ converges to something meaningful for averages of $N(0,1)$, it makes sense to think it is the correct scaling for averages of other i.i.d. random variables, random walks, Brownian motion, some diffusions, some Markov processes, etc.

Suppose we may write $\mathbb{P}_n[A] = g_n(A)e^{-nI(A)}$, where $\frac{1}{n}\log g_n(A) \to 0$ uniformly for all A. What should $I(A)$ look like? What properties should it have? Assume for now that all $\mathbb{P}_n$ are absolutely continuous with respect to some reference measure $\mathbb{P}$. By taking "$A = dx$" we get $p_n(x) = g_n(x)e^{-nI(x)}$, so

$$\frac{1}{n}\log\mathbb{P}_n[A] = \frac{1}{n}\log\int_A e^{-nI(x)}\,\mathbb{P}[dx] + o(1)$$

If f is a non-negative, bounded, measurable function then

$$\left(\int_A f(x)^n\,\mathbb{P}[dx]\right)^{1/n} \xrightarrow{n\to\infty} \operatorname*{esssup}_{\mathbb{P}}\{f(x) : x\in A\}$$

Let $f(x) := e^{-I(x)}$ and combine the previous two equations to get

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[A] = \log\operatorname*{esssup}_{\mathbb{P}}\{e^{-I(x)} : x\in A\} = -\operatorname*{essinf}_{\mathbb{P}}\{I(x) : x\in A\}.$$

This is telling us that we should look for I such that we can write statements of the form

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[A] = -\inf_{x\in A} I(x).$$

The basic properties I should have include non-negativity, $\inf_{x\in\mathcal{X}} I(x) = 0$, and there should exist $x \in A$ which minimizes I over A whenever A is compact. In particular, I should be *lower semicontinuous*, i.e. the set $\{I \le \alpha\}$ is closed for each $\alpha \ge 0$.

1.1.1 Exercise. Show that, in a metric space, I is lower semicontinuous if and only if $\liminf_n I(x_n) \ge I(x)$ whenever $x_n \to x$.

SOLUTION: Assume that I is lower semicontinuous and let $x_n \to x$ be given. Let $\alpha := \liminf_n I(x_n)$. If $\alpha = \infty$ then there is nothing to prove. Assume that α is a finite number. Define $n_1 := 1$ and for each $k > 1$ define n_k to be the smallest index greater than n_{k-1} such that $I(x_{n_k}) \le \inf_{m>n_{k-1}} I(x_m) + 2^{-k}$. Then $x_{n_k} \to x$, since it is a subsequence of a convergent sequence, and by construction $I(x_{n_k}) \to \alpha$. For each $\varepsilon > 0$, eventually $I(x_k) \le \alpha + \varepsilon$, so since $\{I \le \alpha + \varepsilon\}$ is a closed set, it follows that $I(x) \le \alpha + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $I(x) \le \alpha$. If $\alpha = -\infty$ then it is easy to show that $I(x) = -\infty$.

Suppose I has the property that $\liminf_n I(x_n) \ge I(x)$ for all x and all sequences $x_n \to x$. Let $\alpha > 0$ be given and assume for contradiction that the set $\{I \le \alpha\}$ is not closed. Then there is a sequence $\{x_n\}_{n=1}^\infty \subseteq \{I \le \alpha\}$ such that $x_n \to x$ and $x \notin \{I \le \alpha\}$. But then $\alpha \ge \liminf_n I(x_n) \ge I(x) > \alpha$, a contradiction. ✠

1.2 Definition and basic properties

1.2.1 Definition. Let $(\mathcal{X}, \tau)$ be a topological space. $I : \mathcal{X} \to [0,\infty]$ is a *rate function* if it is lower semicontinuous. I is a *good rate function* if $\{I \le \alpha\}$ is compact for all $\alpha \ge 0$.

Remark. Some authors take only good rate functions to be rate functions.

Let I be a rate function. It seems as though it would be nice if

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n[A] = -\inf_{x\in A} I(x)$$

for all measurable A, but this is too restrictive for reasonable applications. For example, it is quite reasonable to have $\mathbb{P}_n[\{x\}] = 0$ for all n, for all $x \in \mathcal{X}$. But this implies that $I(x) = \infty$ everywhere, which is incompatible with the fact that $\inf_{x\in\mathcal{X}} I(x) = 0$ since $\mathbb{P}_n[\mathcal{X}] = 1$.

Recall that $\mathbb{P}_n \xrightarrow{(w)} \mathbb{P}$ does not mean $\lim_{n\to\infty} \mathbb{P}_n[A] = \mathbb{P}[A]$ for all A. Rather, it is equivalent to the pair of inequalities

$$\limsup_n \mathbb{P}_n[F] \le \mathbb{P}[F] \text{ for all } F \text{ closed, and}$$
$$\liminf_n \mathbb{P}_n[G] \ge \mathbb{P}[G] \text{ for all } G \text{ open.}$$

1.2.2 Definition. Let $(\mathscr{X}, \mathscr{B})$ be a measure space and $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ be a family of probability measures. We say that $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ satisfies a *large deviations principle* or *LDP* with (good) rate function I if, for all $\Gamma \in \mathscr{B}$,

$$-\inf_{x\in\Gamma^\circ} I(x) \le \liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon(\Gamma) \le \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon(\Gamma) \le -\inf_{x\in\overline{\Gamma}} I(x)$$

Remark.
(i) There is no reason to assume that $\mathscr{B}$ is the Borel σ-algebra of $\mathscr{X}$.
(ii) In most instances, but not all, $\mathscr{X}$ is a *Polish space* (i.e. a complete separable metric space) and the $\mathbb{P}_\varepsilon$ are Borel measures.
(iii) If all the open sets are measurable then we can write the LDP as

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon(F) \le -\inf_{x\in F} I(x) \text{ for all } F \text{ closed, and}$$
$$\liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon(G) \ge -\inf_{x\in G} I(x) \text{ for all } G \text{ open.}$$

The *full LDP* is the one stated in the definition (or equivalently, the one in the remark when we are in the Borel case). The *weak LDP* in the Borel case is a slight relaxation,

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon(F) \le -\inf_{x\in F} I(x) \text{ for all } F \text{ compact, and}$$
$$\liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon(G) \ge -\inf_{x\in G} I(x) \text{ for all } G \text{ open.}$$

That is to say, "closed" changes to "compact" for the weak LDP.

1.2.3 Exercise. Find a family satisfying a weak LDP but no full LDP.

SOLUTION: From D&Z, page 7, let $\mathbb{P}_\varepsilon := \delta_{1/\varepsilon}$. Then the family $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ satisfies the weak LDP for the *good* rate function $I \equiv \infty$. (The lower bound for open sets is trivial and the upper bound for compact sets follows because eventually $1/\varepsilon$ escapes any fixed compact set. We define $\log 0 := -\infty$.) This family does not satisfy a full LDP for any rate function because $\mathbb{P}_\varepsilon[\{x\}]$ is eventually zero for every x, so the only possible choice of rate function is $I \equiv \infty$, but $\mathbb{P}_\varepsilon[\mathbb{R}] = 1$ implies $\inf_{x\in\mathbb{R}} I(x) = 0$, a contradiction.

This example also shows that having a weak LDP with a good rate function does not imply a full LDP. ✠

1.2.4 Definition. A family $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ is *exponentially tight* if, for every $\alpha > 0$, there is a compact set K_α such that $\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}_\varepsilon[K_\alpha^c] < -\alpha$.

Recall that $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ is a *tight family* if for all $\alpha > 0$ there is a compact set K_α such that $\sup_{\varepsilon>0} \mathbb{P}_\varepsilon[K_\alpha^c] \le \alpha$.

1.2.5 Exercises.
(i) Show that tightness does not imply exponential tightness.

(ii) Show that, for sequences, exponential tightness implies tightness.

SOLUTION:

(i) Let $\mathbb{P}_\varepsilon := \delta_\varepsilon$ for $0 < \varepsilon \le 1$ and $\delta_{1/(1-\varepsilon)}$ for $\varepsilon > 1$. Then the family $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ is not tight (because $1/(1-\varepsilon)$ escapes any fixed compact set for some $\varepsilon > 1$) but is exponentially tight (taking $K_\alpha = [0,1]$ for all $\alpha > 0$ suffices). This example feels cheap.

(ii) Suppose the family $\{\mathbb{P}_n\}_{n\ge1}$ is exponentially tight. Let K be the compact set corresponding to $\alpha = 2$ in the definition of exponential tightness. Then $\limsup_n \frac{1}{n}\log\mathbb{P}_n[K^c] < -2$. It follows that there is N such that, for $n \ge N$, $\frac{1}{n}\log\mathbb{P}_n[K^c] < -1$, i.e. $\mathbb{P}_n[K^c] < e^{-n}$. Let $\alpha > 0$ be given. Choose $M \ge N$ such that $e^{-M} < \alpha$ and a compact set K' such that the (finitely many) probability measures $\{\mathbb{P}_n\}_{n=1}^{M-1}$ all satisfy $\mathbb{P}_n[K'^c] \le \alpha$. The set $K_\alpha := K' \cup K$ witnesses tightness. ✠

Warning: Goodness of I and an LDP do not together imply exponential tightness. It is tempting to note that $K_\alpha := \{I \le \alpha\}$ is compact and write

$$\limsup_{\varepsilon\to0} \varepsilon \log \mathbb{P}_\varepsilon[K_\alpha^c] \le -\inf_{x\in\overline{K_\alpha^c}} I(x),$$

but there is no good reason to assume that the right hand side is at most $-\alpha$, because the closure may introduce some undesirables.

1.2.6 Exercise (4.1.10 in D&Z). Show that if $\mathscr{X}$ is Polish, the $\mathbb{P}_\varepsilon$ are all Borel, and they satisfy an LDP with a good rate function I then, $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ is exponential tight.

1.2.7 Lemma. *Let $a_\varepsilon^i > 0$ for $i = 1,\dots,N$ and $\varepsilon > 0$ be real numbers. Then*

$$\limsup_{\varepsilon\to0} \varepsilon \log\left(\sum_{i=1}^N a_\varepsilon^i\right) = \max_{i=1}^N \limsup_{\varepsilon\to0} \varepsilon \log a_\varepsilon^i$$

PROOF: This follows from the observation that

$$\max_{i=1}^N a_\varepsilon^i \le \sum_{i=1}^N a_\varepsilon^i \le N \max_{i=1}^N a_\varepsilon^i$$

and because the max is being taken over finitely many terms. □

1.2.8 Proposition. *If an exponentially tight family $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ satisfies a weak LDP then it satisfies a full LDP.*

PROOF: There is nothing to prove for open sets since the statements of the weak and full LDP agree. Let a closed set F be given. Let $\alpha > 0$ and K_α be the corresponding compact set from exponential tightness. By well-known properties of measures,

$$\mathbb{P}_\varepsilon[F] = \mathbb{P}_\varepsilon[F \cap K_\alpha] + \mathbb{P}_\varepsilon[F \cap K_\alpha^c] \leq \mathbb{P}_\varepsilon[F \cap K_\alpha] + \mathbb{P}_\varepsilon[K_\alpha^c].$$

Now, $F \cap K_\alpha$ is compact, so by the weak LDP,

$$\limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon[F \cap K_\alpha] \leq -\inf_{x \in F \cap K_\alpha} I(x) \leq -\inf_{x \in F} I(x),$$

and $\limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon[K_\alpha^c] < -\alpha$ by definition of K_α. From this and from 1.2.7,

$$\limsup_{\varepsilon \to 0} \varepsilon \log(\mathbb{P}_\varepsilon[F \cap K_\alpha] + P_\varepsilon[K_\alpha^c]) \leq \max(-\inf_{x \in F} I(x), -\alpha)$$

Taking $\alpha \to \infty$ completes the proof. □

Remark. By 1.2.8, the family given in the solution to 1.2.3 is not exponentially tight.

1.2.9 Lemma. *If an exponentially tight family $\{\mathbb{P}_\varepsilon\}_{\varepsilon > 0}$ satisfies the LDP lower bound for open sets, for a rate function I, then I is a good rate function.*

PROOF: Let $\alpha > 0$ be given and K_α be the corresponding set from exponential tightness. Note that K_α^c is open, so

$$-\alpha > \limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon[K_\alpha^c] \geq \liminf_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon[K_\alpha^c] \geq -\inf_{x \in K_\alpha^c} I(x)$$

so $\inf_{x \in K_\alpha^c} I(x) > \alpha$. Therefore $\{I \leq \alpha\} \subseteq K_\alpha$. Since I is a rate function, $\{I \leq \alpha\}$ is closed, and a closed subset of a compact set is compact, so I is a good rate function. □

1.3 Analogy with weak convergence

1.3.1 Example (Parametric statistics).
Let $\{\xi_n\}_{n=1}^\infty$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. We have the Weak Law of Large Numbers (Bernoulli, 1713),

$$\eta_n := \frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{(p)} \mu.$$

A long while later the Strong Law of Large Numbers was proved, which shows that the convergence is $\mathbb{P}$-a.s. Parametric statistics is concerned with estimating μ. Knowing that the estimators converge to μ is important, but obtaining bounds on the error is also very important.

(i) Confidence intervals are one way of quantifying the error. For given n, how large is $\mathbb{P}[|\eta_n - \mu| \geq a]$?
(ii) Loss functions are another way of quantifying error. A *loss function* is a non-negative radially symmetric measurable function, e.g. $L(x) = x^2, |x|, \mathbf{1}_{|x|\geq a}$, etc. The error can be taken to be $\mathbb{E}[L(\eta_n - \mu)]$. Note that this approach extends the confidence interval approach by considering functions other than indicator functions.

The Central Limit Theorem states that $\sqrt{n}(\eta_n - \mu) \xrightarrow{(w)} \eta \sim N(0,\sigma^2)$. It allows us now to make more precise statements about the convergence.
(i) $\mathbb{P}[|\eta_n - \mu| \geq a/\sqrt{n}] \to \mathbb{P}[|\eta| \geq a]$
(ii) $\mathbb{E}[L(\sqrt{n}(\eta_n - \mu))] \to \mathbb{E}[L(\eta)]$.

1.3.2 Definition. A function $\Pi : E \to [0,1]$ is said to be an *idempotent probability* if $\sup_{x\in E} \Pi(x) = 1$. Define $\Pi[A] := \sup_{x\in A} \Pi(A)$ for $A \subseteq E$. Π is *good*, or *l.s.c.*, if $\Pi(F_n) \to \Pi(\bigcap_{n\geq 1} F_n)$ whenever $\{F_n\}_{n=1}^\infty$ is a decreasing sequence of closed sets.

In the theory of large deviations, we make statements of the form

$$\mathrm{Law}(\eta_n - \mu) \xrightarrow{LD} I \text{ or } \Pi,$$

where I is a rate function and Π is an idempotent probability related according to the relation $\Pi(x) = \exp(-I(x))$. The corresponding statements of convergence may be stated as follows.
(i) $(\mathbb{P}[|\eta_n - \mu| \geq a])^{1/n} \to \exp(-\inf_{|x|\geq a} I(x)) = \Pi[|x| \geq a]$
(ii) $(\mathbb{E}[(L(\eta_n - \mu))^n])^{1/n} \to \sup_{x\in\mathbb{R}} L(x)\Pi(x)$.

1.3.3 Example (Empirical distributions).
Let $\{\xi_n\}_{n=1}^\infty$ be a sequence of i.i.d. random variables distributed as Lebesgue measure on $[0,1]$. The *empirical distribution function* is

$$F_n(x) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\xi_k \leq x}.$$

There is a Law of Large Numbers for this setting, proved by Cantelli,

$$\sup_{x\in[0,1]} |F_n(x) - F(x)| \xrightarrow{(p)} 0.$$

A strong version, that the convergence is a.s., was proved soon after by Glivenko. A Central Limit Theorem was proved by Kolmogorov, in the form

$$\sqrt{n} \sup_{x\in[0,1]} |F_n(x) - F(x)| \xrightarrow{(w)} \max_{t\in[0,1]} |X_t|,$$

where X is Brownian bridge, the solution to the SDE $dX_t = -\frac{X_t}{1-t}dt + dB_t, X_0 = 0$. There are other central limit theorems, but their application is more delicate.

$$\sqrt{n}\int_0^1 |F_n(x) - F(x)|^2 dx \xrightarrow{(w)} \int_0^1 |X_s|^2 ds.$$

Doob observed that as stochastic processes (with x as "time")

$$\sqrt{n}(F_n - F) \xrightarrow{\text{(w)}} X.$$

The analog in Large Deviations will be Sanov's theorem. Vapnik-Chervonenkis theory gives conditions under which

$$\sup_{A \in \mathscr{A}} |\mathbb{P}_n[A] - \mathbb{P}[A]| \to 0$$

in terms of the "information content" of the family of sets $\mathscr{A}$.

Let (E, d) be a Polish space, e.g. $\mathbb{R}^n$, $C[0,1]$, or $D[0,1]$. Recall that $D[0,1]$ is the collection of càdlàg functions with domain $[0,1]$, equipped with the metric

$$\rho(X,Y) = \inf_{\lambda \in \Lambda} \left\{ \sup_{x \in [0,1]} |(X(\lambda(t)) - Y(t)| + \sup_{s,t} \left| \log \frac{|\lambda(t) - \lambda(s)|}{|t-s|} \right| \right\}$$

where Λ is the collection of increasing functions $\lambda : [0,1] \to [0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$. This is the *Skorokhod metric*. Under the more natural looking, equivalent, metric

$$\rho'(X,Y) = \inf_{\lambda \in \Lambda} \left\{ \sup_{x \in [0,1]} |(X(\lambda(t)) - Y(t)| + \sup_{t \in [0,1]} |\lambda(t) - t| \right\}$$

the space fails to be complete. Another important space is $C_b(E)$, the space of bounded continuous functions on a Polish space (E, d).

1.3.4 Definition. A sequence of measures $\{\mathbb{P}_n\}_{n=1}^{\infty}$ on (E, d) converges weakly to $\mathbb{P}$ if $\mathbb{E}_n[f] \to \mathbb{E}[f]$ for all $f \in C_b(E)$.

Remark. The function $f : D[0,1] \to \mathbb{R}$ defined by $f(X) := X(t_0)$ is *not* a continuous function for the Skorokhod topology. It is, of course, continuous for the uniform metric, but with this topology $D[0,1]$ is not separable. We will see that there is a way to get around this, depending on the limiting measure.

1.3.5 Definition. A function f is l.s.c. (resp. u.s.c.) if there is a sequence $\{f_n\}_{n=1}^{\infty}$ of continuous functions such that, for all $x \in E$, $f(x) = \sup_n f_n(x)$ (resp. $f(x) = \inf_n f_n(x)$).

Remark. The sequences in the definition may be taken to be increasing (resp. decreasing) since the max of finitely many continuous functions is itself continuous.

1.3.6 Lemma. *Suppose* $\mathbb{P}_n \xrightarrow{(w)} \mathbb{P}$.

(i) For all bounded l.s.c. functions f, $\liminf_{n\to\infty} \mathbb{E}_n[f] \geq \mathbb{E}[f]$.

(ii) For all open sets G, $\liminf_{n\to\infty} \mathbb{P}_n[G] \geq \mathbb{P}[G]$.

(iii) For all bounded u.s.c. functions f, $\limsup_{n\to\infty} \mathbb{E}_n[f] \leq \mathbb{E}[f]$.

(iv) *For all closed sets F,* $\limsup_{n\to\infty} \mathbb{P}_n[F] \le \mathbb{P}[F]$.

PROOF: Let f be a l.s.c. function and let $\{f_m\}_{m=1}^\infty$ be an increasing sequence of continuous functions such that, for all $x \in E$,

$$f(x) = \sup_m f_m(x) = \lim_{m\to\infty} f_m(x).$$

For each n and m, $\mathbb{E}_n[f] \ge \mathbb{E}_n[f_m]$, so

$$\liminf_{n\to\infty} \mathbb{E}_n[f] \ge \liminf_{n\to\infty} \mathbb{E}_n[f_m] = \mathbb{E}[f_m].$$

Therefore, by Fatou's Lemma,

$$\liminf_{n\to\infty} \mathbb{E}_n[f] \ge \liminf_{m\to\infty} \mathbb{E}[f_m] \ge \mathbb{E}[f].$$

The proof for u.s.c. functions is similar. The last two parts follow because $\mathbf{1}_G$ is l.s.c. when G is open and $\mathbf{1}_F$ is u.s.c when F is closed. □

1.3.7 Definition. Let f be a function on (E,d) and define

$$\underline{f}(x) := \sup\{g(x) : g \le f, g \in C(E)\} \text{ and } \overline{f}(x) := \sup\{g(x) : g \ge f, g \in C(E)\}.$$

Then f is said to be $\mathbb{P}$-*continuous* if $\mathbb{E}[|\overline{f} - \underline{f}|] = 0$, or equivalently,

$$\mathbb{P}[\{x : f \text{ is continuous at } x\}] = 1.$$

1.3.8 Lemma. *If $\mathbb{P}_n \xrightarrow{(w)} \mathbb{P}$ and f is a bounded and $\mathbb{P}$-continuous function on (E,d) then $\mathbb{E}_n[f] \to \mathbb{E}[f]$.*

PROOF: Note that $\overline{f}$ and $\underline{f}$ are u.s.c. and l.s.c., respectively, and they are equal $\mathbb{P}$-a.s., so

$$\liminf_{n\to\infty} \mathbb{E}_n[f] \ge \liminf_{n\to\infty} \mathbb{E}_n[\underline{f}] \ge \mathbb{E}[\underline{f}] = \mathbb{E}[\overline{f}] \ge \limsup_{n\to\infty} \mathbb{E}_n[\overline{f}] \ge \limsup_{n\to\infty} \mathbb{E}_n[f].$$

□

1.3.9 Definition. A sequence of measures $\{\mathbb{P}_n\}_{n=1}^\infty$ on (E,d) converges in the sense of large deviations to Π if $(\mathbb{E}_n[f^n])^{1/n} \to \sup_{x\in E} f(x)\Pi(x)$ for all $f \in C_b^+(E)$. Analogously, a function f is Π-continuous if

$$\sup_{x\in E} \big(\overline{f}(x) - \underline{f}(x)\big)\Pi(x) = 0.$$

1.3.10 Exercise. Look up Ulam's theorem, that probability measures on Polish spaces are (individually) tight, and come up with an LDP analog. Then try for an analog of Prohorov's theorem. Here sequentially precompactness corresponds to exponential tightness.

2 Cramér's Theorem

2.1 Cramér's theorem in $\mathbb{R}^d$

Cramér's theorem gives conditions under which the empirical average of i.i.d. random variables satisfy an LDP. Let $\{X_i\}_{i=1}^{\infty}$ be $\mathbb{R}^d$-valued i.i.d. random vectors with common distribution μ, and let $S_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ be the empirical average. Let μ_n be the distribution of S_n, a probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. The question is whether the family $\{\mu_n\}_{n=1}^{\infty}$ satisfies an LDP, and if so, what is the rate function?

We expect some sort of converge in many situations. Letting $\overline{x} := \mathbb{E}[X_1]$ and $\sigma^2 := \mathrm{Var}(X_1)$, if both of these quantities are finite then $\mathbb{E}[|S_n - \overline{x}|^2] \to 0$ as $n \to \infty$, (this is a version of the law of large numbers) so $S_n \xrightarrow{(p)} \overline{x}$. Yet otherwise said, $\mu_n \xrightarrow{(w)} \delta_{\overline{x}}$.

2.1.1 Definition. The *cumulant generating function* is the logarithm of the moment generating function,

$$\Lambda(\lambda) := \log \mathbb{E}[\exp(\lambda, X_1))].$$

Define $\mathscr{D}_\Lambda := \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$, the *domain* of Λ. The *Fenchel-Legendre transform* of Λ is

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\}.$$

Define $\mathscr{D}_{\Lambda^*} := \{x \in \mathbb{R}^d : \Lambda^*(x) < \infty\}$, the domain of Λ^*.

2.1.2 Examples.

(i) If $X_1 \sim \delta_{x_0}$ then $\Lambda(\lambda) = (\lambda, x_0)$ so $\Lambda^*(x) = 0$ if $x = x_0$ and ∞ otherwise.

(ii) If $X_1 \sim N(0,1)$ then $\Lambda(\lambda) = \frac{1}{2}|\lambda|^2$ and $\Lambda^*(x) = \frac{1}{2}|x|^2$. As usual, the normal distribution is amazingly well-behaved.

(iii) If $X_1 \sim \mathrm{Bernoulli}(p)$ then $\Lambda(\lambda) = \log(pe^{\lambda} + (1-p))$ and

$$\Lambda^*(x) = \begin{cases} x \log \frac{x}{p} + (1-x)\log\frac{1-x}{1-p} & x \in [0,1] \\ \infty & \text{otherwise} \end{cases}$$

Note that $\lim_{x\to 0^+} \Lambda^*(x)$ and $\lim_{x \to 1^-} \Lambda^*(x)$ are both finite, so Λ^* jumps up to ∞ at those points.

2.1.3 Proposition (Properties of Λ).

Suppose that Λ is the cumulant generating function of a random variable X.

(i) $\Lambda(0) = 0$.

(ii) $\Lambda(\lambda) > -\infty$ *for all* λ.

(iii) Hölder's inequality implies that Λ is convex.

(iv) The dominated convergence theorem implies that $\Lambda \in C^\infty(\mathscr{D}_\Lambda^\circ)$.

(v) $\nabla\Lambda(\lambda) = e^{-\Lambda(\lambda)}\mathbb{E}[Xe^{(\lambda,X)}]$ *for* $\lambda \in \mathscr{D}_\Lambda^\circ$.

2.1.4 Proposition (Properties of Λ^*).
Suppose that Λ^* is the Fenchel-Legendre transform of the cumulant generating function of a random variable X.
(i) $\Lambda^*(x) \geq 0$ *for all* x.
(ii) $\Lambda^*(\mathbb{E}[X]) = 0$.
(iii) Λ^* *is convex.*
(iv) Λ^* *is lower semicontinuous, i.e.* Λ^* *is a rate function.*
(Properties (iii) and (iv) hold for all Fenchel-Legendre transforms.)

PROOF:
(i) Consider $\lambda = 0$.
(ii) By Jensen's inequality, $\log \mathbb{E}[e^{(\lambda,X)}] \geq \log e^{(\lambda,\mathbb{E}[X])} = (\lambda, \mathbb{E}[X])$ for all λ, so

$$0 \leq \Lambda^*(\mathbb{E}[X]) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, \overline{x}) - \Lambda(\lambda)\} \leq 0.$$

(iii) Λ^* is a supremum of affine (hence convex) functions.
(iv) Let $x_n \to x$. For all $\lambda \in \mathbb{R}^d$ and all $n \geq 1$, $\Lambda^*(x_n) \geq (\lambda, x_n) - \Lambda(\lambda)$, so

$$\begin{aligned} \liminf_{n\to\infty} \Lambda^*(x_n) &\geq \sup_{\lambda \in \mathbb{R}^d} \{\liminf_{n\to\infty} (\lambda, x_n) - \Lambda(\lambda)\} \\ &= \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\} = \Lambda^*(x). \end{aligned}$$ □

2.1.5 Exercise. Show that if $0 \in \mathscr{D}_\Lambda^\circ$ then Λ^* is a good rate function.
(Hint: consider $\Lambda^*(x)/|x|$.)

SOLUTION: If $0 \in \mathscr{D}_\lambda^\circ$ then there is a small $r > 0$ such that $\Lambda(\lambda) < \infty$ for all $\lambda \in B(0, 2r)$. Since $\Lambda \in C^\infty(\mathscr{D}_\Lambda^\circ)$, $M := \max_{\lambda \in \overline{B}(0,r)} \Lambda(\lambda) < \infty$. For all $x \in \mathbb{R}^d$,

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} (\lambda, x) - \Lambda(\lambda) \geq \left(\frac{r}{|x|} x, x \right) - \Lambda\left(\frac{r}{|x|} x \right) \geq r|x| - M.$$

It follows that the closed set $\{\Lambda^* \leq \alpha\}$ is contained within the ball $B(0, \frac{1}{r}(\alpha + M))$, so it is compact. Therefore Λ^* is a good rate function. ✠

2.1.6 Theorem (Cramér).
If $0 \in \mathscr{D}_\Lambda^\circ$ *then* $\{\mu_n\}_{n=1}^\infty$ *satisfies the LDP with good rate function* Λ^*.

Remark. If $\mathscr{D}_\Lambda^\circ \neq \varnothing$ then $0 \in \mathscr{D}_\Lambda^\circ$.

PROOF: To simplify the proof we make the following assumptions.
(A1) $\mathscr{D}_\Lambda = \mathbb{R}^d$.
(A2) If $\Lambda^*(x) < \infty$ then there is λ_x which maximizes $(\lambda, x) - \Lambda(\lambda)$ over $\lambda \in \mathbb{R}^d$, and in particular $x = \nabla\Lambda(\lambda_x)$. (Note that this assumption fails for the Bernoulli distribution at $x = 0$ and at $x = 1$.)

The first step in the proof is to prove the upper bound for compact sets,

$$\limsup_n \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x\in\Gamma} \Lambda^*(x).$$

This step will not use either assumption. To this end, let $\Gamma \subseteq \mathbb{R}^d$ be a compact set. Let $\delta > 0$ be a small number and define $I^\delta(x) := \min\{\Lambda^*(x) - \delta, 1/\delta\}$.

2.1.7 Exercise. Show $\lim_{\delta\to 0} \inf_{x\in\Gamma} I^\delta(x) = \inf_{x\in\Gamma} \Lambda^*(x)$.

Let $x \in \Gamma$ be given. For any $r > 0$ and $\lambda \in \mathbb{R}^d$, by Chebyshev's inequality,

$$\begin{aligned}
\mathbb{E}[e^{n\Lambda(\lambda)}] &= \mathbb{E}[e^{n(\lambda,S_n)}] && \text{by independence}\\
&\ge \mathbb{E}[e^{n(\lambda,S_n)}\mathbf{1}_{B(x,r)}(S_n)]\\
&\ge e^{\inf_{y\in B(x,r)} n(\lambda,y)}\mu_n(B(x,r)) && \text{by Chebyshev}\\
\Lambda(\lambda) - \inf_{y\in B(x,r)}(\lambda,y) &\ge \frac{1}{n}\log\mu_n(B(x,r))
\end{aligned}$$

Choose $\lambda_x \in \mathbb{R}^d$ such that $(\lambda_x, x) - \Lambda(\lambda_x) \ge I^\delta(x)$. Since $(\lambda_x, \cdot)$ is continuous, we can take r_x small enough so that $\inf_{y\in B(x,r_x)}(\lambda_x, y) \ge (\lambda_x, x) - \delta$, namely, $r_x < \delta/|\lambda_x|$. Therefore

$$\frac{1}{n}\log\mu_n(B(x,r_x)) \le \Lambda(\lambda_x) - (\lambda_x, x) + \delta \le \delta - I^\delta(x).$$

Now, $\Gamma \subseteq \bigcup_{x\in\Gamma} B(x, r_x)$, so there are $x_1, \dots, x_N$ such that $\Gamma \subseteq \bigcup_{i=1}^N B(x_i, r_i)$. Therefore

$$\begin{aligned}
\frac{1}{n}\log\mu_n(\Gamma) &\le \frac{1}{n}\log\sum_{i=1}^N \mu_n(B(x_i,r_i))\\
\limsup_n \frac{1}{n}\log\mu_n(\Gamma) &\le \max_{i=1}^N \limsup_n \frac{1}{n}\log\mu_n(B(x_i,r_i)) && 1.2.7\\
&\le \delta - \min_{i=1}^N I^\delta(x_i) \le \delta - \inf_{x\in\Gamma} I^\delta(x)
\end{aligned}$$

Take $\delta \to 0$ to complete the proof of the upper bound.

The second step in the proof is the lower bound for open sets,

$$\liminf_n \frac{1}{n}\log\mu_n(G) \ge -\inf_{x\in G}\Lambda^*(x).$$

This step will use both assumptions. To this end, let $G \subseteq \mathbb{R}^d$ be an open set. The basic idea is to put $\overline{x}$ into G so that $\mu_n(G) \to 1$, i.e. change measure so that the event G is not rare.

Since G is open it suffices to prove, for all $x \in \mathbb{R}^d$ and for all $\delta > 0$ small, that

$$\liminf_n \frac{1}{n}\log\mu_n(B(x,\delta)) \ge -\Lambda^*(x).$$

Indeed, for all $x \in G$, there is $\delta > 0$ such that $B(x,\delta) \subseteq G$, so if we have the above inequality then

$$\liminf_n \frac{1}{n} \log \mu_n(G) \geq \liminf_n \frac{1}{n} \log \mu_n(B(x,\delta)) \geq -\Lambda^*(x)$$

for each $x \in G$, so $\liminf_n \frac{1}{n} \log \mu_n(G) \geq -\inf_{x \in G} \Lambda^*(x)$.

Let $x \in \mathbb{R}^d$ be given. If $\Lambda^*(x) = \infty$ then there is nothing to prove, so assume that $\Lambda^*(x) < \infty$. By assumption (A2) there is λ_x such that $x = \nabla\Lambda(\lambda_x)$ and $\Lambda^*(x) = (\lambda_x, x) - \Lambda(\lambda_x)$. Define a probability measure $\tilde{\mu}$ on $\mathbb{R}^d$ via its Radon-Nikodym derivative with respect to μ,

$$\frac{d\tilde{\mu}}{d\mu}(z) := e^{(\lambda_x, z) - \Lambda(\lambda_x)}.$$

By (A1), $\tilde{\mu}$ is a probability measure since $\lambda_x \in \mathscr{D}_\Lambda^\circ$, i.e. $\Lambda(\lambda_x) < \infty$ and

$$\int_{\mathbb{R}^d} \tilde{\mu}(dz) = \int_{\mathbb{R}^d} e^{(\lambda_x, z) - \Lambda(\lambda_x)} \mu(dz) = e^{-\Lambda(\lambda_x)} \int_{\mathbb{R}^d} e^{(\lambda_x, z)} \mu(dz) = 1.$$

We have

$$\begin{aligned} \mathbb{E}^{\tilde{\mu}}[X] &= \int_{\mathbb{R}^d} z e^{(\lambda_x, z) - \Lambda(\lambda_x)} \mu(dz) \\ &= e^{-\Lambda(\lambda_x)} \int_{\mathbb{R}^d} z e^{(\lambda_x, z)} \mu(dz) = \nabla\Lambda(\lambda_x) = x \end{aligned}$$

by the formula for the gradient of Λ. The law $\tilde{\mu}_n$ of S_n may be transformed in a compatible way, so that

$$\frac{d\tilde{\mu}_n}{d\mu_n} = e^{n(\lambda, z) - n\Lambda(\lambda)},$$

via the formula

$$\mathbb{P}^{\tilde{\mu}_n}(S_n \in A) := \int_{\frac{1}{n}\sum_{i=1}^n x_i \in A} \tilde{\mu}(dx_1) \cdots \tilde{\mu}(dx_n).$$

Therefore, for any $\delta > 0$ and any η with $\delta > \eta > 0$,

$$\begin{aligned} \frac{1}{n} \log \mu_n(B(x,\delta)) &= \frac{1}{n} \log \int_{B(x,\eta)} e^{-n(\lambda_x, z) + n\Lambda(\lambda_x)} \tilde{\mu}_n(dz) \\ &= \Lambda(\lambda) - (\lambda_x, x) + \frac{1}{n} \log \int_{B(x,\eta)} e^{n(\lambda_x, x - z)} \tilde{\mu}_n(dz) \\ &\geq -\Lambda^*(x) - \eta|\lambda| + \frac{1}{n} \log \tilde{\mu}_n(B(x,\eta)) \end{aligned}$$

since $(\lambda, x - z) \geq -\eta|\lambda|$ over $B(x,\eta)$ by Cauchy-Schwarz. By the law of large numbers, $\lim_{n\to\infty} \mu_n(B(\overline{x},\delta)) = 1$, so $\lim_{n\to\infty} \tilde{\mu}_n(B(x,\delta)) = 1$ by construction. Therefore

$$\liminf_n \frac{1}{n} \log \mu_n(B(x,\delta)) \geq -\Lambda^*(x) - \eta|\lambda|.$$

Taking $\eta \to 0$ completes the proof of the lower bound.

We have now shown that $\{\mu_n\}_{n=1}^\infty$ satisfies the weak LDP with good rate function Λ^*. The third step is the proof is to show that $\{\mu_n\}_{n=1}^\infty$ is exponentially tight. The fact that $0 \in \mathscr{D}_\Lambda^\circ$ is used here. (To be added.) □

2.2 Sanov's theorem

Let $\{Y_i\}_{i=1}^\infty$ be real-valued i.i.d. random variables with common distribution function F. The *empirical distribution function* is defined for $t \in \mathbb{R}$ to be

$$F_n(t) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty,t]}(Y_i).$$

The theorem we will prove today is as follows. Assume that the Y_i's take values in a finite state space $\Sigma = \{a_1,\dots,a_N\}$, with common probability mass function μ. Assume without loss of generality that $\mu(a_j) > 0$ for $j = 1,\dots,N$. Set

$$L_n^Y := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i},$$

the empirical probability mass function. By the Strong Law of Large Numbers,

$$L_n^Y[\{a_j\}] = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{a_j\}}(Y_i) \xrightarrow{a.s.} \mu(a_j)$$

Let $X^i = (\mathbf{1}_{\{a_1\}}(Y_i),\dots,\mathbf{1}_{\{a_N\}}(Y_i))$, the natural embedding of δ_{Y_i} into the probability simplex, so that

$$\frac{1}{n} \sum_{i=1}^n X^i = \frac{1}{n} \sum_{i=1}^n (\mathbf{1}_{\{a_1\}}(Y_i),\dots,\mathbf{1}_{\{a_N\}}(Y_i)) = (L_n^Y[\{a_1\}],\dots,L_n^Y[\{a_n\}])$$

Note that $\{X^i\}_{i=1}^\infty$ is a sequence of i.i.d., bounded, $\mathbb{R}^N$-valued random vectors, so the cumulant generating function exists for all λ and

$$\Lambda(\lambda) = \log \mathbb{E}[e^{(\lambda,X^1)}] = \log \sum_{j=1}^N e^{\lambda_j} \mu(a_j).$$

Calculating Λ^* is an exercise in calculus, and is on the homework.

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^N} (\lambda, x) - \log \sum_{j=1}^N e^{\lambda_j} \mu(a_j).$$

If $x \notin \Delta^N := \{x \in \mathbb{R}^N_+ : \sum_i x_i = 1\}$ then we can appeal to the LDP to show that $\Lambda^*(x) = \infty$. There is $r > 0$ such that $B(x,r) \cap \Delta^N = \varnothing$. By the LDP, $-\inf_{B(x,r)} \Lambda^*(x) \le -\infty$ (so $\Lambda^*(x) = \infty$) since $\mu_n(B(x,r)) = 0$. If $x \in \Delta^N$ then the first order conditions require

$$x_\ell = \frac{e^{\lambda_\ell}\mu(a_\ell)}{\sum_{j=1}^N e^{\lambda_j}\mu(a_j)}.$$

If $x_\ell > 0$ then $\lambda_\ell^* = \log(x_\ell/\mu(a_\ell))$ is the optimizer. If $x_\ell = 0$ then taking $\lambda_\ell \to -\infty$ gives a sequence converging to the optimum. Therefore

$$\Lambda^*(x) = \sum_{j=1}^N x_j \log \frac{x_j}{\mu(a_j)} = H(x|\mu).$$

2.3 Sampling a Brownian motion

Let W be a Brownian motion on $[0,T]$ and $\{0 = t_0 < \cdots < t_N = T\}$ be a partition of the interval. Define $X := (W_{t_1}, \dots, W_{t_N})$, an $\mathbb{R}^N$-valued random variable. Let $\{X^i\}_{i=1}^\infty$ be independent samples of X, and let

$$S_n := \frac{1}{n}\sum_{i=1}^n X^i \sim \frac{1}{\sqrt{n}}(W_{t_1}, \dots, W_{t_N}).$$

(Show this, and think of an LDP for $Z_n = \frac{1}{\sqrt{n}} W$.) Then the cumulant generating function exists,

$$\begin{aligned}
\Lambda(\lambda) &= \log \mathbb{E}[e^{(\lambda, X^1)}] \\
&= \log \mathbb{E}[e^{\sum_{j=1}^N \lambda_j W_{t_j}}] \\
&= \log \mathbb{E}[e^{\sum_{j=1}^N \theta_j (W_{t_j} - W_{t_{j-1}})}] \qquad \text{where } \theta_j := \sum_{i=j}^N \lambda_i \\
&= \frac{1}{2}\sum_{j=1}^N \theta_j^2 (t_j - t_{j-1}).
\end{aligned}$$

Finding Λ^* is a problem in quadratic programming.

$$\begin{aligned}
\Lambda^*(x) &= \sup_\theta \sum_{j=1}^N x_j(\theta_j - \theta_{j+1}) + x_N\theta_N - \frac{1}{2}\sum_{j=1}^N \theta_j^2 (t_j - t_{j-1}) \\
&= \frac{1}{2}\sum_{j=1}^N \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}} \\
&= \frac{1}{2}\sum_{j=1}^N \frac{(x_j - x_{j-1})^2}{(t_j - t_{j-1})^2}(t_j - t_{j-1})
\end{aligned}$$

optimized at $\theta_j^* = (x_j - x_{j-1})/(t_j - t_{j-1})$ (with $x_0 := 0$). The reason for writing Λ^* in the form of the last line will become clear in a moment. We would like to take the limit over partitions. If it is mathematically justifiable to do this then notice the following.

(i) X corresponds to a random path $x \in \mathbb{R}^{[0,T]}$.
(ii) $\Lambda^*(x) = \infty$ if $x(0) \neq 0$ (taking $t_1 \to 0$).
(iii) If $x \notin AC[0,T]$ then $\Lambda^*(x) \to \infty$ (by definition of absolutely continuous).
(iv) If $x \in AC[0,T]$ then $\Lambda^*(x) \to \frac{1}{2}\int_0^T \dot{x}_t^2 dt$.

But what does it all mean?

3 Gärtner-Ellis theorem

The idea now is to forget how S_n was constructed, but keep some of the nice properties and asymptotic conclusions of the cumulant generating function that independence allowed us to make, and see where that gets us. Recall that independence and Chebyshev's inequality applied to $x \mapsto e^{n(\lambda,x)}$ give

$$\begin{aligned} \mu_n(B(x,r)) &\leq e^{-n\inf_y(\lambda,y)}\,\mathbb{E}[e^{n(\lambda,S_n)}] \\ &= e^{-n\inf_y(\lambda,y)+\Lambda(n\lambda)} \\ &= e^{-n(\inf_y(\lambda,y)-\frac{1}{n}\Lambda_n(n\lambda))} \end{aligned}$$

where $\Lambda_n(n\lambda) := \log \mathbb{E}[e^{n(\lambda,S_n)}] = n\Lambda(\lambda)$. This would seem to suggest that if $\{\Lambda_n\}_{n=1}^\infty$ is a sequence of functions such that $\Lambda(\lambda) := \lim_{n\to\infty}\frac{1}{n}\Lambda_n(n\lambda)$ exists then there should be an LDP, probability with rate function Λ^*.

Let $\{Z_n\}_{n=1}^\infty$ be a sequence of random variables, where for each n, Z_n is distributed as μ_n, with cumulant generating function Λ_n. Assume that, for all $\lambda \in \mathbb{R}^d$, $\Lambda(\lambda) := \lim_{n\to\infty}\frac{1}{n}\Lambda_n(\lambda)$ exists as an extended real number, and $0 \in \mathscr{D}_\Lambda^\circ$. This includes the setting for Cramér's theorem as a special case.

Are these assumptions enough on their own for an LDP? No, because we will not be able to obtain the lower bound. In Cramér's theorem we controlled the means of the X_i by changing the measure and then invoked the WLLN. We have neither of these tools in this setting.

3.1 The upper bound

What does carry over?

1) Properties of Λ and Λ^*:
   a) Λ is convex since it is a limit of convex functions.
   b) $\Lambda(\lambda) > -\infty$ for all $\lambda \in \mathbb{R}^d$, because $0 \in \mathscr{D}_\Lambda^\circ$.
   c) Λ^* is non-negative, convex, and l.s.c. since it is a Fenchel-Legendre transform.
   d) Λ^* is a good rate function because $0 \in \mathscr{D}_\Lambda^\circ$.

2) The LDP upper bound for compact sets holds by the same proof as in Cramér's theorem, by applying Chebyshev's inequality to $x \mapsto e^{n(\lambda,x)}$ and Z_n. Indeed,

$$\begin{aligned} \mu_n(B(x,r)) &\le e^{-n(\inf_{y\in B(x,r)}(\lambda,y)-\frac{1}{n}\Lambda_n(n\lambda))} \\ \limsup_{n\to\infty}\frac{1}{n}\log\mu_n(B(x,r)) &\le -(\lambda,x)+r|\lambda|+\Lambda(\lambda) \\ &\le \min\{\Lambda^*(x)-\delta,1/\delta\}+\delta \end{aligned}$$

and then take $\delta \to 0$.

3) It can be shown that $\{\mu_n\}_{n=1}^\infty$ is exponentially tight by the same method as in Cramér's theorem. On $\mathbb{R}$, for $\rho > 0$,

$$\begin{aligned} \mu_n([\rho,\infty)) &\le e^{-n(\lambda\rho-\frac{1}{n}\Lambda_n(n\lambda))} && \text{for } \lambda > 0 \\ \mu_n((-\infty,-\rho]) &\le e^{-n(\lambda\rho-\frac{1}{n}\Lambda_n(n\lambda))} && \text{for } \lambda < 0 \end{aligned}$$

So if $\rho > \limsup_{n\to\infty}\mathbb{E}[Z_n]$ then $\Lambda^*(\rho) = \sup_{\lambda>0}(\lambda\rho - \Lambda(\lambda))$, etc. This with the previous point proves the LDP upper bound for closed sets.

3.2 The lower bound

Recall that, in proving the lower bound, it suffices to show that

$$\frac{1}{n}\log\mu_n(B(y,\delta)) \ge -\Lambda^*(y)$$

for all $y \in \mathscr{D}^\circ_{\Lambda^*}$ and $\delta > 0$ small. In Cramér's theorem we changed the mean of X_1 to y via the change of measure $\frac{d\tilde{\mu}}{d\mu}(z) := \exp((\lambda_y,z)-\Lambda(\lambda_y))$, where $\lambda_y \in \mathbb{R}^d$ was such that $y = \nabla\Lambda(\lambda_y)$. We cannot choose such a nice representative in this setting, but we proceed anyways.

Fix $\lambda^* \in \mathscr{D}^\circ_\Lambda$ and set $\frac{d\tilde{\mu}_n}{d\mu}(z) := \exp(n((\lambda^*,z)-\frac{1}{n}\Lambda_n(n\lambda^*)))$ for large enough n that $\Lambda_n(n\lambda)$ is finite valued. Note that $\tilde{\mathbb{E}}[Z_n] = \frac{1}{n}\nabla\Lambda_n(n\lambda^*)$, but this is immaterial.

$$\begin{aligned} \tilde{\Lambda}_n(\lambda) &= \log\int_{\mathbb{R}^d} e^{n(\lambda,z)+n(\lambda^*,z)-\Lambda_n(n\lambda^*)}\mu_n(dz) \\ &= \Lambda_n(n(\lambda+\lambda^*)) - \Lambda_n(n\lambda^*) \\ \text{so } \frac{1}{n}\tilde{\Lambda}_n(\lambda) &\to \Lambda(\lambda+\lambda^*) - \Lambda(\lambda^*) =: \tilde{\Lambda}(\lambda) \end{aligned}$$

It can be seen that $\tilde{\Lambda}(0) = 0$ and $0 \in \mathscr{D}^\circ_{\tilde{\Lambda}}$ because $\lambda^* \in \mathscr{D}^\circ_\Lambda$. Therefore we may apply the LDP upper bound to the measures $\{\tilde{\mu}_n\}_{n=1}^\infty$.

$$\limsup_{n\to\infty}\frac{1}{n}\log\tilde{\mu}_n(B(y,\delta)^c) \le -\inf_{z\in B(y,\delta)^c}\tilde{\Lambda}^*(z).$$

We assume for simplicity that the infimum is attained for some $z_0 \in B(y,\delta)^c$. If $\tilde{\Lambda}^*(z_0) > 0$ then it can easily be seen that $\mu_n(B(y,\delta)^c) \to 0$ as $n \to \infty$. To this end, note that $z_0 \neq y$. Further,

$$\begin{aligned}
\tilde{\Lambda}^*(z_0) &= \sup_{\theta\in\mathbb{R}^d}((\theta,z_0) - \Lambda(\theta+\lambda^*) + \Lambda(\lambda^*)) \\
&= -(\lambda^*,z_0) + \Lambda(\lambda^*) + \sup_{\theta\in\mathbb{R}^d}((\theta+\lambda^*,z_0) - \Lambda(\theta+\lambda^*)) \\
&= -(\lambda^*,z_0) + \Lambda(\lambda^*) + \Lambda^*(z_0) \\
&\geq -(\lambda^*,z_0) + ((\lambda^*,y) - \Lambda^*(y)) + \Lambda^*(z_0) \\
&= ((\lambda^*,y) - \Lambda^*(y)) - ((\lambda^*,z_0) - \Lambda^*(z_0)) \overset{??}{>} 0
\end{aligned}$$

If we can find $\lambda^* =: \lambda_y$ such that this last line is strictly positive then y is said to be an *exposed point* of Λ^*, with *exposing hyperplane* λ_y. Define $\mathscr{Y} := \{y : y$ is an exposed point with an exposing hyperplane λ_y and $\lambda_y \in \mathscr{D}_\Lambda^\circ\}$. Note

(i) If there is λ such that $y = \nabla\Lambda(\lambda)$ then $y \in \mathscr{Y}$.

(ii) It can be seen that $\mathscr{Y} \subseteq \mathscr{D}_{\Lambda^*}$.

If $y \in \mathscr{Y}$ then there is λ_y such that $\tilde{\mu}_n(B(y,\delta)) \to 1$ for all $\delta > 0$. Let $0 < \eta < \delta$.

$$\begin{aligned}
\log\mu_n(B(y,\delta)) &= \log\int_{B(y,\delta)} e^{-n((\lambda_y,z)+\frac{1}{n}\Lambda_n(n\lambda_y))}\tilde{\mu}_n(dz) \\
&\geq \log\int_{B(y,\eta)} e^{-n((\lambda_y,z)+\frac{1}{n}\Lambda_n(n\lambda_y))}\tilde{\mu}_n(dz) \\
&= \frac{1}{n}\Lambda_n(n\lambda_y) - (\lambda_y,y) + \log\int_{B(y,\eta)} e^{-n(\lambda_y,z-y)}\tilde{\mu}_n(dz) \\
&\geq \frac{1}{n}\Lambda_n(n\lambda_y) - (\lambda_y,y) - |\lambda|\eta + \frac{1}{n}\log\tilde{\mu}_n(B(y,\eta))
\end{aligned}$$

$$\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(B(y,\delta)) \geq \Lambda(\lambda_y) - (\lambda_y,y) - |\lambda|\eta$$

Taking $\eta \to 0$ proves what needed to be proved to get the lower bound. What we have is that, for any open set G,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(G) \geq -\inf_{y\in G\cap\mathscr{Y}}\Lambda^*(y).$$

This is not quite the LDP lower bound if $\mathscr{Y} \neq \mathbb{R}^d$, but there are a variety of conditions that may be placed on Λ that guarantee this, discussed in the textbook.

4 Varadhan's Integral Lemma

Varadhan's integral lemma leverages an LDP to provide approximations to the expected value of certain continuous functions.

4.0.1 Theorem. *Let $\{Z_\varepsilon\}_{\varepsilon>0}$ be a family of random variables taking values in a space $\mathscr{X}$ satisfying an LDP with good rate function I. Let $\phi : \mathscr{X} \to \mathbb{R}$ be a continuous function and suppose there is $\gamma > 1$ such that*

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\gamma\phi(Z_\varepsilon)/\varepsilon}] < \infty.$$

Then

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] = \sup_{x\in\mathscr{X}}\{\phi(x) - I(x)\}.$$

Remark. The requirement involving $\gamma > 1$ is a UI-type criterion, and will hold in particular if ϕ is bounded.

As an example of how the lemma is motivated, suppose Z_ε is real valued and $\frac{d\mu_\varepsilon}{d\mu} = e^{-I(x)/\varepsilon}$ defines a family $\{\mu_\varepsilon\}_{\varepsilon>0}$ absolutely continuous with respect to some reference measure. Then

$$\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon))}] = \int_{\mathbb{R}} e^{\frac{1}{\varepsilon}(\phi(x)-I(x))}dx.$$

Suppose further that $(\phi - I)$ is a smooth concave function with a unique maximizer $\hat{x}$. Then by Taylor's theorem,

$$\phi(x) - I(x) \approx \phi(\hat{x}) - I(\hat{x}) - \frac{1}{2}(x-\hat{x})^2 G(x),$$

so plugging this into the expression above,

$$\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] \approx e^{\frac{1}{\varepsilon}(\phi(\hat{x})-I(\hat{x}))}\int_{\mathbb{R}} e^{\frac{1}{2\varepsilon}(x-\hat{x})^2G(x)}dx.$$

The lemma says that $\varepsilon \log \int_{\mathbb{R}} e^{\frac{1}{2\varepsilon}(x-\hat{x})^2G(x)}dx \to 0$.

4.1 Lower bound for l.s.c. functions

4.1.1 Lemma. *If $\phi : \mathscr{X} \to \mathbb{R}$ is l.s.c. and $\{Z_\varepsilon\}_{\varepsilon>0}$ satisfy the LDP lower bound with (not necessarily good) rate function I then*

$$\liminf_{n\to\infty} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] \geq \sup_{x\in\mathscr{X}}\{\phi(x) - I(x)\}.$$

PROOF: Let $x \in \mathscr{X}$ and $\delta > 0$ small be given. Since ϕ is l.s.c. there is $r > 0$ such that $\inf_{y\in B(x,r)} \phi(y) \geq \phi(x) - \delta$.

$$\begin{aligned}\liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] &\geq \liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{B(x,r)}(Z_\varepsilon)] \\ &\geq (\phi(x) - \delta) - \inf_{y\in B(x,r)} I(y) \\ &\geq (\phi(x) - I(x)) - \delta\end{aligned}$$

Take $\delta \to 0$ and the supremum over x. □

4.2 Upper bound for u.s.c. functions

4.2.1 Lemma. *If $\phi : \mathscr{X} \to \mathbb{R}$ is u.s.c. and $\{Z_\varepsilon\}_{\varepsilon>0}$ satisfy the LDP upper bound (for closed sets) with good rate function I, and if there is $\gamma > 1$ such that*

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\gamma\phi(Z_\varepsilon/\varepsilon)}] < \infty,$$

then

$$\liminf_{n\to\infty} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] \le \sup_{x\in\mathscr{X}}\{\phi(x) - I(x)\}.$$

PROOF: Assume for now that ϕ is bounded above by a constant M. Let $\alpha > 0$ big and $\delta > 0$ small be given. For each $x \in \{I \le \alpha\}$, there is $r_x > 0$ such that both

$$\sup_{\overline{B}(x,r_x)} \phi(y) \le \phi(x) + \delta \quad \text{and} \quad \inf_{\overline{B}(x,r_x)} I(y) \ge I(x) - \delta$$

since ϕ is u.s.c. and I is l.s.c. Then

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{B(x,r_x)}(Z_\varepsilon)] &\le (\phi(x)+\delta) - \inf_{y\in\overline{B}(x,r_x)} I(y)\\ &\le (\phi(x) - I(x)) + 2\delta\end{aligned}$$

Cover $\{I \le \alpha\}$ with balls $B(x_1, r_{x_1}), \dots, B(x_n, r_{x_n})$, which can be done since I is a good rate function, and let $C = \bigcap_{i=1}^n B(x_i, r_{x_i})^c$.

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] &\le \limsup_{\varepsilon\to 0} \varepsilon \log \Bigg(\sum_{i=1}^n \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{B(x_i,r_{x_i})}(Z_\varepsilon)]\\ &\qquad\qquad + \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_C(Z_\varepsilon)]\Bigg)\\ &\le \max\Big(\max_{i=1}^n\{\phi(x_i) - I(x_i) + 2\delta\}, M-\alpha\Big)\\ &\le \max\Big(\sup_{x\in\mathscr{X}}\{\phi(x) - I(x)\}, M-\alpha\Big) + 2\delta\end{aligned}$$

Send $\delta \to 0$ and $\alpha \to \infty$ to complete the proof of the case when ϕ is bounded above. Now assume that, for some $\gamma > 1$,

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\gamma\phi(Z_\varepsilon)/\varepsilon}] < \infty.$$

The remainder of the proof is a series of tedious calculations.

$$\begin{aligned}\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{[M,\infty)}(Z_\varepsilon)] &= e^{M/\varepsilon}\,\mathbb{E}[e^{\frac{1}{\varepsilon}(\phi(Z_\varepsilon)-M)}\mathbf{1}_{[1,\infty)}(e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon - M)})]\\ &\le e^{M/\varepsilon}\,\mathbb{E}[e^{\frac{1}{\varepsilon}(\phi(Z_\varepsilon)-M)}]\\ &\le e^{-(\gamma-1)M/\varepsilon}\,\mathbb{E}[e^{\frac{\gamma}{\varepsilon}\phi(Z_\varepsilon)}]\\ \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{[M,\infty)}(Z_\varepsilon)] &\le -(\gamma-1)M + \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[e^{\frac{\gamma}{\varepsilon}\phi(Z_\varepsilon)}]\end{aligned}$$

In particular,

$$\lim_{M\to\infty}\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{[M,\infty)}(Z_\varepsilon)] = -\infty.$$

Finally, letting ϕ_M denote $\phi \wedge M$,

$$\begin{aligned}
&\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] \\
&\qquad \le \limsup_{\varepsilon\to 0}\varepsilon\log\big(\mathbb{E}[e^{\frac{1}{\varepsilon}\phi_M(Z_\varepsilon)}] + \mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{[M,\infty)}(Z_\varepsilon)]\big) \\
&\qquad \le \max\{\sup_{x\in\mathscr{X}}\{\phi_M(x) - I(x)\}, \limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{[M,\infty)}(Z_\varepsilon)]\} \\
&\qquad \le \max\{\sup_{x\in\mathscr{X}}\phi(x) - I(x), \limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}\mathbf{1}_{[M,\infty)}(Z_\varepsilon)]\}
\end{aligned}$$

The result follows by taking $M \to \infty$. □

4.3 Laplace principles

Recap: If $\{Z_\varepsilon\}_{\varepsilon>0}$ satisfies an LDP with good rate function I and ϕ is continuous and bounded then

$$\lim_{\varepsilon\to 0}\varepsilon\log\mathbb{E}[e^{\frac{1}{\varepsilon}\phi(Z_\varepsilon)}] = \sup_{x\in\mathscr{X}}\{\phi(x) - I(x)\}.$$

Conversely, if this holds for a family $\{Z_n\}_{n=1}^\infty$ and a good rate function I, for all $\phi \in C_b(\mathscr{X};\mathbb{R})$, then we say that $\{Z_n\}_{n=1}^\infty$ satisfies a *Laplace principle* on $\mathscr{X}$ with rate function I. We just saw that the LDP implies the Laplace principle. It can be shown that the Laplace principle implies the LDP if $\mathscr{X}$ is a Polish space.

Bryce (1990) showed that if the measures $\{\mu_n\}_{n=1}^\infty$ are exponentially tight and the limit

$$\Lambda_\phi := \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[e^{n\phi(Z_n)}]$$

exists for all $\phi \in C_b(\mathscr{X};\mathbb{R})$ then $\{Z_n\}_{n=1}^\infty$ satisfy the LDP with good rate function $I(x) := \sup_{\phi\in C_b(\mathscr{X};\mathbb{R})}\{\phi(x) - \Lambda_\phi\}$.

4.4 Asymptotically optimal variance reduction

In *Monte Carlo simulations*, goal is compute $\mathbb{E}^P[G(X)]$, where we know enough about the law of X to sample from it, but not enough to compute the expected value explicitly. Suppose further that G has some properties that make it really bad to simulate under the original measure. (Think insurance or credit derivatives, where there is a tiny chance for a huge payout.) If $P \sim Q$ then

$$\begin{aligned}
\mathbb{E}^P[G] &= \mathbb{E}^Q[G\tfrac{dP}{dQ}] \\
\mathrm{Var}^Q(G\tfrac{dP}{dQ}) &= \mathbb{E}^Q[G^2(\tfrac{dP}{dQ})^2] - \mathbb{E}^Q[G\tfrac{dP}{dQ}]^2 = \mathbb{E}^P[G^2\tfrac{dP}{dQ}] - \mathbb{E}^P[G]^2
\end{aligned}$$

The idea is to change measure in such a way that this variance is as small as possible, and then compute $\mathbb{E}^P[G]$ using this new measure. This problem is not well posed, because taking $\frac{dQ}{dP} = G/\mathbb{E}^P[G]$ would give zero variance, but finding this Q is as hard are the original simulation problem! Instead, we choose a suitable class of measures $\mathscr{Q}$ over which to minimize $\mathbb{E}^Q[G^2(\frac{dP}{dQ})^2]$. One popular class has are those measures that have density with respect to P of the form

$$\frac{dQ^\lambda}{dP}(x) = e^{(\lambda,x)-\Lambda(\lambda)}$$

for a fixed function $\Lambda : \mathscr{X} \to [0,\infty]$, defined for all $\lambda \in \mathscr{D}_\Lambda$. Hence we wish to minimize

$$\min_{\lambda\in\mathscr{X}^*} \mathbb{E}^P[\exp(2\log G - (\lambda,X) + \Lambda(\lambda))],$$

which is ripe for an LDP approximation. Assume that $X = X_1$ for $\{X_\varepsilon\}_{\varepsilon\leq 1}$ which satisfies an LDP with good rate function I.

$$\begin{aligned}\log\mathbb{E}^P&[\exp(2\log G(X) - (\lambda,X) + \Lambda(\lambda))]\\ &= \varepsilon\log\mathbb{E}^P[\exp((2\log G(X_\varepsilon) - (\lambda,X_\varepsilon) + \Lambda(\lambda))/\varepsilon)]\end{aligned}$$

when $\varepsilon = 1$. The goal is to find λ which solves

$$\begin{aligned}\min_\lambda \varepsilon\log\mathbb{E}^P&[\exp((2\log G(X_\varepsilon) - (\lambda,X_\varepsilon) + \Lambda(\lambda))/\varepsilon)]\\ &= \min_\lambda \sup_x\{2\log G(x) - (\lambda,x) + \Lambda(\lambda) - I(x)\}\\ &= \min_\lambda \sup_x\{2\log G(x) + (I(x) - (\lambda,x) + \Lambda(\lambda)) - 2I(x)\}\\ &\geq 2\sup_x\{\log G(x) - I(x)\}\end{aligned}$$

The method is to solve the last problem for the maximizer x^* and find λ^* by attempting to solve $I(x^*) = \sup_\lambda(\lambda, x^*) - \Lambda(\lambda)$. Then check whether

$$\sup_x\{2\log G(x) + (I(x) - (\lambda^*,x) + \Lambda(\lambda^*)) - 2I(x)\} \geq 2\sup_x\{\log G(x) - I(x)\}.$$

If this happens then the resulting change of measure is asymptotically optimal (show this). Note that

$$\mathbb{E}^P[G^2\tfrac{dP}{dQ}] = \mathbb{E}^Q[G^2(\tfrac{dP}{dQ})^2] \geq \mathbb{E}^Q[G(\tfrac{dP}{dQ})]^2 = \mathbb{E}^P[G]^2$$

and

$$\lim_{\varepsilon\to 0} \varepsilon\log\mathbb{E}[e^{\frac{1}{\varepsilon}\log G(X_\varepsilon)}]^2 = 2\sup_{x\in\mathscr{X}}\{\log G(x) - I(x)\}$$

so in some sense the asymptotically optimal change of measure over this class achieves the same asymptotically optimal change over measure over any class.

5 General results

5.1 An existence result for weak LDP

Let $\mathscr{X}$ be a regular topological space and A be a base for the topology of $\mathscr{X}$, and let $\{\mu_n\}_{n=1}^\infty$ be a sequence of Borel measures on $\mathscr{X}$. If $\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A)$ exists for all $A \in A$ then there is a weak LDP. More precisely,

5.1.1 Lemma. *Suppose that*

$$L_A := -\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A)$$

exists for all $A \in A$. Let $I(x) := \sup\{L_A : A \in A, x \in A\}$. If

$$I(x) = \sup_{\substack{A\in A\\ x\in A}} \left(-\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A) \right)$$

then $\{\mu_n\}_{n=1}^\infty$ satisfies a weak LDP.

Proof: First we check that I is a rate function. Clearly $I \geq 0$ since $L_{\mathscr{X}} = 0$. Let $\alpha > 0$ be given. If $x \in \{I > \alpha\}$ then there is $A \in A$ such that $L_A > \alpha$. But then if $y \in A$ then $I(y) > \alpha$, so $L_A \subseteq \{I > \alpha\}$. Therefore $\{I > A\}$ is open, so $\{I \leq \alpha\}$ is closed and I is l.s.c.

Let G be an open set. Then there is $A \in A$ such that $x \in A \subseteq G$.

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -L_A \geq I(x),$$

so we have the LDP lower bound for open sets, pretty much by definition. Up to this point we have not used the assumption on I.

Let Γ be a compact set. As usual, define $I^\delta := \min\{I - \delta, \frac{1}{\delta}\}$. For any $x \in \Gamma$ and $\delta > 0$ we can find $A_{x,\delta} \in A$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A_{x,\delta}) \leq -I^\delta(x).$$

For any fixed $\delta > 0$, there are $x_1, \dots, x_N \in \Gamma$ such that $\Gamma \subseteq \bigcup_{i=1}^N A_{x_i,\delta}$.

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) &\leq \limsup_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^N \mu_n(A_{x_i,\delta}) \\
&= \max_{i=1}^N \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A_{x_i,\delta}) \\
&\leq -\min_{i=1}^N I^\delta(x_i) \leq -\inf_{x\in\Gamma} I^\delta(x)
\end{aligned}$$

Taking $\delta \to 0$ completes the proof of the LDP upper bound for compact sets. □

5.1.2 Example. This result is sometimes applied as follows. When $\mathscr{X}$ is a topological vector space, choose A to be the base of open convex subsets. Let $f(n) := -\log \mu_n(A)$ and try to show that f is sub-additive, i.e. $f(n+m) \le f(n) + f(m)$. If this can be done then $\lim_{n\to\infty} f(n)/n$ either exists (if $\mu_n(A) > 0$ for all n) or is ∞ (if $\mu_n(A) = 0$ eventually). The above lemma then gives a weak LDP.

5.2 Contraction principle

Let $\mathscr{X}$ and $\mathscr{Y}$ be topological spaces. Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of Borel probability measures on $\mathscr{X}$ and $g : \mathscr{X} \to \mathscr{Y}$ be a continuous function. If $\{\mu_n\}_{n=1}^\infty$ satisfies the LDP in $\mathscr{X}$ with good rate function I, then the sequence of induced measures $\{\nu_n = \mu_n g^{-1}\}_{n=1}^\infty$ satisfies the LDP in $\mathscr{Y}$ with good rate function

$$I^g(y) := \inf_{x \in g^{-1}(\{y\})} I(x).$$

PROOF: Clearly $I^g \ge 0$. Note that I being good implies that $I^g(y) = I(\hat{x})$ for some $\hat{x}$ such that $g(\hat{x}) = y$. Whence

$$\{I^g \le \alpha\} = \{g(x) : I(x) \le \alpha\} = g(\{I \le \alpha\}),$$

which is compact, so I^g is a good rate function. (Note that this argument breaks down if I is not good. In fact, if I is not good then I^g is not necessarily a rate function. Consider $I \equiv 0$ and $g = \exp$, which has $I^g = \chi_{[0,\infty)}$ which is not l.s.c.) If G is open then

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \nu_n(G) &= \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(g^{-1}(G)) \\ &\ge -\inf_{x \in g^{-1}(G)} I(x) && \text{since } g^{-1}(G) \text{ is open} \\ &= -\inf_{y \in G} I'(y) \end{aligned}$$

Similarly, if Γ is closed then

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log \nu_n(\Gamma) &= \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(g^{-1}(\Gamma)) \\ &\le -\inf_{x \in g^{-1}(\Gamma)} I(x) && \text{since } g^{-1}(\Gamma) \text{ is closed} \\ &= -\inf_{y \in \Gamma} I'(y) \end{aligned}$$

□

5.3 Approximate contraction principle

Consider the same setting as the last section, but with $g : \mathscr{X} \to \mathscr{Y}$ measurable but not necessarily continuous. Suppose that $\{g_m\}_{m=1}^\infty$ is a sequence of continuous functions $\mathscr{X} \to \mathscr{Y}$. In what sense should $g_m \to g$ so that the contraction principle holds for g?

5.3.1 Theorem. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the underlying probability space, $\mathscr{X}$ and $\mathscr{Y}$ be Polish spaces, $X_n : \Omega \to \mathscr{X}$ be random variables with distribution μ_n, $Y_n = g(X_n)$, and $Y_n^m = g_m(X_n)$. Assume that $g_m \to g$ in the following two ways.*

(i) For all $\delta > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[d_{\mathscr{Y}}(Y_n, Y_n^m) > \delta] = -\infty.$$

In this case we say that $\{Y_n\}_{n=1}^{\infty}$ and $\{\{Y_n^m\}_{n=1}^{\infty}\}_{m=1}^{\infty}$ are approximately exponentially equivalent.

(ii) For all $\alpha > 0$,

$$\lim_{m\to\infty} \sup_{\substack{x\in\mathscr{X} \\ I(x)\le\alpha}} d_{\mathscr{Y}}(g^m(x), g(x)) = 0.$$

This is uniform convergence on level sets of the rate function. Note that this implies that g is continuous where I is finite valued.

If these conditions are satisfied then $\{\nu_n = \mu_n g^{-1}\}_{n=1}^{\infty}$ satisfies an LDP with good rate function $I^g(y) := \inf_{x\in g^{-1}(\{y\})} I(x)$.

5.3.2 Example. One application of this result is as follows (this example will be discussed in detail later on). Suppose X and W are independent stochastic processes and that (X, W) satisfy some sort of LDP. The stochastic integral

$$(X, W) \mapsto \int_0^T f(X_t) dW_t,$$

is a measurable map, but it is not always continuous. Discrete time approximations are often continuous functionals, and they often approach the stochastic integral in the ways required by the theorem.

6 Sample Path LDP

What is the probability that the path of a process is in a given set of paths? We will leverage the finite dimensional results to obtain results about measures on path spaces.

6.1 Mogulskii's theorem

Consider the following setup. Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. $\mathbb{R}^d$ valued random variables distributed as μ. Assume for simplicity that $\mathscr{D}_\Lambda = \mathbb{R}^d$, where Λ is the cumulant generating function of μ, and let Λ^* be the Fenchel-Legendre transform of Λ, as usual. The objects under consideration are the pure jump processes

$$Z_n(t) := \sum_{i=1}^{\lfloor nt \rfloor} X_i$$

defined for $t \in [0,1]$. Let μ_n be the distribution of Z_n on the space of bounded measurable paths, $(L^\infty([0,1]), \|\cdot\|_\infty)$. Mogulskii's theorem states that $\{\mu_n\}_{n=1}^\infty$ satisfies the LDP with good rate function

$$I(\varphi) = \begin{cases} \int_0^1 \Lambda^*(\dot\varphi(t))dt & \text{when } \varphi \in AC([0,1]) \text{ and } \varphi(0) = 0 \\ \infty & \text{otherwise,} \end{cases}$$

where $\dot\varphi$ is the (almost sure) derivative of φ.

First, some motivation. Why should this be true? If

$$Z_n^\Pi = (Z_n(t_1), \dots, Z_n(t_{|\Pi|}))$$

where $\Pi = \{0 < t_1 < \cdots < t_{|\Pi|} \le 1\}$ is a fixed partition and μ_n^Π is the distribution of Z_n^Π on $(\mathbb{R}^d)^{|\Pi|}$, then $\{\mu_n^\Pi\}_{n=1}^\infty$ satisfies the LDP with good rate function

$$I^\Pi(z) := \sum_{i=1}^{|\Pi|} (t_i - t_{i-1}) \Lambda^* \left(\frac{z_i - z_{i-1}}{t_i - t_{i-1}} \right),$$

where $z_0 = 0$ and $t_0 = 0$ are set by definition. (See the problem set associated with lecture 4 for hints toward a proof of this fact.) Taking the size of the partition to zero yields the integral.

We will prove Mogulskii's theorem with the aid of three main lemmas. Recall that X_n and Y_n are *exponentially equivalent* if, for all $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[|X_n - \hat{Y}_n| > \delta] = -\infty.$$

Note that if X_n and Y_n are exponentially equivalent and one satisfies an LDP with rate function I then the other does as well.

6.1.1 Lemma. *Let $\tilde{Z}_n(t) = Z_n(t) + (t - \lfloor nt \rfloor / n) X_{\lfloor nt \rfloor + 1}$ be the polygonal process obtained from the random walk. Then $\tilde{Z}_n$ and Z_n are exponentially equivalent in $L^\infty([0,1])$.*

PROOF: Omitted, but similar to the calculation of the special case for normal random variables found in the proof of Schilder's theorem, below. □

The point of 6.1.1 is that it suffices to prove the LDP for $\tilde{Z}_n$, and this is easier because the $\tilde{Z}_n$ are continuous processes.

6.1.2 Lemma. *$\{\tilde{\mu}_n\}_{n=1}^\infty$ is exponentially tight in $(\mathscr{C}_0([0,1]), \|\cdot\|_\infty)$, the space of continuous functions null at zero.*

PROOF: Show that $K_\alpha := \{I \le \alpha\}$ is compact by proving it is bounded and equicontinuous. □

6.1.3 Lemma. *Set $\mathscr{X} := \{f : [0,1] \to \mathbb{R}^d : f(0) = 0\}$ and endow it with the topology of pointwise convergence (i.e. the product topology on $(\mathbb{R}^d)^{[0,1]}$). Then $\{\tilde{\mu}_n\}_{n=1}^{\infty}$ satisfies an LDP in $\mathscr{X}$ with good rate function I.*

Given these lemmas, the proof of Mogulskii's theorem is as follows. Note first that $\mathscr{D}_I \subseteq \mathscr{C}_0([0,1]) \subseteq \mathscr{X}$, and for all n, $\tilde{\mu}_n(\mathscr{C}_0([0,1])) = 1$. Therefore, by 6.1.3, $\{\tilde{\mu}_n\}_{n=1}^{\infty}$ satisfies an LDP in $\mathscr{C}_0([0,1])$ with the topology inherited from $\mathscr{X}$. This is not the uniform topology, unfortunately. However, every basic open set in this topology, $V_{t,x,g} := \{g \in \mathscr{X} : |g(t) - x| < \delta\} \cap \mathscr{C}_0([0,1])$, is open in the uniform topology on $\mathscr{C}_0$, so the uniform topology is finer than the topology of pointwise convergence. In particular the identity map from the uniform topology to the product topology is continuous.

The *inverse contraction principle* states that if $f : \mathscr{X} \to \mathscr{Y}$ is a continuous bijection and $\{\mu_n\}_{n=1}^{\infty}$ is exponentially tight then if it satisfies an LDP on $\mathscr{Y}$ with rate function I (automatically good by exponential tightness) then it satisfies an LDP on $\mathscr{X}$ with rate function $I \circ f$.

Finally, taking the LDP on $\mathscr{C}_0([0,1])$ to an LDP on $L^{\infty}([0,1])$ is accomplished by noting that $\mathscr{C}_0$ is a closed subset and carries the inherited topology, and that $\mathscr{D}_I \subseteq \mathscr{C}_0$.

Proof (of 6.1.3): The first step is to the show that $\{\tilde{\mu}_n\}_{n=1}^{\infty}$ satisfies an LDP in $\mathscr{X}$ with good rate function

$$I_{\mathscr{X}}(\phi) := \sup_{\Pi} I^{\Pi}(\phi) = \sup_{\Pi} \sum_{i=1}^{|\Pi|} (t_i - t_{i-1}) \Lambda^* \left(\frac{\phi(t_i) - \phi(t_{i-1})}{t_i - t_{i-1}} \right),$$

where the supremum is taken over all partitions Π of $[0,1]$. The second step is to prove that $I = I_{\mathscr{X}}$. To do this, first apply Jensen's inequality to see

$$I^{\Pi}(\phi) = \sum_{i=1}^{|\Pi|} (t_i - t_{i-1}) \Lambda^* \left(\frac{\phi(t_i) - \phi(t_{i-1})}{t_i - t_{i-1}} \right) \le \int_0^1 \Lambda^*(\dot{\phi}(t)) dt = I(\phi)$$

for all partitions Π and all functions ϕ. On the other hand, consider the specific partition

$$\Pi_k = \left\{ 0 < \frac{1}{k} < \cdots < \frac{k-1}{k} \le 1 \right\}.$$

For $\phi \in AC([0,1])$,

$$\begin{aligned} I_{\mathscr{X}}(\phi) &\ge \limsup_{k \to \infty} I^{\Pi_k}(\phi) \\ &= \limsup_{k \to \infty} \sum_{\ell=0}^{k} \frac{1}{k} \Lambda^* \left(k \left(\phi\left(\frac{\ell}{k}\right) - \phi\left(\frac{\ell-1}{k}\right) \right) \right) \ge I(\phi). \end{aligned}$$

If ϕ is not absolutely continuous then there is $\varepsilon > 0$ such that, for all $\delta > 0$, there is a partition $\Pi^{\delta} = \{0 \le s_2 < t_1 \le s_2 < t_2 \le \cdots \le s_k < t_k \le 1\}$ with $\sum_{i=1}^{k} t_i - s_i < \delta$

but $\sum_{i=1}^{k}|\phi(t_i)-\phi(s_i)|\geq\varepsilon$.

$$\begin{aligned} I_{\mathscr{X}}(\phi)\geq I^{\Pi^{\delta}}(\phi) &= \sup_{\lambda\in(\mathbb{R}^d)^k}\sum_{i=1}^{k}(\lambda_i,\phi(t_i)-\phi(s_i))-(t_i-s_i)\Lambda(\lambda_i) \\ &\geq \rho\varepsilon-\delta\sup_{|\lambda|=\rho}\Lambda(\lambda) \end{aligned}$$

where $\rho>0$ is any fixed radius and $\lambda_i:=\rho(\phi(t_i)-\phi(s_i))/|\phi(t_i)-\phi(s_i)|$ is chosen specifically. Since we have assumed $\mathscr{D}_{\Lambda}=\mathbb{R}^d$, the term being subtracted is finite valued, and can be made arbitrarily small by sending $\delta\to 0$. Taking $\rho\to\infty$ then shows that $I_{\mathscr{X}}(\phi)=\infty$.

To prove the first step, we need the concept of a *projective limit*. Let $(J,\leq)$ be a directed set, $(\mathscr{Y}_j)_{j\in J}$ be a family of topological spaces, and $(p_{ij})_{i\leq j}$ be a family of *compatible projections* $p_{ij}:\mathscr{Y}_j\to\mathscr{Y}_i$, i.e. continuous functions such that $p_{ik}=p_{ij}\circ p_{jk}$ for all $i\leq j\leq k$. Set $\mathscr{Y}=\prod_{j\in J}\mathscr{Y}_i$, and define $p_j:\mathscr{Y}\to\mathscr{Y}_j$ to be the projection. Endow $\mathscr{Y}$ with the weak topology defined by $\{p_j\}_{j\in J}$ (i.e. the product topology). Define $\hat{\mathscr{Y}}=\varprojlim\mathscr{Y}_i$ to be the subspace of $\mathscr{Y}$, with the induced topology, such that $y\in\hat{\mathscr{Y}}$ if $p_i(y)=p_{ij}(p_j(y))$, for all $i\leq j$. The following facts are true: (i) $\hat{\mathscr{Y}}$ is closed in $\mathscr{Y}$, (ii) the p_j, restricted to $\hat{\mathscr{Y}}$, are continuous on $\hat{\mathscr{Y}}$ and generate the topology, (iii) the projective limit of closed sets is closed, and (iv) the projective limit of nonempty compact subsets is nonempty and compact. In our setting, J is the collection of all partitions j of $[0,1]$, ordered by containment, $\mathscr{Y}_j=\mathbb{R}^{|j|}$, and $\hat{\mathscr{Y}}=\mathscr{X}$, with $p_j:\varphi\mapsto(\varphi(t_1),\dots,\varphi(t_{|j|}))$. If $\{\mu_n\}_{n=1}^{\infty}$ is a family of measures on $\hat{\mathscr{Y}}$ such that, for all $j\in J$, $\{\mu_n\circ p_j^{-1}\}_{n=1}^{\infty}$ satisfies an LDP with good rate function I^j, then $\{\mu_n\}_{n=1}^{\infty}$ satisfies an LDP with good rate function $\hat{I}(y):=\sup_{j\in J}I^j(p_j(y))$.

Indeed, $\hat{I}$ is a good rate function because

$$\{\hat{I}\leq\alpha\}=\hat{\mathscr{Y}}\cap\bigcap_{j\in J}\{I^j\leq\alpha\}=\varprojlim\{I^j\leq\alpha\},$$

which is compact because each I^j is good. If $G\subseteq\hat{\mathscr{X}}$ is open and $x\in A$ then there is $j\in J$ and $U_j\subseteq\mathscr{Y}_j$ open such that $x\in p_j^{-1}(U_j)$, so

$$\begin{aligned} \liminf_{n\to\infty}\frac{1}{n}\log\mu_n(A) &\geq \liminf_{n\to\infty}\frac{1}{n}\log\mu_n(p_j^{-1}(A)) \\ &\geq -\inf_{y\in U_j}I^j(y)\geq -I^j(p_j(x))\geq -\hat{I}(x). \end{aligned}$$

The upper bound is a bit more work. First show that $F=\varprojlim p_j(F)$. Second let $\alpha<\inf_{x\in F}\hat{I}(x)$ and show that $\{\hat{I}\leq\alpha\}\cap F=\varnothing$. Then $\varprojlim p_j(F)\cap\{I^j\leq\alpha\}=\varnothing$, so there is $j\in J$ such that $p_j(F)\cap\{I^j\leq\alpha\}=\varnothing$. Now, $F\subseteq p_j^{-1}(\overline{p_j(F)})$, so

$$\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(F)\leq\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(p_j^{-1}(\overline{p_j(F)}))\leq -\inf_{y\in p_j(F)}I^j(y)<-\alpha \qquad \square$$

6.2 Schilder's theorem

The objects under consideration now are $W_n(t) := \frac{1}{\sqrt{n}} W(t)$, where W is a Brownian motion on $[0,1]$. Schilder's theorem is essentially a corollary of Mogulskii's theorem, though it was originally proven ten years earlier.

6.2.1 Theorem (Schilder). *Let $(W_t)_{t\in[0,1]}$ be a standard d-dimensional Brownian motion. Set $W_n(t) := \frac{1}{\sqrt{n}} W(t)$ and let μ_n be the distribution of W_n on the space $(\mathscr{C}_0([0,1]), \|\cdot\|_\infty)$. Then $\{\mu_n\}_{n=1}^\infty$ satisfies the LDP with good rate function*

$$I_w(\phi) = \begin{cases} \frac{1}{2}\int_0^1 |\dot\phi(t)|^2 dt & \text{for } \phi \in H_1 \\ \infty & \text{otherwise} \end{cases}$$

where $H_1 := \{\phi \in AC([0,1]) : \phi(t) = \int_0^t \psi(s)ds, \psi \in L^2([0,1])\}$ is Cameron-Martin space, the collection of absolutely continuous functions with (a.s.) derivative in L^2.

Remark. (i) It is not hard to extend this result to Brownian motion on $[0,T]$.

(ii) In Stroock and Deuschel, they prove this result on the space of continuous functions with sublinear growth, $\{\omega \in \mathscr{C}_0([0,\infty)) : \lim_{t\to\infty} \omega(t)/t = 0\}$ under the norm $|\omega| := \sup_{t>0} |\omega(t)|/(1+t)$.

PROOF: Set $\hat W_n(t) = W_n(\lfloor nt \rfloor / n)$. We can identify $\hat W_n$ with $Z_n(t) = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} X_i$, where $X_i \sim N(0,1)$ i.i.d. By Mogulskii's theorem, $\{\hat\mu_n\}_{n=1}^\infty$ satisfies an LDP with good rate function I_w on $(L^\infty([0,1]), \|\cdot\|_\infty)$.

Need to show that W_n and $\hat W_n$ are exponentially equivalent. For any $\delta > 0$,

$$\begin{aligned} \mathbb{P}\Big[\sup_{0\le t\le 1} |W_n(t) - \hat W_n(t)| > \delta\Big] &\le n\,\mathbb{P}\Big[\sup_{0\le t\le 1/n} |W_n(t)| > \delta\Big] \\ &= 2n\,\mathbb{P}\Big[\sup_{0\le t\le 1/n} W(t) > \delta\Big] \\ &= 2n\,\mathbb{P}[W(1/n) > \sqrt{n}\delta] \\ &\approx g(n)e^{-n^2\delta^2}. \end{aligned}$$

(Something is not quite right with the above calculation.) Therefore

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big[\sup_{0\le t\le 1} |W_n(t) - \hat W_n(t)|_\infty > \delta\Big] = -\infty,$$

so $\{\mu_n\}_{n=1}^\infty$ satisfy the LDP with good rate function I_w on $L^\infty([0,1])$. But $C_0([0,1])$ is a closed subspace of $L^\infty([0,1])$ and $\mathscr{D}_{I_w} \subseteq C_0([0,1])$, so it satisfies the LDP on this space. □

6.3 Friedlin-Wentzell theory

In this section we extend Schilder's theorem to Itô diffusions. Consider a sequence of $\mathbb{R}^k$ valued diffusions

$$dX_t^n = b(X_t^n)dt + \frac{1}{\sqrt{n}}\sigma(X_t^n)dW_t, \qquad X_0^n = 0,$$

where W is a d-dimensional Brownian motion (the same for all n). Let μ_n be the distribution of X^n on $\mathscr{C}_0([0,1];\mathbb{R}^k)$. The idea is that $\mu_n \to \delta_\varphi$, where φ solves the ODE $\dot\varphi_t = b(\varphi_t)$, $\varphi_0 = 0$.

6.3.1 Example. Consider the Ornstein-Uhlenbeck process with $k = d = 1$.

$$dX_t^n = -X_t^n dt + \frac{1}{\sqrt{n}}dW_t, \qquad X_0^n = 0.$$

The SDE can be solves explicitly.

$$X_t^n = \int_0^t e^{-(t-s)}\frac{1}{\sqrt{n}}dW_s = \frac{1}{\sqrt{n}}W_t - \int_0^t e^{-(t-s)}\frac{1}{\sqrt{n}}W_s ds =: f\left(\frac{1}{\sqrt{n}}W\right)_t,$$

and in this case $f : \mathscr{C}_0([0,1];\mathbb{R}) \to \mathscr{C}_0([0,1];\mathbb{R})$ is in fact continuous. By the Contraction Principle, we have an LDP for $\{\mu_n\}_{n=1}^\infty$, with

$$I^X(\varphi) = \inf\left\{\frac{1}{2}\int_0^1 \dot\psi_s^2 ds : \psi \in H^1, \varphi = f(\psi)\right\}$$

The infimum is taken over those ψ such that $\varphi_t = \int_0^t e^{-(t-s)}\dot\psi_s ds$ (by integration by parts). This reduces to the differential equation $\dot\psi_t = \dot\varphi_t + \varphi_t$, so we may write

$$I^X(\phi) = \frac{1}{2}\int_0^1 (\dot\varphi_t + \varphi_t)^2 dt = \frac{1}{2}\int_0^1 \frac{\dot\varphi_t - b(\varphi_t))^2}{\sigma(\phi_t)^2}dt.$$

In general X^n is not a continuous image of the Brownian motion. We proceed as though it were and come up with appropriate conditions. If

$$X_t^n = x + \int_0^t b(X_s^n)ds + \int_0^t \sigma(X_s^n)\frac{1}{\sqrt{n}}dW_s = f\left(\frac{1}{\sqrt{n}}W\right)_t$$

and f were continuous then $\{\mu_n\}_{n=1}^\infty$ would satisfy an LDP by the contraction principle, with

$$I^X(\psi) = \inf\left\{\frac{1}{2}\int_0^1 \dot\varphi_s^2 ds : \psi_t = x + \int_0^t b(\psi_s)ds + \int_0^t \sigma(\psi_s)\dot\varphi_s ds\right\}.$$

6.3.2 Theorem. *Assume b and σ are bounded and uniformly Lipschitz. Then $\{\mu_n\}_{n=1}^\infty$ satisfies the LDP in $\mathscr{C}_0([0,1];\mathbb{R}^k)$ with good rate function I^X above.*

Remark. If $\sigma\sigma^T$ is uniformly elliptic then

$$I^X(\varphi) = \begin{cases} \frac{1}{2}\int_0^1 |\sigma^{-1}(\varphi_t)(\dot{\varphi}_t - b(\varphi_t))|^2 dt & \varphi \in H_x^1 \\ \infty & \text{otherwise.} \end{cases}$$

Some of the assumptions on b and σ may be relaxed. In particular, the theorem holds if $\sigma\sigma^T(x) \le K(1+|x|^2)I_d$ and $|b(x)| \le K\sqrt{1+|x|^2}$.

PROOF: The result follows by invoking the Approximate Contraction Principle. Assume for simplicity that $x = 0$. For each m, define the function

$$F^m : \mathscr{C}_0([0,1];\mathbb{R}^d) \to \mathscr{C}_0([0,1];\mathbb{R}^k) \text{ by } F^m(g)_0 = 0 \text{ and}$$
$$F^m(g)_t = F^m(g)_{k/m} + (t - k/m)b(F^m(g)_{k/m}) + \sigma(F^m(g)_{k/m})(g_t - g_{k/m})$$

for $t \in [k/m,(k+1)/m)$ and $k = 0,\dots,m-1$. Define $F : \{I_w < \infty\} \to AC[0,1]$ via

$$F(g)_t = \int_0^t b(F(g)_s)ds + \int_0^t \sigma(F(g)_s)\dot{g}_s ds.$$

Each F^m is continuous (by the boundedness and Lipschitz conditions). Let $X_t^{n,m} = F^m(\frac{1}{\sqrt{n}}W)_t$. To invoke the principle, we need to show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[\|X^{n,m} - X^n\|_\infty > \delta] = -\infty$$

and

$$\lim_{m\to\infty} \sup_{g:I_w(g)\le\alpha} \|F^m(g) - F(g)\|_\infty = 0.$$

□

7 LDP for occupancy times

The remainder of the seminar is dedicated to LDP for occupancy times of ergodic Markov processes. The plan is to first study LDP for Markov chains taking values in a finite state space $\Sigma_N := \{1,\dots,N\}$, using the Gärtner-Ellis theorem. Then we will study Markov chains taking values in Polish spaces, and time permitting, we will talk about continuous parameter Markov processes taking values in Polish spaces.

In each case we will appeal to an abstract result to get an LDP, and then we will develop alternate, friendlier, expressions for the rate function.

7.1 Finite state Markov chains

Let $\{Y_i\}_{i=1}^\infty$ be a Markov chain taking values in Σ_N. Let Π denote the $N \times N$ *transition matrix*, sometimes called a *stochastic matrix*. Then $\Pi_{ij} \ge 0$ for all $1 \le i,j \le N$ and $\Pi\mathbf{1} = \mathbf{1}$. The setup is $\Omega = \Sigma_N^{\mathbb{N}_0}$, $\mathscr{F}$ is the power set, Y is the canonical process, and $\{P_\sigma^\Pi\}_{\sigma\in\Sigma}$ is a (Markov) family of measures such that $P_\sigma^\Pi[Y_0=\sigma]=1$ and

$$P_\sigma^\Pi[Y_{n+1}\in\Gamma|\sigma(Y_m : m\le n)] = \Pi(Y_n,\Gamma)$$

Write $P_\sigma^{\Pi,n}(y_1,\dots,y_n) = \Pi(\sigma,y_1)\Pi(y_1,y_2)\cdots\Pi(y_{n-1},y_n)$.

Let $f:\Sigma\to\mathbb{R}^d$ be a function and consider the sequence of random variables $Z_n = \frac{1}{n}\sum_{i=1}^n f(Y_i)$. For example, $f(i)=(\delta_{i,1},\dots,\delta_{i,n})\in\mathbb{R}^N$ gives $Z_n = L_n^Y$, the setting for Sanov's theorem. Let $\mu_{n,\sigma}$ be the distribution of Z_n on $\mathbb{R}^d$ under $P_\sigma^{\Pi,n}$.

If there were a measure μ such that $\Pi(i,j)=\mu(j)$ for all i and j then the Y_i's would be i.i.d. and we would be in the setting of Cramér's theorem, so there is an LDP with rate function

$$\Lambda^* = \sup_{\lambda\in\mathbb{R}^d}(\lambda,x) - \log\mathbb{E}^\mu[e^{(\lambda,f(Y_1))}].$$

What conditions do we need on Π so that, if $\Lambda_n^{(\sigma)}(\lambda) := \log\mathbb{E}^{P_\sigma^{\Pi,n}}[e^{(\lambda,Z_n)}]$ then $\lim_{n\to\infty}\frac{1}{n}\log\Lambda_n^{(\sigma)}(n\lambda)$ exists. We would also like to know when this limit is independent of σ.

7.1.1 Assumption. Π *is irreducible, i.e. for all* $1\le i,j\le N$ *there is* $m(i,j)$ *such that* $\Pi^{m(i,j)}(i,j)>0$. *In words, the chain can move from any state to any other state with positive probability.*

7.1.2 Theorem (Perron-Frobenius). *Let* B *be an* $N\times N$ *irreducible matrix with non-negative entries. Set*

$$\rho(B) = \sup_{\substack{\|x\|=1\\ x\ge 0}}\min_{1\le i\le N}\frac{(Bx)_i}{x_i}.$$

(i) $\rho(B)$ *is an eigenvalue of* B.
(ii) There are left and right eigenvectors associated to $\rho(B)$ *with strictly positive entries, and the eigenspaces are one dimensional.*
(iii) If λ *is any other eigenvalue of* B *then* $|\lambda|\le\rho(B)$.
(iv) If $\varphi\in\mathbb{R}^N$ *with* $\varphi\gg 0$ *then, for all* i,

$$\lim_{n\to\infty}\frac{1}{n}\log(B^n\varphi)_i = \rho(B) = \lim_{n\to\infty}\frac{1}{n}\log((B^T)^n\varphi)_i.$$

7.1.3 Theorem. *Let* Π *be an irreducible stochastic matrix. For* $\lambda\in\mathbb{R}^d$ *define the* $N\times N$ *irreducible (but not necessarily stochastic) matrix* Π^λ *via*

$$\Pi^\lambda(i,j) = e^{(\lambda,f(j))}\Pi(i,j).$$

Then $\{\mu_{n,\sigma}\}_{n=1}^\infty$ *satisfies the LDP with good rate function*

$$I(x) := \sup_{\lambda\in\mathbb{R}^d}\{(\lambda,x) - \log\rho(\Pi^\lambda)\}$$

PROOF: We need only to show that

$$\lim_{n\to\infty}\frac{1}{n}\Lambda_n(n\lambda)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}^{P_\sigma^{\Pi,n}}[e^{n(\lambda,Z_n)}]$$

exists and is finite for all $\lambda\in\mathbb{R}^d$, and then apply the Gärtner-Ellis theorem.

$$\begin{aligned}\frac{1}{n}\log\mathbb{E}^{P_\sigma^{\Pi,n}}[e^{n(\lambda,Z_n)}] &= \frac{1}{n}\log\mathbb{E}^{P_\sigma^{\Pi,n}}[e^{(\lambda,\sum_{i=1}^n f(i))}]\\ &= \frac{1}{n}\log\sum_{y_1,\dots,y_n}\Pi(\sigma,y_1)e^{(\lambda,f(y_1))}\cdots\Pi(y_{n-1},y_n)e^{(\lambda,f(y_n))}\\ &= \frac{1}{n}\log\sum_{y_n=1}^N(\Pi^\lambda)^n(\sigma,y_n)\\ &= \frac{1}{n}\log\sum_{y_n=1}^N(\Pi^\lambda)^n(\sigma,y_n)\mathbf{1}\\ &= \frac{1}{n}\log((\Pi^\lambda)^n\mathbf{1})_\sigma=\log\rho(\Pi^\lambda)\end{aligned}$$

□

Remark. Note that I does not depend on σ. In the setting of Sanov's theorem, $\Pi^\lambda(i,j)=e^{\lambda_j}\Pi(i,j)$, as might have been suspected. Therefore we have two expressions for the rate function

$$\sup_{\lambda\in\mathbb{R}^N}\{(\lambda,x)-\log\rho(\Pi^\lambda)\}=\sup_{\lambda\in\mathbb{R}^N}\{(\lambda,x)-\Lambda(\lambda)\}$$

where $\Lambda(\lambda)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}^{P_\sigma^{\Pi,n}}[e^{n(\lambda,L_n^Y)}]$. Further, it can be shown these are equal to

$$\hat{I}(x)=\begin{cases}\sup_{\substack{u\in\mathbb{R}^N\\ u\gg 0}}\sum_{i=1}^N x_i\log\left(\frac{u_i}{(u\Pi)_i}\right) & x\in M_1(\Sigma)\\ \infty & \text{otherwise.}\end{cases}$$

Indeed, $L_n^Y\in M_1(\Sigma)=\{x\in\mathbb{R}^n : x\geq 0, x\cdot\mathbf{1}=1\}$, so the LDP lower bound shows that they are equal off of $M_1(\Sigma)$. Fix $x\in M_1(\Sigma)$. For $\lambda\in\mathbb{R}^N$ set u to be the left principle eigenvector of Π^λ. Then

$$(\lambda,x)-\log\rho(\Pi^\lambda)=\sum_{i=1}^N x_i\log\left(\frac{u_i}{(u\Pi)_i}\right).$$

On the other hand, for $u\in\mathbb{R}^N$ with $u\gg 0$, set $\lambda_j:=\log(u_i/(u\Pi)_j)$.

7.2 Markov chains in Polish spaces

Let $\{Y_i\}_{i=1}^\infty$ be a Markov chain taking values in Σ, a complete separable metric space. The setup is $\Omega=\Sigma^{\mathbb{N}_0}$ with the product topology, $\mathscr{F}=\mathscr{B}(\Omega)=(\mathscr{B}(\Sigma))^{\mathbb{N}_0}$, the Borel σ-algebra, Y is the canonical process, $\mathscr{F}_n = \sigma(Y_m : m \leq n)$, and $\{P_\sigma\}_{\sigma\in\Sigma}$ is a (Markov) family of measures such that $P_\sigma[Y_0 = \sigma] = 1$ for all $\sigma \in \Sigma$ and

$$P_\sigma[Y_{n+1} \in \Gamma|\mathscr{F}_n] = \Pi(Y_n, \Gamma)$$

for a fixed stochastic kernel $\Pi : \Sigma \times \mathscr{B}(\Sigma) \to \mathbb{R}$.

The main object under consideration is $L_n^Y(\omega) := \frac{1}{n}\sum_{i=1}^n \delta_{Y_i(\omega)}$, the empirical measure of $\{Y_i\}_{i=1}^n$, a form of "occupancy time." Note that $L_n^Y(\omega) \in M_1(\Sigma)$. Call $\mu_{n,\sigma}$ the distribution of L_n^Y under P_σ. Note that $\mu_{n,\sigma} \in M_1(M_1(\Sigma))$. Yikes!

Structure of $M_1(\Sigma)$

Let $M(\Sigma)$ denote the space of all finite signed measures on $(\Sigma, \mathscr{B}(\Sigma))$. Then $M_1(\Sigma)$ is convex subset of $M(\Sigma)$. We put the topology of weak convergence on $M(\Sigma)$, with sub-basis

$$U_{f,x,\delta} = \{\nu \in M(\Sigma) : |\nu(f) - x| < \delta\}$$

with $x \in \Sigma$, $f \in C_b(\Sigma;\mathbb{R})$, and $\delta > 0$. With this topology $M(\Sigma)$ is locally convex and Hausdorff. Endow $M_1(\Sigma)$ with the relative topology. The topological dual of $M(\Sigma)$ with this topology is $C_b(\Sigma;\mathbb{R})$. Define the *Lévy metric* as follows.

$$d_L(\mu,\nu) := \inf_{\delta>0}\{\mu(F) \leq \nu(F^\delta) + \delta \text{ for all closed sets } F\}$$

With this metric, $(M_1(\Sigma), d_L)$ is a Polish space carrying the same topology as above.

The plan is to find a base $\mathscr{A}$ for $M_1(\Sigma)$ such that, for all $\nu \in M_1(\Sigma)$,

$$\sup_{\substack{A\in\mathscr{A}\\ \nu\in A}}\left(-\limsup_{n\to\infty}\frac{1}{n}\log(\sup_{\sigma\in\Sigma}\mu_{n,\sigma}(A))\right) = \sup_{\substack{A\in\mathscr{A}\\ \nu\in A}}\left(-\liminf_{n\to\infty}\frac{1}{n}\log(\inf_{\sigma\in\Sigma}\mu_{n,\sigma}(A))\right).$$

This will give a weak LDP for $\{\mu_{n,\sigma}\}_{n=1}^\infty$. Then we will prove exponential tightness, giving the full LDP. Finally, we will exhibit alternate, nicer, characterizations of the rate function.

Index