② General Principles

2.1 The Large Deviation Principle

We are going to extend and formalise what we have studied in chapter 1. The large deviation principle (LDP) characterises the limiting behaviour, as $n \to \infty$, of a sequence (family) of probability measures $(\mu_n)_{n \geq 1}$ on $(E, \mathcal{B})$ in terms of a rate function. This characterisation is via asymptotic upper and lower exponential bounds on the values that μ_n assigns to measurable subsets of E.

Throughout, E is a topological space so that open and closed subsets of E are well-defined, and the simplest situation is when elements of $\mathcal{B}_E$, the Borel σ-field on E, are of interest. We always assume that $\mathcal{B}_E \subseteq \mathcal{B}$ unless explicitly stated otherwise.

Definition 1: A function $I : E \longrightarrow [0, \infty]$ is called rate function if

(i) $I \not\equiv \infty$

(ii) I is lower semicontinuous

(iii) I has (compact) closed level sets, i.e. $\forall\, \alpha \geq 0$

$$\Psi_I^\alpha = \{ x : I(x) \leq \alpha \} \quad \text{closed (compact)}$$

The rate function is called a good rate function if the level sets are compact.

$\mathcal{D}_I = \{ x \in E : I(x) < \infty \}$ is the effective domain of I.

In the following, for any set $M \subset E$, $\bar{M}$ denotes the closure of M, and M^0 the interior of M, and M^c the complement. The infimum of a function over an empty set is interpreted as ∞.

Definition 2: A sequence $(\mu_n)_{n \geq 1}$ of probability measures on E is said to satisfy the large deviation principle (LDP) with rate n and with rate function I if, for all $\Gamma \in \mathcal{B}$,

$$(*) \quad -\inf_{x \in \Gamma^0} I(x) \leq \varliminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq \varlimsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x)$$

The right- and left-hand side of $(*)$ are referred to as the upper and lower bounds, respectively.

__Remarks:__

(a) If $\inf_{x\in\Gamma^\circ} I(\cancel{x}) = \inf_{x\in\overline{\Gamma}} I(x) =: I_\Gamma$, then

$$\lim_{n\to\infty} \frac{1}{n}\log \mu_n(\Gamma) = -I_\Gamma .$$

A set Γ with this property is called $\underline{I \text{ continuity}}$ set.
In general LDP implies precise limit only for I continuity sets.

(b) Why can't we just require

$$\lim_{n\to\infty} \frac{1}{n}\log \mu_n(\Gamma) = \inf_{x\in\Gamma} I(x) \quad \text{for all}$$

measurable sets Γ?

Because if we did, we would not be able to use the LDP to describe continuous random variables.

To see that, let the sequence $(\mu_n)_{n\geq 1}$ consist of measures which are non-atomic, i.e., $\mu_n(\{x\}) = 0 \ \forall n\in\mathbb{N}$ and $\forall x\in E$; then it must be that $I(x)=\infty \ \forall x\in E$.

If the lower bound in (4) was to hold with the the infimum over Γ instead of Γ°, it would have to be concluded that $I(x)\equiv\infty$, contradicting the upper bound because of $\mu_n(E)=1 \ \forall n\in\mathbb{N}$.

Thus, some topological restrictions are necessary, and the definition of the LDP codifies a particularly convenient way of stating results that, on the one hand, are accurate enough to be useful and, on the other hand, are loose enough to be correct.

(c) Since $\mu_n(E) = 1$ $\forall\, n \in \mathbb{N}$, it is necessary that $\inf_{x \in E} I(x) = 0$ for the upper bound to hold.

When I is a good rate function, this means that

$\exists\, x \in E$ for which $I(x) = 0$. The upper bound trivially holds whenever $\inf_{x \in \bar{\Gamma}} I(x) = 0$, while the lower bound trivially holds whenever $\inf_{x \in \Gamma^o} I(x) = \infty$.

Thus (*) is equivalent to the following bounds

- upper bound: $\forall\, \alpha < \infty$ and $\forall\, \Gamma \in \mathcal{B}$ with $\bar{\Gamma} \subset (\Psi_I^{\alpha})^c$,

(4) $$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\alpha$$

- lower bound

(5) $\forall\, x \in \mathcal{D}_I$ and any $\Gamma \overset{\subset \mathcal{B}}{}$ with $x \in \Gamma^o$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \geq -I(x)$$

(Note the local structure of the lower bound).

The LDP is equivalent to the following bounds

(6) <u>Upperbound</u> $\forall\, F \subset E$ closed,

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} I(x)$$

(7) <u>Lower bound</u> $\forall$ open sets $G \subset E$,

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} I(x)$$

Having defined what is meant by an LDP, the rest of this subsection is devoted to proving elementary properties and establishing a connection to the weak convergence of probability measures.

<u>**Lemma 3:**</u> Let $N \in \mathbb{N}$ be fixed. Then, for every $a_n^i \geq 0$ $(i \in \mathbb{N}, n \in \mathbb{N})$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \left(\sum_{i=1}^{N} a_n^i \right) = \max_{i \in \{1,\dots,N\}} \limsup_{n\to\infty} \frac{1}{n} \log a_n^i$$

<u>Proof:</u> $$\frac{1}{n} \log\left(\sum_{i=1}^{N} a_n^i\right) = \frac{1}{n} \log\left(\max_{i\in\{1,\dots,N\}} a_n^i\right)\left(1 + \frac{\sum_{i=1}^{N} a_n^i - \max_{i\in\{1,\dots,N\}} a_n^i}{\max_{i\in\{1,\dots,N\}} a_n^i}\right)$$

implies $$0 \leq \frac{1}{n} \log\left(\sum_{i=1}^{N} a_n^i\right) - \max_{i\in\{1,\dots,N\}} \frac{1}{n}\log a_n^i \leq \frac{1}{n} \log N$$

Since N is fixed, $1/n \log N \to 0$ as $n \to \infty$ and

$$\limsup_{n\to\infty} 1/n \max_{i\in\{1,\dots,N\}} 1/n \log a_n^i = \max_{i\in\{1,\dots,N\}} \limsup_{n\to\infty} 1/n \log a_n^i \qquad \square$$

For proving the upper bound of an LDP it is often natural to do so first for compact sets.

<u>**Definition 3:**</u> (Assume all compact sets of E belong to $\mathcal{B}$)

A sequence $(\mu_n)_{n\geq 1}$ of probability measures is said to satisfy the weak LDP with the rate function I and rate n if the upperbound holds for all compact $K \subset \Psi_I^c(\alpha)$, and the lower bound for all measurable sets: (cf. (4) and (5))

upper bounds $\forall \alpha < \infty$ $\forall \Gamma \in \mathcal{B}$ with $\Gamma \in \Psi_I^c(\alpha)$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\alpha$$

Lower bound: $\forall$ $x \in \mathcal{D}_I$ and any $\Gamma \in \mathcal{B}$ with $x \in \Gamma^0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \geq -I(x).$$

Turning the weak LDP to a full LDP requires a way of showing that most of the probability mass (at least on an exponential scale) is concentrated on compact sets. The tool of doing that is the following:

<u>Definition 4</u>: A family / sequence of probability measures $(\mu_n)_{n\geq 1}$ on E is exponentially tight if for every $\alpha < \infty$ there exists a compact set $K_\alpha \subset E$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha.$$

<u>Remark</u>: When E is locally compact or, alternatively, Polish, exponential tightness is implied by the goodness of the rate function.

<u>Lemma 4:</u> Let $(\mu_n)_{n\geq 1}$ be an exponentially tight sequence of probability measures on E.

(a) If the upper bound (4) holds for some $\alpha < \infty$ and all compact subsets of $\Psi_I^c(\alpha)$, then it also holds for all $\Gamma \in \mathcal{B}$ with $\overline{\Gamma} \subset \Psi_I^c(\alpha)$. That is, if (4) holds for all compact sets, then it also holds for all closed sets.

(b) If the lower bound (5) holds for all $\Gamma \in \mathcal{B}$, then I is a good rate function.

<u>Proof</u>: ad (a): Fix $\Gamma \in \mathcal{B}$ and $\alpha < \infty$ such that $\overline{\Gamma} \subset \Psi_I^c(\alpha)$. Let K_α be the compact set in the definition of exponential tightness, and note that $\overline{\Gamma} \cap K_\alpha \in \mathcal{B}$ and $K_\alpha^c \in \mathcal{B}$.

$$\mu_n(\Gamma) \leq \mu_n(\overline{\Gamma} \cap K_\alpha) + \mu_n(K_\alpha^c)$$

$$\overline{\Gamma} \cap K_\alpha \subset \Psi_I^c(\alpha) \implies \inf_{x \in \overline{\Gamma} \cap K_\alpha} I(x) \geq \alpha$$

Combining $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha$ with

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\overline{\Gamma} \cap K_\alpha) \leq -\alpha,$$ results

then in $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\alpha$.

ad (a): Apply the lower bound to the open set $K_\alpha^c \in \mathcal{B}$, and conclude with $\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha$ that $\inf_{x\in K_\alpha^c} I(x) > \alpha$.

Therefore, $\Psi_I(\alpha) \subset K_\alpha$ yields the compactness of the closed level set $\Psi_I(\alpha)$. This argument holds for any $\alpha < \infty$, it follows that I is a good rate function. $\square$

In order to ease notation and to shorten assumptions we will from now on assume that (E, d) is a Polish space (complete separable metric space). Note that this is a restriction, further details for arbitrary topological space are studied in the book by Demlo and Zeitouni, chapter 4.

It will be later be usefull to recall the following definition.

<u>**Definition 5:**</u> Let (E, d) be a Polish space $f: E \to [-\infty, \infty]$ is lower semi-continuous if it satisfies any of the following equivalent properties

(i) $\liminf_{n\to\infty} f(x_n) \geq f(x) \quad \forall (x_n)$ with $x_n \to x$ as $\to \infty$.

(ii) $\lim_{\varepsilon \downarrow 0} \inf_{y\in B_\varepsilon(x)} f(y) = f(x)$

(iii) f has closed level sets, i.e. $\{x\in E : f(x) \leq c\}$ is closed $\forall\, c\in\mathbb{R}$.

Theorem 3: Let $(\mu_n)_{n\geq 1}$ satisfy the LDP. Then the associated rate function I is unique.

Proof: Let I and J be two rate functions for $(\mu_n)_{n\geq 1}$. We show that $I(x) = J(x)$ for all $x \in E$. Fix $x \in E$ and consider the sequence of open balls $B_N = B_{1/N}(x) = \{ y \in E : d(x,y) < 1/N \}$ of radius $1/N$ around x. Then

$$-I(x) \leq -\inf_{y\in B_{N+1}} I(y) \leq \lim_{n\to\infty} \frac{1}{n}\log \mu_n(B_{N+1}) \leq \overline{\lim_{n\to\infty}} \frac{1}{n}\log \mu_n(\overline{B_{N+1}})$$

$$\leq -\inf_{y\in \overline{B_{N+1}}} J(y) \leq -\inf_{y\in B_N} J(y),$$

where we are using $x \in B_{N+1}$ and $B_N \supset \overline{B_{N+1}}$.

Letting $N \to \infty$ and using the l.s.c. of J we conclude $\lim_{N\to\infty} \inf_{y\in B_N} J(y) = J(x)$, and hence $I(x) \geq J(x)$.

The opposite inequality follows from symmetry. □

The role of open and closed sets in the LDP is similar to their role in weak convergence of probability measures. Recall the definition of weak convergence:

<u>Definition 6:</u> A sequence $(\mu_n)_{n\geq 1}$ of probability measures μ_n on E converges weakly to the probability $\mu \in \mathcal{M}_1(E)$ as $n \to \infty$ if

- $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C) \quad \forall C \subset E$ closed.
- $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O) \quad \forall O \subset E$ open.

We can therefore view (the LDP)

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(C) \leq - \inf_{y\in C} I(y) \quad \forall C \subset E \text{ closed},$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(O) \geq - \inf_{y\in O} I(y) \quad \forall O \subset E \text{ open};$$

as analogues of weak convergence on an exponential scale.

Weak convergence of probability measures is equivalent to

$$\int_E f(x)\, \mu_n(dx) \xrightarrow[n\to\infty]{} \int_E f(x)\, \mu(dx)$$

for all $f \in \mathcal{C}_b(E)$, with $\mathcal{C}_b(E)$ the space of bounded continuous functions on E. Henceforth the LDP is ideally suited for handling convergence of integrals of exponential functionals.

The analogy of the LDP with weak convergence has been explored in detail by O'Brien and Vervaat. See the monograph by Dupuis and Ellis as well the more recent one on 'Large deviations for stochastic processes' by Feng and Kurtz.

We now formulate a theorem that enables us to generate one LDP from another one via contraction. Besides being conceptually important, this theorem will turn out to be very useful later on.

<u>Theorem 4:</u> Let $(\mu_n)_{n\geq 1}$ satisfy the LDP with rate n and rate function I. Let Y be another Polish Space, $T: E \to Y$ be a continuous map, and let $Q_n = \mu_n \circ T^{-1}$ denote the image measure.
Then $(Q_n)_{n\geq 1}$ satisfies the LDP on Y with rate n and with rate function J given by

$$J(y) = \inf_{x\in E:\, y=T(x)} I(x) \quad , \text{with } \inf_{\emptyset} I = \infty.$$

<u>Proof.</u> Since T is continuous, T^{-1} maps open sets into open sets and closed sets into closed sets.

Pick $C \subset Y$ closed and write

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(C) = \limsup_{n\to\infty} \frac{1}{n} \mu_n(T^{-1}(C))$$

$$\leq - \inf_{y \in T^{-1}(C)} I(y) = -\inf_{x \in C} \inf_{y \in T^{-1}(\{x\})} I(y)$$

$$= -\inf_{x \in C} J(x).$$ A similar argument works

for $O \subset Y$ open. It remains to prove that J is a rate function.

Clearly, $\mathcal{D}_I = \{x \in E : I(x) < \infty\} \neq \emptyset$ implies

$\mathcal{D}_J = \{y \in Y : J(y) < \infty\} \neq \emptyset$. Since I has

compact level sets, and since a continuous image of a compact set is again compact, also J has compact level sets. $\square$

<u>Example:</u> Sanov – Cramér. E be finite

$E = \mathcal{M}_1(E)$; $Y = \mathbb{R}$; $T : \mathcal{M}_1(E) \to \mathbb{R}$

$$\nu \longmapsto T(\nu) = \sum_{s \in E} s \nu_s$$

$\mu_n^{(\cdot)} = \mathbb{P}(L_n \in \cdot)$; $Q_n = \mathbb{P}(\frac{1}{n} S_n \in \cdot)$

$\forall\, a > 0 \qquad \left(m_\nu = T(\nu) = \sum_{s \in E} s \nu_s \right)$

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \in B_a^c(m_\nu) \right) = -\inf_{z \in B_a^c(m_\nu)} I(z)$$

with $I(z) = \inf_{\substack{\nu \in \mathcal{M}_1(E): \\ m_\nu = z}} I_\mu(\nu)$

$$\frac{1}{n} S_n = z \iff L_n \in \{ \nu \in \mathcal{M}_1(E) : m_\nu = z \}$$

We shall compute $\inf_{\substack{\nu \in \mathcal{M}_1(E): \\ m_\nu = z}} I_\mu(\nu)$:

The infimum is subject to the conditions $\sum_s \nu_s = 1$ and $\sum_s s \nu_s = z$. Using Lagrange multipliers c_1 and c_2, we thus need to compute

$$\inf_{(\nu_s) : \nu_s \geq 0} \sum_s \left[\nu_s \log \frac{\nu_s}{\mu_s} - c_1 \nu_s - c_2 s \nu_s \right]$$

The infimum is attained in $(0,\infty)^{|E|}$ ($x \log x$ is steep at $x = 0$ and $x \to \infty$). We get (differentiating) using $\sum \nu_s = 1$ and $c = c_2 = c_2(z) \in \mathbb{R}$

$$\nu_s = \mu_s e^{cs} \Big/ \sum_s \mu_s e^{cs}$$

Thus $I_\mu(\nu) = \sum_s \nu_s \log \frac{\nu_s}{\mu_s} = cz - \log\left(\sum_s \mu_s e^{cs} \right)$. Optimise over c to get the result.