The Large Deviation Principle

Here we investigate general properties of large deviation phenomena. For the moment we will restrict ourselves to probability measures on a *compact* metrizable space $\mathcal{K}$ (a compact Polish space). Later we will try to remove the compactness hypothesis (for example to be able to handle measures on $\mathbb{R}$!).

LD-convergence

For any $\mu \in \Pi(\mathcal{K})$ we define the semi-norm $\|f\|_{L^p(\mu)} = [\mu(|f|^p)]^{1/p}$ on $C(\mathcal{K})$. It satisfy $\|1\| = 1$,

$$|f| \leqslant |g| \Rightarrow \|f\|_{L^p(\mu)} \leqslant \|g\|_{L^p(\mu)} \tag{1}$$

and

$$f, g \geqslant 0 \Rightarrow \|\max(f, g)\|_{L^p(\mu)} \leqslant 2^{1/p} \max(\|f\|_{L^p(\mu)}, \|g\|_{L^p(\mu)}). \tag{2}$$

Exercise 1. Prove these inequalities.

Exercise 2. Fix $1 \leqslant p < +\infty$. Show that the following two statements are equivalent:

a) $\forall f \in C(\mathcal{K}), \|f\|_{L^p(\mu_n)} \to \|f\|_{L^p(\mu)}$

b) $\mu_n \to \mu$ weakly

for all $\mu_n, \mu \in \Pi(\mathcal{K})$. Convergence of L^p norms for all continuous functions is then equivalent to weak convergence. (b) $\Rightarrow$ (a) is easy. For the reverse implication use $f = |g|^p - |h|^p$.

Let us be given a sequence $(\mu_n)_{n \geqslant 1}$ for probabilities, assume that there exists a sequence of positive numbers $(p_n)_{n \geqslant 1}$ with $p_n \to +\infty$ such that

$$\lim_n \|f\|_{L^{p_n}(\mu_n)}$$

exists for all $f \in C(\mathcal{K})$. Then this limit is another semi-norm on $C(\mathcal{K})$ which we denote by $\|f\|_*$. It will satisfy $\|1\|_* = 1$, the inequality (1) and

$$\|\max(f, g)\|_* \leqslant \max(\|f\|_*, \|g\|_*)$$

for all $f, g \in C(\mathcal{K})$.

Definition 1. *A sequence $(\mu_n)_{n \geqslant 1}$ of probability measures on a compact metrizable space $\mathcal{K}$ is said* LD*-convergent if the limit*

$$\lim_n \|f\|_{L^n(\mu_n)} = \|f\|_*$$

exists for all $f \in C(\mathcal{K})$.

Theorem 2. *Let $\|\cdot\|_*$ be a semi-norm on $C(\mathcal{K})$ satisfying for all $f, g \in C(\mathcal{K})$*

$$\|1\|_* \leqslant 1, \ |f| \leqslant |g| \Rightarrow \|f\|_* \leqslant \|g\|_* \text{ and } \|\max(f, g)\|_* \leqslant \max(\|f\|_*, \|g\|_*). \tag{3}$$

Let $\Pi\colon\mathcal{K}\to[0,1]$ *such that* $1/\Pi(x)=\sup_{\|f\|_*\leqslant 1} f(x)$ *then the function* Π *is usc and*

$$\|f\|_* = \max_{x\in\mathcal{K}} [|f(x)|\Pi(x)] \tag{4}$$

(supremum is reached by upper semi-continuity).

Proof. $1/\Pi(x)$ is the supremum of continuous functions so it is lsc. Let us prove eq. (4). It is enough to consider $f\geqslant 0$. By definition $1\geqslant\Pi(x)f(x)$ for all $x\in\mathcal{K}$ and $f\geqslant 0$ such that $\|f\|_*\leqslant 1$ this implies that $\|f\|_*\geqslant\Pi(x)f(x)$ for all f and then $\|f\|_*\geqslant\sup_x\{\Pi(x)f(x)\}$. To prove the reverse inequality let $C_f=\max_{x\in\mathcal{K}}[f(x)\Pi(x)]$ so for all x, $f(x)\leqslant C_f/\Pi(x)$ and so there exist g with $\|g\|_*\leqslant 1$ such that $f(x)<C_f g(x)+\varepsilon$. By continuity of f and g this continues to hold in a neighborhood of x and by repeating the argument and by the compactness of $\mathcal{K}$ it is possible to find a finite number of functions $\{g_k\}_{k=1,\ldots,n}$ such that $\|g_k\|_*\leqslant 1$ and $f(x)<C_f\max_k g_k(x)+\varepsilon$ for all $x\in\mathcal{K}$. Now by the properties of the semi-norm we have

$$\|f\|_*\leqslant C_f\|\max_k g_k\|_*+\varepsilon\leqslant C_f\max_k\|g_k\|_*+\varepsilon\leqslant C_f+\varepsilon$$

and by the arbitrariness of ε we can conclude. □

Exercise 3. Assume that Π_1, Π_2: $\mathcal{K}\to[0,1]$ are upper semi-continuous. Prove that if $\max_{x\in\mathcal{K}}[|f(x)|\Pi_1(x)]=\max_{x\in\mathcal{K}}[|f(x)|\Pi_2(x)]$ for all $f\in C(\mathcal{K})$ then $\Pi_1=\Pi_2$. (Hint: use $f(x)=(1-Md(x,x_0))_+$ for large M assuming that $\Pi_1(x_0)<\Pi_2(x_0)$).

It is suggestive to introduce the lower semi-continuous function I: $\mathcal{K}\to\mathbb{R}_+$ such that $\Pi(x)=e^{-I(x)}$. Such a function is called a *rate function*. It defines a semi-norm on $C(\mathcal{K})$ by

$$\|f\|_I=\max_{x\in\mathcal{K}}[|f(x)|e^{-I(x)}].$$

Definition 3. *The sequence* $(\mu_n)_{n\geqslant 1}$ *satisfy a LDP (large deviation principle) with a rate function* $I\colon\mathcal{K}\to\mathbb{R}_+$ *(lower semi-continuous) if, for all* $f\in C(\mathcal{K})$*,*

$$\lim_n\|f\|_{L^n(\mu_n)}=\|f\|_I.$$

Then we have proved that

Corollary 4. *If* $(\mu_n)_{n\geqslant 1}$ *is LD-convergent then it satisfy the LDP with one and only one rate function* I *given by* $I(x)=\log\sup_{\|f\|_*\leqslant 1} f(x)$*.*

Theorem 5. *Let* $\{N_n\}_{n\geqslant 1}$ *be a sequence of norms satisfying* $N_n(1)\leqslant 1$ *and* $|f|\leqslant|g|\Rightarrow N_n(f)\leqslant N_n(g)$*. There there exists a converging subsequence* $\{N_n'\}\subseteq\{N_n\}$*, i.e. such that* $\lim_n N_n'(f)=N_\infty'(f)$ *exists for every* $f\in C(\mathcal{K})$*.*

Proof. Take a sequence N_n of such semi-norms. Note that $N_n(f)\leqslant\|f\|_\infty$ since $|f|\leqslant\|f\|_\infty$. Let $\{f_k\}_{k\geqslant 1}$ be a countable dense set in $C(\mathcal{K})$ for the sup norm. Take a subsequence $\{N_n'\}\subseteq\{N_n\}$ so that $S(f_k)=\lim_n N_n'(f_k)$ exists for every k. For every f take ℓ such that $\|f-f_\ell\|_\infty<\varepsilon$ and note that $|N_n'(f)-N_k'(f)|\leqslant|N_n'(f-f_\ell)|+|N_n'(f_\ell)-N_k'(f_\ell)|+|N_k'(f_\ell-f)|\leqslant 2\varepsilon+|N_n'(f_\ell)-N_k'(f_\ell)|\to 2\varepsilon$ for n, $k\to\infty$. The sequence $\{N_n'(f)\}$ is Cauchy for every f and we denote $N_\infty'(f)=\lim_n N'(f)$ its limit. □

Remark 6. In fact the space of all semi-norms N: $C(\mathcal{K})\to\mathbb{R}_+$ satisfying $N(1)\leqslant 1$ and $|f|\leqslant|g|\Rightarrow N(f)\leqslant N(g)$ is a compact metrizable space. (Without proof)

Exercise 4. Let $\mathcal{K} = [0, 1]$ and μ_n the Lebesgue measure for every $n \geqslant 1$. Prove that $(\mu_n)_{n \geqslant 1}$ satisfy the LDP with rate function $I(x) = 0$.

Exercise 5. Let $\mathcal{K} = [0, 1]$ and $\mu_\alpha \in \Pi(\mathcal{K})$ given by $\mu_\alpha(\mathrm{d}x) = (\alpha + 1) x^\alpha \mathrm{d}x$. Prove that each of the following sequences are LD-convergent and find the related rate functions: $(\mu_n)_n$, $(\mu_{2n})_n$, $(\mu_{n^2})_n$, $(\mu_{\sqrt{n}})_n$.

Exercise 6. Prove that if $(\mu_n)_n$ is LD-convergent with rate I, then $(\mu_{2n})_n$ is LD-convergent with rate $2I$. (Hint: $\|f\|_{L^n(\mu_{2n})} = \||f|^{1/2}\|_{L^{2n}(\mu_{2n})}$)

Exercise 7. Let $\mathcal{K} = [0, 1]$ and $(\mu_n)_n$ a sequence satisfying the LDP with rate function $I(x) = \log(1/x)$. Prove that $\mu_n([0, 0.5]) < 0.6^n$ for all n sufficiently large. (Hint: take f such that $f(x) = 1$ in $[0, 0.5]$ and $f(x) = 0$ in $[0.55, 1]$).

Exercise 8. Prove that $\inf_{\mathcal{K}} I = 0$ (Hint: try $f = 1$). Prove that for every $\varepsilon > 0$, $\mu_n(\{x \colon I(x) \leqslant \varepsilon\}) \to 1$ for $n \to \infty$. (Hint: first prove it for I continuous using the Markov inequality and $f = e^I$, then use the fact that any lsc function is the point-wise limit of an increasing sequence of continuous functions).

Back to probabilities

Lemma 7. *Let $\varphi_n, \varphi \colon \mathcal{K} \to \mathbb{R}$ and $\varphi_n \uparrow \varphi$ point-wise, then* $\sup_{\mathcal{K}} \varphi_n \uparrow \sup_{\mathcal{K}} \varphi$. *If $\varphi_n \downarrow \varphi$ and φ_n, φ are usc then* $\max_{\mathcal{K}} \varphi_n \downarrow \max_{\mathcal{K}} \varphi$.

Proof. First statement. For every $\varepsilon > 0$ take x such that $\varphi(x) > \sup_{\mathcal{K}} \varphi - \varepsilon/2$ and n such that $\varphi_n(x) > \varphi(x) - \varepsilon/2$ then $\sup_{\mathcal{K}} \varphi_n > \sup_{\mathcal{K}} \varphi - \varepsilon$. Second statement. Let $c = \lim_n \max_{\mathcal{K}} \varphi_n$. For every $\varepsilon > 0$ the sets $\{x \colon \varphi_n(x) \geqslant c - \varepsilon\}$ form a decreasing sequence of non-empty closed sets. By compactness some x belongs to all the sets, thus $\varphi(x) = \lim_n \varphi_n(x) \geqslant c - \varepsilon$ and $\max_{\mathcal{K}} \varphi > c - \varepsilon$. □

Exercise 9. Semicontinuity is necessary in the second statement of the previous lemma. Find a counter-example.

Lemma 8. *Let $f \colon \mathcal{K} \to \mathbb{R}$, $f \geqslant 0$*

a) If f is lsc then $\liminf_n \|f\|_{L^n(\mu_n)} \geqslant \sup_{\mathcal{K}} (f e^{-I})$;

b) If f is usc then $\limsup_n \|f\|_{L^n(\mu_n)} \leqslant \max_{\mathcal{K}} (f e^{-I})$.

Proof. a) Take $f_n \in C(\mathcal{K})$, $0 \leqslant f_n \uparrow f$. As $j \to \infty$

$$\liminf_n \|f\|_{L^n(\mu_n)} \geqslant \liminf_n \|f_j\|_{L^n(\mu_n)} = \max_{\mathcal{K}} (f_j e^{-I}) \uparrow \sup_{\mathcal{K}} (f e^{-I}).$$

Part b) is similar using the semi-continuity of the limit f. □

Corollary 9.

a) $\liminf_n (\mu_n(G))^{1/n} \geqslant \exp(-\inf_G I)$ *for every open set $G \subset \mathcal{K}$;*

b) $\limsup_n (\mu_n(F))^{1/n} \leqslant \exp(-\inf_F I)$ *for every closed set $F \subset \mathcal{K}$;*

c) If an open set $G \subset \mathcal{K}$ *satisfy* $\inf_G I = \inf_{\bar{G}} I$ *then*

$$\lim_n (\mu_n(G))^{1/n} = \lim_n (\mu_n(\bar{G}))^{1/n} = \exp(-\inf_G I) = \exp(-\min_{\bar{G}} I).$$

Exercise 10. Note that $\inf_G I = \inf_{\bar{G}} I$ for open G does not imply that $\mu_n(G)/\mu_n(\bar{G}) \to 1$. Find a counterexample. (Hint: try $\mathcal{K} = [0, 1]$, $G = (0, 1]$ and choose μ_n as a mixture of Lebesgue measure and an atom in 0 with appropriate weights).

Choose a metric d on $\mathcal{K}$ and consider open and closed balls $B_{x,r-} = \{y \in \mathcal{K}: d(x, y) < r\}$ and $B_{x,r+} = \{y \in \mathcal{K}: d(x, y) \leqslant r\}$. We can describe the rate function I in terms of probability decay of such balls. In particular the following holds.

Proposition 10. *For every* $x \in \mathcal{K}$ *there exists a function* $N: (0, 1) \to \mathbb{N}$ *such that*

$$\lim_{r \to 0+} \sup_{n \geqslant N(r)} \left| \frac{1}{n} \log \mu_n(B_{x,r\pm}) + I(x) \right| = 0.$$

Proof. By the probability decay rate estimates we have that there exists $N(r)$ such that for all $n \geqslant N(r)$ we have

$$\exp(-\inf_{B_{x,r-}} I) - r \leqslant (\mu_n(B_{x,r-}))^{1/n} \leqslant (\mu_n(B_{x,r-}))^{1/n} \leqslant \exp(-\inf_{B_{x,r+}} I) + r.$$

By lsc-ity of I, $\inf_{B_{x,r\pm}} I \to I(x)$ for $r \to 0+$, therefore the statement. □

In fact decay of small balls determine LD-convergence.

Proposition 11. *Assume that for every* $x \in \mathcal{K}$

$$\liminf_{r \to 0+} \liminf_n \log(\mu_n(B_{x,r-}))^{1/n} = \limsup_{r \to 0+} \limsup_n \log(\mu_n(B_{x,r+}))^{1/n} = -I(x)$$

then the sequence $(\mu_n)_n$ *obey the LDP with rate function* I*.*

Proof. By compactness we can extract a sub-sequence obeying the LDP with rate function I' but by the above proposition

$$-I(x) \leqslant \lim_{r \to 0+} \liminf_{n_k} \log(\mu_{n_k}(B_{x,r-}))^{1/n_k} = -I'(x) = \lim_{r \to 0+} \limsup_{n_k} \log(\mu_{n_k}(B_{x,r+}))^{1/n_k} \leqslant -I(x).$$

□

The Gärtner-Ellis theorem and Cramérs theorem in $\mathbb{R}^d$

Let us give a first application of the sequential compactness result above: LD-convergence needs convergence of $\|f\|_{L^n(\mu_n)}$ for all continuous functions. Here we will see that in the vector space case and with enough regularity it is enough to check convergence of exponentials of linear functionals.

Theorem 12. (Gärtner-Ellis) *Let* $\mathcal{K} \subseteq \mathbb{R}^d$ *be compact and let* X_n *be* $\mathcal{K}$*-valued r.v.s such that for all* $\lambda \in \mathbb{R}^d$

$$\lim_n \frac{1}{n} \log \mathbb{E}[e^{n \langle \lambda, X_n \rangle}] = G(\lambda)$$

where $G: \mathbb{R}^d \to \mathbb{R}_+$ *is a differentiable (convex) function. Then* $\{X_n\}_{n\geqslant 1}$ *obey the LDP with the (convex) rate function* $I(x) = \sup_{\lambda \in \mathbb{R}^d} [\langle \lambda, x\rangle - G(\lambda)]$.

Proof. Given the existence of the limit above, convexity of G is just a consequence of Hölder inequality. By compactness there exists a subsequence $\{X_{n_k}\}$ which satisfy the LDP for some rate function $I': \mathbb{R}^d \to \mathbb{R}_+$ which can be taken equal to $+\infty$ outside $\mathcal{K}$. Then by definition

$$G(\lambda) = \lim_k \frac{1}{n_k} \log \mathbb{E}[e^{n_k \langle \lambda, X_{n_k}\rangle}] = \sup_{x \in \mathcal{K}} (\langle \lambda, x\rangle - I'(x)) = \sup_{x \in \mathbb{R}^d} (\langle \lambda, x\rangle - I'(x))$$

To conclude that $I = I'$ we need to prove that the Fenchel-Legendre transform is invertible at G. Our hypothesis guarantees that for each x there exists $\lambda_x \in \mathbb{R}^d$ such that $I(x) = \langle \lambda_x, x\rangle - G(\lambda_x)$ and that $I(y) > I(x) + \langle \lambda_x, y - x\rangle$ for all $y \neq x$ (prove it!). Then for all $y \in \mathbb{R}^d$, $G(\lambda_x) \geqslant \langle \lambda_x, y\rangle - I'(y)$ so that $I'(x) \geqslant \langle \lambda_x, x\rangle - G(\lambda_x) = I(x)$. On the other hand, by compactness, there exists $\hat{y} \in \mathcal{K}$ such that, if $\hat{y} \neq x$,

$$\langle \lambda_x, \hat{y} - x\rangle - I'(\hat{y}) = G(\lambda_x) - \langle \lambda_x, x\rangle = -I(x) > -I(\hat{y}) + \langle \lambda_x, \hat{y} - x\rangle$$

which means that $I'(\hat{y}) < I(\hat{y})$, in contradiction to the fact that $I' \geqslant I$ on $\mathcal{K}$. So we must have $x = \hat{y}$ but then $I(x) = \langle \lambda_x, x\rangle - G(\lambda_x) = I'(x)$ which concludes the proof. □

Of course a basic corollary is the multidimensional Cramér theorem.

Corollary 13. (Cramér-Varadhan) *Let* $(X_n)_{n\geqslant 1}$ *be an iid sequence with values in the compact* $\mathcal{K} \subseteq \mathbb{R}^d$. *Let* $\Lambda(\lambda) = \log \mathbb{E}[\exp\langle \lambda, X_1\rangle]$, $I(x) = \sup_\lambda [\langle \lambda, x\rangle - \Lambda(\lambda)]$ *and* μ_n *be the law of the empirical mean* $S_n = (X_1 + \cdots + X_n)/n$. *Then the sequence* $(\mu_n)_n$ *satisfy the LDP with the convex rate function* I.

Proof. Exercise. You need only to verify that $G(\lambda) = \Lambda(\lambda)$ and justify its smoothness. □

The rate function is the Γ-limit of relative entropy

Given a probability measure μ on the compact $\mathcal{K}$ we can define the relative entropy $H_\mu: \Pi(\mathcal{K}) \to [0, +\infty]$ as

$$H_\mu(\nu) = \sup_\varphi [\nu(\varphi) - \log \mu(e^\varphi)]$$

where the supremum is taken wrt all continuous functions on $\mathcal{K}$. Then

$$\mathrm{H}_\mu(\nu) = \begin{cases} \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\mathrm{d}\mu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

The function H_μ is strictly convex and $\mathrm{H}_\mu(\nu) = 0$ iff $\nu = \mu$, moreover

$$\log \mu(e^\varphi) = \sup_\nu [\nu(\varphi) - H_\mu(\nu)]$$

where the supremum runs over all probability measures on $\mathcal{K}$. Consider a sequence $(\mu_n)_n$ and define $H_n(\nu) = H_{\mu_n}(\nu)/n$, then $H_n(\nu) = \sup_\varphi [\nu(\varphi) - \log \|e^\varphi\|_{L^n(\mu_n)}]$ and

$$\log \|e^\varphi\|_{L^n(\mu_n)} = \sup_{\nu \in \Pi(\mathcal{K})} [\nu(\varphi) - H_n(\nu)]$$

Assume that $(\mu_n)_n$ obey the LDP with rate I, then we have

$$\lim_n \log \|e^{\varphi}\|_{L^n(\mu_n)} = \sup_{x \in \mathcal{K}} (\varphi(x) - I(x)) = \sup_{\nu \in \Pi(\mathcal{K})} [\nu(\varphi) - \nu(I)]$$

where we introduce a maximization over probabilities to stress the similarity with the variational representation at finite n.

Rate function can be seen as asymptotic entropy in Γ-convergence sense. Γ-convergence is a variational convergence introduced by E. de Giorgi. In a metric space $\mathcal{M}$ the Γ-convergence of a sequence of positive functionals $I_n: \mathcal{M} \to \mathbb{R}_+$ is defined via

$$\Gamma\liminf_n \; I_n(x) = \inf \{\liminf_n \; I_n(x_n) : x_n \to x\}$$

and

$$\Gamma\limsup_n I_n(x) = \inf \{\limsup_n I_n(x_n) : x_n \to x\}$$

When $\Gamma\limsup I_n = \Gamma\liminf I_n$ we denote the common limit as $\Gamma\lim I_n$. The Γ limits are lsc functions. Here, as before, we restrict ourselves to a compact space $\mathcal{M}$. We will be interested in the case where $\mathcal{M} = \Pi(\mathcal{K})$ for some compact metrizable $\mathcal{K}$.

Theorem 14. *For every $J: \mathcal{M} \to \mathbb{R}$ and $x \in \mathcal{M}$*

a) $J(x) \leqslant \Gamma\liminf_n \; I_n(x) \Leftrightarrow \forall x_n \to x$: $\liminf_n I_n(x_n) \geqslant J(x)$ *;*

b) $\Gamma\limsup_n \; I_n(x) \leqslant J(x) \Leftrightarrow \exists x_n \to x : \limsup_n I_n(x_n) \leqslant J(x)$ *.*

$\Gamma\liminf$ *and* $\Gamma\limsup$ *are respectively the largest and smallest lsc functions satisfying the above properties.*

Then $\Gamma\lim_n I_n = I$ if and only if for all x_n, x such that $x_n \to x$ we have $\liminf_n I_n(x_n) \geqslant I(x)$ and moreover for every x there exists $x_n \to x$ such that $\limsup_n I_n(x_n) \leqslant I(x)$.

Theorem 15. *Properties of Γ-convergence :*

a) $\min_{\mathcal{M}} I = \lim_n (\min_{\mathcal{M}} \; I_n)$

b) If $I_n \xrightarrow{\Gamma} I$, x_n minimizer of I_n and $x_n \to x \;\Rightarrow\; x$ minimizer of I

c) If G is continuous and $I_n \xrightarrow{\Gamma} I$ then $I_n + G \xrightarrow{\Gamma} I + G$

d) If $\mathcal{M}$ is separable from every sequence I_n we can extract a Γ-converging subsequence.

Example 16. Let $\mathcal{M} = [0, 1]$ and $I_n(x) = \sin^2(2\pi x)$ then $\Gamma\lim_n I_n(x) = 0$ for all $x \in \mathcal{M}$.

Theorem 17. (μ_n) *obey the LDP with rate function I iff* $\Gamma\lim_n H_n = I$.

Proof. Let us prove that LDP $\Rightarrow$ Γ-convergence. By the variational characterization $H_n(\nu_n) \geqslant \nu_n(\varphi) - \log \|e^{\varphi}\|_n$ for every $\varphi \in C(\mathcal{K})$. If $\nu_n \to \nu$ then, by LD convergence, $\liminf_n \; H_n(\nu_n) \geqslant \nu(\varphi) - \sup_{\mathcal{K}} (\varphi - I)$. Since I is lsc there exists a sequence of continuous functions $\varphi_n \uparrow I$. Then

$$\liminf_n \; H_n(\nu_n) \geqslant \lim_n [\nu(\varphi_n) - \sup_{\mathcal{K}} (\varphi_n - I)] = \nu(I)$$

since $\sup_{\mathcal{K}} (\varphi_n - I) \to 0$. Now for every φ there exists $(\nu_n)_n$ such that $\|e^{\varphi}\|_n = \nu_n(\varphi) - H_n(\nu_n)$. By compactness we can take a subsequence still denoted ν_n such that $\nu_n \to \nu$ for some ν. By LD-convergence we get $\lim_n H_n(\nu_n) = \nu(\varphi) - \sup_{\mathcal{K}} (\varphi - I)$. But by the previous bound $\nu(\varphi) - \sup_{\mathcal{K}} (\varphi - I) \geqslant \nu(I)$ which implies that $\nu(\varphi - I) \geqslant \sup_{\mathcal{K}} (\varphi - I)$, so the limit point ν must be concentrated on the maxima of $\varphi - I$. Take φ such that this function has a unique maximum at x to obtain that for every $x \in \mathcal{K}$ there exists $\nu_n \to \delta_x$ for which $\lim_n H_n(\nu_n) \leqslant I(x)$. By convexity, for all $\nu \in \Pi(\mathcal{K})$, $\Gamma\mathrm{limsup}_n H_n(\nu) = \Gamma\mathrm{limsup}_n H_n(\int \delta_y \nu(\mathrm{d}y)) \leqslant \int \nu(\mathrm{d}y) \Gamma\mathrm{limsup}_n H_n(\delta_y) \leqslant \nu(I)$. The reverse implication is a direct consequence of the variational properties of Γ-convergence:

$$\sup_{\nu} (\nu(\varphi) - \nu(I)) = \lim_n \sup_{\nu} (\nu(\varphi) - H_n(\nu)) = \lim_n \log \|e^{\varphi}\|_n$$

for every $\varphi \in C(\mathcal{K})$ which by density implies LD-convergence with rate function I. □

Remark 18. Compactness of LD-convergence can be obtained as consequence of the compactness of Γ-convergence.

Properties of LD-converging sequences

Theorem 19. *(*CONTRACTION PRINCIPLE*) Let $\mathcal{K}'$ be another compact metrizable space and $F\colon \mathcal{K} \to \mathcal{K}'$ be a continuous function, $(\mu_n)_n$ a sequence of measures on $\mathcal{K}$ satisfying the LDP with rate function I, and $\nu_n = F_* \mu_n \in \Pi(\mathcal{K}')$ the image measure of μ_n wrt. F, i.e. $\nu_n(A) = \mu_n(F^{-1}(A))$ for all Borel sets $A \subseteq \mathcal{K}'$. Then $(\nu_n)_n$ satisfies the LDP with rate function*

$$I'(y) = \min_{F^{-1}(y)} I = \min \{I(x) : x \in F^{-1}(\{y\})\}$$

where the minimum is by definition $+\infty$ if $F^{-1}(\{y\}) = \varnothing$. Otherwise the minimum is attained since $F^{-1}(\{y\})$ is compact and I is lsc.

Proof. Exercise. Use the fact that for all $g \in C(\mathcal{K}')$ the change of variables formula gives $\|g\|_{L^n(\nu_n)} = \|g \circ F\|_{L^n(\mu_n)}$. □

Actually, in the contraction principle only continuity of F at the points where $I < +\infty$ matters:

Theorem 20. *(*CONTRACTION PRINCIPLE II*) Let $F\colon \mathcal{K} \to \mathcal{K}'$ be a function which is continuous on $\mathcal{K}_0 \subseteq \mathcal{K}$. Let $(\mu_n)_n \subseteq \Pi(\mathcal{K})$ obey the LDP with rate function I such that $I(x) = +\infty$ for every $x \notin \mathcal{K}_0$. Then $\nu_n = F_* \mu_n$ satisfy the LDP with rate function $I'(y) = \min_{F^{-1}(y)} I$.*

Proof. Take a function $g > 0$ continuous on $\mathcal{K}_0$ and consider the functions $f = g \circ F$ and

$$(f)_k^+(x) = \sup_{x' \in \mathcal{K}} [e^{-kd(x,x')} f(x')], \qquad (f)_k^-(x) = \inf_{x' \in \mathcal{K}} [e^{kd(x,x')} f(x')]$$

The functions $f_k^{\pm}$ are continuous, $f_k^+ \geqslant f \geqslant f_k^-$ and $f_k^+ \downarrow f$, $f_k^- \uparrow f$ point-wise on $\mathcal{K}_0$ as $k \to \infty$ (check it: use the triangular inequality for d). We have $\|(f)_k^-\|_{L^n(\mu_n)} \leqslant \|g \circ F\|_{L^n(\nu_n)} \leqslant \|(f)_k^+\|_{L^n(\mu_n)}$ so it remains to prove that

$$\sup_{\varepsilon} \sup_{\mathcal{K}} [(f)_k^- e^{-I}] = \inf_{\varepsilon} \sup_{\mathcal{K}} [(f)_k^+ e^{-I}] = \sup_{\mathcal{K}_0} [(g \circ F) e^{-I'}] = \sup_{\mathcal{K}'} [g e^{-I'}]$$

to get existence of the limit and the identification of the rate function I'. Now you can check that by Lemma 7 we have $\sup_{\mathcal{K}} [(f)_k^+ e^{-I}] = \sup_{\mathcal{K}_0} [(f)_k^+ e^{-I}] \downarrow \sup_{\mathcal{K}_0} [e^{-I} f]$ since the functions $(f)_k^+$, f are continuous on $\mathcal{K}_0$ and $\sup_{\mathcal{K}} [(f)_{\varepsilon}^- e^{-I}] = \sup_{\mathcal{K}_0} [(f)_{\varepsilon}^- e^{-I}] \uparrow \sup_{\mathcal{K}_0} [f e^{-I}]$ by the (increasing) pointwise convergence of $(f)_k^-$ towards f. □

Theorem 21. *(*Change of measure*) Let $(\mu_n)_n$ and $(\nu_n)_n$ two sequences of probabilities on $\mathcal{K}$ such that, for all n,*

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n} = c_n \, e^{-n h}$$

for some $h \in C(\mathcal{K})$ and $(c_n)_n$. Then if $(\mu_n)_n$ obey the LDP with rate function I then $(\nu_n)_n$ obey the LDP with rate function J given by

$$J = I + h - \min_{\mathcal{K}} (I + h) = I + h - \lim_n \frac{\log c_n}{n}.$$

Proof. Exercise. Use the fact that $\|f\|_{L^n(\nu_n)} = \|f e^{-h}\|_{L^n(\mu_n)} / \|1\|_{L^n(\mu_n)}$. □