Basics of the Theory of Large Deviations

Michael Scheutzow

February 1, 2018

Abstract

In this note, we collect basic results of the theory of large deviations. Missing proofs can be found in the monograph [1].

1 Introduction

We start by recalling the following computation which was done in the course *Wahrscheinlichkeitstheorie I* (and which is also done in the course *Versicherungsmathematik*).

Assume that $X,\ X_1,\ X_2, \ldots$ are i.i.d real-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $x \in \mathbb{R}$, $\lambda > 0$. Then, by Markov's inequality,

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \geq nx\Big) \leq \exp\{-\lambda nx\}\Big(\mathbb{E}\big(\mathrm{e}^{\lambda X_1}\big)\Big)^n = \exp\{-n(\lambda x - \Lambda(\lambda))\},$$

where $\Lambda(\lambda) := \log \mathbb{E} \exp\{\lambda X\}$.

Defining $I(x) := \sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\}$, we therefore get

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \geq nx\Big) \leq \exp\{-nI(x)\},$$

which often turns out to be a good bound.

2 Cramer's theorem for real-valued random variables

Definition 2.1. a) Let X be a real-valued random variable. The function $\Lambda : \mathbb{R} \to (-\infty, \infty]$ defined by

$$\Lambda(\lambda) := \log \mathbb{E}\mathrm{e}^{\lambda X}$$

is called *cumulant generating function* or *logarithmic moment generating function*. Let $D_\Lambda := \{\lambda : \Lambda(\lambda) < \infty\}$.

b) $\Lambda^* : \mathbb{R} \to [0, \infty]$ defined by

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

is called *Fenchel-Legendre transform* of Λ. Let $D_{\Lambda^*} := \{\lambda : \Lambda^*(\lambda) < \infty\}$.

In the following we will often use the convenient abbreviation

$$\Lambda^*(F) := \inf_{x \in F} \Lambda^*(x)$$

for a subset $F \subseteq \mathbb{R}$ (with the usual convention that the infimum of the empty set is $+\infty$).

Lemma 2.2. a) *Λ is convex.*

b) *Λ^* is convex.*

c) *Λ^* is lower semi-continuous, i. e. $\{x \in E : \Lambda^*(x) \leq \alpha\}$ is closed for every $\alpha \in \mathbb{R}$.*

d) *If $D_\Lambda = \{0\}$, then we have $\Lambda^* \equiv 0$.*

e) *If $\Lambda(\lambda) < \infty$ for some $\lambda > 0$, then we have $\mathbb{E}X < \infty$ (but possibly $\mathbb{E}X = -\infty$).*

f) *If $\mathbb{E}X < \infty$ (but possibly $\mathbb{E}X = -\infty$), then we have*

$$\Lambda^*(x) := \sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\}, \; x \geq \mathbb{E}X$$

and Λ^* is nondecreasing on $[\mathbb{E}X, \infty)$.

g) $\mathbb{E}|X| < \infty$ *implies* $\Lambda^*(\mathbb{E}X) = 0$.

h) $\inf_{x\in\mathbb{R}} \Lambda^*(x) = 0$.

i) Λ *is differentiable in the interior of* D_Λ *with derivative*

$$\Lambda'(\eta) = \frac{1}{\mathbb{E}\exp(\eta X)}\mathbb{E}(X\mathrm{e}^{\eta X}).$$

Further, $\Lambda'(\eta) = y$ *implies* $\Lambda^*(y) = \eta y - \Lambda(\eta)$.

Proof. [1] □

Examples 2.3. a) $\mathcal{L}(X) = \mathrm{Poisson}(\theta)$, $\theta > 0$. Then

$$\Lambda^*(x) = -x + \theta + x\log\left(\frac{x}{\theta}\right), \quad x \geq 0.$$

b) $\mathcal{L}(X) = \mathrm{Bernoulli}(p)$, $0 < p < 1$. Then

$$\Lambda^*(x) = x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}, \quad 0 \leq x \leq 1.$$

c) $\mathcal{L}(X) = \mathrm{Exp}(\theta)$, $\theta > 0$. Then

$$\Lambda^*(x) = \theta x - 1 - \log(\theta x), \quad x \geq 0.$$

d) $\mathcal{L}(X) = \mathcal{N}(0,\sigma^2)$, $\sigma > 0$. Then

$$\Lambda^*(x) = \frac{x^2}{2\sigma^2}, \quad x \in \mathbb{R}.$$

In all cases $\Lambda^*(x)$ is ∞ for all other values of x.

Now we are ready to formulate and prove Cramér's Theorem.

Theorem 2.4. *Let* $X_1, X_2, \ldots$ *be a sequence of independent and identically distributed real-valued random variables and let* $\mu_n := \mathcal{L}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)$, $n \in \mathbb{N}$. *Then the sequence* $\{\mu_n\}_{n\in\mathbb{N}}$ *satisfies the following properties.*

a) $\limsup_{n\to\infty} \frac{1}{n}\log\mu_n(F) \leq -\Lambda^*(F)$ *for every closed set* $F \subseteq \mathbb{R}$.

b) $\liminf_{n\to\infty} \frac{1}{n}\log\mu_n(G) \geq -\Lambda^*(G)$ *for every open set* $G \subseteq \mathbb{R}$.

c) *For a closed set $F \subseteq \mathbb{R}$ we even have $\mu_n(F) \leq 2\exp\{-n\Lambda^*(F)\}$ for every $n \in \mathbb{N}$ and for a closed interval F of $\mathbb{R}$ we even have $\mu_n(F) \leq \exp\{-n\Lambda^*(F)\}$ for every $n \in \mathbb{N}$.*

Proof. Obviously c) implies a), so it suffices to prove c) and b). We always assume that X is a random variable with law $\mu := \mu_1$.

c) The assertions are clearly true in case $F = \emptyset$, so we assume that F is closed and nonempty. The assertions are also clear in case $\Lambda^*(F) = \inf_{x\in F} \Lambda^*(x) = 0$, so we assume $\Lambda^*(F) > 0$. It follows from Lemma 2.2d) that there exists some $\bar{\lambda} \neq 0$ such that $\Lambda(\bar{\lambda}) < \infty$. Assume that $\bar{\lambda} > 0$ (the case $\bar{\lambda} < 0$ is treated analogously). Lemma 2.2e) shows that $\mathbb{E}X < \infty$. For $\lambda \geq 0$ and $x \in \mathbb{R}$ we get:

$$\begin{aligned}
\mu_n([x,\infty)) &= \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i \geq x\right\} \\
&= \mathbb{P}\left\{\exp\left(\lambda\sum_{i=1}^{n} X_i\right) \geq \exp(\lambda n x)\right\} \\
&\leq \exp(-\lambda n x)\,\mathbb{E}\left(\exp\left(\lambda\sum_{i=1}^{n} X_i\right)\right) \\
&= \mathrm{e}^{-n(\lambda x - \Lambda(\lambda))}.
\end{aligned}$$

Since $\mathbb{E}X < \infty$, Lemma 2.2f) shows that for $x \geq \mathbb{E}X$ we have

$$\mu_n([x,\infty)) \leq \mathrm{e}^{-n\Lambda^*(x)}. \tag{1}$$

Case 1: $\mathbb{E}|X| < \infty$ (i. e. $\mathbb{E}X > -\infty$). Since $\Lambda^*(F) > 0$, Lemma 2.2g) implies $\mathbb{E}X \in F^c$. Let (x_-, x_+) be the largest interval in F^c which contains $\mathbb{E}X$. Since F is nonempty, at least one of the numbers x_-, x_+ is finite. If x_+ is finite, then $x_+ \in F$ and

$$\mu_n(F \cap [x_+,\infty)) \leq \mu_n([x_+,\infty)) \leq \exp\{-n\Lambda^*(x_+)\} \leq \exp\{-n\Lambda^*(F)\}.$$

If $x_- > -\infty$, then $x_- \in F$ and

$$\mu_n(F \cap (-\infty, x_-]) \leq \mu_n(-\infty, x_-]) \leq \exp\{-n\Lambda^*(x_-)\} \leq \exp\{-n\Lambda^*(F)\}.$$

Hence $\mu_n(F) \leq 2\exp\{-n\Lambda^*(F)\}$. In case F is an interval either $x_- = -\infty$ or $x_+ = \infty$.

Case 2: $\mathbb{E}X = -\infty$. Lemma 2.2f) shows that the function $x \mapsto \Lambda^*(x)$ is nondecreasing and Lemma 2.2h) implies that $\lim_{x\to-\infty} \Lambda^*(x) = 0$. Since $\Lambda^*(F) > 0$ and $F \neq \emptyset$, there exists some $x_+ \in \mathbb{R}$ such that $F \subseteq [x_+, \infty)$ and $x_+ \in F$. Now (1) implies

$$\mu_n(F) \leq \mu_n([x_+, \infty)) \leq \exp\{-n\Lambda^*(x_+)\} \leq \exp\{-n\Lambda^*(F)\}.$$

This proves part c).

b) Below we will show, that for every $\delta > 0$ (and every law μ) we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n((-\delta, \delta)) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) \, (= -\Lambda^*(0)) \, . \tag{2}$$

Assume this holds and let $x \in \mathbb{R}$ and $Y := X - x$. Then we have

$$\Lambda_Y(\lambda) = \log \mathbb{E}e^{\lambda Y} = -\lambda x + \Lambda(\lambda)$$

and

$$\Lambda_Y^*(z) = \sup \{\lambda z - \Lambda_Y(\lambda)\} = \sup \{\lambda z + \lambda x - \Lambda(\lambda)\} = \Lambda^*(z + x).$$

Using (2), this implies that for any $x \in \mathbb{R}$ and $\delta > 0$, we have

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \mu_n((x - \delta, x + \delta)) &= \liminf_{n\to\infty} \frac{1}{n} \log \mu_n^{(Y)}((-\delta, \delta)) \\ &\geq -\Lambda_Y^*(0) = -\Lambda^*(x). \end{aligned}$$

If $G = \emptyset$, then assertion b) is obvious. So assume that G is open and nonempty and $x \in G$. Then we have $(x - \delta, x + \delta) \subset G$ for some $\delta > 0$ and hence

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n((x - \delta, x + \delta)) \geq -\Lambda^*(x),$$

and – since $x \in G$ was arbitrary – we have

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

It remains to show (2).

Case 1: $\mu((-\infty, 0)) > 0$, $\mu((0, \infty)) > 0$ and μ has compact support.

Since μ has compact support, there exists some $a > 0$ such that $\mu([-a, a]) = 1$. Further, $\Lambda(\lambda) \leq |\lambda| a < \infty$ for every $\lambda \in \mathbb{R}$. For the rest of the proof, see [1].

Case 2: $\mu((-\infty, 0)) > 0$, $\mu((0, \infty)) > 0$ and μ has unbounded support. Let M be so large that both $\mu([-M, 0))$ and $\mu((0, M])$ are strictly positive. Below we will let M tend to infinity. Define the probability measure ν on the Borel sets of $\mathbb{R}$ by

$$\nu(A) := \frac{\mu(A \cap [-M, M])}{\mu([-M, M])}.$$

Clearly ν satisfies the assumptions of Case 1. Defining ν_n in analogy to μ_n and letting Λ_M denote the logarithmic moment generating function associated to ν and defining

$$\Lambda^{(M)}(\lambda) := \log \int_{-M}^{M} \mathrm{e}^{\lambda y} d\mu(y),$$

we get

$$\mu_n((-\delta, \delta)) \geq \nu_n((-\delta, \delta)) \mu([-M, M])^n$$

and

$$\begin{aligned} \liminf \frac{1}{n} \log \mu_n((-\delta, \delta)) &\geq \liminf \frac{1}{n} \log \nu_n((-\delta, \delta)) + \log \mu([-M, M]) \\ &\geq \Lambda_M(\lambda) + \log \mu([-M, M]) = \Lambda^{(M)}(\lambda). \end{aligned}$$

It suffices to prove

$$I^* := \lim_{M \to \infty} \inf_{\lambda \in \mathbb{R}} \Lambda^{(M)}(\lambda) \geq \inf_{\lambda \in \mathbb{R}} \Lambda(\lambda). \tag{3}$$

Since $M \mapsto \inf_{\lambda \in \mathbb{R}} \Lambda^{(M)}(\lambda)$ is nondecreasing, the sets

$$\{\lambda : \Lambda^{(M)}(\lambda) \leq I^*\}$$

are nonempty, compact and decreasing in M, so the intersection of all these sets is nonempty. If λ_0 is in the intersection, then

$$\Lambda(\lambda_0) = \lim_{M \to \infty} \Lambda^{(M)}(\lambda_0) \leq I^*.$$

This finishes Case 2.

Case 3: Either $\mu((-\infty,0)) = 0$ or $\mu((0,\infty)) = 0$. In this case, $\lambda \mapsto \Lambda(\lambda)$ is either nonincreasing or nondecreasing and $\inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = \log\mu(\{0\})$. Therefore

$$\mu_n((-\delta,\delta)) \geq \mu_n(\{0\}) = (\mu(\{0\}))^n ,$$

and hence

$$\frac{1}{n}\log\mu_n((-\delta,\delta)) \geq \log\mu(\{0\}) = \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda).$$

This proves (2) and hence Cramer's theorem.

□

3 Basic concepts of the theory of large deviations

In the following E denotes a topological space and $\mathcal{E}$ the Borel sets of E.

Definition 3.1. $I : E \to [0,\infty]$ is called a *rate function*, in case I is lower semi-continuous (i. e. $\{x \in E : I(x) \leq \alpha\}$ is closed for every $\alpha \geq 0$). I is called a *good rate function* if – in addition – $\{x \in E : I(x) \leq \alpha\}$ is compact for every $\alpha \geq 0$.

Again we will use the abbreviation $I(G) := \inf\{I(x);\, x \in G\}$ for any subset G of E.

Definition 3.2. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a family of probability measures on $(E,\mathcal{E})$ and let I be a rate function. $\{\mu_n\}_{n\in\mathbb{N}}$ is said to satisfy a *large deviation principle* (LDP) with rate function I, if

a) $\liminf_{n\to\infty} \frac{1}{n}\log\mu_n(G) \geq -I(G)$ for every open set G

b) $\limsup_{n\to\infty} \frac{1}{n}\log\mu_n(F) \leq -I(F)$ for every closed set F.

$\{\mu_n\}_{n\in\mathbb{N}}$ is said to satisfy a *weak large deviation principle* with rate function I, if a) holds and b) holds with "closed" replaced by "compact".

Remark 3.3. Cramér's Theorem says that the sequence of the laws μ_n of the average of n i. i. d. random variables satisfies an LDP with rate function Λ^*. Λ^* may or may not be a good rate function (depending on the law of the X_i).

Remark 3.4. If the topological space E is Hausdorff, then every compact set is closed and hence a Borel set. If E is not Hausdorff, then a compact set need not be Borel and that causes a problem when formulating a weak LDP. One way out is to replace "compact" by "compact and closed". Further below, we will assume that E is Hausdorff and therefore this will not be a problem for us.

Remark 3.5. Why do we require that a rate function I be lower semi-continuous? Well, assume that I is any function (not necessarily lower semi-continuous) from E to $[-\infty, \infty]$ such that $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies a (weak) large deviation principle with function I. For $x \in E$, define

$$\widetilde{I}(x) := \sup_{x\in G,\, G\,\text{open}} I(G).$$

Then $\widetilde{I}(x) \leq I(x)$, since $\inf_{y\in G} I(y) \leq I(x)$ for every open set G containing x. Further, if G is an open set containing x, then $\widetilde{I}(x) \geq I(G)$ which implies $\widetilde{I}(G) \geq I(G)$, so we have $\widetilde{I}(G) = I(G)$ for every open set G containing x. Therefore $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies a (weak) large deviation principle with function $\widetilde{I}$ as well. Furthermore $\widetilde{I}$ is lower semi-continuous: fix $\alpha \in \mathbb{R}$; we show that $\{x \in E : \widetilde{I}(x) > \alpha\}$ is open. Let $x \in E$ satisfy $\widetilde{I}(x) = \beta > \alpha$. By definition of $\widetilde{I}$ there exists an open set G containing x such that $I(G) > \alpha$ and therefore $\widetilde{I}(G) > \alpha$ showing that $\widetilde{I}$ is a rate function. $\widetilde{I}$ is called the *lower semi-continuous regularization* of I.

Remark 3.6. Suppose $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies an LDP with rate functions I and I'. Is it then true that $I = I'$, i. e. is the rate function uniquely determined by $\{\mu_n\}_{n\in\mathbb{N}}$? For any "reasonable" topological space E this is true. It is more than enough to assume that E is a metric space. So, let us suppose that E is a metric space and $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies a weak LDP with rate function I. Fix $x \in E$. Then for any open set G containing x we have

$$-I(x) \leq -I(G) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G). \tag{4}$$

On the other hand, for any $\varepsilon > 0$ we find an open set G_1 containing x such that $I(G_1) > I(x) - \varepsilon$ since I is lower semi-continuous. Since E is a metric space we can find another open set G_2 containing x such that $\bar{G}_2$ (the closure of G_2) is contained in G_1. Therefore

$$\begin{aligned} -I(x) &\geq -I(\bar{G}_2) - \varepsilon \geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\bar{G}_2) - \varepsilon \\ &\geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(G_2) - \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary this shows – together with (4) for $G = G_2$ – that $I(x)$ is uniquely determined by the family $\{\mu_n\}_{n\in\mathbb{N}}$. The argument goes through for topological spaces which are *regular* (see [1]).

The example below shows that uniqueness does not hold on every topological space.

Example 3.7. Let $E := \{a, b\}$ and let $\mathcal{T} = \{\emptyset, \{a\}, E\}$ be a topology on E, $(E, \mathcal{T})$ is called *Sierpinski space*. Observe that $(E, \mathcal{T})$ is not a Hausdorff space. Let $\mu_n := \delta_a$ be a unit point mass at a for every n. Then $\{\mu_n\}_{n\in\mathbb{N}}$ satifies a LDP with (good) rate function I for *any* function $I : E \to [0, \infty]$ which satisfies $I(b) = 0$, i. e. $I(a)$ can be chosen arbitrarily in $[0, \infty]$.

From now on we will assume that the space $(E, \mathcal{T})$ is Hausdorff. This guarantees in particular that every compact set is closed and hence measurable.

Theorem 3.8. (Contraction Principle) *Let $(E_1, \mathcal{T}_1)$ and $(E_2, \mathcal{T}_2)$ be (Hausdorff) topological spaces and let $T : E_1 \to E_2$ be continuous. Further assume that $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies an LDP with good rate function I_1 on $(E_1, \mathcal{T}_1)$. Define $I_2 : E_2 \to [0, \infty]$ by*

$$I_2(y) := \inf \{I_1(x) : x \in E_1, T(x) = y\} .$$

Then $\{\mu_n \circ T^{-1}\}_{n\in\mathbb{N}}$ satisfies an LDP with good rate function I_2.

Proof. We first check that I_2 is a good rate function. Let $\alpha \geq 0$. Then

$$\{y : I_2(y) \leq \alpha\} = T(\{x : I_1(x) \leq \alpha\});$$

"$\supseteq$" is clear and "$\subseteq$" follows since I_1 is good, which implies that the infimum in the definition of I_2 is attained whenever $I_2(y) < \infty$. Since the continuous map T maps compact subsets of E_1 to compact subsets of E_2, it follows that I_2 is a good rate function (we don't need to check lower semi-continuity since we assumed E_2 to be Hausdorff).

Observe that for any $A \subseteq \mathbb{E}_2$, we have

$$\inf_{y\in A} I_2(y) = \inf_{x\in T^{-1}(A)} I_1(x).$$

Since T is continuous, $T^{-1}(A)$ is open (resp. closed) if A is open (resp. closed). Therefore the LDP for $\{\mu_n \circ T^{-1}\}_{n\in\mathbb{N}}$ follows as a consequence of the LDP for $\{\mu_n\}_{n\in\mathbb{N}}$. □

Remark 3.9. If I_1 is a rate function but not a good one, then I_2 as defined in the previous theorem need not even be a rate function. As an example, take $E_1 = E_2 = \mathbb{R}$, $I_1 \equiv 0$ and $T(x) = \exp(x)$.

Definition 3.10. $\{\mu_n\}_{n\in\mathbb{N}}$ is called *exponentially tight*, if for every $0 < \alpha < \infty$ there exists a compact set K_α such that

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

Remark 3.11. If $\{\mu_n\}_{n\in\mathbb{N}}$ is exponentially tight and E is Polish, then $\{\mu_n\}_{n\in\mathbb{N}}$ is tight. To see this, pick $0 < \varepsilon < 1$ and define $\alpha := -\log \varepsilon$. By assumption there exists a compact set K such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K^c) < -\alpha = \log \varepsilon,$$

so there exists some n_0 such that for all $n \geq n_0$ we have

$$\mu_n(K^c) < \exp\{-\alpha n\} = \varepsilon^n < \varepsilon.$$

Since E is Polish, a single probability measure (and hence every finite set of probability measures) is always tight and therefore the family $\{\mu_n\}_{n\in\mathbb{N}}$ is tight.

Lemma 3.12. *Assume that $\{\mu_n\}_{n\in\mathbb{N}}$ is exponentially tight and that I is a rate function.*

a) *If*

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

for every compact set F, then the same is true for every closed set F (this is even true without the assumption that I is a rate function).

b) *If*

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

for every open set G, then I is a good rate function.

So: If $\{\mu_n\}_{n\in\mathbb{N}}$ is exponentially tight and satisfies a weak LDP with rate function I, then $\{\mu_n\}_{n\in\mathbb{N}}$ even satisfies an LDP and I is good.

Proof. a) Let F be a closed set in E and $\alpha := \inf_{x\in F} I(x)$. First assume that $\alpha < \infty$. Let K_α be as in the definition of exponential tightness. Clearly

$$F \cap K_\alpha \subseteq F \subseteq \{I \geq \alpha\}$$

and

$$\mu_n(F) \leq \mu_n(F \cap K_\alpha) + \mu_n(K_\alpha^c).$$

Hence

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) &\leq \limsup_{n\to\infty} \frac{1}{n} \log \left(2 \max \{\mu_n(F \cap K_\alpha), \mu_n(K_\alpha^c)\}\right) \\
&= \limsup_{n\to\infty} \max \left\{ \frac{1}{n} \log \mu_n(F \cap K_\alpha), \frac{1}{n} \log \mu_n(K_\alpha^c) \right\} \\
&\leq -\alpha = -\inf_{x\in F} I(x).
\end{aligned}$$

If $\alpha = \infty$, then replace α by M in the above arguments and then let $M \to \infty$.

b) Fix $\alpha \geq 0$. Let K_α be as in the definition of exponential tightness. Then K_α^c is open and

$$-\inf_{x\in K_\alpha^c} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha$$

and therefore $\{I \leq \alpha\} \subseteq K_\alpha$. Since $\{I \leq \alpha\}$ is closed by assumption (I is a rate function) and every closed subset of a compact set is compact (this holds on any topological space), we see that I is a good rate function.

□

The following proposition is a partial converse of the previous lemma.

Proposition 3.13. *Let E be a locally compact Hausdorff space and assume that $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies an LDP with good rate function I. Then $\{\mu_n\}_{n\in\mathbb{N}}$ is exponentially tight.*

Proof. Fix $\alpha \geq 0$. Then the set $\{x \in E : I(x) \leq \alpha\}$ is compact (since I is good). By local compactness every $x \in E$ has a compact neighborhood. We cover $\{x \in E : I(x) \leq \alpha\}$ by the family of interiors of all these compact neighborhoods for all $x \in \{x \in E : I(x) \leq \alpha\}$. By compactness there exists a finite subcover and the union of the corresponding compact neighborhoods is itself compact (finite unions of compact sets are always compact). Let K denote this compact set. Then

$$\{x \in E : I(x) \le \alpha\} \subset \mathrm{int}(K) \subset K$$

and hence

$$K^c \subset \mathrm{cl}(K^c) \subset \{x \in E : I(x) > \alpha\},$$

where int(A) and cl(A) denote the interior and the closure of a set A. We get

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K^c) &\le \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\mathrm{cl}(K^c)) \\ &\le -\inf_{x\in \mathrm{cl}(K^c)} I(x) \le -\alpha.\end{aligned}$$

□

4 Sanov's theorem

Our goal in this section is to prove Sanov's Theorem, which states an LDP for the empirical distribution of a sequence of i. i. d. random variables. The rate function turns out to be the well-known relative entropy (with respect to the joint law of the random variables). In this section we always assume that E is a Polish space with complete metric ρ. We denote the set of all probability measures on $(E, \mathcal{E})$ by $\mathcal{M}_1(E)$. We will need a topology on $\mathcal{M}_1(E)$.

Proposition 4.1. *Define $d : \mathcal{M}_1(E) \times \mathcal{M}_1(E) \to \mathbb{R}$ by*

$$d(\mu, \nu) := \inf\{\delta > 0 : \mu(F) \le \nu(F^\delta) + \delta \textit{ for all closed sets } F \subseteq E\},$$

where $F^\delta := \{x \in E : \rho(x, F) < \delta\}$. Then d is a metric (the Lévy metric*) on $\mathcal{M}_1(E)$. With this metric $\mathcal{M}_1(E)$ is a Polish space. Furthermore convergence with respect to this metric is the same as weak convergence.*

In the following, $X_1, X_2, \ldots$ will denote an E-valued sequence of i. i. d. random variables with law μ.

Proposition 4.2.

$$\mu_n(\omega) := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)} \to \mu \quad \textit{almost surely.}$$

Proof. Let $f : E \to \mathbb{R}$ be a bounded measurable function. By the strong law of large numbers, we have

$$\int f(x)\mu_n(\omega, \mathrm{d}x) = \frac{1}{n}\sum_{i=1}^{n} f(X_i(\omega)) \to \mathbb{E}f(X_1) = \int f(x)\mu(\mathrm{d}x) \text{ a. s.} \tag{5}$$

In particular this holds for any bounded and continuous function. Unfortunately we cannot deduce immediately that $\mu_n(\omega, .)$ converges to μ weakly almost surely because the exceptional sets of measure zero depend on the function f. Therefore we will show the following:

$$\mathbb{P}\left(\liminf_{n\to\infty} \mu_n(\omega, G) \geq \mu(G) \text{ for all open sets } G\right) = 1, \tag{6}$$

from which the assertion follows by the Portmanteau Theorem (WT II). Let $G_1, G_2, \ldots$ be a countable base of the topology of E, i. e. a countable family of open sets such that any open set is the union of a subfamily of the G_i. We can and will assume that the family is closed with respect to taking finite unions. By (5) we have

$$\mathbb{P}\left(\lim_{n\to\infty} \mu_n(\omega, G_i) = \mu(G_i) \text{ for all } i\right) = 1.$$

Let $A := \{\omega : \mu_n(\omega, G_i) \to \mu(G_i) \text{ for all } i \in \mathbb{N}\}$. Then $\mathbb{P}(A) = 1$. Any open set G can be written as a union $G = \cup_{k=1}^{\infty} G_{i_k}$. Then for any $\omega \in A$ we have

$$\mu_n(\omega, G) = \mu_n\left(\omega, \cup_{k=1}^{\infty} G_{i_k}\right) \geq \mu_n\left(\omega, \cup_{k=1}^{N} G_{i_k}\right) \to \mu\left(\cup_{k=1}^{N} G_{i_k}\right)$$

and hence

$$\liminf_{n\to\infty} \mu_n(\omega, G) = \mu(G).$$

This shows (6) and the proposition is proved. □

Let μ_n be defined as in the previous proposition and denote the law of the $\mathcal{M}_1(E)$-valued random variable μ_n by L_n (L_n is a probability measure on $\mathcal{M}_1(E)$).

Lemma 4.3. *$\{L_n\}_{n\in\mathbb{N}}$ is exponentially tight.*

Proof. $\{\mu\}$ is tight, since E is Polish. Hence there exist compact subsets $\Gamma_l \subseteq E$, $l \in \mathbb{N}$ such that $\mu\left(\Gamma_l^c\right) \leq e^{-2l^2}(e^l - 1)$. Now define

$$\begin{aligned} K^l \quad &:= \quad \left\{\nu : \nu\left(\Gamma_l\right) \geq 1 - \frac{1}{l}\right\} \qquad \text{and} \\ K_L \quad &:= \quad \cap_{l=L}^{\infty} K^l, \qquad L = 2, 3, \ldots \end{aligned}$$

Now it is not hard to see, that the set K_L is tight and closed and therefore – by Prochorov's Theorem – compact and that

$$\limsup_{n\to\infty} \frac{1}{n} \log L_n\left(K_L^c\right) \leq -L$$

showing that $\{L_n\}_{n\in\mathbb{N}}$ is exponentially tight. For details, see [1]. $\square$

Definition 4.4. Let $\mu, \nu \in \mathcal{M}_1(E)$. Then

$$H(\nu|\mu) := \begin{cases} \int_E f \log f d\mu & \text{if } \nu \ll \mu \text{ and } f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}, \\ \infty & \text{otherwise.} \end{cases}$$

is called *relative entropy* (or *Kullback-Leibler distance*) of ν relative to μ.

Remark 4.5.

- $H(\mu|\mu) = 0$ for every $\mu \in \mathcal{M}_1(E)$.
- In general $H(\nu|\mu) \neq H(\mu|\nu)$.
- $H(\nu|\mu) \geq 0$, since – in case $\nu \ll \mu$ –

$$\begin{aligned} H(\nu|\mu) &= \int \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu \\ &\geq \int \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \mathrm{d}\mu \, \log \int \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \mathrm{d}\mu = 0, \end{aligned}$$

by Jensen's inequality applied to the convex function $x \mapsto x \log x$ on $[0, \infty)$.

Lemma 4.6.

$$\begin{aligned} H(\nu|\mu) &= \sup \left(\int_E f \mathrm{d}\nu - \log \int_E \mathrm{e}^f \mathrm{d}\mu \right) \\ &= \sup_{f \in C_b(E)} \left(\int_E f \mathrm{d}\nu - \log \int_E \mathrm{e}^f \mathrm{d}\mu \right), \end{aligned}$$

where the first supremum is taken over all measurable functions $f : E \to \mathbb{R}$ such that $\int_E \mathrm{e}^f \mathrm{d}\mu < \infty$ and $\int_E f \mathrm{d}\nu$ is defined and where $C_b(E)$ denotes the set of bounded continuous real-valued functions on E.

Proof. We will see in the proof, that the first equality is even true in an arbitrary measurable space $(E, \mathcal{E})$.

Step 1: We prove "$\geq$" in the first inequality. There is nothing to prove in case ν is not absolutely continuous with respect to μ, so we assume $\nu \ll \mu$. Let $f : E \to \mathbb{R}$ be measurable such that $\int \mathrm{e}^f \mathrm{d}\mu < \infty$ and $\int f \mathrm{d}\nu > -\infty$ (there is nothing to prove if $\int f \mathrm{d}\nu = -\infty$). Since $\int \mathrm{e}^f \mathrm{d}\mu > 0$, the formula

$$\mathrm{d}\mu_f(x) := \frac{\mathrm{e}^{f(x)}}{\int \mathrm{e}^f d\mu} \mathrm{d}\mu(x)$$

defines a probability measure which is equivalent to μ (i. e. $\mu_f \ll \mu$ and $\mu \ll \mu_f$). Since $\nu \ll \mu$ we have

$$\begin{aligned} H(\nu|\mu) &= \int \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right) \mathrm{d}\mu = \int \left(\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right) \mathrm{d}\nu \\ &= \int \left(\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu_f} \frac{\mathrm{d}\mu_f}{\mathrm{d}\mu} \right) \mathrm{d}\nu \\ &= \int \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu_f} \mathrm{d}\nu + \int \log \frac{\mathrm{e}^f}{\int \mathrm{e}^f \mathrm{d}\mu} \mathrm{d}\nu \\ &= H(\nu|\mu_f) + \int_E f \mathrm{d}\nu - \log \int_E \mathrm{e}^f \mathrm{d}\mu \\ &\geq \int_E f \mathrm{d}\nu - \log \int_E \mathrm{e}^f \mathrm{d}\mu, \end{aligned}$$

where we have used that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu_f} \frac{\mathrm{d}\mu_f}{\mathrm{d}\mu}$ μ- and hence ν-almost surely (WT II, 1.46).
Step 2: We prove "$\leq$" in the first inequality. In case $\nu \ll \mu$ we define $f := \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Then

$$H(\nu|\mu) = \int_E f \mathrm{d}\nu - \log \int_E \mathrm{e}^f \mathrm{d}\mu$$

so we have equality in this case. It remains to show "$\leq$" in case ν is *not* absolutely continuous with respect to μ. Take a set $A \in \mathcal{E}$ such that $\nu(A) > 0$, but $\mu(A) = 0$. Define $f_M(x) := M\mathbf{1}_A(x)$, $M \in \mathbb{N}$. Then $H(\nu|\mu) = \int_E f \mathrm{d}\nu - \log \int_E \mathrm{e}^f \mathrm{d}\mu = M\nu(A) \to \infty$ as $M \to \infty$. This completes Step 2.
Step 3: We prove the second inequality. Clearly we have "$\geq$" (we take the supremum over a smaller set of functions on the right hand side of the equality), so it only remains to show "$\leq$" (see [1]). $\square$

Corollary 4.7. *For every $\mu \in \mathcal{M}_1(E)$, $H(.|\mu)$ is a rate function.*

Proof. We already showed that $H(.|\mu)$ is nonnegative. To show that $H(.|\mu)$ is lower semi-continuous, fix $\alpha \geq 0$. Then

$$\{\nu : H(\nu|\mu) \leq \alpha\} = \bigcap_{f \in C_b(E)} \left\{\nu : \int f d\nu \leq \log \int_E \mathrm{e}^f \mathrm{d}\mu + \alpha\right\}$$

is an intersection of closed sets and hence closed. $\square$

Now we formulate Sanov's Theorem.

Theorem 4.8. **(Sanov)** *Let $X_1, X_2, \ldots$ be an E-valued sequence of i. i. d. random variables with law μ. Define $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $n \in \mathbb{N}$ and $L_n := \mathcal{L}(\mu_n)$. Then the sequence $\{L_n\}_{n\in\mathbb{N}}$ satisfies an LDP with good rate function $H(.|\mu)$.*

Proof. We know from the previous corollary that $H(.|\mu)$ is a rate function and from Lemma 4.3 that $\{L_n\}_{n\in\mathbb{N}}$ is exponentially tight. It remains to verify properties a) and b) in Lemma 3.12 for $\{L_n\}_{n\in\mathbb{N}}$ (see [1]). $\square$

References

[1] A. DEMBO, O. ZEITOUNI (1998). *Large Deviations Techniques*, Springer, Berlin/Heidelberg/New York.

[2] F. DEN HOLLANDER (2000). *Large Deviations*, Fields Institute Monographs, AMS, Providence.