

6. LDP ON TOPOLOGICAL VECTOR SPACES

In the previous sections, we have obtained from the Markov's inequality a natural candidate of rate function Λ^* . In fact, this candidate is the rate function for many general processes, and for this reason we will study the function systematically in this section. Throughout the section, \mathcal{X} is a Hausdorff real topological vector space.

Definition 6.1. Let \mathcal{X} be a Hausdorff real topological vector space and $\mu_\varepsilon \in M_1(\mathcal{X})$ be the probability measure associated with Z_ε .

- Define the *logarithmic generating function* as $\Lambda_{\mu_\varepsilon} : \mathcal{X}^* \rightarrow (-\infty, \infty]$

$$\Lambda_{\mu_\varepsilon}(\lambda) = \log \mathbb{E}(e^{\langle \lambda, Z_\varepsilon \rangle})$$

- Let $\bar{\Lambda}(\lambda) = \limsup_{\varepsilon \rightarrow 0} \varepsilon \Lambda_{\mu_\varepsilon}(\frac{\lambda}{\varepsilon})$ and write $\Lambda(\lambda) = \bar{\Lambda}(\lambda)$ if the limit exists.
- Define the *Fenchel-Legendre transform* $\bar{\Lambda}^*$ of $\bar{\Lambda}$ as

$$\bar{\Lambda}^*(x) = \sup_{\lambda \in \mathcal{X}^*} [\langle \lambda, x \rangle - \bar{\Lambda}(\lambda)].$$

6.1. General upper bound.

Theorem 6.2.

- (1) $\bar{\Lambda}$ is a convex on \mathcal{X}^*
- (2) $\bar{\Lambda}^*$ is a convex rate function.
- (3) For any compact set $\Gamma \subset \mathcal{X}$,

$$\limsup_{n \rightarrow \infty} \varepsilon \log \mu_\varepsilon(\Gamma) \leq - \inf_{x \in \Gamma} \bar{\Lambda}^*(x)$$

Proof. (1) By Hölder's inequality, given any $t, t' \in [0, 1]$ satisfying $t + t' = 1$,

$$\log \mathbb{E}(e^{\langle t\lambda + t'\lambda', Z_\varepsilon \rangle}) \leq t \log \mathbb{E}(e^{\langle \lambda, Z_\varepsilon \rangle}) + t' \log \mathbb{E}(e^{\langle \lambda', Z_\varepsilon \rangle}),$$

proving the convexity of $\Lambda_{\mu_\varepsilon}$ and hence $\bar{\Lambda}$.

(2) The convexity of $\bar{\Lambda}^*$ follows immediately from definition. To see $\bar{\Lambda}^*$ is rate function, note that (a) it is non-negative

$$\bar{\Lambda}^*(x) \geq \langle 0, x \rangle - \bar{\Lambda}(0) = 0,$$

and that (b) it is lower semicontinuous, as $x \mapsto \langle x, \lambda \rangle - \Lambda(\lambda)$ is continuous for all $\lambda \in \mathcal{X}^*$.

- (3) Choose for each $x \in \Gamma$ an $\lambda_x \in \mathcal{X}^*$ and a open neighborhood A_x of x such that

$$\begin{aligned} \langle \lambda_x, x \rangle - \bar{\Lambda}(x) &\geq I^\delta(x), \\ \inf_{y \in A_x} \{\langle \lambda_x, y \rangle - \langle \lambda_x, x \rangle\} &\geq -\delta, \end{aligned}$$

where I^δ is the δ -rate function associated with $\bar{\Lambda}^*$ as defined in (3.8). Applying the Markov's inequality yields

$$\mu_\varepsilon(A_x) \leq \mathbb{E}\left(e^{\langle \frac{\lambda_x}{\varepsilon}, Z_\varepsilon \rangle - \langle \frac{\lambda_x}{\varepsilon}, x \rangle}\right) \exp\left(- \inf_{y \in A_x} \left\langle \frac{\lambda_x}{\varepsilon}, y \right\rangle - \left\langle \frac{\lambda_x}{\varepsilon}, x \right\rangle\right),$$

which in turn implies

$$\varepsilon \log \mu_\varepsilon(A_x) \leq \delta - \left[\langle \lambda_x, x \rangle - \varepsilon \Lambda_{\mu_\varepsilon}\left(\frac{\lambda_x}{\varepsilon}\right) \right].$$

Now that Γ is a compact set, one can find a finite cover $\{A_{x_i}\}_{i=1}^N$ with $x_i \in \Gamma$ such that

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mu_\varepsilon(\Gamma) \leq \delta - \min_{1 \leq i \leq N} I^\delta(x_i) \leq \delta - \inf_{x \in \Gamma} I^\delta(x),$$

proving the theorem by letting $\delta \rightarrow 0$. \square

6.2. Rate function and Fenchel-Legendre transform.

A large class of topological vector spaces, for example when \mathcal{X} is the space of finite measures on Polish space, is actually locally convex. For this reason, we expect good properties regarding LDP for such spaces.

Theorem 6.3. *Let \mathcal{X} be a locally convex Hausdorff topological vector space. Assume that μ_ε satisfies the LDP with a good rate function I . Suppose in addition that*

$$\bar{\Lambda}(\lambda) = \limsup_{\varepsilon \rightarrow 0} \varepsilon \Lambda_{\mu_\varepsilon} \left(\frac{\lambda}{\varepsilon} \right), \quad \forall \lambda \in \mathcal{X}^*. \quad (6.1)$$

(1) *For each $\lambda \in \mathcal{X}^*$, $\lim_{\varepsilon \rightarrow 0} \varepsilon \Lambda_{\mu_\varepsilon} \left(\frac{\lambda}{\varepsilon} \right)$ exists, is finite, and*

$$\Lambda(\lambda) = \sup_{x \in \mathcal{X}} [\langle \lambda, x \rangle - I(x)].$$

(2) *If I is convex, then it is the Fenchel-Legendre transform of Λ , namely,*

$$I(x) = \Lambda^*(x) := \sup_{\lambda \in \mathcal{X}^*} [\langle \lambda, x \rangle - \Lambda(\lambda)].$$

(3) *If I is not convex, then Λ^* is the affine regularization of I , i.e.,*

$$\Lambda^* = \sup \{ f : f \leq I, f \text{ is convex} \}.$$

Proof. (1) Fix $\lambda \in \mathcal{X}^*$ and $\gamma > 1$. By assumption, Varadhan's lemma (Theorem 5.4) asserts

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \Lambda_{\mu_\varepsilon} \left(\frac{\lambda}{\varepsilon} \right) = \sup_{\lambda \in \mathcal{X}^*} [\langle \lambda, x \rangle - I(x)] < \infty.$$

Since Λ is convex and $\Lambda(0) = 0$, it follows that $\Lambda > -\infty$.

(2) It follows from the duality lemma (Lemma D.1).

(3) By (1),

$$\Lambda^*(x) = I^{**}(x) = \sup_{\lambda \in \mathcal{X}^*} \inf_{y \in \mathcal{X}} [\langle \lambda, x - y \rangle + I(y)] \leq I(x).$$

By Hahn-Banach separation theorem (Theorem C.2),

$$\sup \{ f : f \leq I, f \text{ is convex} \} = \sup \{ \lambda + c : \lambda + c \leq I, \lambda \in \mathcal{X}^*, c \in \mathbb{R} \}.$$

If $I \geq \lambda + c$, then it is necessary $\inf_{y \in \mathcal{X}} I(y) - \langle \lambda, y \rangle \geq c$, or equivalently,

$$\inf_{y \in \mathcal{X}} \langle \lambda, x - y \rangle + I(y) \geq \langle \lambda, x \rangle + c,$$

from which the claim follows. \square

Theorem 6.4. *Suppose that $\{\mu_\varepsilon\}$ satisfies a weak LDP with a convex rate function I , and that \mathcal{X} is a locally convex, Hausdorff topological vector space. Assume that for each $\lambda \in \mathcal{X}$, the limits $\Lambda_\lambda(t) = \lim_{\varepsilon \rightarrow 0} \varepsilon \Lambda_{\mu_\varepsilon}(t\lambda/\varepsilon)$ exist as extended real numbers, and that $\Lambda_\lambda(t)$ is a lower semicontinuous function of $t \in \mathbb{R}$. If for every $\lambda \in \mathcal{X}^*$ and every $a \in \mathbb{R}$,*

$$\inf_{x:\langle\lambda,x\rangle>a} I(x) \leq \inf_{z>a} \Lambda_\lambda^*(z),$$

then $I = \Lambda^*$, and consequently, Λ^* controls a weak LDP associated with $\{\mu_\varepsilon\}$.

Proof. By the assumption,

$$\begin{aligned} \sup_{x \in \mathcal{X}} \{\langle\lambda, x\rangle - I(x)\} &\geq \sup_{a \in \mathbb{R}} \sup_{x:\langle\lambda,x\rangle>a} \{\langle\lambda, x\rangle - I(x)\} \\ &\geq \sup_{a \in \mathbb{R}} \left[a - \inf_{x:\langle\lambda,x\rangle>a} I(x) \right] \\ &\geq \sup_{a \in \mathbb{R}} \left[a - \inf_{z>a} \Lambda_\lambda^*(z) \right] = \sup_{z \in \mathbb{R}} [z - \Lambda_\lambda^*(z)]. \end{aligned}$$

Note that Λ_λ is convex with $\Lambda_\lambda(0) = 0$ and is lower semicontinuous. Therefore, it can not attain the value $-\infty$. By the duality lemma (Lemma D.1), $\sup_{z \in \mathbb{R}} [z - \Lambda_\lambda^*(z)] = \Lambda_\lambda(1) = \Lambda(\lambda)$, proving

$$\sup_{x \in \mathcal{X}} \{\langle\lambda, x\rangle - I(x)\} \geq \Lambda(\lambda).$$

By Lemma 5.5, we have

$$\Lambda(\lambda) \geq \sup_{x \in \mathcal{X}} [\langle\lambda, x\rangle - I(x)].$$

Combining the above gives $I^* = \Lambda$. Applying the duality lemma (Lemma D.1) again gives $\Lambda^* = I^{**} = I$. \square

6.3. Convexity consideration.

Lemma 6.5. *Let \mathcal{A} be a base for a Hausdorff topological vector space \mathcal{X} , such that (1) $\bar{I} = \underline{I}$ and that (2) for every $A_1, A_2 \in \mathcal{A}$, $\bar{\mathcal{L}}_{A_1+A_2} \geq \frac{1}{2}(\underline{\mathcal{L}}_{A_1} + \underline{\mathcal{L}}_{A_2})$. Then the rate function I of the weak LDP for $\{\mu_\varepsilon\}$ is convex.*

Proof. It suffices to show that $I\left(\frac{x+y}{2}\right) \leq \frac{I(x)+I(y)}{2}$ for all $x, y \in \mathcal{X}$, as the assertion amounts to that $I(tx + (1-t)y) \leq tI(x) + (1-t)I(y)$ for all $x, y \in \mathcal{X}$ and all $t = k \cdot 2^{-n}$ with $k = 0, 1, \dots, 2^n$ and $n \in \mathbb{N}$. By lower semicontinuity, the inequality further extends to hold for all $t \in [0, 1]$.

To prove the claim, our starting point is the observation that the Hausdorff TVS is regular, which follows from the fact that $+: \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X}$ is continuous. By virtue of the regularity and lower semicontinuity, we can find for every $x, y \in \mathcal{X}$ a neighborhoods $A \in \mathcal{A}$ of $\frac{x+y}{2}$, $A_1 \in \mathcal{A}$ of x and $A_2 \in \mathcal{A}$ of y such that $A \supset \bar{A}_1 + \bar{A}_2$ satisfying

$$-I^\delta\left(\frac{x+y}{2}\right) \geq \bar{\mathcal{L}}_A \geq \frac{1}{2}(\bar{\mathcal{L}}_{A_1+A_2}) \geq \underline{\mathcal{L}}_{A_1} + \underline{\mathcal{L}}_{A_2} \geq \frac{1}{2}I(x) + \frac{1}{2}I(y).$$

This concludes the proof. \square

Corollary 6.6. *Suppose that $\{\mu_\varepsilon\}$ is exponentially tight, satisfies (6.1) and the assumptions of Lemma 6.5. Then $\{\mu_\varepsilon\}$ satisfies in \mathcal{X} the LDP with the good, convex rate function Λ^* .*

Proof. By Lemma 6.5, $\{\mu_\varepsilon\}$ satisfies a weak LDP with a convex rate function. As $\{\mu_\varepsilon\}$ is exponentially tight, it is deduced that it satisfies the full LDP with a convex, good rate function. The proof is concluded by Theorem 6.3 (1) and (2). \square

7. APPLICATION - CRAMÉR'S THEOREM IN POLISH SPACE

For convenience, we write

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^n X_i \text{ and } \hat{S}_m^n = \frac{1}{n-m} \sum_{i=m+1}^n X_i.$$

Theorem 7.1. *Let \mathcal{X} is a locally convex, Hausdorff, topological real vector space, and X_i , μ_n , Λ , and Λ^* be as defined. Assume the following hold.*

- (1) *There is a closed, convex subset \mathcal{E} of \mathcal{X} such that $\mu(\mathcal{E}) = 1$ and \mathcal{E} can be made into a Polish space with respect to the topology induced by \mathcal{X} .*
- (2) *The closed convex hull of each compact $K \subset \mathcal{E}$ is compact.*

Then $\{\mu_n\}$ satisfies in \mathcal{X} (and \mathcal{E}) a weak LDP with rate function Λ^ . Moreover, for every open, convex subset $A \subset \mathcal{X}$,*

$$\lim_{n \rightarrow \infty} \log \mu_n(A) = -\inf \Lambda^*(x).$$

Remark 7.2. *The theorem, in particular, applies to separable Banach spaces, or $M_1(\Sigma)$ with Σ a Polish space.*

As a corollary of the theorem, we generalize the Cramer's theorem to the following.

Corollary 7.3 (Cramér). *The sequence $\{\mu_n\}$ of the laws of empirical means of \mathbb{R}^d -valued i.i.d. random variables satisfies a weak LDP with the convex rate function Λ^* . Moreover, if $0 \in \overset{\circ}{\mathcal{D}}_\Lambda$, then $\{\mu_n\}$ satisfies the full LDP with the good, convex rate function Λ^* .*

Proof. The weak LDP is a consequence of Theorem 7.1, from which the full LDP follows due to the exponential tightness. \square

7.1. Proofs.

Lemma 7.4. *Let Theorem 7.1(1) hold true. Then, the sequence $\{\mu_n\}$ satisfies the weak LDP in \mathcal{X} with a convex rate function I .*

Proof. The lemma follows from Lemma 6.5 with a base \mathcal{A} of topology consisting of convex (open) sets, as argued in the following.

Note that for all convex set A , $\mu_n(A) = \mu_n(A \cap \mathcal{E})$ and thus $\overline{\mathcal{L}}_A = \underline{\mathcal{L}}_A$ exists by subadditivity. In addition, we know that for any convex open sets A_1 and A_2 ,

$$\begin{aligned} \mu_n(A_1)\mu_n(A_2) &= P_\mu(\hat{S}_n \in A_1, \hat{S}_n^{2n} \in A_2) \\ &\leq P_\mu\left(\hat{S}_{2n} \in \frac{1}{2}(A_1 + A_2)\right) = \mu_{2n}\left(\frac{1}{2}(A_1 + A_2)\right), \end{aligned}$$

from which the weak LDP holds for $\{\mu_n\}$ with rate function

$$I(x) = -\inf_{A \in \mathcal{A}: x \in A} \underline{\mathcal{L}}_A.$$

\square

Lemma 7.5. *Let Theorem 7.1(1)(2) hold true. Then, for every open, convex subset $A \subset \mathcal{X}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x \in A} I(x),$$

where I is the convex rate function in Lemma 7.4.

Proof. By definition,

$$-I(x) \leq \underline{\mathcal{L}}_A = \overline{\mathcal{L}}_A \Rightarrow -\inf_{x \in A} I(x) \leq \underline{\mathcal{L}}_A = \overline{\mathcal{L}}_A.$$

Hence, it suffices to prove the other inequality. To this end, we first make use of the subadditivity to approximate the limit and tightness to approximate the set A . Specifically, for all $\delta > 0$, there exist $N \in \mathbb{N}$ and compact set $C \subset A$ such that

$$\underline{\mathcal{L}}_A \leq \frac{1}{N} \log \mu_N(A) + \delta \leq \frac{1}{N} \log \mu_N(\overline{\text{co}}(C)) + 2\delta \leq \overline{\mathcal{L}}_{\overline{\text{co}}(C)} \leq -\inf_{x \in A} I(x),$$

where the third inequality follows from subadditivity. This completes the proof. \square

Proof of Theorem 7.1. In light of Theorem 6.4, it suffices to show that

$$\frac{1}{n} \log \mathbb{E} \left[e^{t \langle \lambda, n \hat{S}_n \rangle} \right] = \log \mathbb{E} \left[e^{t \langle \lambda, X_1 \rangle} \right] =: \Lambda_\lambda(t)$$

is lower semicontinuous and

$$\inf_{x: \langle \lambda, x \rangle > a} I(x) \leq \inf_{z > a} \Lambda_\lambda^*(z) \text{ for all } a \in \mathbb{R} \text{ and } \lambda \in \mathcal{X}^*.$$

The former follows obviously from Fatou's lemma. As for the latter, we consider the following two cases: (1) $\lambda = 0$ and (2) $\lambda \neq 0$.

(1) We assume $\lambda = 0$, so that

$$\Lambda_\lambda^*(x) = \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{if } x \neq 0, \end{cases}$$

Given that $\mu_n(\mathcal{X}) = 1$, it follows from Lemma 7.5 that $\inf_{x \in \mathcal{X}} I(x) = 0$. Hence, the proposed inequality holds.

(2) Note that the half-plane $H_a := \{x : \langle \lambda, x \rangle > a\}$ is convex and open. By Lemma 7.5,

$$-\inf_{x \in H_a} I(x) = \overline{\mathcal{L}}_{H_a} \geq \sup_{\delta > 0} \overline{\mathcal{L}}_{H_{a+\delta}}.$$

To conclude the proof, define $Y_i = \langle \lambda, X_i \rangle$ and $\hat{Z}_n = \sum_{i=1}^n Y_i$ so that the logarithmic moment generating function of Y_1 is exactly $\Lambda_\lambda(t)$. By one-dimensional Cramér's theorem (Theorem 3.3), we have

$$\inf_{\delta > 0} \overline{\mathcal{L}}_{H_{a+\delta}} = \sup_{\delta > 0} \left[-\inf_{z > a+\delta} I(x) \right] = -\inf_{z > a} I(x),$$

This proves the theorem. \square