8. Gärtner-Ellis theorem.

Throughout this section, the space $\mathcal{X}$ is a Hausdorff topological vector space.

8.1. **Baldi's theorem.**

In this subsection, we consider a generalization of Sanov's theorem on topological vector space.

Definition 8.1. A point $x \in \mathcal{X}$ is called a an *exposed point* of $\overline{\Lambda}^*$ if there exists an *exposing hyperplane* $\lambda \in \mathcal{X}^*$ such that for every $z \neq x$,

$$\langle \lambda, x\rangle - \overline{\Lambda}^*(x) > \langle \lambda, z\rangle - \overline{\Lambda}^*(z).$$

Theorem 8.2 (Baldi)**.** *Suppose $\{\mu_\varepsilon\}$ are exponentially tight family of probability measures on $\mathcal{X}$.*

1. *For every closed set $F \subset \mathcal{X}$, $\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(F) \leq -\inf_{x\in F} \overline{\Lambda}^*(x)$.*
2. *Let $\mathcal{F}$ be the set of exposed points of $\overline{\Lambda}^*$ with an exposing hyperplane λ for which*

$$\Lambda(\lambda) = \lim_{\varepsilon\to 0} \varepsilon \Lambda_{\mu_\varepsilon}\left(\frac{\lambda}{\varepsilon}\right) \text{ exists and } \overline{\Lambda}(\gamma\lambda) < \infty \text{ for some } \gamma > 1. \tag{8.1}$$

Then, for every open set $G \subset \mathcal{X}$,

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(G) \geq -\inf_{x\in G\cap\mathcal{F}} \overline{\Lambda}^*(x).$$

3. *If for every open set $G \subset \mathcal{X}$, $\inf_{x\in G\cap\mathcal{F}} \overline{\Lambda}^*(x) = \inf_{x\in G} \overline{\Lambda}^*(x)$, then $\{\mu_\varepsilon\}$ satisfies the LDP with the good rate function $\overline{\Lambda}^*$.*

Proof. (1) It follows from the exponential tightness together with Theorem 6.2.

(2) The case $\overline{\Lambda}(\lambda) = -\infty$ for some $\lambda \in \mathcal{X}^*$ is trivial, for $\overline{\Lambda}^*(\cdot) = \infty$ unanimously. Assuming $\overline{\Lambda}(\lambda) < \infty$ for all $\lambda \in \mathcal{X}^*$, one need to translate $\overline{\Lambda}$ by λ in the following manner. Let G be any fixed open set, $y \in G \cap \mathcal{F}$, $\delta > 0$, and $\eta \in \mathcal{X}^*$ be an exposing hyperplane for $\overline{\Lambda}^*$ such that (8.1) holds. By continuity of η, we choose an open neighborhood $B_\delta \subset G$ of y such that

$$\sup_{z\in B_\delta} \langle \eta, z - y\rangle < \delta.$$

According to (8.1), for every sufficiently small $\varepsilon > 0$, $\Lambda_{\mu_\varepsilon}\left(\frac{\eta}{\varepsilon}\right)$ is well-defined, for each of which we may defined a probability measure $\tilde{\mu}_\varepsilon$ equivalent to μ_ε (i.e., $\tilde{\mu}_\varepsilon$ and μ_ε are absolutely continuous to each other):

$$\frac{d\tilde{\mu}_\varepsilon}{d\mu_\varepsilon}(z) = \exp\left[\left\langle \frac{\eta}{\varepsilon}, z\right\rangle - \Lambda_{\mu_\varepsilon}\left(\frac{\eta}{\varepsilon}\right)\right]$$

so that

$$\begin{aligned}\varepsilon \log \mu_\varepsilon(B_\delta) &= \varepsilon \log \int_{B_\delta} \exp\left[-\left\langle \frac{\eta}{\varepsilon}, z\right\rangle + \Lambda_{\mu_\varepsilon}\left(\frac{\eta}{\varepsilon}\right)\right] d\tilde{\mu}_\varepsilon(z)\\ &= \varepsilon \Lambda_{\mu_\varepsilon}\left(\frac{\eta}{\varepsilon}\right) - \langle \eta, y\rangle + \varepsilon \log \int_{B_\delta} \exp\left[\left\langle \frac{\eta}{\varepsilon}, y - z\right\rangle\right] d\tilde{\mu}_\varepsilon(z),\end{aligned}$$

yielding

$$\liminf_{\varepsilon\to 0}\varepsilon\log\mu_\varepsilon(B_\delta)\geq\overline{\Lambda}(\eta)-\langle\eta,y\rangle+\lim_{\delta\to 0}\liminf_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon(B_\delta),$$
$$\geq-\overline{\Lambda}^*(y)+\lim_{\delta\to 0}\liminf_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon(B_\delta).$$

It remains to show that $\lim_{\delta\to 0}\liminf_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon(B_\delta)=0$. To this end, we may choose, due to exponential tightness, for each $\alpha<\infty$ a compact sets K_α such that

$$\limsup_{\varepsilon\to 0}\varepsilon\log\mu_\varepsilon(K_\alpha^c)<-\alpha$$

and claim that

$$\limsup_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon(B_\delta^c\cap K_\alpha)<0\text{ for all }\delta>0,\alpha>0,\tag{8.2}$$

$$\limsup_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon(K_\alpha^c)<0\text{ for all large }\alpha>0,\tag{8.3}$$

which together implies the desired estimate.

To prove (8.2), define a function for every $\theta\in\mathcal{X}^*$

$$\tilde{\Lambda}(\theta):=\limsup_{\varepsilon\to 0}\varepsilon\Lambda_{\tilde{\mu}_\varepsilon}\left(\frac{\theta}{\varepsilon}\right)$$

and observe that

$$\tilde{\Lambda}(\theta)=\overline{\Lambda}(\theta+\eta)-\Lambda(\eta),$$
$$\tilde{\Lambda}^*(z)=\overline{\Lambda}^*(z)+\Lambda(\eta)-\langle\eta,z\rangle.$$

Combining these with Theorem 6.2(3), one deduces

$$\limsup_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon(B_\delta^c\cap K_\alpha)\leq-\inf_{z\in B_\delta^c\cap K_\alpha}\tilde{\Lambda}^*(z)<0$$

since $\tilde{\Lambda}^*$ is lower semicontinuous and η is an exposing hyperplane.

For (8.3), define open half-spaces

$$H_\rho=\{z\in\mathcal{X}:\langle\eta,z\rangle<\rho\}$$

For every $\beta>0$, applying Markov's inequality to $\tilde{\Lambda}(\beta\eta)$ yields

$$\varepsilon\log\tilde{\mu}_\varepsilon\big(H_\rho^c\big)\leq\varepsilon\log\left[\int_{H_\rho^c}e^{\left\langle\frac{\beta\eta}{\varepsilon},x\right\rangle-\beta\rho}d\tilde{\mu}_\varepsilon(x)\right]\leq\varepsilon\Lambda_{\tilde{\mu}_\varepsilon}\left(\frac{\beta\eta}{\varepsilon}\right)-\beta\rho,\tag{8.4}$$

and thus, according to (8.1), for all large $\rho>0$,

$$\limsup_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon\big(H_\rho^c\big)\leq\inf_{\beta>0}\big[\tilde{\Lambda}(\beta\eta)-\beta\rho\big]=\inf_{\beta>0}[\Lambda((1+\beta)\eta)-\Lambda(\eta)-\beta\rho]<0.$$

On the other hand, applying Markov's inequality to $\Lambda(\eta)$ gives that for any ρ and all sufficiently large $\alpha=\alpha(\rho)$,

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\varepsilon\log\tilde{\mu}_\varepsilon\big(K_\alpha^c\cap H_\rho\big)\\&\leq\limsup_{\varepsilon\to 0}\varepsilon\log\int_{K_\alpha^c\cap H_\rho}e^{\left\langle\frac{\eta}{\varepsilon},z\right\rangle-\Lambda_{\mu_\varepsilon}\left(\frac{\eta}{\varepsilon}\right)}d\mu_\varepsilon(z)\\&<\rho-\Lambda(\eta)-\alpha<0.\end{aligned}\tag{8.5}$$

Combining (8.4) and (8.5) proves (8.3).

(3) It follows immediately from (1) and (2). □

In the study of large deviations, a general criterion for the lower bound to hold is the differentiability of Λ in the following sense.

Definition 8.3 (Gateaux differentiable)**.** Let $\mathcal{X}$ be a topological vector space. A function $f : \mathcal{X}^* \to \mathbb{R}$ is said to be *Gateaux differentiable* if for all $\lambda, \theta \in \mathcal{X}^*$, the function $t \in \mathbb{R} \mapsto f(\lambda + t\theta)$ is differentiable at $t = 0$.

Corollary 8.4. *Let μ_ε be exponentially tight probability measures on the Banach space $\mathcal{X}$. Suppose that $\Lambda(\cdot) = \lim_{\varepsilon\to 0} \Lambda_{\mu_\varepsilon}\left(\frac{\cdot}{\varepsilon}\right)$ is finite valued, Gateaux differentiable, and lower semicontinuous in $\mathcal{X}^*$ with respect to the weak* topology. Then $\{\mu_\varepsilon\}$ satisfies the LDP with the good rate function Λ^*.*

Proof. It suffices to show that for any $x \in \mathcal{D}_{\Lambda^*}$, there exists a sequence of exposed points x_k such that $x_k \to x$ and $\Lambda^*(x_k) \to \Lambda^*(x)$. Define

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

and

$$\partial\Lambda^* = \{x \in \mathcal{X} : \partial\Lambda^*(x) \neq \emptyset\}.$$

Assume that $\mathcal{D}_{\Lambda^*} \neq \emptyset$. Since Λ^* is convex and lower semicontinuous, by the Brønsted–Rockafellar theorem (Theorem D.4), for every $x \in \mathcal{D}_{\Lambda^*}$, there exists a sequence $x_k \in \partial\Lambda^*$ such that $x_k \to x$ and $\Lambda^*(x_k) \to \Lambda^*(x)$.

It is enough to prove that when Λ is Gateaux differentiable and weak* lower semicontinuous, any point in dom $\partial\Lambda^*$ is also an exposed point. To this end, observe that

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

Therefore, by Gateaux differentiability, for every $\theta \in \mathcal{X}^*$,

$$\langle \theta, x\rangle \leq \lim_{t\to 0^+} \frac{1}{t}[\Lambda(\lambda + t\theta) - \Lambda(\lambda)] =: D\Lambda(\theta) \quad \text{and} \quad D\Lambda(\theta) = -D\Lambda(-\theta).$$

Consequently, $\langle \theta, x\rangle = D\Lambda(\theta)$ for all $\theta \in \mathcal{X}^*$. Due to this property, if there exists $y \in \mathcal{X}$ satisfying

$$\Lambda(\lambda) = \langle \lambda, y\rangle - \Lambda^*(y) \Rightarrow \langle \lambda, y\rangle = D\Lambda(\lambda),$$

then $\langle \theta, y\rangle = \langle \theta, x\rangle$ for all $\theta \in \mathcal{X}^*$, implying $x = y$. This completes the proof. □

8.2. **Gärtner-Ellis theorem.**

For the finite dimensional vector space, the Baldi's theorem reduces to the Gärtner-Ellis theorem. Let Λ_n be the logarithmic moment generating function associated with random variables Z_n, which can be defined as

$$\Lambda_n(\lambda) := \log \mathbb{E}\left[e^{\langle \lambda, Z_n\rangle}\right].$$

The Gärtner-Ellis theorem states the following.

Definition 8.5. A convex function $\Lambda : \mathbb{R}^d \to (-\infty, \infty]$ is *essentially smooth* if:

(1) $\mathring{\mathcal{D}}_\Lambda$ is nonempty.
(2) Λ is differentiable throughout $\mathring{\mathcal{D}}_\Lambda$.
(3) Λ is steep, namely, $\lim_{n\to\infty}|\nabla\Lambda(\lambda_n)| = \infty$ whenever $\{\lambda_n\}$ is a sequence in $\mathring{\mathcal{D}}_\Lambda$ converging to a boundary point of $\mathring{\mathcal{D}}_\Lambda$.

Definition 8.6. The *relative interior* of a nonempty convex set C is defined as

$$\mathrm{ri}\; C = \{y \in C : \text{for all } x \in C, y - \varepsilon(x-y) \in C \text{ for some } \varepsilon > 0\}.$$

Theorem 8.7 (Gärtner-Ellis). *Suppose that $\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$ exists for all $\lambda \in \mathbb{R}^d$ as extended real numbers and $0 \in \mathring{\mathcal{D}}_\Lambda$.*

(1) For any closed set F,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \le -\inf_{x\in F} \Lambda^*(x).$$

(2) For any open set G,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \ge -\inf_{x\in G\cap\mathcal{F}} \Lambda^*(x),$$

where $\mathcal{F}$ is the set of exposed points of Λ^* admitting an exposing hyperplane belongs to $\mathring{\mathcal{D}}_\Lambda$.

(3) If Λ is an essentially smooth, lower semicontinuous function, then the LDP holds with the good rate function Λ^*.

Lemma 8.8. *Under the same assumption as in Theorem 8.7, the following hold.*

(1) $\Lambda(\lambda)$ is a convex function, $\Lambda(\lambda) > -\infty$ everywhere, and $\Lambda^*(x)$ is a good convex rate function.

(2) Suppose that $y = \nabla\Lambda(\eta)$ for some $\eta \in \mathring{\mathcal{D}}_\Lambda$. Then $\Lambda^*(y) = \langle \eta, y\rangle - \Lambda(\eta)$. Moreover $y \in \mathcal{F}$, with η being the exposing hyperplane for y.

Proof. (1) The convexity follows from Hölder's inequality as in Lemma 3.4.

If $\Lambda(\lambda) = -\infty$ for some $\lambda \in \mathbb{R}^d$, then by convexity $\Lambda(\alpha\lambda) = -\infty$ for all $\alpha \in (0,1]$. Since $\Lambda(0) = 0$, it follows by convexity that $\Lambda(-\alpha\lambda) = \infty$ for all $\alpha \in (0,1]$, contradicting the assumption that $0 \in \mathring{\mathcal{D}}$. Thus, $\Lambda > -\infty$ everywhere.

Since $0 \in \mathring{\mathcal{D}}_\Lambda$, it follows that $B_\delta(0) \subset \mathring{\mathcal{D}}_\Lambda$ for some $\delta > 0$, and $c = \sup_{\lambda\in B_\delta(0)} \Lambda(\lambda) < \infty$ since the convex function Λ is continuous in $\mathring{\mathcal{D}}_\Lambda$. Therefore,

$$\Lambda^*(x) \ge \sup_{\lambda\in\overline{B}_\delta(0)} [\langle\lambda, x\rangle - \Lambda(\lambda)] \ge |\delta||x| - \sup_{\lambda\in\overline{B}_\delta(0)} \Lambda(x),$$

implying $\Psi_{\Lambda^*}(\alpha)$ is bounded for every $\alpha < \infty$. The function Λ^* is convex and lower semicontinuous by a similar argument as that in Lemma 3.4. Combining these implies that Λ^* is a good convex rate function.

(2) Suppose now that for some $x \in \mathbb{R}^d$,

$$\Lambda(\eta) = \langle\eta, y\rangle - \Lambda^*(y) \le \langle\eta, x\rangle - \Lambda^*(x).$$

Then, for every $\theta \in \mathbb{R}^d$,

$$\langle\theta, x\rangle \le \Lambda(\eta+\theta) - \Lambda(\eta).$$

In particular,

$$\langle\theta, x\rangle \le \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[\Lambda(\eta+\varepsilon\theta) - \Lambda(\eta)] = \langle\theta, \nabla\Lambda(\eta)\rangle.$$

Since the inequality holds for all $\theta \in \mathbb{R}^d$, $x = \nabla\Lambda(\eta) = y$. Hence, y is an exposed point of Λ^* with exposing hyperplane $\eta \in \mathring{\mathcal{D}}_\Lambda$. □

Lemma 8.9 (Rockafellar). *If $\Lambda : \mathbb{R}^d \to (-\infty, \infty]$ is an essentially smooth, lower semicontinuous, convex function, then* ri $\mathcal{D}_{\Lambda^*} \subseteq \mathcal{F}$.

Proof. Assume without loss of generality that $\mathcal{D}_{\Lambda^*} \neq \emptyset$ for the lemma is automatic otherwise. Under the circumstances, fix henceforth $x \in \text{ri } \mathcal{D}_{\Lambda^*}$ and define a function

$$f(\lambda) := \Lambda(\lambda) - \langle \lambda, x \rangle + \Lambda^*(x).$$

Observe that $f : \mathbb{R}^d \to [0, \infty]$ is convex, lower semincontinuous, and $\inf_{\lambda \in \mathbb{R}^d} f(\lambda) = 0$. Moreover, $f^*(\cdot) = \Lambda^*(\cdot + x) - \Lambda^*(x)$. Therefore, from $x \in \text{ri } \mathcal{D}_{\Lambda^*}$ it follows that $0 \in \text{ri } \mathcal{D}_{f^*}$. By Lemma D.2, there exists $\eta \in \mathcal{D}_\Lambda$ such that $f(\eta) = 0$. Let $\tilde{\Lambda}(\cdot) = \Lambda(\cdot + \eta) - \Lambda(\eta)$, so that $\tilde{\Lambda}$ is an essentially smooth, convex function and $\tilde{\Lambda}(0) = 0$. Consequently, by Lemma D.3, $\tilde{\Lambda}$ is finite in a neighborhood of the origin and thus $\eta \in \mathring{\mathcal{D}}_\Lambda$, at which f is differentiable by the assumption. Hence, $f(\eta) = \inf_{\lambda \in \mathbb{R}^d} f(\lambda)$, implying that $\nabla f(\eta) = 0$, i.e., $x = \nabla \Lambda(\eta)$. It now follows from Lemma 8.8(2) that $x \in \mathcal{F}$. Since $x \in \text{ri } \mathcal{D}_{\Lambda^*}$ is arbitrary, the proof is complete. □

Proof of Theorem 8.7. (1) and (2) follow from Theorem 8.2 immediately once exponential tightness is proved. To that end, we apply the Markov's inequality to the marginal μ_ε^j on the j-th coordinate of μ_ε. Denote by $\boldsymbol{u}_j \in \mathbb{R}^d$ the j-th unit vector. Since $0 \in \mathring{\mathcal{D}}_\Lambda$ we can find $\eta_j, \theta_j > 0$ such that

$$\mu_\varepsilon^j[\rho, \infty) \leq \exp\bigl(-n\eta_j \rho + \Lambda_n\bigl(-n\eta_j \boldsymbol{u}_j\bigr)\bigr),$$
$$\mu_\varepsilon^j(-\infty, -\rho] \leq \exp\bigl(-n\theta_j \rho + \Lambda_n\bigl(-n\theta_j \boldsymbol{u}_j\bigr)\bigr).$$

Therefore,

$$\lim_{\rho \to \infty} \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon[\rho, \infty) = -\infty,$$
$$\lim_{\rho \to \infty} \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(-\infty, -\rho] = -\infty.$$

The claim is thus proved.

We now prove (3). To this end, it suffices to show that

$$\inf_{x \in G \cap \text{ ri } \mathcal{D}_{\Lambda^*}} \Lambda^*(x) \leq \inf_{x \in G} \Lambda^*(x).$$

We shall further assume, without loss of generality, that $G \cap \mathcal{D}_{\Lambda^*}$ is nonempty, for the inequality is obvious otherwise. Under the assumption, ri $\mathcal{D}_{\Lambda^*}$ is also nonempty, and given any fixed $y \in G \cap \mathcal{D}_{\Lambda^*}$ and $z \in \text{ri } \mathcal{D}_{\Lambda^*}$, we have that for all sufficiently small $\varepsilon > 0$,

$$(1 - \varepsilon)y + \varepsilon z \in G \cap \text{ri } \mathcal{D}_{\Lambda^*},$$

which in turn implies

$$\inf_{x \in G \cap \mathcal{D}_{\Lambda^*}} \Lambda^*(x) \leq \limsup_{\varepsilon \to 0} \Lambda^*((1 - \varepsilon)y + \varepsilon z) \leq \Lambda^*(y).$$

The theorem follows immediately as $y \in G \cap \mathcal{D}_{\Lambda^*}$ is arbitrary. □