3. Cramér's theorem in $\mathbb{R}^d$

The aim of this section is to establish the large deviation principle for i.i.d. random vectors in $\mathbb{R}^d$. Let $(X_i)_{i=1}^{\infty}$ be i.i.d. d-dimensional random vectors each of which is distributed with respect to the law $\mu \in M_1(\mathbb{R}^d)$, and let μ_n be the law of $\hat{S}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Write $\overline{x} := \mathbb{E}X_1$.

Definition 3.1. The *logarithmic moment generating function* associated with the law μ is defined as

$$\Lambda(\lambda) := \log \mathbb{E}\big(e^{\langle \lambda, X_1\rangle}\big). \tag{3.1}$$

Definition 3.2. The *Fenchel-Legendre transform* of $\Lambda(\lambda)$ is

$$\Lambda^*(x) := \sup_{\lambda\in\mathbb{R}^d} [\langle \lambda, x\rangle - \Lambda(\lambda)]. \tag{3.2}$$

3.1. Cramér's theorem in $\mathbb{R}$.

The subsection is devoted to the proof of the following theorem.

Theorem 3.3 (Cramér)**.** *Let X_i, μ_n, Λ, and Λ^* be as defined on $\mathbb{R}$. Then, $\{\mu_n\}$ satisfies the LDP with the convex rate function Λ^*, namely,*

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

Furthermore,

(3) If $0 \in \mathring{\mathcal{D}}_\Lambda$, then Λ^* is good.

Before delving into the exposition, we should first demonstrate the role played by the logarithmic moment generating function. For real-valued random variables, the key inequality in need to establish the upper bound of the LDP (i.e., (1.2)) is Markov's inequality: For all $\lambda \ge 0$,

$$\mu_n[x,\infty) \le e^{-\lambda x}\mathbb{E}\left(e^{\lambda \hat{S}_n}\right) = e^{-\lambda x}\big(\mathbb{E}\big(e^{\lambda X_1}\big)\big)^n$$

$$\Rightarrow \limsup_{n\to\infty} \frac{1}{n}\log\mu_n[x,\infty) \le \inf_{\lambda\ge 0} -\lambda x + \Lambda(\lambda).$$

Similarly,

$$\limsup_{n\to\infty} \frac{1}{n}\log\mu_n(-\infty,x] \le \inf_{\lambda\le 0} -\lambda x + \Lambda(\lambda).$$

These observations hinted at a candidate of rate function and urges a closer inspection on them.

The following lemma provides clear a picture of Λ and Λ^*.

Lemma 3.4. *Let Λ and Λ^* be as defined. Then,*

(1) Λ is a convex function and Λ^* is a convex rate function.
(2) Either of following holds.
  - *If $\Lambda(\lambda) < \infty$ only when $\lambda = 0$, then Λ^* is identically zero.*
  - *If $\Lambda(\lambda) < \infty$ for some $\lambda > 0$ (respectively, $\lambda < 0$), then $\overline{x} < \infty$ (respectively, $\overline{x} > -\infty$) is well-defined. Under the circumstances,*

$$\Lambda^*(x) = \begin{cases} \sup_{\lambda \geq 0}[\lambda x - \Lambda(\lambda)] \text{ if } x \geq \overline{x}, \\ \sup_{\lambda \leq 0}[\lambda x - \Lambda(\lambda)] \text{ if } x \leq \overline{x}, \end{cases} \tag{3.3}$$

which satisfies

- Λ^* *is decreasing on* $(-\infty, \overline{x}]$ *and increasing on* $[\overline{x}, \infty)$, *and*
- $\inf_{x\in\mathbb{R}} \Lambda^*(x) = 0$ *and if* $\overline{x} \in \mathbb{R}$, *then* $\Lambda^*(\overline{x}) = 0$.

(3) Λ *is differentiable in* $\mathring{\mathcal{D}}_\Lambda$ *with* $\Lambda'(\lambda) = \big(\mathbb{E}\big(e^{\lambda X_1}\big)\big)^{-1}\mathbb{E}\big(X_1 e^{\lambda X_1}\big)$ *and*

$$\Lambda'(\lambda) = x \Rightarrow \Lambda^*(x) = \lambda x - \Lambda(\lambda).$$

(4) If $0 \in \mathring{\mathcal{D}}_\Lambda$, *then* Λ^* *is a good rate function. Moreover, if* $\mathcal{D}_\Lambda = \mathbb{R}$, *then* $\lim_{|x|\mapsto\infty} \frac{\Lambda^*(x)}{|x|} = \infty$.

Proof. (1) By Hölder's inequality, given any $\lambda, \lambda' \in \mathbb{R}$ and any $t, t' \in [0,1]$ satisfying $t + t' = 1$,

$$\log \mathbb{E}\Big(e^{\langle t\lambda + t'\lambda', X_1\rangle}\Big) \leq t \log \mathbb{E}\big(e^{\langle \lambda, X_1\rangle}\big) + t' \log \mathbb{E}\Big(e^{\langle \lambda', X_1\rangle}\Big),$$

proving the convexity of Λ. The convexity of Λ^* follows from definition:

$$\sup_{\lambda\in\mathbb{R}}[\langle \lambda, tx + t'x\rangle - \Lambda(\lambda)] \leq t \sup_{\lambda\in\mathbb{R}}[\langle \lambda, x\rangle - \Lambda(\lambda)] + t' \sup_{\lambda\in\mathbb{R}}[\langle \lambda, x'\rangle - \Lambda(\lambda)].$$

To prove the lower semicontinuity, observe that for any $\lambda \in \mathbb{R}$ and any $x \in \mathbb{R}$,

$$\liminf_{y\to x} \langle \lambda, y\rangle - \Lambda(\lambda) = \langle \lambda, x\rangle - \Lambda(\lambda),$$

implying $\liminf_{y\to x} \Lambda^*(y) \geq \Lambda^*(x)$. The non-negativity of Λ^* follows from the fact $\Lambda^*(x) \geq \langle 0, x\rangle - \Lambda(0) = 0$.

(2) The case $\mathcal{D}_\Lambda = \{0\}$ is automatic. If $\Lambda(\lambda) < \infty$ for some $\lambda > 0$, then, due to the fact that $e^x \geq x$ for all $x \geq 0$,

$$\mathbb{E}\Big[X_1 \mathbb{1}_{\{X_1 \geq 0\}}\Big] \leq \lambda^{-1}\mathbb{E}\Big[e^{\langle \lambda, X_1\rangle}\mathbb{1}_{\{X_1\geq 0\}}\Big] \leq \lambda^{-1}\exp(\Lambda(\lambda)) < \infty,$$

meaning both $\mathbb{E}(\langle \lambda, X_1\rangle)$ and $\Lambda(\lambda)$ are well-defined. Hence, by Jensen's inequality,

$$\overline{x} = \lambda^{-1} \log \circ \exp(\mathbb{E}(\langle \lambda, X_1\rangle)) \leq \lambda^{-1}\Lambda(\lambda) < \infty.$$

The arguments proceeds similarly for $\Lambda(\lambda) > -\infty$ whenever $\lambda < 0$.

We then proceed to prove (3.3) and its related properties. Since now $\overline{x}$ is well-defined, by Jensen's inequality, for all $\eta < 0$ and $x \geq \overline{x}$ (similar for $\eta > 0$ and $x \leq \overline{x}$),

$$\langle \eta, x\rangle - \Lambda(\eta) = \log \mathbb{E}\big(e^{\langle \eta, x - X_1\rangle}\big) \leq \langle \eta, x - \overline{x}\rangle \leq 0 = \langle 0, x\rangle - \Lambda(0), \tag{3.4}$$

from which (3.3) follows. Moreover, (3.3) implies the monotonicity on $(\overline{x}, \infty)$ and $(\infty, \overline{x})$. Finally, if $\overline{x} \in \mathbb{R}$, then by (3.4), $\inf_{x\in\mathbb{R}} \Lambda^*(x) = \Lambda^*(\overline{x}) = 0$. If $\overline{x} = -\infty$ (similar for $\overline{x} = -\infty$), we deduce from Markov's inequality that for all $\lambda \geq 0$

$$\log \mu[x, \infty) \leq \log e^{-\langle \lambda, x\rangle}\mathbb{E}\big(e^{\langle \lambda, X_1\rangle}\big),$$

from which it follows that $0 \leq \inf_{x\in\mathbb{R}} \Lambda^*(x) \leq \lim_{x\to\overline{x}} \Lambda^*(x) \leq 0$ as desired.

(3) The differentiability follows from the dominated convergence theorem. Let $f_\varepsilon(x) = \frac{e^{(\lambda+\varepsilon)x} - e^{\lambda x}}{\varepsilon}$ so that $f_\varepsilon(x) \to xe^{\lambda x}$ as $\varepsilon \to 0$ and $|f_\varepsilon(x)| \leq \frac{e^{\lambda x}\left(e^{\delta\,|x|} - 1\right)}{\delta} =: h_\delta(x)$ whenever $|\varepsilon| < \delta$. Since $\mathbb{E}(h_\delta(X_1)) < \infty$ for all sufficiently small δ, we may apply the dominated convergence theorem to derive the derivative of Λ. Finally, the function

$g(\eta) := \langle \eta, x\rangle - \Lambda(\eta)$ is concave and thus $g'(\lambda) = 0$ implies $g(\lambda) = \sup_{\eta\in\mathbb{R}} g(\eta)$, proving the proposed property.

(4) Suppose $[\lambda_-, \lambda_+] \subset \mathcal{D}_\Lambda$ is a non-degenerate interval containing 0. Then, for $\lambda \in [\lambda_-, \lambda_+]$

$$\liminf_{|x|\mapsto\infty} \frac{\Lambda^*(x)}{|x|} \geq \liminf_{|x|\mapsto\infty} \left[\lambda \operatorname{sign}(x) - \frac{\Lambda(\lambda)}{|x|}\right] \geq \min\{-\lambda_-, \lambda_+\} > 0,$$

implying $\Lambda^*(x) \to \infty$ as $|x| \to \infty$. Hence, the sublevel set $\Psi_{\Lambda^*}(\alpha)$ is closed and bounded for all $\alpha < \infty$, and Λ^* is good. When $\mathcal{D}_\Lambda = \mathbb{R}$, then the result follows by letting $\lambda_+ = -\lambda_- \to \infty$. □

Proof of Theorem 3.3. (1) For economy, let $I_F = \inf_{x\in F} \langle \lambda, x\rangle - \Lambda(\lambda)$. If $I_F = 0$, then the inequality is trivial. Hence, we assume $I_F > 0$. Under the circumstances, the numbers $x_+ = \inf [\overline{x}, \infty) \cap F$ and $x_- = \sup(-\infty, \overline{x}] \cap F$ are different from $\overline{x}$. By Markov's inequality, for all $\lambda > 0$ and $\lambda' < 0$,

$$\mu_n(F) \leq \mu_n[x_+, \infty) + \mu_n(-\infty, x_-] \leq e^{-n(\langle \lambda, x_+\rangle - \Lambda(\lambda))} + e^{-n(\langle \lambda', x_-\rangle - \Lambda(\lambda'))}.$$

By Lemma 3.4,

$$\mu_n(F) \leq e^{-n\Lambda^*(x_+)} + e^{-n\Lambda^*(x_-)} \leq 2e^{I_F}.$$

Taking the normalized logarithm and letting $n \to \infty$ yields the desired inequality.

(2) It suffices to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(-\delta, \delta) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0), \tag{3.5}$$

for it is once proved, one may simply consider the translation $Y = X - x$ to deduce that the logarithmic moment generating function $\Lambda_Y(\lambda) = \Lambda(\lambda) - \langle \lambda, x\rangle$ and that $\Lambda_Y^*(z) = \Lambda^*(z + x)$, which in turn imply

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

This proves the desired lower bound.

To prove (3.5), assume first that (a) $\mu(-\infty, 0) > 0$, (b) $\mu(0, \infty) > 0$, and (c) μ is boundedly supported. Under the circumstances, (a) and (b) imply $\Lambda(\lambda) \to \infty$ as $|\lambda| \to \infty$, and (c) implies Λ is finite on $\mathbb{R}$. Consequently, by Lemma 3.4, there exists $\eta \in \mathbb{R}$ such that

$$\Lambda'(\lambda) = 0 \quad \text{and} \quad \Lambda(\lambda) = \inf_{\eta\in\mathbb{R}} \Lambda(\eta).$$

define a probability measure μ by

$$\frac{d\tilde{\mu}}{d\mu}(x) = e^{\langle \lambda, x\rangle - \Lambda(\lambda)}.$$

Observe that

$$\begin{aligned}\mu_n(-\varepsilon,\varepsilon) &= \int_{|\sum_{i=1}^n x_i|<n\varepsilon} d\mu(x_1)\cdots d\mu(x_n)\\ &= \int_{|\sum_{i=1}^n x_i|<n\varepsilon} e^{\sum_{i=1}^n[-\langle\lambda,x_i\rangle+\Lambda(\lambda)]} d\tilde{\mu}(x_1)\cdots d\tilde{\mu}(x_n).\\ &\geq e^{n\Lambda(\lambda)-n\varepsilon|\lambda|}\tilde{\mu}_n(-\varepsilon,\varepsilon).\end{aligned}$$

By Lemma 3.4,

$$\mathbb{E}_{\tilde{X}\sim\tilde{\mu}}(\tilde{X}) = \int x e^{\langle\lambda,x\rangle-\Lambda(\lambda)} d\mu(x) = \big(\mathbb{E}(e^{\langle\lambda,x\rangle})\big)^{-1}\mathbb{E}\big(X_1 e^{\langle\lambda,X_1\rangle}\big) = \Lambda'(\lambda) = 0.$$

Hence, by the law of large numbers,

$$\lim_{n\to\infty}\tilde{\mu}_n(-\varepsilon,\varepsilon) = 1.$$

Hence, for all $0<\varepsilon<\delta$,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(-\delta,\delta) \geq \liminf_{n\to\infty}\frac{1}{n}\log\mu_n(-\varepsilon,\varepsilon) \geq \Lambda(\lambda)-\varepsilon|\lambda|,$$

proving (3.5) by letting $\varepsilon\to 0$.

Now if support of μ is not bounded, fix $M>0$ such that (a) $\mu[-M,0)>0$, (b) $\mu(0,M]>0$. Hence, by letting ν be the normalized law of X_1 on $\{|X_1|\leq M\}$, we have that

$$\begin{aligned}\mu_n(-\delta,\delta) &\geq \int_{\frac{1}{n}\sum_{i=1}^n x_i\in(-\delta,\delta)} \prod_{i=1}^n \mathbb{1}_{[-M,M]}(x_i) d\mu(x_1)\cdots d\mu(x_n)\\ &= \nu_n(-\delta,\delta)\cdot\mu[-M,M]^n\end{aligned}$$

and that the logarithmic moment generating function associated with ν is

$$\log\int_{-M}^{M} e^{\langle\lambda,x\rangle}d\mu(x) - \log\mu[-M,M] =: \Lambda_M(\lambda) - \log\mu[-M,M].$$

Hence, by the case of bounded support,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(-\varepsilon,\varepsilon) \geq \log\mu[-M,M] + \liminf_{n\to\infty}\frac{1}{n}\log\nu_n(-\varepsilon,\varepsilon) \geq \inf_{\lambda\in\mathbb{R}}\Lambda_M(\lambda),$$

and therefore, by writing $I_M = -\inf_{\lambda\in\mathbb{R}}\Lambda_M(\lambda)$ and $I^* = \limsup_{M\to\infty} I_M$,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(-\varepsilon,\varepsilon) \geq -I^*. \tag{3.6}$$

Since Λ_M is increasing in M, so is $-I_M$. Therefore, $I^*>-\infty$ and the level sets $\{\lambda:\Lambda_M(\lambda)\leq -I^*\}$ are decreasing compacted sets, admitting some λ_0 in their intersection. By monotone convergence theorem,

$$\Lambda(\lambda_0) = \lim_{M\to\infty}\Lambda_M(\lambda_0) \leq -I^*. \tag{3.7}$$

Combining (3.6) and (3.7) proves (3.5).

Finally, if $\mu(-\infty,0)=0$ or $\mu(0,\infty)=0$, then Λ is monotone and $\inf_{\lambda\in\mathbb{R}}\Lambda(\lambda) = \log\mu(\{0\})$. Therefore, $\mu_n(-\delta,\delta)\geq\mu_n(\{0\}) = \mu(\{0\})^n$, implying (3.5).

(3) It follows from Lemma 3.4(4). □

3.2. **Cramér's theorem in $\mathbb{R}^d$.**

In this section, we prove the Cramér's theorem in higher dimensions. In contrast with its one-dimensional counterpart, we need additional assumptions for the full LDP.

Theorem 3.5 (Cramér). *Let X_i, μ_n, Λ, and Λ^* be as defined on $\mathbb{R}^d$ with $d \geq 2$. Assume further that $\mathcal{D}_\Lambda = \mathbb{R}^d$. Then, $\{\mu_n\}$ satisfies the LDP with the convex rate function Λ^*, namely,*

(1) For any closed set F, $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x)$.

(2) For any open set G, $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} \Lambda^*(x)$.

Furthermore,

(3) If $0 \in \mathring{\mathcal{D}}_\Lambda$, then Λ^* is good.

Remark 3.6. *The theorem will be revisited, if time permits, later with the following generalizations.*

(1) Even without the assumption $\mathcal{D}_\Lambda = \mathbb{R}^d$, the weak LDP holds.

(2) If $0 \in \mathring{\mathcal{D}}_\Lambda$, then the full LDP holds.

Our starting point is a number of properties regarding logarithmic generating function.

Lemma 3.7. *Let Λ and Λ^* be as defined. Then,*

(1) Λ is a convex function and Λ^* is a convex rate function.

(2) Λ is differentiable in $\mathring{\mathcal{D}}_\Lambda$ with $\nabla\Lambda(\lambda) = \big(\mathbb{E}\big(e^{\langle\lambda,X_1\rangle}\big)\big)^{-1}\mathbb{E}\big(X_1 e^{\langle\lambda,X_1\rangle}\big)$ and

$$\nabla\Lambda(\lambda) = x \Rightarrow \Lambda^*(x) = \langle\lambda, x\rangle - \Lambda(\lambda).$$

(3) If $0 \in \mathring{\mathcal{D}}_\Lambda$, then Λ^* is a good rate function. Moreover, if $\mathcal{D}_\Lambda = \mathbb{R}^d$, then $\lim_{|x|\mapsto\infty} \frac{\Lambda^*(x)}{|x|} = \infty$.

Proof. (1) Same as Lemma 3.4(1).

(2) The gradient can be obtained through a similar calculation as in Lemma 3.4(3), from which the differentiability follows naturally as it is continuous $\mathring{\mathcal{D}}_\Lambda$. To prove the remaining property, define

$$g_\eta(t) = \langle t\eta + (1-t)\lambda, x\rangle - \Lambda(t\eta + (1-t)\lambda)$$

to paraphrase the property as $g_\eta(1) \leq g_\eta(0)$ for all $\eta \in \mathcal{D}_\Lambda$ and all $t \in [0,1]$. To this end, observe that $g_\eta(t)$ is concave in t and that

$$g'_\eta(0) = \langle\eta - \lambda, x\rangle - \langle\eta - \lambda, \nabla\Lambda(\lambda)\rangle = 0,$$

proving the proposed inequality.

(3) Note that if $0 \in \mathring{\mathcal{D}}_\Lambda$, then there exists $\rho > 0$ such that the ρ-ball centered at the origin is contained in $\rho(\mathcal{D}_\Lambda)$ and

$$\Lambda^*(x) \geq \rho|x| - \sup_{|\lambda|\,=\rho} \Lambda(\lambda) \to \infty \text{ as } |x| \to \infty.$$

The case $\mathcal{D}_\Lambda = \mathbb{R}^d$ follows similarly. □

Definition 3.8. The δ-rate function associated with I is

$$I^\delta(x) = \min\Big\{I(x) - \delta, \frac{1}{\delta}\Big\}. \tag{3.8}$$

Proof of Theorem 3.5 for $\mathbb{R}^d$. (1) To prove the upper bound, it suffices to show that for every closed set F and every $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \le -\inf_{x\in F} I^\delta(x), \tag{3.9}$$

where I^δ is the δ-rate function associated with Λ^* defined as in (3.8).

We first prove (3.9) for any compact set F. Choose for each $x \in \Gamma$ an $\lambda_x \in \mathbb{R}^d$ and a open neighborhood A_x of x such that

$$\langle \lambda_x, x\rangle - \Lambda(x) \ge I^\delta(x),$$
$$\inf_{y\in A_x} \{\langle \lambda_x, y\rangle - \langle \lambda_x, x\rangle\} \ge -\delta,$$

Applying the Markov's inequality yields

$$\mu_n(A_x) \le \mathbb{E}\Big(e^{n\cdot\left(\langle \lambda_x, \cdot \hat{S}_n\rangle - \langle \lambda_x, x\rangle\right)}\Big) \exp\Big(-n \cdot \inf_{y\in A_x} \langle \lambda_x, y\rangle - \langle \lambda_x, x\rangle]\Big),$$

which in turn implies

$$\frac{1}{n} \log \mu_n(A_x) \le \delta - [\langle \lambda_x, x\rangle - \Lambda(\lambda_x)].$$

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

$$\limsup_{\varepsilon\to 0} \frac{1}{n} \log \mu_n(F) \le \delta - \min_{1\le i\le N} I^\delta(x_i) \le \delta - \inf_{x\in F} I^\delta(x),$$

proving (3.9) by letting $\delta \to 0$.

It remains to show that μ_n is an exponentially tight family (See Remark 1.6). Let $H_\rho = [-\rho, \rho]^d$ and μ^j denote the marginal law of μ on the j-th coordinate, so that $\mu_n\big(H_\rho^c\big) \le \sum_{j=1}^d \mu_n^j[-\rho,\rho]^c$. Since $0 \in \mathring{\mathcal{D}}_\Lambda$, Lemma 3.7 then implies

$$\lim_{\rho\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mu_n\big(H_\rho^c\big) \le -\infty.$$

(2) For the lower bound, we need to prove that for every $x \in \mathcal{D}_{\Lambda^*}$ and every $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B_\delta(x)) \ge -\Lambda^*(y). \tag{3.10}$$

To prove (3.10), assume first that $x = \nabla\Lambda(\lambda)$ for some $\lambda \in \mathbb{R}^d$ and define a probability measure μ by

$$\frac{d\tilde{\mu}}{d\mu}(y) = e^{\langle \lambda, y\rangle - \Lambda(\lambda)}.$$

Observe that

$$\begin{aligned}\frac{1}{n} \log \mu_n(B_\delta(x)) &= \Lambda(\lambda) - \langle \lambda, x\rangle + \frac{1}{n} \log \int_{B_\delta(x)} e^{n\cdot\langle \lambda, x-y\rangle} d\tilde{\mu}_n(y)\\ &\ge \Lambda(\lambda) - \langle \lambda, x\rangle - \delta|\lambda| + \frac{1}{n} \log \tilde{\mu}_n(B_\delta(x)).\end{aligned}$$

By the dominated convergence theorem and Lemma 3.7,

$$\mathbb{E}_{\tilde{X}_1\sim\tilde{\mu}}(\tilde{X}_1) = \int xe^{\langle\lambda,x\rangle-\Lambda(\lambda)}d\mu(x) = \big(\mathbb{E}(e^{\langle\lambda,x\rangle})\big)^{-1}\mathbb{E}\big(X_1e^{\langle\lambda,X_1\rangle}\big) = \nabla\Lambda(\lambda) = y,$$

and thus, by the weak law of large numbers, $\lim_{n\to\infty}\log\tilde{\mu}_n(B_\delta(y)) = 1$. Consequently,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(B_\delta(y)) \geq \Lambda(\lambda) - \langle\lambda,x\rangle - \delta|\lambda| = -\Lambda^*(x) - \delta|\lambda|,$$

which proves (3.10) as

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(B_\delta(y)) \geq \lim_{\delta\to0}\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(B_\delta(y)) \geq -\Lambda^*(x).$$

To extend the lower bound (3.10) to cover any $x \in \mathcal{D}_{\Lambda^*}$, we introduce i.i.d. multivariate normal random variables V_i that are independent of X_i and define $Y_i = X_i + \frac{V_i}{\sqrt{M}}$, $M > 0$. Denote by Λ_M the logarithmic moment generating function of Y_1 and ν_n the law governing $\hat{S}^{(M)} = \frac{1}{n}\sum_{j=1}^n Y_j$. Clearly,

$$\Lambda_M(\eta) = \Lambda(\eta) + \frac{1}{2M}\ |\eta|^2,$$

and

$$\Lambda^*(x) \geq \sup_{\eta\in\mathbb{R}^d}[\langle\eta,x\rangle - \Lambda_M(x)].$$

Since $0 \in \mathring{\mathcal{D}}_\Lambda$ implies $\overline{x}$ is finite, we have, by Jensen's inequality, $\langle\eta,x\rangle - \Lambda(x) \leq 0$ and

$$\limsup_{\rho\to\infty}\sup_{|\eta|>\rho}\langle\eta,x\rangle - \Lambda_M(\eta) \leq \limsup_{\rho\to\infty}\sup_{|\eta|>\rho} -\frac{1}{2M}\ |\eta|^2 = -\infty$$

Hence, the supremum of $g(\lambda) := \langle\lambda,x\rangle - \Lambda(\lambda)$ is attained at some finite $\lambda \in \mathbb{R}^d$, and, by differentiability of Λ_M,

$$0 = \nabla g(\lambda) = x - \nabla\Lambda_M(\lambda).$$

Thus, by the discussion above, for every $\delta > 0$,

$$\liminf_{n\to\infty}\frac{1}{n}\log\nu_n(B_\delta(x)) \geq -\Lambda^*(x) > -\infty. \tag{3.11}$$

Now that $\hat{S}_n^{(M)}$ has the same distribution as $\hat{S}_n + \frac{V}{\sqrt{Mn}}$, where $V \sim \text{Normal}(0, I)$ is independent of $\hat{S}_n$, we derive

$$\mu_n(B_{2\delta}(x)) \geq \nu_n(B_\delta(x)) - \mathbb{P}\Big(|V| \geq \sqrt{Mn}\delta\Big) \tag{3.12}$$

with

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(|V| \geq \sqrt{Mn}\delta\Big) \leq -\frac{M\delta^2}{2}. \tag{3.13}$$

Combining (3.11)-(3.13) proves (3.10). □