Large Deviations

April 10, 2002

This chapter serves as an introduction to the theory of large deviations. Sometimes theorems are not stated in its full generality to illustrate main ideas. Interested students should refer to the following textbooks:

1. Dembo, A. & Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer-Verlag, New York.

2. Dupuis, P. & Ellis, R. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley, New York.

Let $(X_1, X_2, \cdots)$ be a sequence of iid random variables with mean $\mu = \mathsf{E}X_1$. Define the sample mean

$$\bar{X}_n \doteq \frac{1}{n}\sum_{j=1}^{n} X_j; \qquad \forall \;\; n \geq 1.$$

It follows from SLLN that $\bar{X}_n \to \mu$ almost surely as $n \to \infty$. Thus, for any $A \subseteq \mathsf{R}$ such that $\mu \notin \bar{A}$ (the closure of A), we have

$$\mathsf{P}(\bar{X}_n \in A) \quad \to \quad 0;$$

in other words, $\{\bar{X}_n \in A\}$ is a rare event. The question is that *how fast is the decay of the probability of the rare event* $\{\bar{X}_n \in A\}$. We call such study of rare events **large deviations**, as opposed to the *small deviation*. For example, a typical set is $A = \{x \in \mathsf{R};\ |x-\mu| \geq \epsilon\}$ for some positive real number ϵ. The *small deviation* (of order $\frac{1}{\sqrt{n}}$) is indeed the CLT, which claims

$$\mathsf{P}\left(|\bar{X}_n - \mu| \geq \frac{\epsilon\sigma}{\sqrt{n}}\right) \quad \to \quad 2\big(1-\Phi(\epsilon)\big)$$

under mild conditions. On the other hand, from SLLN we know

$$\mathsf{P}(\bar{X}_n \in A) = \mathsf{P}\left(|\bar{X}_n - \mu| \geq \epsilon\right) \quad \to \quad 0;$$

and under some mild conditions, Cramér's theorem says the decay is exponential:

$$\mathsf{P}\left(|\bar{X}_n - \mu| \geq \epsilon\right) = e^{-n\left(K(A)+o(1)\right)} \qquad \text{or} \qquad \frac{1}{n}\log \mathsf{P}\left(\bar{X}_n \in A\right) \quad \to \quad K(A),$$

where $K(\cdot)$ is nicely characterized. Indeed,

$$K(A) = \inf_{x\in A} I(x), \qquad \text{where} \;\; I(x) = \sup_{\theta\in\mathsf{R}}\left[\theta x - \log \mathsf{E}e^{\theta X_1}\right]$$

The precise statement of Cramér theorem can be found below.

Remark: If we denote by μ_n the probability measure on $(\mathsf{R}, \mathcal{B}(\mathsf{R}))$ by $\bar{X}_n$, then we have

$$\mathsf{P}\left(\bar{X}_n \in A\right) = \mu_n(A).$$

Large deviation results are usually stated in terms of (μ_n).

We should start with the following example.

Example: Suppose $(X_1, X_2, \cdots)$ are iid $N(0, \sigma^2)$ random variables. With $S_n \doteq \sum_{j=1}^n X_j$, it follows that $S_n \sim N(0, n\sigma^2)$. Let $\bar{X}_n = \frac{S_n}{n}$, we have $\bar{X}_n \sim N(0, \sigma^2/n)$, and for every $\epsilon > 0$,

$$\mathsf{P}\left(\bar{X}_n \geq \epsilon\right) = \mathsf{P}\left(\frac{\sqrt{n}}{\sigma}\bar{X}_n \geq \frac{\sqrt{n}}{\sigma}\epsilon\right) = 1 - \Phi\left(\frac{\sqrt{n}}{\sigma}\epsilon\right).$$

Thanks to the inequality

$$\frac{x}{1+x^2}e^{-\frac{x^2}{2}} \leq \int_x^\infty e^{-\frac{z^2}{2}}\,dz \leq \frac{1}{x}e^{-\frac{x^2}{2}},$$

we have

$$\frac{1}{\sqrt{2\pi}}\frac{x}{1+x^2}e^{-\frac{x^2}{2}} \leq 1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}$$

or

$$-\log\sqrt{2\pi} - \log\left(\frac{\epsilon\sqrt{n}}{\sigma} + \frac{\sigma}{\epsilon\sqrt{n}}\right) - \frac{\epsilon^2 n}{2\sigma^2} \leq \log \mathsf{P}\left(\bar{X}_n \geq \epsilon\right) \leq -\log\sqrt{2\pi} - \log\frac{\epsilon\sqrt{n}}{\sigma} - \frac{\epsilon^2 n}{2\sigma^2},$$

which implies that

$$\frac{1}{n}\log \mathsf{P}\left(\bar{X}_n \geq \epsilon\right) = -\frac{\epsilon^2}{2\sigma^2}.$$

But in this case, $A = [\epsilon, \infty)$, and

$$I(x) = \sup_{\theta \in \mathsf{R}}\left(\theta x - \log \mathsf{E}e^{\theta X_1}\right) = \sup_{\theta \in \mathsf{R}}\left(\theta x - \frac{1}{2}\sigma^2\theta^2\right) = \frac{x^2}{2\sigma^2}.$$

Therefore,

$$K(A) = \inf_{x \in A} I(x) = \frac{\epsilon^2}{2\sigma^2},$$

which verify the Cramér theorem in this special case. □

1 Large deviation principle and Cramér theorem

Let X be a complete seperable metric space (e.g. R), and (μ_n) is a sequence of probability measures on $(X, \mathcal{B}(X))$.

Definition: We say (μ_n) satisfies the **large deviation principle** with *rate function* $I : X \to [0, \infty]$, if

1. the function I is lower-semicontinuous; that is, for every $x \in X$ and every sequence $(x_n) \subseteq X$ with $x_n \to x$, the inequality $I(x) \leq \liminf_n I(x_n)$ holds;

2. for every $a \in [0, \infty]$, the set $\{x \in X;\ I(x) \leq a\}$ is compact in X;
3. for each closed set $C \subseteq X$,

$$\limsup_n \frac{1}{n} \log \mu_n(C) \leq - \inf_{x \in C} I(x);$$

4. for each open set $G \subseteq X$,

$$\liminf_n \frac{1}{n} \log \mu_n(G) \geq - \inf_{x \in G} I(x);$$

Remark: One can establish easily that if $A \in \mathcal{B}(X)$ such that

$$\inf_{x \in A^o} I(x) = \inf_{x \in A} I(x) = \inf_{x \in \bar{A}} I(x),$$

then

$$\lim_n \frac{1}{n} \log \mu_n(A) = - \inf_{x \in A} I(x).$$

Here A^o and $\bar{A}$ are the *interior* and *closure* of set A, respectively.

Exercise: Show that the infimum of rate function I over closed set is always achieved.

Exercise: Show that the supremum of any collection of lower-semicontinuous function is lower-semicontinuous.

The simplest setting of large deviation is as follows. Let (X_n) be a sequence of iid random variables, and denote by P_n the probability measure on $(\mathsf{R}, \mathcal{B}(\mathsf{R}))$ induced by the sample mean $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$. In particular, $\mathsf{P}_1 := \mathsf{P}$ stands for the probability measure induced by X_1 (the distribution). For simplicity, we assume that: (1) the moment generating

$$M(\theta) = \mathsf{E} e^{\theta X_1} = \int_{\mathsf{R}} e^{\theta x} d\mathsf{P}(x)$$

is finite for all θ; (2) the random variable is neither bounded from below nor from above, that is, if I is an closed interval and $\mathsf{P}(I) = 1$, then $I = \mathsf{R}$. We have the following result.

Cramér theorem: The sequnece (P_n) satisfies the large deviation principle with rate function

$$I(x) = \sup_{\theta \in \mathsf{R}} \big[\theta x - \log M(\theta)\big], \qquad \forall\ x \in \mathsf{R}.$$

Remark: The function $\log M(\theta)$ is convex (check), and I is indeed the conjuate of $\log M(\theta)$.

Proof. It is not difficult to see that the function I satisfies (1), (2) of the large deviation principle. It remains to show (3) and (4). Let $\mu \doteq \mathsf{E} X_1$. It follows from Jensen's inequality that

$$\log M(\theta) = \log \left(\mathsf{E} e^{\theta X_1} \right) \geq \log \left(e^{\mathsf{E} \theta X_1} \right) = \theta \mu, \quad \forall\ \theta \in \mathsf{R},$$

and when $\theta = 0$, the equality holds. Therefore, $I(\mu) = 0$, in other words, $I(x)$ achieves its minimum at point $x = \mu$. However, I is clearly convex, whence it is non-decreasing for $x > \mu$ and non-increasing for $x < \mu$. Furthermore, it is not difficult to check that

$$I(x) = \begin{cases} \sup_{\theta>0}\big[\theta x - \log M(\theta)\big] & ; \quad \text{if } x > \mu \\ \sup_{\theta<0}\big[\theta x - \log M(\theta)\big] & ; \quad \text{if } x < \mu \end{cases}$$

(3) upper bound: take an interval $J_y \doteq [y, \infty)$ for $y > \mu$. Then for every $\theta > 0$ we have

$$\mathsf{P}_n(J_y) = \int_{[y,\infty)} d\mathsf{P}_n(x) \le e^{-n\theta y} \int_{[y,\infty)} e^{n\theta x} d\mathsf{P}_n(x) \le e^{-n\theta y} \mathsf{E} e^{-n\theta \bar{X}_n} = e^{-n\theta y}\big(M(\theta)\big)^n,$$

which implies that

$$\frac{1}{n} \log \mathsf{P}_n(J_y) \le -\theta y + \log M(\theta), \qquad \forall\, \theta > 0,$$

or

$$\limsup_n \frac{1}{n} \log \mathsf{P}_n(J_y) \le -\sup_{\theta>0}\big[\theta y - \log M(\theta)\big] = -I(y) = -\inf_{x\in J_y} I(x).$$

Similarly,

$$\limsup_n \frac{1}{n} \log \mathsf{P}_n(\tilde{J}_y) \le -\inf_{x\in \tilde{J}_y} I(x)$$

where $\tilde{J}_y \doteq (-\infty, y]$ for $y < \mu$. Let C be an arbitrary closed set. If $\mu \in C$, then $-\inf_{x\in C} I(x) = -I(\mu) = 0$ and the inequality (3) is trivial. Now assume $\mu \notin C$. Let (y_1, y_2) be the largest interval around μ such that $C \cap (y_1, y_2) = \emptyset$. It follows that $C \subseteq \tilde{J}_{y_1} \cup J_{y_2}$, and

$$\begin{aligned} \limsup_n \frac{1}{n} \log \mathsf{P}_n(C) &\le \limsup_n \frac{1}{n} \log \mathsf{P}_n(\tilde{J}_{y_1} \cup J_{y_2}) \\ &= \max\left(\limsup_n \frac{1}{n} \log \mathsf{P}_n(\tilde{J}_{y_1}),\ \limsup_n \frac{1}{n} \log \mathsf{P}_n(J_{y_2})\right) \qquad \text{(see exercise below)} \\ &\le -\min\left[I(y_1),\ I(y_2)\right] = -\inf_{x\in C} I(x). \end{aligned}$$

(4) lower bound: In order to prove the lower bound, it is sufficient to show that for every $\delta > 0$ and interval $U \doteq (y - \delta, y + \delta)$ for an arbitrary $y \in \mathsf{R}$, we have

$$\liminf_n \frac{1}{n} \log \mathsf{P}_n(U) \ge -I(y).$$

¿From the assumption that X is neither bounded from below nor from above, one can check (exercise!) that

$$\lim_{\theta\to\infty} \frac{\log M(\theta)}{|\theta|} = \infty$$

and

$$I(y) = \sup_{\theta\in\mathsf{R}} \left[\theta y - \log M(\theta)\right]$$

is achieved at some $\theta_0 \in \mathsf{R}$, with

$$\frac{M'(\theta_0)}{M(\theta_0)} = y.$$

The trick here is *change of probability measure*, such that under the new probability measure, the sequence (X_n) is iid with mean y. Define the following measure

$$\mathsf{P}^{(\theta_0)}(B) \doteq \frac{1}{M(\theta_0)} \int_B e^{\theta_0 x}\, d\mathsf{P}(x) \quad \text{or} \quad \frac{d\mathsf{P}^{(\theta_0)}}{d\mathsf{P}}(x) = \frac{1}{M(\theta_0)} e^{\theta_0 x}.$$

The mean under this probability measure is

$$\int_{\mathsf{R}} x\, d\mathsf{P}^{(\theta_0)}(x) = \int_{\mathsf{R}} x \cdot \frac{1}{M(\theta_0)} e^{\theta_0 x} d\mathsf{P}(x) = \frac{M'(\theta_0)}{M(\theta_0)} = y,$$

and it follows from law of large numbers that, for every $\epsilon > 0$,

$$\int\!\!\int_{\left\{\left|\frac{x_1+\cdots+x_n}{n} - y\right| < \epsilon\right\}} d\mathsf{P}^{(\theta_0)}(x_1) \cdots d\mathsf{P}^{(\theta_0)}(x_n) \quad \to \quad 1.$$

Now, for any $0 < \epsilon < \delta$, we have

$$\begin{aligned}
\mathsf{P}_n(U) &= \int\!\!\int_{\left\{\left|\frac{x_1+\cdots+x_n}{n} - y\right| < \delta\right\}} d\mathsf{P}(x_1) \cdots d\mathsf{P}(x_n) \\
&\geq \int\!\!\int_{\left\{\left|\frac{x_1+\cdots+x_n}{n} - y\right| < \epsilon\right\}} d\mathsf{P}(x_1) \cdots d\mathsf{P}(x_n) \\
&\geq e^{-n(\theta_0 y + \epsilon|\theta_0|)} \int\!\!\int_{\left\{\left|\frac{x_1+\cdots+x_n}{n} - y\right| < \epsilon\right\}} e^{\theta_0(x_1+\cdots+x_n)}\, d\mathsf{P}(x_1) \cdots d\mathsf{P}(x_n) \\
&= e^{-n(\theta_0 y + \epsilon|\theta_0|)} \cdot \big(M(\theta_0)\big)^n \int\!\!\int_{\left\{\left|\frac{x_1+\cdots+x_n}{n} - y\right| < \epsilon\right\}} d\mathsf{P}^{(\theta_0)}(x_1) \cdots d\mathsf{P}^{(\theta_0)}(x_n).
\end{aligned}$$

It follows that

$$\liminf_n \frac{1}{n} \log \mathsf{P}_n(U) \geq -\theta_0 y - \epsilon|\theta_0| + \log M(\theta_0).$$

Letting $\epsilon \to 0$, we have

$$\liminf_n \frac{1}{n} \log \mathsf{P}_n(U) \geq -\theta_0 y + \log M(\theta_0) = -I(y).$$

This completes the proof. $\square$

Exercise: Suppose (x_n), (y_n) are sequences of non-negative real numbers. Show that

$$\limsup_n \frac{1}{n} \log(x_n + y_n) = \max\left(\limsup_n \frac{1}{n} \log x_n,\ \limsup_n \frac{1}{n} \log y_n\right).$$

In particular, if (P_n) is a sequence of probability measures and

$$\phi(B) \doteq \limsup_n \frac{1}{n} \log \mathsf{P}_n(B), \qquad \forall\ B \in \mathcal{F},$$

then

$$\phi\left(\cup_{j=1}^m B_j\right) = \max_{1 \leq j \leq m} \phi(B_j).$$

Remark: The Cramér theorem holds in much weaker conditions; see e.g. Dembo & Zeitouni (1998).

Example: Suppose (X_n) is a sequence of iid Binomial random variables with

$$\mathsf{P}(X_1 = 1) = \mathsf{P}(X_1 = 0) = \frac{1}{2}.$$

Its moment generating function is

$$M(\theta) = \frac{1}{2}(1 + e^{\theta}) \quad \Rightarrow \quad \log M(\theta) = \log(1 + e^{\theta}) - \log 2,$$

and the rate function is

$$I(x) = \sup_{\theta} \big[\theta x - \log M(\theta)\big] = \begin{cases} \log 2 + x \log x + (1-x)\log(1-x) & ; \text{ if } x \in [0,1] \\ +\infty & ; \text{ if } x \notin [0,1] \end{cases}$$

Interested students might want to establish the large deviation principle directly for this special case by using Stirling formula:

$$n! = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \cdot e^{\epsilon_n}, \qquad \text{where } \frac{1}{12n+1} < \epsilon_n < \frac{1}{12n}; \quad \forall\, n \geq 2.$$

Example: Suppose that (X_n) is a sequence of iid exponential random variables with rate 1; i.e. the common density is $f(x) = e^{-x}$. It is easy to check that

$$M(\theta) = \frac{1}{1-\theta}, \quad \forall\, \theta < 1, \quad \text{and} \quad M(\theta) = \infty, \quad \forall\, \theta \geq 1,$$

and

$$I(x) = \sup_{\theta} \big[\theta x - \log M(\theta)\big] = \sup_{\theta < 1} \big[\theta x + \log(1-\theta)\big] = \begin{cases} x - 1 - \log x & ; \text{ if } x > 0 \\ +\infty & ; \text{ if } x \leq 0 \end{cases}$$

Interested students might want to establish the large deviation principle directly by using Stirling formula. Note that here the probability measure P_n has a density

$$d\mathsf{P}_n = \frac{n^n}{(n-1)!} e^{-nx} x^{n-1}\, dx, \qquad \forall\, x \geq 0.$$

2 Laplace principle

Another way of studying the large deviations is the following Laplace principle, which is equivalent to the large deviation principle. Combined with weak convergence, the Laplace principle is a very powerful tool, as exploited in Dupuis & Ellis (1997).

Definition: We say (μ_n) satisfies the **Laplace principle** with *rate function* $I : X \to [0, \infty]$, if

1. the function I is lower-semicontinuous; that is, for every $x \in X$ and every sequence $(x_n) \subseteq X$ with $x_n \to x$, the inequality $I(x) \leq \liminf_n I(x_n)$ holds;
2. for every $a \in [0, \infty]$, the set $\{x \in X;\ I(x) \leq a\}$ is compact in X;

3. for every bounded and continuous function $h : X \to \mathsf{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh(x)}\, d\mu_n(x) = -\inf_{x\in X} \big[h(x) + I(x)\big]$$

Theorem: The large deviation principle and Laplace principle are equivalent (with the same rate function).

Proof: "Large deviation principle $\Rightarrow$ Laplace principle". *Upper bound*: Suppose $-M \le h(x) \le M$ for some positive constant M. For an arbitrary positive integer k, and $1 \le j \le k$, define the closed sets

$$F_{k,j} \doteq \left\{ x \in X;\ -M + \frac{2(j-1)M}{k} \le -h(x) \le -M + \frac{2jM}{k} \right\}.$$

The lareg deviation upper bound implies that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh(x)}\, d\mu_n(x) &\le \limsup_{n\to\infty} \frac{1}{n} \log \left(\sum_{j=1}^k \int_{F_{k,j}} e^{-nh(x)}\, d\mu_n(x) \right) \\
&= \max_j \left(\limsup_n \frac{1}{n} \log \int_{F_{k,j}} e^{-nh(x)} \mu_n(F_{k,j}) \right) \\
&\le \max_j \left(-M + \frac{2jM}{k} - \inf_{x\in F_{k,j}} I(x) \right) \\
&\le \max_j \sup_{x\in F_{k,j}} (-h(x) - I(x)) + \frac{2M}{k} \\
&= \sup_{x\in X} (-h(x) - I(x)) + \frac{2M}{k}.
\end{aligned}$$

Letting $k \to \infty$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh(x)}\, d\mu_n(x) \le -\inf_{x\in X} \big[h(x) + I(x)\big].$$

Lower bound: For any $x \in X$, consider an open set

$$G \doteq \{ y \in X;\ h(y) < h(x) + \epsilon \},$$

here ϵ is an arbitrary positive constant. It follows that

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh(y)}\, d\mu_n(y) &\ge \limsup_{n\to\infty} \frac{1}{n} \log \int_G e^{-nh(y)}\, d\mu_n(y) \\
&\ge -h(x) - \epsilon + \liminf_n \frac{1}{n} \log \mu_n(G) \ge -h(x) - \epsilon - \inf_{y\in G} I(y) \\
&\ge -h(x) - I(x) - \epsilon.
\end{aligned}$$

Letting $\epsilon \to 0$, we complete the proof.

"Laplace principle $\Rightarrow$ Large deviation principle". *Upper bound:* Suppose F is an arbitrary closed set. Let $d(x, F)$ denote the distant between $x \in X$ and the closed set F. For every $k \in \mathsf{N}$, define

$$h_k(x) \doteq k\big(d(x, F) \wedge 1\big), \quad \forall\, x \in X.$$

It is easy to check that h_k is a bounded, continuous function, taking value 0 for all $x \in F$. Therefore, we have

$$- \inf_{x \in X} \big[h_k(x) + I(x)\big] = \lim_n \frac{1}{n} \log \int_X e^{-nh_k(x)} \, d\mu_n(x) \geq \limsup_n \frac{1}{n} \log \mu_n(F), \qquad \forall \; k.$$

It remains to show that

$$\lim_{k \to \infty} \inf_{x \in X} \big[h_k(x) + I(x)\big] = \inf_{x \in F} I(x).$$

Clearly, we have

$$\limsup_k \inf_{x \in X} \big[h_k(x) + I(x)\big] \leq \limsup_k \inf_{x \in F} \big[h_k(x) + I(x)\big] = \inf_{x \in F} I(x).$$

Now we only need to show

$$\liminf_k \inf_{x \in X} \big[h_k(x) + I(x)\big] \geq \inf_{x \in F} I(x).$$

Since

$$\inf_{x \in X} \big[h_k(x) + I(x)\big] = \inf_{x \in F} I(x) \wedge \inf_{x \notin F} \big[h_k(x) + I(x)\big],$$

it suffices to show

$$\liminf_k \inf_{x \notin F} \big[h_k(x) + I(x)\big] \geq \inf_{x \in F} I(x).$$

Suppose that the above inequality does not hold. In other words, there exists a sufficiently small positive number $\epsilon > 0$ such that

$$h_k(x_k) + I(x_k) \leq \inf_{x \in F} I(x) - \epsilon$$

for a subsequence of $k \in \mathsf{N}$ (still denoted by k), and a sequence $(x_k) \subseteq F^c$. (Note, if $\inf_{x \in F} I(x) = \infty$, we should replace the RHS by some constant M, and the rest of the proof should be similarly modified without difficulty). It is now easy to see that

$$d(x_k, F) \quad \to \quad 0, \qquad \text{as } k \to \infty.$$

Furthermore, the sequence (x_k) belong to the *compact level set*

$$C \doteq \left\{ y \in X; \quad I(y) \leq \inf_{x \in F} I(x) - \epsilon \right\}.$$

Therefore, there exists a further subsequence of (x_k) that converges to, say $x^* \in C$. However, it also holds that $x^* \in F$ since $d(x^*, F) = \lim d(x_k, F) = 0$. We have

$$\inf_{x \in F} I(x) \leq I(x^*) \leq \inf_{x \in F} I(x) - \epsilon.$$

A contradiction.

Lower bound: Suppose G is an open set. We should assume $\inf_{x \in G} I(x) < \infty$; otherwise the lower bound automatically holds. For any $x \in G$, there exists a constant $\epsilon > 0$ such that $B(x; \epsilon) \subseteq G$;

here $B(x;\epsilon)$ is the open ball with radius ϵ and center x. Let $M > 0$ be an arbitrary constant, and consider

$$h(y) \doteq M \cdot \left(\frac{d(y,x)}{\epsilon} \wedge 1 \right).$$

Clearly h is bounded and continuous, with $h(x) = 0$ and $h(y) = M$ for $y \notin B(x;\epsilon)$. Moreover, we have

$$\int_X e^{-nh(y)}\, d\mu_m(y) \le e^{-nM} \mu_n\left(B(x;\epsilon)^c\right) + \mu_n\big(B(x;\epsilon)\big) \le e^{-nM} + \mu_n\big(B(x;\epsilon)\big).$$

It follows from the assumption that

$$\begin{aligned} -I(x) &= -h(x) - I(x) \le -\inf_{y\in X}\big[h(y) + I(y)\big] \\ &\le \liminf_n \frac{1}{n} \log\left(e^{-nM} + \mu_n\big(B(x;\epsilon)\big)\right) = \max\left(-M, \liminf_n \frac{1}{n}\log \mu_n\big(B(x;\epsilon)\big)\right). \end{aligned}$$

Verify that the last equality holds (exercise!) Letting $M \to \infty$, we have

$$\liminf_n \frac{1}{n} \log \mu_n\big(B(x;\epsilon)\big) \ge -I(x) \quad \Rightarrow \quad \liminf_n \frac{1}{n}\log \mu_n(G) \ge -I(x).$$

The lower bound follows readily. □

Remark *(The uniqueness of the rate function I)***:** The rate function is unique; indeed, observe from the above proof that

$$\lim_{k\to\infty} \inf_{x\in X} \big[h_k(x) + I(x)\big] = \inf_{x\in F} I(x);$$

where

$$h_k(x) \doteq k\big(d(x,F) \wedge 1\big), \quad \forall\, x \in X.$$

Indeed, for any $y \in X$, choose $F = \{y\}$, we have a representation

$$I(y) = \lim_{k\to\infty} \inf_{x\in X}\big[h_k(x) + I(x)\big] = -\lim_{k\to\infty}\lim_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh_k(x)}\, d\mu(x).$$

Remark: Suppose for (μ_n), the Laplace limit

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh(x)}\, d\mu_n(x) = -\inf_{x\in X}\big[h(x) + I(x)\big]$$

holds for all bounded, *Lipschitz* continuous function h. Then (μ_n) satisfies the Large deviation principle and the Laplace principle. Indeed, in the above proof, all the tese functions $(h_k), h$ are bounded and Lipschitz. Therefore the Large deviation principle is implied, which further imply the Laplace principle.

3 A collection of exercises

Exercise: Suppose $h : [0,1] \to \mathsf{R}$ is a bounded and continuous function, then

$$\lim_n \frac{1}{n} \log \int_0^1 e^{-nh(x)}\, dx = -\min_{x\in[0,1]} h(x).$$

Contraction Principle: Suppose X, Y are two Polish spaces (i.e.; complete and seperable metric spaces), and $f : X \to Y$ is a continuous mapping. Suppose (μ_n) is a sequence of probability measures on X that satisfy the Laplace (large deviation) principle. Then the measures induced on Y through f, i.e.

$$\nu_n(B) \doteq \mu_n\left(x \in X;\quad f(x) \in B\right), \qquad \forall B \in \mathcal{B}(Y);$$

satisy the Laplace (large deviation) principle with rate function

$$J(y) = \inf\big\{I(x);\quad f(x) = y,\ x \in X\big\}, \qquad \text{with convention} \quad \inf\{\emptyset\} = +\infty.$$