CHAPTER I

LARGE DEVIATIONS FOR I.I.D. SEQUENCES: PART 1

Chapters I and II are devoted to large deviation theorems for i.i.d. real-valued random variables. The reader will get acquainted with the basic results via explicit computation.

I.1 Introduction

We begin our journey on familiar territory. Let $X_1, X_2, \ldots$ be i.i.d. random variables on a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$, where $\mathcal{B}(\mathbb{R})$ is the Borel sigma-field on $\mathbb{R}$. Write $\mathbb{E}$ to denote expectation under $\mathbb{P}$, let

$$\begin{aligned} \mathbb{E}\, X_1 &= \mu \in \mathbb{R}, \\ \mathrm{Var}\, X_1 &= \sigma^2 \in (0, \infty), \end{aligned}$$

and let $S_n = X_1 + \cdots + X_n$ $(n \in \mathbb{N})$ be the partial sums. In standard textbooks on probability theory two fundamental theorems dealing with such sequences can be found:

Strong Law of Large Numbers (SLLN)

$$\frac{1}{n} S_n \underset{n \to \infty}{\longrightarrow} \mu \qquad \mathbb{P}\text{–a.s.}$$

Central Limit Theorem (CLT)

$$\frac{1}{\sigma \sqrt{n}} (S_n - \mu n) \underset{n \to \infty}{\longrightarrow} Z \qquad \text{in law w.r.t. } \mathbb{P},$$

where Z is a standard normal random variable.

While the SLLN asserts that the empirical average $\frac{1}{n} S_n$ converges to μ as $n \to \infty$, the CLT quantifies the probability that S_n differs from μn by an amount of order $\sqrt{n}$. Deviations of this size are called "normal".

In these lecture notes we deal with events where S_n differs from μn by an amount of order n, so well beyond what is described by the CLT. Deviations of this size are called "large". An example is the event

$$\{S_n \geq (\mu + a) n\}, \qquad a > 0,$$

whose probability tends to zero as $n \to \infty$. It is our task to quantify the *rate* at which this occurs. We will see that, under a certain condition on the tail of the distribution of X_1, the decay is exponential in n:

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\Big(S_n \geq (\mu + a)\, n\Big) = -I(a) < 0, \qquad a > 0.$$

This is because, typically, a positive fraction of the components $X_1, \ldots, X_n$ in the sum S_n must deviate from μ to create the large deviation an away from μn. It

will turn out that knowledge of the rate function $a \mapsto I(a)$ is crucial for a correct evaluation of integrals of exponential functionals of S_n as $n \to \infty$.

As we go along, we also consider other functionals of $X_1, \ldots, X_n$, such as the empirical measure $L_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n})$, with δ_x the point-measure at $x \in \mathbb{R}$. It turns out that L_n has a large deviation behavior similar to that of $\frac{1}{n}S_n$ with its own rate function. We will go even further and look at large deviations for the empirical measure of words of length two and larger, thus obtaining results at ever higher levels of detail. Our goal will be to describe a theory for handling such large deviation questions in general.

Before we start we make an observation. Given two sequences of positive numbers (α_n) and (β_n), we write

$$\alpha_n \simeq \beta_n \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \frac{1}{n}(\log \alpha_n - \log \beta_n) = 0, \tag{I.1}$$

i.e., the symbol $\simeq$ means that the two sequences are logarithmically equivalent. The following elementary fact plays an important role throughout the exposition:

$$\alpha_n + \beta_n \simeq \alpha_n \vee \beta_n. \tag{I.2}$$

The reader should think of (I.2) as a "largest-exponent-wins" principle. Obviously, (I.2) can be iterated to apply to finitely many sequences.

I.2 An example: coin tossing

We begin with an example that serves as a warm-up.

THEOREM I.3 *Let (X_i) be i.i.d. random variables with $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$. Let $S_n = \sum_{i=1}^n X_i$. Then, for all $a > \frac{1}{2}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a),$$

where

$$I(z) = \begin{cases} \log 2 + z \log z + (1-z)\log(1-z) & \text{if } z \in [0,1], \\ \infty & \text{otherwise.} \end{cases}$$

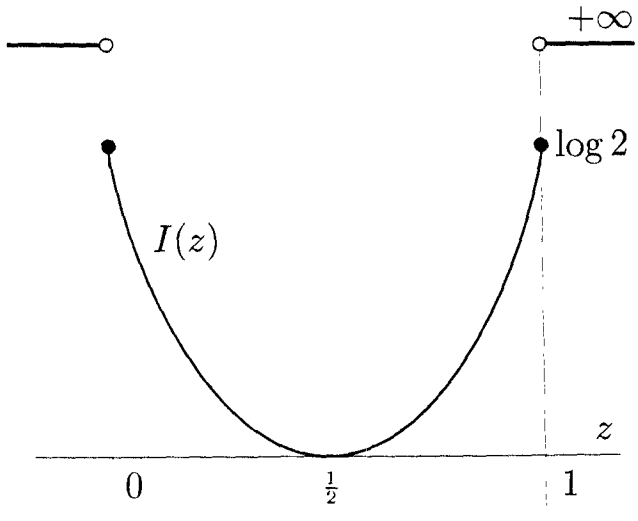


FIG. 1. The rate function for coin tossing

PROOF. The claim is trivial for $a > 1$. For $a \in (\frac{1}{2}, 1]$ we observe that $\mathbb{P}(S_n \geq an) = 2^{-n} \sum_{k \geq an} \binom{n}{k}$, which yields the estimate

$$2^{-n} Q_n(a) \leq \mathbb{P}(S_n \geq an) \leq (n+1) 2^{-n} Q_n(a),$$

where

$$Q_n(a) = \max_{k \geq an} \binom{n}{k}.$$

The maximum is attained at $k = \lceil an \rceil$, the smallest integer $\geq an$. Stirling's formula $n! = n^n e^{-n} \sqrt{2\pi n} \, (1 + O(\frac{1}{n}))$ therefore allows us to infer that

$$\lim_{n \to \infty} \frac{1}{n} \log Q_n(a) = -a \log a - (1-a) \log(1-a).$$

Noting that the upper and the lower bound merge on an exponential scale as $n \to \infty$, we arrive at the desired statement. □

Since $\mathbb{E} X_1 = \frac{1}{2}$ and $a > \frac{1}{2}$, Theorem I.3 deals with large deviations in the upward direction. It is clear from symmetry that the same holds for $\mathbb{P}(S_n \leq an)$ with $a < \frac{1}{2}$. This is manifested by the symmetry relation $I(1-z) = I(z)$.

The function $z \mapsto I(z)$ is called the *rate function*. Note that it is infinite outside $[0,1]$, finite and strictly convex inside $[0,1]$, and has a unique zero at $z = \frac{1}{2}$ (see Fig. 1). This zero corresponds to the SLLN. Indeed, Theorem I.3 implies that

$$\sum_{n \in \mathbb{N}} \mathbb{P}\big(|\tfrac{1}{n} S_n - \tfrac{1}{2}| > \delta\big) < \infty \qquad \forall \delta > 0,$$

and so the SLLN follows via the Borel-Cantelli lemma. The curvature of the rate function at $z = \frac{1}{2}$ corresponds to the CLT, as will become clear later.

I.3 Cramér's Theorem for the empirical average

We now formulate the first basic result of large deviation theory, which goes back to Cramér [D12]. This result, which generalizes Theorem I.3, identifies the large deviation behavior of the *empirical average* $\frac{1}{n} S_n$ under a certain condition on the tail of the distribution of X_1.

THEOREM I.4 *Let (X_i) be i.i.d. $\mathbb{R}$-valued random variables satisfying*

$$\varphi(t) = \mathbb{E}\, e^{tX_1} < \infty \qquad \forall t \in \mathbb{R}. \tag{I.5}$$

Let $S_n = \sum_{i=1}^n X_i$. Then, for all $a > \mathbb{E} X_1$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a),$$

where

$$I(z) = \sup_{t \in \mathbb{R}} [zt - \log \varphi(t)]. \tag{I.6}$$

PROOF. We may suppose without loss of generality that $a = 0$ and $\mathbb{E} X_1 < 0$. Namely, the substitution $X_1 \to X_1 + a$ gives $\varphi(t) \to e^{at} \varphi(t)$ and, consequently, $I(a) \to I(0)$. We may also suppose that X_1 is non-degenerate, because the claim is trivial otherwise.

EXERCISE I.7 *Assume $\mathbb{P}(X_1 = a) = 1$. Check that $I(a) = 0$ and $I(z) = \infty$ for $z \neq a$.*

Henceforth we abbreviate

$$\rho = \inf_{t \in \mathbb{R}} \varphi(t).$$

Note that $I(0) = -\log \rho$ (with $I(0) = \infty$ if $\rho = 0$), so we must prove that

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \log \rho.$$

Let $F(x) = \mathbb{P}(X_1 \leq x)$ be the distribution function of X_1. It follows from (I.5) that $\varphi \in C^\infty(\mathbb{R})$, the space of smooth functions, with

$$\begin{aligned} \varphi'(t) &= \int_{\mathbb{R}} x e^{tx} dF(x), \\ \varphi''(t) &= \int_{\mathbb{R}} x^2 e^{tx} dF(x). \end{aligned}$$

Hence, φ is strictly convex and $\varphi'(0) = \mathbb{E} X_1 < 0$. According to where the mass of $\mathbb{P}$ is situated, we may distinguish three subcases:

(i) $\mathbb{P}(X_1 < 0) = 1$.
Then φ is strictly decreasing and $\lim_{t \to \infty} \varphi(t) = \rho = 0$. Since $\mathbb{P}(S_n \geq 0) = 0$, the claim follows.

(ii) $\mathbb{P}(X_1 \leq 0) = 1$ and $\mathbb{P}(X_1 = 0) > 0$.
Then φ is strictly decreasing and $\lim_{t \to \infty} \varphi(t) = \rho = \mathbb{P}(X_1 = 0) > 0$. Since $\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_1 = \cdots = X_n = 0) = \rho^n$, the claim follows.

(iii) $\mathbb{P}(X_1 < 0) > 0$ and $\mathbb{P}(X_1 > 0) > 0$.
Then $\lim_{t \to \pm\infty} \varphi(t) = \infty$, and since φ is strictly convex there exists a unique $\tau \in \mathbb{R}$, satisfying $\tau > 0$, such that (see Fig. 2)

$$\varphi(\tau) = \rho, \quad \varphi'(\tau) = 0.$$

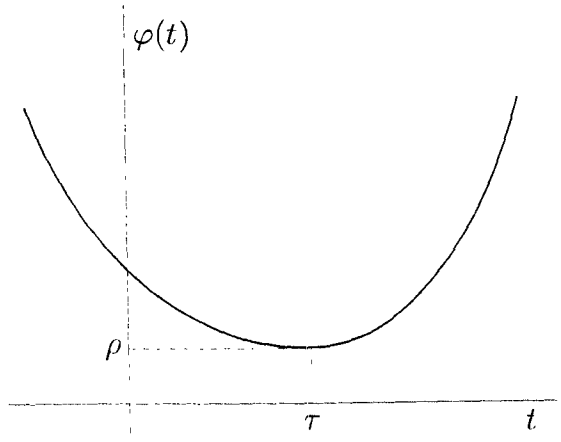


FIG. 2. Relation between φ, τ and ρ

We proceed with the analysis of (iii). By the exponential Chebyshev inequality, we have the following upper bound:

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(e^{\tau S_n} \geq 1) \leq \mathbb{E}\, e^{\tau S_n} = [\varphi(\tau)]^n = \rho^n.$$

Hence

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \leq \log \rho.$$

To get the lower bound we resort to a technique that will find an interesting generalization later on. Let $(\widehat{X}_i)$ be an i.i.d. sequence of random variables with distribution function

$$\widehat{F}(x) = \frac{1}{\rho}\int_{(-\infty,x]} e^{\tau y} dF(y).$$

The function $\widehat{F}$ is called the *Cramér-transform* of F. Note that $\rho = \varphi(\tau) = \int_{\mathbb{R}} e^{\tau y} dF(y)$. The proof proceeds via three lemmas:

LEMMA I.8 $\mathbb{E}\,\widehat{X}_1 = \widehat{\mu} = 0$ *and* $\operatorname{Var}\widehat{X}_1 = \widehat{\sigma}^2 \in (0,\infty)$.

PROOF. Let $\widehat{\varphi}(t) = \mathbb{E}\, e^{t\widehat{X}_1}$. Then

$$\widehat{\varphi}(t) = \int_{\mathbb{R}} e^{tx} d\widehat{F}(x) = \frac{1}{\rho}\int_{\mathbb{R}} e^{tx} e^{\tau x} dF(x) = \frac{1}{\rho}\varphi(t+\tau) < \infty \qquad \forall t \in \mathbb{R}.$$

This implies that $\widehat{\varphi} \in C^\infty(\mathbb{R})$ and

$$\mathbb{E}\,\widehat{X}_1 = \widehat{\varphi}'(0) = \tfrac{1}{\rho}\varphi'(\tau) = 0,$$

$$\operatorname{Var}\widehat{X}_1 = \widehat{\varphi}''(0) = \tfrac{1}{\rho}\varphi''(\tau) \in (0,\infty).$$

□

LEMMA I.9 *Let* $\widehat{S}_n = \sum_{i=1}^n \widehat{X}_i$. *Then* $\mathbb{P}(S_n \geq 0) = \rho^n \mathbb{E}\big(e^{-\tau\widehat{S}_n} 1_{\{\widehat{S}_n \geq 0\}}\big)$.

PROOF. Write

$$\begin{aligned}\mathbb{P}(S_n \geq 0) &= \int_{\{x_1+\cdots+x_n\geq 0\}} dF(x_1)\ldots dF(x_n)\\ &= \int_{\{x_1+\cdots+x_n\geq 0\}} [\rho e^{-\tau x_1} d\widehat{F}(x_1)]\ldots[\rho e^{-\tau x_n} d\widehat{F}(x_n)],\end{aligned}$$

which gives the expression. □

LEMMA I.10 $\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{E}\left(e^{-\tau\widehat{S}_n} 1_{\{\widehat{S}_n\geq 0\}}\right) \geq 0.$

PROOF. First we observe that, by Lemma I.8, the CLT can be applied to $\widehat{S}_n$. Pick any number $C > 0$ such that $\frac{1}{\sqrt{2\pi}}\int_0^C e^{-x^2/2}dx > \frac{1}{4}$ and estimate

$$\mathbb{E}\left(e^{-\tau\widehat{S}_n} 1_{\{\widehat{S}_n\geq 0\}}\right) \geq e^{-\tau C\widehat{\sigma}\sqrt{n}}\,\mathbb{P}\Big(\frac{\widehat{S}_n}{\widehat{\sigma}\sqrt{n}} \in [0,C)\Big).$$

The probability in the RHS is at least $\frac{1}{4}$ for n large enough. □

The proof of Theorem I.4 is now concluded by observing that Lemmas I.9 and I.10 provide the desired lower bound for case (iii):

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq 0) \geq \log\rho.$$

□

The following exercises identify the rate function I in (I.6) for a number of choices for the distribution of X_1.

EXERCISE I.11 *Compute I for the following distributions: Poisson, Exponential and Normal.*

EXERCISE I.12 *Deduce Theorem I.3 from Theorem I.4.*

EXERCISE I.13 *Compute I when $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 3) = \frac{1}{3}$. Note that the resulting formula is not as nice as for coin tossing.*

The rate function I in Theorem I.4 has the following properties:

LEMMA I.14 *Assume* (I.5). *Then:*
(i) *I is lower semi-continuous and convex on $\mathbb{R}$.*
(ii) *I has compact level sets.*
(iii) *I is continuous and strictly convex on $\text{int}(\mathcal{D}_I)$, where $\mathcal{D}_I = \{z \in \mathbb{R}\colon I(z) < \infty\}$ and $\text{int}(\mathcal{D}_I)$ is the interior of $\mathcal{D}_I$.*
(iv) *I is smooth on $\text{int}(\mathcal{D}_I)$.*
(v) *$I(z) \geq 0$ with equality if and only if $z = \mu$.*
(vi) *$I''(\mu) = \frac{1}{\sigma^2}$.*

REMARKS I.15
1. The level sets of I are the sets $I^{-1}([0, c]) = \{z \in \mathbb{R}\colon I(z) \leq c\}$ with $c \in [0, \infty)$.
2. Lower semi-continuity is equivalent to the level sets being closed.
3. The convexity of I implies that $\mathcal{D}_I$ is an interval (possibly infinite).

EXERCISE I.16 *Prove Lemma* I.14.

Thus, the typical picture of I is qualitatively similar as in Fig. 1, except that there are various possibilities for $\mathcal{D}_I$ and for the behavior of I near $\partial\mathcal{D}_I$, the boundary of $\mathcal{D}_I$ (see Exercise I.11).

I.4 Comments

Let us have a closer look at what has happened in Section I.3.

(1) The same statement as in Theorem I.4 holds for $\mathbb{P}(S_n \leq an)$ and $a < \mathbb{E}X_1$, with the same formula for I. This is easily checked via the mirror reflection $X_1 \to -X_1$.

(2) Note that the key idea in the proof of Theorem I.4 is the "exponential tilting" of the probability measure (through the Cramér transform) in combination with the CLT. Under the tilted probability measure the large deviation event $\{S_n \geq 0\}$ becomes "typical", as is seen from Lemmas I.8–I.10. This type of argument is very powerful and will recur in later chapters in various guises. In fact, it will play the role of a guiding principle, which will become clear as we go along.

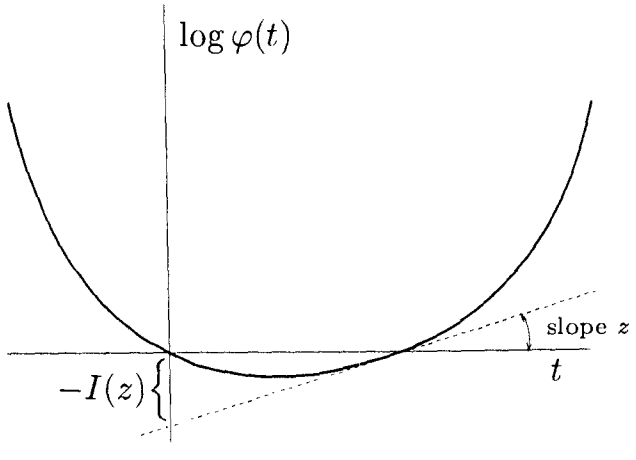


FIG. 3. Geometrical interpretation of the Legendre transform

(3) Equation (I.6) says that the rate function I is the *Legendre transform* of the cumulant generating function $\log\varphi$ (see Fig. 3). This important relationship will turn up later in various different guises. See, in particular, Chapter V.

(4) The requirement (I.5) in Theorem I.4 can be weakened to the condition that

$$0 \in \text{int}\,(\mathcal{D}_\varphi) \quad \text{with} \quad \mathcal{D}_\varphi = \{t \in \mathbb{R}\colon\ \varphi(t) < \infty\}. \tag{I.17}$$

We will not prove this generalization (see e.g. Dembo and Zeitouni [A2] Section 2.2.1). However, the reader can easily check that the proof in Section I.3 (and its mirror version for the large deviations in the downward direction) carries over when, in addition to (I.17), we have

$$\lim_{t\to\partial\mathcal{D}_\varphi:\ t\in\mathcal{D}_\varphi} \bigl|[\log\varphi]'(t)\bigr| = \infty, \tag{I.18}$$

a condition referred to as "$\log\varphi$ is steep at $\partial\mathcal{D}_\varphi$". Indeed, this property implies that $t \mapsto e^{at}\varphi(t)$ achieves a minimum in int $(\mathcal{D}_\varphi)$ for every $a \in \mathbb{R}$, which is needed to handle case (iii) after the substitution $X_1 \to X_1 + a$. See also Section V.2.

EXERCISE I.19 *Give examples for which* $\log\varphi$ *is: (i) steep; (ii) not steep.*

EXERCISE I.20 *Show that steepness of* $\log\varphi$ *implies* $\mathcal{D}_I = \mathbb{R}$.

(5) Theorem I.4 is in fact true even without (I.17), as can be shown with the help of a truncation argument (Bahadur [B1]). However, without (I.17) it looses much of its power. For instance, if $\mathcal{D}_\varphi = \{0\}$, then $I \equiv 0$. But in that case all Theorem I.4 tells us is that the large deviation probabilities decay slower than exponential. It gives us no information on how slow the decay actually is. Similarly, if $\mathcal{D}_\varphi$ contains a left-neighborhood of 0 but not a right-neighborhood, then $I(x) = 0$ for all $x \geq 0$, in which case Theorem I.4 gives interesting information only for large deviations that go downwards. Vinogradov [A9] gives a detailed analysis of all the subexponential cases. The decay in these cases depends heavily on the law of X_1 and in general is not captured by as elegant a formula as the Legendre transform in Theorem I.4.

(6) The following needs no condition at all.

EXERCISE I.21 *Use Hölder's inequality and Fatou's lemma to show that* $\log\varphi$ *is convex and lower-semicontinuous on* $\mathbb{R}$.

Because of these two properties we have the inverse formula (see Rockafellar [C7] Theorem 12.2)

$$\log\varphi(t) = \sup_{z\in\mathbb{R}}[tz - I(z)], \tag{I.22}$$

which is the dual of the formula in (I.6). This fact shows that φ and I are in a one-to-one relation. As soon as (I.17) holds, the law of X_1 is uniquely determined by φ. Thus, each law has its own unique rate function.

(7) Lemma I.14(v) is, of course, related to the SLLN (see the end of Section I.2). Lemma I.14(vi), on the other hand, is related to the CLT. Indeed, replacing a by $a_n = \mu + \frac{b\sigma}{\sqrt{n}}$ $(b > 0)$ and pretending that the result in Theorem I.4 is still applicable, we find

$$\mathbb{P}\Bigl(\frac{S_n - \mu n}{\sigma\sqrt{n}} \in db\Bigr) = e^{-nI(\mu+\frac{b\sigma}{\sqrt{n}})\,+\,\text{c.t.}}db = e^{-b^2/2\,+\,\text{c.t.}}db \qquad n \to \infty,$$

which is the standard normal density (modulo the normalization, which is hidden in the correction term). This argument can be made more precise by appealing to Lemmas I.8–I.10.

(8) Bahadur and Ranga Rao [D2] prove that if (I.17) holds and the law of X_1 is non-lattice (i.e., is not concentrated on a periodic grid), then the following refined estimate holds:

$$\mathbb{P}(S_n \geq an) = \frac{1}{\sqrt{2\pi\widehat{\sigma}(a)^2 n}}\, e^{-nI(a)}\,(1+o(1))\,, \qquad a \in \operatorname{int}(\mathcal{D}_I),\ a > \mu,$$

with $\widehat{\sigma}(a)^2 = \frac{[I'(a)]^2}{I''(a)}$, and similarly for $a \in \operatorname{int}(\mathcal{D}_I)$, $a < \mu$. Again, this behavior can be traced back to Lemmas I.8–I.10. In these lecture notes we will not address such "higher-order" results and focus on the leading exponential term only.

(9) By Lemma I.14(i,v), if $a > \mu$, then $I(z) \geq I(a)$ for all $z \geq a$. Hence the result of Theorem I.4 can be rewritten as the statement

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n} S_n \in A\Big) = -\inf_{z\in A} I(z) \qquad \text{with } A = [a,\infty). \tag{I.23}$$

This formula should be interpreted as saying that the large deviation $\{\frac{1}{n}S_n \in A\}$ is essentially carried by the event where $\frac{1}{n}S_n$ is close to $\bar{z}$, with $\bar{z}$ the minimizer of $I(z)$ on A (which happens to be $\bar{z} = a$ when $a > \mu$). Namely, the latter event costs $\exp[-nI(\bar{z}) + o(n)]$ and therefore is the cheapest realization of A. This fact illustrates a key principle in large deviation theory:

ANY LARGE DEVIATION IS DONE IN THE LEAST UNLIKELY OF ALL THE UNLIKELY WAYS!

Equation (I.23) of course suggests that the same statement holds when $A = [a,\infty)$ is replaced by a sufficiently nice subset of $\mathbb{R}$. This is indeed the case, but we will postpone such topological considerations to Chapter III.

EXERCISE I.24 *Show that* (I.23) *holds: (i) for all* $A = [a,\infty)$ *with* $a \in \mathbb{R}$; *(ii) for all* $A = (a,\infty)$ *with* $a \in \mathbb{R} \setminus \partial\mathcal{D}_I$; *(iii) for all* $A = (a,b)$ *with* $a, b \in \mathbb{R} \setminus \partial\mathcal{D}_I$.

I.5 A toy application of Cramér

Let us demonstrate how Theorem I.4 can be used to study integrals of exponential functionals of S_n. For that we cook up the following artificial example. A more serious application will be described in Chapter VI.

Let $\mathbb{P}(X_1 = \frac{1}{2}) = \mathbb{P}(X_1 = \frac{3}{2}) = \frac{1}{2}$. Suppose that we are interested in the behavior of $\mathbb{E}\left[(\frac{1}{n}S_n)^n\right]$ for large n. Naively we might think that $\frac{1}{n}\log \mathbb{E}\left[(\frac{1}{n}S_n)^n\right] \to 0$ because $\frac{1}{n}S_n \to 1$ by the SLLN. However, this intuition is wrong. To see why, we

compute (recall the notation introduced in (I.1))

$$\begin{aligned}
\mathbb{E}\left[(\tfrac{1}{n}S_n)^n\right] &= \int_0^\infty a^n\,\mathbb{P}(\tfrac{1}{n}S_n \in da) \\
&= \int_0^\infty n a^{n-1}\mathbb{P}(\tfrac{1}{n}S_n \geq a)\,da \\
&\simeq \int_0^\infty \exp\bigl\{n[\log a - J(a)]\bigr\}\,da \\
&\simeq \exp\Bigl\{n \sup_{a>0}[\log a - J(a)]\Bigr\},
\end{aligned}$$

where J is the rate function given by Cramér's Theorem, i.e., $J(a) = I(a - \frac{1}{2})$ with I as in Theorem I.3. Hence we find

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[(\tfrac{1}{n}S_n)^n\right] = b \quad \text{with} \quad b = \sup_{a>0}[\log a - J(a)].$$

EXERCISE I.25 *Let a^* be the maximizer of the variational expression in the last display. Show that $a^* \neq 1$ and $b > 0$.*

Thus, the expectation $\mathbb{E}\left[(\frac{1}{n}S_n)^n\right]$ is not dominated by the almost sure behavior, but rather by the rare event where $\frac{1}{n}S_n$ is in the vicinity of a^*.

The above calculation, which is nothing other than an example of Laplace's method for exponential integrals, can easily be made rigorous by appealing to (I.2). Indeed, cut $[\frac{1}{2}, \frac{3}{2}]$ into finitely many small intervals and apply (I.2) iteratively, in combination with the fact that I is continuous on $[\frac{1}{2}, \frac{3}{2}]$. We leave the details to the reader. In Section III.3 we will encounter a more powerful way to approach this kind of argument.