An Information-Spectrum Approach to Large Deviation Theorems

Te Sun HAN[†]

January 27, 2014

[†]Te Sun Han is with the Graduate School of Information Systems, University of Electro-Communications, Chofugaoka 1-5-1, Chofu, Tokyo 182-8585, Japan. E-mail: han@is.uec.ac.jp

Abstract: In this paper we show a some new look at large deviation theorems from the viewpoint of the information-spectrum (IS) methods, which has been first exploited in information theory, and also demonstrate a new basic formula for the large deviation rate function in general, which is a pair of the lower and upper IS rate functions. In particular, we are interested in establishing the general large deviation rate functions that can be derivable as the Fenchel-Legendre transform of the cumulant generating function. The final goal is to show a necessary and sufficient condition for the rate function to be of Cramér-Gärtner-Ellis type.

1 Introduction

The present paper is intended to show a basic new look at problems in large deviation theory. As is well known, in many fields such as probability theory, computer sciences, communication sciences, and cryptography, etc., we face the problem to pertinently evaluate behaviors of the tail probability of asymptotic distributions, and it is now not only of theoretical importance but also of practical interest.

In such a situation, therefore, it would also be useful to elucidate and look into *basic* common and/or "elementary" properties as well as structures underlying the large deviation problems.

For this purpose, as a basic tool for analyzing the tail probabilities in large deviation, we introduce the concept of a *general* source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ that is general in the sense that our scope is not only in stationary/ergodic sources but also in non-ergodic and/or nonstationary sources. In order to enable us to deal with this kind of general sources, we define a pair of two kinds of large deviation rate functions, say, a pair of the lower and upper information-spectrum (=IS) rate functions basically without any assumptions on its probabilistic memory structures. This is because sources in consideration may be non-ergodic and/or nonstationary. It should be noted here that, by convention, large deviation rate functions have been supposed to be assigned with one each source. In this way, we try to reveal general and/or simple skeletons of those basic problems. Incidentally, such an IS approach to various kinds of information-theoretical problems has been devised by Han and Verdú [1], Verdú and Han [2], and Han [3], etc.

A first study on large deviations in this "general" direction was made by Chen [5], which has shown a possible generalization of the Gärtner-Ellis theorem using a kind of tilted distribution techniques. In particular, Chen [5] has illustrated that the rate function is *not* necessarily convex.

In the present paper, in the same "general" spirit, the notions of lower/upper IS rate functions are systematically formulated to demonstrate that the pertinent large deviation rate measure can be reasonably described in terms of these IS rate functions.

In Section 2 we prepare necessary notions to establish the pair of basic fundamental lower/upper large deviation rate functions (Theorem 2.1, Theorem 2.2), which are substantially reminiscent of the conventional basic theorems (Dembo and Zeitouni [4]).

In Section 3 it is shown, as applications of the theorem stated in Section 2, that the rate function for *mixed* sources is *not* convex, and also that the lower and upper IS rate functions do *not* coincide with each other for *nonstationary* sources.

In Section 4, a pleasant generalization from the real space $\mathbf{R} \equiv (-\infty, +\infty)$ to general topological spaces is pointed out, although this generalization is not used in the subsequent sections.

In Section 5, as a second crucial step, we proceed to elucidate structural correspondences between the lower/upper IS rate functions and the inferior/superior limits of the normalized cumulant generating functions to demonstrate the formula for computing the cumulant generating function using the lower/upper IS rate functions, of which the inverse function, i.e., the rate function of Cramér-Gärtner-Ellis type is given

in Theorem 5.4. Here, a conclusion (Theorem 5.5) is stated that both are connected equivalently with each other via the Fenchel-Legendre transformation, under a mild assumption, if only and if the former IS rate functions are closed and convex, which is called the reduction theorem.

Finally, in Section 6 we give the proofs of Theorem 5.1 and Lemma 5.1.

2 A General formula for Large Deviations

Let $(Z_1, Z_2, \cdots)$ be any sequence of random variables taking values in $\mathbf{R} \equiv (-\infty, +\infty)$, and call $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ a *general* source. Here, any probabilistic dependency structures are basically not assumed about $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, where Z_n is supposed typically to be the arithmetic mean $\frac{S_n}{n}$ with another underlying "sum" source $\mathbf{S} = \{S_n\}_{n=1}^{\infty}$. We are interested in large deviation behaviors of those $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$.

Let $\pi_1 > \pi_2 > \cdots \to 0$ be an arbitrarily prescribed sequence. With this π_i ($i = 1, 2, \cdots$) define the shrinking open intervals as follows:

$$\Phi_i(R) \equiv (R - \pi_i, R + \pi_i) \quad (i = 1, 2, \cdots), \tag{2.1}$$

and also define as follows:

Definition 2.1

$$\underline{H}_i(R) = \liminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in \Phi_i(R)\}}, \tag{2.2}$$

$$\overline{H}_i(R) = \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in \Phi_i(R)\}}, \tag{2.3}$$

$$\underline{H}(R) = \lim_{i \to \infty} \underline{H}_i(R), \tag{2.4}$$

$$\overline{H}(R) = \lim_{i \to \infty} \overline{H}_i(R). \tag{2.5}$$

$\square$

Clearly, $\underline{H}_i(R)$ and $\overline{H}_i(R)$ are increasing functions in i. It is also obvious that $\underline{H}_i(R) \leq \overline{H}_i(R)$ ($i = 1, 2, \cdots$) and $\underline{H}(R) \leq \overline{H}(R)$. We call these $\underline{H}(R), \overline{H}(R)$ the lower/upper IS rate functions of $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$. It should also be remarked that $\underline{H}(R), \overline{H}(R)$ do not depend on the choice of the sequence $\pi_1 > \pi_2 > \cdots \to 0$.

Throughout in this paper all relevant quantities such as $\underline{H}(R), \overline{H}(R)$ are allowed to take values $\pm\infty$.

In order to establish fundamental formulas for "Large Deviation Principle" (=LDP), we need here the following notion.

Definition 2.2 (*E-tight*) *Let* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *be a general source. If*

$$\lim_{K \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} = -\infty \tag{2.6}$$

holds, then we say that the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *is exponentially tight (abbreviated as E-tight; cf. Dembo and Zeitouni [4]).*

Here, we have the following fundamental theorem for LDP, although it has a rather conventional form. This theorem can also be regarded as forming a pair with Theorem 2.2 below.

Theorem 2.1 *If a general source* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *is E-tight, then for any measurable set* Γ *it holds that*

$$- \inf_{R \in \Gamma^{\circ}} \underline{H}(R) \le \limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \le - \inf_{R \in \overline{\Gamma}} \underline{H}(R), \qquad (2.7)$$

where Γ°, $\overline{\Gamma}$ *are the interior and the closure of* Γ, *respectively.*

Remark 2.1 *As will be seen in the proof below, the lower bound in inequality (2.7):*

$$- \inf_{R \in \Gamma^{\circ}} \underline{H}(R) \le \limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \qquad (2.8)$$

holds without the E-tightness assumption. On the other hand, as for another type of upper bound for (2.7), see Remark 5.7 in Section 5 (also see Corollary 2.2 below).

Remark 2.2 *The E-tightness (Definition 2.2) of the source has the following meaning for the lower IS rate function* $\underline{H}(R)$: *let* $K > 0$ *be an arbitrarily large number, then for any* R *such that* $|R| > K$ *it holds that*

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} \ge -\underline{H}(R)$$

in view of the definition of $\underline{H}(R)$. *Hence,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} \ge - \inf_{|R| > K} \underline{H}(R).$$

Therefore, from the E-tightness (2.6) we have.

$$\lim_{K \to \infty} \inf_{|R| > K} \underline{H}(R) = +\infty,$$

that is,

$$\liminf_{R \to +\infty} \underline{H}(R) = +\infty, \quad \liminf_{R \to -\infty} \underline{H}(R) = +\infty. \qquad (2.9)$$

HoweverC(2.9) does not necessarily imply the E-tightness. For example, if we consider the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *such that* $\Pr\{Z_n = n\} = \Pr\{Z_n = -n\} = \frac{1}{2}$ *(*$\forall n = 1, 2, \cdots$*), this leads to* $\underline{H}(R) = \overline{H}(R) = +\infty$ *(*$\forall R \in \mathbf{R}$*), and so in this case (2.9) holds but (2.6) does not holdD* $\qquad \square$

Proof of Theorem 2.1: It is enough to show

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \le - \inf_{R \in \overline{\Gamma}} \underline{H}(R) \qquad (2.10)$$

3

and (2.8).

a) First we show (2.8). For notational simplicity we use here the notation $\Gamma_\delta(R) \equiv (R - \delta CR + \delta)$ (cf. (2.1)). Then, for any small number $\delta > 0$ it holds that

$$\Gamma_\delta(R) \supset \Phi_i(R) \quad (\forall i \geq i_0(\delta)). \tag{2.11}$$

On the other hand, from the definition (2.4) of $\underline{H}(R)$ it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R)\} \geq -(\underline{H}(R) + \gamma) \quad (\forall i \geq i_0(R)), \tag{2.12}$$

where $\gamma > 0$ is an arbitrary small number. Therefore, there exists a sequence $n_1 < n_2 < \cdots \to \infty$ (dependent on R) such that

$$\frac{1}{n_k} \log \Pr\{Z_{n_k} \in \Phi_i(R)\} \geq -(\underline{H}(R) + 2\gamma) \quad (\forall i \geq i_0(R)). \tag{2.13}$$

Thus, in view of (2.11) and (2.13),

$$\frac{1}{n_k} \log \Pr\{Z_{n_k} \in \Gamma_\delta(R)\} \geq -(\underline{H}(R) + 2\gamma).$$

Hence,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\}$$
$$\geq \limsup_{k\to\infty} \frac{1}{n_k} \log \Pr\{Z_{n_k} \in \Gamma_\delta(R)\}$$
$$\geq -(\underline{H}(R) + 2\gamma).$$

As $\gamma > 0$ is arbitrary, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\} \geq -\underline{H}(R) \quad (\forall \delta > 0). \tag{2.14}$$

Here for any $R \in \Gamma^\circ$ let $\delta > 0$ be small enough to satisfy $\Gamma_\delta(R) \subset \Gamma^\circ$. Then, by means of (2.14),

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \geq \limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma^\circ\}$$
$$\geq \limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\}$$
$$\geq -\underline{H}(R).$$

As R is an arbitrary internal point of Γ°, we conclude that

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \geq -\inf_{R\in\Gamma^\circ} \underline{H}(R).$$

4

b) Next we will show (2.10). It follows from the definition (2.4) of $\underline{H}(R)$ that

$$\frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R)\} \leq -(\underline{H}(R) - 2\gamma) \qquad (\forall n \geq n_0(R, i); \ \forall i \geq i_0(R)),$$

(2.15)

where $\gamma > 0$ is an arbitrarily small number. For any constant $K > 0$ set $\overline{\Gamma}_K \equiv \overline{\Gamma} \cap [-K, K]$ and consider an arbitrary $R \in \overline{\Gamma}_K$, then (2.15) with $i = i_0 \equiv i_0(R)$ reduces to

$$\frac{1}{n} \log \Pr\{Z_n \in \Phi_0(R)\} \leq -(\underline{H}(R) - 2\gamma) \quad (\forall n \geq n_0(R, i_0); \forall R \in \overline{\Gamma}_K), \qquad (2.16)$$

where we have put $\Phi_0(R) = \Phi_{i_0}(R)$. It is evident that $\Phi_0(R)$ is an open set. Now we notice that $\overline{\Gamma}_K$ is a bounded closed set and hence satisfies the compactness (owing to Heine-Borel theorem). As a consequence, we can choose a finite number of points $R_1, R_2, \cdots, R_{m_K}$ in $\overline{\Gamma}_K$ so that

$$\overline{\Gamma}_K \subset \bigcup_{l=1}^{m_K} \Phi_0(R_l).$$

Therefore, for

$$\forall n \geq \max\left(n_0(R_1, i_0), n_0(R_2, i_0), \cdots, n_0(R_{m_K}, i_0)\right).$$

we have

$$\frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\}$$

$$\leq \quad \frac{1}{n} \log \left(\sum_{l=1}^{m_K} \Pr\{Z_n \in \Phi_0(R_l)\}\right)$$

$$\leq \quad \frac{1}{n} \log \left(\max_{1 \leq l \leq m_K} \Pr\{Z_n \in \Phi_0(R_l)\}\right) + \frac{1}{n} \log m_K.$$

Hence, by (2.16) we have

$$\frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\}$$

$$\leq \quad - \min_{1 \leq l \leq m_K} \underline{H}(R_l) + \frac{1}{n} \log m_K + 2\gamma$$

$$\leq \quad - \inf_{R \in \overline{\Gamma}_K} \underline{H}(R) + \frac{1}{n} \log m_K + 2\gamma,$$

from which it follows that

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\} \leq - \inf_{R \in \overline{\Gamma}_K} \underline{H}(R) + 2\gamma.$$

Then, as $\gamma > 0$ is arbitrary,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\} \leq - \inf_{R \in \overline{\Gamma}_K} \underline{H}(R). \tag{2.17}$$

On the other hand, the assumed E-tightness condition of $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ implies that for any large $L > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} \leq -L \quad (\forall K \geq K_0(L)). \tag{2.18}$$

Hence, combining (2.17) and (2.18) results in

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}\}$$

$$\leq \limsup_{n\to\infty} \frac{1}{n} \log \left(\Pr\{Z_n \in \overline{\Gamma}_K\} + \Pr\{|Z_n| > K\} \right)$$

$$\leq \limsup_{n\to\infty} \left(\frac{1}{n} \log \max \left(\Pr\{Z_n \in \overline{\Gamma}_K\}, \Pr\{|Z_n| > K\} \right) + \frac{1}{n} \log 2 \right)$$

$$= \limsup_{n\to\infty} \left(\max \left(\frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\}, \frac{1}{n} \log \Pr\{|Z_n| > K\} \right) \right)$$

$$= \max \left(\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\}, \limsup_{n\to\infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} \right)$$

$$\leq \max \left(- \inf_{R \in \overline{\Gamma}_K} \underline{H}(R), -L \right)$$

$$= - \min \left(\inf_{R \in \overline{\Gamma}_K} \underline{H}(R), L \right)$$

$$\leq - \min \left(\inf_{R \in \overline{\Gamma}} \underline{H}(R), L \right). \tag{2.19}$$

We notice here that $L > 0$ is arbitrarily large, so letting $L \to \infty$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq \limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}\}$$
$$\leq - \inf_{R \in \overline{\Gamma}} \underline{H}(R),$$

thus completing the proof of (2.10). □

Let us now proceed to show the second fundamental theorem on large deviation. To do so, we need the following notion.

Definition 2.3 (σ-convergence) *Let $f_n(R)$ ($n = 1, 2, \cdots$) be a sequence of functions of R on $\mathbf{R}$. If for any bounded closed subset $\mathcal{D}$ of $\mathbf{R}$ and for any $\gamma > 0$ there exists a sequence $n_1 < n_2 < \cdots \to +\infty$ (independent of $R \in \mathcal{D}$) such that*

$$f_{n_k}(R) \geq \limsup_{n\to\infty} f_n(R) - \gamma \quad (\forall k \geq k_0(R, \gamma); \forall R \in \mathcal{D}), \tag{2.20}$$

6

then we say that $\{f_n(R)\}_{n=1}^{\infty}$ is σ-convergent.

Remark 2.3 *It is easy to check that $\{f_n(R)\}_{n=1}^{\infty}$ is σ-convergent if $\{f_n(R)\}_{n=1}^{\infty}$ converges on $\mathbf{R}$. As for the IS meaning of Definition 2.3, refer to Definition 2.4 below.*
□

Definition 2.4 (σ-convergent) *Given a general source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, set*

$$f_n(R) \equiv \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in \Phi_i(R)\}} \quad (\text{with any fixed } i), \tag{2.21}$$

where $\Phi_i(R)$ is defined as in (2.1). If $\{f_n(R)\}_{n=1}^{\infty}$ is σ-convergent in the sense of Definition 2.3, then we say that the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is σ-convergent . □

Remark 2.4 *It is not difficult to check that $\overline{H}(R) = \underline{H}(R)$ ($\forall R \in \mathbf{R}$) is a sufficient condition for the σ-convergence of $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$.*
□

With these definitions, we have the following second fundamental theorem, which can be regarded as providing the pair with Theorem 2.1.

Theorem 2.2 *If a general source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is E-tight and σ-convergent, then for any measurable set Γ it holds that*

$$-\inf_{R \in \Gamma^\circ} \overline{H}(R) \leq \liminf_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq -\inf_{R \in \overline{\Gamma}} \overline{H}(R), \tag{2.22}$$

where Γ°, $\overline{\Gamma}$ are the interior and the closure of Γ, respectivelyD □

Remark 2.5 *As will be seen from the proof below, the lower bound in inequality (2.22):*

$$-\inf_{R \in \Gamma^\circ} \overline{H}(R) \leq \liminf_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \tag{2.23}$$

holds without the assumptions of Theorem 2.2D □

Proof of Theorem 2.2: It suffices to show

$$\liminf_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq -\inf_{R \in \overline{\Gamma}} \overline{H}(R) \tag{2.24}$$

and (2.23).

a) First we show (2.23). Although the proof of this part basically parallels that of part a) of Theorem 2.1 with due modifications, we write it again for the reader's

convenience. Here too, we use the notation that $\Gamma_\delta(R) \equiv (R - \delta CR + \delta)$ $(\delta > 0)$.
Then, we have

$$\Gamma_\delta(R) \supset \Phi_i(R) \quad (\forall i \geq i_0(\delta)). \tag{2.25}$$

Moreover, by the definition (2.5) of $\overline{H}(R)$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R)\} \geq -(\overline{H}(R) + \gamma) \quad (\forall i \geq i_0(R)), \tag{2.26}$$

where $\gamma > 0$ is an arbitrarily small number. Hence,

$$\frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R)\} \geq -(\overline{H}(R) + 2\gamma) \quad (\forall n \geq n_0(R, i); \ \forall i \geq i_0(R)). \tag{2.27}$$

ThereforeCby (2.25) and (2.27),

$$\frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\} \geq -(\overline{H}(R) + 2\gamma) \quad (\forall n \geq n_0(R, \delta)).$$

Hence,

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\} \geq -(\overline{H}(R) + 2\gamma).$$

As $\gamma > 0$ is arbitrary, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\} \geq -\overline{H}(R). \tag{2.28}$$

Now, for any $R \in \Gamma^\circ$ we can choose a small $\delta > 0$ so that $\Gamma_\delta(R) \subset \Gamma^\circ$. Then, by (2.28),

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} &\geq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma^\circ\} \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma_\delta(R)\} \\
&\geq -\overline{H}(R).
\end{aligned}$$

Since $R \in \Gamma^\circ$ is arbitrary, we conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \geq -\inf_{R\in\Gamma^\circ} \overline{H}(R),$$

which implies (2.23).

b) Next we show (2.24). With an arbitraily large $K > 0$ we set $\overline{\Gamma}_K \equiv \overline{\Gamma} \cap [-K, K]$.
By the definition (2.5) of $\overline{H}(R)$,

$$\begin{aligned}
\overline{H}_i(R) &\equiv \limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in \Phi_i(R)\}} \\
&\geq \overline{H}(R) - \gamma \quad (\forall i \geq i_0(R)). \tag{2.29}
\end{aligned}$$

Then, by the assumed σ-convergence of $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, there exists a sequence $n_1 < n_2 < \cdots \to +\infty$ (independent of $R \in \overline{\Gamma}_K$) such that

$$\frac{1}{n_k} \log \frac{1}{\Pr\{Z_{n_k} \in \Phi_i(R)\}} \geq \overline{H}_i(R) - 2\gamma$$

$$(\forall k \geq k_0(R,i); \ \forall i \geq i_0(R); \ \forall R \in \overline{\Gamma}_K). \quad (2.30)$$

As a consequence, combining (2.29) and (2.30) yields

$$\Pr\{Z_{n_k} \in \Phi_i(R)\} \leq \exp[-n_k(\overline{H}(R) - 3\gamma)]$$

$$(\forall k \geq k_0(R,i); \ \forall i \geq i_0(R); \forall R \in \overline{\Gamma}_K). \quad (2.31)$$

Consider the special case of (2.31) with $i = i_0 \equiv i_0(R)$, and put $\Phi_0(R) = \Phi_{i_0}(R)$. Then, (2.31) reduces to

$$\frac{1}{n_k} \log \Pr\{Z_{n_k} \in \Phi_0(R)\}$$

$$\leq -(\overline{H}(R) - 3\gamma) \quad (\forall k \geq k_0(R); \ \forall R \in \overline{\Gamma}_K). \quad (2.32)$$

We note here that $\overline{\Gamma}_K$ is a bounded closed set and hence is compact (owing to Heine-Borel theorem). Therefore, there exists a finite number of $R_1, R_2, \cdots, R_{m_K} \in \overline{\Gamma}_K$ such that

$$\overline{\Gamma}_K \subset \bigcup_{l=1}^{m_K} \Phi_0(R_l).$$

Thus,

$$\frac{1}{n_k} \log \Pr\{Z_{n_k} \in \overline{\Gamma}_K\}$$

$$\leq \frac{1}{n_k} \log \left(\sum_{l=1}^{m_K} \Pr\{Z_{n_k} \in \Phi_0(R_l)\} \right)$$

$$\leq \frac{1}{n_k} \log \left(\max_{1 \leq l \leq m_K} \Pr\{Z_{n_k} \in \Phi_0(R_l)\} \right) + \frac{1}{n_k} \log m_K,$$

from which together with (2.32) it follows that

$$\frac{1}{n_k} \log \Pr\{Z_{n_k} \in \overline{\Gamma}_K\}$$

$$\leq - \min_{1 \leq l \leq m_K} \overline{H}(R_l) + \frac{1}{n_k} \log m_K + 3\gamma$$

$$\leq - \inf_{R \in \overline{\Gamma}_K} \overline{H}(R) + \frac{1}{n_k} \log m_K + 3\gamma.$$

Hence,

$$\liminf_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\} \leq \liminf_{k \to \infty} \frac{1}{n_k} \log \Pr\{Z_{n_k} \in \overline{\Gamma}_K\}$$

$$\leq - \inf_{R \in \overline{\Gamma}_K} \overline{H}(R) + 3\gamma.$$

As $\gamma > 0$ is arbitrary, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\} \leq - \inf_{R\in\overline{\Gamma}_K} \overline{H}(R). \qquad (2.33)$$

On the other hand, by the assumed E-tightness condition of the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, for any large $L > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} \leq -L \quad (\forall K \geq K_0(L)). \qquad (2.34)$$

ThenCfrom (2.33) and (2.34),

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}\}$$

$$\leq \liminf_{n\to\infty} \frac{1}{n} \log \left(\Pr\{Z_n \in \overline{\Gamma}_K\} + \Pr\{|Z_n| > K\} \right)$$

$$\leq \liminf_{n\to\infty} \left(\frac{1}{n} \log \max \left(\Pr\{Z_n \in \overline{\Gamma}_K\}, \Pr\{|Z_n| > K\} \right) + \frac{1}{n} \log 2 \right)$$

$$= \liminf_{n\to\infty} \left(\max \left(\frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\}, \frac{1}{n} \log \Pr\{|Z_n| > K\} \right) \right)$$

$$\leq \max \left(\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}_K\}, \limsup_{n\to\infty} \frac{1}{n} \log \Pr\{|Z_n| > K\} \right)$$

$$\leq \max \left(- \inf_{R\in\overline{\Gamma}_K} \overline{H}(R), -L \right)$$

$$= - \min \left(\inf_{R\in\overline{\Gamma}_K} \overline{H}(R), L \right)$$

$$\leq - \min \left(\inf_{R\in\overline{\Gamma}} \overline{H}(R), L \right). \qquad (2.35)$$

We notice here that $L > 0$ is arbitrarily large, so letting $L \to \infty$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \overline{\Gamma}\}$$

$$\leq - \inf_{R\in\overline{\Gamma}} \overline{H}(R),$$

which implies (2.24)D $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

So far we have demonstrated two fundamental formulas for large deviation (Theorem 2.1 and Theorem 2.2), which are quite basic from the viewpoint of information-spectra. It should be noted here that

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \quad \text{and} \quad \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\}$$

are both given their own lower and upper bounds, respectively, while in usual large deviation theorems $\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\}$ and $\liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\}$ are altogether given a pair of lower and upper bounds.

From this rather conventional standpoint, we can summarize Theorem 2.1, Theorem 2.2 along with Remark 2.1, Remark 2.5 as the following two corollaries (*full* LDP and *weak* LDP):

Corollary 2.1 (Full LDP) *If a source* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *is E-tight, then any measurable set* Γ,

$$
\begin{aligned}
- \inf_{R\in\Gamma^\circ} \overline{H}(R) &\leq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq - \inf_{R\in\overline{\Gamma}} \underline{H}(R). \quad (2.36)
\end{aligned}
$$

Remark 2.6 *The E-tightness condition in Corollary 2.1 is actully needed in order to make the full LDP hold.. For example, let us consider the source* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *as was shown in Remark 2.2. Then,*

$$
\underline{H}(R) = \overline{H}(R) = +\infty \quad (\forall R \in \mathbf{R}),
$$

$$
\lim_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} = 0 \quad (\Gamma = \mathbf{R}),
$$

which obviously contradicts (2.36). □

Corollary 2.2 (Weak LDP: Dembo and Zeitouni [4]) *Without the E-tightness assumption, the following* weak *LDP holds as well:*

1) For any set Γ,

$$
- \inf_{R\in\Gamma^\circ} \overline{H}(R) \leq \liminf_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\}; \quad (2.37)
$$

2) For any compact set Γ,

$$
\limsup_{n\to\infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq - \inf_{R\in\Gamma} \underline{H}(R). \quad (2.38)
$$

3 Examples

In this section, as illustrative examples of the fundamental formulas as shown in the foregoing section, let us consider the large deviation behavior of *mixed* sources and/or *nonstationary sources* (cf. Han [3]).

A. Mixed sources:

First, let

$$\mathbf{X}_1 = (X_{1,1}, X_{1,2}, \cdots), \tag{3.1}$$

$$\mathbf{X}_2 = (X_{2,1}, X_{2,2}, \cdots) \tag{3.2}$$

be two stationary memoryless Gaussian sources with values in $\mathbf{R}$, and let the probability distributions of

$$X_1^n = (X_{1,1}, X_{1,2}, \cdots, X_{1,n}),$$

$$X_2^n = (X_{2,1}, X_{2,2}, \cdots, X_{2,n})$$

be denoted by $P_{X_1^n}(\cdot)$, $P_{X_2^n}(\cdot)$, respectively. Moreover, let the *mixed* source of $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{X} = (X_1, X_2, \cdots) \quad (X^n = (X_1, X_2, \cdots, X_n)) \tag{3.3}$$

be defined as the source with the probability distribution:

$$P_{X^n}(d\mathbf{x}) = \alpha_1 P_{X_1^n}(d\mathbf{x}) + \alpha_2 P_{X_2^n}(d\mathbf{x}) \quad (n = 1, 2, \cdots; \ \mathbf{x} \in \mathbf{R}^n), \tag{3.4}$$

where $\alpha_1 > 0, \alpha_2 > 0$ are constants such that $\alpha_1 + \alpha_2 = 1$. The mixed source $\mathbf{X}$ thus defined is not memoryless but stationary.

Setting

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{3.5}$$

we are interested in the large deviation behavior of $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$. Put

$$Z_{1,n} = \frac{1}{n} \sum_{i=1}^{n} X_{1,i}, \tag{3.6}$$

$$Z_{2,n} = \frac{1}{n} \sum_{i=1}^{n} X_{2,i} \tag{3.7}$$

and with a fixed number R_0 let $\Phi_i(R_0)$ be such as defined in (2.1) in Section 2D First, by Cramér's theorem for the arithmetic mean of a stationary memoryless source (cf. Dembo and Zeitouni [4]), we have

$$
\begin{aligned}
- \inf_{R \in \Phi_i(R_0)} I_1(R) \quad &\leq \quad \liminf_{n \to \infty} \frac{1}{n} \log \Pr\{Z_{1,n} \in \Phi_i(R_0)\} \\
&\leq \quad \limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_{1,n} \in \Phi_i(R_0)\} \leq - \inf_{R \in \overline{\Phi_i(R_0)}} I_1(R),
\end{aligned}
\tag{3.8}
$$

where $I_1(R)$ is the large deviation rate function of the stationary memoryless source $\mathbf{Z}_1 = \{Z_{1,n}\}_{n=1}^{\infty}$ that is defined as

$$I_1(R) = \sup_{\theta}(\theta R - \varphi_1(\theta)) \tag{3.9}$$

12

in terms of the cumulant generating function $\varphi_1(\theta) \equiv \log \mathrm{E}(e^{\theta X_{1,1}})$ of $X_{1,1}$. Since $I_1(R)$ is a closed convex function (and hence a continuous function) of R and $\Phi_i(R_0)$ is a nonempty open interval, we have

$$\inf_{R \in \Phi_i(R_0)} I_1(R) = \inf_{R \in \overline{\Phi_i(R_0)}} I_1(R).$$

Then, by virtue of (3.8), we see that $\frac{1}{n} \log \Pr\{Z_{1,n} \in \Phi_i(R_0)\}$ has the limit (as $n \to \infty$), and (3.8) is written as

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr\{Z_{1,n} \in \Phi_i(R_0)\} = -\inf_{R \in \Phi_i(R_0)} I_1(R). \tag{3.10}$$

In an analogous manner, we have

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr\{Z_{2,n} \in \Phi_i(R_0)\} = -\inf_{R \in \Phi_i(R_0)} I_2(R), \tag{3.11}$$

where $I_2(R)$ is the large deviation rate function of the stationary memoryless source $\mathbf{Z}_2 = \{Z_{2,n}\}_{n=1}^{\infty}$ that is defined as

$$I_2(R) = \sup_{\theta}(\theta R - \varphi_2(\theta)) \tag{3.12}$$

in terms of the cumulant generating function $\varphi_2(\theta) \equiv \log \mathrm{E}(e^{\theta X_{2,1}})$ of $X_{2,1}$. On the other hand, in view of (3.4) we see that

$$\Pr\{Z_n \in \Phi_i(R_0)\} = \alpha_1 \Pr\{Z_{1,n} \in \Phi_i(R_0)\} + \alpha_2 \Pr\{Z_{2,n} \in \Phi_i(R_0)\},$$

from which together with (3.10), (3.11) it follows that $\frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R_0)\}$ has the limit such as

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R_0)\} = -\min\left(\inf_{R \in \Phi_i(R_0)} I_1(R), \inf_{R \in \Phi_i(R_0)} I_2(R)\right). \tag{3.13}$$

As a consequence, from the definition of $\underline{H}(R)$, $\overline{H}(R)$ as in Section 2, we have

$$\begin{aligned} \underline{H}(R_0) &= -\lim_{i \to \infty} \lim_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Phi_i(R_0)\} \\ &= \lim_{i \to \infty} \min\left(\inf_{R \in \Phi_i(R_0)} I_1(R), \inf_{R \in \Phi_i(R_0)} I_2(R)\right) \\ &= \min(I_1(R_0), I_2(R_0)), \end{aligned}$$

where we have again invoked the closed convexity (and hence the continuity) of the functions $I_1(R), I_2(R)$ (cf. Rockafeller [6]). Therefore,

$$\underline{H}(R) = \min(I_1(R), I_2(R)) \quad (\forall R \in \mathbf{R}). \tag{3.14}$$

13

Similarly,

$$\overline{H}(R) = \min(I_1(R), I_2(R)) \quad (\forall R \in \mathbf{R}). \tag{3.15}$$

These (3.14), (3.15) are the lower/upper IS rate functions of the arithmetic mean $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ for the mixed source $\mathbf{X}$. It should be remarked here that $I_($R), $I_2(R)$ are convex functions but $\min(I_1(R), I_2(R))$ is not necessarily convex, which means that the rate functions $\underline{H}(R)$, $\overline{H}(R)$ are not necessarily convex. Also, it is easy to check that the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is E-tight. Thus, Corollary 2.1 together with (3.14), (3.15) yields the large deviation formula in the case of mixed sources:

$$- \inf_{R \in \Gamma^\circ} \min(I_1(R), I_2(R))$$

$$\leq \quad \liminf_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\}$$

$$\leq \quad \limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\}$$

$$\leq \quad - \inf_{R \in \overline{\Gamma}} \min(I_1(R), I_2(R)). \tag{3.16}$$

On the other hand, we recall that the cumulant generating function of $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is given by

$$\varphi_n(\theta) = \frac{1}{n} \log \int P_{Z_n}(dz) e^{n\theta z},$$

which is here written as

$$\varphi_n(\theta) = \frac{1}{n} \log \int \left(\alpha_1 P_{Z_{1,n}}(dz) e^{n\theta z} + \alpha_2 P_{Z_{2,n}}(dz) e^{n\theta z} \right).$$

It is not difficult to verify that the limit $\varphi(\theta) \equiv \lim_{n \to \infty} \varphi_n(\theta)$ exists with

$$\varphi(\theta) = \max \left(\varphi_1(\theta), \varphi_2(\theta) \right). \tag{3.17}$$

We notice that the functions $\varphi(\theta), \varphi_1(\theta), \varphi_2(\theta)$ are always convex. With this $\varphi(\theta)$ let us here define, as usual, the "rate function" by

$$I(R) \equiv \sup_{\theta}(\theta R - \varphi(\theta)).$$

However, this "rate function" $I(R)$ is always convex and hence is different from the lower/upper IS rate functions $\underline{H}(R), \overline{H}(R)$, because $\underline{H}(R), \overline{H}(R)$ are not necessarily convex. Thus, in the case of mixed sources $I(R)$ cannot be a "pertinent" large deviation rate measure.

B. Nonstationary sources:

Here too, we consider two Gaussian sources

$$\mathbf{X}_1 = (X_{1,1}, X_{1,2}, \cdots), \tag{3.18}$$

$$\mathbf{X}_2 = (X_{2,1}, X_{2,2}, \cdots) \tag{3.19}$$

such as defined in (3.1) and (3.2). Define the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ by

$$Z_n = \begin{cases} Z_{1,n} \equiv \frac{1}{n}\sum_{i=1}^{n} X_{1,i} & \text{if } n \text{ is odd}, \\ \\ Z_{2,n} \equiv \frac{1}{n}\sum_{i=1}^{n} X_{2,i} & \text{if } n \text{ is even}. \end{cases} \tag{3.20}$$

Then, it is not difficult to verify that

$$\begin{aligned} \overline{H}(R_0) &= \lim_{i\to\infty} \limsup_{n\to\infty} \frac{1}{n}\log\frac{1}{\Pr\{Z_n \in \Phi_i(R_0)\}} \\ &= \max(\overline{H}_1(R_0), \overline{H}_2(R_0)), \end{aligned} \tag{3.21}$$

where $\overline{H}(R_0)$, $\overline{H}_1(R_0)$, $\overline{H}_2(R_0)$ are the upper IS rate functions for $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, $\mathbf{Z}_1 = \{Z_{1,n}\}_{n=1}^{\infty}$, $\mathbf{Z}_2 = \{Z_{2,n}\}_{n=1}^{\infty}$, respectively.

Similarly, it is not difficult to verify also that

$$\begin{aligned} \underline{H}(R_0) &= \lim_{i\to\infty} \liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\Pr\{Z_n \in \Phi_i(R_0)\}} \\ &= \min(\underline{H}_1(R_0), \underline{H}_2(R_0)), \end{aligned} \tag{3.22}$$

where $\underline{H}(R_0)$, $\underline{H}_1(R_0)$, $\underline{H}_2(R_0)$ are the lower IS rate functions for $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, $\mathbf{Z}_1 = \{Z_{1,n}\}_{n=1}^{\infty}$, $\mathbf{Z}_2 = \{Z_{2,n}\}_{n=1}^{\infty}$, respectively. Thus, we see that $\underline{H}(R_0) \neq \overline{H}_1(R_0)$ in general.

Now, (3.21) and (3.22) are rewritten as

$$\begin{aligned} \underline{H}(R_0) &= \min(I_1(R), I_2(R)), \tag{3.23}\\ \overline{H}(R_0) &= \max(I_1(R), I_2(R)), \tag{3.24} \end{aligned}$$

because

$$\begin{aligned} \underline{H}_1(R_0) &= \overline{H}_1(R_0) = I_1(R_0),\\ \underline{H}_2(R_0) &= \overline{H}_2(R_0) = I_2(R_0). \end{aligned}$$

Since it is easy to check that the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is E-tight, C-tight and σ-convergent, Theorem 2.1 and Theorem 2.2 yield

$$\begin{aligned} &-\inf_{R\in\Gamma^\circ} \min(I_1(R), I_2(R))\\ &\leq \limsup_{n\to\infty} \frac{1}{n}\log\Pr\{Z_n \in \Gamma\}\\ &\leq -\inf_{R\in\overline{\Gamma}} \min(I_1(R), I_2(R)), \end{aligned} \tag{3.25}$$

and

$$\begin{aligned} &-\inf_{R\in\Gamma^\circ} \max I_1(R), I_2(R))\\ &\leq \liminf_{n\to\infty} \frac{1}{n}\log\Pr\{Z_n \in \Gamma\}\\ &\leq -\inf_{R\in\overline{\Gamma}} \max(I_1(R), I_2(R)). \end{aligned} \tag{3.26}$$

15

Thus, in this nonstationary case, the large deviation principle cannot be specified only with a *single* rate function but can be with a *pair* of rate functions as in (3.25) and (3.26). Notice that Corollary 2.1 for this case does not work as well.

4 Note on Generalizations

So far we have established two fundamental theorems (Theorem 2.1, Theorem 2.2) assuming that random variables Z_n take values in the real space $\mathbf{R}$. Actually, however, we can generalize these theorems to the case where Z_n takes values in a general topological space $\mathcal{X}$. To see this, we extend Definitions 2.1, 2.2, 2.4 as follows:

Definition 4.1

$$\overline{H}(R) = \sup_{v(R)} \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in v(R)\}}, \tag{4.1}$$

$$\underline{H}(R) = \sup_{v(R)} \liminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in v(R)\}}, \tag{4.2}$$

where $\sup_{v(R)}$ denotes the supremum over all the neighborhoods $v(R)$ of R. □

Definition 4.2 (*E*-tight) *If for any L there exists a compact set $\mathcal{A}_L \subset \mathcal{X}$ such that*

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \mathcal{A}_L^c\} \le -L \tag{4.3}$$

where c indicates the complement of a set, then we say that the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is exponentially tight (abbreviated as E-tight; cf. Dembo and Zeitouni [4]). □

Definition 4.3 (*σ*-convergent) *Given a general source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$, set*

$$f_n(R) \equiv \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in v(R)\}} \quad \text{(with any fixed neighborhood } v(R)). \tag{4.4}$$

If $\{f_n(R)\}_{n=1}^{\infty}$ is σ-convergent in the sense of Definition 2.3, then we say that the source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is σ-convergent . □

Remark 4.1 *As will be easily seen, Definitions 4.1, Definition 4.2, Definition 4.3 reduce to Definition 2.1, Definition 2.2, Definition 2.4, respectively, in the case of $\mathcal{X} = \mathbf{R}$.* □

Theorem 4.1 *Therem 2.1 and Theorem 2.2 as well as Corollary 2.1 and Corollary 2.2 hold also in the case with any topological space $\mathcal{X}$.*

Proof: It suffices basically to parallel the proofs of Theorem 2.1 and Theorem 2.2 with due modifications. □

5 Cumulant Generating Functions and Information-Spectrum Rate Functions

Thus far having established the fundamental formulas (Theorem 2.1, Theorem 2.2, Corollary 2.1, Corollary 2.2) on general large deviation problems described in terms of the lower/upper IS rate functions $\underline{H}(R)$, $\overline{H}(R)$, we are now interested in the problem of how to compute $\underline{H}(R)$, $\overline{H}(R)$ when the random variables Z_n take values in $\mathbf{R}$.

In many "simple" source cases, for example, as Cramér's theorem and Gärtner-Ellis' theorem tell us, a desirable large deviation "rate function" $I(R)$ is computed as the Fenchel-Legendre transforms (cf. Rockafeller [6]) of the cumulant generatiing function $\varphi(\theta)$ (or something like that):

$$I(R) = \sup_\theta (\theta R - \varphi(\theta)). \tag{5.1}$$

In such cases, the problem of computing the large deviation function $I(R)$ reduces to how to compute the cumulant generating function $\varphi(\theta)$. On the other hand, we notice here that, in the light of Theorem 2.1 and Theorem 2.2, $\underline{H}(R)$ and $\overline{H}(R)$ also should be regarded as "rate functions," which might suggest that $\underline{H}(R)$, $\overline{H}(R)$ could be set to be equal to $I(R)$ as:

$$\underline{H}(R) = \overline{H}(R) = \sup_\theta (\theta R - \varphi(\theta)). \tag{5.2}$$

However, in most general cases that are not necessarily elementary and/or typical, the right-hand side of (5.1) does not give rise to a desirable large deviation rate function any more, as was already seen in the foregoing section, i.e., in general cases,

$$\underline{H}(R), \overline{H}(R) \neq \sup_\theta (\theta R - \varphi(\theta)). \tag{5.3}$$

Our main concern with (5.2), (5.3) then addresses the problem of how to elucidate under what conditions (5.2) holds and/or under what conditions (5.2) does not hold; furthermore, also if not then how not.

In this section we address this problem. Let $\mathbf{Z} = \{Z_n\}_{n=1}^\infty$ be a general source. To this end, let us start with any fixed closed interval $M \equiv [M_1, M_2]$ ($M_1 < M_2$) and define the M-truncated cumulant generating function by

$$\varphi_n^{(M)}(\theta) \equiv \frac{1}{n} \log \int_M P_{Z_n}(dz) e^{n\theta z}, \tag{5.4}$$

and also, define

$$\overline{\varphi}_M(\theta) \equiv \limsup_{n \to \infty} \varphi_n^{(M)}(\theta) \quad (\forall \theta \in \mathbf{R}), \tag{5.5}$$

$$\underline{\varphi}_M(\theta) \equiv \liminf_{n \to \infty} \varphi_n^{(M)}(\theta) \quad (\forall \theta \in \mathbf{R}). \tag{5.6}$$

Moreover, define the M-truncated lower/upper IS rate functions $\underline{H}_M(R)$, $\overline{H}_M(R)$ by

$$\underline{H}_M(R) = \begin{cases} \underline{H}(R) & \text{for} \quad R \in M, \\ +\infty & \text{for} \quad R \notin M, \end{cases} \tag{5.7}$$

$$\overline{H}_M(R) = \begin{cases} \overline{H}(R) & \text{for} \quad R \in M, \\ +\infty & \text{for} \quad R \notin M, \end{cases} \tag{5.8}$$

where $\underline{H}(R)$, $\overline{H}(R)$ are the lower/upper IS rate functions. Then, we have the following fundamental formulas.

Theorem 5.1 *For any general source* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *it holds that*

$$\overline{\varphi}_M(\theta) = \sup_R(\theta R - \underline{H}_M(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.9}$$

$$\underline{\varphi}_M(\theta) \geq \sup_R(\theta R - \overline{H}_M(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.10}$$

The proof of Theorem 5.1 is given in Section 6. $\qquad\qquad\square$

Remark 5.1 *In case* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *is σ-convergent (cf. Definition 2.4) in Theorem 5.1, the following holds:*

$$\underline{\varphi}_M(\theta) = \sup_R(\theta R - \overline{H}_M(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.11}$$

Let us now consider the special case of (5.4)$\sim$ (5.11) with $M = [-K, K]$ $(K > 0)$. This special case is indicated, with an abuse of notation, simply by "K" in place of "M." Then, an immediate consequence of Theorem 5.1 with K in place of M (under the limiting operation $K \to \infty$) is the following theorem, where we have set

$$\overline{\varphi}^{\circ}(\theta) \equiv \lim_{K \to \infty} \overline{\varphi}_K(\theta) \quad (\forall \theta \in \mathbf{R}), \tag{5.12}$$

$$\underline{\varphi}^{\circ}(\theta) \equiv \lim_{K \to \infty} \underline{\varphi}_K(\theta) \quad (\forall \theta \in \mathbf{R}). \tag{5.13}$$

Theorem 5.2 *For any general source* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *it holds that*

$$\overline{\varphi}^{\circ}(\theta) = \sup_R(\theta R - \underline{H}(R)) \quad (\forall \theta \in \mathbf{R}), \tag{5.14}$$

$$\underline{\varphi}^{\circ}(\theta) \geq \sup_R(\theta R - \overline{H}(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.15}$$

Remark 5.2 *In case the σ-convergence property is satisfied in Theorem 5.2, the following holds:*

$$\underline{\varphi}^{\circ}(\theta) = \sup_R(\theta R - \overline{H}(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.16}$$

Proof of Theorem 5.2:

In view of Theorem 5.1 it suffices to take account of the definition of $\overline{\varphi}^{\circ}(\theta)$, $\underline{\varphi}^{\circ}(\theta)$ and to notice that

$$\lim_{K \to \infty} \sup_R(\theta R - \underline{H}_K(R)) = \sup_R(\theta R - \underline{H}(R)) \quad (\forall \theta \in \mathbf{R}), \tag{5.17}$$

$$\lim_{K \to \infty} \sup_R(\theta R - \overline{H}_K R)) = \sup_R(\theta R - \overline{H}(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.18}$$

18

$\square$

So far in Theorem 5.1 we have shown a relation between the M-truncated cumulant generating functions and the M-truncated lower/upper IS rate functions. We now want to see the direct (*not* via M-truncation) relation between the non-truncated cumulant generating functions and the non-truncated lower/upper IS rate functions.

The non-truncated cumulant generating functions are defined by

$$\varphi_n(\theta) \equiv \frac{1}{n} \log \int_{-\infty}^{+\infty} P_{Z_n}(dz) e^{n\theta z}, \tag{5.19}$$

and

$$\overline{\varphi}(\theta) \equiv \limsup_{n \to \infty} \varphi_n(\theta) \quad (\forall \theta \in \mathbf{R}), \tag{5.20}$$

$$\underline{\varphi}(\theta) \equiv \liminf_{n \to \infty} \varphi_n(\theta) \quad (\forall \theta \in \mathbf{R}). \tag{5.21}$$

Paralleling the previous functions $\varphi_n^{(K)}(\theta), \varphi^\circ(\theta)$, we define the following "tail" functions with an arbitrary $K > 0$:

$$\varphi_n^{(\vee K)}(\theta) \equiv \frac{1}{n} \log \int_{|z|>K} P_{Z_n}(dz) e^{n\theta z}, \tag{5.22}$$

$$\overline{\varphi}^\vee(\theta) \equiv \lim_{K \to \infty} \limsup_{n \to \infty} \varphi_n^{(\vee K)}(\theta).$$

Definition 5.1 (*C*-tight) Let $\mathbf{Z} = \{Z_n\}_{n=1}^\infty$ be a general source. If

$$\lim_{K \to \infty} \varphi^{(\vee K)}(\theta) = -\infty \quad (\forall \theta \in \mathbf{R}), \tag{5.23}$$

then we say that $\mathbf{Z} = \{Z_n\}_{n=1}^\infty$ is cumulatively tight (abbreviated as C-tight). $\square$

Then, we now have the following lemma that relates the truncated cumulant generating functions to the non-truncated cumulant generating functions:

Lemma 5.1 *If a general source* $\mathbf{Z} = \{Z_n\}_{n=1}^\infty$ *is C-tight (cf. Definition 2.2), then it holds that*

$$\overline{\varphi}^\circ(\theta) = \overline{\varphi}(\theta), \quad \underline{\varphi}^\circ(\theta) = \underline{\varphi}(\theta) \quad (\forall \ \theta \in \mathbf{R}). \tag{5.24}$$

The proof of this lemma is given in Section 6. $\square$

Now, Lemma 5.1, together with Theorem 5.2 and Remark 5.2, immediately leads to the following Theorem 5.3 and Remark 5.3, respectively.

Theorem 5.3 *If a general source* $\mathbf{Z} = \{Z_n\}_{n=1}^\infty$ *is C-tight, then*

$$\overline{\varphi}(\theta) = \sup_R (\theta R - \underline{H}(R)) \quad (\forall \theta \in \mathbf{R}), \tag{5.25}$$

$$\underline{\varphi}(\theta) \geq \sup_R (\theta R - \overline{H}(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.26}$$

Remark 5.3 *In case* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *is C-tight and σ-convergent in Theorem 5.3, the following holds:*

$$\underline{\varphi}(\theta) = \sup_{R}(\theta R - \overline{H}(R)) \quad (\forall \theta \in \mathbf{R}). \tag{5.27}$$

Remark 5.4 *Theorem 5.3 along with Remark 5.3 is reminiscent of Varadhan's integral lemma [7] where θR can be replaced by an arbitrary bounded continuous function $\phi(R)$. However, in this general case, the existence of a* good *rate function is required.* □

Before proceeding to show the IS formulas for rate functions described in terms of the cumulant generating functions, we need two definitions and one lemma.

Definition 5.2 (Rockafeller [6]) *Given a function f on $\mathbf{R}$, we define the closed convex hull function $\sqcup f$ of f as the pointwise supremum of the collection of all affine functions h on $\mathbf{R}$ such that $h(R) \leq f(R)$ ($\forall R \in \mathbf{R}$). It is evident that $\sqcup f(R) \leq f(R)$ for all $R \in \mathbf{R}$. If $\sqcup f(R) = f(R)$ for all $R \in \mathbf{R}$, we say that f is a closed convex function.*

Definition 5.3 (Fenchel-Legendre transform: Rockafeller [6]) *If*

$$g(\theta) = \sup_{R}(\theta R - f(R)) \quad (\forall \theta \in \mathbf{R}), \tag{5.28}$$

then we say that g is the conjugate of f, and denote the g by f^*.

Lemma 5.2 (Rockafeller [6]) *The conjugate f^* is always a closed convex function. Moreover, it holds that $f^* = (\sqcup f)^*$ and $f^{**} = \sqcup f$.* □

Thus, applying Lemma 5.2 to Theorem 5.3 and Remark 5.3 immediately yields the following inverse formulas:

Theorem 5.4 (Inverse formula) *For any C-tight source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ it holds that*

$$\sqcup \underline{H}(R) = \sup_{\theta}(\theta R - \overline{\varphi}(\theta)) \quad (\forall R \in \mathbf{R}), \tag{5.29}$$

$$\sqcup \overline{H}(R) \geq \sup_{\theta}(\theta R - \underline{\varphi}(\theta)) \quad (\forall R \in \mathbf{R}). \tag{5.30}$$

Remark 5.5 (Inverse formula) *In case the σ-convergence property is also satisfied in Theorem 5.4, the following holds:*

$$\sqcup \overline{H}(R) = \sup_{\theta}(\theta R - \underline{\varphi}(\theta)) \quad (\forall R \in \mathbf{R}). \tag{5.31}$$

Remark 5.6 *Theorem 5.4 as well as Remark 5.5 is reminiscent of Bryc's inverse Varadhan lemma [8] where θR can be replaced by an arbitray bounded continuous function $f(R)$. However, in this latter case, the E-tightness and the existence of the limit of the cumulant generating function are required.* □

A direct consequence of Theorem 5.3, Theorem 5.4, Remark 5.3 and Remark 5.5 is the following corollary which states a key relation between the lower/upper IS rate functions $\underline{H}(R), \overline{H}(R)$ and the cumulant generating functions $\underline{\varphi}(\theta), \overline{\varphi}(\theta)$:

Corollary 5.1 (A condition for the normalized cumulant generating function to have the limit)

1) Let a source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ be C-tight. Then,

$$\sqcup\underline{H}(R) = \sqcup\overline{H}(R) \quad (\forall R \in \mathbf{R}) \tag{5.32}$$

implies

$$\underline{\varphi}(\theta) = \overline{\varphi}(\theta) \quad (\forall \theta \in \mathbf{R}), \tag{5.33}$$

that is, the cumulant generating function $\varphi_n(\theta)$ defined by (5.19) has the limit

$$\varphi(\theta) \equiv \underline{\varphi}(\theta) = \overline{\varphi}(\theta) \quad (\forall \theta \in \mathbf{R}). \tag{5.34}$$

2) If a source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ is not only C-tight but also is σ-convergent, then (5.32) is the necessary and sufficient condition for (5.33). □

Proof: 1) Suppose that (5.32) holds. By Lemma 5.2 combined with Theorem 5.3 we see that

$$\overline{\varphi}(\theta) \quad = \quad \sqcup\overline{\varphi}(\theta) = \sup_R(\theta R - \sqcup\underline{H}(R)) \quad \text{for } \forall \theta \in \mathbf{R}, \tag{5.35}$$

$$\underline{\varphi}(\theta) \quad = \quad \sqcup\underline{\varphi}(\theta) \geq \sup_R(\theta R - \sqcup\overline{H}(R)) \quad \text{for } \forall \theta \in \mathbf{R}, \tag{5.36}$$

which together with (5.32) yields $\overline{\varphi}(\theta) = \underline{\varphi}(\theta)$ ($\forall \theta \in \mathbf{R}$), where we have used the fact $\overline{\varphi}(\theta) \geq \underline{\varphi}(\theta)$.

2) Suppose that (5.34) holds. Then, from (5.29) and (5.31) we have (5.32). □

In some sense, formulas (5.29) and (5.31) may be regarded as providing formulas for computing the lower/upper IS rate functions $\underline{H}(R), \overline{H}(R)$ as the Fenchel-Legendre transforms of the cumulative generating functions $\underline{\varphi}(\theta), \overline{\varphi}(\theta)$. To see this more, let us define the following rate functions $\underline{I}(R), \overline{I}(R)$ of Cramér-Gärtner-Ellis type by

$$\underline{I}(R) \equiv \sup_\theta(\theta R - \overline{\varphi}(\theta)) \quad (\forall R \in \mathbf{R}), \tag{5.37}$$

$$\overline{I}(R) \equiv \sup_\theta(\theta R - \underline{\varphi}(\theta)) \quad (\forall R \in \mathbf{R}). \tag{5.38}$$

Then, from (5.29) and (5.31) we have

$$\sqcup\underline{H}(R) = \underline{I}(R), \quad \sqcup\overline{H}(R) = \overline{I}(R) \quad (\forall R \in \mathbf{R}). \tag{5.39}$$

However, in view of Theorem 2.1, Theorem 2.2 and Corollary 2.1, Corollary 2.2, the formulas that we wanted to obtain were those for computing $\underline{H}(R) = \overline{H}(R)$ but not

for $\sqcup\underline{H}(R), \sqcup\overline{H}(R)$. Formula (5.39) tells us that the "rate function" $\underline{I}(R), \overline{I}(R)$ can capture, as well, relevant structures of large deviation probabilities that are reflected via the nature that $\underline{I}(R), \overline{I}(R)$ are closed convex functions (cf. Lemma 5.2). In other words, $\underline{I}(R), \overline{I}(R)$ overlook all the finer structures that cannot be grasped via the closed convexity of $\underline{I}(R), \overline{I}(R)$ alone. We should be reminded that $\underline{H}(R), \overline{H}(R)$ are not necessarily closed convex functions.

Thus, we do not yet reach the relevant formulas for computing $\underline{H}(R), \overline{H}(R)$ *via* the cumulant generating function, which remains to be further investigated. On the other hand, even without such relevant computation formulas, we could enjoy insightful general view, demonstrated so far in this paper, at basic large deviation problems. This is an advantage of the IS approach.

These observations can formally be summarized as:

Theorem 5.5 (Reduction theorem) *Let a general source* $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ *be C-tight and σ-convergent. Then, it holds that*

$$\sqcup\underline{H}(R) = \underline{I}(R), \quad \sqcup\overline{H}(R) = \overline{I}(R) \quad (\forall R \in \mathbf{R}). \tag{5.40}$$

Moreover, it holds that

$$1) \quad \underline{H}(R) = \underline{I}(R) \quad (\forall R \in \mathbf{R}) \tag{5.41}$$

if and only if $\underline{H}(R)$ is closed and convex; and also that

$$2) \quad \overline{H}(R) = \overline{I}(R) \quad (\forall R \in \mathbf{R}) \tag{5.42}$$

if and only if $\overline{H}(R)$ is closed and convex (cf. Definition 5.2). (Thus, in this case, the computation problems for $\underline{H}(R)$, $\overline{H}(R)$ completely reduces to those for $\underline{I}(R)$, $\overline{I}(R)$.)

Proof: The former part is the same one as in (5.39). The latter part follows if we observe that $\underline{I}(R), \overline{I}(R)$ are always closed convex functions. □

Theorem 5.6 (General note) *Let $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ be a general source and $\overline{\varphi}(\theta)$ be the cumulant generating function defined by (5.19), (5.20). Then,*

$$\underline{H}(R) \geq \sqcup\underline{H}(R) \geq \underline{I}(R) \quad (\forall R \in \mathbf{R}), \tag{5.43}$$

where $\underline{I}(R)$ was defined in (5.37).

Proof: The Fenchel-Legendre transformation of (5.14) gives

$$\sqcup\underline{H}(R) = \sup_{\theta}(\theta R - \overline{\varphi}^{\circ}(\theta)),$$

from which together with $\overline{\varphi}(\theta) \geq \overline{\varphi}^{\circ}(\theta)$ it follows that

$$\underline{H}(R) \geq \sqcup\underline{H}(R) \geq \sup_{\theta}(\theta R - \overline{\varphi}(\theta)) = \underline{I}(R).$$

Thus, in this general case, the computation problem for $\underline{H}(R)$, $\overline{H}(R)$ does *not* reduce to that for $\underline{I}(R)$.

Remark 5.7 *It is evident that $\underline{I}(R)$ is a closed convex function. Application of (5.43) to the right-most term in Theorem 2.1 yields a Cramér-Gärtner-Ellis type of upper bound (though in general much looser):*

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr\{Z_n \in \Gamma\} \leq - \inf_{R \in \overline{\Gamma}} \underline{I}(R) \tag{5.44}$$

for any compact set Γ (cf. Dembo and Zeitouni [4]). □

The following final remark concerns the "locality" of the truncated Fenchel-Legendre transforms:

Remark 5.8 (Locality) *Let a source $\mathbf{Z} = \{Z_n\}_{n=1}^{\infty}$ be C-tight. We define $\underline{I}_M(R)$ as*

$$\underline{I}_M(R) = \left\{ \begin{array}{lll} \underline{I}(R) & \text{for} & R \in M, \\ +\infty & \text{for} & R \notin M, \end{array} \right. \tag{5.45}$$

and suppose that $\underline{H}(R)$ is a closed convex function, i.e., $\sqcup\underline{H}(R) = \underline{H}(R)$. Then, from (5.29) we have $\underline{H}(R) = \underline{I}(R)$, and hence $\underline{H}_M(R) = \underline{I}_M(R)$ ($\forall M, \forall R \in \mathbf{R}$). On the other hand, the he Fenchel-Legendre tansform of (5.9) turns out to be

$$\sqcup(\underline{H}_M)(R) = \sup_{\theta}(\theta R - \overline{\varphi}_M(R)).$$

Thus, in view of $\sqcup(\underline{H}_M)(R) = \underline{H}_M(R) = \underline{I}_M(R)$, we obtain

$$\underline{I}_M(R) = \sup_{\theta}(\theta R - \overline{\varphi}_M(\theta)) \quad (\forall M = [M_1, M_2], \forall R \in \mathbf{R}). \tag{5.46}$$

Then, (5.46) means that, if we want to calculate the value of the rate function $\underline{I}(R)$ at some $R = a_0$, it is not necessary to calculate the values of $\overline{\varphi}(\theta)$ over all $\theta \in \mathbf{R}$ and transform it. Instead, choose a small interval (c, d) contaning a_0 then it suffices to compute the cumulant generating function $\overline{\varphi}_M(\theta)$ only over the domain $M = [c, d]$ no matter how small it is. This demonstrates the "locality" of the rate function $\underline{I}R)$. Similarly for $\overline{I}(R)$, $\underline{\varphi}(\theta)$. □

6 Proofs

In this section we give the proofs of Theorem 5.1 and Lemma 5.1.

6.1 Proof of Theorem 5.1

The proof of Theorem 5.1 consists of several steps. The mainstream is to directly compute the cumulant generating function with $M \equiv [M_1, M_2]$ ($M_1 < M_2$):

$$\varphi_n^{(M)}(\theta) \equiv \frac{1}{n} \log \int_M P_{Z_n}(dz) e^{n\theta z} \tag{6.1}$$

in terms of the quantities $\pi_1, \pi_2, \cdots; \Phi_i(R), \overline{H}_i(R), \underline{H}_i(R), \overline{H}(R), \underline{H}(R)$ defined as in the beginning of Section 2.

Step 1:

For each $i = 1, 2, \cdots$, set $\pi_i = 2^{-i}$ and define an open interval $\Phi_i(R)$ (cf. Section 2) by

$$\Phi_i(R) = (R - \pi_i, R + \pi_i).$$

Then, since $M = [M_1, M_2]$ is compact, for each i there exists a finite number L_i of open intervals

$$\Phi_i(a_i^{(j)}) \quad (j = 1, 2, \cdots, L_i) \tag{6.2}$$

such that

$$a_i^{(j)} \in M \quad (j = 1, 2, \cdots, L_i) \tag{6.3}$$

and

$$M \subset \bigcup_{j=1}^{L_i} \Phi_i(a_i^{(j)}). \tag{6.4}$$

The collection (6.2) of such open intervals is called a finite *cover* of M, simply denoted by Φ_i.

Hereafter, for notational simplicity, we write $I_i^{(j)}$ instead of $\Phi_i(a_i^{(j)})$. Then, the integral of (6.1) is upper bounded as

$$\frac{1}{n} \log \int_M P_{Z_n}(dz) e^{n\theta z} \leq \frac{1}{n} \log \sum_{j=1}^{L_i} \int_{I_i^{(j)}} P_{Z_n}(dz) e^{n\theta z}. \tag{6.5}$$

On the other hand, by definition,

$$\underline{H}_i(a_i^{(j)}) = \liminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in I_i^{(j)}\}}, \tag{6.6}$$

so that, for an arbitrarily small number $\delta > 0$,

$$\begin{aligned} \Pr&\{Z_n \in I_i^{(j)}\} \\ &\leq \exp[-n(\underline{H}_i(a_i^{(j)}) - \delta)] \quad (\forall n \geq \exists n_i^{(j)}; \ \forall j = 1, 2, \cdots, L_i). \end{aligned} \tag{6.7}$$

ThereforeC(6.5) is evaluated as follows.

$$\begin{aligned} \int_M P_{Z_n}(dz) e^{n\theta z} &\leq \sum_{j=1}^{L_i} \int_{I_i^{(j)}} P_{Z_n}(dz) e^{n\theta z} \\ &\leq \sum_{j=1}^{L_i} \exp[-n(\underline{H}_i(a_i^{(j)}) - \delta)] \exp[n\theta a_i^{(j)} + n\pi_i |\theta|] \\ &= \sum_{j=1}^{L_i} \exp[-n(\underline{H}_i(a_i^{(j)}) - \delta)] \exp[n\theta a_i^{(j)} + n2^{-i} |\theta|] \end{aligned}$$

24

$$= \sum_{j=1}^{L_i} \exp[n(\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)}))] \exp[n(\delta + 2^{-i}|\theta|)]$$

$$\leq L_i \exp[n \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)}))] \exp[n(\delta + 2^{-i}|\theta|)]. \tag{6.8}$$

Substituting (6.8) into the right-hand side (6.5) yields

$$\varphi_n^{(M)}(\theta) \leq \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) + \delta + 2^{-i}|\theta| + \frac{1}{n} \log L_i. \tag{6.9}$$

We now define the function $\underline{H}_i^{(M)}(R)$ on $\mathbf{R}$ by

$$\underline{H}_i^{(M)}(R) = \begin{cases} \underline{H}_i(R) & \text{for} & R \in M, \\ +\infty & \text{for} & R \notin M. \end{cases} \tag{6.10}$$

Then, (6.9) can be written as

$$\varphi_n^{(M)}(\theta) \leq \sup_R (\theta R - \underline{H}_i^{(M)}(R)) + \delta + 2^{-i}|\theta| + +\frac{1}{n} \log L_i, \tag{6.11}$$

where $\sup_R$ means the supremum over $\mathbf{R}$. Hence,

$$\limsup_{n \to \infty} \varphi_n^{(M)}(\theta) \leq \sup_R (\theta R - \underline{H}_i^{(M)}(R)) + \delta + 2^{-i}|\theta|. \tag{6.12}$$

It should be noted here that the function $\underline{H}_i^{(M)}(R)$ is monotone increasing in i, that isC

$$\underline{H}_i^{(M)}(R) \leq \underline{H}_{i+1}^{(M)}(R) \quad (\forall R \in \mathbf{R}; \ \forall i = 1, 2, \cdots). \tag{6.13}$$

Therefore, we have the limit function (as was already defined by (5.7)):

$$\underline{H}_M(R) = \lim_{i \to \infty} \underline{H}_i^{(M)}(R) \quad (\forall R \in \mathbf{R}), \tag{6.14}$$

where the value $+\infty$ is also allowedD NowCby (6.12),

$$\begin{aligned} \limsup_{n \to \infty} \varphi_n^{(M)}(\theta) &\leq \lim_{i \to \infty} \left(\sup_R (\theta R - \underline{H}_i^{(M)}(R) + \delta + 2^{-i}|\theta|) \right) \\ &\leq \lim_{i \to \infty} \sup_R (\theta R - \underline{H}_i^{(M)}(R)) + \delta. \end{aligned} \tag{6.15}$$

Since $\delta > 0$ is arbitrarily smallCit follows from (6.15) that

$$\begin{aligned} \overline{\varphi}_M(\theta) \equiv \limsup_{n \to \infty} \varphi_n^{(M)}(\theta) &\leq \lim_{i \to \infty} \sup_R (\theta R - \underline{H}_i^{(M)}(R)) \\ &= \lim_{i \to \infty} \sup_R (\theta R - \sqcup \underline{H}_i^{(M)}(R)), \end{aligned} \tag{6.16}$$

25

where, for a function $f(R)$ on $\mathbf{R}$, the function $\sqcup f$ is defined in Definition 5.2 (also see Lemma 5.2).

Step 2:
Define
$$g_\theta^{(i)}(R) = \theta R - \sqcup \underline{H}_i^{(M)}(R) \tag{6.17}$$
and set
$$g_\theta^{(i)} = \sup_R g_\theta^{(i)}(R), \tag{6.18}$$
$$g_\theta = \lim_{i\to\infty} \sup_R g_\theta^{(i)}(R). \tag{6.19}$$

Suppose here that $g_\theta = -\infty$, then it trivially holds that
$$
\begin{aligned}
\lim_{i\to\infty} \sup_R \left(\theta R - \underline{H}_i^{(M)}(R) \right) &= \lim_{i\to\infty} \sup_R \left(\theta R - \sqcup \underline{H}_i^{(M)}(R) \right) \\
&\leq \sup_R \left(\theta R - \underline{H}_M(R) \right),
\end{aligned}
\tag{6.20}
$$

Next, let us consider the case of $g_\theta > -\infty$ and define the set:
$$\mathcal{G}_\theta(i) = \left\{ R \in \mathbf{R} \, \middle| \, g_\theta^{(i)}(R) \geq g_\theta \right\}.$$

Then, we see that $\sup_R$ on the right-hand side of (6.18) is attained at some point $R = R_0 \in M$; and $g_\theta^{(i)}(R)$ is a closed concave function; so that $\mathcal{G}_\theta(i)$ $(i = 1, 2, \cdots)$ is a sequence of monotone shrinking closed intervals. Therefore, there must exist at least a point $R_1 \in \mathbf{R}$ such that
$$R_1 \in \bigcap_{i=1}^\infty \mathcal{G}_\theta(i)$$
and
$$\theta R_1 - \sqcup \underline{H}_M(R_1) = \lim_{i\to\infty} \left(\theta R_1 - \sqcup \underline{H}_i^{(M)}(R_1) \right) \geq g_\theta, \tag{6.21}$$

where we have used the monotonicity of $\sqcup \underline{H}_i^{(M)}(R)$ in i:
$$\lim_{i\to\infty} \sqcup \underline{H}_i^{(M)}(R) = \sqcup \underline{H}_M(R) \quad (\forall R \in \mathbf{R})$$

Thus, from (6.17)$\sim$(6.19) and (6.21), we have
$$\lim_{i\to\infty} \sup_R \left(\theta R - \sqcup \underline{H}_i^{(M)}(R) \right) \leq \sup_R \left(\theta R - \sqcup \underline{H}_M(R) \right). \tag{6.22}$$

Then, taking account of
$$\sup_R \left(\theta R - \sqcup \underline{H}_i^{(M)}(R) \right) = \sup_R \left(\theta R - \underline{H}_i^{(M)}(R) \right),$$
$$\sup_R \left(\theta R - \sqcup \underline{H}_M(R) \right) = \sup_R \left(\theta R - \underline{H}_M(R) \right),$$

we see that (6.22) is equivalent to

$$\lim_{i\to\infty} \sup_R \left(\theta R - \underline{H}_i^{(M)}(R)\right) \le \sup_R \left(\theta R - \underline{H}_M(R)\right). \tag{6.23}$$

ConsequentlyСby means of (6.16) and (6.23), it is concluded that

$$\overline{\varphi}_M(\theta) \le \sup_R \left(\theta R - \underline{H}_M(R)\right). \tag{6.24}$$

Step 3:

Next, it follows from the definition of $\underline{H}_i(R)$ as in (2.2) that, for any small $\delta > 0$, there exists a sequence of positive integers $n_1^{(j)} < n_2^{(j)} < \cdots \to \infty$, which may depend on $\delta > 0$ and i, such that

$$\Pr\{Z_{n_k^{(j)}} \in I_i^{(j)}\} \ge \exp[-n_k^{(j)}(\underline{H}_i(a_i^{(j)}) + \delta)]$$
$$(\forall k \ge k_0(i,\delta); \ \forall i \ge i_0(\delta); \ \forall j = 1, 2, \cdots, L_i). \tag{6.25}$$

Then,

$$\max_j \left(\frac{1}{n_k^{(j)}} \log \int_M P_{Z_{n_k^{(j)}}}(dz) e^{n_k^{(j)}\theta z} \right)$$

$$\ge \max_j \left(\frac{1}{n_k^{(j)}} \log \int_{I_i^{(j)}} P_{Z_{n_k^{(j)}}}(dz) e^{n_k^{(j)}\theta z} \right)$$

$$\ge \max_j \left(\frac{1}{n_k^{(j)}} \log \left(\exp[-n_k^{(j)}(\underline{H}_i(a_i^{(j)}) + \delta)] \exp[n_k^{(j)}\theta a_i^{(j)} - n_k^{(j)}\pi_i|\theta|] \right) \right)$$

$$= \max_j \left(\frac{1}{n_k^{(j)}} \log \left(\exp[-n_k^{(j)}(\underline{H}_i(a_i^{(j)}) + \delta)] \exp[n_k^{(j)}\theta a_i^{(j)} - n_k^{(j)}2^{-i}|\theta|] \right) \right)$$

$$= \max_j \left((\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) - (\delta + 2^{-i}|\theta|) \right)$$

$$= \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) - (\delta + 2^{-i}|\theta|).$$
$$\tag{6.26}$$

As a consequence,

$$\max_j \varphi_{n_k^{(j)}}^{(M)}(\theta) \ge \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) - \delta - 2^{-i}|\theta|. \tag{6.27}$$

Set

$$\varphi_{n_k}^{(M)}(\theta) = \max_j \varphi_{n_k^{(j)}}^{(M)}(\theta) \quad (n_k = n_k^{(j)}; \ \exists j = 1, 2, \cdots, L_i),$$

then (6.27) yields

$$\varphi_{n_k}^{(M)}(\theta) \ge \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) - \delta - 2^{-i}|\theta|. \tag{6.28}$$

27

We observe here that the left-hand side of (6.28) does not depend on the choice of a finite cover Φ_i of M, so that

$$\varphi_{n_k}^{(M)}(\theta) \geq \sup_{\Phi_i} \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) - \delta - 2^{-i}|\theta|, \qquad (6.29)$$

where $\sup_{\Phi_i}$ means the supremum over all the finite covers Φ_i of M. It is easy to check that

$$\sup_{\Phi_i} \max_j (\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) = \sup_R (\theta R - \underline{H}_i^{(M)}(R)).$$

Thus,

$$\varphi_{n_k}^{(M)}(\theta) \geq \sup_R (\theta R - \underline{H}_i^{(M)}(R)) - \delta - 2^{-i}|\theta|. \qquad (6.30)$$

Noting that $n_k^{(j)} \geq k$ $(\forall j = 1, 2, \cdots, L_i)$ and taking $\limsup_{k \to \infty}$ in both sides of (6.30), we have

$$\begin{aligned}
\overline{\varphi}_M(\theta) &\equiv \limsup_{n \to \infty} \varphi_n^{(M)}(\theta) \\
&\geq \limsup_{k \to \infty} \varphi_{n_k}^{(M)}(\theta) \\
&\geq \sup_R (\theta R - \underline{H}_i^{(M)}(R)) - \delta - 2^{-i}|\theta| \\
&\geq \sup_R (\theta R - \underline{H}_M(R)) - \delta - 2^{-i}|\theta|, \qquad (6.31)
\end{aligned}$$

where we have taken account of the monotonicity of $\underline{H}_i^{(M)}(R)$. Moreover, taking account of $\lim_{i \to \infty}$ in both sides of (6.31) and recalling that $\delta > 0$ is arbitrary, we have

$$\overline{\varphi}_M(\theta) \geq \sup_R (\theta R - \underline{H}_M(R)). \qquad (6.32)$$

Then, it follows from (6.24) and (6.32) that

$$\overline{\varphi}_M(\theta) \equiv \limsup_{n \to \infty} \varphi_n^{(M)}(\theta) = \sup_R (\theta R - \underline{H}_M(R)) \quad (\forall \theta \in \mathbf{R}), \qquad (6.33)$$

which is nothing but (5.9).

Step 4:
    Accordingly to (6.6), we set

$$\overline{H}_i(a_i^{(j)}) = \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{\Pr\{Z_n \in I_i^{(j)}\}} \qquad (6.34)$$

and define the function $\overline{H}_i^{(M)}(R)$ on $\mathbf{R}$ as

$$\overline{H}_i^{(M)}(R) = \begin{cases} \overline{H}_i(R) & \text{for} \quad R \in M, \\ +\infty & \text{for} \quad R \notin M. \end{cases} \qquad (6.35)$$

28

Taking account of the monotonicity in i of $\overline{H}_i^{(M)}(R)$, we obtain the limit function

$$\overline{H}_M(R) = \lim_{i \to \infty} \overline{H}_i^{(M)}(R) \quad (\forall R \in \mathbf{R}). \tag{6.36}$$

Then, in the same manner as in deriving (6.32), we have

$$\underline{\varphi}_M(\theta) \equiv \liminf_{n \to \infty} \varphi_n^{(M)}(\theta) \geq \sup_R (\theta R - \overline{H}_M(R)), \tag{6.37}$$

which is nothing but (5.10).

Step 5:

On the other hand, the opposite inequality:

$$\underline{\varphi}_M(\theta) \equiv \liminf_{n \to \infty} \varphi_n^{(M)}(\theta) \leq \sup_R (\theta R - \overline{H}_M(R)) \tag{6.38}$$

does *not* necessarily hold. In the sequel we will show that the assumed σ-convergence on the interval $\mathcal{D} = M$ (cf. Definitions 2.3, 2.4) is a sufficient condition for (6.38) to hold.

In view of (6.34) we see that the σ-convergence (with any small δ) means the existence of a sequence $n_1 < n_2 < \cdots \to \infty$, which may depend on i and δ but must not on j, such that

$$\frac{1}{n_k} \log \frac{1}{\Pr\{Z_{n_k} \in I_i^{(j)}\}} \geq \overline{H}_i(a_i^{(j)}) - \delta$$
$$(\forall k \geq k_0(i, j, \delta); \ \forall i \geq i_0(\delta); \ \forall j = 1, 2, \cdots, L_i). \tag{6.39}$$

We rewrite this as

$$\Pr\{Z_{n_k} \in I_i^{(j)}\} \leq \exp[-n_k(\overline{H}_i(a_i^{(j)}) - \delta)]$$
$$(\forall k \geq k_0(i, j, \delta); \ \forall i \geq i_0(\delta); \ \forall j = 1, 2, \cdots, L_i). \tag{6.40}$$

Then,

$$\int_{I_i^{(j)}} P_{Z_{n_k}}(dz) e^{n_k \theta z}$$
$$\leq \ \exp[-n_k(\overline{H}_i(a_i^{(j)}) - \delta)] \exp[n_k \theta a_i^{(j)} + n_k 2^{-i}|\theta|]$$
$$= \ \exp[n_k(\theta a_i^{(j)} - \overline{H}_i(a_i^{(j)}))] \exp[n_k(\delta + 2^{-i}|\theta|)].$$
$$\tag{6.41}$$

Consequently,

$$\int_M P_{Z_{n_k}}(dz) e^{n_k \theta z}$$
$$\leq \ \sum_{j=1}^{L_i} \int_{I_i^{(j)}} P_{Z_{n_k}}(dz) e^{n_k \theta z}$$

$$\leq \sum_{j=1}^{L_i} \exp[n_k(\theta a_i^{(j)} - \overline{H}_i(a_i^{(j)}))] \exp[n_k(\delta + 2^{-i}|\theta|)]$$

$$\leq L_i \exp[n_k \max_j(\theta a_i^{(j)} - \overline{H}_i(a_i^{(j)}))] \exp[n_k(\delta + 2^{-i}|\theta|)]$$

$$(6.42)$$

ThereforeC

$$
\begin{aligned}
\varphi_{n_k}^{(M)}(\theta) &\equiv \frac{1}{n_k} \log \int_M P_{Z_{n_k}}(dz) e^{n_k \theta z} \\
&\leq \max_j(\theta a_i^{(j)} - \underline{H}_i(a_i^{(j)})) + \delta + 2^{-i}|\theta| + \frac{\log L_i}{n_k} \\
&\leq \sup_R(\theta R - \overline{H}_i^{(M)}(R)) + \delta + 2^{-i}|\theta| + \frac{\log L_i}{n_k}.
\end{aligned}
$$

$$(6.43)$$

Hence,

$$
\begin{aligned}
\underline{\varphi}_M(\theta) &\equiv \liminf_{n \to \infty} \varphi_n^{(M)}(\theta) \\
&\leq \liminf_{k \to \infty} \varphi_{n_k}^{(M)}(\theta) \\
&= \sup_R(\theta R - \overline{H}_i^{(M)}(R)) + \delta + 2^{-i}|\theta|.
\end{aligned}
$$

$$(6.44)$$

Now, taking $\lim_{i \to \infty}$ and letting $\delta \to 0$ in both of (6.44), we have

$$\underline{\varphi}_M(\theta) \equiv \liminf_{n \to \infty} \varphi_n^{(M)}(\theta) \leq \lim_{i \to \infty} \sup_R(\theta R - \overline{H}_i^{(M)}(R)).$$

Then, in an analogous manner as in the argument for a sequence of monotone shrinking intervals (Step 2), we conclude that

$$\underline{\varphi}_M(\theta) \leq \sup_R(\theta R - \overline{H}_M(R)) \quad (\forall \theta \in \mathbf{R}),$$

$$(6.45)$$

which together with (6.37) yields

$$\underline{\varphi}_M(\theta) \equiv \liminf_{n \to \infty} \varphi_n^{(M)}(\theta) = \sup_R(\theta R - \overline{H}_M(R)) \quad (\forall \theta \in \mathbf{R}).$$

$$(6.46)$$

6.2 Proof of Lemma 5.1

Since

$$\varphi_n(\theta) = \frac{1}{n} \log \int P_{Z_n}(dz) e^{n\theta z}$$

$$
= \quad \frac{1}{n} \log \left(\int_{|z| \leq K} P_{Z_n}(dz) e^{n\theta z} + \int_{|z| > K} P_{Z_n}(dz) e^{n\theta z} \right)
$$

$$
\leq \quad \frac{1}{n} \log \max \left(\int_{|z| \leq K} P_{Z_n}(dz) e^{n\theta z}, \int_{|z| > K} P_{Z_n}(dz) e^{n\theta z} \right) + \frac{1}{n} \log 2
$$

$$
= \quad \max \left(\frac{1}{n} \log \int_{|z| \leq K} P_{Z_n}(dz) e^{n\theta z}, \frac{1}{n} \log \int_{|z| > K} P_{Z_n}(dz) e^{n\theta z} \right) + \frac{1}{n} \log 2
$$

$$
= \quad \max \left(\varphi_n^{(K)}(\theta), \varphi_n^{(\vee K)}(\theta) \right) + \frac{1}{n} \log 2,
$$

we have

$$
\begin{aligned}
\overline{\varphi}(\theta)@ \quad &\equiv \quad \limsup_{n \to \infty} \varphi_n(\theta) \\
&= \quad \limsup_{n \to \infty} \frac{1}{n} \log \int P_{Z_n}(dz) e^{n\theta z} \\
&\leq \quad \max \left(\limsup_{n \to \infty} \varphi_n^{(K)}(\theta), \limsup_{n \to \infty} \varphi_n^{(\vee K)}(\theta) \right).
\end{aligned} \tag{6.47}
$$

Letting $K \to \infty$ in both sides of (6.47), it follows from the assumed C-tightness that

$$
\begin{aligned}
\overline{\varphi}(\theta) \quad &\leq \quad \max \left(\lim_{K \to \infty} \limsup_{n \to \infty} \varphi_n^{(K)}(\theta), \lim_{K \to \infty} \limsup_{n \to \infty} \varphi_n^{(\vee K)}(\theta) \right) \\
&= \quad \max \left(\overline{\varphi}^{\circ}(\theta), -\infty \right) \\
&= \quad \overline{\varphi}^{\circ}(\theta).
\end{aligned} \tag{6.48}
$$

On the other hand, $\varphi(\theta) \geq \varphi^{\circ}(\theta)$ always holds, so that we conclude

$$
\overline{\varphi}(\theta) = \overline{\varphi}^{\circ}(\theta) \quad (\forall \theta \in \mathbf{R}). \tag{6.49}
$$

Furthermore, in the same way as above, we have

$$
\begin{aligned}
\underline{\varphi}(\theta) \quad &\equiv \quad \liminf_{n \to \infty} \varphi_n(\theta) \\
&\leq \quad \max \left(\liminf_{n \to \infty} \varphi_n^{(K)}(\theta), \limsup_{n \to \infty} \varphi_n^{(\vee K)}(\theta) \right).
\end{aligned} \tag{6.50}
$$

Again, letting $K \to \infty$ in (6.50) yields

$$
\begin{aligned}
\underline{\varphi}(\theta) \quad &\leq \quad \max \left(\lim_{K \to \infty} \liminf_{n \to \infty} \varphi_n^{(K)}(\theta), \lim_{K \to \infty} \limsup_{n \to \infty} \varphi_n^{(\vee K)}(\theta) \right) \\
&= \quad \max \left(\underline{\varphi}^{\circ}(\theta), -\infty \right) \\
&= \quad \underline{\varphi}^{\circ}(\theta),
\end{aligned} \tag{6.51}
$$

where we have used again the assumed C-tightness. Since $\underline{\varphi}(\theta) \geq \underline{\varphi}^{\circ}(\theta)$ always holds, we conclude that

$$
\underline{\varphi}(\theta) = \underline{\varphi}^{\circ}(\theta) \quad (\forall \theta \in \mathbf{R}). \tag{6.52}
$$

31

References

[1] T.S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol.IT-39, no.3, pp. 752-772, 1993

[2] S. Verdú and T.S. Han, "A general formula for channel capacity," IEEE Transactions on Information Theory, vol.IT-40, no.4, pp.1147-1157, 1994

[3] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer-Verlag, Berlin, New York, 2003; The Japanese edition was published by Baifukan-publisher, Tokyo, 1998

[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers, Boston, 1993

[5] Po-Ning Chen, "Generalization of Gärtner-Ellis Therem." *IEEE Transactions on Information Theory*, vol.IT-46, no.7, pp. 2752-2760, 2000

[6] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, New Jersey, 1970

[7] S.R.S. Varadhan, *Large deviations and Applications*, SIAM, Philadelphia, 1980

[8] W. V. Bryc, "Large deviations by the asymptotic value method," In M. Pinsker, editor, [*Defusion Processes and Related Problems in Analysis," pp. 447-472, Birkhäuser, Basel, 20:1004-1030, 1992*