Csiszár's Cutoff Rates for Arbitrary Discrete Sources*

Po-Ning Chen[†] and *Fady Alajaji*[‡]

[†]Department of Communication Engineering
National Chiao-Tung University
HsinChu, Taiwan, R.O.C.

[‡]Department of Mathematics and Statistics
Queen's University
Kingston, Ontario K7L 3N6, Canada

Abstract

Csiszár's forward β-cutoff rate (given a fixed $\beta > 0$) for a discrete source is defined as the smallest number R_0 such that for every $R > R_0$, there exists a sequence of fixed-length codes of rate R with probability of error asymptotically vanishing as $e^{-n\beta(R-R_0)}$. For a discrete memoryless source, the forward β-cutoff rate is shown by Csiszár [6] to be equal to the source Rényi entropy. An analogous concept of reverse β-cutoff rate regarding the probability of correct decoding is also characterized by Csiszár in terms of the Rényi entropy.

In this work, Csiszár's results are generalized by investigating the β-cutoff rates for the class of arbitrary discrete sources with memory. It is demonstrated that the limsup and liminf Rényi entropy rates provide the formulas for the forward and reverse β-cutoff rates, respectively. Consequently, new fixed-length source coding operational characterizations for the Rényi entropy rates are established.

Index terms – Source reliability function, Rényi's entropy rates, cutoff rates, arbitrary sources with memory, fixed-length source coding, probability of error.

* *IEEE Transactions on Information Theory*: to appear. This work was supported in part by an ARC grant from Queen's University, the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant OGP0183645, and the National Science Council of Taiwan, R.O.C., under Grant NSC 88-2219-E-009-004.

1 Introduction

In [6], Csiszár establishes the concept of generalized fixed-length coding cutoff rates (forward and reverse) for discrete memoryless sources (DMS's). More specifically, given $\beta > 0$, he defines the forward β-cutoff rate for a source $\{X_i\}_{i=1}^{\infty}$ as the number R_0 that provides the best possible lower bound in the form $\beta(R - R_0)$ to the source reliability function. This definition implies that the source error probability is guaranteed to exponentially decay with a linear exponent of specified slope β for $R > R_0$. He also provides a similar definition for the reverse β-cutoff rate (where $\beta > 0$) with respect to the source unreliability function (the exponent of the vanishing probability of correct decoding). He then demonstrates that the forward and reverse β-cutoff rates are respectively given by $H_{1/(1+\beta)}(X_1)$ and $H_{1/(1-\beta)}(X_1)$, where $H_\alpha(X_1)$ denotes the Rényi entropy of order α [15]. This result provides a new operational significance for Rényi's entropy.

Previous operational characterizations of Rényi's entropy were established by Arikan [1] for the theory of guessing, by Jelinek [12] and others (e.g., [14]) for the buffer overflow problem in lossless source coding, and by Campbell [5] for the lossless variable-length coding problem with an exponential cost constraint for a DMS. Recently, Erez and Zamir [9] demonstrated that for discrete memoryless modulo additive noise channels with side information at the transmitter, Gallager's random coding error exponent as well as the sphere-packing error exponent can be written in terms of the Rényi entropy. Finally, Campbell's work was generalized in [16] for the class of Markov sources of arbitrary order.

In this work, we extend Csiszár's results [6] by investigating the β-cutoff rate for arbitrary (not necessarily, stationary, ergodic, etc.) discrete-time finite-alphabet sources $\boldsymbol{X} \triangleq \{X^n = (X_1^{(n)}, \ldots, X_n^{(n)})\}_{n=1}^{\infty}$. We demonstrate that the limsup and liminf Rényi entropy rates provide the expressions for the forward and reverse β-cutoff rates, respectively. These results also provide simple, and in certain cases, computable lower bounds to the source reliability and unreliability functions.

The rest of this paper is organized as follows. In Section 2, relevant previous results by Han on the reliability and unreliability functions of arbitrary sources are briefly reviewed. The general expression for the forward β-cutoff rate and the reverse β-cutoff rates are proved in Sections 3 and 4, respectively. Finally, concluding remarks are stated in Section 5.

2 Preliminaries: Source Reliability and Unreliability Functions

In this section, we briefly review the previous results by Han [10, 11] on the general expressions for the reliability and unreliability functions of arbitrary discrete-time finite-alphabet sources (for previous work on the source coding error exponent, see [7], [13], [2], [8], [3] and [11]).

Consider a discrete-time source $\boldsymbol{X}$ defined by a sequence of finite dimensional distributions [10]: $\boldsymbol{X} \triangleq \left\{X^n = \left(X_1^{(n)}, \ldots, X_n^{(n)}\right)\right\}_{n=1}^{\infty}$. We assume that the source alphabet $\mathcal{X}$ is finite.

Definition 1 (Fixed-length source code) An (n, M) fixed-length source code for X^n is a collection of M n-tuples $\mathscr{C}_n = \{c_1^n, \ldots, c_M^n\}$. The error probability of the code is

$$P_e(\mathscr{C}_n) \triangleq P_{X^n}\left[X^n \notin \mathscr{C}_n\right].$$

Definition 2 (Source reliability function [10, Definitions 1.12 and 1.13]) Fix $e > 0$. $R > 0$ is *e-achievable* for a source $\boldsymbol{X}$, if there exists a sequence of (n, M_n) fixed-length source code $\mathscr{C}_n$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R \quad \text{and} \quad \liminf_{n\to\infty} -\frac{1}{n} \log P_e(\mathscr{C}_n) \ge e.$$

The infimum of all e-achievable rates for source $\boldsymbol{X}$ is denoted by $R(e|\boldsymbol{X})$. The reliability function for source $\boldsymbol{X}$, $E(R|\boldsymbol{X})$ is the dual of $R(e|\boldsymbol{X})$. More specifically,

$$E(R|\boldsymbol{X}) \triangleq \sup\{e > 0 : R \text{ is } e\text{-achievable for } \boldsymbol{X}\},$$

and $E(R|\boldsymbol{X}) = 0$ if the above set is empty.

Note that since the source alphabet is finite, $R(e|\boldsymbol{X}) \leq \log|\mathcal{X}| < \infty$ for every $e > 0$; this implies that $E(R|\boldsymbol{X}) = \infty$ for $R > \log|\mathcal{X}|$. Furthermore, $E(R|\boldsymbol{X})$ is non-decreasing in R but non-convex in general.

Theorem 1 ([10, Theorem 1.15]) Fix $e > 0$. For any source $\boldsymbol{X}$,

$$R(e|\boldsymbol{X}) = \sup\{R - \sigma(R) \ : \ R \in (0,\infty) \text{ and } \sigma(R) < e\},$$

where

$$\sigma(R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq R\right\}.$$

Definition 3 (Source unreliability function [10, Definitions 1.14 and 1.15]) Fix $e > 0$. $R > 0$ is *reverse e-achievable* for source $\boldsymbol{X}$, if there exists a sequence of (n, M_n) fixed-length source code $\mathcal{C}_n$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n \leq R \quad \text{and} \quad \liminf_{n\to\infty} -\frac{1}{n}\log(1 - P_e(\mathcal{C}_n)) \leq e.$$

The infimum of all reverse e-achievable rates for source $\boldsymbol{X}$ is denoted by $\underline{R}^*(e|\boldsymbol{X})$. Therefore for any $0 < R < \underline{R}^*(e|\boldsymbol{X})$, every code sequence $\mathcal{C}_n$ with $\limsup_{n\to\infty}(1/n)\log M_n \leq R$ satisfies $P_e(\mathcal{C}_n) > 1 - \exp\{-ne\}$ for all sufficiently large n. This is a *pessimistic* view point, since we require that all code sequences are "bad" for all sufficiently large n[1]. The unreliability function for source $\boldsymbol{X}$, $\underline{E}^*(R|\boldsymbol{X})$ is the dual of $\underline{R}^*(e|\boldsymbol{X})$. More specifically,

[1]Note that this is consistent with our terminology for $\underline{R}^*(e|\boldsymbol{X})$ as an unreliability function. However, one could also regard our definition from the *optimistic* point of view [10] if the quantity of interest is the probability of correct decoding as opposed to the probability of error. In this case, one would require "good" codes for infinitely many n.

$$\underline{E}^*(R|\boldsymbol{X}) \triangleq \inf\{e > 0 : R \text{ is reverse } e\text{-achievable for } \boldsymbol{X}\}.$$

Under slight modification, the following result follows from [10, Theorem 1.16].

Theorem 2 Fix $e > 0$. For any source $\boldsymbol{X}$,

$$\underline{R}^*(e|\boldsymbol{X}) = \inf\{h > 0 : \inf_{R>0}(\underline{\lambda}(R) + [R - \underline{\lambda}(R) - h]^+) \leq e\},$$

where

$$\underline{\lambda}(R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\},$$

and $[x]^+ \triangleq \max(x, 0)$.

3 Forward β-Cutoff Rate

Definition 4 (Forward β-cutoff rate) Fix $\beta > 0$. $R_0 \geq 0$ is a *forward β-achievable rate* for a source $\boldsymbol{X}$ if

$$E(R|\boldsymbol{X}) \geq \beta(R - R_0)$$

for every $R > 0$, or equivalently,

$$R(e|\boldsymbol{X}) \leq \frac{1}{\beta}e + R_0,$$

for every $e > 0$. The *forward β-cutoff rate* for $\boldsymbol{X}$ is defined as the infimum of all forward β-achievable rates, and is denoted by $R_0^{(f)}(\beta|\boldsymbol{X})$. A graphical illustration of $R_0^{(f)}(\beta|\boldsymbol{X})$ is provided in Figure 1.

It is important to remark that the above definition of the forward β-cutoff rate is equivalent to the first part of Csiszár's definition (cf. [6, Definition 1]).

Before providing the general expression of the forward β-cutoff rate, we prove the following lemma, which is a consequence of Theorem 1.

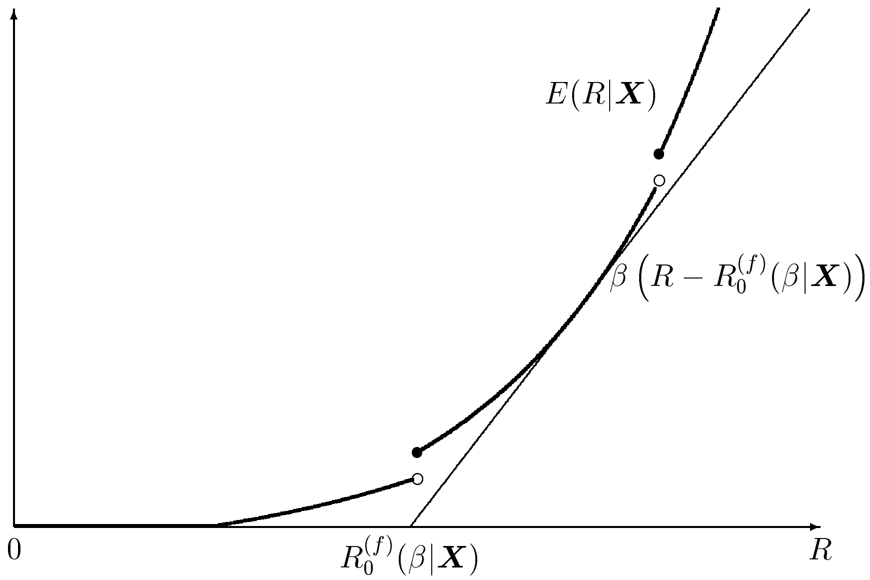

Figure 1: Forward β-cutoff rate for an arbitrary source $\boldsymbol{X}$.

Lemma 1 The following two conditions are equivalent.

$$(\forall\ R > 0) \quad \sigma(R) \geq \frac{\beta}{1+\beta}(R - R_0) \tag{3.1}$$

and

$$(\forall\ e > 0) \quad R(e|\boldsymbol{X}) \leq \frac{1}{\beta}e + R_0. \tag{3.2}$$

Proof:

1. *Forward part*: (3.1)$\Rightarrow$(3.2)

   For any $e > 0$, we obtain by Theorem 1 that

$$(\forall\ \delta > 0)(\exists\ R_\delta \text{ with } \sigma(R_\delta) < e) \quad R(e|\boldsymbol{X}) - \delta \leq R_\delta - \sigma(R_\delta).$$

$$\begin{aligned}\Rightarrow R(e|\boldsymbol{X}) &\leq R_\delta - \sigma(R_\delta) + \delta \\ &\leq R_\delta - \frac{\beta}{1+\beta}(R_\delta - R_0) + \delta \qquad (3.3)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{1+\beta}R_\delta + \frac{\beta}{1+\beta}R_0 + \delta \\
&< \frac{1}{1+\beta}\left(\frac{1+\beta}{\beta}e + R_0\right) + \frac{\beta}{1+\beta}R_0 + \delta \qquad (3.4)\\
&= \frac{1}{\beta}e + R_0 + \delta,
\end{aligned}
$$

where (3.3) follows by (3.1), and (3.4) holds because

$$e > \sigma(R_\delta) \geq \frac{\beta}{1+\beta}(R_\delta - R_0).$$

The proof is then completed by noting that δ can be made arbitrarily small (independently of e).

2. *Converse part*: (3.2) $\Rightarrow$ (3.1)

(3.1) holds trivially for those R satisfying $\sigma(R) = \infty$. For any $R > 0$ with $\sigma(R) < \infty$, let $e_\delta \stackrel{\triangle}{=} \sigma(R) + \delta$ for some $\delta > 0$. Then (by Theorem 1)

$$R(e_\delta|\boldsymbol{X}) \geq R - \sigma(R).$$

$$
\begin{aligned}
\Rightarrow \sigma(R) &\geq R - R(e_\delta|\boldsymbol{X}) \\
&\geq R - \frac{1}{\beta}e_\delta - R_0 \qquad (3.5)\\
&= R - \frac{1}{\beta}\sigma(R) - \frac{\delta}{\beta} - R_0,
\end{aligned}
$$

where (3.5) follows by (3.2). Thus,

$$\sigma(R) \geq \frac{\beta}{1+\beta}(R - R_0) - \frac{\delta}{1+\beta}.$$

The proof is then completed by noting that δ can be made arbitrarily small.

$\square$

Remark: The above lemma actually identifies the forward β-cutoff rate $R_0^{(f)}(\beta|\boldsymbol{X})$ as the R-axis intercept of the *support line* with slope $\beta/(1+\beta)$ to the large deviation spectrum curve $\sigma(R)$. We next establish an expression for $R_0^{(f)}(\beta|\boldsymbol{X})$ by showing that the limsup Rényi entropy rate of order $1/(1+\beta)$ is indeed the above intercept.

Theorem 3 (Forward β-cutoff rate formula) Fix $\beta > 0$. For an arbitrary source $\boldsymbol{X}$,

$$R_0^{(f)}(\beta|\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} H_{1/(1+\beta)}(X^n),$$

where

$$H_\alpha(X^n) \triangleq \frac{1}{1-\alpha} \log \sum_{x^n \in \mathcal{X}^n} P_{X^n}^\alpha(x^n)$$

is the (n-dimensional) Rényi entropy of order α.

Proof:

1. *Forward part*: $R_0^{(f)}(\beta|\boldsymbol{X}) \le \limsup_{n\to\infty}(1/n)H_{1/(1+\beta)}(X^n)$

By the equivalence of conditions (3.1) and (3.2), it suffices to show that

$$(\forall\ R \ge 0)\ \sigma(R) \ge \frac{\beta}{1+\beta}\left(R - \limsup_{n\to\infty} \frac{1}{n} H_{1/(1+\beta)}(X^n)\right).$$

$$\begin{aligned}
Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \ge R\right] &= Pr\left[e^{-t\log P_{X^n}(X^n)} \ge e^{ntR}\right],\ \text{for } t > 0 \\
&\le e^{-ntR} \sum_{x^n \in \mathcal{X}^n} P_{X^n}^{1-t}(x^n),\ \text{for } t > 0 \text{ (by Markov's inequality)} \\
&= \exp\left\{-nt\left(R - \frac{1}{n}H_{1-t}(X^n)\right)\right\}, \quad \text{for } 0 < t < 1.
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \sigma(R) &\ge t\left(R - \limsup_{n\to\infty}\frac{1}{n}H_{1-t}(X^n)\right), \quad \text{for } 0 < t < 1 \\
&= \frac{\beta}{1+\beta}\left(R - \limsup_{n\to\infty}\frac{1}{n}H_{1/(1+\beta)}(X^n)\right), \quad \text{for } \beta \triangleq \frac{t}{1-t} > 0.
\end{aligned}$$

2. *Converse part*: $R_0^{(f)}(\beta|\boldsymbol{X}) \geq \limsup_{n\to\infty}(1/n)H_{1/(1+\beta)}(X^n)$

The converse part holds trivially if $\limsup_{n\to\infty}(1/n)H_{1/(1+\beta)}(X^n) = 0$. Without loss of generality, we assume that $\limsup_{n\to\infty}(1/n)H_{1/(1+\beta)}(X^n) > 0$.

By the equivalence of conditions (3.1) and (3.2), it suffices to show that for any $\delta > 0$ arbitrarily small, there exists $\underline{R} = \underline{R}(\delta) > 0$ such that

$$\sigma(\underline{R}) \leq \frac{\beta}{1+\beta}\left(\underline{R} - \limsup_{n\to\infty}\frac{1}{n}H_{1/(1+\beta)}(X^n) + 3\delta\right).$$

Consider the tilted distribution (e.g., [4, 3]) with parameter t of the random variable $-\log P_{X^n}(X^n)$, defined as

$$\begin{aligned} P_{X^n}^{(t)}(x^n) &\triangleq \frac{e^{t(-\log P_{X^n}(x^n))}P_{X^n}(x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} e^{t(-\log P_{X^n}(\hat{x}^n))}P_{X^n}(\hat{x}^n)} \\ &= \frac{P_{X^n}^{1-t}(x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{X^n}^{1-t}(\hat{x}^n)} = \frac{P_{X^n}^{1-t}(x^n)}{\exp\{tH_{1-t}(X^n)\}} \\ &= \exp\{-t[\log P_{X^n}(x^n) + H_{1-t}(X^n)]\}P_{X^n}(x^n), \end{aligned} \tag{3.6}$$

where $t = \beta/(1+\beta)$. By definition of limsup, there exists an increasing sequence of positive integers $\mathcal{J}\triangleq\{n_j\}_{j\geq 1}$ satisfying

$$\lim_{n\to\infty,\ n\in\mathcal{J}}\frac{1}{n}H_{1/(1+\beta)}(X^n)\triangleq\lim_{j\to\infty}\frac{1}{n_j}H_{1/(1+\beta)}(X^{n_j}) = \limsup_{n\to\infty}\frac{1}{n}H_{1/(1+\beta)}(X^n).$$

Also define

$$\tau\triangleq\inf\left\{R\in[0,\infty)\ :\ \sigma_{\mathcal{J}}^{(t)}(R) > 0\right\}, \tag{3.7}$$

where[2]

$$\sigma_{\mathcal{J}}^{(t)}(R)\triangleq\liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) \geq R\right].$$

[2]Recall that for any sequence $\{a_n\}$,

$$\liminf_{n\to\infty,\ n\in\mathcal{J}} a_n\triangleq\liminf_{j\to\infty} a_{n_j} = \lim_{j\to\infty}\inf_{k\geq j} a_{n_k}.$$

In Appendix A (cf. Lemmas 4 and 5), we show that for $0 < t < 1$,

$$\tau \leq \frac{1}{1-t}\log|\mathcal{X}|,$$

and

$$\lim_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n}H_{1/(1+\beta)}(X^n) > 0 \ \Rightarrow \ \tau \geq \lim_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n}H_{1/(1+\beta)}(X^n) > 0.$$

Hence, we can choose a fixed $\delta \in (0, \tau]$ such that

$$\sigma_{\mathcal{J}}^{(t)}(\tau+\delta) = \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) \geq \tau+\delta\right] > 0.$$

The above inequality implies the existence of $\gamma > 0$ such that

$$-\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) \geq \tau+\delta\right] > \gamma$$

for all $n \in \mathcal{J}$ sufficiently large. Thus, for those n satisfying the above inequality,

$$P_{X^n}^{(t)}\left[x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < \tau+\delta\right] > 1 - e^{-n\gamma}.$$

Let

$$I_k \triangleq [b_{k-1}, b_k) \quad \text{for } 1 \leq k \leq K \triangleq \lceil (\tau+\delta)/(2\delta) \rceil,$$

where

$$(\forall\ 1 \leq k < K) \quad b_k = 2k\delta, \quad \text{and} \quad b_K = \tau+\delta.$$

Note that $b_k - b_{k-1} = 2\delta$ for every $1 \leq k < K$ and $0 < b_K - b_{K-1} \leq 2\delta$. Since $-\log P_{X^n}(X^n) \geq 0$ with probability 1, then

$$\begin{aligned} P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) < \tau+\delta\right] &= \sum_{k=1}^{K} P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \in I_k\right] \\ &> 1 - e^{-n\gamma} \quad \text{for all } n \in \mathcal{J} \text{ sufficiently large.} \end{aligned}$$

Hence, there exists $k(n) \in [1, K]$ for all sufficiently large $n \in \mathcal{J}$ such that

$$P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \in I_{k(n)}\right] \geq \frac{1-e^{-n\gamma}}{K}. \tag{3.8}$$

Let $\underline{R} \triangleq \liminf_{n\to\infty,\ n\in\mathcal{J}} b_{k(n)-1} - \delta$ (here, we assume that by choosing $\delta > 0$ small enough, we can make $\underline{R} > 0$. We will substantiate this assumption later). Then by noting that $\underline{R} < b_{k(n)-1}$ for all sufficiently large $n \in \mathcal{J}$, we obtain that

$$P_{X^n}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \geq \underline{R}\right] \geq P_{X^n}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \in I_{k(n)}\right]$$

for all $n \in \mathcal{J}$ sufficiently large. But,

$$\begin{aligned}
& P_{X^n}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \in I_{k(n)}\right] \\
= & \sum_{[x^n : -(1/n)\log P_{X^n}(x^n) \in I_{k(n)}]} P_{X^n}(x^n) \\
= & \sum_{[x^n : -(1/n)\log P_{X^n}(x^n) \in I_{k(n)}]} \exp\{t[\log P_{X^n}(x^n) + H_{1-t}(X^n)]\} P_{X^n}^{(t)}(x^n) \quad \text{(by (3.6))} \\
\geq & \exp\{-ntb_{k(n)} + tH_{1-t}(X^n)\} \sum_{[x^n : -(1/n)\log P_{X^n}(x^n) \in I_{k(n)}]} P_{X^n}^{(t)}(x^n) \\
= & \exp\left\{-nt\left(b_{k(n)} - \frac{1}{n}H_{1-t}(X^n)\right)\right\} P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \in I_{k(n)}\right] \\
\geq & \frac{1-e^{-n\gamma}}{K}\exp\left\{-nt\left(b_{k(n)} - \frac{1}{n}H_{1-t}(X^n)\right)\right\} \quad \forall\, n \in \mathcal{J} \text{ sufficiently large,}
\end{aligned} \tag{3.9}$$

where the last inequality follows from (3.8). Consequently,

$$\begin{aligned}
\sigma(\underline{R}) &= \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \geq \underline{R}\right] \\
&\leq \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \geq \underline{R}\right] \\
&\leq t\left(\liminf_{n\to\infty,\ n\in\mathcal{J}} b_{k(n)} - \lim_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n}H_{1-t}(X^n)\right) \\
&\leq t\left(\liminf_{n\to\infty,\ n\in\mathcal{J}} b_{k(n)-1} + 2\delta - \limsup_{n\to\infty} \frac{1}{n}H_{1-t}(X^n)\right) \\
&= t\left(\underline{R} - \limsup_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) + 3\delta\right).
\end{aligned} \tag{3.10}$$

Now it remains to validate the claim on $\underline{R}$ that it can be made positive by choosing δ small enough. We prove this assumption by contradiction. Suppose that $\underline{R}$ cannot be made positive for any $\delta > 0$; i.e., $\liminf_{n\to\infty, n\in\mathcal{J}} b_{k(n)-1} = 0$ for arbitrarily small $\delta > 0$. Then by following a similar procedure as in (3.9) and (3.10), we obtain

$$\begin{aligned}
0 &\le \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \ge 0\right] \\
&\le \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \ge 0\right] \\
&\le t\left(2\delta - \lim_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n}H_{1-t}(X^n)\right) \\
&= t\left(2\delta - \limsup_{n\to\infty} \frac{1}{n}H_{1-t}(X^n)\right),
\end{aligned}$$

which implies that

$$\limsup_{n\to\infty} \frac{1}{n}H_{1-t}(X^n) = 0$$

since δ can be made arbitrarily small, thus contradicting the positivity assumption on $\limsup_{n\to\infty} \frac{1}{n}H_{1-t}(X^n)$. The proof is therefore completed. $\square$

Observation:

It is important to point out that the proofs of the forward and converse parts do not directly depend on Theorem 1 or on source coding concepts. While the proof of the forward part is straightforward, the proof of the converse is more involved. More specifically, the objective of the converse part is to demonstrate that if $\limsup_{n\to\infty}(1/n)H_{1-t}(X^n)$ is slightly nudged to the left (by a factor of 3δ), then there exists a coordinate $\underline{R}$ on the R-axis such that a straight line of slope $\beta/(1+\beta)$ given by

$$y = \frac{\beta}{1+\beta}\left[R - \left(\limsup_{n\to\infty} \frac{1}{n}H_{1-t}(X^n) - 3\delta\right)\right]$$

lies above the curve of $\sigma(R)$ at $R = \underline{R}$, thus violating its status of support line for $\sigma(R)$.

This proof is established by observing that the desired coordinate $\underline{R}$ lies in a small neighborhood of τ, where τ is the largest point for which the spectrum $\sigma_{\mathcal{J}}^{(t)}(R)$ of the tilted distribution with parameter t for the random variable $-\log P_{X^n}(X^n)$ vanishes. A key point is to choose the tilted parameter t to be equal to $\beta/(1+\beta)$ which is the slope of the support line to $\sigma(R)$. We graphically illustrate this observation (based on a true example) in Figure 2.

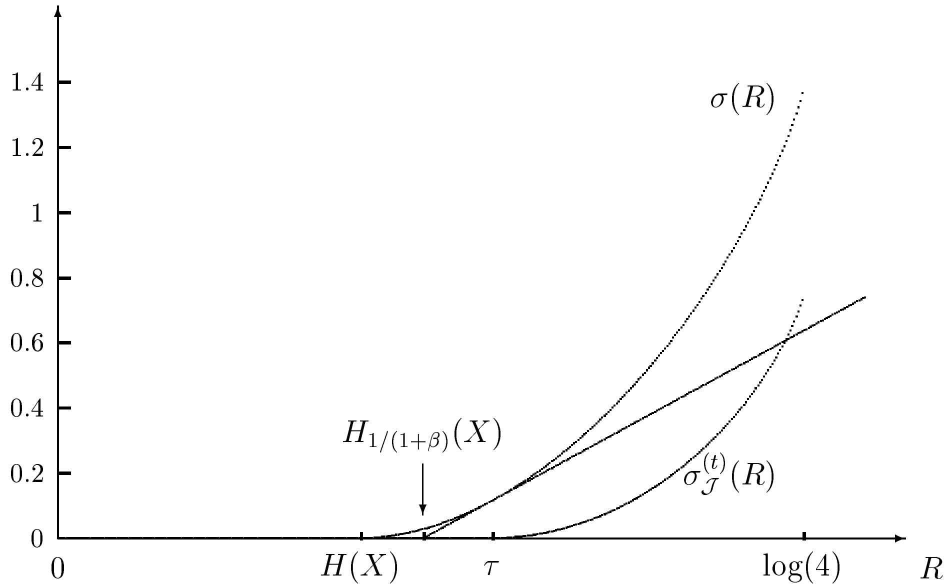

Figure 2: Functions $\sigma(R)$, $\sigma_{\mathcal{J}}^{(t)}(R)$ and $[\beta/(1+\beta)](R - \limsup_{n\to\infty}(1/n)H_{1/(1+\beta)}(X^n))$ for an i.i.d. binary source with $P_X(0) = 1 - P_X(1) = 1/4$ and $\beta = 9$ (or equivalently, $t = 0.9$). When $R > \log(4)$, $\sigma(R) = \sigma_{\mathcal{J}}^{(t)}(R) = \infty$.

4 Reverse β-cutoff rate

Definition 5 (Reverse β-cutoff rate) Fix $\beta > 0$. $R_0 \geq 0$ is a *reverse β-achievable rate* for a source $\boldsymbol{X}$ if

$$\underline{E}^*(R|\boldsymbol{X}) \geq -\beta(R - R_0)$$

for every $R > 0$, or equivalently,

$$\underline{R}^*(e|\boldsymbol{X}) \geq -\frac{1}{\beta}e + R_0,$$

for every $e > 0$. The *reverse β-cutoff rate* for $\boldsymbol{X}$ is defined as the supremum of all reverse β-achievable rates, and is denoted by $R_0^{(r)}(\beta|\boldsymbol{X})$. A graphical illustration of $R_0^{(r)}(\beta|\boldsymbol{X})$ is provided in Figure 3.

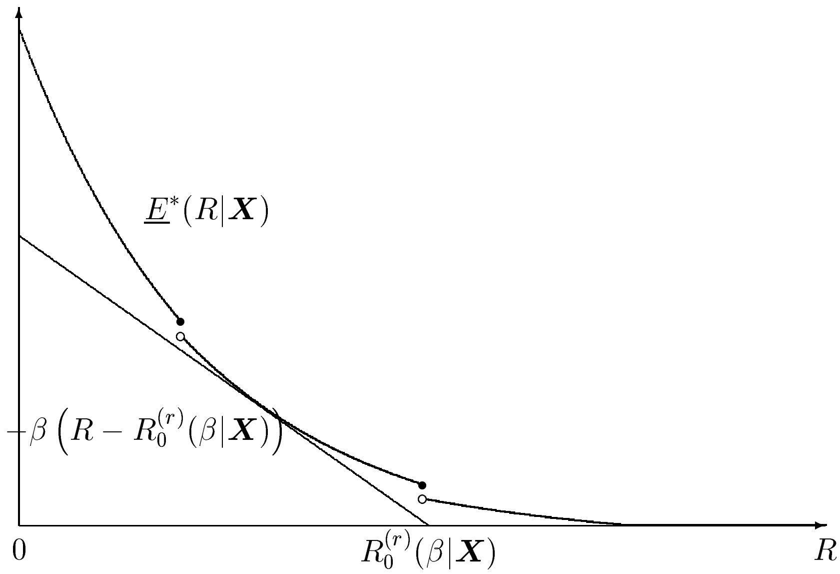

Figure 3: Reverse β-cutoff rate for an arbitrary source $\boldsymbol{X}$.

We observe that the above definition of β-cutoff rate is equivalent to Csiszár's definition of largest $\tilde{\beta}$-unachievable rate in [6, Definition 1], where $\tilde{\beta} = -\beta$.

We first prove the following two lemmas.

Lemma 2 Consider $\underline{\lambda}(R)$ defined in Theorem 2. Then the following properties hold.

1. For any $R > 0$ satisfying $\underline{\lambda}(R) > R$, $\underline{\lambda}(R) = \infty$.

2. If $\underline{\lambda}(R) > 0$ for some $R > 0$, then

$$R_\infty \stackrel{\triangle}{=} \sup\left\{R \geq 0 : \underline{\lambda}(R) > R\right\} > 0, \tag{4.11}$$

and for very $0 < R < R_\infty$,

$$\left|\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right\}\right| = 0 \quad \text{for all sufficiently large } n. \tag{4.12}$$

***Proof*:**

1. Let us prove this property by contradiction. Suppose there exists $R > 0$ such that $\underline{\lambda}(R) > R$, and that there exists $L < \infty$ with

$$\underline{\lambda}(R) = \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} < L.$$

   Then

$$Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} \geq e^{-nL} > 0 \quad \text{infinitely often in } n.$$

$$\Rightarrow \quad \left|\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right\}\right| \geq 1 \quad \text{infinitely often in } n.$$

   On the other hand, $\underline{\lambda}(R) > R$ implies the existence of $\delta > 0$ with $\underline{\lambda}(R) > R + \delta$, which implies that for all sufficiently large n

$$Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} \leq e^{-n(R+\delta)}.$$

   Consequently, for infinitely many n,

$$\begin{aligned} 1 &\leq \left|\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right\}\right| \\ &\leq e^{nR} Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} \\ &\leq e^{nR}e^{-n(R+\delta)} = e^{-n\delta} < 1. \end{aligned}$$

   Hence, the desired contradiction is obtained.

2. (4.11) is an immediate consequence of the non-increasing property of $\underline{\lambda}(R)$. We next prove (4.12) by contradiction. We know that for every $0 < R < R_\infty$, $\underline{\lambda}(R) = \infty$. Now suppose that

$$\left|\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right\}\right| \geq 1 \quad \text{infinitely often in } n.$$

Then for infinitely many n,

$$
\begin{aligned}
Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} &\geq e^{-nR}\left|\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right\}\right| \\
&\geq e^{nR}.
\end{aligned}
$$

Thus

$$
\underline{\lambda}(R) = \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} \leq R,
$$

which contradicts the fact that $\underline{\lambda}(R) = \infty$. □

Lemma 3 The following two conditions are equivalent.

$$
(\forall\ R > 0) \quad \underline{\lambda}(R) \geq -\frac{\beta}{1-\beta}(R - R_0) \tag{4.13}
$$

and

$$
(\forall\ e > 0) \quad \underline{R}^*(e|\boldsymbol{X}) \geq -\frac{1}{\beta}e + R_0, \tag{4.14}
$$

under $\beta \in (0,1)$.

Proof:

1. *Forward part*: (4.13) $\Rightarrow$ (4.14)

   For any $e > 0$, we obtain from Theorem 2 that $(\forall\ \delta > 0)(\exists\ h_\delta > 0)$

   $$
   \inf_{R>0}\{\underline{\lambda}(R) + [R - \underline{\lambda}(R) - h_\delta]^+\} \leq e \quad \text{and} \quad \underline{R}^*(e|\boldsymbol{X}) + \delta \geq \underline{h}_\delta,
   $$

   which in turn implies the existence of R_δ satisfying

   $$
   \begin{aligned}
   \inf_{R>0}\{\underline{\lambda}(R) + [R - \underline{\lambda}(R) - h_\delta]^+\} + \delta &\geq \underline{\lambda}(R_\delta) + [R_\delta - \underline{\lambda}(R_\delta) - h_\delta]^+ \\
   &= \max\{\underline{\lambda}(R_\delta), R_\delta - h_\delta\}.
   \end{aligned}
   $$

   Thus

   $$
   e + \delta \geq \underline{\lambda}(R_\delta) \quad \text{and} \quad e + \delta \geq R_\delta - h_\delta.
   $$

$$\begin{aligned}
\Rightarrow \underline{R}^*(e|\boldsymbol{X}) &\geq h_\delta - \delta \\
&\geq (R_\delta - e - \delta) - \delta \\
&\geq \left[-\frac{1-\beta}{\beta}\underline{\lambda}(R_\delta) + R_0\right] - e - 2\delta \\
&\geq -\frac{1-\beta}{\beta}(e+\delta) + R_0 - e - 2\delta \\
&= -\frac{1}{\beta}e + R_0 - \frac{1+\beta}{\beta}\delta.
\end{aligned}$$

The proof is then completed by noting that δ can be made arbitrarily small.

2. *Converse part:* (4.14) $\Rightarrow$ (4.13)

The claim holds trivially when $R < \underline{\lambda}(R)$ since it implies by Lemma 2 that $\underline{\lambda}(R) = \infty$. It remains to prove the claim under $R \geq \underline{\lambda}(R)$. Let

$$h_R \triangleq R - \underline{\lambda}(R) \quad \text{and} \quad e_R \triangleq \underline{\lambda}(R).$$

$$\Rightarrow \quad \inf_{a>0}\left(\underline{\lambda}(a) + [a - \underline{\lambda}(a) - h_R]^+\right) \leq \underline{\lambda}(R) + [R - \underline{\lambda}(R) - h_R]^+ = \underline{\lambda}(R) = e_R.$$

$$\Rightarrow \quad \underline{R}^*(e_R|\boldsymbol{X}) \leq h_R.$$

Therefore,

$$-\frac{1}{\beta}\underline{\lambda}(R) + R_0 = -\frac{1}{\beta}e_R + R_0 \leq \underline{R}^*(e_R|\boldsymbol{X}) \leq h_R = R - \underline{\lambda}(R).$$

Hence,

$$\underline{\lambda}(R) \geq -\frac{\beta}{1-\beta}(R - R_0).$$

□

Theorem 4 (Reverse β-cutoff rate formula) Fix $0 < \beta < 1$. For any source $\boldsymbol{X}$,

$$R_0^{(r)}(\beta|\boldsymbol{X}) = \liminf_{n\to\infty}\frac{1}{n}H_{1/(1-\beta)}(X^n). \tag{4.15}$$

Proof: The theorem holds if $\underline{\lambda}(R) = 0$ for all $R > 0$, in which case both the reverse β-cutoff rate and the liminf Rényi entropy rate are zero[3]. Without loss of generality, we assume that $\underline{\lambda}(R) > 0$ for some $R > 0$.

1. Forward part: $R_0^{(r)}(\beta|\boldsymbol{X}) \geq \liminf_{n\to\infty}(1/n)H_{1/(1-\beta)}(X^n)$

By (4.13), it suffices to show that

$$(\forall\ R > 0)\ \ \underline{\lambda}(R) \geq -\frac{\beta}{1-\beta}\left(R - \liminf_{n\to\infty}\frac{1}{n}H_{1/(1-\beta)}(X^n)\right).$$

$$\begin{aligned} Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) < R\right] &= Pr\left[P_{X^n}^t(X^n) > e^{-ntR}\right],\ \text{for } t > 0 \\ &\leq e^{ntR}\sum_{x^n\in\mathcal{X}^n} P_{X^n}^{1+t}(x^n),\ \text{for } t > 0 \text{ (by Markov's inequality)} \\ &= \exp\left\{nt\left(R - \frac{1}{n}H_{1+t}(X^n)\right)\right\}, \quad \text{for } t > 0. \end{aligned}$$

$$\begin{aligned} \Rightarrow\ \underline{\lambda}(R) &= \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) < R\right] \\ &\geq -t\left(R - \liminf_{n\to\infty}\frac{1}{n}H_{1+t}(X^n)\right), \quad \text{for } t > 0 \qquad (4.16) \\ &= -\frac{\beta}{1-\beta}\left(R - \liminf_{n\to\infty}\frac{1}{n}H_{1/(1-\beta)}(X^n)\right), \quad \text{for } 0 < \beta \triangleq \frac{t}{1+t} < 1. \end{aligned}$$

2. *Converse part:* $R_0^{(r)}(\beta|\boldsymbol{X}) \leq \liminf_{n\to\infty}(1/n)H_{1/(1-\beta)}(X^n)$

[3]It is straightforward to verify that the reverse β-cutoff rate is zero. We herein show that $\liminf_{n\to\infty}(1/n)H_{1/(1-\beta)}(X^n) = 0$. For $\alpha = 1/(1-\beta) > 1$,

$$P_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} = P_{X^n}\left\{P_{X^n}^{\alpha-1}(X^n) > e^{-(\alpha-1)nR}\right\} \leq \frac{E\left[P_{X^n}^{\alpha-1}(X^n)\right]}{e^{-(\alpha-1)nR}} = e^{(\alpha-1)nR}e^{(1-\alpha)H_\alpha(X^n)}.$$

Since $\underline{\lambda}(R) = 0$ for any $R > 0$, we get that

$$0 = \underline{\lambda}(R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) < R\right\} \geq (1-\alpha)R + (\alpha-1)\liminf_{n\to\infty}\frac{1}{n}H_\alpha(X^n).$$

Thus, for any $R > 0$, $R \geq \liminf_{n\to\infty}(1/n)H_\alpha(X^n)$. Therefore, $\liminf_{n\to\infty}(1/n)H_\alpha(X_n) = 0$. $\square$

By (4.13), it suffices to show that for any $\delta > 0$ arbitrarily small, there exists R_1 such that

$$\underline{\lambda}(R_1) \leq -\frac{\beta}{1-\beta}\left(R_1 - \liminf_{n\to\infty}\frac{1}{n}H_{1/(1-\beta)}(X^n) - 3\delta\right).$$

Define the tilted distribution

$$\begin{aligned} P_{X^n}^{(t)}(x^n) &\triangleq \frac{P_{X^n}^{1-t}(x^n)}{\sum_{\hat{x}^n\in\mathcal{X}^n} P_{X^n}^{1-t}(\hat{x}^n)} \\ &= \exp\{-t[\log P_{X^n}(x^n) + H_{1-t}(X^n)]\}P_{X^n}(x^n), \end{aligned} \tag{4.17}$$

where $t = -\beta/(1-\beta) < 0$. Also define

$$\tau \triangleq \sup\left\{R \in [0,\infty) \ : \ \underline{\lambda}^{(t)}(R) > 0\right\}$$

and

$$m_n^{(t)} \triangleq \sum_{x^n\in\mathcal{X}^n} P_{X^n}^{(t)}(x^n)[-\log P_{X^n}(x^n)],$$

where

$$\underline{\lambda}^{(t)}(R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right].$$

We first point out that τ is positive and finite. Our assumption about the existence of $R > 0$ such that $\underline{\lambda}(R) > 0$ implies via Lemma 2 that

$$R_\infty \triangleq \sup\{R \geq 0 : \underline{\lambda}(R) > R\} > 0$$

and for $0 < R < R_\infty$,

$$\left|\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < R\right\}\right| = 0$$

for all sufficiently large n. Thus $\underline{\lambda}^{(t)}(R) = \infty$ for $0 < R < R_\infty$. Therefore,

$$\tau \triangleq \sup\left\{R \geq 0 : \underline{\lambda}^{(t)}(R) > 0\right\} \geq R_\infty > 0.$$

Furthermore, we show in Appendix B (cf. Lemma 6) that $\tau \leq \log|\mathcal{X}|$.

We next observe that

$$(1-t)\log P_{X^n}(x^n) = \log P_{X^n}^{(t)}(x^n) + tH_{1-t}(X^n).$$

Hence,

$$\begin{aligned}
m_n^{(t)} &= \frac{1}{1-t}\sum_{x^n\in\mathcal{X}^n} P_{X^n}^{(t)}(x^n)\log\frac{1}{P_{X^n}^{(t)}(x^n)} + \frac{-t}{1-t}H_{1-t}(X^n) \\
&\le \frac{1}{1-t}\log|\mathcal{X}|^n + \frac{-t}{1-t}\log|\mathcal{X}|^n = \log|\mathcal{X}|^n. \qquad (4.18)
\end{aligned}$$

Since $0<\tau\le\log|\mathcal{X}|$, it follows from the definition of τ that for any $0<\delta<\min\{\tau, 2\log|\mathcal{X}|-\tau\}$, there exists $\varepsilon>0$ such that

$$\underline{\lambda}^{(t)}(\tau-\delta) = \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) < \tau-\delta\right] > \varepsilon > 0.$$

Thus

$$P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) \ge \tau-\delta\right] > 1-e^{-n\varepsilon} \quad \text{for all sufficiently large } n.$$

Therefore, for those n satisfying the above inequality,

$$\begin{aligned}
& P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : 2\log|\mathcal{X}| > -\frac{1}{n}\log P_{X^n}(x^n) \ge \tau-\delta\right] \\
\ge\ & P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : \frac{2}{n}m_n^{(t)} > -\frac{1}{n}\log P_{X^n}(x^n) \ge \tau-\delta\right] \quad \text{(by (4.18))} \\
=\ & P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : \frac{1}{n}\log P_{X^n}(x^n) \ge \tau-\delta\right] \\
& -P_{X^n}^{(t)}\left[x^n\in\mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) \ge 2\frac{m_n^{(t)}}{n}\right] \\
\ge\ & 1-e^{-n\varepsilon}-\frac{1}{2} \quad \text{(by Markov's inequality)} \\
=\ & \frac{1-2e^{-n\varepsilon}}{2}.
\end{aligned}$$

Let

$$I_k \triangleq [b_{k-1}, b_k) \quad \text{for } 1\le k\le L \triangleq \left\lfloor \frac{2\log|\mathcal{X}|-\tau+\delta}{2\delta}\right\rfloor,$$

where $b_k \triangleq (\tau - \delta) + 2k\delta$ for $1 \le k < L$, and $b_L \triangleq 2\log|\mathcal{X}|$. Note that $b_k - b_{k-1} = 2\delta$ for every $1 \le k < L$ and $b_L - b_{L-1} \ge 2\delta$. Therefore there exists $1 \le k(n) \le L$ such that

$$P_{X^n}^{(t)}\left[-\frac{1}{n}\log P_{X^n}(X^n) \in I_{k(n)}\right] \ge \frac{1-2e^{-n\varepsilon}}{2L} \qquad \text{for all sufficiently large } n.$$

Then, by letting $R_1 \triangleq \limsup_{n\to\infty} b_{k(n)} + \delta$ and noting that $R_1 \ge b_{k(n)}$ for all sufficiently large n, we obtain that

$$Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) < R_1\right] \ge Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \in I_{k(n)}\right] \text{ for all } n \text{ sufficiently large.}$$

However, for all sufficiently large n, we have that

$$\begin{aligned}
& Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) \in I_{k(n)}\right] \\
= & \sum_{x^n \in \{x^n \in \mathcal{X}^n : -(1/n)\log P_{X^n}(x^n) \in I_{k(n)}\}} P_{X^n}(x^n) \\
= & \sum_{x^n \in \{x^n \in \mathcal{X}^n : -(1/n)\log P_{X^n}(x^n) \in I_{k(n)}\}} e^{t[\log P_{X^n}(x^n) + H_{1-t}(X^n)]} P_{X^n}^{(t)}(x^n) \qquad \text{(by (4.17))} \\
\ge & e^{-nt[b_{k(n)-1} - (1/n)H_{1-t}(X^n)]} \sum_{x^n \in \{x^n \in \mathcal{X}^n : -(1/n)\log P_{X^n}(x^n) \in I_{k(n)}\}} P_{X^n}^{(t)}(x^n) \qquad (4.19) \\
= & e^{-nt[b_{k(n)-1} - (1/n)H_{1-t}(X^n)]} P_{X^n}^{(t)}\left[-\frac{1}{n}\log P_{X^n}(X^n) \in I_{k(n)}\right] \\
\ge & \frac{1-2e^{-n\varepsilon}}{2L} e^{-nt[b_{k(n)-1} - (1/n)H_{1-t}(X^n)]},
\end{aligned}$$

where (4.19) follows from the fact that $b_{k(n)-1} \le -(1/n)\log P_{X^n}(x^n) < b_{k(n)}$ and that $t < 0$. Consequently,

$$\begin{aligned}
\underline{\lambda}(R_1) &= \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left[-\frac{1}{n}\log P_{X^n}(X^n) < R_1\right] \\
&\le t\left(\limsup_{n\to\infty} b_{k(n)-1} - \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n)\right) \\
&\le t\left(\limsup_{n\to\infty} b_{k(n)} - \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) - 2\delta\right) \\
&= t\left(R_1 - \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) - 3\delta\right). \qquad (4.20)
\end{aligned}$$

This completes the proof. □

Remark: For the case of $\beta \geq 1$, the expression of the reverse β-cutoff rate is no longer provided by (4.15). It can actually be shown that for $\beta \geq 1$,

$$R_0^{(r)}(\beta|\boldsymbol{X}) = \frac{1}{\beta} \liminf_{n\to\infty} \frac{1}{n} H_\infty(X^n),$$

where

$$H_\infty(X^n) = \lim_{\alpha\uparrow\infty} H_\alpha(X^n) = -\log \max_{x^n \in \mathcal{X}^n} P_{X^n}(x^n)$$

is the Rényi entropy of infinite order.

5 Concluding remarks

In this paper, general expressions for the forward and reverse β-cutoff rates, $R_0^{(f)}(\beta|\boldsymbol{X})$ and $R_0^{(r)}(\beta|\boldsymbol{X})$ respectively, for an arbitrary discrete-time finite-alphabet source $\boldsymbol{X}$ were established. More specifically, it was demonstrated that

$$R_0^{(f)}(\beta|\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} H_{1/(1+\beta)}(X^n)$$

and

$$R_0^{(r)}(\beta|\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} H_{1/(1-\beta)}(X^n).$$

These results – which provide a new operational characterization for the Rényi entropy rates (in addition to the variable length source coding characterization under exponential cost constraints investigated in [16]) – generalize Csiszár's previous work [6] on the β-cutoff rates, where he only considered the case of memoryless sources. It can be directly verified that if the source $\boldsymbol{X}$ is memoryless, then Theorems 3 and 4 simplify to Csiszár's result [6, Theorem 1]. In closing, we would like to make the following observations.

- It is important to point out that if the source $\boldsymbol{X}$ is a time-invariant Markov source of arbitrary order, then its Rényi entropy rate exists and can be computed [16, 17]. Thus in this case, the β-cutoff rates for this source can be obtained.

- It directly follows from the definition of the source reliability function $E(R|\boldsymbol{X})$ of $\boldsymbol{X}$ that a convex lower bound can be obtained on $E(R|\boldsymbol{X})$. It consists of the supremum of all the support lines with slope β which pass through the point $(R_0^{(f)}(\beta|\boldsymbol{X}), 0)$: for each $R > 0$,

$$E(R|\boldsymbol{X}) \geq \sup_{\beta>0}[\beta(R - R_0^{(f)}(\beta|\boldsymbol{X}))]. \tag{5.21}$$

  Note that since the right-hand side of (5.21) is the best convex lower bound to $E(R|\boldsymbol{X})$, then the inequality given by (5.21) becomes tight whenever $E(R|\boldsymbol{X})$ is convex. This is the case for irreducible Markov sources [18, 17]. Furthermore, for the class of sources $\boldsymbol{X}$ for which $E(R|\boldsymbol{X})$ is not known but its Rényi entropy rate can be calculated (e.g., the class of non-irreducible Markov sources [17]), a computable lower bound to $E(R|\boldsymbol{X})$ can also be obtained. A similar remark applies for the source unreliability function $\underline{E}^*(R|\boldsymbol{X})$.

Appendix A

Lemma 4 For $t = \beta/(1+\beta) \in (0,1)$,

$$\inf\left\{R : \sigma_{\mathcal{J}}^{(t)}(R) > 0\right\} \leq \frac{1}{1-t}\log|\mathcal{X}|,$$

for every increasing sequence of positive integers $\mathcal{J} = \{n_j\}_{j\geq 1}$.

***Proof*:** Let us prove the result by contradiction. Suppose that

$$\sigma_{\mathcal{J}}^{(t)}\left(\frac{\log|\mathcal{X}|+\delta}{1-t}\right) = 0$$

for some positive δ. Then

$$\begin{aligned} 0 &= \sigma_{\mathcal{J}}^{(t)}\left(\frac{\log|\mathcal{X}|+\delta}{1-t}\right) \\ &\triangleq \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}(x^n) \geq \frac{\log|\mathcal{X}|+\delta}{1-t}\right] \end{aligned}$$

$$
\begin{aligned}
&= \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{1-t}(x^n) \ge \log|\mathcal{X}| + \delta\right] \\
&= \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) - \frac{t}{n}H_{1-t}(X^n) \ge \log|\mathcal{X}| + \delta\right] \\
&= \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \frac{t}{n}H_{1-t}(X^n) + \log|\mathcal{X}| + \delta\right] \\
&\ge \liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right],
\end{aligned}
$$

where the last step follows since $H_{1-t}(X^n) \ge 0$. Thus we can say that

$$
\liminf_{n\to\infty,\ n\in\mathcal{J}} -\frac{1}{n}\log P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right] < \frac{\delta}{2},
$$

$$
\Rightarrow \quad P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right] > e^{-n\delta/2} \quad \text{infinitely often in } n \in \mathcal{J}.
$$

For those n satisfying the above inequality, the set

$$
\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right]
$$

is non-empty, and hence,

$$
\begin{aligned}
& P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right] \\
\le\ & \left|\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right]\right| \frac{1}{|\mathcal{X}|^n e^{n\delta}}.
\end{aligned}
$$

Finally, we obtain the contradiction by observing that

$$
\begin{aligned}
|\mathcal{X}|^n &\ge \left|\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right]\right| \\
&\ge |\mathcal{X}|^n e^{n\delta} \cdot P_{X^n}^{(t)}\left[x^n : -\frac{1}{n}\log P_{X^n}^{(t)}(x^n) \ge \log|\mathcal{X}| + \delta\right] \\
&> |\mathcal{X}|^n e^{n\delta} e^{-n\delta/2} = |\mathcal{X}|^n e^{n\delta/2} \quad \text{for infinitely many } n \in \mathcal{J}.
\end{aligned}
$$

□

Lemma 5 For $t \in (0,1)$ and every increasing sequence of positive integers $\mathcal{J} = \{n_j\}_{j\ge1}$, if

$$
\limsup_{n\to\infty,\ n\in\mathcal{J}} (1/n)H_{1-t}(X^n) > 0,
$$

then

$$\inf\left\{R : \sigma^{(t)}_{\mathcal{J}}(R) > 0\right\} \geq \limsup_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n} H_{1-t}(X^n).$$

Proof: $(\forall\ \mu > 0)$,

$$\begin{aligned}
& P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq \limsup_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n} H_{1-t}(X^n) - 2\mu\right\} \\
\geq\ \ & P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq \frac{1}{n} H_{1-t}(X^n) - \mu\right\} \text{ for infinitely many } n \in \mathcal{J}.
\end{aligned}$$

But

$$\begin{aligned}
& P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq \frac{1}{n} H_{1-t}(X^n) - \mu\right\} \\
=\ \ & P^{(t)}_{X^n}\left\{\frac{1}{n}\left(-t\left[\log P_{X^n}(X^n) + H_{1-t}(X^n)\right]\right) \geq -\mu t\right\} \\
=\ \ & P^{(t)}_{X^n}\left\{\frac{1}{n}\log\frac{P^{(t)}_{X^n}(X^n)}{P_{X^n}(X^n)} \geq -\mu t\right\} \\
=\ \ & 1 - P^{(t)}_{X^n}\left\{\frac{1}{n}\log\frac{P^{(t)}_{X^n}(X^n)}{P_{X^n}(X^n)} < -\mu t\right\} \\
=\ \ & 1 - P^{(t)}_{X^n}\left\{P^{(t)}_{X^n}(X^n) < e^{-n\mu t} P_{X^n}(X^n)\right\} \\
\geq\ \ & 1 - e^{-n\mu t} \cdot P_{X^n}\left\{P^{(t)}_{X^n}(X^n) < e^{-n\mu t} P_{X^n}(X^n)\right\} \\
\geq\ \ & 1 - e^{-n\mu t}.
\end{aligned}$$

Thus

$$\begin{aligned}
& P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq \limsup_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n} H_{1-t}(X^n) - 2\mu\right\} \\
\geq\ \ & 1 - e^{-n\mu t} \text{ for infinitely many } n \in \mathcal{J}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
& (\forall\ \mu > 0)\ \sigma^{(t)}_{\mathcal{J}}\left(\limsup_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n} H_{1-t}(X^n) - 2\mu\right) = 0 \\
\Rightarrow\ \ & \inf\{R : \sigma^{(t)}_{\mathcal{J}}(R) > 0\} \geq \limsup_{n\to\infty,\ n\in\mathcal{J}} \frac{1}{n} H_{1-t}(X^n) - 2\mu.
\end{aligned}$$

$\square$

Appendix B

Lemma 6 For $t < 0$,

$$\sup\left\{R : \underline{\lambda}^{(t)}(R) > 0\right\} \leq \liminf_{n\to\infty} \frac{1}{n} H_{1-t}(X^n) \leq \log|\mathcal{X}|.$$

Proof: For any $\mu > 0$,

$$\begin{aligned} & P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) \geq \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) + 2\mu\right\} \\ \leq \;& P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) > \frac{1}{n}H_{1-t}(X^n) + \mu\right\} \quad \text{for infinitely many } n. \end{aligned}$$

But

$$\begin{aligned} & P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) > \frac{1}{n}H_{1-t}(X^n) + \mu\right\} \\ = \;& P^{(t)}_{X^n}\left\{\frac{1}{n}\left(-t\left[\log P_{X^n}(X^n) + H_{1-t}(X^n)\right]\right) < \mu t\right\} \quad \text{for } t < 0 \\ = \;& P^{(t)}_{X^n}\left\{\frac{1}{n}\log\frac{P^{(t)}_{X^n}(X^n)}{P_{X^n}(X^n)} < \mu t\right\} \\ = \;& P^{(t)}_{X^n}\left\{P^{(t)}_{X^n}(X^n) < e^{n\mu t}P_{X^n}(X^n)\right\} \\ \leq \;& e^{n\mu t}P_{X^n}\left\{P^{(t)}_{X^n}(X^n) < e^{n\mu t}P_{X^n}(X^n)\right\} \\ \leq \;& e^{n\mu t}. \end{aligned}$$

Thus for infinitely many n,

$$P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) < \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) + 2\mu\right\} \;\geq\; 1 - e^{n\mu t},$$

which implies

$$\underline{\lambda}^{(t)}\left(\liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) + 2\mu\right)$$

$$
\begin{aligned}
&= \liminf_{n\to\infty} -\frac{1}{n}\log P^{(t)}_{X^n}\left\{-\frac{1}{n}\log P_{X^n}(X^n) < \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n)+2\mu\right\} \\
&\le \limsup_{n\to\infty} -\frac{1}{n}\log\left(1-e^{n\mu t}\right) = 0 \qquad \text{(since } t<0\text{)}.
\end{aligned}
$$

Consequently,

$$
\sup\left\{R : \underline{\lambda}^{(t)}(R) > 0\right\} \le \liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) + 2\mu.
$$

The proof is completed by noting that μ can be made arbitrarily small and that

$$
\liminf_{n\to\infty}\frac{1}{n}H_{1-t}(X^n) \le \log|\mathcal{X}|.
$$

□

Acknowledgments

The authors would like to thank the reviewers and Professor Csiszár for their constructive advice that helped improve this paper.

References

[1] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, No. 1, pp. 99–105, January 1996.

[2] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, pp. 405–417, 1974.

[3] R. E. Blahut, *Principle and Practice of Information Theory*, Addison-Wesley, MA, 1987.

[4] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*, Wiley, New York, 1990.

[5] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, 8, pp. 423–429, 1965.

[6] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on Information Theory*, vol. 41, pp. 26–34, January 1995.

[7] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Sci. Math. Hungar.*, vol. 6, pp. 181–191, 1971.

[8] L. D. Davisson, G. Longo and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Transactions on Information Theory*, vol. 27, pp. 431–438, July 1981.

[9] U. Erez and R. Zamir, "Noise prediction for channels with side information at the transmitter: error exponents," *Proc. Information Theory Workshop*, Metsovo, Greece, June 1999.

[10] T. S. Han, *Information-Spectrum Methods in Information Theory*, (in Japanese), Baifukan Press, Tokyo, 1998.

[11] T. S. Han, "The reliability functions of the general source with fixed-length coding," *IEEE Transactions on Information Theory*, vol. 46, pp. 2117-2132, Sept. 2000.

[12] F. Jelinek, "Buffer overflow in variable length coding of fixed rate sources," *IEEE Transactions on Information Theory*, vol. IT-14, pp. 490–501, May 1968.

[13] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 20, pp. 197–199, 1974.

[14] N. Merhav, "Universal coding with minimum probability of codeword length overflow," *IEEE Transactions on Information Theory*, vol. 37, pp. 556–563, May 1991.

[15] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. on Math. Statist. Probability*, Berkeley, CA, vol. 1, pp. 547–561, 1961.

[16] Z. Rached, F. Alajaji and L. L. Campbell, "Rényi's entropy rate for discrete Markov sources," *Proc. CISS'99*, March 17-19, Baltimore, MD, 1999.

[17] Z. Rached, F. Alajaji and L. L. Campbell, "Rényi's divergence and entropy rates for finite alphabet Markov sources," *IEEE Transactions on Information Theory*, to be published May 2001.

[18] K. Vasek, "On the error exponent for ergodic Markov source," *Kybernetika*, vol. 16, no. 4, pp. 318–329, 1980.