Csiszár's Cutoff Rates for the General Hypothesis Testing Problem

Fady Alajaji Po-Ning Chen Ziad Rached

Abstract

In [6], Csiszár established the concept of forward β-cutoff rate for the error exponent hypothesis testing problem based on independent and identically distributed (i.i.d.) observations. Given $\beta < 0$, he defined the forward β-cutoff rate as the number $R_0 \geq 0$ that provides the best possible lower bound in the form $\beta(E - R_0)$ to the type 1 error exponent function for hypothesis testing where $0 < E < R_0$ is the rate of exponential convergence to 0 of the type 2 error probability. He then demonstrated that the forward β-cutoff rate is given by $D_{1/(1-\beta)}(X\|\bar{X})$, where $D_\alpha(X\|\bar{X})$ denotes the Rényi α-divergence [19], $\alpha > 0$, $\alpha \neq 1$. Similarly, for $0 < \beta < 1$, Csiszár also established the concept of reverse β-cutoff rate for the correct exponent hypothesis testing problem.

In this work, we extend Csiszár's results by investigating the forward and reverse β-cutoff rates for the hypothesis testing between two arbitrary sources with memory. We demonstrate that the liminf Rényi α-divergence rate provides the expression for the forward β-cutoff rate. We also show that if the log-likelihood large deviation spectrum admits a limit, then the reverse β-cutoff rate equals the liminf α-divergence rate, where $\alpha = \frac{1}{1-\beta}$ and $0 < \beta < \beta_{\max}$, where $\beta_{\max}$ is the largest $\beta < 1$ for which the liminf $\frac{1}{1-\beta}$-divergence rate is finite. For $\beta_{\max} \leq \beta < 1$, we show that the reverse cutoff rate is in general only upper bounded by the liminf Rényi divergence rate. Unlike [4] where the alphabet for the source coding cutoff rate problem was assumed to be finite, we assume arbitrary (countable or continuous) source alphabet. We also provide several examples to illustrate our forward and reverse β-cutoff rates results and the techniques employed to establish them.

Index Terms: Arbitrary sources with memory, α-divergence rate, forward and reverse cutoff rates, hypothesis testing error and correct exponents, information spectrum, large deviation theory.

IEEE Transactions on Information Theory: Submitted September 2002, Revised August 2003. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Premier's Research Excellence Award of Ontario, and the National Science Council of Taiwan. This work was presented in part at the 2002 and 2003 International Symposiums on Information Theory.

F. Alajaji is with the Department of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada. P.-N. Chen is with the Department of Communcation Engineering, National Chiao Tung Univrsity, Hsin Chu, Taiwan 30050, R.O.C. Z. Rached was with Department of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada; he is now with the Department of Mathematics & Statistics, Notre Dame University, Zouk Mosbeh, Keserouan, P. O. Box 72 Zouk Mikael, Lebanon.

1 Introduction

In [6], Csiszár established the concept of forward β-cutoff rate for the hypothesis testing problem based on independent and identically distributed (i.i.d.) observations. Given $\beta < 0$, he defined the forward β-cutoff rate as the number $R_0 \geq 0$ that provides the best possible lower bound in the form $\beta(E - R_0)$ to the type 1 error exponent function for hypothesis testing where $0 < E < R_0$ is the rate of exponential convergence to 0 of the type 2 error probability. He then demonstrated that the forward β-cutoff rate is given by $D_{1/(1-\beta)}(X\|\bar{X})$, where $D_\alpha(X\|\bar{X})$ denotes Rényi's α-divergence, $\alpha > 0$, $\alpha \neq 1$ [19]. Csiszár also established the concept of reverse β-cutoff rate for the hypothesis testing problem based on i.i.d. observations. Given $\beta > 0$, he defined the reverse β-cutoff rate as the number $R_0 \geq 0$ that yields the best possible lower bound in the form $\beta(E - R_0)$ to the hypothesis testing type 1 correct exponent (or reliability) function where $0 < R_0 < E$ is the rate of exponential convergence to 0 of the type 2 error probability. He then showed that the reverse β-cutoff rate is equal to $D_{1/(1-\beta)}(X\|\bar{X})$. These results provide a new operational significance for the α-divergence.

The error exponent for the binary hypothesis testing problem has been thoroughly studied for finite state i.i.d. sources and Markov sources. The results for i.i.d. sources can be found in [7], [10], [12], and for irreducible Markov sources in [1], [15]. The error exponent for testing between ergodic Markov sources with continuous state-space under certain additional restrictions was established in [13]. In its full generality, i.e., for arbitrary sources (not necessarily, stationary, ergodic, etc.), the error exponent was studied in [5], [8], [9], [14] and [11]. Specifically, in [8, 9], Han established expressions for the type 2 error and correct exponents. While Han's error exponent formula holds in full generality, his reverse exponent characterization requires some technical assumptions on the log-likelihood large deviation spectrum. In [14], Nagaoka and Hayashi generalized Han's results in the larger quantum hypothesis testing setting and without requiring any conditions. Recently, Iriyama [11] revisited Han's results without imposing any asumptions and provided alternative expressions for the type 2 error and correct exponents in terms of the spectral inf/sup divergence rates.

In the sequel, we extend Csiszár's results by investigating the forward and reverse β-cutoff rates for the hypothesis testing between two arbitrary sources with memory. We demonstrate that the liminf Rényi α-divergence rate provides the expression for the forward β-cutoff rate. Our proof relies in part on the formulas established in [8], and considerable extensions of the techniques used in [4] that generalize Csiszár's source coding cutoff rate results for arbitrary discrete sources with memory. Unlike [4] where the source alphabet was assumed to be finite, we assume arbitrary (countable or continuous) source alphabet. The methods used in our proof are a mixture of the techniques used in deriving the forward and reverse β-cutoff rates for source coding [4]. However, some new techniques are also needed to obtain our result.

We also investigate the reverse β-cutoff rate problem for arbitrary sources with memory. We show that if the log-likelihood ratio large deviation spectrum, $\rho(R)$, admits a limit, then the liminf α-divergence rate with $\alpha = \frac{1}{1-\beta}$ provides the expression for the reverse β-cutoff rate for $0 < \beta < \beta_{\max}$, where $\beta_{\max}$ is the largest $\beta < 1$ for which the liminf $\frac{1}{1-\beta}$-divergence rate is finite. We also observe (via an example), that if the limit of $\rho(R)$ does not hold, the reverse cutoff rate *may not* equal the liminf Rényi divergence rate in general. For $\beta_{\max} \leq \beta < 1$, we show that the reverse β-cutoff rate is in general only upper bounded by the liminf Rényi divergence rate.

The rest of the paper is organized as follows. In Section 2, we briefly recall previous results by Han [8] on the general expression for the Neyman-Pearson type 2 error subject to an exponential bound on the type 1 error. In Section 3, we establish the formula for the forward β-cutoff rate, and we illustrate it via some examples in Section 4. In Section 5, we recall the general expression for the reliability function of the type 2 probability of correct decoding [8] and formulate the reverse β-cutoff rate problem by carefully examining the inconsistency of definitions between [6] and [8]. In Section 6, we investigate the reverse β-cutoff rate, and in Section 7 we illustrate it with some examples. Finally, we conclude in Section 8.

2 Hypothesis Testing Error Exponent

Let us first define the general source as an infinite sequence

$$\boldsymbol{X} = \{X^n\}_{n=1}^{\infty} \triangleq \left\{ X^n = \left(X_1^{(n)}, \ldots, X_n^{(n)} \right) \right\}_{n=1}^{\infty}$$

of n-dimensional random variables X^n where each component random variable $X_i^{(n)}$ $(1 \leq i \leq n)$ takes values in an arbitrary (countable or continuous) set $\mathcal{X}$ that we call the source alphabet. Given two arbitrary sources $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ and $\bar{\boldsymbol{X}} = \{\bar{X}^n\}_{n=1}^{\infty}$ taking values in the same source alphabet $\{\mathcal{X}^n\}_{n=1}^{\infty}$, we may define the general hypothesis testing problem with $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ as the null hypothesis and $\bar{\boldsymbol{X}} = \{\bar{X}^n\}_{n=1}^{\infty}$ as the alternative hypothesis.

Let $\mathcal{A}_n$ be any subset of $\mathcal{X}^n$, $n = 1, 2, \ldots$ that we call the acceptance region of the hypothesis testing, and define

$$\mu_n \triangleq \Pr\{X^n \notin \mathcal{A}_n\} \quad \text{and} \quad \lambda_n \triangleq \Pr\{\bar{X}^n \in \mathcal{A}_n\}$$

where μ_n, λ_n are called type 1 error probability and type 2 error probability, respectively.

Definition 1 [6] Fix $E > 0$. A rate r is called E-achievable if there exists a sequence of acceptance regions $\mathcal{A}_n$ such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mu_n \geq r \quad \text{and} \quad \liminf_{n\to\infty} -\frac{1}{n}\log \lambda_n \geq E.$$

Definition 2 [6] The supremum of all E-achievable rates is denoted by $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$:

$$D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \triangleq \sup\{r > 0 : r \text{ is } E\text{-achievable}\},$$

and $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 0$ if the above set is empty (which is a degenerate uninteresting case). The dual of the function $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is defined as:

$$B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \triangleq \sup\{E > 0 : E \text{ is } r\text{-achievable}\},$$

and $B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 0$ if the above set is empty.

Proposition 1 [8] Fix $r > 0$. For the general hypothesis testing problem, we have that

$$B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \inf_{R\in\mathbb{R}}\{R + \eta(R) : \eta(R) < r\},$$

where[1]

$$\eta(R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n \ : \ \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \leq R\right\}$$

is the large deviation spectrum of the normalized log-likelihood ratio.

For the sake of simplicity, we assume throughout that the source alphabet is countable. However, we will point out the necessary modifications in the proofs for the case of a continuous alphabet. The above proposition is the main tool for our key lemma in the following section.

[1] If the source alphabet $\mathcal{X}$ is continuous, then $P_{X^n}(X^n)$ plays the role of the density function $f_{X^n}(X^n)$, when it exists.

3 Forward β-Cutoff Rate for Arbitrary Sources

Definition 3 [**6**] Fix $\beta < 0$. $R_0 \geq 0$ is a forward β-achievable rate for the general hypothesis testing problem if

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

for every $E > 0$, or equivalently,

$$B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \geq R_0 + \frac{r}{\beta},$$

for every $r > 0$. The forward β-cutoff rate is defined as the supremum of all forward β-achievable rates, and is denoted by $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$.

Note that in the degenerate case where $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is identically 0, we have that $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 0$. We herein assume that $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is not 0 for all values of E. A graphical illustration of the forward β-cutoff rate, $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, for testing between two arbitrary sources $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ is given in Figure 1.

Before stating our main result, we first observe in the next lemma that the forward β-cutoff rate $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is indeed the R-axis intercept of a support line of slope $\frac{\beta}{1-\beta}$ to the large deviation spectrum $\eta(R)$.

Lemma 1 Fix $\beta < 0$. The following conditions are equivalent.

$$(\forall R \in \mathbb{R}) \quad \eta(R) \geq \frac{\beta}{\beta - 1}(R_0 - R) \tag{1}$$

and

$$(\forall r > 0) \quad B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \geq R_0 + \frac{r}{\beta}. \tag{2}$$

Proof:

a) (1) $\Rightarrow$ (2). For any $r > 0$, we obtain by Proposition 1 that

$$(\forall \delta > 0)(\exists R_\delta \text{ with } \eta(R_\delta) < r) \quad B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) + \delta \geq R_\delta + \eta(R_\delta).$$

Therefore

$$\begin{aligned}
B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) &\geq R_\delta + \eta(R_\delta) - \delta \\
&\geq R_\delta - \delta + \frac{\beta}{\beta-1}(R_0 - R_\delta) \qquad (3)\\
&= -\delta + \frac{\beta}{\beta-1}R_0 - \frac{R_\delta}{\beta-1} \\
&\geq -\delta + \frac{\beta}{\beta-1}R_0 - \frac{R_0}{\beta-1} + \frac{r}{\beta} \qquad (4)\\
&= \frac{r}{\beta} + R_0 - \delta, \qquad (5)
\end{aligned}$$

where (3) follows from (1), and (4) holds because

$$r > \eta(R_\delta) \geq \frac{\beta}{\beta-1}(R_0 - R_\delta).$$

Since δ can be made arbitrarily small, the proof of the forward part is completed.

b) (2) $\Rightarrow$ (1). Inequality (1) holds trivially for those R satisfying $\eta(R) = \infty$. For any $R \in \mathbb{R}$ with $\eta(R) < \infty$, let $r_\delta \triangleq \eta(R) + \delta$ for some $\delta > 0$. Then (by Proposition 1)

$$B_e(r_\delta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \leq R + \eta(R).$$

Therefore

$$\begin{aligned}
\eta(R) &\geq B_e(r_\delta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) - R \\
&\geq R_0 + \frac{r_\delta}{\beta} - R \qquad (6)\\
&= R_0 + \frac{\eta(R)}{\beta} + \frac{\delta}{\beta} - R,
\end{aligned}$$

where (6) follows by (2). Thus,

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

Since δ can be made arbitrarily small, the proof of the converse part is completed.

$\square$

Theorem 1 (Forward β-cutoff rate formula). Fix $\beta < 0$. For the general hypothesis testing problem,

$$R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \liminf_{n\to\infty} \frac{1}{n} D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n),$$

where

$$D_\alpha(X^n\|\bar{X}^n) \triangleq \frac{1}{\alpha-1}\log\left(\sum_{x^n\in\mathcal{X}^n}[P_{X^n}(x^n)]^\alpha[P_{\bar{X}^n}(x^n)]^{1-\alpha}\right)$$

is the n-dimensional α-divergence.[2]

Proof: Note that $\eta(R) > 0$ for some[3] $R \in \mathbb{R}$.

1. *Forward part*: $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \geq \liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)$.

   By the equivalence of conditions (1) and (2), it suffices to show that

$$(\forall R\in\mathbb{R})\ \eta(R) \geq \frac{\beta}{\beta-1}\left(\liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n) - R\right).$$

Indeed, we have the following.

$$\begin{aligned}
\Pr\left[\frac{1}{n}\log\frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R\right] &= \Pr\left[e^{-t\log\frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)}} \geq e^{-ntR}\right], \text{ for } t>0 \\
&\leq e^{ntR}\left(\sum_{x^n\in\mathcal{X}^n}[P_{X^n}(x^n)]^{1-t}[P_{\bar{X}^n}(x^n)]^t\right) \qquad (7)\\
&= \exp\left\{-nt\left(\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - R\right)\right\},
\end{aligned}$$

for $0 < t < 1$, where (7) follows by Markov's inequality. Therefore

$$\begin{aligned}
\eta(R) &\geq t\left(\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - R\right)\\
&= \frac{\beta}{\beta-1}\left(\liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n) - R\right), \text{ for } \beta \triangleq \frac{t}{t-1} < 0.
\end{aligned}$$

---

[2] If the source alphabet is continuous, and it admits densities under X^n and $\bar{X}^n$, then the n-dimensional α-divergence is given by

$$D_\alpha(X^n\|\bar{X}^n) \triangleq \frac{1}{\alpha-1}\log\left(\int [f_{X^n}(x^n)]^\alpha[f_{\bar{X}^n}(x^n)]^{1-\alpha}dx^n\right).$$

[3] If $\eta(R) = 0$ for all $R\in\mathbb{R}$, then

$$B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \inf_{R\in\mathbb{R}}\{R+\eta(R)|\eta(R)<r\} = \inf_{R\in\mathbb{R}}\{R\} = -\infty,$$

contradicting that $B_e(r|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is, by definition, an exponent and should be always non-negative.

2. *Converse part*: $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \le \liminf_{n\to\infty} \frac{1}{n} D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)$.

The converse holds trivially if $\liminf_{n\to\infty} \frac{1}{n} D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)$ is infinite. Hence we can assume that $\liminf_{n\to\infty} \frac{1}{n} D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n) < K$, where K is some constant. By the equivalence of conditions (1) and (2), it suffices to show that for any $\delta > 0$ arbitrarily small, there exists $\underline{R} = \underline{R}(\delta) \in \mathbb{R}$ such that

$$\eta(\underline{R}) \le \frac{\beta}{\beta-1}\left(3\delta + \liminf_{n\to\infty} \frac{1}{n} D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n) - \underline{R}\right).$$

Consider the twisted distribution defined as:

$$\begin{aligned} P_{X^n}^{(t)}(x^n) &\triangleq \frac{[P_{\bar{X}^n}(x^n)]^t [P_{X^n}(x^n)]^{1-t}}{\sum_{\hat{x}^n\in\mathcal{X}^n} [P_{\bar{X}^n}(\hat{x}^n)]^t [P_{X^n}(\hat{x}^n)]^{1-t}} \\ &= \exp\left\{t\left[\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + D_{1-t}(X^n\|\bar{X}^n)\right]\right\} P_{X^n}(x^n), \end{aligned} \tag{8}$$

where $t = \beta/(\beta-1)$. Note that $0 < t < 1$. Let $\mathcal{N}$ be a set of positive integers such that

$$\lim_{n\in\mathcal{N}, n\to\infty} \frac{1}{n} D_{1/(1-\beta)}(X^n\|\bar{X}^n) = \liminf_{n\to\infty} \frac{1}{n} D_{1/(1-\beta)}(X^n\|\bar{X}^n), \tag{9}$$

and define

$$\tau \triangleq \sup\{R \in \mathbb{R} : \eta^{(t)}(R) > 0\},$$

where

$$\eta^{(t)}(R) \triangleq \liminf_{n\in\mathcal{N}, n\to\infty} -\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R\right\}$$

is the twisted large deviation spectrum of the normalized log-likelihood ratio with parameter t, and τ satisfies (cf. Lemmas 3 and 4 in Appendix A) that

$$-\infty < \tau \le \lim_{n\in\mathcal{N}, n\to\infty} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n) = \liminf_{n\to\infty} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n) < K.$$

We then note by definition of $\eta^{(t)}(\cdot)$ and the finiteness property of τ that for any $\delta > 0$, there exists $\varepsilon > 0$ such that:

$$\eta^{(t)}(\tau-\delta) = \liminf_{n\in\mathcal{N}, n\to\infty} -\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le \tau-\delta\right\} > \varepsilon > 0.$$

As a result,

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

On the other hand, define

$$\bar{\eta}^{(t)}(R) \triangleq \liminf_{n\in\mathcal{N}, n\to\infty} -\frac{1}{n}\log P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \geq R\right\}$$

and

$$\bar{\tau} \triangleq \inf\{R \in \mathbb{R} : \bar{\eta}^{(t)}(R) > 0\}.$$

Then by noting that

$$\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} = D_{1-t}(X^n\|\bar{X}^n) - \frac{1}{t}\log\frac{P^{(t)}_{X^n}(x^n)}{P_{X^n}(x^n)},$$

we have:

$$\bar{\eta}^{(t)}(R) = \sigma\left(-tR + \frac{t}{n}D_{1-t}(X^n\|\bar{X}^n)\right)$$

and

$$\begin{aligned}
\bar{\tau} &= -\frac{1}{t}\sup\{R \in \mathbb{R} : \sigma(R) > 0\} + \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) \\
&\leq \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) && (10)\\
&< K \quad \text{for } n \in \mathcal{N} \text{ sufficiently large,} && (11)
\end{aligned}$$

where

$$\sigma(R) \triangleq \liminf_{n\in\mathcal{N}, n\to\infty} -\frac{1}{n}\log P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P^{(t)}_{X^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\},$$

(10) follows from Lemma 5 in Appendix A, and (11) holds by definition of K. This indicates the existence of $\bar{\varepsilon} > 0$ such that $\bar{\eta}^{(t)}(K) > \bar{\varepsilon}$, which immediately gives that for $n \in \mathcal{N}$ sufficiently large,

$$P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \geq K\right\} \leq e^{-n\bar{\varepsilon}}.$$

Therefore, for $n \in \mathcal{N}$ sufficiently large,

$$\begin{aligned}
& P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : K > \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \tau - \delta\right\} \\
=\ & P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \tau - \delta\right\} \\
& - P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \geq K\right\} \\
\geq\ & 1 - e^{-n\varepsilon} - e^{-n\bar{\varepsilon}}. && (12)
\end{aligned}$$

Let $I_1 \triangleq (\tau - \delta, b_1)$, and

$$I_k \triangleq [b_{k-1}, b_k) \quad \text{for} \quad 2 \le k \le L \triangleq \left\lceil \frac{K - \tau + \delta}{2\delta} \right\rceil,$$

where $b_k \triangleq (\tau - \delta) + 2k\delta$ for $1 \le k < L$, and $b_L \triangleq K$. By (12), there exists $1 \le k(n) \le L$ such that

$$P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\} \ge \frac{1 - e^{-n\varepsilon} - e^{-n\bar{\varepsilon}}}{L}, \tag{13}$$

for $n \in \mathcal{N}$ sufficiently large. Then, by letting $R_1 \triangleq \limsup_{n \in \mathcal{N}, n \to \infty} b_{k(n)} + \delta$, we obtain that for $n \in \mathcal{N}$ sufficiently large

$$P_{X^n} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R_1 \right\} \ge P_{X^n} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\}.$$

However, for sufficiently large $n \in \mathcal{N}$, we have that

$$\begin{aligned}
& P_{X^n} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\} \\
= & \sum_{\left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\}} P_{X^n}(x^n) \\
= & \sum_{\left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\}} e^{-t\left(\log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + D_{1-t}(X^n \| \bar{X}^n) \right)} P_{X^n}^{(t)}(x^n) && (14) \\
> & e^{-nt\left(-b_{k(n)-1} + \frac{1}{n} D_{1-t}(X^n \| \bar{X}^n) \right)} \sum_{\left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\}} P_{X^n}^{(t)}(x^n) \\
= & e^{-nt\left(-b_{k(n)-1} + \frac{1}{n} D_{1-t}(X^n \| \bar{X}^n) \right)} P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \in I_{k(n)} \right\} \\
\ge & \frac{1 - e^{-n\varepsilon} - e^{-n\bar{\varepsilon}}}{L} e^{-nt\left(-b_{k(n)-1} + \frac{1}{n} D_{1-t}(X^n \| \bar{X}^n) \right)}, && (15)
\end{aligned}$$

where (14) follows from (8), and (15) follows from (13). Consequently

$$\begin{aligned}
\eta(R_1) &= \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R_1\right\} \\
&\le \liminf_{n\in\mathcal{N}, n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R_1\right\} \\
&\le t\left(-\limsup_{n\in\mathcal{N}, n\to\infty} b_{k(n)-1} + \liminf_{n\in\mathcal{N}, n\to\infty} \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)\right) \\
&\le t\left(-\limsup_{n\in\mathcal{N}, n\to\infty} b_{k(n)} + 2\delta + \liminf_{n\to\infty} \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)\right) \\
&= t\left(3\delta + \liminf_{n\to\infty} \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - R_1\right).
\end{aligned}$$

Since δ can be made arbitrarily small, the proof is completed. □

Observations:

A. While the proof of the forward part is straightforward, the proof of the converse part is considerably more complex. The objective of the converse part is to demonstrate that if $\liminf_{n\to\infty} \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)$ is slightly shifted to the right (by a factor of 3δ), then there exists a coordinate $\underline{R}$ such that a straight line of slope $\beta/(1-\beta)$ given by

$$y = \frac{\beta}{\beta-1}\left(3\delta + \liminf_{n\to\infty} \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - R\right)$$

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

This proof is established by observing that the desired coordinate $\underline{R}$ lies in a small neighborhood of τ, where τ is the smallest point for which $\eta^{(t)}(R)$ vanishes. A key point is to choose the twisted parameter t to be equal to $\beta/(\beta-1)$ which is the negative slope of the support line to $\eta(R)$. We graphically illustrate this observation (based on a true example involving binary memoryless or i.i.d. sources) in Figure 2. The computational details are described in Example 1 (cf. Section 4).

B. Note also that the proof holds if the alphabet is *countable* or *continuous* as opposed to the source coding forward and reverse β-cutoff rates results [4] where the finiteness property of the alphabet is necessary. The modifications in the proof for the continuous case are clear. Simply, replace the probability mass function by the proper probability function and the summation by integration. We graphically illustrate this observation (based on a true example involving memoryless Gaussian sources) in Figure 3. The computational details are described in Example 2 (cf. Section 4).

C. The proof of the hypothesis testing forward β-cutoff rate is more involved than the proof of the source coding forward β-cutoff rate result given in [4]. The main difficulty arises from the formula in Proposition 1 where the infimum for R is taken over the entire real line contrary to Proposition 1 in [4] for source coding where R ranges from 0 to ∞. This requires us to deal separately with the degenerate case $\tau = -\infty$ (cf. Lemma 4 in Appendix A). Also, the technique used to prove the forward β-cutoff rate for hypothesis testing relies on the proofs of *both* the source coding forward and reverse β-cutoff rates, but in major parts though similar to the reverse source coding β-cutoff rate.

D. If the sources $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ are arbitrary (not necessarily stationary, irreducible) time-invariant finite-alphabet Markov sources of arbitrary order, then we know that the α-divergence rate exists and can be computed [17, 18]. Thus in this case, the forward β-cutoff rate for testing between Markov sources can be obtained. Also, from the definition of $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, it follows directly that for all $E > 0$,

$$D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \geq \sup_{\beta<0} \left[\beta(E - R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}))\right].$$

Note that this convex lower bound is *computable* for the entire class of Markov sources, while $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is not necessarily computable in general (it is computable for irreducible Markov sources [1], [15], see Figure 4). We graphically illustrate this observation for testing between irreducible Markov sources in Figure 4 and arbitrary Markov sources (not necessarily stationary, irreducible) in Figure 5. The computational details are described in Examples 3 and 4 (cf. Section 4).

4 Examples for the Forward β-cutoff rate

Throughout this section, the natural logarithm is used.

Example 1 *Finite-alphabet memoryless sources:* Naturally, in this case, the forward cutoff rate expression is readily known as Theorem 1 reduces to Csiszár's result [6]. But this example is useful to illustrate the proof method discussed in Observation A of the previous section since $\eta(R)$ and $\eta^{(t)}(R)$ can be explicitly derived.

Consider the binary hypothesis testing between two memoryless (i.i.d.) sources $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ and $\bar{\boldsymbol{X}} = \{\bar{X}_i\}_{i=1}^{\infty}$ under the distributions $(1/2, 1/2)$ and $(1/4, 3/4)$ respectively. Then the log-likelihood ratio $Z = \log \frac{P_X(X)}{P_{\bar{X}}(X)}$ has the following distribution:

$$\Pr\{Z = \log(2)\} = 1 - \Pr\{Z = \log(2/3)\} = 1/2.$$

By Cramer's Theorem [3, p. 9], we get that

$$\eta(R) = \begin{cases} \infty, & R < -\log(3/2) \\ \log(2), & R = -\log(3/2) \\ \frac{\log(\log(3/2)+R)-\log(\log(2)-R)}{\log(3)} R & \\ +\frac{\log(3/2)}{\log(3)} \log(\log(3/2)+R) & \\ +\frac{\log(2)}{\log(3)} \log(\log(2)-R) & \\ +\log(2) - \log(\log(3)), & -\log(3/2) < R < E[Z] = \log(2) - \log(3)/2 \\ 0, & \text{otherwise,} \end{cases}$$

where $E[Z]$ denotes the expectation of the random variable Z. Let R' be the rate at which the line of slope $\beta/(1-\beta)$ is tangent to $\eta(R)$. By straightforward calculations, we get that

$$R' = \log 2 - \frac{\log 3}{1 + 3^{\frac{\beta}{1-\beta}}},$$

and that the forward β-cutoff rate, $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, which is the R-axis intercept of the tangent line of slope $\beta/(1-\beta)$ to $\eta(R)$, is given by

$$R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \frac{2\beta - 1}{\beta} \log 2 - \frac{\beta - 1}{\beta} \log\left(1 + 3^{\frac{\beta}{1-\beta}}\right) - \log 3.$$

On the other hand, the α-divergence between $\boldsymbol{X}$ under $\bar{\boldsymbol{X}}$ is given by

$$D_\alpha(X\|\bar{X}) = \frac{1}{\alpha - 1}\left((\alpha - 2)\log 2 + \log(1 + 3^{1-\alpha})\right),$$

which yields

$$D_{\frac{1}{1-\beta}}(X\|\bar{X}) = \frac{2\beta - 1}{\beta} \log 2 - \frac{\beta - 1}{\beta} \log\left(1 + 3^{\frac{\beta}{1-\beta}}\right) - \log 3.$$

Note that the forward β-cutoff rate, $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, and the $\liminf$ α-divergence rate (which is equal to the α-divergence since the sources are memoryless) of order $\alpha = 1/(1-\beta)$ are equal as expected from Theorem 1. Let us now derive τ in order to check that $\tau = R'$. First, we need to compute $\eta^{(t)}(R)$. The set $\mathcal{N}$ is equal to the set of natural numbers in this case. Note that the distribution of the random variable $Z^{(t)}$ under the twisted distribution with parameter $0 < t < 1$ is given by

$$P^{(t)}\{Z = \log 2\} = 1/(1+3^t) \quad \text{and} \quad P^{(t)}\{Z = \log(2/3)\} = 3^t/(1+3^t).$$

By Cramer's Theorem [3, p. 9], we get that

$$\eta^{(t)}(R) = \begin{cases} \infty, & R < -\log(3/2) \\ \log(1+3^t), & R = -\log(3/2) \\ t(R - \log 2) & \\ +\frac{\log(\log(3/2)+R) - \log(\log(2)-R)}{\log(3)} R & \\ +\frac{\log(3/2)}{\log(3)} \log(\log(3/2) + R) & \\ +\frac{\log(2)}{\log(3)} \log(\log(2) - R) & \\ +\log(1+3^t) - \log(\log(3)), & -\log(3/2) < R < E[Z^{(t)}] = \frac{\log 2}{1+3^t} + \log(2/3)\frac{3^t}{1+3^t} \\ 0, & \text{otherwise,} \end{cases}$$

where $E[Z^{(t)}]$ denotes the expectation of the random variable $Z^{(t)}$. Therefore

$$\tau = \frac{\log 2}{1+3^t} + \log(2/3)\frac{3^t}{1+3^t}.$$

It is easy to check that indeed we have $\tau = R'$ when the twisted parameter t is chosen to be $\beta/(\beta - 1)$. This example is illustrated in Figure 2 for $\beta = -7$.

Example 2 *Continuous alphabet memoryless sources:* Consider the hypothesis testing problem between two memoryless sources $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ and $\bar{\boldsymbol{X}} = \{\bar{X}_i\}_{i=1}^{\infty}$ under the Gaussian distributions $N(\nu, 1)$ and $N(-\nu, 1)$ respectively, where $N(a, b)$ represents a Gaussian distribution with mean a and variance b. It is easy to check that the log-likelihood ratio $Z = \log \frac{f_X(X)}{f_{\bar{X}}(X)}$ is Gaussian distributed with mean $2\nu^2$ and variance $4\nu^2$, which gives that the moment generating function of Z is $E[e^{\theta Z}] = e^{2\nu^2\theta + 2\nu^2\theta^2}$. By Cramer's Theorem, we get that

$$\eta(R) = \begin{cases} \frac{1}{8\nu^2}(R - 2\nu^2)^2, & R < 2\nu^2 \\ 0, & \text{otherwise.} \end{cases}$$

Let R' be the rate at which the line of slope $\beta/(1-\beta)$ is tangent to $\eta(R)$. We have that

$$R' = 2\nu^2 \frac{1+\beta}{1-\beta}.$$

Thus, the forward β-cutoff rate, $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, which is the R-axis intercept of the tangent line of slope $\beta/(1-\beta)$ to $\eta(R)$, is given by

$$R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 2\nu^2\frac{1}{1-\beta}.$$

On the other hand, the α-divergence between $\boldsymbol{X}$ under $\bar{\boldsymbol{X}}$ is given by $D_\alpha(X\|\bar{X}) = 2\nu^2\alpha$, which yields

$$D_{\frac{1}{1-\beta}}(X\|\bar{X}) = 2\nu^2\frac{1}{1-\beta}.$$

Note that the forward β-cutoff rate, $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, and the lim inf α-divergence rate (which is equal to the α-divergence since the sources are memoryless) of order $\alpha = 1/(1-\beta)$ are equal as expected from Theorem 1.

Now, let us compute $\eta^{(t)}(R)$. The set $\mathcal{N}$ in this case is equal to the set of natural numbers. For some normalization constant C,

$$\begin{aligned} P_{X^n}^{(t)}(x^n) &= C\cdot\exp\left\{-\frac{t}{2}\sum_{i=1}^n(x_i+\nu)^2\right\}\exp\left\{-\frac{1-t}{2}\sum_{i=1}^n(x_i-\nu)^2\right\} \\ &= C\cdot\exp\left\{-\frac{1}{2}\sum_{i=1}^n[t(x_i+\nu)^2+(1-t)(x_i-\nu)^2]\right\} \\ &= C\cdot\exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i^2+2(2t-1)\nu x_i+\nu^2)\right\}, \end{aligned}$$

which is a Gaussian distribution with mean $(1-2t)\nu$ and unit variance. Similarly, by invoking Cramer's Theorem, we get that,

$$\eta^{(t)}(R) = \begin{cases} \frac{1}{8\nu^2}(R+(2t-1)2\nu^2)^2, & R<(1-2t)2\nu^2 \\ 0, & \text{otherwise.} \end{cases}$$

Hence, $\tau = (1-2t)2\nu^2$. It is straightforward to check that $\tau = R'$ when the twisted parameter t is chosen to be $\beta/(\beta-1)$. This example is depicted in Figure 3 for $\beta = -0.5$.

Example 3 *Irreducible finite-alphabet Markov sources:* Suppose that $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ are two irreducible Markov sources with arbitrary initial distributions and probability transition matrices P and Q defined as follows:

$$P = \begin{pmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{pmatrix}, \qquad Q = \begin{pmatrix} 1/5 & 4/5 \\ 5/6 & 1/6 \end{pmatrix}.$$

Define a new matrix $R = (r_{ij})$ by

$$r_{ij} = p_{ij}^{\alpha} q_{ij}^{1-\alpha}, \quad i,j = 0,1.$$

The α-divergence rate between $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ exists and is given by

$$\lim_{n\to\infty} \frac{1}{n} D_\alpha(X^n \| \bar{X}^n) = \frac{1}{\alpha - 1} \log \lambda,$$

where λ is the largest positive real eigenvalue of R [17], [18]. Hence the computation of the convex lower bound for $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is easily obtained as shown in Figure 4 for the values $\beta = -5, -3, -2, -4/3, -1, -2/3, -1/2, -2/5$ (proceeding from left to right), where $\alpha = \frac{1}{1-\beta}$. Note that in this case the convex lower bound is tight [1], [15].

Example 4 *Arbitrary finite-alphabet Markov sources:* Suppose that $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ are two arbitrary Markov sources with arbitrary initial distributions and probability transition matrices P and Q defined as follows:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \\ 0 & 0 & 3/5 & 2/5 & 0 \\ 0 & 1/6 & 5/6 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/2 \end{pmatrix}, \qquad Q = \begin{pmatrix} 1/5 & 4/5 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 1/6 & 5/6 & 0 & 0 \\ 1/8 & 0 & 1/2 & 0 & 3/8 \end{pmatrix}.$$

Define a new matrix $R = (r_{ij})$ by

$$r_{ij} = p_{ij}^{\alpha} q_{ij}^{1-\alpha}, \quad i,j = 0,1,2,3,4.$$

The α-divergence rate between $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ can be computed [17], [18]. Hence, the convex lower bound for $D_e(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ can be easily derived as shown in Figure 5 for the values $\beta = -5, -3, -2, -1, -2/3, -1/2, -2/5, -1/6$ (proceeding from left to right), where $\alpha = \frac{1}{1-\beta}$.

5 Hypothesis Testing Correct Exponent and Problem Formulation

In [6], Csiszár investigated the hypothesis testing problem between i.i.d. observations by considering the β-cutoff rate for the exponent of the best correct probability of type 1 with exponential constraint on the probability of type 2 error. More formally, he used the following definitions.

Definition 4 **[6]** Fix $E > 0$. A rate r is called E-unachievable if there exists a sequence of acceptance regions $\mathcal{A}_n$ such that

$$\limsup_{n\to\infty} -\frac{1}{n}\log(1-\mu_n) \le r \quad \text{and} \quad \liminf_{n\to\infty} -\frac{1}{n}\log\lambda_n \ge E,$$

where μ_n and λ_n are type 1 and type 2 error probabilities respectively. The infimum of all E-unachievable rates is defined as:

$$D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \stackrel{\triangle}{=} \inf\{r > 0 : r \text{ is } E\text{-unachievable}\},$$

and $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \infty$ if the above set is empty.

For $0 < r < D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, every acceptable region $\mathcal{A}_n$ with $\liminf_{n\to\infty} -\frac{1}{n}\log\lambda_n \ge E$ satisfies $\mu_n > 1 - e^{-nr}$ for n infinitely often.

Definition 5 **[6]** Fix $\beta > 0$. $R_0 \ge 0$ is a reverse β-achievable rate for the general hypothesis testing problem if

$$D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \ge \beta(E - R_0)$$

for every $E > 0$. The reverse β-cutoff rate is defined as the infimum of all reverse β-achievable rates, and is denoted by $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$.

However, in [14, Section 7], Nagaoka and Hayashi investigated (using Han's definitions [8]) the general hypothesis testing problem between arbitrary sources with memory by considering the exponent of the best correct probability of type 2 with exponential constraint on the probability of type 1 error. More formally, they used the following definition.

Definition 6 **[8, 14]** Fix $r > 0$. A rate E is called r-unachievable if there exists a sequence of acceptance regions $\mathcal{A}_n$ such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log\mu_n \ge r \quad \text{and} \quad \limsup_{n\to\infty} -\frac{1}{n}\log(1-\lambda_n) \le E.$$

The infimum of all r-unachievable rates is denoted by $B_e^*(r|\boldsymbol{X}\|\bar{\boldsymbol{X}})$:

$$B_e^*(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \stackrel{\triangle}{=} \inf\{E > 0 : E \text{ is } r\text{-unachievable}\},$$

and $B_e^*(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \infty$ if the above set is empty.

Proposition 2 [14, Thm. 14, Eq. (56)] Fix $r > 0$. For the general hypothesis testing problem, we have that

$$B_e^*(r|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = r + \inf\{R \in \mathbb{R} \ : \ \bar{\rho}(R) \le r\},$$

where

$$\bar{\rho}(R) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log P_{\bar{X}^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R\right\} - R.$$

Remark 1: Note that Csiszár's and Nagaoka/Hayashi's definitions seem different at first glance. In our investigation, we realized that in order to establish our results on the reverse β-cutoff rate for general sources with memory, a formula for the reliability function of the type 1 probability of correct decoding (or type 1 correct exponent) $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, is needed. However, in [14], Nagaoka and Hayashi provided a formula for the reliability function of the type 2 probability of correct decoding, $B_e^*(r|\boldsymbol{X}\|\bar{\boldsymbol{X}})$. This turned out to be an obstacle, since we were not able to derive the reverse β-cutoff rate formula by directly using the formula for $B_e^*(r|\boldsymbol{X}\|\bar{\boldsymbol{X}})$. To overcome this obstacle, we observed that if we interchange the role of the null and alternative hypotheses distributions (i.e., $\boldsymbol{X} \leftrightarrow \bar{\boldsymbol{X}}$), and also r with E (i.e., $r \leftrightarrow E$) in Nagaoka/Hayashi's definition (Definition 6), then a formula for $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ can be readily obtained from Nagaoka and Hayashi's result. More specifically, we have the following.

Definition 7 Fix $E > 0$. A rate r is called E-unachievable if there exists a sequence of acceptance regions $\mathcal{A}'_n = \mathcal{A}_n^c$ (complement of $\mathcal{A}_n$) such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log\lambda_n \ge E \quad \text{and} \quad \limsup_{n\to\infty} -\frac{1}{n}\log(1-\mu_n) \le r,$$

where

$$\lambda_n = \Pr\{\bar{X}^n \notin \mathcal{A}'_n\} = \Pr\{\bar{X}^n \in \mathcal{A}_n\} \quad \text{and} \quad \mu_n = \Pr\{X^n \in \mathcal{A}'_n\} = \Pr\{X^n \notin \mathcal{A}_n\}.$$

The infimum of all E-unachievable rates is given by

$$B_e^*(E|\bar{\boldsymbol{X}}\|\boldsymbol{X}) = \inf\{r > 0 : r \text{ is } E\text{-unachievable}\},$$

and $B_e^*(E|\bar{\boldsymbol{X}}\|\boldsymbol{X}) = \infty$ if the above set is empty.

With Definition 7, Proposition 2 becomes as follows.

Proposition 3 For any $E > 0$,

$$B_e^*(E|\bar{\boldsymbol{X}}\|\boldsymbol{X}) = E + \inf\{R \in \mathbb{R} \ : \ \rho(R) \leq E\},$$

where

$$\rho(R) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\} - R.$$

Remark 2: We can now clearly observe that Definitions 7 and 4 are identical. This indicates that Nagaoka and Hayashi's $B_e^*(E|\bar{\boldsymbol{X}}\|\boldsymbol{X})$ is in fact Csiszár's $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$. Hence, using Definition 4, Proposition 3 should be as follows.

Proposition 4 For any $E > 0$,

$$D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = E + \inf\{R \in \mathbb{R} \ : \ \rho(R) \leq E\},$$

where

$$\rho(R) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\} - R.$$

The above proposition is a key ingredient for our main results in the following section.

6 Reverse β-Cutoff Rate for Arbitrary Sources

In the degenerate case where $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 0$, we have that $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \infty$. Similarly, if $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \infty$, then $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 0$. A graphical illustration of $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is given in Figure 6. We first show the following lemma, which will provide us the key mechanism to establish our reverse β-cutoff rate result.

Lemma 2 Fix $t < 0$. The following two conditions are equivalent.

$$(\forall\ R \in \mathbb{R}) \quad \rho(R) \geq -R(1-t) + tR_0 \tag{16}$$

and

$$(\forall\ E > 0) \quad D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \geq \frac{t}{t-1}(E - R_0). \tag{17}$$

Proof:

a) (16)⇒(17).

$$\begin{aligned} D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) &= E + \inf\{R \in \mathbb{R} : \rho(R) \leq E\} \\ &\geq E + \inf\{R \in \mathbb{R} : -R(1-t) + tR_0 \leq E\} \\ &= \frac{t}{t-1}\left(E - R_0\right), \end{aligned}$$

where the inequality follows from (16).

b) (17)⇒(16).

$$\inf\{R \in \mathbb{R} : \rho(R) \leq E\} \geq \frac{t}{t-1}(E - R_0) - E = \frac{1}{t-1}E - \frac{t}{t-1}R_0.$$

Thus

$$E \leq \rho\left(\frac{1}{t-1}E - \frac{t}{t-1}R_0\right),$$

since by [14, Remark 15], $\rho(\cdot)$ is strictly decreasing. Letting

$$R = \frac{1}{t-1}E - \frac{t}{t-1}R_0,$$

or

$$E = -R(1-t) + tR_0,$$

the above inequality can be rewritten as

$$\rho(R) \geq -R(1-t) + tR_0,$$

where $R \in \mathbb{R}$. □

We next employ Lemma 2 to show our main result regarding the reverse β-cutoff rate.

Theorem 2 (Reverse $\boldsymbol{\beta}$-cutoff rate formula). The following hold for the general hypothesis testing problem.

1. $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \leq \liminf_{n\to\infty} \frac{1}{n} D_{1/(1-\beta)}(X^n\|\bar{X}^n)$ for $0 < \beta < 1$.

2. If $\rho(R)$ admits a limit,

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \geq \liminf_{n\to\infty} \frac{1}{n} D_{1/(1-\beta)}(X^n\|\bar{X}^n) \quad \text{for } 0 < \beta < \beta_{\max},$$

where

$$\beta_{\max} \triangleq \sup\left\{\gamma \in (0,1) : \liminf_{n\to\infty} \frac{1}{n} D_{1/(1-\gamma)}(X^n\|\bar{X}^n) < \infty\right\}.$$

Note that from the above two inequalities, we directly obtain that for $0 < \beta < \beta_{\max}$,

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

if $\rho(R)$ admits a limit.

Proof:[4]

1. *Forward part:* $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \leq \liminf_{n\to\infty}(1/n)D_{1/(1-\beta)}(X^n\|\bar{X}^n)$ for $0 < \beta < 1$.

By the equivalence of conditions (16) and (17), it suffices to show that

$$(\forall R \in \mathbb{R}) \;\; \rho(R) \geq -R(1-t) + t \cdot \liminf_{n\to\infty} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n),$$

for $t = -\beta/(1-\beta) < 0$.

Consider the twisted distribution defined as:

$$\begin{aligned} P_{X^n}^{(t)}(x^n) &\triangleq \frac{[P_{\bar{X}}(x^n)]^t [P_{X^n}(x^n)]^{1-t}}{\sum_{\hat{x}^n \in \mathcal{X}^n} [P_{\bar{X}}(\hat{x}^n)]^t [P_{X^n}(\hat{x}^n)]^{1-t}} \\ &= \exp\left\{t \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + tD_{1-t}(X^n\|\bar{X}^n)\right\} P_{X^n}(x^n). \end{aligned} \tag{18}$$

Then for $t < 0$,

$$\begin{aligned} & P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\} \\ = & \sum_{\left\{x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\}} P_{X^n}(x^n) \\ = & \sum_{\left\{x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\}} \exp\left\{-t \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} - tD_{1-t}(X^n\|\bar{X}^n)\right\} P_{X^n}^{(t)}(x^n) \\ \leq & \exp\left\{-tnR - tD_{1-t}(X^n\|\bar{X}^n)\right\} \sum_{\left\{x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \leq R\right\}} P_{X^n}^{(t)}(x^n) \\ \leq & \exp\left\{-tnR - tD_{1-t}(X^n\|\bar{X}^n)\right\}. \end{aligned}$$

[4]For the proof of the continuous alphabet case, the same remark given in Observation B of Section 3 applies.

So,

$$\begin{aligned}
\rho(R) &= \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} - R \\
&\ge tR + t\cdot \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - R \\
&= -R(1-t) + t\cdot\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n). \qquad (19)
\end{aligned}$$

2. *Converse part:* $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \ge \liminf_{n\to\infty}\frac{1}{n}D_{1/(1-\beta)}(X^n\|\bar{X}^n)$ for $0<\beta<\beta_{\max}$.

By the equivalence of (16) and (17), it suffices to show the existence of R_1 for any $\delta > 0$ such that

$$\rho(R_1) \le -R_1(1-t) + t\left(\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - \delta\right), \qquad (20)$$

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

$$\zeta(R) \triangleq \rho(R) + R = \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n\in\mathcal{X}^n \ : \ \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\}.$$

Then (20) is equivalent to:

$$\zeta(R_1) = \rho(R_1) + R_1 \le tR_1 + t\left(\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - \delta\right),$$

Now, define

$$\lambda \triangleq \sup\{R\in\mathbb{R} : \zeta^{(t)}(R) > 0\},$$

where

$$\zeta^{(t)}(R) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\}$$

is the twisted large deviation spectrum of the normalized log-likelihood ratio with parameter t. It can be shown that λ satisfies $-\infty < \lambda \le 0$ (cf. Lemmas 6 and 7 in Appendix B).

By $\zeta^{(t)}(\lambda - \delta/2) > \varepsilon$ for some $\varepsilon > 0$, we know that

$$P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le \lambda - \delta/2\right\} \le e^{-n\varepsilon}$$

for infinitely many n. Furthermore, $\zeta^{(t)}(\lambda+\delta/2) = 0$ implies that for any $0 < \bar{\varepsilon} < \varepsilon$,

$$P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le \lambda + \delta/2\right\} \ge e^{-n\bar{\varepsilon}} \qquad (21)$$

for all sufficiently large n. Therefore,

$$P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \lambda - \delta/2 < \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le \lambda + \delta/2\right\} \ge e^{-n\bar{\varepsilon}} - e^{-n\varepsilon}$$

for infinitely many n, or equivalently,

$$-\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}\le-\frac{1}{n}\log\left[e^{-n\bar{\varepsilon}}-e^{-n\varepsilon}\right],$$

for infinitely many n, which (by the fact that $a_n\le b_n$ for infinitely many n implies that $\liminf_{n\to\infty}a_n\le\limsup_{n\to\infty}b_n$ for any two sequences $\{a_n\}$ and $\{b_n\}$) immediately gives that

$$\begin{aligned}
&\liminf_{n\to\infty}-\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}\\
\le\;&\limsup_{n\to\infty}-\frac{1}{n}\log\left[e^{-n\bar{\varepsilon}}-e^{-n\varepsilon}\right]=\bar{\varepsilon}.
\end{aligned}$$

As $\bar{\varepsilon}$ can be made arbitrarily small, we obtain that

$$\liminf_{n\to\infty}-\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}=0. \tag{22}$$

Then, by letting $R_1\triangleq\lambda+\delta/2$, we obtain that

$$\begin{aligned}
&P_{X^n}\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le R_1\right\}\\
\ge\;&P_{X^n}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}\\
=\;&\sum_{\left\{x^n\in\mathcal{X}^n:\lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}}P_{X^n}(x^n)\\
=\;&\sum_{\left\{x^n\in\mathcal{X}^n:\lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}}e^{-tD_{1-t}(X^n\|\bar{X}^n)}e^{-t\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}}P_{X^n}^{(t)}(x^n)\\
\ge\;&e^{-tD_{1-t}(X^n\|\bar{X}^n)}e^{-tn(\lambda-\delta/2)}\sum_{\left\{x^n\in\mathcal{X}^n:\lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}}P_{X^n}^{(t)}(x^n)\\
=\;&e^{-tD_{1-t}(X^n\|\bar{X}^n)}e^{-tn(R_1-\delta)}P_{X^n}^{(t)}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2<\frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}\le\lambda+\delta/2\right\}.
\end{aligned}$$

Consequently,

$$
\begin{aligned}
\zeta(R_1) &\triangleq \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R_1\right\}\\
&= \liminf_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R_1\right\}\\
&\le \liminf_{n\to\infty}\Bigg(t\,\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + t\,(R_1-\delta)\\
&\qquad\qquad -\frac{1}{n}\log P^{(t)}_{X^n}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2 < \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le \lambda+\delta/2\right\}\Bigg)\\
&\le \limsup_{n\to\infty} t\,\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + t\,(R_1-\delta)\\
&\quad + \liminf_{n\to\infty} -\frac{1}{n}\log P^{(t)}_{X^n}\left\{x^n\in\mathcal{X}^n : \lambda-\delta/2 < \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le \lambda+\delta/2\right\}\\
&= t\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + t\,R_1 - \delta t,
\end{aligned}
$$

where the second equality holds since $\rho(R)$ admits a limit, and the last equality follows from (22).

□

Observations:

A. Unlike Theorem 1, where the forward cutoff rate formula holds in full generality, the reverse cutoff rate expression of Theorem 2 requires one assumption that $\rho(R)$ admits a limit. This assumption is not too stringent; for example, it holds for the class of sources with memory that satisfy the Gärtner-Ellis Theorem [3, p. 15]. However, it is important to point out that if this assumption does not hold, the reverse cutoff rate expression formula is not valid in general (see Example 7 below).

B. Remark also that if $\beta_{\max} \le \beta < 1$, it is possible that Part 2 of Theorem 2 may no longer hold (cf. Example 8 below). However, it can be directly verified that for sources (with general alphabet) satisfying the Gärtner-Ellis Theorem, $\beta_{\max} = 1$ and hence Part 2 of Theorem 2 is always valid. Evidently, this is also the case for finite-alphabet sources.

C. The above two observations directly imply that for finite-alphabet i.i.d. sources, Theorem 2 reduces to Csiszár's result [6]; i.e., the reverse β-cutoff rate is given by the Rényi divergence with parameter $1/(1-\beta)$ for all $0 < \beta < 1$.

D. Finally, it directly follows from the definition of $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ and Part 1 of Theorem 2 (which holds without any assumption), that a convex lower bound for the reliability function $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$ can be established in terms of the liminf Rényi divergence rate, as noted in Observation D of Section 3 (with the exception that the supremum is over $0 < \beta < 1$). Hence, for any sources with memory whose liminf Rényi divergence rate is explicitly known (e.g., for general time-invariant finite-alphabet Markov sources [17, 18]), a computable lower bound to $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$ can be determined.

7 Examples for the Reverse β-cutoff rate

We herein provide four examples to illustrate and facilitate the understanding of Theorem 2. In Example 5, we present a case (involving discrete alphabet sources) where $\rho(R)$ admits a limit and $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is equal to the Rényi divergence rate. In Example 6, we illustrate the situation (involving continuous alphabet sources) where the limit of $\rho(R)$ does not exist while $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \liminf_{n\to\infty}(1/n)D_{1/(1-\beta)}(X^n\|\bar{X}^n)$. In Example 7, we construct an example where $\rho(R)$ does not admit a limit and $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \neq \liminf_{n\to\infty}(1/n)D_{1/(1-\beta)}(X^n\|\bar{X}^n)$. This example shows that *without* the assumption on the existence of the limit of $\rho(R)$, Theorem 2 may *not* hold in general (particularly Part 2 of the theorem). In all these examples, the results hold for all values of β (i.e., $\beta_{\max} = 1$). In the last example (Example 8), we depict a case where $\beta_{\max} < 1$. In all examples, we employ the natural logarithm.

Example 5: Let $P_{\bar{X}^n}(a_n) = 1 - e^{-2n}$ and $P_{\bar{X}^n}(b_n) = e^{-2n}$, where $a_n \neq b_n$ and $a_n, b_n \in \mathcal{X}^n$. Also, let $P_{X^n}(a_n) = 1 - e^{-cn}$ and $P_{X^n}(b_n) = e^{-cn}$, where $0 < c < 2$. Then, the log-likelihood ratio, Z_n, is given by

$$Z_n = \log\frac{P_{\bar{X}^n}(X^n)}{P_{X^n}(X^n)} = \begin{cases} \log\dfrac{1-e^{-2n}}{1-e^{-cn}}, & \text{with probability (in } P_{X^n}\text{) } 1-e^{-cn} \\[2ex] -(2-c)n, & \text{with probability (in } P_{X^n}\text{) } e^{-cn}, \end{cases}$$

which implies that

$$\zeta(R) \;\triangleq\; \rho(R) + R = \lim_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{\frac{1}{n}Z_n \leq R\right\} = \begin{cases} 0, & \text{for } R > 0 \\ c, & \text{for } -(2-c) \leq R \leq 0 \\ \infty, & \text{for } R < -(2-c). \end{cases}$$

Let us first compute the α-divergence rate between X^n and $\bar{X}^n$, where $\alpha > 1$. The normalized n-dimensional α-divergence is given by

$$\frac{1}{n}D_\alpha(X^n\|\bar{X}^n) = \frac{1}{n(\alpha-1)}\log\left[(1-e^{-cn})^\alpha(1-e^{-2n})^{1-\alpha} + e^{-cn\alpha}e^{-2n(1-\alpha)}\right].$$

We have the following three cases.

1. $c\alpha+2-2\alpha > 0$. Note that e^{-cn} and e^{-2n} approach 0 as $n \to \infty$ and that $e^{-cn\alpha}e^{-2n(1-\alpha)} = e^{-n(c\alpha+2-2\alpha)}$, which also approaches 0 as $n \to \infty$. Hence, the α-divergence rate is equal to 0 since the argument of the logarithm $\to 1$ as $n \to \infty$.

2. $c\alpha+2-2\alpha < 0$. In this case, since $e^{-n(c\alpha+2-2\alpha)} \to \infty$ as $n \to \infty$, the argument of the logarithm, for large n, is dominated by $e^{-n(c\alpha+2-2\alpha)}$. Hence

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}D_\alpha(X^n\|\bar{X}^n) &= \lim_{n\to\infty} -\frac{n(c\alpha+2-2\alpha)}{n(\alpha-1)}\\ &= \frac{c\alpha+2-2\alpha}{1-\alpha}.\end{aligned}$$

3. $c\alpha+2-2\alpha = 0$. Clearly, the α-divergence rate is equal to 0 in this case.

We next determine $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ using Proposition 4.

$$D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = E + \inf\{R\in\mathbb{R} : \zeta(R) - R \le E\} = \begin{cases} E, & \text{for } 0 < E \le c\\ c, & \text{for } c < E \le 2\\ E-2+c, & \text{for } 2 < E.\end{cases}$$

Now, the reverse β-cutoff rate is the E-axis intercept of the line of slope β passing by the point $(2,c)$ as illustrated in Figure 7. A straightforward calculation yields

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = -\frac{c}{\beta} + 2.$$

For $\alpha = 1/(1-\beta)$, we get that

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \frac{c\alpha+2-2\alpha}{1-\alpha}.$$

Since by definition, $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \ge 0$, we directly obtain that

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \lim_{n\to\infty}\frac{1}{n}D_{1/(1-\beta)}(X^n\|\bar{X}^n) \quad \text{for } 0 < \beta < 1.$$

Note that for this example, $\beta_{\max} = 1$ since the α-divergence rate is always finite.

Example 6: Consider the hypothesis testing problem between two i.i.d. (memoryless) Gaussian sources $\boldsymbol{X} = \{X_n\}_{n=1}^{\infty}$ and $\bar{\boldsymbol{X}} = \{\bar{X}_n\}_{n=1}^{\infty}$ such that for n odd, $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ have distributions $N(-1,1)$ and $N(1,1)$, respectively; for n even, $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ have distributions $N(-2,1)$ and $N(2,1)$, respectively. Then

$$z(x) = \log \frac{dP_{\bar{X}}(x)}{dP_X(x)} = \begin{cases} \log \dfrac{e^{-(x-1)^2/2}}{e^{-(x+1)^2/2}} = 2x, & \text{if } n \text{ is odd} \\ \log \dfrac{e^{-(x-2)^2/2}}{e^{-(x+2)^2/2}} = 4x, & \text{if } n \text{ is even.} \end{cases}$$

Hence, under distribution $P_{\boldsymbol{X}}$, the log-likelihood ratios $Z_1, Z_2, \ldots$ are i.i.d. with $N(-2,4)$ for n odd, and with $N(-8,16)$ for n even. So, it follows using the Gärtner-Ellis Theorem that

$$\begin{aligned} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} &= -\frac{1}{n}\log \Pr\left[\frac{Z_1 + Z_2 + \cdots + Z_n}{n} \le R\right] \\ &\to \begin{cases} \dfrac{(R+2)^2}{8}\mathbf{1}\{R < -2\}, & \text{if } n \text{ is odd} \\ \dfrac{(R+8)^2}{32}\mathbf{1}\{R < -8\}, & \text{if } n \text{ is even,} \end{cases} \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. We then obtain that

$$\rho(R) = \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} - R = \frac{(R+2)^2}{8}\mathbf{1}\{R < -2\} - R.$$

Given the above $\rho(R)$ (note that the limit of $\rho(R)$ does not exist), we can determine the correct exponent $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ using Proposition 4 as follows.

$$D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = E + \inf\{R \in \mathbb{R} : \rho(R) \le E\} = (\sqrt{E} - \sqrt{2})^2\mathbf{1}\{E \ge 2\}.$$

Solving for the E-axis intercept of the support line to $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ of slope β yields the reverse β-cutoff rate:

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \frac{2}{1-\beta}.$$

Finally, a simple calculation gives that

$$\liminf_{n\to\infty} \frac{1}{n} D_{1/(1-\beta)}(X^n\|\bar{X}^n) = \frac{2}{1-\beta},$$

which is finite for $0 < \beta < \beta_{\max} = 1$ (note that limsup Rényi divergence rate equals $8/(1-\beta)$). Hence we obtain that $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is equal to the liminf Rényi divergence rate for all $0 < \beta < 1$.

Example 7: Consider the hypothesis testing problem between sources $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ and $\bar{\boldsymbol{X}} = \{\bar{X}_i\}_{i=1}^{\infty}$, where for odd n, $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ are respectively unit-variance memoryless Gaussian with means $-\nu$ and ν, respectively, and for even n, $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ have respective distributions $P_{X^n}(x_0^n) = e^{-\theta n}$, $P_{X^n}(x_1^n) = 1 - e^{-\theta n}$, and $P_{\bar{X}^n}(x_0^n) = e^{-2\theta n}$, $P_{\bar{X}^n}(x_1^n) = 1 - e^{-2\theta n}$ where $x_0^n \neq x_1^n$. Then for odd n,

$$Z_n(x^n) = \frac{1}{n} \log \frac{dP_{\bar{X}^n}(x^n)}{dP_{X^n}(x^n)} \quad = \quad \frac{2\nu}{n} \sum_{i=1}^{n} x_i,$$

and for even n,

$$Z_n(x^n) = \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \quad = \quad \begin{cases} \dfrac{1}{n} \log \dfrac{e^{-2\theta n}}{e^{-\theta n}} = -\theta, & \text{if } x^n = x_0^n \\ \dfrac{1}{n} \log \dfrac{1 - e^{-2\theta n}}{1 - e^{-\theta n}} = \dfrac{1}{n} \log(1 + e^{-\theta n}), & \text{if } x^n = x_1^n. \end{cases}$$

Hence, under distribution $P_{\boldsymbol{X}}$, the normalized log-likelihood ratio $Z_n(X^n)$ is normal distributed with mean $-2\nu^2$ and variance $4\nu^2/n$ for odd n; for even n,

$$\Pr\left[Z_n(X^n) = -\theta\right] = e^{-\theta n} \text{ and } \Pr\left[Z_n(X^n) = \frac{1}{n} \log\left(1 + e^{-\theta n}\right)\right] = 1 - e^{-\theta n}.$$

In addition, under distribution $P_{\bar{\boldsymbol{X}}}$, the normalized log-likelihood ratio $Z_n(\bar{X}^n)$ is normal distributed with mean $2\nu^2$ and variance $4\nu^2/n$ for odd n; for even n,

$$\Pr\left[Z_n(\bar{X}^n) = -\theta\right] = e^{-2\theta n} \text{ and } \Pr\left[Z_n(\bar{X}^n) = \frac{1}{n} \log\left(1 + e^{-\theta n}\right)\right] = 1 - e^{-2\theta n}.$$

Direct Derivation of the Correct Exponent: It is known from the Neyman-Person Lemma [16] that the log-likelihood test is optimal. Hence, the type 1 and type 2 error probabilities (μ_n and λ_n) can be written as

$$\mu_n = \Pr[Z_n(X^n) > R_n] + (1 - \omega_n) \Pr[Z_n(X^n) = R_n]$$

and

$$\lambda_n = \Pr[Z_n(\bar{X}_n) < R_n] + \omega_n \Pr[Z_n(\bar{X}^n) = R_n],$$

where $\omega_n \in [0, 1]$ is some randomization factor, and R_n is the log-likelihood ratio threshold.

Now the requirement in Definition 4 for $\liminf_{n\to\infty} -(1/n) \log \lambda_n = E > 0$ can be met for

$$R_n = 2\nu^2 - 2\nu\sqrt{2E} \quad \text{and} \quad \omega_n = 0$$

if n is odd, and for

$$\begin{cases} R_n = \frac{1}{n}\log(1+e^{-\theta n}) \text{ and } \omega_n = e^{-En}, & \text{if } 0 < E < 2\theta \\ R_n = \frac{1}{n}\log(1+e^{-\theta n}) \text{ and } \omega_n \le e^{-2\theta n}, & \text{if } E = 2\theta \\ R_n = -\theta \text{ and } \omega_n = e^{-(E-2\theta)n}, & \text{if } E \ge 2\theta, \end{cases}$$

if n is even. Thus if we take $\theta = 18\nu^2$ and apply the Gärtner-Ellis Theorem, we get the following (type 1) correct exponent

$$\begin{aligned} D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) &= \limsup_{n\to\infty} -\frac{1}{n}\log(1-\mu_n) \\ &= \limsup_{n\to\infty} -\frac{1}{n}\log\left(\Pr[Z_n(X^n) < R_n] + \omega_n \Pr[Z_n(X^n) = R_n]\right) \\ &= \max\Big\{(\sqrt{E}-\sqrt{2}\nu)^2 \cdot \mathbf{1}\{E > 2\nu^2\}, \\ &\qquad E\cdot\mathbf{1}\{0 < E \le \theta\} + \theta\cdot\mathbf{1}\{\theta < E \le 2\theta\} + (E-\theta)\cdot\mathbf{1}\{E > 2\theta\}\Big\} \\ &= \begin{cases} E, & \text{if } 0 < E \le 18\nu^2 \\ 18\nu^2, & \text{if } 18\nu^2 < E \le 32\nu^2 \\ (\sqrt{E}-\sqrt{2}\nu)^2, & \text{if } 32\nu^2 < E \le 50\nu^2 \\ E - 18\nu^2, & \text{for } E \ge 50\nu^2. \end{cases} \end{aligned}$$

The Correct Exponent using Nagaoka and Hayashi's Formula: We next verify our above direct derivation of the correct exponent $D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ by applying Nagaoka and Hayashi's formula in Proposition 4.

$$\begin{aligned} \zeta(R) &= \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} \\ &= \limsup_{n\to\infty} -\frac{1}{n}\log \Pr\left[Z_n(X^n) \le R\right] \\ &= \max\left(\frac{(R+2\nu^2)^2}{8\nu^2}\cdot\mathbf{1}\{R < -2\nu^2\},\quad \theta\cdot\mathbf{1}\{-\theta \le R \le 0\} + \infty\cdot\mathbf{1}\{R < -\theta\}\right) \\ &= \begin{cases} 0, & \text{for } R > 0 \\ \theta, & \text{for } -2\nu^2 - 2\nu\sqrt{2\theta} \le R \le 0 \\ \dfrac{(R+2\nu^2)^2}{8\nu^2}, & \text{for } -\theta \le R < -2\nu^2 - 2\nu\sqrt{2\theta} \\ \infty, & \text{for } R < -\theta. \end{cases} \end{aligned}$$

Setting $\theta = 18\nu^2$, the following expression for $\rho(R)$ is obtained.

$$\rho(R) \stackrel{\triangle}{=} \zeta(R) - R = \begin{cases} -R, & \text{for } R > 0 \\ 18\nu^2 - R, & \text{for } -14\nu^2 \le R \le 0 \\ \dfrac{(R-2\nu^2)^2}{8\nu^2}, & \text{for } -18\nu^2 \le R < -14\nu^2 \\ \infty, & \text{for } R < -18\nu^2. \end{cases}$$

Hence, by Proposition 4,

$$\begin{aligned} D_e^*(E|\boldsymbol{X}\|\bar{\boldsymbol{X}}) &= E + \inf\{R \in \mathbb{R} : \zeta(R) - R \le E\} \\ &= \begin{cases} E, & \text{for } 0 < E < 18\nu^2 \\ 18\nu^2, & \text{for } 18\nu^2 \le E < 32\nu^2 \\ (\sqrt{E} - \sqrt{2}\nu)^2, & \text{for } 32\nu^2 \le E \le 50\nu^2 \\ E - 18\nu^2, & \text{for } E > 50\nu^2, \end{cases} \end{aligned}$$

and indeed our direct derivation coincides with Nagaoka and Hayashi's expression.

Examination of the Reverse Cutoff Rate: By Lemma 2, the β-reverse cutoff rate $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$ is the smallest R_0 satisfying $\zeta(R) = \rho(R) + R \ge t(R + R_0)$, where $t = \beta/(\beta - 1)$ for $0 < \beta < 1$. We thus get that

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \begin{cases} 0, & \text{for } -\dfrac{9}{7} \le t < 0 \\ \dfrac{7t+9}{t} \cdot 2\nu^2, & \text{for } -3 \le t < -\dfrac{9}{7} \\ (1-t) \cdot 2\nu^2, & \text{for } -4 \le t < -3 \\ \dfrac{9t+16}{t} \cdot 2\nu^2, & \text{for } t < -4. \end{cases}$$

We next compute the liminf $(1-t)$-divergence rate (where $\beta = t/(t-1)$) to compare it with $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$.

$$\begin{aligned} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n) &= -\frac{1}{nt} \log \left(\int_{x^n \in \mathcal{X}^n} e^{t \log[dP_{\bar{X}^n}(x^n)/dP_{X^n}(x^n)]} \, dP_{X^n}(x^n) \right) \\ &= -\frac{1}{nt} \log E\left[e^{ntZ_n(X^n)} \right] \\ &= \begin{cases} 2(1-t)\nu^2, & \text{for } n \text{ odd} \\ -\dfrac{1}{nt} \log\left(e^{-(t+1)\theta n} + (1 + e^{-\theta n})^t (1 - e^{-\theta n}) \right), & \text{for } n \text{ even.} \end{cases} \end{aligned}$$

Since

$$\lim_{n\to\infty} -\frac{1}{nt}\log\left(e^{-(t+1)\theta n} + (1+e^{-\theta n})^t(1-e^{-\theta n})\right) \;=\; \theta\,\frac{t+1}{t}\,\mathbf{1}\{t<-1\},$$

we obtain (with $\theta = 18\nu^2$) that

$$\begin{aligned}
\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) &= \min\left\{2(1-t)\nu^2, \frac{t+1}{t}\cdot 18\nu^2\cdot\mathbf{1}\{t<-1\}\right\} \\
&= \begin{cases} 0, & \text{for } t\ge -1 \\ \dfrac{t+1}{t}\cdot 18\nu^2, & \text{for } -4+\sqrt{7}\le t<-1 \\ (1-t)\cdot 2\nu^2, & \text{for } -4-\sqrt{7}\le t<-4+\sqrt{7} \\ \dfrac{t+1}{t}\cdot 18\nu^2, & \text{for } t<-4-\sqrt{7}. \end{cases}
\end{aligned}$$

Note from this example, where $\rho(R)$ does not admit a limit, that

$$R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \le \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)$$

for $t<0$ (i.e., Part 1 of Theorem 2 is valid), while Part 2 of Theorem 2 does *not* hold (e.g., for $t=-8$, $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 14\nu^2$ while $\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) = 15.75\nu^2$). Hence if the limit of $\rho(R)$ does not exist, $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \ne \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)$ in general.

Example 8: In this example, we illustrate the situation where $\beta_{\max} = 1/2 < 1$ and show that if $\beta_{\max} < \beta < 1$, Part 2 of Theorem 2 may not hold.

Consider the hypothesis testing problem between two sources $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ and $\bar{\boldsymbol{X}} = \{\bar{X}_i\}_{i=1}^\infty$. Both $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ have a point mass on $\tilde{x}^n$, and have a density over the hyper-line $a\cdot\hat{x}^n = (a\cdot\hat{x}_1, a\cdot\hat{x}_2, \ldots, a\cdot\hat{x}_n)$ for $a\ge 0$, where $\tilde{x}^n \notin \{x^n\in\mathbb{R}^n : x^n = a\cdot\hat{x}^n$ for some $a\ge 0\}$ and $\hat{x}^n \ne \mathbf{0}$ is fixed. The densities for $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ over $\{x^n\in\mathbb{R}^n : x^n = a\cdot\hat{x}^n$ for some $a\ge 0\}$ are described by $p_{X^n}(a\cdot\hat{x}^n) = \theta n e^{-(a+1)\theta n}$, and $p_{\bar{X}^n}(a\cdot\hat{x}^n) = 2\theta n e^{-2(a+1)\theta n}$, respectively. Furthermore, $P_{X^n}(\tilde{x}^n) = 1-e^{-\theta n}$, and $P_{\bar{X}^n}(\tilde{x}^n) = 1-e^{-2\theta n}$. Then under distribution $P_{\boldsymbol{X}}$,

$$\begin{aligned}
Z_n(a\cdot\hat{x}^n) &= \frac{1}{n}\log\frac{p_{\bar{X}^n}(a\cdot\hat{x}^n)}{p_{X^n}(a\cdot\hat{x}^n)} = \frac{1}{n}\log\frac{2\theta n e^{-2(a+1)\theta n}}{\theta n e^{-(a+1)\theta n}} \\
&= \frac{1}{n}\log(2) - (a+1)\theta \quad \text{ with density } \; \theta n e^{-(a+1)\theta n}\cdot\mathbf{1}\{a\ge 0\},
\end{aligned}$$

and

$$Z_n(\tilde{x}^n) = \frac{1}{n}\log\frac{P_{\bar{X}^n}(\tilde{x}^n)}{P_{X^n}(\tilde{x}^n)} = \frac{1}{n}\log\frac{1-e^{-2\theta n}}{1-e^{-\theta n}} = \frac{1}{n}\log(1+e^{-\theta n}) \text{ with prob. } 1-e^{-\theta n}.$$

Then by the Gärtner-Ellis Theorem,

$$
\begin{aligned}
\zeta(R) &= \limsup_{n\to\infty} -\frac{1}{n}\log P_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} \\
&= \limsup_{n\to\infty} -\frac{1}{n}\log \Pr\left[Z_n(X^n) \le R\right] \\
&= \begin{cases} 0, & R > 0 \\ \theta, & \text{for } -\theta < R \le 0 \\ \limsup_{n\to\infty} -\frac{1}{n}\log \int_{\log(2)/(n\theta)-R/\theta-1}^{\infty} \theta n e^{-\theta n(a+1)} da, & \text{for } R \le -\theta \end{cases} \\
&= \theta \cdot \mathbf{1}\{-\theta < R \le 0\} - R \cdot \mathbf{1}\{R \le -\theta\},
\end{aligned}
$$

Note that the limit of $\rho(R)$ (or $\zeta(R)$) does exit. A direct application of Lemma 2 yields that $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = 0$ for $-1 \le t < 0$, and that $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) = \infty$ for $t < -1$ (where $t = \beta/(\beta-1)$). We next compute the liminf $(1-t)$-divergence rate, where $\beta = t/(t-1)$.

$$
\begin{aligned}
\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) &= -\frac{1}{nt}\log\left(\int_{x^n\in\mathcal{X}^n} e^{t\log[dP_{\bar{X}^n}(x^n)/dP_{X^n}(x^n)]}\, dP_{X^n}(x^n)\right) \\
&= -\frac{1}{nt}\log E\left[e^{ntZ_n(X^n)}\right] \\
&= -\frac{1}{nt}\log\left[\int_0^{\infty} e^{nt((1/n)\log(2)-(a+1)\theta)}\cdot\theta n e^{-(a+1)\theta n}\, da + (1+e^{-\theta n})^t(1-e^{-\theta n})\right] \\
&= -\frac{1}{nt}\log\left[\theta n 2^t e^{-(t+1)\theta n}\int_0^{\infty} e^{-a(t+1)\theta n}\, da + (1+e^{-\theta n})^t(1-e^{-\theta n})\right] \\
&\to 0\cdot\mathbf{1}\{-1 < t < 0\} + \infty\cdot\mathbf{1}\{t \le -1\}.
\end{aligned}
$$

Therefore, $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) \le \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)$ for all $t < 0$ (i.e., $0 < \beta < 1$) and Part 1 of Theorem 2 is satisfied. However, $R_0^{(r)} \ge \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n)$ only for $-1 < t < 0$; in other words, Part 2 of Theorem 2 is only valid for $0 < \beta < \beta_{\max} = 1/2$. Notably, for $t = -1$,

$$
0 = R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}}) < \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) = \infty.
$$

8 Conclusions

We examined the forward and reverse β-cutoff rates for the hypothesis testing problem between arbitrary sources with memory (not necessarily Markovian, ergodic, stationary, etc.) and arbitrary alphabet (countable or continuous). We showed that the forward β-cutoff rate is given by the $\liminf$ Rényi α-divergence rate, where $\alpha = \frac{1}{1-\beta}$ and $\beta < 0$. We also demonstrated that if the large deviation spectrum $\rho(R)$ admits a limit, then the reverse β-cutoff rate is given by the lim inf α-divergence rate, where $\alpha = \frac{1}{1-\beta}$ and $0 < \beta < \beta_{\max}$. For $\beta_{\max} \le \beta < 1$, we provided an upper bound on the reverse β-cutoff rate. As expected, for finite-alphabets i.i.d. sources, our theorems reduce to Csiszár's result [6]. Our forward and reverse β-cutoff rates results and the methods used to derive them were illustrated via several examples. An interesting consequence of this work is that while the formula of the type 1 error (respectively correct) exponent is not explicitly known for general sources with memory, our forward (respectively reverse) β-cutoff rate result, expressed in terms of the liminf Rényi divergence rate, can be used to provide a useful convex lower bound (which is computable in principle) to the error (resp. correct) exponent.

Future work may include the study of Csiszár's channel coding β-cutoff rates [6] for arbitrary discrete channels with memory using our information spectrum techniques.

Appendix A: Properties of τ and $\sigma(R)$

Lemma 3 For $0 < t < 1$,

$$\tau \triangleq \sup\{R : \eta^{(t)}(R) > 0\} \le \liminf_{n\to\infty} \frac{1}{n} D_{1-t}(X^n \| \bar{X}^n).$$

Proof: For any $\nu > 0$,

$$\begin{aligned} & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + 2\nu\right\} \\ \le\ & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + \nu\right\} \text{ for } n \in \mathcal{N} \text{ sufficiently large,} \end{aligned}$$

where $\mathcal{N}$ is defined in (9). But

$$\begin{aligned} & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + \nu\right\} \\ =\ & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\left(\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + D_{1-t}(X^n\|\bar{X}^n)\right) > \nu\right\} \\ =\ & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{t}{n}\left(\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + D_{1-t}(X^n\|\bar{X}^n)\right) < -\nu t\right\} \\ =\ & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}^{(t)}(x^n)}{P_{X^n}(x^n)} < -\nu t\right\} \qquad (23) \\ =\ & P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : P_{X^n}^{(t)}(x^n) < e^{-n\nu t}P_{X^n}(x^n)\right\} \\ \le\ & e^{-n\nu t}P_{X^n}\left\{x^n \in \mathcal{X}^n : P_{X^n}^{(t)}(x^n) < e^{-n\nu t}P_{X^n}(x^n)\right\} \\ \le\ & e^{-n\nu t}, \end{aligned}$$

where (23) follows from (8). Thus for $n \in \mathcal{N}$ sufficiently large,

$$P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + 2\nu\right\} \ge 1 - e^{-n\nu t},$$

which implies

$$\begin{aligned}
&\eta^{(t)}\left(\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + 2\nu\right)\\
=\ &\liminf_{n\in\mathcal{N},n\to\infty} -\frac{1}{n}\log P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + 2\nu\right\}\\
\le\ &\limsup_{n\in\mathcal{N},n\to\infty} -\frac{1}{n}\log\left(1 - e^{-n\nu t}\right) = 0.
\end{aligned}$$

Consequently,

$$\sup\left\{R : \eta^{(t)}(R) > 0\right\} \le \liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + 2\nu.$$

The proof is completed by noting that ν can be made arbitrarily small. □

Lemma 4 For $0 < t < 1$, if $\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) < K$, then

$$\tau \triangleq \sup\{R : \eta^{(t)}(R) > 0\} > -\infty.$$

Proof: By (8), we get that

$$P_{X^n}^{(t)}(x^n) \quad = \quad e^{tD_{1-t}(X^n\|\bar{X}^n)}e^{(1-t)\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)}}P_{\bar{X}^n}(x^n).$$

Hence,

$$\begin{aligned}
&P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R\right\}\\
\le\ &e^{tD_{1-t}(X^n\|\bar{X}^n)}e^{(1-t)nR}P_{\bar{X}^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \le R\right\}\\
\le\ &e^{tD_{1-t}(X^n\|\bar{X}^n)}e^{(1-t)nR},
\end{aligned}$$

which implies that

$$\eta^{(t)}(R) \ge -t\limsup_{n\in\mathcal{N},n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) - (1-t)R.$$

Therefore,

$$\tau \ge -\frac{t}{1-t}\limsup_{n\in\mathcal{N},n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n).$$

This shows that $\tau = -\infty$ implies that

$$\limsup_{n\in\mathcal{N},n\to\infty} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n) = \lim_{n\in\mathcal{N},n\to\infty} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n) = \liminf_{n\to\infty} \frac{1}{n} D_{1-t}(X^n\|\bar{X}^n) = \infty,$$

contradicting the assumption that $\liminf_{n\to\infty}(1/n)D_{1-t}(X^n\|\bar{X}^n) < K$. $\square$

Lemma 5 We have the following:

$$\sup\{R \in \mathbb{R} : \sigma(R) > 0\} \geq 0.$$

Proof: For any $\nu > 0$,

$$\begin{aligned}
P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{X^n}^{(t)}(x^n)}{P_{X^n}(x^n)} \leq -\nu\right\} &= P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : P_{X^n}^{(t)}(x^n) \leq e^{-n\nu}P_{X^n}(x^n)\right\} \\
&\leq e^{-n\nu}P_{X^n}\left\{x^n \in \mathcal{X}^n : P_{X^n}^{(t)}(x^n) \leq e^{-n\nu}P_{X^n}(x^n)\right\} \\
&\leq e^{-n\nu},
\end{aligned}$$

which implies $\sigma(-\nu) \geq \nu$. Hence, the lemma holds. $\square$

Appendix B: Properties of λ

Lemma 6 For $t < 0$, $\lambda \leq 0$.

Proof: Observe that for $R > 0$,

$$\begin{aligned}
& P_{X^n}^{(t)}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} > R\right\} \\
\leq\ & e^{-nR(1-t)+tD_{1-t}(X^n\|\bar{X}^n)}P_{\bar{X}^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} > R\right\} \\
\leq\ & e^{-nR(1-t)+tD_{1-t}(X^n\|\bar{X}^n)} \\
\leq\ & e^{-nR(1-t)},
\end{aligned}$$

where the last inequality follows from the non-negativity of $D_{1-t}(X^n\|\bar{X}^n)$ and the negativity of t. This implies that for $R > 0$,

$$\zeta^{(t)}(R) \leq \limsup_{n\to\infty} -\frac{1}{n}\log\left(1 - e^{-nR(1-t)}\right) = 0,$$

which immediately implies that $\lambda \leq 0$. $\square$

Lemma 7 For $0 > t > t_{\min} \triangleq \beta_{\max}/(\beta_{\max}-1)$, $\lambda > -\infty$.

Proof: If $\lambda = -\infty$, then $\zeta^{(t)}(R) = 0$ for every $R \in \mathbb{R}$. Hence, by choosing any $v = (t - t_{\min})/2$, we have:

$$\begin{aligned}
& P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} \\
\le\ & e^{tD_{1-t}(X^n\|\bar{X}^n)-(t-v)D_{1-(t-v)}(X^n\|\bar{X}^n)+vnR}\, P^{(t-v)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} \\
\le\ & e^{-(t-v)D_{1-(t-v)}(X^n\|\bar{X}^n)+vnR},
\end{aligned}$$

which implies that

$$\begin{aligned}
0 = \zeta^{(t)}(R) &\ge \limsup_{n\to\infty}\left[(t-v)\frac{1}{n}D_{1-(t-v)}(X^n\|\bar{X}^n)\right] - vR \\
&= (t-v)\liminf_{n\to\infty}\frac{1}{n}D_{1-(t-v)}(X^n\|\bar{X}^n) - vR.
\end{aligned}$$

This indicates that

$$\liminf_{n\to\infty}\frac{1}{n}D_{1-(t-v)}(X^n\|\bar{X}^n) \ge \frac{v}{t-v}R \quad \text{for every } R \in \mathbb{R},$$

or equivalently,

$$\liminf_{n\to\infty}\frac{1}{n}D_{1-(t-v)}(X^n\|\bar{X}^n) = \infty.$$

A contradiction to the definition of $t_{\min}$ is obtained, since by Lyapounov's inequality [2, Eq. (5.37)], for $0 > \alpha > \gamma$,

$$\begin{aligned}
\exp\left\{-\alpha D_{1-\alpha}(X^n\|\bar{X}^n)\right\} &= \left(\sum_{x^n\in\mathcal{X}^n}\left(\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)}\right)^{1-\alpha} P_{\bar{X}^n}(x^n)\right)^{1/(1-\alpha)} \\
&\le \left(\sum_{x^n\in\mathcal{X}^n}\left(\frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)}\right)^{1-\gamma} P_{\bar{X}^n}(x^n)\right)^{1/(1-\gamma)} \\
&= \exp\left\{-\gamma D_{1-\gamma}(X^n\|\bar{X}^n)\right\},
\end{aligned}$$

or equivalently,

$$|\alpha| D_{1-\alpha}(X^n\|\bar{X}^n) \le |\gamma| D_{1-\gamma}(X^n\|\bar{X}^n),$$

and hence by definition of $t_{\min}$,

$$\liminf_{n\to\infty}\frac{1}{n}D_{1-(t-v)}(X^n\|\bar{X}^n) < \infty.$$

□

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments; in particular, they are grateful to one of the reviewers for informing them about [14] whose results were instrumental in significantly improving Theorem 2.

References

[1] V. Anantharam, "A large deviations approach to error exponents in source coding and hypothesis testing," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 938–943, July 1990.

[2] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, New Work, 1995.

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

[4] P.-N. Chen and F. Alajaji, "Csiszár's cutoff rates for arbitrary discrete sources,"*IEEE Transactions on Information Theory*, vol. 47, pp. 330–338, January 2001.

[5] P.-N. Chen, "General formulas for the Neyman-Pearson type 2 error exponent subject to fixed and exponential type 1 error bounds," *IEEE Transactions on Information Theory*, vol. 42, January 1996.

[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 Scientiarum Mathematicarum Hungarica*, vol. 6, pp. 181–191, 1971.

[8] T. S. Han, "Hypothesis testing with the general source," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2415–2427, November 2000.

[9] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer-Verlag, Berlin, 2003.

[10] W. Hoeffding, "Asymptotically optimal test for multinomial distributions," *Annals of Mathematical Statistics*, vol. 36, pp. 369–400, 1965.

[11] K. Iriyama, "Error exponent for hypothesis testing of general sources," submitted for publication, 2002.

[12] F. Jelinek, *Probabilistic Information Theory*, McGraw-Hill, New York, 1968.

[13] L. H. Koopmans, "Asymptotic rate of discrimination for Markov processes," *Annals of Mathematical Statistics*, vol. 31, pp. 982–994, 1960.

[14] H. Nagaoka and M. Hayashi, "An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses," *e-print quant-ph/0206185*, June 2002. Available at *http://arxiv.org/abs/quant-ph/0206185.*

[15] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Transactions on Information Theory*, vol. 31, pp. 360–365, May 1985.

[16] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Royal Soc. London,* Series A, vol. 231, pp. 289-337, 1933.

[17] T. Nemetz, "On the α-divergence rate for Markov-dependent hypotheses," *Problems of Control and Information Theory,* vol. 3, no. 2, pp. 147–155, 1974.

[18] Z. Rached, F. Alajaji and L. L. Campbell, "Rényi's divergence and entropy rates for finite-alphabet Markov sources," *IEEE Transactions on Information Theory*, vol. 47, pp. 1553–1561, May 2001.

[19] A. Rényi, "On measures of entropy and information," *Proc. Fourth Berkeley Symposium on Mathematics and Statistics* **1**, 547–561, University of California Press, Berkeley, 1961.

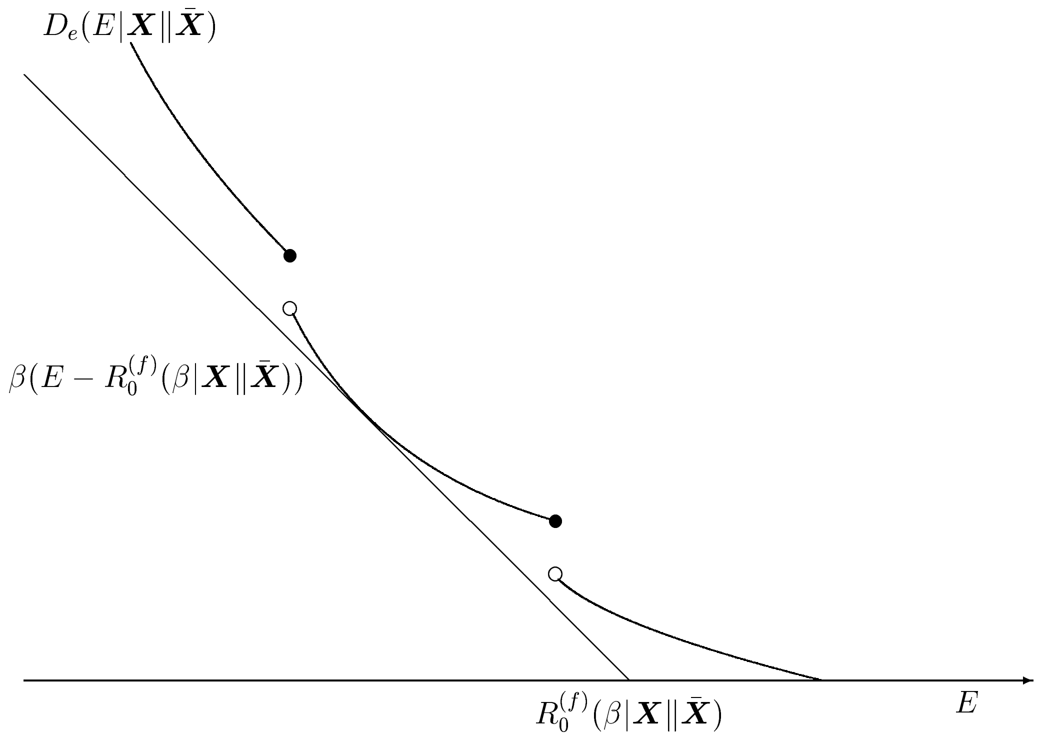

Figure 1: A graphical illustration of the forward β-cutoff rate, $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, for testing between two arbitrary sources $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$.

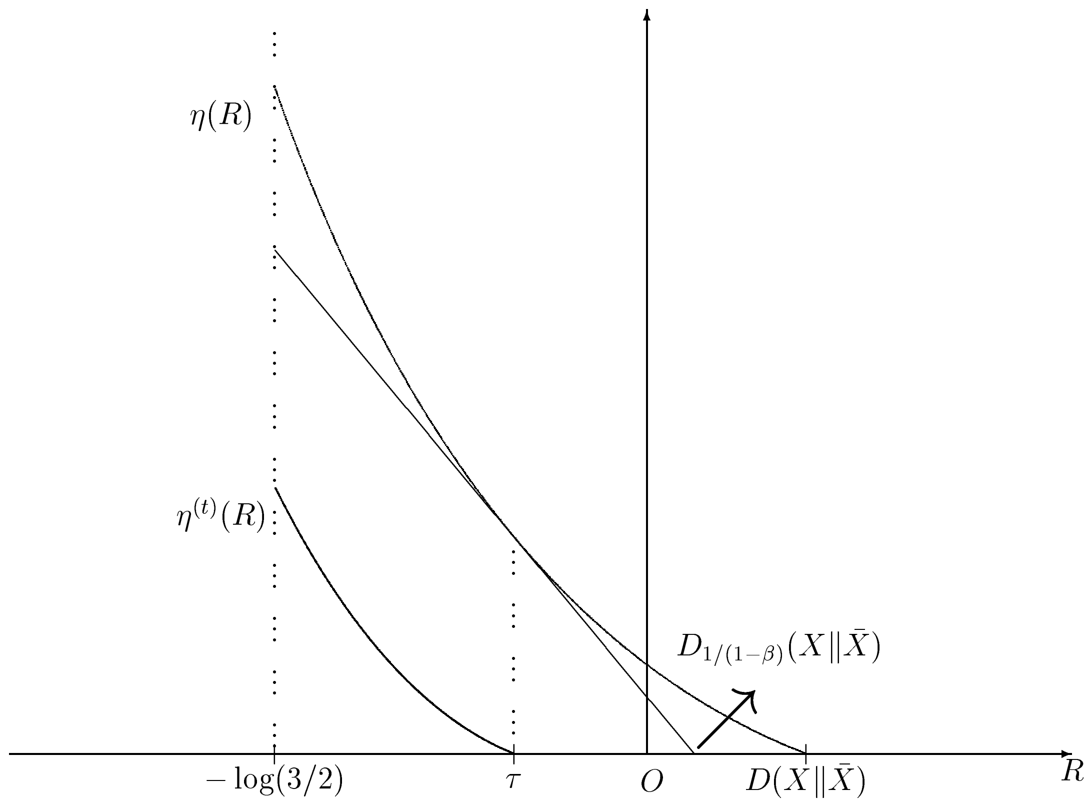

Figure 2: Functions $\eta(R)$, $\eta^{(t)}(R)$ and $(\beta/(\beta-1))\left[\liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)-R\right]$ for testing between two binary memoryless sources $\boldsymbol{X}=\{X_i\}_{i=1}^{\infty}$ and $\bar{\boldsymbol{X}}=\{\bar{X}_i\}_{i=1}^{\infty}$ under the distributions $(1/2,1/2)$ and $(1/4,3/4)$ respectively, and with $\beta=-7$. When $R<-\log(3/2)$, $\eta(R)=\eta^{(t)}(R)=\infty$.

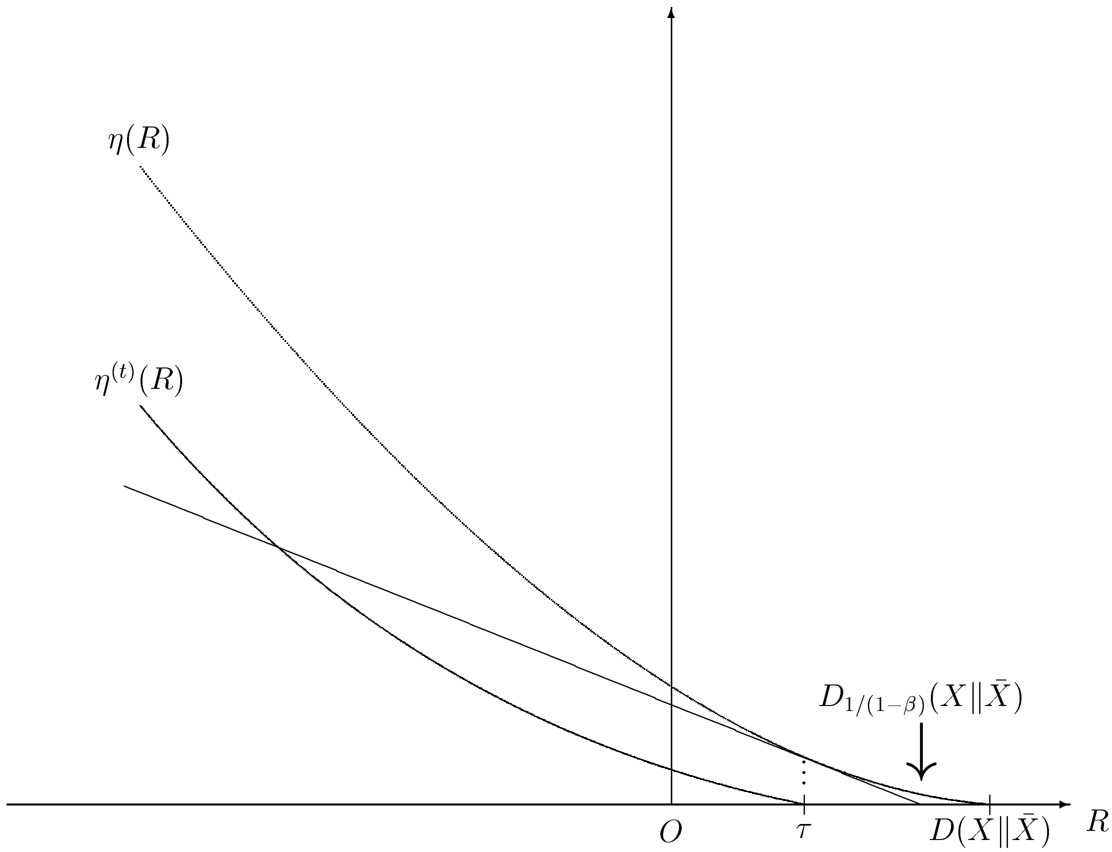

Figure 3: Functions $\eta(R)$, $\eta^{(t)}(R)$ and $(\beta/(\beta-1))\left[\liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)-R\right]$ for testing between two memoryless sources $\boldsymbol{X}=\{X_i\}_{i=1}^{\infty}$ and $\bar{\boldsymbol{X}}=\{\bar{X}_i\}_{i=1}^{\infty}$ under the Gaussian distributions $N(\nu,1)$ and $N(-\nu,1)$ respectively, and with $\beta=-0.5$.

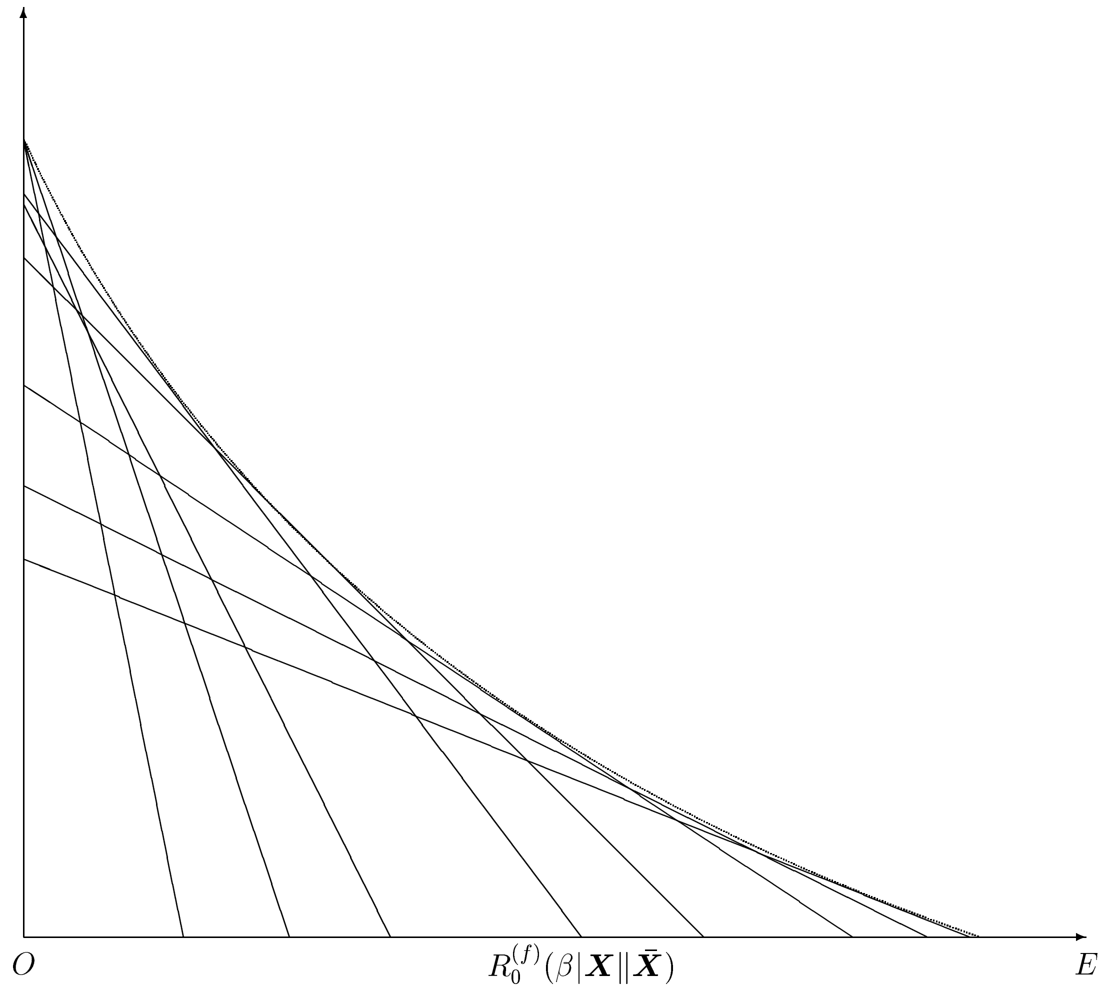

Figure 4: Convex lower bound for testing between irreducible Markov sources. Each line of slope β intersects the E-axis at $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$. Proceeding from left to right, the values of β are: $-5, -3, -2, -4/3, -1, -2/3, -1/2, -2/5$.

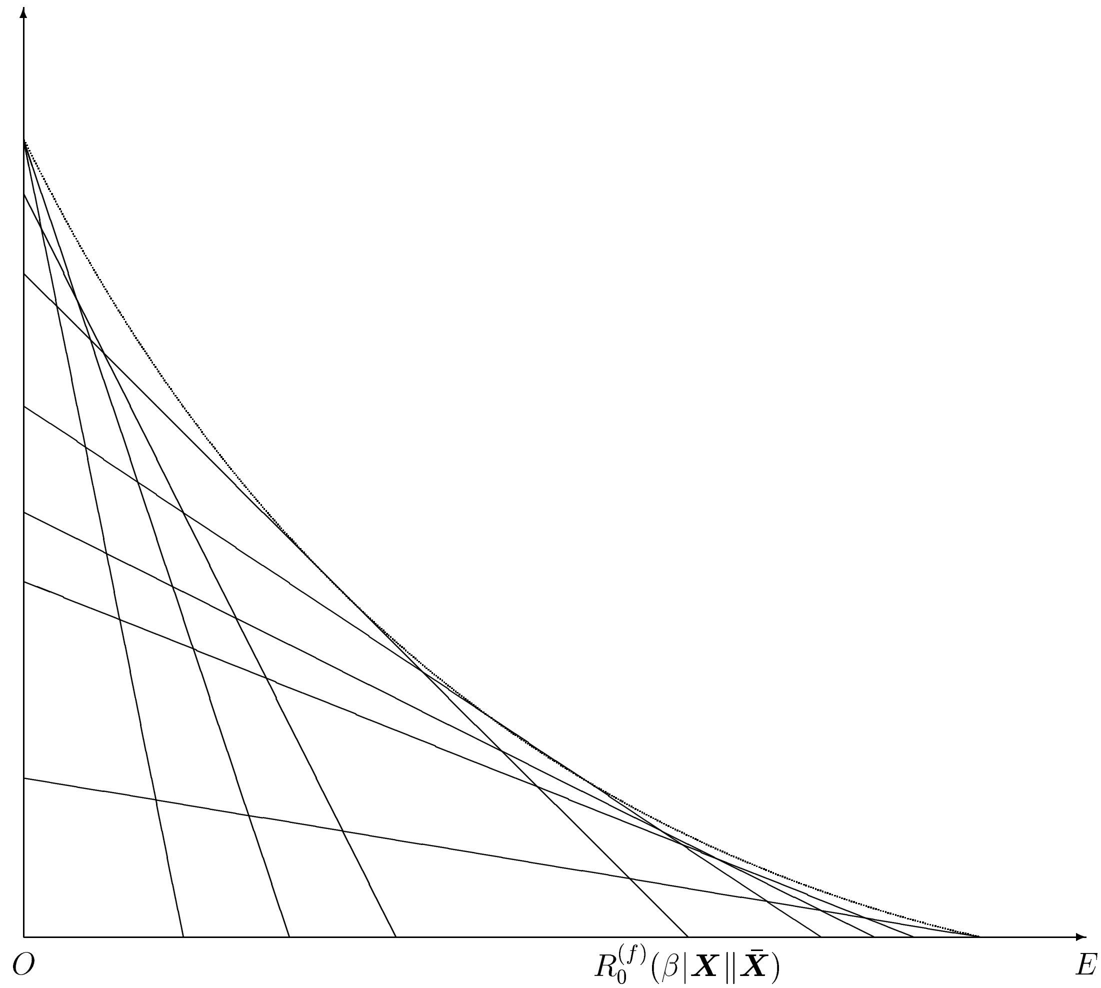

Figure 5: Convex lower bound for testing between arbitrary Markov sources. Each line of slope β intersects the E-axis at $R_0^{(f)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$. Proceeding from left to right, the values of β are: $-5, -3, -2, -1, -2/3, -1/2, -2/5, -1/6$.

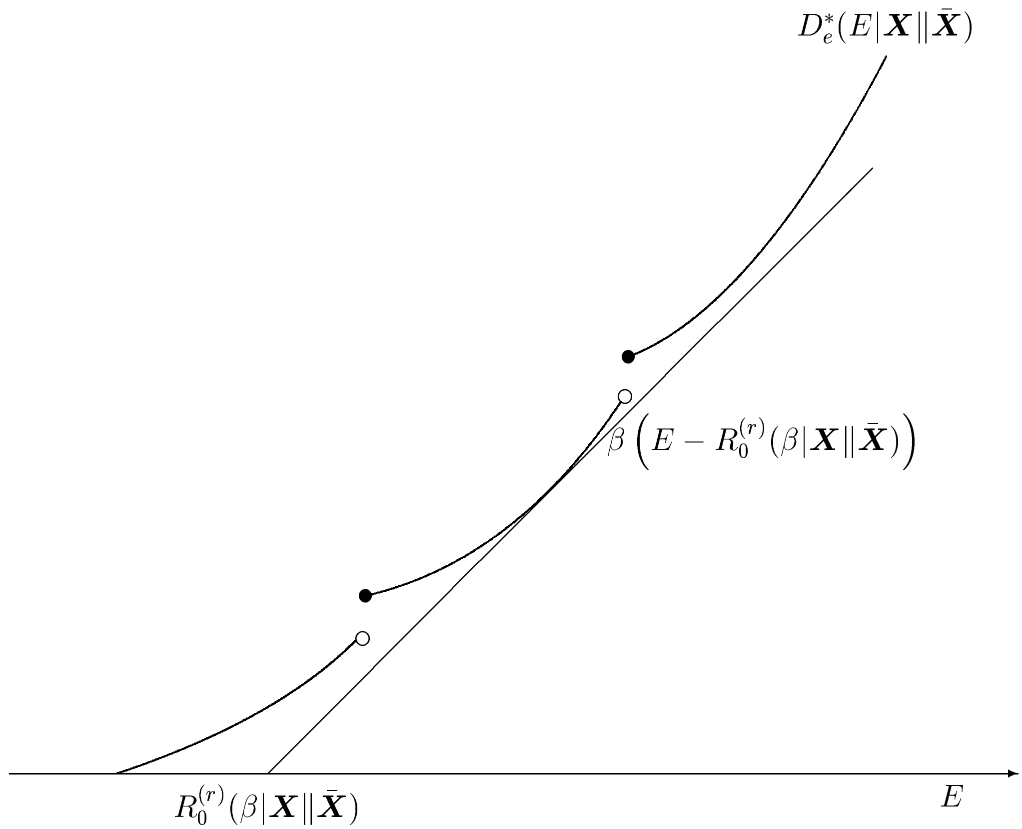

Figure 6: A graphical illustration of the reverse β-cutoff rate, $R_0^{(r)}(\beta|\boldsymbol{X}\|\bar{\boldsymbol{X}})$, for testing between two arbitrary sources $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$.

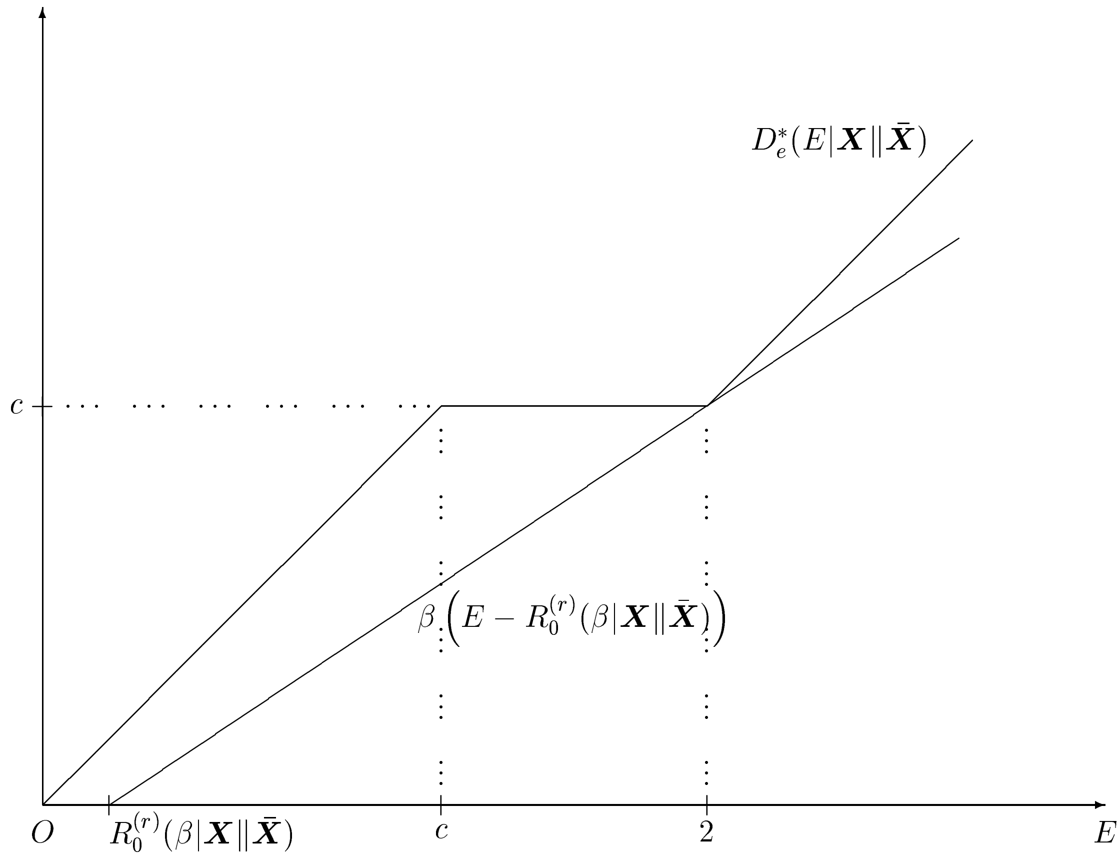

Figure 7: Reliability function of the type 1 probability of correct decoding for testing between the two sources $P_{X^n}(\cdot)$ and $P_{\bar{X}^n}(\cdot)$ as given in Example 5.