Mathematics and Engineering Communications Laboratory

Technical Report

Csiszár's Hypothesis Testing Cutoff Rates for Arbitrary Sources with Memory

F. Alajaji, *P.-N. Chen*, and *Z. Rached*

December 2002

Csiszár's Hypothesis Testing Cutoff Rates for Arbitrary Sources with Memory

Fady Alajaji Po-Ning Chen Ziad Rached *

Abstract

In [5], 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 α-divergence [15], $\alpha > 0$, $\alpha \neq 1$. Similarly, for $\beta > 0$, 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 α-divergence rate provides the expression for the forward β-cutoff rate. Under two conditions on the large deviation spectrum, $\rho(R)$, we show that the reverse β-cutoff rate is given by the 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 $\limsup$ $\frac{1}{1-\beta}$-divergence rate is finite. In particular, we examine i.i.d. observations and sources that satisfy the hypotheses of the Gärtner-Ellis Theorem. Unlike [3] 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 numerical examples to illustrate our forward and reverse β-cutoff rates results.

Index Terms: Hypothesis testing error and correct exponent, forward and reverse β-cutoff rates, information spectrum, α-divergence rate, arbitrary observations with memory.

---

*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.

F. Alajaji and Z. Rached are with the Dept. of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada. P.-N. Chen is with the Dept. of Commun. Eng., National Chiao Tung Univ., Taiwan.

1 Introduction

In [5], 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 the α-divergence, $\alpha > 0$, $\alpha \neq 1$. Similarly, 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 defines the reverse 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 correct exponent (or reliability) function for hypothesis testing where $0 < R_0 < E$ is the rate of exponential convergence to 0 of the type 2 error probability. He then demonstrated that the reverse β-cutoff rate is given by $D_{1/(1-\beta)}(X\|\bar{X})$. 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 [6], [9], [10], and for irreducible Markov sources in [1], [12]. The error exponent for testing between ergodic Markov sources with continuous state-space under certain additional restrictions was established in [11]. In its full generality, i.e., for arbitrary sources (not necessarily stationary, ergodic, etc.), the error exponent was studied in [4], [7], [8].

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 α-divergence rate provides the expression for the forward β-cutoff rate. Our proof relies in part on the formulas established in [7], and extensions of the techniques used in [3] that generalize Csiszár's source coding cutoff rate results for arbitrary discrete sources with memory. Unlike [3] where the source alphabet was assumed to be finite, we assume arbitrary (countable or continuous) source alphabet. The techniques used in our proof are a mixture of the techniques used in deriving the forward and reverse β-cutoff rates for source coding [3]. 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, $\rho(R)$, is convex and if there exists an $R \in \mathbb{R}$ such that $\rho(R) + R = 0$, then the 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 lim sup $\frac{1}{1-\beta}$-divergence rate is finite. For $1 > \beta \geq \beta_{\max}$, we only provide an upper bound to the reverse β-cutoff rate. However, our result does reduce to Csiszár's result for finite-alphabet i.i.d. observations for $0 < \beta < 1$. We also examine sources with memory (countable or continuous alphabet) which satisfy the Gärtner-Ellis Theorem. We show that in this case, the above conditions on $\rho(R)$ are satisfied and that the reverse β-cutoff rate is given by the $\frac{1}{1-\beta}$-divergence rate.

The rest of the paper is organized as follows. In Section 2, we briefly recall previous results by Han [7] 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 illustrate it numerically in Section 4. In Section 5, we recall the general expression for the reliability function of the type 2 probability of correct decoding [7] and formulate the reverse β-cutoff rate problem by carefully examining the inconsistency of definitions in [5] and [7]. In Section 6, we investigate the reverse β-cutoff rate and illustrate it numerically. Finally, in Section 7, we conclude with a summary along with several directions for future work.

2 Hypothesis Testing Error Exponent

Let us first define the general source as an infinite sequence

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

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 $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ and $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^{\infty}$ taking values in the same source alphabet $\{\mathcal{X}^n\}_{n=1}^{\infty}$, we may define the general hypothesis testing problem with $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ as the null hypothesis and $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^{\infty}$ 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 [5] 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 The supremum of all E-achievable rates is denoted by $D_e(E|\mathbf{X}\|\bar{\mathbf{X}})$:

$$D_e(E|\mathbf{X}\|\bar{\mathbf{X}}) \stackrel{\triangle}{=} \sup\{r > 0 : r \text{ is } E\text{-achievable}\},$$

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

$$B_e(r|\mathbf{X}\|\bar{\mathbf{X}}) \stackrel{\triangle}{=} \sup\{E > 0 : E \text{ is } r\text{-achievable}\},$$

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

Proposition 1 [7] Fix $r > 0$. For the general hypothesis testing problem, we have that

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

where[1]

$$\eta(R) \stackrel{\triangle}{=} \liminf_{n\to\infty} -\frac{1}{n}\log Pr\left\{\frac{1}{n}\log\frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R\right\},$$

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 $f_{X^n}(X^n)$.

3 Forward β-Cutoff Rate Between Arbitrary Hypotheses

Definition 3 [5] Fix $\beta < 0$. $R_0 \geq 0$ is a forward β-achievable rate for the general hypothesis testing problem if

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

for every $E > 0$, or equivalently,

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

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

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

Before stating our main result, we first observe in the next lemma that the forward β-cutoff rate $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$ 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

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

Proof:

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

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

Therefore

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

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)

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

Therefore

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

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,

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

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

   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, i.e., it admits a density $f_{X^n}(\cdot)$, 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

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

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

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

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 5 and 6 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

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

and

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

Then by noting that

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

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

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

(10) follows from Lemma 7 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,

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

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

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

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

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

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 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 [3] 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 probability density 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 [3]. 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 [3] for source coding where R ranges from 0 to ∞. This requires us to deal separately with the degenerate case $\tau = -\infty$ (cf. Lemma 6 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 $\mathbf{X}$ and $\bar{\mathbf{X}}$ 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 [13], [14]. Thus in this case, the forward β-cutoff rate for testing between Markov sources can be obtained. Also, from the definition of $D_e(E|\mathbf{X}\|\bar{\mathbf{X}})$, it follows directly that for all $E > 0$,

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

Note that this convex lower bound is *computable* for the entire class of Markov sources, while $D_e(E|\mathbf{X}\|\bar{\mathbf{X}})$ is not necessarily computable in general (it is computable for irreducible Markov sources [1], [12], 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 Numerical Examples for the Forward β-cutoff rate

Throughout this section, the natural logarithm is used.

Example 1 *Finite-alphabet memoryless sources:* Consider the binary hypothesis testing between two memoryless sources $\mathbf{X} = \{X_i\}_{i=1}^{\infty}$ and $\bar{\mathbf{X}} = \{\bar{X}_i\}_{i=1}^{\infty}$ 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 [2, 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, $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$, which is the R-axis intercept of the tangent line of slope $\beta/(1-\beta)$ to $\eta(R)$, is given by

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

On the other hand, the α-divergence between $\mathbf{X}$ under $\bar{\mathbf{X}}$ 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, $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$, and the $\liminf$ α-divergence rate (which is equal to the α-divergence since the sources are DMS) 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 [2, 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 $\mathbf{X} = \{X_i\}_{i=1}^{\infty}$ and $\bar{\mathbf{X}} = \{\bar{X}_i\}_{i=1}^{\infty}$ 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, $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$, which is the R-axis intercept of the tangent line of slope $\beta/(1-\beta)$ to $\eta(R)$, is given by

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

On the other hand, the α-divergence between $\mathbf{X}$ under $\bar{\mathbf{X}}$ 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, $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$, and the lim inf α-divergence rate (which is equal to the α-divergence since the sources are DMS) 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 $\mathbf{X}$ and $\bar{\mathbf{X}}$ 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 $\mathbf{X}$ and $\bar{\mathbf{X}}$ 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 [13], [14]. Hence the computation of the convex lower bound for $D_e(E|\mathbf{X}\|\bar{\mathbf{X}})$ 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], [12].

Example 4 *Arbitrary finite-alphabet Markov sources:* Suppose that $\mathbf{X}$ and $\bar{\mathbf{X}}$ 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 $\mathbf{X}$ and $\bar{\mathbf{X}}$ can be computed [13], [14]. Hence, the convex lower bound for $D_e(E|\mathbf{X}\|\bar{\mathbf{X}})$ 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 [5], 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 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:

$$D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) \triangleq \inf\{r > 0 : r \text{ is } E\text{-unachievable}\},$$

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

For $0 < r < D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$, 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 Fix $\beta > 0$. $R_0 \ge 0$ is a reverse β-achievable rate for the general hypothesis testing problem if

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

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

However, in [7], Han investigated 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, he used the following definition.

Definition 6 **[7]** 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 $B_e^*(r|\mathbf{X}\|\bar{\mathbf{X}})$:

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

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

Proposition 2 [7] Fix $r > 0$. For the general hypothesis testing problem, we have that

$$B_e^*(r|\mathbf{X}\|\bar{\mathbf{X}}) = \inf_{R\in\mathbb{R}} \{R + \bar{\rho}(R) + [r - \bar{\rho}(R)]^+\},$$

where

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

$[x]^+ = \max\{x, 0\}$, provided the limit in $\bar{\rho}(R)$ exists, and for any $M > 0$, there exists $K > 0$ such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log P_{\bar{X}^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \ge K\right\} \ge M.$$

Remark 1: Note that Csiszár's and Han'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, $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$, is needed. However, in [7], Han provided a formula for the reliability function of the type 2 probability of correct decoding, $B_e^*(r|\mathbf{X}\|\bar{\mathbf{X}})$. 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 $B_e^*(r|\mathbf{X}\|\bar{\mathbf{X}})$. To overcome this obstacle, we observed that if we interchange the role of the null and alternative hypotheses distributions (i.e., $\mathbf{X} \leftrightarrow \bar{\mathbf{X}}$), and also r with E (i.e., $r \leftrightarrow E$) in Han's definition (Definitions 6), then a formula for $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$ can be readily obtained from Han'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

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

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

With Definition 7, Proposition 2 becomes as follows.

Proposition 3 For any $E > 0$,

$$B_e^*(E|\bar{\mathbf{X}}\|\mathbf{X}) = \inf_{R\in\mathbb{R}} \left\{R + \rho(R) + [E - \rho(R)]^+\right\},$$

where

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

provided the limit in $\rho(R)$ exists, and for any $M > 0$, there exists $K > 0$ such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log P_{X^n} \left\{x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \geq K\right\} \geq M.$$

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

Proposition 4 For any $E > 0$,

$$D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) = \inf_{R\in\mathbb{R}} \left\{R + \rho(R) + [E - \rho(R)]^+\right\},$$

where

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

provided the limit in $\rho(R)$ exists, and for any $M > 0$, there exists $K > 0$ such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log P_{X^n} \left\{x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \geq K\right\} \geq M. \tag{16}$$

Condition (16) evidently holds in the unrestrictive case where $P_{X^n}(\cdot)$ is absolutely continuous with respect to $P_{\bar{X}^n}(\cdot)$. The above proposition is a key ingredient for our main results in the following section.

6 Reverse β-Cutoff Rate Between Arbitrary Hypotheses

In the degenerate case where $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) = 0$, we have that $R_0^{(r)}(\beta|\mathbf{X}\|\bar{\mathbf{X}}) = \infty$. Similarly, if $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) = \infty$, then $R_0^{(r)}(\beta|\mathbf{X}\|\bar{\mathbf{X}}) = 0$. A graphical illustration of $R_0^{(r)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$ is given in Figure 6. Without loss of generality, we herein assume that $P_{X^n}(\cdot)$ is absolutely continuous with respect to $P_{\bar{X}^n}(\cdot)$.

We first show the following lemmas, which will provide us the key mechanism to establish our reverse β-cutoff rate result.

Lemma 2 Assume that the limit in $\rho(R)$ exists. For all $E > 0$, we have that

$$D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) \le E + \inf\{R \in \mathbb{R} : \rho(R) \le E\}.$$

Proof: We have the following.

$$\begin{aligned}
D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) &= \inf_{R\in\mathbb{R}} \left\{R + \rho(R) + [E - \rho(R)]^+\right\} \quad \text{(by Proposition 4)} \\
&= \min\left\{\inf_{\rho(R)\le E}\{R + E\}, \inf_{\rho(R)>E}\{R + \rho(R)\}\right\} \\
&\le \inf_{\rho(R)\le E}\{R + E\} \\
&= E + \inf\{R \in \mathbb{R} : \rho(R) \le E\}.
\end{aligned}$$

□

Lemma 3 Assume that $\rho(R)$ admits a limit and is convex, and that there exists an R such that $R + \rho(R) = 0$. Then for those E satisfying $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) > 0$,

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

Proof: Since $\rho(R)$ is decreasing by definition and it is assumed to be convex, then it is continuous and strictly decreasing. Let R^* be the smallest one that satisfies $R + \rho(R) = 0$. Then for $E \le \rho(R^*)$,

$$\begin{aligned}
D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) &= \inf_{R\in\mathbb{R}} \left\{R + \rho(R) + [E - \rho(R)]^+\right\} \quad \text{(by Proposition 4)} \\
&\le R^* + \rho(R^*) + [E - \rho(R^*)]^+ = 0.
\end{aligned}$$

Hence, the set of values of E such that $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) > 0$ does not include $E \leq \rho(R^*)$. Now as $\rho(R)$ is assumed convex, its slope is strictly increasing, which implies that the slope of $\rho(R)$ is less than -1 for $R < R^*$. This immediately gives that the slope of the function $R + \rho(R)$ is negative for $R < R^*$. Consequently, for any $E > \rho(R^*)$ (which corresponds to $R < R^*$ since $\rho(R)$ is strictly decreasing),

$$\begin{aligned} \inf_{\{R:\rho(R)>E\}} \{R + \rho(R)\} &= \{R + \rho(R)\}|_{R=\rho^{-1}(E)} \\ &= \rho^{-1}(E) + E = \inf_{\rho(R)\leq E} \{R + E\}, \end{aligned}$$

where

$$\rho^{-1}(E) \triangleq \inf\{a : \rho(a) \leq E\},$$

is the quantile or inverse of $\rho(\cdot)$. Thus,

$$\begin{aligned} D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) &= \inf_{R\in\mathbb{R}} \left\{R + \rho(R) + [E - \rho(R)]^+\right\} \\ &= \min\left\{\inf_{\rho(R)\leq E}\{R + E\}, \inf_{\rho(R)>E}\{R + \rho(R)\}\right\} \\ &= \inf_{\rho(R)\leq E}\{R + E\} \\ &= E + \inf\{R \in \mathbb{R} : \rho(R) \leq E\}. \end{aligned}$$

$\square$

Lemma 4 Fix $t < 0$. Also, assume that $\rho(R)$ admits a limit and is convex, and that there exists an R such that $R + \rho(R) = 0$. The following two conditions are equivalent.

$$(\forall\ R \in \mathbb{R}) \quad \rho(R) \geq -R(1 - t) + tR_0 \tag{17}$$

and

$$(\forall\ E > 0) \quad D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) \geq \frac{t}{t-1}(E - R_0). \tag{18}$$

Proof:

a) (17)⇒(18). By Lemma 3, for those E satisfying $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) > 0$, we have that

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

where the inequality follows from (17). This implies that

$$\inf\{E > 0 : D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) > 0\} \leq R_0.$$

Hence, for these E satisfying $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) = 0$, the claim also holds since $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$ is increasing.

b) (18)$\Rightarrow$(17). By Lemma 2 and (18), for $E > 0$, we have that

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

Thus

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

since $\rho(\cdot)$ 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 $R \in \mathbb{R}$. $\square$

We next employ Lemma 4 to show our main result regarding the reverse β-cutoff rate.

Theorem 2 (Reverse $\boldsymbol{\beta}$-cutoff rate formula). Assume that $\rho(R)$ admits a limit and is convex, and that there exists an R such that $R + \rho(R) = 0$. For the general hypothesis testing problem,

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

and

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

where

$$\beta_{\max} = \sup\left\{\beta \in (0,1) : \limsup_{n\to\infty} \frac{1}{n} D_{1/(1-\gamma)}(X^n\|\bar{X}^n) < \infty \text{ for every } 0 < \gamma < \beta\right\},$$

and

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

is the n-dimensional Rényi α-divergence. Note that the above two inequalities directly imply that

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

for $0<\beta<\beta_{\max}$.

Proof:[4]

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

By the equivalence of conditions (17) and (18), it suffices to show that

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

Consider the twisted distribution defined as:

$$\begin{aligned}
P_{X^n}^{(t)}(x^n) &\triangleq \frac{[P_{\bar{X}}(x^n)]^t[P_{X^n}(x^n)]^{1-t}}{\sum_{\hat{x}^n\in\mathcal{X}^n}[P_{\bar{X}}(\hat{x}^n)]^t[P_{X^n}(\hat{x}^n)]^{1-t}} \\
&= \exp\left\{(t-1)\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + tD_{1-t}(X^n\|\bar{X}^n)\right\}P_{\bar{X}^n}(x^n). \qquad (19)
\end{aligned}$$

Then for $t<0$,

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

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

So, since $\rho(R)$ admits a limit, we have

$$\begin{aligned}
\rho(R) &= \limsup_{n\to\infty} -\frac{1}{n}\log P_{\bar{X}^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \le R\right\} \\
&\ge -R(1-t) + t\cdot\liminf_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) \\
&= -R(1-t) + t\cdot\liminf_{n\to\infty}\frac{1}{n}D_{1/(1-\beta)}(X^n\|\bar{X}^n), \quad \text{for } \beta \triangleq \frac{t}{t-1} \in (0,1).
\end{aligned}$$

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

By the equivalence of (17) and (18), it suffices to show the existence of $\bar{R}$ for any $\delta > 0$ such that

$$\rho(\bar{R}) \le -\bar{R}(1-t) + t\left(\limsup_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) + \frac{(1-t)}{t}3\delta\right),$$

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

$$\lim_{n\in\mathcal{N}, n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) = \limsup_{n\to\infty}\frac{1}{n}D_{1-t}(X^n\|\bar{X}^n) \tag{20}$$

and define

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

where

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

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 8 and 9 in Appendix B). We then note by definition of $\rho^{(t)}(\cdot)$ and the finiteness property of λ that for any $\delta > 0$, there exists $\epsilon > 0$ such that

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

As a result,

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

On the other hand, define

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

and

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

Then by noting that

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

we have:

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

where

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

and

$$\begin{aligned} \bar{\lambda} &= \frac{1}{t} \sup\{R \in \mathbb{R} : \sigma(R) > 0\} - \frac{1}{n} D_{1-t}(X^n \| \bar{X}^n) \\ &\le 0, \end{aligned} \tag{21}$$

where (21) follows from Lemma 7 in Appendix A, and the non-negativity [5] of the Rényi divergence $D_{1-t}(X^n \| \bar{X}^n)$. This indicates the existence of $\bar{\epsilon} > 0$ such that $\bar{\rho}^{(t)}(\delta) > \bar{\epsilon}$, which immediately gives that for $n \in \mathcal{N}$ sufficiently large,

$$P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \ge \delta \right\} \le e^{-n\bar{\epsilon}}.$$

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

$$\begin{aligned} & P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \delta > \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \lambda - \delta \right\} \\ \ge\ & P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \lambda - \delta \right\} \\ & - P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} \ge \delta \right\} \\ \ge\ & 1 - e^{-n\epsilon} - e^{-n\bar{\epsilon}}. \end{aligned} \tag{22}$$

Let $I_1 \triangleq (\lambda - \delta, b_1)$, and[5]

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

where $b_k \triangleq (\lambda - \delta) + 2k\delta$ for $1 \le k < L$, and $b_L \triangleq \delta$. By (22), there exists $1 \le k(n) \le L$ such that

$$P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \in I_{k(n)} \right\} \ge \frac{1 - e^{-n\epsilon} - e^{-n\bar{\epsilon}}}{L}, \tag{23}$$

[5]Note that when $\lambda < 0$, $L \ge 2$; so the definition is well-established. However, in case $\lambda = 0$, we just take $L = 1$, and $I_1 = (-\delta, \delta)$.

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,

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

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

$$\begin{aligned}
& P_{\bar{X}^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \in I_{k(n)}\right\} \\
= & \sum_{\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \in I_{k(n)}\right\}} P_{\bar{X}^n}(x^n) \\
= & \sum_{\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \in I_{k(n)}\right\}} e^{-tD_{1-t}(X^n\|\bar{X}^n)} e^{(1-t)\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}} P^{(t)}_{X^n}(x^n) \\
\ge & \; e^{-tD_{1-t}(X^n\|\bar{X}^n)} e^{(1-t)nb_{k(n)-1}} \sum_{\left\{x^n\in\mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \in I_{k(n)}\right\}} P^{(t)}_{X^n}(x^n) \\
= & \; e^{-tD_{1-t}(X^n\|\bar{X}^n)} e^{(1-t)nb_{k(n)-1}} P^{(t)}_{X^n}\left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} \in I_{k(n)}\right\} \\
\ge & \; \frac{1 - e^{-n\epsilon} - e^{-n\bar{\epsilon}}}{L} e^{-tD_{1-t}(X^n\|\bar{X}^n)} e^{(1-t)nb_{k(n)-1}}.
\end{aligned}$$

Consequently,

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

where equality (a) holds since $\rho(R)$ admits a limit, and equality (b) follows from the definition of $\mathcal{N}$ in (20). $\square$

We observe that the convexity condition for $\rho(R)$ given in the above theorem is not necessary for the expression of the reverse β-cutoff rate to be given by the $\frac{1}{1-\beta}$-divergence rate. This is illustrated in the following example, where we show that $\rho(R)$ is not convex while

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

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

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

which implies that

$$\rho(R) \;=\; \lim_{n\to\infty} -\frac{1}{n}\log\Pr\left\{\frac{1}{n}Z_n \le R\right\} = \begin{cases} 0, & \text{for } R\ge 0 \\ 2, & \text{for } -(2-c)\le R<0 \\ \infty, & \text{for } R< -(2-c). \end{cases}$$

Note that $\rho(R)$ in not convex but $R+\rho(R) = 0$ for $R=0$. Note also that condition (16) is satisfied since $P_{X^n}(\cdot)$ and $P_{\bar{X}^n}(\cdot)$ are absolutely continuous with respect to each other. 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.

Let us now compute the reverse β-cutoff rate. First, we need to compute $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}})$ using Proposition 4. We have the following cases.

- $E > 2$. We have that

$$R + \rho(R) + [E - \rho(R)]^+ = \begin{cases} R + E, & \text{for } R \geq 0 \\ R + E, & \text{for } -(2-c) \leq R < 0 \\ \infty, & \text{for } R < -(2-c). \end{cases}$$

  Hence

$$\begin{aligned} D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) &= \inf_{R \in \mathbb{R}} \left\{ R + \rho(R) + [E - \rho(R)]^+ \right\} \\ &= E - 2 + c. \end{aligned}$$

- $0 < c < E \leq 2$. In this case

$$R + \rho(R) + [E - \rho(R)]^+ = \begin{cases} R + E, & \text{for } R \geq 0 \\ R + 2, & \text{for } -(2-c) \leq R < 0 \\ \infty, & \text{for } R < -(2-c). \end{cases}$$

  Hence, $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) = c$.

- $0 < E \leq c$. In this case

$$R + \rho(R) + [E - \rho(R)]^+ = \begin{cases} R + E, & \text{for } R \geq 0 \\ R + 2, & \text{for } -(2-c) \leq R < 0 \\ \infty, & \text{for } R < -(2-c). \end{cases}$$

  Hence, $D_e^*(E|\mathbf{X}\|\bar{\mathbf{X}}) = E$.

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. By straightforward calculation, we get that

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

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

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

Since, by definition, $R_0^{(r)}(\beta|\mathbf{X}\|\bar{\mathbf{X}}) \geq 0$, it is straightforward to check that

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

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

We finally present a class of sources with memory (countable or continuous alphabet) for which the reverse β-cutoff rate is given by the Rényi $\frac{1}{1-\beta}$-divergence rate for *all* $0 < \beta < 1$.

Corollary 1 Consider the hypothesis testing problem between sources with memory such that the log-likelihood ratio process $\{Z_n\}$ where $Z_n = \log \frac{P_{\bar{X}^n}(\bar{X}^n)}{P_{X^n}(\bar{X}^n)}$, satisfies both hypotheses of the Gärtner-Ellis Theorem [2, p. 15]:

- $\phi(\theta) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n}\phi_n(\theta)$ exists for all $\theta \in \mathbb{R}$,
- ϕ is differentiable on d_φ, where $d_\varphi \stackrel{\triangle}{=} \{\theta : \phi(\theta) < \infty\}$,

where $\phi_n(\theta) \stackrel{\triangle}{=} \log E_{P_{\bar{X}^n}}[e(\theta Z_n)]$. Then the reverse β-cutoff rate satisfies

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

Proof: We will prove the result for countable alphabets. The required modifications for the continuous alphabet case are straightforward. We need to show that for sources satisfying the Gärtner-Ellis Theorem, the Rényi divergence rate exists, that the conditions of Theorem 2 hold and that $\beta_{\max} = 1$. First, the Rényi divergence rate exists and $\beta_{\max} = 1$ from the first hypothesis of the Gärtner-Ellis Theorem and the fact that

$$\frac{1}{n} D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n) = \frac{1-\beta}{\beta}\frac{1}{n}\phi_n\left(\frac{1}{\beta-1}\right).$$

Next, by the Gärtner-Ellis Theorem, we have that

$$\rho(R) = \sup_{\theta\leq 0} \{\theta R - \phi(\theta)\}.$$

Clearly, $\rho(R)$ admits a limit and is convex in R. Let us show that there exists an R such that $R+\rho(R)=0$. First, note that $E_{P_{\bar{X}^n}}[e(\theta Z_n)]=1$ for $\theta=0$ and $\theta=-1$. Hence, $\phi(\theta)=0$ for $\theta=0$ and $\theta=-1$. This implies that

$$\rho(R)+R=\sup_{\theta\leq 0}\left[(\theta+1)R-\phi(\theta)\right]\geq\left[(\theta+1)R-\phi(\theta)\right]\Big|_{\theta=-1}=0. \tag{24}$$

Observe that

$$\begin{aligned} e^{\phi_n(\theta)}=E[e^{\theta Z_n}] &= \sum e^{\theta\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}}P_{\bar{X}^n}(x^n) \\ &= \sum e^{(\theta+1)\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}}P_{X^n}(x^n) \\ &\geq e^{(\theta+1)\sum\log\frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)}P_{X^n}(x^n)}=e^{(\theta+1)[-D(X^n\|\bar{X}^n)]}, \end{aligned} \tag{25}$$

where $D(\cdot\|\cdot)$ is the Kullback-Leibler divergence and (25) follows by Jensen's inequality. This implies that

$$\left\{(\theta+1)R_n-\frac{1}{n}\phi_n(\theta)\right\}\leq 0 \quad \text{for all } \theta\leq 0,$$

where $R_n=-\frac{1}{n}D(X^n\|\bar{X}^n)$. As a result,

$$\begin{cases} \limsup_{n\to\infty}\left\{(\theta+1)R_n-\frac{1}{n}\phi_n(\theta)\right\}=(\theta+1)\liminf_{n\to\infty}R_n-\phi(\theta)\leq 0, & \text{for } \theta<-1 \\ \liminf_{n\to\infty}\left\{(\theta+1)R_n-\frac{1}{n}\phi_n(\theta)\right\}=(\theta+1)\liminf_{n\to\infty}R_n-\phi(\theta)\leq 0, & \text{for } 0\geq\theta\geq -1 \end{cases}$$

Therefore

$$\rho(\liminf_{n\to\infty}R_n)+\liminf_{n\to\infty}R_n\leq 0.$$

On the other hand, by (24), we have that

$$\rho(\liminf_{n\to\infty}R_n)+\liminf_{n\to\infty}R_n\geq 0.$$

Hence

$$\rho(\liminf_{n\to\infty}R_n)+\liminf_{n\to\infty}R_n=0.$$

$\square$

Remark 3: By Corollary 1, for i.i.d. finite-alphabet observations (in this case, Gärtner-Ellis Theorem reduces to Cramer's Theorem), our result in Theorem 2 reduces to Csiszár's result [5]; i.e., the reverse β-cutoff rate is given by the Rényi divergence with parameter $\frac{1}{1-\beta}$, for $0<\beta<1$.

Remark 4: The above corollary holds for the class of finite-alphabet irreducible Markov sources since the latter satisfy both hypotheses of the Gärtner-Ellis Theorem. Indeed, the α-divergence rate exists and is differentiable; furthermore, it admits a simple computable expression [14].

Numerical Examples:

We briefly present two examples of memoryless sources where we explicitly verify the existence of R such that $R + \rho(R) = 0$.

Example 6: *Finite-alphabet memoryless sources:* Consider Example 1 in Section 4 where $\mathbf{X}$ and $\bar{\mathbf{X}}$ are interchanged. Note that $\rho(R)$ is equal to $\eta(R)$ in this case. It is straightforward to check that $R + \rho(R) = 0$ for R approximately -0.13.

Example 7: *Continuous alphabet memoryless sources:* Consider Example 2 in Section 4 where $\mathbf{X}$ and $\bar{\mathbf{X}}$ are interchanged. Note that $\rho(R)$ is equal to $\eta(R)$ in this case. By straightforward calculation we get that $R + \rho(R) = 0$ for $R = -2\nu^2$.

7 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.) of arbitrary alphabet (countable or continuous). We showed that the forward β-cutoff rate is given by the $\liminf$ α-divergence rate, where $\alpha = \frac{1}{1-\beta}$ and $\beta < 0$. Under two conditions on the large deviation spectrum, $\rho(R)$, we showed that the reverse β-cutoff rate is given by the α-divergence rate, where $\alpha = \frac{1}{1-\beta}$ and $0 < \beta < \beta_{\max}$. For $\beta_{\max} \leq \beta < 1$, we provided an upper bound to the reverse β-cutoff rate. We also investigated sources with memory (countable or continuous alphabet) that satisfy the hypotheses of the Gärtner-Ellis Theorem. We showed that the conditions on $\rho(R)$ are satisfied and that the reverse β-cutoff rate is given by the Rényi divergence rate. A direct consequence is that, for i.i.d. observations, our result indeed reduces to Csiszár's result, hence providing a simple expression for the reverse β-cutoff rate in terms of the Rényi divergence. Another consequence is that, for finite-alphabet irreducible Markov sources, the reverse β-cutoff rate is given by the Rényi divergence rate which can be computed using Perron-Frobenius theory [14]. We also provided several numerical examples to illustrate our forward and reverse β-cutoff rates results.

Future work may include the study of Csiszár's channel coding β-cutoff rates [5] for arbitrary discrete channels with memory using our information spectrum techniques.

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

Lemma 5 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

$$\begin{aligned}
& P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}(x^n)}{P_{\bar{X}^n}(x^n)} > \frac{1}{n} D_{1-t}(X^n \| \bar{X}^n) + \nu \right\} \\
=\ & P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : -\frac{1}{n} \left(\log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + D_{1-t}(X^n \| \bar{X}^n) \right) > \nu \right\} \\
=\ & P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{t}{n} \left(\log \frac{P_{\bar{X}^n}(x^n)}{P_{X^n}(x^n)} + D_{1-t}(X^n \| \bar{X}^n) \right) < -\nu t \right\} \\
=\ & P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log \frac{P_{X^n}^{(t)}(x^n)}{P_{X^n}(x^n)} < -\nu t \right\} \qquad (26) \\
=\ & P_{X^n}^{(t)} \left\{ x^n \in \mathcal{X}^n : P_{X^n}^{(t)}(x^n) < e^{-n\nu t} P_{X^n}(x^n) \right\} \\
\le\ & e^{-n\nu t} P_{X^n} \left\{ x^n \in \mathcal{X}^n : P_{X^n}^{(t)}(x^n) < e^{-n\nu t} P_{X^n}(x^n) \right\} \\
\le\ & e^{-n\nu t},
\end{aligned}$$

where (26) 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,

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

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

$\square$

Lemma 6 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

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

Hence,

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

which implies that

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

Therefore,

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

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 7 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. □

Appendix B: Properties of λ

Lemma 8 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 $D_{1-t}(X^n \| \bar{X}^n)$. This implies that for $R > 0$,

$$\rho^{(t)}(R) \;\leq\; \liminf_{n \in \mathcal{N}, n \to \infty} -\frac{1}{n} \log \left(1 - e^{-nR(1-t)}\right) = 0,$$

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

Lemma 9 For $0 > t > \beta_{\max}/(\beta_{\max}-1)$, $\lambda > -\infty$.

Proof: If $\lambda = -\infty$, then $\rho^{(t)}(R) = 0$ for every $R \in \mathbb{R}$. Hence, by choosing any $\delta > 0$ satisfying $t > t-\delta > \beta_{\max}/(\beta_{\max}-1)$, we have:

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

which implies that

$$0 = \rho^{(t)}(R) \ge (t-\delta)\limsup_{n\in\mathcal{N}, n\to\infty}\frac{1}{n}D_{1-(t-\delta)}(X^n\|\bar{X}^n) - \delta R.$$

This indicates that

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

or equivalently,

$$\limsup_{n\to\infty}\frac{1}{n}D_{1-(t-\delta)}(X^n\|\bar{X}^n) = \infty,$$

which contradicts the assumption on $\beta_{\max}$. $\square$

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

[3] 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.

[4] 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.

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

[6] 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.

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

[8] T. S. Han, *Information-Spectrum Methods in Information Theory*, Tokyo, Japan: BaifuKan, 1998 (in Japanese).

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

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

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

[12] 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.

[13] T. Nemetz, "On the α-divergence rate for Markov-dependent hypotheses," *Problems of Control and Information Theory,* vol. 3 (2), pp. 147–155, 1974.

[14] 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.

[15] 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.

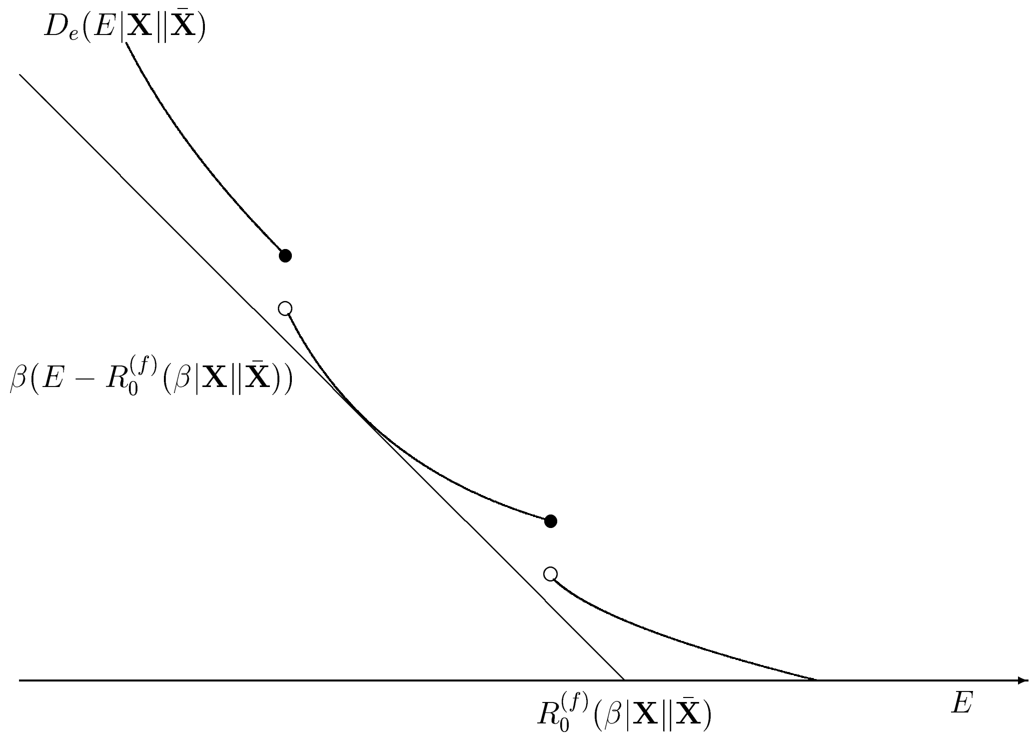

Figure 1: A graphical illustration of the forward β-cutoff rate, $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$, for testing between two arbitrary sources $\mathbf{X}$ and $\bar{\mathbf{X}}$.

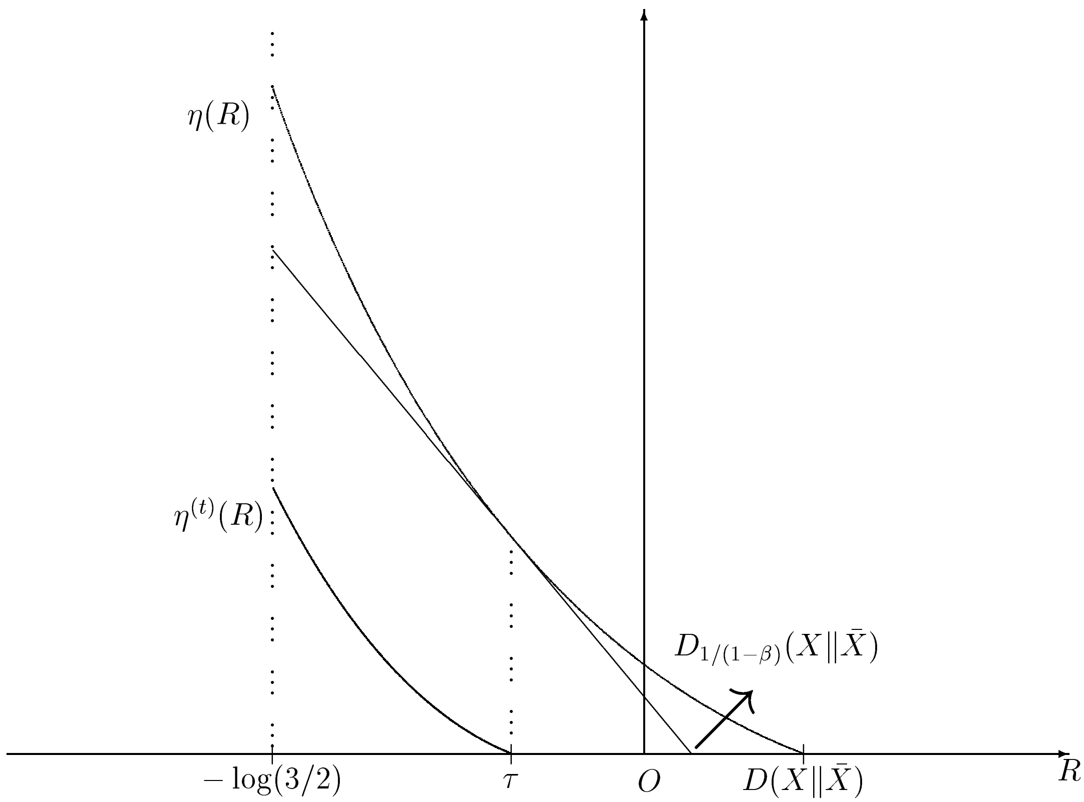

Figure 2: Functions $\eta(R)$, $\eta^{(t)}(R)$ and $(\beta/(\beta-1))\left[\liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)-R\right]$ for testing between two binary memoryless sources $\mathbf{X}=\{X_i\}_{i=1}^{\infty}$ and $\bar{\mathbf{X}}=\{\bar{X}_i\}_{i=1}^{\infty}$ under the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

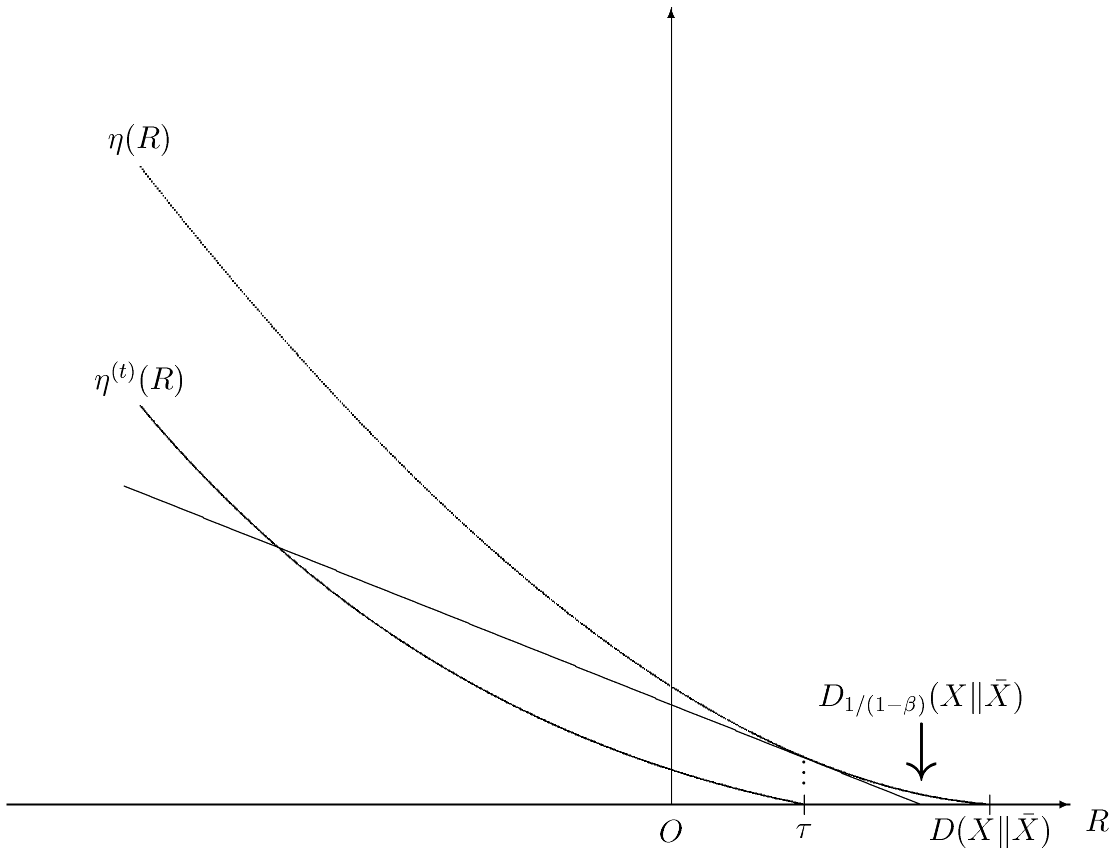

Figure 3: Functions $\eta(R)$, $\eta^{(t)}(R)$ and $(\beta/(\beta-1))\left[\liminf_{n\to\infty}\frac{1}{n}D_{\frac{1}{1-\beta}}(X^n\|\bar{X}^n)-R\right]$ for testing between two memoryless sources $\mathbf{X}=\{X_i\}_{i=1}^{\infty}$ and $\bar{\mathbf{X}}=\{\bar{X}_i\}_{i=1}^{\infty}$ under the Gaussian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

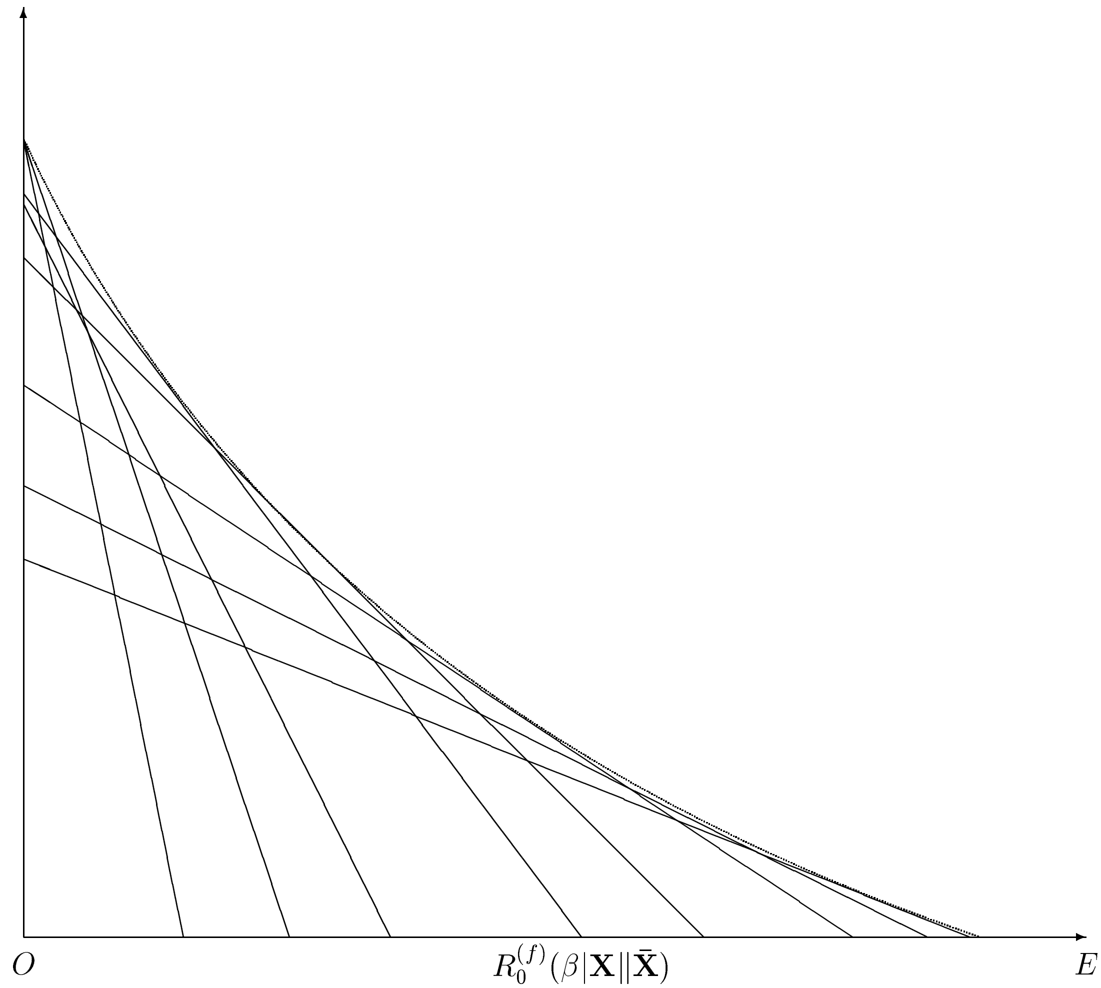

Figure 4: Convex lower bound for testing between irreducible Markov sources. Each line of slope β intersects the E-axis at $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$. Proceeding from left to right, the values of β are: $-5, -3, -2, -4/3, -1, -2/3, -1/2, -2/5$.

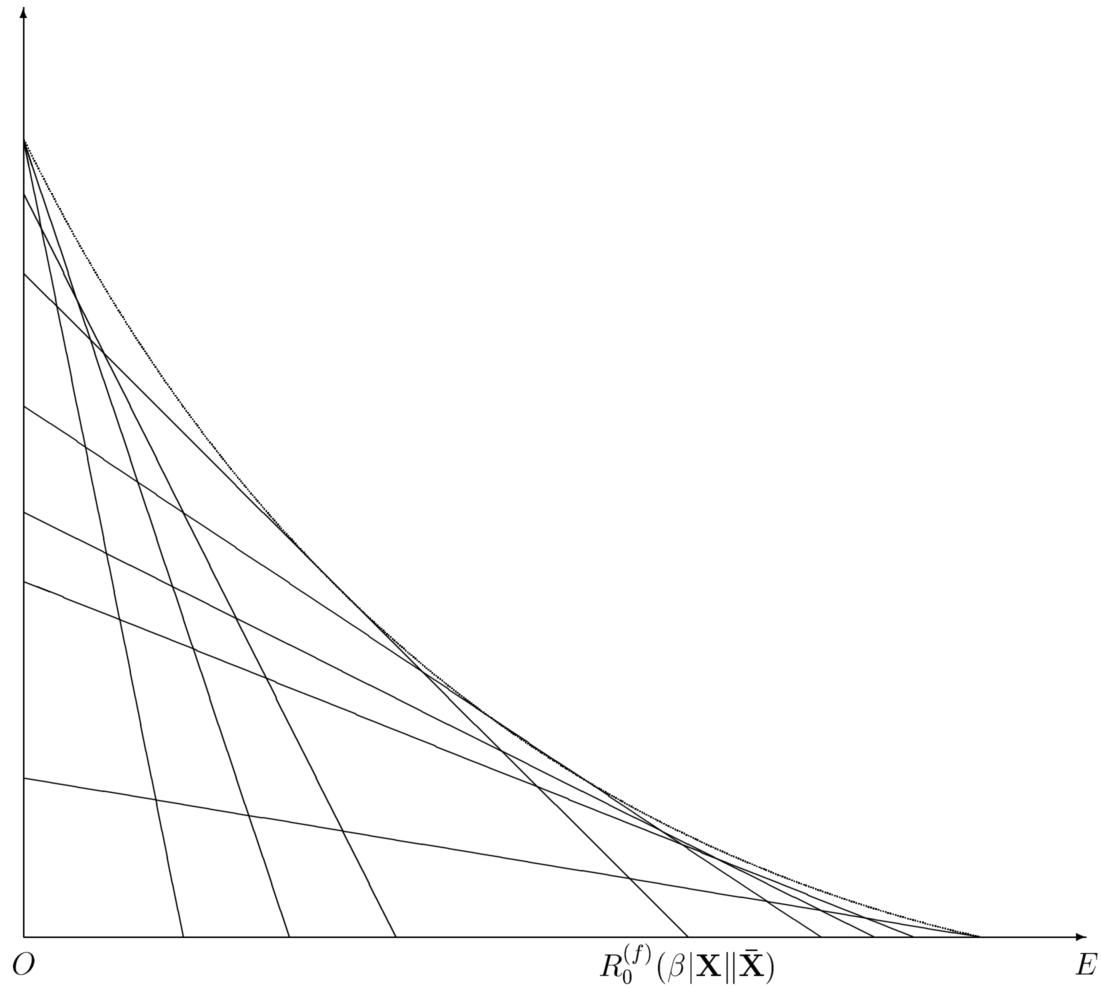

Figure 5: Convex lower bound for testing between arbitrary Markov sources. Each line of slope β intersects the E-axis at $R_0^{(f)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$. Proceeding from left to right, the values of β are: $-5, -3, -2, -1, -2/3, -1/2, -2/5, -1/6$.

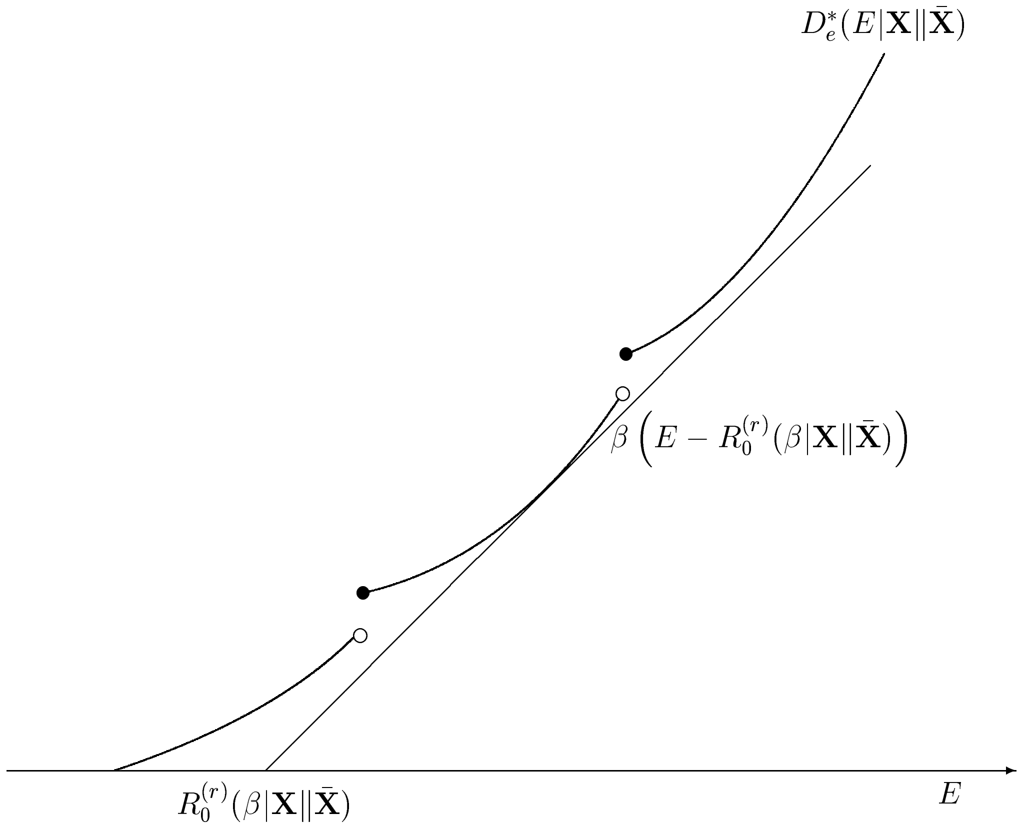

Figure 6: A graphical illustration of the reverse β-cutoff rate, $R_0^{(r)}(\beta|\mathbf{X}\|\bar{\mathbf{X}})$, for testing between two arbitrary sources $\mathbf{X}$ and $\bar{\mathbf{X}}$.

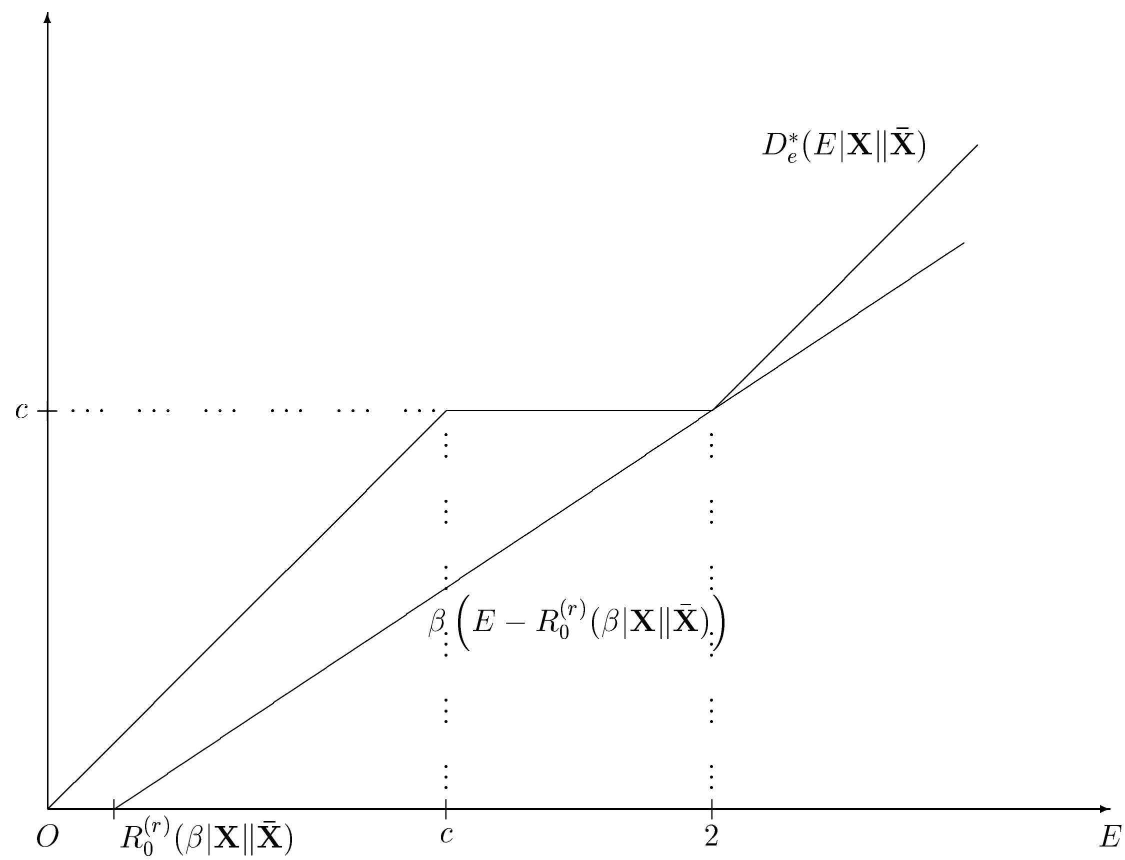

Figure 7: Reliability function of the type 1 probability of correct decoding for testing between the two sources $P_{X^n}(\cdot)$ and $P_{\bar{X}^n}(\cdot)$ as given in Example 5.