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

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

[11] ——, "The reliability functions of the general source with fixed-length coding," *IEEE Trans. Information Theory*, vol. 46, pp. 2117–2132, Sept 2000.

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

[13] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 197–199, Mar. 1974.

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

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

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

[17] ——, "Rényi's divergence and entropy rates for finite alphabet Markov sources," *IEEE Trans. Inform. Theory*, May. 2001, to be published.

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

The First-Order Asymptotic of Waiting Times with Distortion Between Stationary Processes

Zhiyi Chi

***Abstract*—Let X and Y be two independent stationary processes on general metric spaces, with distributions P and Q, respectively. The first-order asymptotic of the waiting time $W_n(D)$ between X and Y, allowing distortion, is established in the presence of one-sided ψ-mixing conditions for Y. With probability one, $n^{-1} \log W_n(D)$ has the same limit as $-n^{-1} \log Q(B(X_1^n, D))$, where $Q(B(X_1^n, D))$ is the Q-measure of the D-ball around $(X_1, \ldots, X_n)$, with respect to a given distortion measure. Large deviations techniques are used to get the convergence of $-n^{-1} \log Q(B(X_1^n, D))$. First, a sequence of functions R_n in terms of the marginal distributions of X_1^n and Y_1^n as well as D are constructed and demonstrated to converge to a function $R(P, Q, D)$. The functions R_n and $R(P, Q, D)$ are different from rate distortion functions. Then $-n^{-1} \log Q(B(X_1^n, D))$ is shown to converge to $R(P, Q, D)$ with probability one.**

***Index Terms*—Large deviation, ψ-mixing, relative entropy, string matching, waiting times.**

I. Introduction

This correspondence generalizes the results of Yang and Kieffer (1998), Dembo and Kontoyiannis (1999), and Yang and Zhang (1999) on the first-order asymptotics of waiting times between two independent stationary processes to the case allowing distortion and more general mixing conditions.

In recent years, there has been increasing interest in the asymptotic properties of waiting times between stationary processes, due to their applications in coding theory and DNA analysis. The idea of using waiting times for coding was first introduced in [4]. Most of the previous studies focus on the case of exact matching (see [4]–[8]; and the references therein). This correspondence will be mainly concerned with the first-order asymptotics of the waiting times for matching allowing distortion, which have been studied in several recent publications ([9], [1]–[3]).

Manuscript received August 23, 1999; revised July 20, 2000.

The author is with the Department of Statistics, The University of Chicago, Chicago, IL 60637 USA (e-mail: chi@galton.uchicago.edu).

Communicated by I. Csisár, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(01)00600-9.

Let $X = \{X_n, n \geq 1\}$ and $Y = \{Y_n, n \geq 1\}$ be independent processes taking values in Polish spaces $(\mathcal{A}_X^\infty, \mathcal{F}_X)$ and $(\mathcal{A}_Y^\infty, \mathcal{F}_Y)$, with distributions P and Q, respectively. Given $a = \{a_n, n \geq 1\}$, denote $a_i^j = (a_i, \ldots, a_j)$. Denote by $\sigma(X_i^j)$ the σ-field generated by X_i^j, and likewise for Y_i^j.

Given a measurable function ρ: $\mathcal{A}_X \times \mathcal{A}_Y \to [0, \infty)$, the distortion measure ρ_n for any $x_1^n \in \mathcal{A}_X^n$ and $y_1^n \in \mathcal{A}_Y^n$ is defined as

$$\rho_n(x_1^n, y_1^n) = \frac{1}{n} \sum_{i=1}^n \rho(x_i, y_i). \tag{1.1}$$

For $x_1^n \in \mathcal{A}_X^n$ and $D \geq 0$, denote

$$B(x_1^n, D) = \{y_1^n \in \mathcal{A}_Y^n\colon \rho_n(x_1^n, y_1^n) \leq D\}. \tag{1.2}$$

Given $D \geq 0$, for samples x, y from X, Y, respectively, the waiting time $W_n(D)$ until a D-close version of x_1^n first appears in y is

$$W_n(D) = W_n(x_1^n, y, D) = \inf\{k \geq 1\colon y_k^{k+n-1} \in B(x_1^n, D)\}.$$

For finite $\mathcal{A}_X$ and $\mathcal{A}_Y$, Yang and Kieffer (1998) proved that if Y is an independent and identically distributed (i.i.d.) process or a kth-order Markov process with positive transition probabilities, and X is a stationary ergodic process, then

$$\frac{1}{n} \log W_n(D) \to R(P, Q, D) \qquad (P \times Q)\text{–a.s.} \tag{1.3}$$

for some function $R(P, Q, D)$, given as the solution of a variational problem in terms of relative entropy. Similar results were obtained in [9] as well. More recently, using large deviations techniques, Dembo and Kontoyiannis (1999) and Yang and Zhang (1999) independently proved (1.3) for general spaces $\mathcal{A}_X$ and $\mathcal{A}_Y$, when Y is i.i.d., while the second paper established finer higher order asymptotics for $W_n(D)$. We will use large deviations techniques to generalize these authors' results on the first-order asymptotic of $\log W_n(D)$ to the case where Y is not i.i.d., which often comes up in applications and significantly complicates analysis. Specifically, we will consider Y which satisfies the following mixing condition: there are constants $C \geq 1$, $d \geq 1$, such that for any $A \in \sigma(Y_1^n)$ and $B \in \sigma(Y_{n+d+1}^\infty)$

$$(\psi_-)\colon Q(A)Q(B) < CQ(A \cap B)$$
$$(\psi_+)\colon Q(A \cap B) < CQ(A)Q(B).$$

The above "one-sided" ψ-mixing conditions (ψ_-) and (ψ_+) were explored in detail in [10] and [11]. It is known that for Y ergodic and mixing, condition (ψ_-) plus condition (ψ_+) is equivalent to ψ-mixing ([12]), which is satisfied by kth-order Markov processes on finite spaces with positive transition probabilities. Recall that Y is called ψ-mixing if $\lim_{k\to\infty} \psi(k) = 0$, where

$$\psi(k) = \sup\left\{\left|\frac{Q(A \cap B)}{Q(A)Q(B)} - 1\right|\colon A \in \sigma(Y_1^r),\right.$$
$$\left. B \in \sigma(Y_{r+k+1}^\infty),\ r \geq 1\right\}.$$

Following the methodology in [1], the first step to (1.3) is to establish a strong approximation of $W_n(D)$ by $Q(B(X_1^n, D))$.

Proposition 1: Let X and Y be two independent stationary processes on $\mathcal{A}_X^\infty$ and $\mathcal{A}_Y^\infty$, with distributions P and Q, respectively. Suppose $Q(B(X_1^n, D)) > 0$ eventually P-a.s. If X is ergodic and Y satisfies condition (ψ_-), then

$$\frac{1}{n}\log W_n(D) = -\frac{1}{n}\log Q(B(X_1^n, D)) + o(1), \qquad n\to\infty,\ (P\times Q)\text{–a.s.} \tag{1.4}$$

We omit the proof for Proposition 1 because it follows the same line as the one in [8], which is for a strong approximation when the matching is exact and Y is ψ-mixing. A similar result was also proved in [9]. Equation (1.4) has also been proved under summable ϕ-mixing condition for Y, i.e., $\sum \phi(k) < \infty$, with ([1], [2])

$$\phi(k) = \sup\{|Q(B|A) - Q(B)|: A\in\sigma(Y_1^r),\ B\in\sigma(Y_{r+k+1}^\infty),\ r\ge 1\}.$$

While ψ-mixing implies ϕ-mixing, in general there are no further implications between them. More details on the mixing conditions can be found in [12].

We now come to the first main issue in establishing the asymptotics of the waiting times, for Y not being i.i.d. By (1.4), in order to prove (1.3), it is enough to get

$$\log Q(B(X_1^n, D)) \to -R(P, Q, D).$$

For Y being i.i.d., $R(P, Q, D)$ is obtained as the solution to a variational problem in terms of P_1 and Q_1, where for $n\ge 1$, P_n and Q_n are the marginals of X_1^n and Y_1^n, respectively. In contrast, for Y only satisfying the one-sided ψ-mixing conditions, the definition and existence of $R(P, Q, D)$ need to be addressed. To begin with, given $D>0$ and $n\ge 1$, define

$$R_n(P_n, Q_n, D) = \frac{1}{n}\inf\left\{\int_{\mathcal{A}_X^n} D(\nu(\cdot|x_1^n)\|Q_n(\cdot))\,dP_n(x_1^n):\ \nu\in\mathcal{P}_n,\ \nu_1 = P_n,\ \rho_\nu^{(n)}\le D\right\} \tag{1.5}$$

where $\mathcal{P}_n = \mathcal{P}(\mathcal{A}_X^n\times\mathcal{A}_Y^n)$ is the set of probability measures on $\mathcal{A}_X^n\times\mathcal{A}_Y^n$, ν_1 the $\mathcal{A}_X^n$-marginal of ν

$$\rho_\nu^{(n)} = \int \rho_n(x_1^n, y_1^n)\,d\nu(x_1^n, y_1^n)$$

and $D(\mu\|\nu)$ the relative entropy between μ and ν, defined as[1]

$$D(\mu\|\nu) = \begin{cases}\int\left(\log\dfrac{d\mu}{d\nu}\right)d\mu, & \text{if } \frac{d\mu}{d\nu} \text{ exists}\\ \infty, & \text{otherwise.}\end{cases}$$

We then define the function $R(P, Q, D)$ as

$$R(P, Q, D) = \lim_{n\to\infty} R_n(P_n, Q_n, D). \tag{1.6}$$

$R(P, Q, D)$ is a generalization of the joint entropy. Indeed, if $\mathcal{A}_X = \mathcal{A}_Y$ and discrete, then with $\rho(x, y) = \mathbf{1}_x(y)$, we get exact matching and

$$R_n(P_n, Q_n, 0) = n^{-1}H(P_n) + n^{-1}D(P_n\|Q_n).$$

[1]When $n>1$, $R_n(P_n, Q_n, D)$ defined here are different from those in [1].

Therefore, in this special case, if $H(X)<\infty$, then $R(P, Q, 0) = H(X) + D(X\|Y)$, provided either $R(P, Q, 0)$ or $D(X|Y)$ exists.

Under regular conditions

$$R_n(P_n, Q_n, D) = \frac{1}{n}\inf\left\{D(\nu\|P_n\times Q_n): \nu\in\mathcal{P}_n,\ \nu_1 = P_n,\ \rho_\nu^{(n)}\le D\right\}.$$

It is worth noting that the functions $R_n(P_n, Q_n, D)$ are different from the n-symbol rate distortion functions $R_n(P_n, D)$, defined as ([13], [14])

$$R_n(P_n, D) = \frac{1}{n}\inf\left\{D(\nu\|P_n\times\nu_2): \nu\in\mathcal{P}_n, \nu_1 = P_n, \rho_\nu^{(n)}\le D\right\}.$$

Indeed, in [3], it was shown that

$$R_n(P_n, D) = \inf\{R_n(P_n, \mu, D): \mu\in\mathcal{P}(\mathcal{A}_Y^n)\}.$$

In addition, if X satisfies $\limsup_{n\to\infty} R_n(P_n, D) < \infty$, then for the rate distortion function $R(P, D) = \inf_{n\ge 1} R_n(P_n, D)$, there is $R(P, D) = \lim_{n\to\infty} R_n(P_n, D)$ ([14, p. 93]).

Similar to the rate-distortion function, the convergence in (1.6) in general is a nontrivial problem, except for the case where Y is an i.i.d. process. Denote

$$D_{\rm av} = E\rho(X_1, Y_1), \qquad D_{\max} = \operatorname*{ess\,sup}_{(X_1, Y_1)}\rho(X_1, Y_1)$$
$$D_{\min}^{(n)} = E_P\left[\operatorname*{ess\,inf}_{Y_1^n}\rho_n(X_1^n, Y_1^n)\right], \qquad D_{\min} = \lim_{n\to\infty} D_{\min}^{(n)}. \tag{1.7}$$

Obviously, $E\rho_n(X_1^n, Y_1^n) = D_{\rm av}$. We will assume

$$D_{\max} < \infty, \quad D_{\min}^{(n)} < D_{\rm av}, \qquad \text{for all } n\ge 1. \tag{1.8}$$

It is easy to show that the limit in (1.7) exists. Therefore, $D_{\min}$ is finite as well.

Using log-moment generating functions ([2, Proposition 1]; [3, Property 1]), it can be shown that for $D\in(D_{\min}^{(n)}, D_{\rm av})$

$$R_n(P_n, Q_n, D) = \Lambda_n^*(D) \tag{1.9}$$

where $\Lambda_n^*(D) = \sup_{\lambda\in\mathbf{R}}[\lambda D - \Lambda_n(\lambda)]$ is the Fenchel–Legendre transform of Λ, with

$$\Lambda_n(\lambda) = \frac{1}{n}\int_{\mathcal{A}_X^n}\log\left\{\int_{\mathcal{A}_Y^n} e^{n\lambda\rho_n(x_1^n, y_1^n)}\,dQ_n(y_1^n)\right\}dP_n(x_1^n). \tag{1.10}$$

The supremum in the definition of $\Lambda_n^*(D)$ can be achieved over $\lambda\le 0$, i.e.,

$$\begin{aligned}R_n(P_n, Q_n, D) &= \sup_{\lambda\le 0}[\lambda D - \Lambda_n(D)]\\ &= \sup_{\lambda\le 0}\left[\lambda D - \frac{1}{n}\int_{\mathcal{A}_X^n}\log\left\{\int_{\mathcal{A}_Y^n} e^{n\lambda\rho_n(x_1^n, y_1^n)}\,dQ_n(y_1^n)\right\}\cdot dP_n(x_1^n)\right].\end{aligned} \tag{1.11}$$

For $D\ge D_{\rm av}$, (1.9) still holds. First

$$R_n(P_n, Q_n, D) = 0, \qquad D\ge D_{\rm av}$$

as can be seen by taking

$$d\nu(x_1^n,\, y_1^n) = \, dP_n(x_1^n)\, dQ_n(y_1^n)$$

in (1.5). Second, since $\Lambda_n'(0) = E\rho_n(x_1^n,\, y_1^n) \le D$ and $\Lambda_n''(\lambda) \ge 0$, $\lambda D - \Lambda_n(\lambda)$ is nondecreasing on $(\infty,\, 0]$, and achieves its supremum at $\lambda = 0$, thus, $\sup_{\lambda \le 0}[\lambda D - \Lambda_n(\lambda)] = 0$.

We will show that the condition (ψ_-) guarantees the convergence of $R_n(P_n,\, Q_n,\, D)$.

Theorem 1: Let X and Y be two independent stationary processes with distributions P and Q, respectively. If X is ergodic and Y satisfies condition (ψ_-), then $R_n(P_n,\, Q_n,\, D)$ and $R(P,\, Q,\, D)$ are continuous functions on $(D_{\min},\, \infty)$, and

$$R_n(P_n,\, Q_n,\, D) \xrightarrow{\text{u.c.}} R(P,\, Q,\, D)$$

where $\xrightarrow{\text{u.c.}}$ stands for uniform convergence on any compact set. In addition, for $D \ge D_{\text{av}}$, $R(P,\, Q,\, D) = 0$.

For the case of exact matching, argument similar to that for Theorem 1 leads to to the following.

Proposition 2: $P \ll Q$ and Y satisfies condition (ψ_+), then $n^{-1} D(P_n \| Q_n) \to D(X \| Y)$.

The proof for Proposition 2 is rather simple. Because this correspondence is mainly concerned with matching allowing distortion, the proof will be relegated to the Appendix.

The second main issue in establishing the first-order asymptotic of the waiting times is to show

$$\lim_{n\to\infty} \log\, Q(B(x_1^n,\, D)) = -R(P,\, Q,\, D)$$

for almost all random sample from X. To make a connection between the desired convergence and the Large Deviation Principle (LDP), consider the special case where $\mathcal{A}_X = \{\theta\}$ is a singleton. Consequently, $\mathcal{A}_X^\infty$ is a singleton. Let $Z = \{Z_n,\, n \ge 1\}$ be a process on $[0,\, \infty)$ such that $Z_n = \rho(\theta,\, Y_n)$. Then the desired convergence is implied by the LDP of $\{\mu_n\}$, with μ_n the law of $(Z_1 + \ldots + Z_n)/n$. Furthermore, by (1.11) and (1.6), $R(P,\, Q,\, D)$ is the rate function for the LDP of $\{\mu_n\}$, provided it satisfies the condition of Gärtner–Ellis theorem ([14, pp. 43–54]). It is known that if Z satisfies condition (ψ_-), then $\{\mu_n\}$ indeed satisfies LDP ([10]).

Theorem 2: Let X and Y be two independent stationary processes, with distributions P and Q, respectively. If X is ergodic and Y satisfies both conditions (ψ_-) and (ψ_+), then for $D \in (D_{\min},\, \infty)$, there is

$$\frac{1}{n} \log\, Q(B(X_1^n,\, D)) \to -R(P,\, Q,\, D), \qquad P\text{–a.s.} \tag{1.12}$$

The above results immediately lead to the first-order asymptotic of the waiting times $W_n(D)$.

Corollary 1: Let X and Y be two independent stationary processes on $\mathcal{A}_X$ and $\mathcal{A}_Y$, respectively. If X is ergodic and Y satisfies both conditions (ψ_-) and (ψ_+), then $W_n(D)$ have the first-order asymptotic (1.3).

The proof of Theorem 2 consists of two parts. First, we show the asymptotic upper bound

$$\limsup_{n\to\infty} n^{-1} \log\, Q(B(X_1^n,\, D)) \le -R(P,\, Q,\, D), \qquad P\text{–a.s.}$$

Given large $k \ge 1$, define two auxiliary processes

$$U^{(k)} = \left\{U_n^{(k)},\, n \ge 1\right\}$$

and

$$V^{(k)} = \left\{V_n^{(k)},\, n \ge 1\right\}$$

such that $U^{(k)}$ is given by

$$U_n^{(k)} = (X_n,\, \ldots,\, X_{n+k-1}) \in \mathcal{A}_X^k$$

and $V^{(k)}$ is an i.i.d. process on $\mathcal{A}_Y^k$, with $V_1^{(k)} \sim Q_k$. By condition (ψ_-) and the continuity of $R(P,\, Q,\, D)$ in D, the proof for the upper bound in terms of P and Q can be reduced to those for a sequence of upper bounds in terms of the distributions of $U^{(k)}$ and $V^{(k)}$. Then we can apply the results in [2] and [3] to $U^{(k)}$ and $V^{(k)}$, $k = 1,\, 2,\, \ldots,$ and eventually prove the asymptotic upper bound for $n^{-1} \log\, Q(B(X_1^n,\, D))$.

Second, we show

$$\liminf_{n\to\infty} n^{-1} \log\, Q(B(X_1^n,\, D)) \ge -R(P,\, Q,\, D), \qquad P\text{–a.s.}$$

The idea is again to reduce the proof to the case where the second process is i.i.d. The large deviations technique used for the proof is the standard exponential change of measures based on the log-moment generating functions ([14, pp. 31–34]). To prove the lower bound for almost all random samples x from X, we define a series of empirical log-moment generating functions in terms of x and Y. Because Y is not i.i.d., it is essential that the log-moment generating functions be devised appropriately to have several asymptotic properties necessary for the proof of the lower bound. Also because Y is not i.i.d., a random perturbation argument is employed in the proof.

In Section II, the convergence of $R_n(P_n,\, Q_n,\, D)$ to $R(P,\, Q,\, D)$ is proved. In Section III, the P-almost sure convergence

$$n^{-1} \log\, Q(B(X_1^n,\, D)) \to -R(P,\, Q,\, D)$$

is established. In the proofs we will use Hammersley's lemma on approximate subadditivity, which can be found in [11]. For convenience, part of the result is quoted as follows.

Lemma 1: Assume $\{h_n,\, n \ge 1\}$ is a sequence such that

$$h_{m+n} \le h_m + h_n + \Delta_{m+n}, \qquad m,\, n \ge 1$$

with $\{\Delta_n,\, n \ge 1\}$ is a nondecreasing nonnegative sequence satisfying

$$\sum_{n=1}^\infty \frac{\Delta_n}{n(n+1)} < \infty.$$

Then $L = \lim_{n\to\infty}[h_n/n]$ exists and $L < \infty$.

II. Convergence of $R_n(P_n,\, Q_n,\, D)$ and Asymptotics of $R(P,\, Q,\, D)$

Proof of Theroem 1: The proof is based on the following lemmas.

Lemma 2: Suppose Y satisfies condition (ψ_-). Fix $C > 0$ and $d \ge 1$ such that $Q(A)Q(B) < CQ(A \cap B)$, $n \ge 1$, $A \in \sigma(Y_1^n)$, $B \in \sigma(Y_{n+d+1}^\infty)$. Then for any $m,\, n \ge 1$ with $m + n > d$, and $\lambda \le 0$

$$(m+n)\Lambda_{m+n}(\lambda) \ge m\Lambda_m(\lambda) + n\Lambda_n(\lambda) - \log\, C + dD_{\max}\lambda, \tag{2.1}$$

Furthermore, given $k \ge 1$, for $n \ge 1$, denoting

$$n = s(k+d) + r, \quad s = \left\lfloor \frac{n}{k+d} \right\rfloor, \qquad 0 \le r < k+d \tag{2.2}$$

we have

$$\Lambda_n(\lambda) \ge \frac{sk}{n}\, \Lambda_k(\lambda) - \frac{s}{n} \log\, C + \frac{(sd+r)D_{\max}\lambda}{n}. \tag{2.3}$$

Lemma 3: If Λ_n are continuous convex functions on $\boldsymbol{R}$ and $\Lambda_n(\lambda) \to \Lambda(\lambda)$, with $|\Lambda(\lambda)| < \infty$, for all $\lambda \in \boldsymbol{R}$, then $\Lambda_n \xrightarrow{\text{u.c.}} \Lambda$.

Lemma 4: If Y satisfies (ψ_-), then

$$\Lambda(\lambda) = \lim_{n \to \infty} \Lambda_n(\lambda) \tag{2.4}$$

exists and is convex and continuous. Furthermore, $\Lambda_n(\lambda) \xrightarrow{\text{u.c.}} \Lambda(\lambda)$ and

$$\lim_{\lambda \to -\infty} \frac{\Lambda(\lambda)}{\lambda} = D_{\min}. \tag{2.5}$$

Assuming the above three lemmas for now, since

$$R_n(P_n, Q_n, D) = \lambda^*(D)$$

is convex in D, by Lemma 3, we only need to prove that $R_n(P_n, Q_n, D)$ converge pointwise to $R(P, Q, D)$ on $(D_{\min}, \infty)$, and $R(P, Q, D) < \infty$.

We will prove that given D, $R_n(P_n, Q_n, D)$ is an approximately subadditive sequence. To this end, given $\epsilon \in (0, (D - D_{\min})/4)$, as n is large enough, $D_{\min}^{(n)} < D_{\min} + \epsilon < D - 3\epsilon$, which implies that as $\lambda \to -\infty$

$$\lambda D - \Lambda_n(\lambda) \sim \lambda \left(D - D_{\min}^{(n)}\right) \to -\infty.$$

Then $\lambda D - \Lambda_n(\lambda)$ can achieve its supremum $\Lambda_n^*(D)$ on $(-\infty, 0]$. Let $S_n = \{\lambda \le 0 : \lambda D - \Lambda_n(\lambda) = \Lambda_n^*(D)\}$. We show that for some $N \ge 1$, $\bigcup_{n \ge N} S_n$ is bounded. From (2.2), it is seen that

$$\lim_{n \to \infty} [s/n] = 1/(k+d).$$

Fix k and N large enough, such that $D_{\min}^{(k)} < D_{\min} + \epsilon$, and for all $n = s(k+d) + r \ge N$, with $0 \le r < k+d$, there is $(sd+r)D_{\max} \le \epsilon n$. Since $\Lambda_k(\lambda) \le skn^{-1}\Lambda_k(\lambda) \le 0$, (2.3) leads to

$$\begin{aligned}\lambda D - \Lambda_n(\lambda) &\le \lambda D - \frac{sk}{n}\Lambda_k(\lambda) + \frac{s}{n}\log C - \frac{(sd+r)D_{\max}\lambda}{n} \\ &\le \lambda D - \Lambda_k(\lambda) + \frac{s}{N}\log C - \epsilon\lambda. \end{aligned}\tag{2.6}$$

If $\lambda < 0$ has large enough absolute value, then the right-hand side of (2.6) is bounded from above by

$$\lambda D - \lambda\left(D_{\min}^{(k)} + \epsilon\right) - \lambda\epsilon \le \lambda D - \lambda(D_{\min} + 2\epsilon) - \lambda\epsilon = \lambda(D - D_{\min} - 3\epsilon) < 0.$$

Together with (2.6), this implies that there is $\lambda_0 < 0$, such that for all $n \ge N$

$$\lambda_0 D - \Lambda_n(\lambda_0) < 0. \tag{2.7}$$

Since $\Lambda_n^*(D) \ge 0$, this shows that for all $n \ge N$, $S_n \subset [\lambda_0, 0]$, and hence $\bigcup_{n \ge N} S_n$ is bounded.

By (2.1), we get

$$\begin{aligned}&(m+n)(\lambda D - \Lambda_{m+n}(\lambda)) \\ &\quad\le m(\lambda D - \Lambda_m(\lambda)) + n(\lambda D - \Lambda_n(\lambda)) + \log C + dD_{\max}|\lambda|.\end{aligned}$$

Because $\bigcup_{n \ge N} S_n$ is bounded, and

$$R_k(P_k, Q_k, D) = \min_{\lambda \le 0}[\lambda D - \Lambda_k(\lambda)]$$

for all $k \ge 1$, letting r_0 be an upper bound of $dD_{\max}|\lambda|$ over $\bigcup_{n \ge N} S_n$, the above inequality implies that for $m, n \ge N$

$$\begin{aligned}&(m+n)R_{m+n}(P_{m+n}, Q_{m+n}, D) \\ &\quad\le mR_m(P_m, Q_m, D) + nR_n(P_n, Q_n, D) + r_0 + |\log C|.\end{aligned}$$

Therefore, $nR_n(P_n, Q_n, D)$, as a sequence in n, is approximately subadditive. By Hammersley's lemma on approximate subadditivity, $R_n(P_n, Q_n, D)$ is convergent, and the limit $R(P, Q, D) \ge 0$ is finite.

Because all $R_n(P_n, Q_n, D)$ are convex in D, so is $R(P, Q, D)$. Then, as $R(P, Q, D)$ is finite, it is continuous. Therefore, by Lemma 3, $R_n(P_n, Q_n, D) \xrightarrow{\text{u.c.}} R(P, Q, D)$.

Finally, for $D \ge D_{\text{av}}$, by the discussion following (1.11), $R_n(P_n, Q_n, D) = 0$, $n \ge 1$, implying $R(P, Q, D) = 0$. □

We now prove the lemmas given at the beginning of this section. First we introduce some notation. Given a finite set $J \subset \boldsymbol{N}$, for $a = \{a_n, n \ge 1\}$, denote $a_J = \{a_i, i \in J\}$. Denote

$$\rho(J) = \rho(x_J, y_J) = \sum_{j \in J} \rho(x_j, y_j). \tag{2.8}$$

Furthermore, denote by P_J and Q_J the marginal distributions of X_J and Y_J, respectively.

Proof of Lemma 2: Given $m, n \ge 1$ with $m + nd$, let

$$I = \{1, \ldots, m\} \qquad J = \{m + d + 1, \ldots, m + n\}.$$

Then $(P_{m+n} \times Q_{m+n})$-almost everywhere on $\mathcal{A}_X^{m+n} \times \mathcal{A}_Y^{m+n}$, there is

$$\sum_{i=1}^{m+n} \rho(x_i, y_i) \le \rho(I) + \rho(J) + dD_{\max}.$$

Since $\lambda < 0$, then

$$\begin{aligned}&(m+n)\Lambda_{m+n}(\lambda) \\ &\quad= \int \log\left\{\int e^{[\lambda \sum_{i=1}^{m+n} \rho(x_i, y_i)]}\, dQ_{m+n}(y_1^{m+n})\right\} dP_{m+n}(x_1^{m+n}) \\ &\quad\ge \int \log\left\{\int e^{[\lambda\rho(I) + \lambda\rho(J) + \lambda dD_{\max}]}\, dQ_{m+n}(y_1^{m+n})\right\} \\ &\qquad\qquad \cdot dP_{m+n}(x_1^{m+n}) \\ &\quad= \int \log\left\{\int e^{\lambda\rho(I) + \lambda\rho(J)}\, dQ_{I \cup J}(y_I, y_J)\right\} dP_{m+n}(x_1^{m+n}) \\ &\qquad + dD_{\max}\lambda \\ &\quad\overset{(a)}{\ge} \int \log\left\{\int e^{\lambda\rho(I) + \lambda\rho(J)} C^{-1}\, dQ_I(y_I)\, dQ_J(y_J)\right\} dP_{m+n}(x_1^{m+n}) \\ &\qquad + dD_{\max}\lambda \\ &\quad= \int \log\left\{\int e^{\lambda\rho(I)}\, dQ_I(y_I)\right\} dP_I(x_I) \\ &\qquad + \int \log\left\{\int e^{\lambda\rho(J)}\, dQ_I(y_J)\right\} dP_J(x_J) \\ &\qquad - \log C + dD_{\max}\lambda \\ &\quad\overset{(b)}{=} m\Lambda_m(\lambda) + (n-d)\Lambda_{n-d}(\lambda) - \log C + dD_{\max}\lambda \end{aligned}\tag{2.9}$$

where (a) is due to condition (ψ_-), and (b) the stationarity of both X and Y. On the other hand, because $\lambda < 0$ and $\rho(x, y) \ge 0$, it is not hard to see that $n\Lambda_n(\lambda) \le (n-d)\Lambda_{n-d}(\lambda)$, which, together with (2.9), implies (2.1).

As to (2.3), by (2.9) and induction, it is easy to prove that, if $n = s(k+d) + r$, with $0 \le r < k + d$

$$n\Lambda_n(\lambda) \ge sk\Lambda_k(\lambda) - s\log C + s\, dD_{\max}\lambda + r\Lambda_r(\lambda).$$

Because $r\Lambda_r(\lambda) \ge rD_{\max}\lambda$, then by the above inequality, (2.3) is proved. □

Proof of Lemma 3: Given $M > 0$, let $a = \Lambda(M+1) - \Lambda(M)$, $b = \Lambda(-M) - \Lambda(-M-1)$. Since $\Lambda_n(\lambda) \to \Lambda(\lambda)$ pointwise

$$\Lambda_n(M+1) - \Lambda_n(M) \to a, \quad \Lambda_n(-M) - \Lambda_n(-M-1) \to b.$$

Let $c=\max\{|a|,\,|b|\}+1$. Then for large n, $|\Lambda_n(M+1)-\Lambda_n(M)|\le c$, and $|\Lambda_n(-M)-\Lambda_n(-M-1)|\le c$. Since Λ_n are convex, then for any $y,\,z\in[-M,\,M]$, $|\Lambda_n(y)-\Lambda_n(z)|\le c|y-z|$. This shows that Λ_n are equicontinuous on $[-M,\,M]$, which implies uniform convergence to Λ on $[-M,\,M]$. □

Proof of Lemma 4: First, for any $m,\,n\ge 1$, it is easy to see that

$$(m+n)\operatorname*{ess\,inf}_{Y_1^n}\rho_{m+n}(x_1^{m+n},\,Y_1^{m+n}) \ge m\operatorname*{ess\,inf}_{Y_1^m}\rho_m(x_1^m,\,Y_1^m)+n\operatorname*{ess\,inf}_{Y_1^n}\rho_n(x_1^n,\,Y_1^n).$$

Take expectation on both sides to get

$$(m+n)D_{\min}^{(m+n)}\ge mD_{\min}^{(m)}+nD_{\min}^{(n)}.$$

Then $-nD_{\min}^{(n)}$ satisfies approximate subadditivity, and hence $D_{\min}=\lim_{n\to\infty}D_{\min}^{(n)}$ exists. Clearly, $0\le D_{\min}<\infty$.

We show that Λ_n converge pointwise on $\boldsymbol{R}$. Given $\lambda\in\boldsymbol{R}$, let $n\to\infty$ in (2.3). By (2.2), $\lim_{n\to\infty}[s/n]=1/(k+d)$. Then there is

$$\liminf_{n\to\infty}\Lambda_n(\lambda)\ge\frac{k}{k+d}\,\Lambda_k(\lambda)-\frac{\log C}{k+d}+\frac{dD_{\max}\lambda}{k+d}.$$

Let $k\to\infty$ to get

$$\liminf_{n\to\infty}\Lambda_n(\lambda)\ge\limsup_{k\to\infty}\Lambda_k(\lambda).$$

Therefore, $\Lambda(\lambda)=\lim_{n\to\infty}\Lambda_n(\lambda)$ exists. Since Λ_n is convex on $\boldsymbol{R}$ and $|\Lambda_n(\lambda)|\le|\lambda|D_{\max}$, $\Lambda(\lambda)$ is convex and $|\Lambda(\lambda)|\le|\lambda|D_{\max}$ as well. Then by Lemma 3, Λ is continuous and $\Lambda_n\xrightarrow{\text{u.c.}}\Lambda$.

To prove (2.5), let $f_n(\lambda)=\Lambda_n(\lambda)/\lambda$ and $f(\lambda)=\Lambda(\lambda)/\lambda$. It is easy to see that both f_n and f are nondecreasing. Let $L=\lim_{\lambda\to-\infty}f(\lambda)$. It is clear that $L>-\infty$. Given $\epsilon>0$, fix $\lambda_0<0$ such that $f(\lambda_0)<L+\epsilon$. For n large enough and $\lambda<\lambda_0$

$$f_n(\lambda)\le f_n(\lambda_0)<f(\lambda_0)+\epsilon\le L+2\epsilon.$$

Let $\lambda\downarrow-\infty$. By $f_n(\lambda)\downarrow D_{\min}^{(n)}$, there is $D_{\min}^{(n)}\le L+2\epsilon$, which, by Lemma 4, implies $D_{\min}\le L$.

To prove $L\le D_{\min}$, divide (2.3) by λ, with $\lambda<0$, to get

$$\begin{aligned}f_n(\lambda)&\le\frac{sk}{n}\,f_k(\lambda)+\frac{s\log C}{n|\lambda|}-\frac{(sd+r)D_{\max}}{n}\\&\le f_k(\lambda)+\frac{s\log C}{n|\lambda|}+\frac{(sd+r)D_{\max}}{n}.\end{aligned}$$

Noting that $\lambda<0$, $f_k(\lambda)=\Lambda_k(\lambda)/\lambda\ge 0$, let $n\to\infty$ to get

$$f(\lambda)\le f_k(\lambda)+\frac{\log C}{(k+d)|\lambda|}+\frac{dD_{\max}}{k+d}.$$

Letting $\lambda\to-\infty$, then $k\to\infty$, there is $L\le D_{\min}$, which completes the proof. □

III. The First-Order Asymptotic of $\log Q(B(X_1^n,\,D))$

In this section the goal is to prove Theorem 2. That is, given two independent stationary processes X and Y, if X is ergodic and Y satisfies both conditions (ψ_-) and (ψ_+), then for $D\in(D_{\min},\,\infty)$, there is

$$\lim_{n\to\infty}\frac{1}{n}\log Q(B(X_1^n,\,D))=-R(P,\,Q,\,D),\qquad P\text{–a.s.}.\tag{3.1}$$

We first show that if (3.1) holds for $D\in(D_{\min},\,D_{\rm av})$, then it also holds for $D\ge D_{\rm av}$. Indeed, given any $D'\in(D_{\min},\,D_{\rm av})$, by

$$Q(B(X_1^n,\,D'))\le Q(B(X_1^n,\,D))$$

it is easy to see if (3.1) holds for D', then there is

$$\limsup_{n\to\infty}|\log Q(B(X_1^n,\,D))/n|\le R(P,\,Q,\,D').$$

Letting $D'\to D_{\rm av}$, by the continuity of $R(P,Q,D)$, and $R(P,\,Q,\,D_{\rm av})=0$, there is then

$$\lim_{n\to\infty}[\log Q(B(X_1^n,\,D))/n]=0.$$

Inequality (3.1) remains to be shown for $D\in(D_{\min},\,D_{\rm av})$. It is enough to show that

$$\limsup_{n\to\infty}\frac{1}{n}\log Q(B(X_1^n,\,D))\le-R(P,\,Q,\,D),\qquad P\text{–a.s.}\tag{3.2}$$

$$\liminf_{n\to\infty}\frac{1}{n}\log Q(B(X_1^n,\,D))\ge-R(P,\,Q,\,D),\qquad P\text{–a.s.}\tag{3.3}$$

We will prove (3.2) and (3.3) in subsequent subsections. Before doing this, we fix notation. From now on $d\ge 1$ is fixed and assumed to satisfy conditions (ψ_-) and (ψ_+) for Y. Define $c(n,\,k)$ and $e(n,\,k)$ on $\boldsymbol{N}\times\boldsymbol{N}$ such that for $n=s(k+d)+r$, $0\le r<k+d$

$$c(n,\,k)=\left\lfloor\frac{n}{k+d}\right\rfloor\qquad e(n,\,k)=\left\lfloor\frac{n}{k+d}\right\rfloor d+r.$$

Note that as $n\to\infty$, $e(n,\,k)/n\to d/k$. Therefore, for any $\epsilon>0$, if k is large enough, then for all n large enough, $e(n,\,k)/n<\epsilon$. We will continue to use the notation in (2.8). If $I\subset\boldsymbol{N}$, and $j\ge 0$, then define $I+j=\{i+j,\,i\in I\}$. Also, if $a,\,b\in\boldsymbol{N}$, and $a\le b$, then denote $[a,\,b]=\{a,\,a+1,\,\ldots,\,b\}$.

A. Proof of the Upper Bound (3.2)

Given $\epsilon\in(0,\,D_{\rm av}-D)$, fix $k>d$, such that $D_{\min}^{(k)}<D$, and $d+\log C<\epsilon k$. For all $t\ge 1$, let

$$I_t=I_t(k)=\{j\colon (t-1)(k+d)+1\le j\le tk+td\}.\tag{3.4}$$

Then $I_1=[1,\,k]$, and $I_t=I_{t-1}+k+d$. Given n large enough, such that $s=c(n,\,k)\ge 2$, for $j=0,\,\ldots,\,k-1$, $I_1+j,\,\ldots,\,I_s+j$ are disjoint subsets in $[1,\,n]$. Then by

$$\rho([1,\,n])\ge\sum_{t=1}^{s-1}\rho(I_t+j)$$

there is $B(x_1^n,\,D)\subset A_j$, where

$$A_j=\left\{y_1^n\in\mathcal{A}_Y^n\colon\,\sum_{t=1}^{s-1}\rho(x_{I_t+j},\,y_{I_t+j})<nD\right\}.$$

Since the distance between I_t+j and $I_{t'}+j$ is at least d, for $t\ne t'$, by condition (ψ_+)

$$\begin{aligned}Q(B(x_1^n,\,D))&\le Q(A_j)\\&\le C^{s-1}Q_k^{s-1}\Bigg\{(z_0,\,\ldots,\,z_{s-2})\colon z_t\in\mathcal{A}_Y^k,\\&\qquad\sum_{t=1}^{s-1}\rho(x_{I_t+j},\,z_t)<nD\Bigg\}.\end{aligned}\tag{3.5}$$

Let

$$\mathcal{I}=\{I_t+j,\ t=1,\,\ldots,\,s-1,\ j=0,\,1,\,\ldots,\,k-1\}.$$

Then $\mathcal{I}$ contains $k(s-1)$ intervals of length k. Enumerate $\mathcal{I}$ such that its ith element is $[a_i, a_i+k-1]$, with a_i increasing. It is easy to see that $a_i \in I_1 \cup \ldots \cup I_{s-1}$. Then (3.5) leads to

$$\begin{aligned} Q(B(x_1^n, D)) &\le [Q(A_0)\ldots Q(A_{k-1})]^{1/k} \\ &\le C^{s-1}\left(\prod_{j=0}^{k-1} Q_k^{s-1}\left\{(z_1, \ldots, z_{s-1})\colon z_t \in \mathcal{A}_Y^k,\right.\right. \\ &\qquad \left.\left.\sum_{t=1}^{s-1}\rho(x_{I_t+j}, z_t) < nD\right\}\right)^{1/k} \\ &\le C^{s-1}\left(Q_k^{k(s-1)}\left\{(z_1, \ldots, z_{|\mathcal{I}|})\colon z_i \in \mathcal{A}_Y^k,\right.\right. \\ &\qquad \left.\left.\frac{1}{k}\sum_{i=1}^{|\mathcal{I}|}\rho(x_{I_i}, z_i) \le nD\right\}\right)^{1/k}. \end{aligned} \tag{3.6}$$

Let $\mathcal{J} = \{[x, x+k-1], x \in [1, n]\}$. Clearly, $\mathcal{I} \subset \mathcal{J}$ and $|\mathcal{J}\backslash\mathcal{I}| = e(n, k)+k$. Then by (3.6), it is easy to see that P-almost surely

$$\begin{aligned} Q(B(x_1^n, D)) \le C^{s-1}&\left(Q_k^n\left\{(z_1, \ldots, z_n)\colon z_i \in \mathcal{A}_Y^k,\right.\right. \\ &\frac{1}{nk}\sum_{i=1}^{n}\rho(x_i^{i+k-1}, z_i) \le D \\ &\left.\left.+\ \frac{(e(n, k)+k)D_{\max}}{n}\right\}\right)^{1/k} \end{aligned} \tag{3.7}$$

Define a process $U = \{U_i, i \ge 1\}$ such that

$$U_i = (X_i, \ldots, X_{i+k-1}) \in \mathcal{A}_U = \mathcal{A}_X^k.$$

Since X is stationary and ergodic, so is U, and $U_i \sim P_k$. Define an i.i.d. process $V = \{V_i, i \ge 1\}$, such that $V_i \in \mathcal{A}_V = \mathcal{A}_Y^k$ and $V_i \sim Q_k$. Denote the probability distributions of U and V by $\tilde{P}$ and $\tilde{Q}$, respectively. Finally, let $\tilde{\rho} = \rho_k$ be the function on $\mathcal{A}_U \times \mathcal{A}_V$ whereby the distortion measures for $u_1^n \in \mathcal{A}_U^n$ and $v_1^n \in \mathcal{A}_V^n$ are defined.

By the choice of k and (3.7), for n large enough

$$Q(B(X_1^n, D)) \le C^{s-1}\left(\tilde{Q}(B(U_1^n, D+\epsilon))\right)^{1/k}$$

where the ball $B(U_1^n, D+\epsilon)$ is defined according to $\tilde{\rho}_n$. Then

$$\frac{1}{n}\log Q(B(X_1^n, D)) \le \epsilon - \frac{1}{kn}\log \tilde{Q}(B(U_1^n, D+\epsilon)), \qquad P\text{–a.s.} \tag{3.8}$$

From the construction of U and V,

$$E\tilde{\rho}(U_1, V_1) = D_{\text{av}} \quad \text{and} \quad E_{\tilde{P}}[\text{ess inf}_{V_1}\tilde{\rho}(U_1, V_1)] = D_{\min}^{(k)}.$$

Because 1) $D+\epsilon \in (D_{\min}^{(k)}, D_{\text{av}})$, 2) U is stationary ergodic, and 3) V is an i.i.d. process, [2, Corollary 1] or [3, Corollary 1] applies to U, V, and $\tilde{Q}(B(U_1^n, D+\epsilon))$. Then, as $n \to \infty$, $\tilde{P}$-almost surely, the right-hand side of (3.8) tends to $-\epsilon + k^{-1}R_1(\tilde{P}_1, \tilde{Q}_1, D+\epsilon)$. Thus P-almost surely

$$\begin{aligned} \limsup_{n\to\infty}\frac{1}{n}\log Q(B(X_1^n, D)) &\le \epsilon - k^{-1}R_1(\tilde{P}_1, \tilde{Q}_1, D+\epsilon) \\ &= \sup_{\lambda \le 0}\{\lambda(D+\epsilon) - \tilde{\Lambda}(\lambda)\} \end{aligned} \tag{3.9}$$

where $\tilde{\Lambda}(\lambda) = \int \log\left\{\int e^{\lambda\tilde{\rho}(u, v)}\, d\tilde{Q}_1(u)\right\} d\tilde{P}_1(v)$. Because

$$\begin{aligned} \tilde{\Lambda}(\lambda) &= \int \log\left\{\int \exp\left(\frac{\lambda}{k}\sum_{i=1}^{k}\rho(x_i, y_i)\right) dQ_k(y_1^k)\right\} dP_k(x_1^k) \\ &= k\Lambda_k(\lambda/k) \end{aligned}$$

there is

$$R_1(\tilde{P}_1, \tilde{Q}_1, D+\epsilon) = \sup_{\lambda \le 0}\{\lambda(D+\epsilon) - k\Lambda_k(\lambda/k)\} = k\Lambda_k^*(D+\epsilon)$$

and hence

$$k^{-1}R_1(\tilde{P}_1, \tilde{Q}_1, D+\epsilon) = R_k(P_k, Q_k, D+\epsilon). \tag{3.10}$$

Let $k \to \infty$ and then $\epsilon \to 0$. Because $R(P, Q, D)$ is continuous in D, by (3.9) and (3.10), and (3.2) is proved.

B. Proof of the Lower Bound (3.3): A Special Case

In this subsection we prove the asymptotic lower bound (3.3) for a special case, where (3.13) is satisfied. Then in the next subsection, we prove the lower bound for the general case, which can be reduced to the special case by a random perturbation argument. The proof is based on the following two lemmas.

Lemma 5: For $k \ge 1$ and random sample x from X, denote $u_t = x_{I_t}$, where I_t is given by (3.4). Let $\hat{P}_{n,k}$ be the empirical measure of $\mathcal{A}_X^k$ induced by $u_1, \ldots, u_n$. Define the following empirical log-moment generating function with respect to Q_k and $\hat{P}_{n,k}$:

$$\begin{aligned} \Lambda_{\hat{P}_{n,k}}(\lambda) &= \frac{1}{kn}\sum_{s=1}^{n}\log\left\{\int_{\mathcal{A}_Y^k} e^{\lambda\rho(u_s, z)}\, dQ_k(z)\right\} \\ &= \frac{1}{k}\int_{\mathcal{A}_X^k}\log\left\{\int_{\mathcal{A}_Y^k} e^{\lambda\rho(u, z)}\, dQ_k(z)\right\} d\hat{P}_{n,k}(u). \end{aligned} \tag{3.11}$$

Given $D \in (D_{\min}, D_{\text{av}})$, let

$$\lambda_{n,k} = \arg\max_{\lambda \in \boldsymbol{R}}\{\lambda D - \Lambda_{\hat{P}_{n,k}}(\lambda)\}. \tag{3.12}$$

Then P-almost surely, for k is large enough, there are

1) for all n large enough, $\lambda_{n,k}$ exists, is nonpositive, and $|\lambda_{n,k}|$ is bounded away from ∞;
2) $\liminf_{n\to\infty}|\lambda_{n,k}| > 0$;
3) $\limsup_{n\to\infty}\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k}) < \infty$.

Lemma 6: Given X and Y as above

$$\limsup_{k\to\infty}\limsup_{n\to\infty}\Lambda^*_{\hat{P}_{n,k}}(D) \le R(P, Q, D).$$

The lemmas will be proved in Section III-D. Assuming they are true for now, we prove (3.3) under the following additional condition:

$$\liminf_{n\to 0}\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k}) > 0. \tag{3.13}$$

Given $\epsilon \in (0, D-D_{\min})$, fix $k \ge 1$ with $1+D_{\max} < \epsilon k$. Given N with $(1+D_{\max})e(N, k) < \epsilon N$, let $n = c(N, k)$. Then on $\mathcal{A}_X^N \times \mathcal{A}_Y^N$

$$\begin{aligned} \rho(x_1^N, y_1^N) &\le \sum_{t=1}^{n}\rho(x_{I_t}, y_{I_t}) + e(N, k)D_{\max} \\ &\le \sum_{t=1}^{n}\rho(x_{I_t}, y_{I_t}) + \epsilon N \quad (P \times Q)\text{–almost everywhere} \end{aligned}$$

where I_t is given by (3.4). Therefore, by $nk \le N$, there is

$$B(x_1^N, D) \supset \left\{ y_1^N \in \mathcal{A}_Y^N \colon \sum_{t=1}^n \rho(x_{I_t}, y_{I_t}) \le (D-\epsilon)kn \right\}. \quad (3.14)$$

Let V be the same i.i.d. process as in the proof of the upper bound and $\tilde{Q}$ its distribution. Write $u_t = x_{I_t}$. Then by (3.14) and condition (ψ_-)

$$Q(B(x_1^N, D)) \ge C^{-n} \tilde{Q}(B(u_1^n, (D-\epsilon)k)). \quad (3.15)$$

Note that $(u_1, u_2, \ldots)$ is a random sample of process $U = \{U_n, n \ge 1\}$, with $U_n = X_{I_n}$. In general, U is neither stationary nor ergodic. Despite this, the argument of [2] still applies to $(u_1, u_2, \ldots)$ and V. To be specific, let $\zeta_t = \rho(u_t, V_t)$, $T_n = \sum_{t=1}^n \zeta_t$, and $\hat{T}_n = T_n/n$. Denote by μ_n the law of $(\zeta_1, \ldots, \zeta_n)$. Then

$$\tilde{Q}(B(u_1^n, (D-\epsilon)k)) = \Pr(\hat{T}_n \le (D-\epsilon)k).$$

Noting that $\hat{T}_n$ depends on x, we will show that P-almost surely, for k large enough, there exists $K = K(k, x) > 0$, such that for $\delta > 0$ and

$$J_n = e^{nk\Lambda^*_{\hat{P}_{n,k}}(D-\epsilon)} \Pr(\hat{T}_n \le (D-\epsilon)k)$$

there is

$$\log J_n \ge -K\delta\sqrt{n}. \quad (3.16)$$

Once (3.16) is proved, it follows that for n large enough

$$e^{nk\Lambda^*_{\hat{P}_{n,k}}(D-\epsilon)} \tilde{Q}(B(u_1^n, (D-\epsilon)k)) \ge e\left(-K\delta\sqrt{n}\right),$$

and, therefore, by (3.15)

$$\frac{1}{N} \log Q(B(x_1^N, D)) \ge -\frac{c(N,k)C}{N} - \frac{K\delta\sqrt{n}}{N} - \frac{nk}{N}\Lambda^*_{\hat{P}_{n,k}}(D-\epsilon).$$

Let $N \to \infty$ then $k \to \infty$. By Lemma 6

$$\liminf_{N\to\infty} \frac{1}{N} \log Q(B(x_1^N, D)) \ge -R(P, Q, D-\epsilon).$$

Let $\epsilon \to 0$ and apply Theorem 1 to complete the proof of (3.3).

To prove (3.16), by $D - \epsilon \in (D_{\min}, D_{\text{av}})$ and Lemma 5, if k is chosen large enough, then P-almost surely,

$$\lambda_{n,k} = \arg\max_{\lambda \in \boldsymbol{R}} \{\lambda(D-\epsilon) - \Lambda_{\hat{P}_{n,k}}(\lambda)\}$$

exists for all n large enough, and $\lambda_{n,k} \le 0$. Define a probability measure on $[0, \infty)^n$

$$d\nu_n(\zeta_1^n) = \exp\left\{\lambda_{n,k}\sum_{t=1}^n \zeta_t - nk\Lambda_{\hat{P}_{n,k}}(\lambda_{n,k})\right\} d\mu_n(\zeta_1^n) \quad (3.17)$$

with $\Lambda_{\hat{P}_{n,k}}(\lambda)$ being a log-moment generating function. It is easy to check that under ν_n, for $T_n = \sum_{t=1}^n \zeta_t$, $E_{\nu_n}[T_n] = nk\Lambda'_{\hat{P}_{n,k}}(\lambda_{n,k})$ and

$$\text{Var}_{\nu_n}[T_n] = nk\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k}).$$

On the other hand, because $\lambda_{n,k}$ is a maximal point of $(D-\epsilon)\lambda - \Lambda_{\hat{P}_{n,k}}(\lambda)$, there is $D - \epsilon = \Lambda'_{\hat{P}_{n,k}}(\lambda_{n,k})$, and hence $E_{\nu_n}[T_n] = (D-\epsilon)nk$. Let

$$G_n = -\frac{T_n - (D-\epsilon)nk}{\sqrt{nk\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k})}}.$$

Then by the above results, when $\zeta_1^n \sim \nu_n$, G_n has mean 0 and variance 1, and hence by (3.17)

$$\begin{aligned} J_n &= e^{nk\Lambda^*_{\hat{P}_{n,k}}(D-\epsilon)} \\ &\quad \cdot E_{\nu_n}\left\{1_{\{\hat{T}_n \le (D-\epsilon)k\}} e^{-n\lambda_{n,k}\hat{T}_n + nk\Lambda_{\hat{P}_{n,k}}(\lambda_{n,k})}\right\} \\ &= E_{\nu_n}\left\{1_{\{G_n \ge 0\}} e^{\left[\lambda_{n,k}\sqrt{nk\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k})}\, G_n\right]}\right\} \\ &\ge E_{\nu_n}\left\{1_{\{0 < G_n < \delta\}} e^{\left(-\beta_n\sqrt{n}G_n\right)}\right\} \\ &\ge e^{\left(-\beta_n\sqrt{n}\,\delta\right)} \Pr_{\nu_n}(0 < G_n < \delta) \end{aligned}$$

for any $\delta > 0$, where

$$\beta_n = \lambda_{n,k}\sqrt{k\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k})} > 0.$$

Due to the boundedness of ρ, the random variables ζ_t are uniformly bounded. Therefore, by the assumption (3.13), it is seen that the Linderberg condition for the central limit theorem is satisfied by G_n, and hence

$$\lim_{n\to\infty} \Pr_{\nu_n}(0 < G_n < \delta) = \rho > 0.$$

Lemma 5 also implies with probability 1, for n large enough, β_n is bounded away from ∞. Choose $K > 0$ so that $K - \beta_n$ is bounded away from 0, then

$$\liminf_{n\to\infty} \log[e^{(K\sqrt{n}\delta)} J_n] \ge \log \rho > -\infty$$

implying

$$\liminf_{n\to\infty} \sqrt{n}\left[K\delta + \frac{1}{\sqrt{n}} \log J_n\right] > -\infty$$

which holds only if

$$\liminf_{n\to\infty} \left[\frac{1}{\sqrt{n}} \log J_n\right] > -K\delta.$$

This complete the proof of (3.16). □

C. Proof of the Lower Bound (3.3): The General Case

Now we prove (3.3) without assuming the condition (3.13). Let Y be an arbitrary stationary process satisfying conditions (ψ_-) and (ψ_+). Given $\epsilon \in (0, \min\{D - D_{\min}, D_{\text{av}} - D\}/2)$, let $Z = (Y, W)$, with process $W = \{W_n, n \ge 1\}$ an i.i.d. process independent of Y such that W_n has uniform distribution on $[0, \epsilon]$. Then Z is a process defined on $\mathcal{A}_Z = \mathcal{A}_Y \times [0, \epsilon]$. Note that Z also satisfies conditions (ψ_-) and (ψ_+). Let $\hat{Q}$ be the distribution of Z.

Define $\hat{\rho}$ on $\mathcal{A}_X \times \mathcal{A}_Z$, such that for $x \in \mathcal{A}_X$ and $z = (y, u) \in \mathcal{A}_Z$, $\hat{\rho}(x, z) = \rho(x, y) + u$. Then define

$$\hat{B}(x_1^n, D) = \{z_1^n \in \mathcal{A}_Z^n \colon \hat{\rho}_n(x_1^n, z_1^n) \le D\}.$$

With Y being replaced by Z, ρ by $\hat{\rho}$, and Q by $\hat{Q}$, (3.11) becomes

$$\begin{aligned} \Lambda_{\hat{P}_{n,k}}(\lambda) &= \frac{1}{kn}\sum_{t=1}^n \log\left\{\int_{\mathcal{A}_Z^k} e^{\lambda\hat{\rho}(u_t, z)}\, d\hat{Q}_k(z)\right\} \\ &= \frac{1}{kn}\sum_{t=1}^n \log\left\{\int_{\mathcal{A}_Y^k} e^{\lambda\rho(u_t, v)}\, dQ_k(v)\right\} \\ &\quad + \frac{1}{kn}\sum_{t=1}^n \log\left\{\int_{[0,1]^k} e^{\lambda\epsilon(w_1 + \cdots + w_k)}\, dw_1 \cdots dw_k\right\} \\ &= F(\lambda) + \log\left(\int_0^1 e^{\lambda\epsilon u}\, du\right) \\ &= F(\lambda) + G(\lambda). \end{aligned}$$

It is easy to see that $F''(\lambda) \geq 0$ and $G''(\lambda) > 0$. Therefore, $\Lambda''_{\hat{P}_{n,k}}(\lambda) > 0$, for all λ. By Lemma 5 (1), given $D \in (D_{\min}, D_{\text{av}})$, P-almost surely, for k large enough

$$\lambda_{n,k} = \operatorname*{argsup}_{\lambda \leq 0}\{\lambda D - \Lambda_{\hat{P}_{n,k}}(\lambda)\}$$

exists for all n large enough and bounded from $-\infty$. Therefore, $G''(\lambda_{n,k})$ is bounded below from 0, and hence

$$\liminf_{n\to\infty} \Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k}) > 0$$

i.e., for the triple $\{X, Z, \rho\}$, the condition (3.13) is satisfied.

By the last subsection, for $D' \in (D_{\min}, D_{\text{av}})$

$$\liminf_{n\to\infty} [\log \hat{Q}(\hat{B}(x_1^n, D'))/n] \geq -R(P, \hat{Q}, D')$$

P-almost surely. On the other hand, because $Z = (Y, W)$, with Y and W independent and $W_n \in [0, \epsilon]$

$$\hat{Q}(\hat{B}(x_1^n, D)) \leq Q(B(x_1^n, D)).$$

Then it is seen that

$$\liminf_{n\to\infty} \frac{1}{n} \log Q(B(X_1^n, D)) \geq -R(P, \hat{Q}, D). \tag{3.18}$$

For $n \geq 1$, letting

$$\hat{\Lambda}_n(\lambda) = [1/n] \int_{\mathcal{A}_X^n} \log \left\{ \int_{A_Z^n} e^{\lambda \hat{\rho}(x_1^n, z_1^n)} \, d\hat{Q}_n(z_1^n) \right\} dP_n(x_1^n)$$

we get

$$R_n(P_n, \hat{Q}_n, D) = \sup_{\lambda \leq 0}\{\lambda D - \hat{\Lambda}_n(\lambda)\}. \tag{3.19}$$

By

$$\hat{\rho}(x_1^n, z_1^n) = \rho(x_1^n, y_1^n) + \epsilon \sum_{i=1}^n w_i$$

and $\hat{Q} = Q \times \nu$, with ν the uniform distribution on $[0, \epsilon]^N$, it is easy to see that

$$\lambda D - \hat{\Lambda}_n(\lambda) \leq \lambda(D - \epsilon) - \Lambda_n(\lambda), \qquad \text{for all } \lambda \leq 0.$$

Since $D - \epsilon > D_{\min}$, then for n large enough, the supremum in (3.19) as well as that of $\lambda(D-\epsilon) - \Lambda_n(\lambda)$ is achievable for $\lambda \leq 0$, leading to $R_n(P_n, \hat{Q}_n, D) \leq R_n(P_n, Q_n, D - \epsilon)$. Therefore, letting $n \to \infty$, there is $R(P, \hat{Q}, D) \leq R(P, Q, D - \epsilon)$. Then by (3.18)

$$\liminf_{n\to\infty} \frac{1}{n} \log Q(B(X_1^n, D)) \geq -R(P, Q, D - \epsilon).$$

Let $\epsilon \to 0$ and apply the continuity of $R(P, Q, D)$ to complete the proof of (3.3) for the general case. □

D. Proofs of Lemmas 5 and 6

Proof of Lemma 5:

1) Fix $\epsilon \in (0, 1/2)$ small enough, and integers h, k large enough, so that

$$\epsilon k \geq (1-\epsilon)h \qquad e(h, k)D_{\max} < \epsilon k \qquad c(k, h) \log C < \epsilon k$$
$$(D - \epsilon)(1 - \epsilon) > D_{\min}^{(h)}. \tag{3.20}$$

First we show that P-almost surely, with k given as above, for n large enough, $\Lambda'_{\hat{P}_{n,k}}(0) < D$. Define

$$E_n = (kn)^{-1} \sum_{t=1}^n \int \rho(u_t, z) \, dQ_k(z).$$

Then $\Lambda'_{\hat{P}_{n,k}}(0) = E_n$. Let $S_n = I_1 \cup \ldots \cup I_n$. Then $S_n \subset [1, b_n]$, with $b_n = nk + (n-1)d$. Noting

$$E_n = (kn)^{-1} \sum_{i \in S_n} \int \rho(x_i, y) \, dQ_1(y)$$

by the choice of k, for n large enough

$$\begin{aligned} E_n - \frac{1}{b_n} \sum_{i=1}^{b_n} \int \rho(x_i, z) \, dQ_1(z) &\geq -\frac{1}{b_n} \sum_{i \in [1, b_n] \setminus S} \int \rho(x_i, z) \, dQ_1(z) \\ &\geq -\frac{e(b_n, k) D_{\max}}{b_n} > -\epsilon. \end{aligned}$$

Because X is ergodic, P-almost surely

$$\liminf_{n\to\infty} E_n \geq \lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^{b_n} \int \rho(x_i, z) \, dQ_1(z) - \epsilon = D_{\text{av}} - \epsilon > D \tag{3.21}$$

and hence $\Lambda'_{\hat{P}_{n,k}}(0) > D$. Because $\lambda D - \Lambda_{\hat{P}_{n,k}}(\lambda)$ is concave, it is then seen that for n large enough, if $\lambda_{n,k}$ exists, then it has to be nonpositive.

Next we show the existence as well as the boundedness of $\lambda_{n,k}$. For $r \geq 1$, let $J_r = I_r(h)$ defined by (3.4). Then, for any $u \in \mathcal{A}_X^k$ and $z \in \mathcal{A}_Y^k$, it is easy to check that for $j = 0, \ldots, h-1$

$$\begin{aligned} \rho(u, z) &\leq \sum_{r=1}^{c(k,h)} \rho(u_{J_r+j}, z_{J_r+j}) + e(k, h) D_{\max} \\ &\leq \sum_{r=1}^{c(k,h)} \rho(u_{J_r+j}, z_{J_r+j}) + \epsilon k \end{aligned}$$

where for $J_r + j$ not totally included by $[1, k]$, the value of ρ is defined as 0. Then, for any $u \in \mathcal{A}_X^k$ and $\lambda \leq 0$

$$\begin{aligned} &\log \left\{ \int_{\mathcal{A}_Y^k} e^{\lambda \rho(u, z)} \, dQ_k(z) \right\} \\ &\quad\geq \frac{1}{h} \sum_{j=0}^{h-1} \log \left\{ \int_{\mathcal{A}_Y^k} e^{\left[\lambda\left(\sum_{r=1}^{c(k,h)} \rho(u_{J_r+j}, z_{J_r+j}) + \epsilon k\right)\right]} \cdot dQ_k(z) \right\} \\ &\quad\overset{(a)}{\geq} \frac{1}{h} \sum_{j=0}^{h-1} \log \left\{ \int_{\mathcal{A}_Y^k} C^{-c(k,h)} e^{\left[\lambda\left(\sum_{r=1}^{c(k,h)} \rho(u_{J_r+j}, z_{J_r+j}) + \epsilon k\right)\right]} \cdot \prod_{r=1}^{c(k,h)} dQ_h(z_{J_r+j}) \right\} \\ &\quad= -c(k, h) \log C + \epsilon k \lambda \\ &\qquad+ \frac{1}{h} \sum_{j=0}^{h-1} \sum_{r=1}^{c(k,h)} \log \left\{ \int_{\mathcal{A}_Y^h} e^{\lambda \rho(u_{J_r+j}, z)} \, dQ_h(z) \right\} \end{aligned} \tag{3.22}$$

where (a) is because Q satisfies condition (ψ_-). In particular, for each $u = x_{I_t}$, $t = 1, \ldots, n$, (3.22) holds. Then by the choice of k

$$\begin{aligned} \Lambda_{\hat{P}_{n,k}}(\lambda) = &\frac{1}{nk} \sum_{t=1}^n \log \left\{ \int_{\mathcal{A}_Y^k} e^{\lambda \rho(x_{I_t}, z)} \, dQ_k(z) \right\} \geq -\epsilon + \epsilon\lambda \\ &+ \frac{1}{hkn} \sum_{t=1}^n \sum_{j=0}^{h-1} \sum_{r=1}^{c(k,h)} \log \left\{ \int_{\mathcal{A}_Y^h} e^{\lambda \rho(u_{t, J_r+j}, z)} \, dQ_h(z) \right\} \end{aligned}$$

where

$$u_{t, J_r+j} = (x_{I_t})_{J_r+j} = x_L$$

with $L = J_r + (t-1)(k+d)+j$. Let $\mathcal{L}$ be the collection of all such L's. It is not hard to see that $\mathcal{L}$ consists of intervals with length h and left endpoint in $I_1 \cup \ldots \cup I_n$. Define $S = \bigcup_{L\in\mathcal{L}} L$. Then $S \subset [1, b_n]$. Because $\lambda \le 0$, by the choice of ϵ, h, and k, for n large enough

$$\begin{aligned}\lambda D - \Lambda_{\hat{P}_{n,k}}(\lambda) &\le \lambda(D-\epsilon) \\ &\quad + \epsilon - \frac{1}{hkn}\sum_{i\in S}\log\left\{\int_{\mathcal{A}_Y^h} e^{\lambda\rho(x_i^{i+h-1}, z)}\, dQ_h(z)\right\} \\ &\le \lambda(D-\epsilon) \\ &\quad + \epsilon - \frac{1}{hkn}\sum_{i=1}^{b_n}\log\left\{\int_{\mathcal{A}_Y^h} e^{\lambda\rho(x_i^{i+h-1}, z)}\, dQ_h(z)\right\} \\ &\le \lambda(D-\epsilon)+\epsilon-\frac{1}{h(1-\epsilon)} \\ &\quad \cdot \frac{1}{b_n}\sum_{i=1}^{b_n}\log\left\{\int_{\mathcal{A}_Y^h} e^{\lambda\rho(x_i^{i+h-1}, z)}\, dQ_h(z)\right\}.\end{aligned} \tag{3.23}$$

Denote

$$f(\lambda) = \lambda(D-\epsilon) + \epsilon - (1-\epsilon)^{-1}\Lambda_h(\lambda).$$

By $(D-\epsilon)(1-\epsilon) > D_{\min}^{(h)}$, $\lim_{\lambda\to-\infty} f(\lambda) = -\infty$. Given λ_0 with $f(\lambda_0) < 0$, let $\lambda = \lambda_0$ in (3.23). Because X is ergodic, P-almost surely, the right-hand side of (3.23) converges to $f(\lambda_0)$, which implies for n large enough, $\lambda_0 D - \Lambda_{\hat{P}_{n,k}}(\lambda_0) < 0$. Since

$$\sup_{\lambda\le 0}\{\lambda D - \Lambda_{\hat{P}_{n,k}}(\lambda)\} \ge 0$$

this implies the supremum is achieved in $[\lambda_0, 0]$, hence proving 1).

2) Fix ϵ, h, and k as above. Assume with positive probability $\liminf_{n\to\infty} |\lambda_{n,k}| = 0$. Then

$$\begin{aligned}D &= \Lambda'_{\hat{P}_{n,k}}(\lambda_{n,k}) \\ &= \frac{1}{kn}\sum_{t=1}^{n}\int \rho(u_t, z)e^{\lambda_{n,k}\rho(u_t, z)}\, dQ_k(z) \\ &\quad \cdot \left(\int e^{\lambda_{n,k}\rho(u_t, z)}\, dQ_k(z)\right)^{-1}.\end{aligned}$$

By $\rho(u_t, z) \le kD_{\max}$

$$\begin{aligned}\frac{1}{kn}e^{-|\lambda_{n,k}|kD_{\max}}\sum_{t=1}^{n}\int \rho(u_t, z)\, dQ_k(z) &\le D \\ &\le \frac{1}{kn}\, e^{|\lambda_{n,k}|kD_{\max}}\sum_{t=1}^{n}\int \rho(u_t, z)\, dQ_k(z).\end{aligned}$$

Let $n(l)$ be a subsequence such that $\lambda_{n(l),k} \to 0$. Because $\rho(u_t, z)$ is bounded, by the dominated convergence theorem, the above bounds implies $\lim_{l\to\infty} E_{n(l)} = D$, which is a contradiction to (3.21). Therefore, $|\lambda_{n,k}|$ is bounded below away from 0.

3) To show that $\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k})$ are bounded from above, first it is easy to see that

$$\begin{aligned}\Lambda_{\hat{P}_{n,k}}(\lambda_{n,k}) &\le \frac{1}{kn}\sum_{t=1}^{n}\int \rho^2(u_t, z)e^{\lambda_{n,k}\rho(u_t, z)}\, dQ_k(z) \\ &\quad \cdot \left(\int e^{\lambda_{n,k}\rho(u_t, z)}\, dQ_k(z)\right)^{-1}.\end{aligned}$$

By 1), given k large enough, for n large enough, $|\lambda_{n,k}|$ is bounded. Fix $M \limsup_{n\to\infty}|\lambda_{n,k}|$. Then for n large enough

$$\begin{aligned}\Lambda''_{\hat{P}_{n,k}}(\lambda_{n,k}) &\le \frac{1}{kn}\sum_{t=1}^{n}\int (kD_{\max})^2\, dQ_k(z) \\ &\quad \cdot \left(\int e^{-MkD_{\max}}\, dQ_k(z)\right)^{-1} \\ &= ke^{kMD_{\max}}D_{\max}^2,\end{aligned}$$

which completes the proof of 3). □

Proof of Lemma 6: From (3.23)

$$\begin{aligned}\lambda D - \Lambda_{\hat{P}_{n,k}}(\lambda) &\le \lambda(D-\epsilon) + \epsilon - \frac{1}{h(1-\epsilon)} \\ &\quad \cdot \frac{1}{b_n}\sum_{i=1}^{b_n}\log\left\{\int_{\mathcal{A}_Y^h} e^{\lambda(\rho(x_i^{i+h-1}, z))}\, dQ_h(z)\right\}.\end{aligned}$$

Given k, choose a subsequence $n(i)$ such that

$$\begin{aligned}\lim_{i\to\infty}\Lambda^*_{\hat{P}_{n(i),k}}(D) &= \lim_{i\to\infty}[\lambda_{n(i),k}D - \Lambda_{\hat{P}_{n(i),k}}(\lambda_{n(i),k})] \\ &= \limsup_{n\to\infty}\Lambda^*_{\hat{P}_{n,k}}(D).\end{aligned}$$

By Lemma 6, $\lambda_{n(i),k}$ is bounded, and hence has a limit point, say, λ_0. Then by the ergodicity of X and the boundedness of ρ, it is not hard to get from (3.23) that

$$\begin{aligned}&\limsup_{n\to\infty}\Lambda^*_{\hat{P}_{n,k}}(D) \\ &\quad \le \lambda_0(D-\epsilon)+\epsilon \\ &\qquad - \frac{1}{h(1-\epsilon)}\int_{\mathcal{A}_X^h}\log\left\{\int_{\mathcal{A}_Y^h} e^{\lambda_0(\rho(x, z))}\, dQ_h(z)\right\} dP_h(x) \\ &\quad = \lambda_0(D-\epsilon)+\epsilon - \frac{1}{1-\epsilon}\Lambda_h(P_h, Q_h, \lambda_0) \\ &\quad \le \epsilon + \frac{1}{1-\epsilon}R_h(P_h, Q_h, (D-\epsilon)(1-\epsilon)).\end{aligned}$$

Let $k\to\infty$, then $h\to\infty$, and finally $\epsilon\to 0$ to complete the proof. □

Appendix

The First-Order Asymptotic of $D(P_n\|Q_n)$ with $\mathcal{A}_X = \mathcal{A}_Y$ Discrete and Y Satisfying Condition (ψ_+)

Proof of Proposition 2: Suppose $\mathcal{A}_X = \mathcal{A}_Y = \mathcal{A}$, with $\mathcal{A}$ discrete. Since

$$\begin{aligned}\frac{1}{n}D(P_n\|Q_n) &= \frac{1}{n}\sum_{x_1^n}P(x_1^n)\log\frac{P(x_1^n)}{Q(x_n^1)} \\ &= -\frac{H(P_n)}{n} + \frac{1}{n}\sum_{x_1^n}P(x_1^n)\log\frac{1}{Q(x_1^n)} \\ &= -\frac{H(P_n)}{n} + \frac{R_n}{n}\end{aligned}$$

and $n^{-1}H(P_n) \to H(X)$, it is enough to show that $n^{-1}R_n$ converges.

Because X satisfies the condition (ψ_+), there are constants C, $d > 1$, such that for $A \in \sigma(X_1^n)$ and $B \in \sigma(X_{n+d+1}^\infty)$

$$Q(A\cap B) < CQ(A)Q(B).$$

Given k, define $I_t = I_t(k)$ by (3.4). For n large enough, let $s = c(n, k)$. Then for $x_1^n \in \mathcal{A}^n$

$$Q(x_1^n) \le Q(x_{I_1}, \ldots, x_{I_s}) \le C^s\prod_{j=1}^{s}Q(x_{I_j}).$$

Therefore,

$$\begin{aligned}\frac{R_n}{n} &\ge \frac{1}{n}\sum_{x_1^n}P(x_1^n)\log\frac{C^{-s}}{\prod_{j=1}^{s}Q(x_{I_j})} \\ &= -\frac{s\log C}{n} + \frac{s}{n}R_k \\ &= -\frac{s\log C}{n} + \frac{s}{n}R_k.\end{aligned}$$

Let $n\to\infty$. Then by $s/n \to 1/k$, there is

$$\liminf_{n\to\infty}[R_n/n] \ge -[\log C/k] + [R_k/k].$$

Let $k\to\infty$ to get

$$\liminf_{n\to\infty}[R_n/n] \ge \limsup_{k\to\infty}[R_k/k].$$

Therefore, $n^{-1}R_n$ converges. □

ACKNOWLEDGMENT

The author wishes to thank A. Dembo and I. Kontoyiannis for much valuable advice and pointing out several important papers for reference.

REFERENCES

[1] E.-h. Yang and J. C. Kieffer, "On the performance of data compression algorithms based upon string matching," *IEEE Trans. Inform. Theory*, vol. 44, pp. 47–65, Jan. 1998.

[2] A. Dembo and I. Kontoyiannis, "The asymptotics of waiting times between stationary processes, allowing distortion," *Ann. Appl. Probab.*, vol. 9, pp. 413–429, May 1999.

[3] E.-h. Yang and Z. Zhang, "On the redundancy of lossy source coding with abstract alphabets," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1092–1110, May 1999.

[4] A. D. Wyner and J. Ziv, "Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1250–1258, Nov. 1989.

[5] P. C. Shields, "Waiting times: Positive and negative results on the Wyner–Ziv problem," *J. Theor. Probab.*, vol. 6, no. 3, pp. 499–519, 1993.

[6] K. Marton and P. C. Shields, "Almost-sure waiting times results for weak and very weak Bernoulli processes," *Ergodic Theory Dyn. Syst.*, vol. 15, pp. 951–960, 1995.

[7] W. Szpankowski, "Asymptotic properties of data compression and suffix trees," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1647–1659, Sept. 1993.

[8] I. Kontoyiannis, "Asymptotic recurrence and waiting times for stationary processes," *J. Theor. Probab.*, vol. 11, pp. 795–811, July 1998.

[9] T. Łuczak and W. Szpankowski, "A suboptimal lossy data compression based on approximate pattern matching," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1439–1451, Sept. 1997.

[10] W. Bryc, "On large deviations for uniformly strong mixing sequences," *Stochastic Processes Their Applic.*, vol. 41, pp. 191–202, 1992.

[11] W. Bryc and A. Dembo, "Large deviations and strong mixing," *Ann. l'Inst. Henri Poincaré—Probab. et Statist.*, vol. 32, no. 4, pp. 549–569, 1996.

[12] R. C. Bradley, "Basic properties of strong mixing conditions," in *Dependence in Probability and Statistics*, E. Eberlein and M. Taqqu, Eds. Basel, Switzerland: Birkhäuser, 1986, pp. 165–192.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[14] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, MA: Jones and Bartlett, 1992.

Algebraic-Geometry Codes with Asymptotic Parameters Better than the Gilbert–Varshamov and the Tsfasman–Vlăduţ–Zink Bounds

Chaoping Xing

***Abstract*—In this correspondence, we show that both the Gilbert–Varshamov and the Tsfasman–Vlăduţ–Zink bounds can be improved by Goppa geometric codes around two points where these two bounds intersect.**

***Index Terms*—Algebraic-geometry codes, algebraic curves, Gilbert–Varshamov bound, rational points, Tsfasman–Vlăduţ–Zink bound.**

I. INTRODUCTION

Around 1981, Goppa [1] discovered a beautiful construction of linear codes based on curves over finite fields with many rational points. Nowadays, these codes are called Goppa geometric codes or algebraic-geometry codes. One of the exciting results from Goppa's construction is that the well-known Gilbert–Varshamov bound can be improved by Goppa geometry codes over finite fields of some composite order [9], [3]. For example, if $q \geq 49$ is a square or $q \geq (576 \times 27)^3$ is a cube, then the Gilbert–Varshamov bound is improved in an open interval.

Before proceeding with the ideas and results of the correspondence, we need to introduce some technical notation.

For a linear code C over $\boldsymbol{F}_q$ we denote by $n(C)$, $k(C)$, and $d(C)$ the length, the dimension, and the minimum distance of C, respectively. Let U_q^{lin} be the set of ordered pairs $(\delta, R) \in \boldsymbol{R}^2$ for which there exists an infinite sequence $C_1, C_2, \ldots$ of linear codes over $\boldsymbol{F}_q$ with $n(C_i) \to \infty$ and

$$\delta = \lim_{i\to\infty} \frac{d(C_i)}{n(C_i)} \qquad R = \lim_{i\to\infty} \frac{k(C_i)}{n(C_i)}.$$

The following description of U_q^{lin} can be found in [8, Sec. 1.3.1].

Proposition 1.1: There exists a continuous function $\alpha_q^{\text{lin}}(\delta)$, $\delta \in [0, 1]$, such that

$$U_q^{\text{lin}} = \{(\delta, R) \in \boldsymbol{R}^2 \colon 0 \leq R \leq \alpha_q^{\text{lin}}(\delta),\ 0 \leq \delta \leq 1\}.$$

Moreover, $\alpha_q^{\text{lin}}(0) = 1$, $\alpha_q^{\text{lin}}(\delta) = 0$ for $\delta \in [(q-1)/q, 1]$, and $\alpha_q^{\text{lin}}(\delta)$ decreases on the interval $[0, (q-1)/q]$.

For $0 < \delta < 1$ define the q-ary entropy function

$$H_q(\delta) = \delta \log_q (q-1) - \delta \log_q \delta - (1-\delta) \log_q (1-\delta)$$

where $\log_q$ is the logarithm to the base q, and put

$$R_{\text{GV}}(q, \delta) = 1 - H_q(\delta).$$

Then the Gilbert–Varshamov bound says that

$$\alpha_q^{\text{lin}}(\delta) \geq R_{\text{GV}}(q, \delta), \qquad \text{for all } \delta \in \left(0, \frac{q-1}{q}\right). \tag{1}$$

In order to introduce the asymptotic bound from Goppa geometric codes, we have to define some notation regarding the number of rational points of curves over finite fields.

When we speak of an algebraic curve over $\boldsymbol{F}_q$, we always mean a smooth, projective, absolutely irreducible algebraic curve defined over

Manuscript received January 7, 2000; revised June 14, 2000. This work was supported in part by MOD-ARF under Research Grant R146-000-018-112 and by NSTB-MOE under Research Grant R252-000-015-203.

The author is with the Department of Mathematics, National University of Singapore, Singapore 117543 (e-mail: matxcp@nus.edu.sg).

Communicated by I. F. Blake, Associate Editor for Coding Theory.

Publisher Item Identifier S 0018-9448(01)00594-6.