The First Order Asymptotics of Waiting Times between Stationary Processes under Nonstandard Conditions

Matthew Harrison

Division of Applied Mathematics
Brown University
Providence, RI 02912 USA
Matthew_Harrison@Brown.EDU

April 2, 2003

Abstract

We give necessary and sufficient conditions for the almost sure convergence of

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

when $(X_n)_{n\geq 1}$ is stationary and ergodic and when Q is stationary and satisfies certain strong mixing conditions. $B(x_1^n, D)$ is the single letter, additive distortion ball of radius D at the point $x_1^n := (x_1, \ldots, x_n)$. The asymptotic behavior of this quantity arises frequently in rate distortion theory, particularly when looking at the asymptotics of waiting times until a match (allowing distortion) between two stationary processes.

1 Introduction

Given two independent processes $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ with distributions P and Q, respectively, we are interested in the behavior of the waiting time until $Y_{k+1}^{k+n} := (Y_{k+1}, \ldots, Y_{k+n})$ matches X_1^n to within an allowable distortion D. In particular, we are interested in

$$W(X_1^n, Y_1^\infty, D) := \inf\left\{k \geq 1 : Y_k^{k+n-1} \in B(X_1^n, D)\right\},$$

where $B(x_1^n, D)$ is the set of y_1^n that match x_1^n to within distortion D. Typically, we have a nonnegative function ρ that measures the distortion between a single x and y and we define the distortion ball $B(x_1^n, D)$ in terms of the average distortion

$$B(x_1^n, D) := \left\{y_1^n : \frac{1}{n}\sum_{k=1}^{n} \rho(x_k, y_k) \leq D\right\}.$$

The asymptotic properties of these waiting times as n gets large have applications in rate distortion theory, analysis of DNA sequences and other areas. They have been studied in several recent papers [3, 4, 12, 14].

In each of these papers, the authors investigated these waiting times by showing that asymptotically

$$\log W(X_1^n, Y_1^\infty, D) \overset{\text{a.s.}}{\approx} -\log Q(B(X_1^n, D))$$

and then studying the quantity on the right. We adopt the same approach. All that is needed is

Proposition 1.1. *Suppose $(Y_n)_{n\geq 1}$ is a random process on $T^{\mathbb{N}}$ with distribution Q that is stationary and ψ_--mixing.*[1]*. Define $W_n := \inf\{k \geq 1 : Y_k^{k+n-1} \in A_n\}$ for a sequence of measurable sets $(A_n)_{n\geq 1}$, $A_n \in T^n$. If $(c_n)_{n\geq 1}$ is a nonnegative sequence with $\sum_n e^{-c_n} < \infty$, then*

$$\text{Prob}\left\{-\log_e Q(A_n) - c_n \leq \log_e W_n \leq -\log_e Q(A_n) + c_n + \log_e n \ \textit{eventually}\right\} = 1.$$

With $A_n := B(x_1^n, D)$ we can relate waiting times to the probabilities of distortion balls. Once we note that $\text{Prob}\{\log W_n = \log Q(A_n)\} = 1$ whenever $Q(A_n) \in \{0, 1\}$, then the proof of Proposition 1.1 follows almost exactly from a similar result in Kontoyiannis (1998) [11]. We give a proof in the Appendix for completeness.

A common assumption is that the distortion function ρ is bounded or that it satisfies certain moment conditions. Unfortunately, the bounded assumption rules out squared error distortion $\rho(x, y) = \|x - y\|^2$ on $\mathbb{R}^d$, which is common in practice, and the moment conditions depend on the source distribution, which may not be known. This last shortcoming can be critical when studying universal lossy data compression or statistical methods in lossy data compression and is our main motivation for trying to relax these assumptions.

Here we investigate the limiting behavior of

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

without restrictions on ρ. We obtain necessary and sufficient conditions for the a.s. convergence of this quantity and we precisely characterize its behavior when convergence fails. We also relax several other assumptions that often appear in the literature. The source $(X_n)_{n\geq 1}$ is assumed to be stationary and ergodic, as opposed to independent and identically distributed (i.i.d.), the reproduction $(Y_n)_{n\geq 1}$ is allowed to have some memory, as opposed to i.i.d., both the source and the reproduction take values in arbitrary alphabets, as opposed to finite alphabets, and the range of distortion values D is not constrained in the usual manner.

2 Main Results

We begin with the setup used throughout the remainder of the paper. $(S, \mathcal{S})$ and $(T, \mathcal{T})$ are standard measurable spaces.[2] $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ are independent stationary random processes on the sequence spaces $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$ and $(T^{\mathbb{N}}, \mathcal{T}^{\mathbb{N}})$ with distributions P and Q, respectively. We assume that P is ergodic and that Q satisfies the following strong mixing condition:

$$C^{-1}Q(A)Q(B) \leq Q(A \cap B) \leq CQ(A)Q(B)$$

[1] Q is ψ_--mixing if there exists finite $C, d \geq 1$ such that $Q(A)Q(B) \leq CQ(A \cap B)$ for all $A \in \sigma(Y_1^n)$ and $B \in \sigma(Y_{n+d}^\infty)$ and any n. (See Chi (2001) [3] and the references therein.)

[2] Standard measurable spaces include Polish spaces and let us avoid uninteresting pathologies while working with random sequences [8].

for some fixed $1 \leq C < \infty$ and any $A \in \sigma(Y_1^n)$ and $B \in \sigma(Y_{n+1}^\infty)$ and any n.[3] Notice that this includes the cases where Q is i.i.d. ($C = 1$) and where Q is a finite state Markov chain with all positive transition probabilities.

Let $\rho : S \times T \to [0, \infty)$ be an $\mathcal{S} \times \mathcal{T}$-measurable function ($\mathcal{S} \times \mathcal{T}$ denotes the smallest product σ-algebra). We define the following standard quantities:

$$B(x_1^n, D) := \left\{ y_1^n \in T^n : \frac{1}{n} \sum_{k=1}^n \rho(x_k, y_k) \leq D \right\},$$

$$\Lambda_n(\lambda) := \frac{1}{n} E_P \log E_Q e^{\lambda \sum_{k=1}^n \rho(X_k, Y_k)}, \qquad \Lambda_\infty(\lambda) := \limsup_{n\to\infty} \Lambda_n(\lambda),$$

$$\Lambda_n^*(D) := \sup_{\lambda \leq 0} \left[\lambda D - \Lambda_n(\lambda) \right], \quad n = 1, \ldots, \infty,$$

$$\rho_Q(x) := \operatorname*{ess\,inf}_Q \rho(x, Y_1), \qquad D_{\min} := E\rho_Q(X_1), \qquad D_{\text{ave}} := E\rho(X_1, Y_1).$$

We always assume that $D \in \mathbb{R}$ and log denotes the natural logarithm $\log_e$. Notice that $0 \leq D_{\min} \leq D_{\text{ave}} \leq \infty$. $B(x_1^n, D)$ is called the distortion ball of radius D at x_1^n and ρ is called the single letter distortion function.

In the special case where Q is i.i.d. it is easy to see that $\Lambda_n = \Lambda_1$ for all $n = 1, \ldots, \infty$, and similarly that $\Lambda_n^* = \Lambda_1^*$ for all n. In the general case we have

$$\begin{aligned} \Lambda_\infty(\lambda) &= \lim_{n\to\infty} \Lambda_n(\lambda), \quad \lambda \leq 0, \\ \Lambda_\infty^*(D) &= \lim_{n\to\infty} \Lambda_n^*(D), \end{aligned} \tag{2.1}$$

which we prove in Section 3. We also show that

$$\Lambda_n^*(D) = R_n(P_n, Q_n, D) := \frac{1}{n} \inf_{W_n} H(W_n \| P_n \times Q_n), \tag{2.2}$$

where the infimum is over all probability measures W_n on $S^n \times T^n$ that have marginal distribution P_n on S^n and with $E_{W_n} n^{-1} \sum_{k=1}^n \rho(X_k, Y_k) \leq D$. P_n and Q_n denote the nth marginals of P and Q, respectively. $H(\mu\|\nu)$ denotes the relative entropy in nats

$$H(\mu\|\nu) := \begin{cases} E_\mu \log \frac{d\mu}{d\nu} & \text{if } \mu \ll \nu, \\ \infty & \text{otherwise.} \end{cases}$$

This alternative characterization of Λ^* is well known [4], although we prove it here without any restrictions on D. Notice that (2.1) and (2.2) give

$$R_\infty(P, Q, D) := \lim_{n\to\infty} R_n(P_n, Q_n, D) = \Lambda_\infty^*(D).$$

We are interested in the asymptotic behavior of $-\log Q(B(X_1^n, D))$. An easy result is

$$\begin{aligned} &\text{Prob}\left\{-\log Q(B(X_1^n, D)) = \infty \text{ eventually}\right\} = 1 \quad \text{if } D < D_{\min}, \\ &\text{Prob}\left\{-\log Q(B(X_1^n, D)) < \infty \text{ eventually}\right\} = 1 \quad \text{if } D > D_{\min}. \end{aligned} \tag{2.3}$$

The main result of the paper is the following:

[3] In the notation of Chi (2001) [3] this implies that Q is $\psi_\pm$-mixing, but is stronger because we require that $d = 1$.

Theorem 2.1. *If $D \neq D_{\min}$ or $\Lambda_\infty^*(D) = \infty$ or $\rho_Q(X_1)$ is a.s. constant then*

$$\lim_{n\to\infty} -\frac{1}{n}\log Q(B(X_1^n, D)) \overset{a.s.}{=} \Lambda_\infty^*(D). \tag{2.4}$$

Otherwise, $0 < D = D_{\min} < \infty$, and

$$\text{Prob}\left\{-\log Q(B(X_1^n, D)) = \infty \textit{ infinitely often}\right\} > 0, \tag{2.5a}$$

$$\text{Prob}\left\{-\log Q(B(X_1^n, D)) < \infty \textit{ infinitely often}\right\} = 1, \tag{2.5b}$$

$$\lim_{m\to\infty} -\frac{1}{n_m}\log Q(B(X_1^{n_m}, D)) \overset{a.s.}{=} \Lambda_\infty^*(D) < \infty, \tag{2.5c}$$

where $(n_m)_{m\geq 1}$ is the (a.s.) infinite subsequence of $(n)_{n\geq 1}$ for which $-\log Q(B(X_1^n, D))$ is finite, or (a.s.) equivalently, the subsequence where $\sum_{k=1}^n \rho_Q(X_k) \leq nD$.

Proposition 1.1 shows that we can replace $-\log Q(B(X_1^n, D))$ with $\log W(X_1^n, Y_1^\infty, D)$ in Theorem 2.1 and in (2.3). In this case Prob and a.s. will refer to the joint probability of $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$. Also, notice that (2.5) implies that the limit in (2.4) does not exist with positive probability, so the conditions for (2.4) are necessary and sufficient (and similarly with (2.5)). Finally, we point out that (2.4) always holds when $D = 0$, because either $D < D_{\min}$ or $D_{\min} = 0$, which means $\rho_Q(X_1) \overset{\text{a.s.}}{=} 0$. This clears up part of the difficulty mentioned in Dembo and Kontoyiannis (2002) [4][pp. 1593–1594] when trying to think of (2.4) as a lossy generalization of the lossless AEP (Asymptotic Equipartition Property).

The generalized AEP (2.4) can be used to show that (a sequence of) random lossy codebooks generated by Q can have asymptotic (pointwise) rates of $R_\infty(P, Q, D)$. See Kontoyiannis and Zhang (2002) [12] for the details and for the assumptions that make all of this precise. Briefly, these codebooks send the index of the first time that a random Y_1^n is in $B(X_1^n, D)$. This random match is used as the distorted version of X_1^n. These codebooks enforce the condition that X_1^n is distorted by no more than D. (2.5) shows that this will not be possible in certain cases when even when $R_\infty(P, Q, D)$ is finite. What does this mean for the intuition that we can always use random lossy codebooks based on Q to achieve asymptotic rates of $R_\infty(P, Q, D)$?

Consider the situation where $D = D_{\min}$ and $R_\infty(P, Q, D)$ is finite. Notice that this includes all situations where (2.4) does not hold. Define the set

$$A(x_1^n) := \left\{ y_1^n \in T^n : \frac{1}{n}\sum_{k=1}^n \rho(x_k, y_k) = \operatorname*{ess\,inf}_Q \frac{1}{n}\sum_{k=1}^n \rho(x_k, Y_k) \right\}. \tag{2.6}$$

Suppose that we modify the random coding scheme mentioned above so that the distorted X_1^n must match $A(X_1^n)$ instead of $B(X_1^n, D)$. In Section 3.5 we show that

$$\operatorname*{ess\,inf}_Q \frac{1}{n}\sum_{k=1}^n \rho(X_k, Y_k) \to D_{\min} \quad \text{and} \quad -\frac{1}{n}\log Q(A(X_1^n)) \to R_\infty(P, Q, D_{\min}),$$

where in both cases the convergence holds a.s. and in expectation. We also show that

$$E_P\left[\operatorname*{ess\,inf}_Q \frac{1}{n}\sum_{k=1}^n \rho(X_k, Y_k)\right] = D_{\min}.$$

This random coding scheme ensures that the distortion converges to $D = D_{\min}$, although a particular X_1^n might be distorted by more than D, and the asymptotic rate is $R_\infty(P,Q,D)$. It also ensures that the expected distortion is always D. Thus, it is still possible to generate codebooks based on Q with distortion D and rate $R_\infty(P,Q,D)$, but we must use a weaker notion of distortion.[4]

The proof of Theorem 2.1 proceeds in several stages. The lower bound in (2.4) is a consequence of Chebyshev's inequality. The upper bound for the case $D_{\min} < D \le D_{\text{ave}}$ when Q is i.i.d. follows from a large deviations argument. The outline for this argument comes from Dembo and Kontoyiannis (2002) [4][Theorem 1], but there they assumed that $D_{\text{ave}} < \infty$. The upper bound for the general case when $D_{\min} < D \le D_{\text{ave}}$ is derived from the i.i.d. case with a blocking argument and a result from ergodic theory. The case where $D > D_{\text{ave}}$ is a simple application of Chebyshev's inequality. The behavior when $D = D_{\min}$ comes from the subadditive ergodic theorem ($\rho_Q(X_1)$ a.s. constant) and from the recurrence properties of random walks with stationary increments ($\rho_Q(X_1)$ not a.s. constant).

3 Proof of Theorem 2.1

Throughout the proofs we use the stationarity and mixing properties of Q to apply the bounds

$$C^{-1}E_Q f(Y_1^n)E_Q g(Y_1^m) \le E_Q f(Y_1^n)g(Y_{n+1}^{n+m}) \le C E_Q f(Y_1^n)E_Q g(Y_1^m)$$

for nonnegative functions f and g. We also make use of the fact that

$$B(x_1^{n+m}, D) \supset B(x_1^n, D) \cap B(x_{n+1}^{n+m}, D),$$

where here we abuse notation and think of $B(x_{n+1}^{n+m}, D)$ as being an element of $\sigma(Y_{n+1}^{n+m})$. Lastly, we make frequent use of the regularity properties of Λ_n and Λ_n^* found in the Appendices. We do not necessarily point out each place where these ideas are put to use.

Let us first establish (2.1). Define

$$\Lambda_n(x_1^n, \lambda) := \frac{1}{n}\log E_Q e^{\lambda \sum_{k=1}^n \rho(x_k, Y_k)}.$$

We have

$$\begin{aligned}
(n+m)\Lambda_{n+m}(x_1^{n+m}, \lambda) + \log C &= \log E_Q e^{\lambda \sum_{k=1}^{n+m} \rho(x_k, Y_k)} + \log C \\
&= \log E_Q e^{\lambda \sum_{k=1}^{n} \rho(x_k, Y_k)} e^{\lambda \sum_{k=n+1}^{n+m} \rho(x_k, Y_k)} + \log C \\
&\le \log E_Q e^{\lambda \sum_{k=1}^{n} \rho(x_k, Y_k)} + \log C + \log E_Q e^{\lambda \sum_{k=1}^{m} \rho(x_{n+k}, Y_k)} + \log C \\
&= \left[n\Lambda_n(x_1^n, \lambda) + \log C\right] + \left[m\Lambda_m(x_{n+1}^{n+m}, \lambda) + \log C\right].
\end{aligned}$$

If $\lambda \le 0$, then all of these terms are bounded above by $\log C$ and the subadditive ergodic theorem gives

$$\lim_{n\to\infty} \Lambda_n(X_1^n, \lambda) \overset{\text{a.s.}}{=} \lim_{n\to\infty} \Lambda_n(\lambda) = \inf_{m \ge M}\left[\Lambda_m(\lambda) + \frac{\log C}{m}\right] \tag{3.1}$$

[4]Notice that $A(X_1^n)$ and $B(X_1^n, D_{\min})$ differ by a Q-null set a.s. when $\rho_Q(X_1)$ is a.s. constant. In this special case, the modified coding scheme does not change and is why (2.4) remains true.

for any $M \geq 1$. This shows that $\Lambda_\infty = \lim_n \Lambda_n$.

We have

$$\Lambda_\infty^*(D) = \sup_{\lambda \leq 0} \left[\lambda D - \lim_{n\to\infty} \Lambda_n(\lambda)\right] = \sup_{\lambda \leq 0} \left[\lambda D - \inf_{m \geq M} \left(\Lambda_m(\lambda) + \frac{\log C}{m}\right)\right]$$
$$= \sup_{m \geq M} \sup_{\lambda \leq 0} \left[[\lambda D - \Lambda_m(\lambda)] - \frac{\log C}{m}\right] = \lim_{n\to\infty} \left[\Lambda_n^*(D) - \frac{\log C}{n}\right],$$

since M is arbitrary. This shows that $\Lambda_\infty^* = \lim_n \Lambda_n^*$.

The proof of (2.2) essentially comes from Dembo and Kontoyiannis (2002) [4]. We first show that

$$R_n(P_n, Q_n, D) \geq \Lambda_n^*(D). \tag{3.2}$$

Fix any probability measure W_n on $S^n \times T^n$ with

$$S^n\text{-marginal equal to } P_n, \qquad E_{W_n} \frac{1}{n} \sum_{k=1}^n \rho(X_k, Y_k) \leq D, \tag{3.3}$$

and with $H(W_n \| P_n \times Q_n) < \infty$. Since all our spaces are standard, regular conditional probability distributions exist and we have $H(W_n \| P_n \times Q_n) = E_{P_n} H(W_n(\cdot|X_1^n) \| Q_n)$, where $W_n(\cdot|x_1^n)$ is the conditional probability of W_n on T^n given $x_1^n \in S^n$.

Let $\psi : T^n \to (-\infty, 0]$ be measurable. Then [4][pp.1595]

$$H(Q_n' \| Q_n) \geq E_{Q_n'}(\psi(Y_1^n)) - \log E_{Q_n} e^{\psi(Y_1^n)}$$

for any probability measure Q_n' on T^n. Applying this with $\psi(y_1^n) := \lambda \sum_{k=1}^n \rho(x_k, y_k)$ for $\lambda \leq 0$ gives

$$H(W_n(\cdot|x_1^n) \| Q_n) \geq \lambda E_{W_n(\cdot|x_1^n)} \sum_{k=1}^n \rho(x_k, Y_k) - \log E_{Q_n} e^{\lambda \sum_{k=1}^n \rho(x_k, Y_k)}.$$

Taking expected values and using (3.3) gives

$$n^{-1} H(W_n \| P_n \times Q_n) \geq \lambda D - \Lambda_n(\lambda).$$

Taking the supremum over $\lambda \leq 0$ and then the infimum over W_n satisfying (3.3) gives (3.2).

(3.2) immediately gives (2.2) whenever $\Lambda_n^*(D) = \infty$. This includes the case $D < D_{\min}$. When $\Lambda_n^*(D) < \infty$, we will construct W_n satisfying (3.3) that have $n^{-1} H(W_n \| P_n \times Q_n) \leq \Lambda_n^*(D)$ to complete the proof of (2.2).

Suppose $D \geq D_{\text{ave}}$. Then $W_n := P_n \times Q_n$ satisfies (3.3). Notice that $H(W_n \| P_n \times Q_n) = 0 \leq \Lambda_n^*(D)$. Combining this with (3.2) gives (2.2).

Now suppose that $D_{\min} < D < D_{\text{ave}}$ and $\Lambda_n^*(D) < \infty$. The Appendix shows that we can choose finite $\lambda_D < 0$ so that $\Lambda_n'(\lambda_D) = D$. Define W_n by

$$\left[\frac{dW_n}{d(P_n \times Q_n)}\right](x_1^n, y_1^n) := \frac{e^{\lambda_D \sum_{k=1}^n \rho(x_k, y_k)}}{E_{Q_n} e^{\lambda_D \sum_{k=1}^n \rho(x_k, Y_k)}}.$$

W_n has S^n-marginal P_n and the Appendix shows that

$$E_{W_n} \frac{1}{n} \sum_{k=1}^n \rho(X_k, Y_k) = \Lambda_n'(\lambda_D) = D.$$

Evaluating $H(W_n \| P_n \times Q_n)$ gives

$$n^{-1} H(W_n \| P_n \times Q_n) = \lambda_D D - \Lambda_n(\lambda_D) \leq \Lambda_n^*(D).$$

Combining this with (3.2) gives (2.2).

Finally, suppose that $D = D_{\min}$ and $\Lambda_n^*(D) < \infty$. The Appendix shows that $\Lambda_n^*(D) = n^{-1} E[-\log Q_n(A(X_1^n))]$, where $A(x_1^n)$ is defined in (2.6). Define W_n by

$$\left[\frac{dW_n}{d(P_n \times Q_n)}\right](x_1^n, y_1^n) := \frac{I_{A(x_1^n)}(y_1^n)}{Q(A(x_1^n))}.$$

$I_A(z)$ denotes the indicator function that $z \in A$. Since $E[-\log Q(A(X_1^n))]$ is finite, the denominator is positive P-a.s. and W_n is well defined. The S^n-marginal of W_n is P_n. From the definition of $A(x_1^n)$ and the mixing properties of Q we see that

$$E_{W_n} \frac{1}{n} \sum_{k=1}^{n} \rho(X_k, Y_k) = E_{W_n} \left[\operatorname*{ess\,inf}_{Q_n} \frac{1}{n} \sum_{k=1}^{n} \rho(X_k, Y_k)\right] = E_{W_n} \frac{1}{n} \sum_{k=1}^{n} \rho_Q(X_k) = D_{\min} = D,$$

so (3.3) holds. Evaluating $H(W_n \| P_n \times Q_n)$ gives

$$n^{-1} H(W_n \| P_n \times Q_n) = n^{-1} E\left[-\log Q_n(A(X_1^n))\right] = \Lambda_n^*(D).$$

Combining this with (3.2) gives (2.2). This completes the proof of (2.2).

Now we will establish (2.3). We have the following implications:

$$\begin{aligned}
\operatorname*{ess\,inf}_{Q} \frac{1}{n} \sum_{k=1}^{n} \rho(x_k, Y_k) > D \quad &\Longrightarrow \quad Q(B(x_1^n, D)) = 0, \\
\operatorname*{ess\,inf}_{Q} \frac{1}{n} \sum_{k=1}^{n} \rho(x_k, Y_k) < D \quad &\Longrightarrow \quad Q(B(x_1^n, D)) > 0.
\end{aligned} \tag{3.4}$$

The properties of Q show that

$$\operatorname*{ess\,inf}_{Q} \frac{1}{n} \sum_{k=1}^{n} \rho(X_k, Y_k) = \frac{1}{n} \sum_{k=1}^{n} \rho_Q(X_k) \stackrel{\text{a.s.}}{\to} D_{\min}$$

by the ergodic theorem, so

$$\begin{aligned}
\text{Prob}\left\{\operatorname*{ess\,inf}_{Q} \frac{1}{n} \sum_{k=1}^{n} \rho(X_k, Y_k) > D \text{ eventually}\right\} &= 1 \quad \text{if } D < D_{\min}, \\
\text{Prob}\left\{\operatorname*{ess\,inf}_{Q} \frac{1}{n} \sum_{k=1}^{n} \rho(X_k, Y_k) < D \text{ eventually}\right\} &= 1 \quad \text{if } D > D_{\min}.
\end{aligned}$$

Combining these with (3.4) gives (2.3).

3.1 Proof: Lower bound

For any $\lambda \leq 0$ we have

$$-\frac{1}{n} \log Q(B(x_1^n, D)) \geq -\frac{1}{n} \log E_Q e^{\lambda \sum_{k=1}^{n} \rho(x_k, Y_k) - \lambda n D} = \lambda D - \Lambda_n(x_1^n, \lambda).$$

Taking limits, applying (3.1) and (2.1), and optimizing over $\lambda \le 0$ (λ rational) gives

$$\liminf_{n\to\infty} -\frac{1}{n}\log Q(B(X_1^n, D)) \overset{\text{a.s.}}{\ge} \Lambda_\infty^*(D). \tag{3.5}$$

The reason we can restrict the supremum to rational $\lambda \le 0$, is that $\lambda D - \Lambda_\infty(\lambda)$ is concave in λ. This proves half of (2.4) and completes the proof when $\Lambda_\infty^*(D) = \infty$. Note that this includes the cases where $D < D_{\min}$.

Henceforth we will assume that $\Lambda_\infty^*(D) < \infty$. This assumption implies several things that are worth pointing out. First, we have

$$\Lambda_n(\lambda) = \frac{1}{n}E_P \log E_Q \prod_{k=1}^n e^{\lambda\rho(X_k,Y_k)} \le \frac{1}{n}\sum_{k=1}^n \left[E_P \log E_Q e^{\lambda\rho(X_k,Y_k)} + \log C\right] = \Lambda_1(\lambda) + \log C.$$

Similarly, we have

$$\Lambda_n(\lambda) \ge \Lambda_1(\lambda) - \log C,$$

so

$$\Lambda_1(\lambda) - \log C \le \Lambda_n(\lambda) \le \Lambda_1(\lambda) + \log C, \quad 1 \le n \le \infty.$$

These inequalities immediately imply

$$\Lambda_1^*(D) - \log C \le \Lambda_n^*(D) \le \Lambda_1^*(D) + \log C, \quad 1 \le n \le \infty. \tag{3.6}$$

So $\Lambda_\infty^*(D) < \infty$ implies that $\Lambda_n^*(D) < \infty$ for all n and this implies that $\Lambda_n(\lambda)$ is finite for all $\lambda \le 0$ and all n.

3.2 Proof: Upper bound, i.i.d. Q, $\mathbf{D_{\min} < D \le D_{\text{ave}}}$

In this section, we assume that Q is i.i.d., that is, it is a product measure. We also assume that $D_{\min} < D \le D_{\text{ave}}$. We allow for the case $D_{\text{ave}} = \infty$ (in which case $D < D_{\text{ave}}$). We want to prove that

$$\limsup_{n\to\infty} -\frac{1}{n}\log Q(B(X_1^n, D)) \overset{\text{a.s.}}{\le} \Lambda_1^*(D), \tag{3.7}$$

A proof is outlined in Dembo and Kontoyiannis (2002) [4][Theorem 1] under the added assumption that $D < D_{\text{ave}} < \infty$. The proof is essentially an application of the lower bound of the Gärtner-Ellis Theorem [6][Theorem V.6(b)] for large deviations.

Let $\Lambda^*(d) := \sup_{\lambda\in\mathbb{R}}[\lambda d - \Lambda_1(\lambda)]$. We have $\Lambda_1^* = \Lambda^*$ on $(-\infty, D_{\text{ave}}]$. Notice that

$$\frac{1}{n}\log E_Q e^{\lambda\sum_{k=1}^n \rho(X_k,Y_k)} = \frac{1}{n}\sum_{k=1}^n \log E_Q e^{\lambda\rho(X_k,Y_1)} \overset{\text{a.s.}}{\to} \Lambda_1(\lambda) \tag{3.8}$$

by the assumption that Q is a product measure and by the ergodic theorem.

Fix a realization $(x_n)_{n\ge1}$ of $(X_n)_{n\ge1}$ such that (3.8) holds for all $\lambda \in \mathbb{R}$. (We can choose the exceptional sets independent of λ since Λ_1 is increasing.) Define the sequence of random variables $(W_n)_{n\ge1}$ by

$$W_n := \frac{1}{n}\sum_{k=1}^n \rho(x_k, Y_k)$$

and let R_n denote the distribution of W_n. Note that $\log Q(B(x_1^n, D)) = \log R_n((-\infty, D])$, so we are interested in $\liminf_n n^{-1} R_n((-\infty, D]))$.

At this point we would like to invoke the Gärtner-Ellis Theorem. Unfortunately, not all of the assumptions are satisfied. We need to verify that the proof of the lower bound of the Gärtner-Ellis Theorem can still be carried through in our case. Here are the details. They closely follow the proof of the Gärtner-Ellis Theorem found in den Hollander (2000) [6].

For each $\lambda \le 0$ define the new sequence of probability distributions $(R_n^\lambda)_{n\ge1}$ by

$$\left[\frac{dR_n^\lambda}{dR_n}\right](w) := \frac{e^{\lambda n w}}{Ee^{\lambda n W_n}}.$$

Fix $\epsilon > 0$ such that $D_{\min} < D - \epsilon < D_{\text{ave}}$. We have

$$\begin{aligned}\log Q(B(x_1^n, D)) &= \log R_n((-\infty, D]) \ge \log R_n((D-\epsilon, D)) \\ &= \log \int I_{(D-\epsilon,D)}(w)\frac{Ee^{\lambda n W_n}}{e^{\lambda n w}} R_n^\lambda(dw) \\ &\ge \log Ee^{\lambda n W_n} + \log e^{-\lambda n (D-\epsilon)} + \log R_n^\lambda((D-\epsilon, D)),\end{aligned}$$

where $I_A(w)$ is the indicator function of $w \in A$. Dividing by n, taking limits and applying (3.8) gives

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log Q(B(x_1^n, D)) &\ge \Lambda_1(\lambda) - \lambda(D-\epsilon) + \liminf_{n\to\infty} \frac{1}{n} \log R_n^\lambda((D-\epsilon, D)) \\ &\ge -\Lambda_1^*(D-\epsilon) + \liminf_{n\to\infty} \frac{1}{n} \log R_n^\lambda((D-\epsilon, D)). \end{aligned} \tag{3.9}$$

If we can choose $\tilde\lambda \le 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log R_n^{\tilde\lambda}((D-\epsilon, D)) \ge 0, \tag{3.10}$$

then we will be finished. To see this, notice that (3.9) and (3.10) give

$$\liminf_{n\to\infty} \frac{1}{n} \log Q(B(x_1^n, D)) \ge -\Lambda_1^*(D-\epsilon).$$

Letting $\epsilon \downarrow 0$, using the fact that Λ_1^* is continuous at D and noticing that $(x_n)_{n\ge1}$ was a.s. arbitrary completes the proof of (3.7).

Now we will prove (3.10). Choose $\tilde\lambda < 0$ such that

$$\Lambda^*(D-\epsilon/2) = \Lambda_1^*(D-\epsilon/2) = \tilde\lambda(D-\epsilon/2) - \Lambda_1(\tilde\lambda).$$

Define

$$\tilde\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{n} \log E_{R_n^{\tilde\lambda}} e^{\lambda n W_n} = \lim_{n\to\infty} \frac{1}{n} \log \frac{E_{R_n} e^{\lambda n W_n} e^{\tilde\lambda n W_n}}{E_{R_n} e^{\tilde\lambda n W_n}} = \Lambda_1(\lambda + \tilde\lambda) - \Lambda_1(\tilde\lambda)$$

by (3.8). Notice that $\tilde\Lambda$ is finite on $(-\infty, -\tilde\lambda]$ which includes a neighborhood of 0. Define

$$\begin{aligned}\tilde\Lambda^*(d) &:= \sup_{\lambda\in\mathbb{R}} \left[\lambda d - \tilde\Lambda(\lambda)\right] = \sup_{\lambda\in\mathbb{R}} \left[(\lambda+\tilde\lambda)d - \Lambda_1(\lambda+\tilde\lambda)\right] - \tilde\lambda d + \Lambda_1(\tilde\lambda) \\ &= \Lambda^*(d) - \left[\tilde\lambda d - \Lambda_1(\tilde\lambda)\right].\end{aligned}$$

Notice that $\tilde{\Lambda}^*(D-\epsilon/2)=0$.

$\Lambda^* = \Lambda_1^*$ is strictly convex on $(D_{\min}, D_{\text{ave}})$, so $\tilde{\Lambda}^*$ is also. Furthermore, since $\Lambda^* = \Lambda_1^*$ is strictly decreasing on $(D_{\min}, D_{\text{ave}})$ it must have supporting planes with strictly negative slopes. This implies that $\tilde{\Lambda}^*$ has supporting planes with slopes strictly less than $-\tilde{\lambda}$ on $(D_{\min}, D_{\text{ave}})$. Recalling that $\tilde{\Lambda}$ is finite on $(-\infty, -\tilde{\lambda})$, we can apply the Gärtner-Ellis Theorem [6][Theorem V.6(b)] to get

$$\liminf_{n\to\infty} \frac{1}{n} \log R_n^{\tilde{\lambda}}((D-\epsilon, D)) \geq - \inf_{d\in(D-\epsilon,D)} \tilde{\Lambda}^*(d) \geq -\tilde{\Lambda}^*(D-\epsilon/2) = 0,$$

which completes the proof.

Before continuing, we need to modify (3.7) slightly. Let $M \geq 0$ be any integer valued random variable. Then we also have

$$\limsup_{n\to\infty} -\frac{1}{n} \log Q(B(X_{M+1}^{M+n}, D)) \overset{\text{a.s.}}{\leq} \Lambda_1^*(D). \tag{3.11}$$

The stationarity of P and (3.7) show that (3.11) holds for any fixed, constant M. So (3.11) holds for all (fixed, constant) M simultaneously, and therefore it also holds for any random M independent of n.

3.3 Proof: Upper bound, $\mathbf{D}_{\min} < \mathbf{D} \leq \mathbf{D}_{\text{ave}}$

We no longer assume that Q is a product measure, however we still assume that $D_{\min} < D \leq D_{\text{ave}}$ and we want to use (3.7) to derive the general upper bound

$$\limsup_{n\to\infty} -\frac{1}{n} \log Q(B(X_1^n, D)) \overset{\text{a.s.}}{\leq} \Lambda_\infty^*(D). \tag{3.12}$$

We first derive some bounds that let us establish (3.13), which essentially says that we can shift the sequence $(X_n)_{n\geq 1}$ in certain ways without decreasing the above lim sup (or even the lim sup along a subsequence).

Let Q' be the distribution of an i.i.d. process with the same first marginal as Q. The mixing properties of Q show that for any set $A \in \sigma(Y_1^n)$, we have $C^{-n}Q'(A) \leq Q(A) \leq C^n Q'(A)$. This lets us use (3.7) to immediately see that

$$\begin{aligned}\limsup_{n\to\infty} -\frac{1}{n} \log Q(B(X_1^n, D)) &\leq \limsup_{n\to\infty} -\frac{1}{n} \log Q'(B(X_1^n, D)) + \log C \\ &\overset{\text{a.s.}}{=} \Lambda_1^*(D) + \log C < \infty,\end{aligned}$$

since Q and Q' have the same Λ_1 and since $\Lambda_\infty^*(D) < \infty$ implies that $\Lambda_1^*(D) < \infty$. We can thus find an integer valued random variable N such that

$$\sup_{n\geq N} -\frac{1}{n} \log Q(B(X_1^n, D)) \overset{\text{a.s.}}{\leq} \Lambda_1^*(D) + \log C + 1 < \infty.$$

Let $(a_n)_{n\geq 1}$ be a strictly increasing, positive integer sequence and let $M \overset{\text{a.s.}}{\geq} N$ be an integer valued random variable. We have

$$
\begin{aligned}
&\limsup_{n\to\infty} -\frac{1}{a_n}\log Q(B(X_1^{a_n},D)) \\
&\quad\overset{\text{a.s.}}{\le} \limsup_{n\to\infty} -\frac{1}{a_n}\left[\log Q(B(X_1^M,D)) - \log C + \log Q(B(X_{M+1}^{a_n},D))\right] \\
&\quad\overset{\text{a.s.}}{\le} \limsup_{n\to\infty} -\frac{1}{a_n}\left[-M(\Lambda_1^*(D)+\log C+1) - \log C + \log Q(B(X_{M+1}^{a_n},D))\right] \\
&\quad= \limsup_{n\to\infty} -\frac{1}{a_n - M}\log Q(B(X_{M+1}^{a_n},D)). \qquad (3.13)
\end{aligned}
$$

Now we will use a blocking argument so that we can apply (3.7), actually (3.11). Fix $m \ge 1$ and $0 \le r < m$. Define $\hat{S} := S^m$, $\hat{T} := T^m$, $\hat{\rho} : \hat{S}\times\hat{T} \to [0,\infty)$ by

$$
\begin{gathered}
\hat{\rho}(\hat{x},\hat{y}) := \frac{1}{m}\sum_{k=1}^m \rho(x_k,y_k), \quad \hat{x} := (x_1,\dots,x_m),\ \hat{y} := (y_1,\dots,y_m), \\
\hat{B}(\hat{x}_1^n,D) := \left\{ \hat{y}_1^n \in \hat{T}^n : \frac{1}{n}\sum_{k=1}^n \hat{\rho}(\hat{x}_k,\hat{y}_k) \le D \right\}, \\
\hat{\Lambda}_1(\lambda) := E_{\hat{P}} \log E_{\hat{Q}} e^{\lambda\hat{\rho}(\hat{X}_1,\hat{Y}_1)}, \qquad \hat{\Lambda}_1^*(D) := \sup_{\lambda\le 0}\left[\lambda D - \hat{\Lambda}_1(\lambda)\right], \\
\hat{\rho}_{\hat{Q}}(\hat{x}) := \operatorname*{ess\,inf}_{\hat{Q}} \hat{\rho}(\hat{x},\hat{Y}_1), \qquad \hat{D}_{\min} := E\hat{\rho}_{\hat{Q}}(\hat{X}_1), \qquad \hat{D}_{\text{ave}} := E\hat{\rho}(\hat{X}_1,\hat{Y}_1),
\end{gathered}
$$

where $\hat{P}$ is the distribution of $(\hat{X}_k)_{k\ge1}$, $\hat{X}_k := (X_{r+1+(k-1)m},\dots,X_{r+km})$, and $\hat{Q}$ is the distribution of $(\hat{Y}_k)_{k\ge1}$, $\hat{Y}_k := (Y_{r+1+(k-1)m},\dots,Y_{r+km})$. Notice that $\hat{P}$, $\hat{Q}$ and all of the above quantities do not depend on r (except of course for the specific realizations of $(\hat{X}_k)_{k\ge1}$ and $(\hat{Y}_k)_{k\ge1}$). Let $(\tilde{Y}_k)_{k\ge1}$ be i.i.d. random variables with joint distribution $\tilde{Q}$ on $(\hat{T}^{\mathbb{N}},\hat{\mathcal{T}}^{\mathbb{N}})$ such that $\tilde{Y}_1$ has the same distribution as $\hat{Y}_1$. Notice that we can replace $\hat{Q}$ with $\tilde{Q}$ in the definitions of $\hat{\Lambda}_1$, $\hat{\Lambda}_1^*$, $\hat{D}_{\min}$ and $\hat{D}_{\text{ave}}$ without changing anything since they only depend on the distribution of $\hat{Y}_1$. Notice also that $\hat{D}_{\min} = D_{\min}$ (because of the mixing properties of Q) and $\hat{D}_{\text{ave}} = D_{\text{ave}}$.

Choose an integer valued random variable M so that $r + Mm \overset{\text{a.s.}}{\ge} N$. Using (3.13) gives

$$
\begin{aligned}
\limsup_{s\to\infty} -\frac{1}{r+sm}\log Q(B(X_1^{r+sm},D)) &\overset{\text{a.s.}}{\le} \limsup_{s\to\infty} -\frac{1}{(s-M)m}\log Q(B(X_{r+Mm+1}^{r+sm},D)) \\
= \limsup_{s\to\infty} -\frac{1}{(s-M)m}\log \hat{Q}(\hat{B}(\hat{X}_{M+1}^s,D)) &= \limsup_{s\to\infty} -\frac{1}{sm}\log \hat{Q}(\hat{B}(\hat{X}_{M+1}^{M+s},D)) \\
\le \limsup_{s\to\infty} -\frac{1}{sm}\log \tilde{Q}(\hat{B}(\hat{X}_{M+1}^{M+s},D)) + \frac{s\log C}{sm}, & \qquad (3.14)
\end{aligned}
$$

where we switched from $\hat{Q}$ to $\tilde{Q}$ in the last step.

We would like to be able to immediately apply (3.11) to the final expression in (3.14) to get

$$
\limsup_{s\to\infty} -\frac{1}{r+sm}\log Q(B(X_1^{r+sm},D)) \overset{\text{a.s.}}{\le} \frac{1}{m}\hat{\Lambda}_1^*(D) + \frac{\log C}{m}.
$$

Unfortunately, unless P is totally ergodic, $\hat{P}$ need not be ergodic, although it is stationary, and we cannot immediately apply (3.7). However, Berger (1971) [2][pp. 278–9] and

Gallager (1968) [7][pp. 495–497] show that $\hat{P}$ can be decomposed into m (not necessarily unique) equally likely, stationary and ergodic components[5]

$$\hat{P} = \frac{1}{m}\sum_{j=1}^{m} \hat{P}^{(j)}.$$

Letting $J_r \in \{1, \ldots, m\}$ be the random variable that (a.s.) indicates which ergodic component generated $(\hat{X}_k)_{k\geq 1}$, we can apply (3.11) to (3.14) separately for each ergodic component to get

$$\limsup_{s\to\infty} -\frac{1}{r+sm}\log Q(B(X_1^{r+sm}, D)) \overset{\text{a.s.}}{\leq} \frac{1}{m}\hat{\Lambda}^*_{1,J_r}(D) + \frac{\log C}{m}, \tag{3.15}$$

where

$$\hat{\Lambda}_{1,j}(\lambda) := E_{\hat{P}^{(j)}} \log E_{\hat{Q}} e^{\lambda\hat{\rho}(\hat{X}_1,\hat{Y}_1)} = E_{\hat{P}^{(j)}} \log E_{\tilde{Q}} e^{\lambda\hat{\rho}(\hat{X}_1,\tilde{Y}_1)},$$
$$\hat{\Lambda}^*_{1,j}(D) := \sup_{\lambda\leq 0}\left[\lambda D - \hat{\Lambda}_{1,j}(\lambda)\right].$$

Recall that $0 \leq r < m$ was arbitrary, so (3.15) gives

$$\begin{aligned}
\limsup_{n\to\infty} -\frac{1}{n}\log Q(B(X_1^n, D)) &= \max_{0\leq r<m}\limsup_{s\to\infty} -\frac{1}{r+sm}\log Q(B(X_1^{r+sm}, D)) \\
&\overset{\text{a.s.}}{\leq} \max_{0\leq r<m}\frac{1}{m}\hat{\Lambda}^*_{1,J_r}(D) + \frac{\log C}{m} \leq \max_{1\leq j\leq m}\frac{1}{m}\hat{\Lambda}^*_{1,j}(D) + \frac{\log C}{m}.
\end{aligned} \tag{3.16}$$

We will now use the same notation and blocking technique to show that

$$\max_{1\leq j\leq m}\frac{1}{m}\hat{\Lambda}^*_{1,j}(D) \leq \Lambda^*_\infty(D) + \frac{\log C}{m}. \tag{3.17}$$

Indeed, combining (3.16) and (3.17) and letting $m \to \infty$ gives (3.12) as desired.

Beginning with (3.1) and using the same arguments as before gives ($\lambda \leq 0$)

$$\begin{aligned}
\Lambda_\infty(\lambda) &\overset{\text{a.s.}}{=} \lim_{n\to\infty}\frac{1}{n}\log E_Q e^{\lambda\sum_{k=1}^n \rho(X_k,Y_k)} \\
&= \lim_{s\to\infty}\frac{1}{r+sm}\log E_Q e^{\lambda\sum_{k=1}^r \rho(X_k,Y_k)} e^{\lambda\sum_{k=r+1}^{r+sm}\rho(X_k,Y_k)} \\
&\leq \liminf_{s\to\infty}\frac{1}{r+sm}\left[\log E_Q e^{\lambda\sum_{k=1}^r \rho(X_k,Y_k)} + \log C + \log E_Q e^{\lambda\sum_{k=r+1}^{r+sm}\rho(X_k,Y_k)}\right] \qquad (3.18)\\
&\overset{\text{a.s.}}{=} \liminf_{s\to\infty}\frac{1}{sm}\log E_Q e^{\lambda\sum_{k=r+1}^{r+sm}\rho(X_k,Y_k)} = \liminf_{s\to\infty}\frac{1}{sm}\log E_{\hat{Q}} e^{\lambda m\sum_{k=1}^s \hat{\rho}(\hat{X}_k,\hat{Y}_k)} \\
&\leq \liminf_{s\to\infty}\frac{1}{sm}\log E_{\tilde{Q}} e^{\lambda m\sum_{k=1}^s \hat{\rho}(\hat{X}_k,\tilde{Y}_k)} + \frac{s\log C}{sm} \\
&= \liminf_{s\to\infty}\frac{1}{sm}\sum_{k=1}^s \log E_{\tilde{Q}} e^{\lambda m\hat{\rho}(\hat{X}_k,\tilde{Y}_1)} + \frac{\log C}{m} \overset{\text{a.s.}}{=} \frac{1}{m}\hat{\Lambda}_{1,J_r}(m\lambda) + \frac{\log C}{m}, \qquad (3.19)
\end{aligned}$$

[5]Here is an illustrative example: $(X_1, X_2, \ldots)$ is equally likely either $(0, 1, 0, 1, \ldots)$ or $(1, 0, 1, 0, \ldots)$, which is stationary and ergodic (an irreducible, periodic Markov chain). For $m = 2$, $(\hat{X}_1, \hat{X}_2, \ldots)$ is equally likely either $((0, 1), (0, 1), \ldots)$ or $((1, 0), (1, 0), \ldots)$ for any r, which is stationary but not ergodic (a mixture of two different constant, and thus stationary and ergodic, sequences).

where we are able to ignore the first term in (3.18) because it has finite expectation (namely $r\Lambda_r(\lambda)$) and is thus a.s. finite. The last equality comes from the ergodic theorem applied to each ergodic component of $\hat{P}$. As we vary r, the random variables $(J_r)_{0\le r<m}$ indicate (a.s.) each of the distinct ergodic components at least once [2, 7]. Thus (3.19) implies that

$$\Lambda_\infty(\lambda) \le \min_{1\le j\le m} \frac{1}{m}\hat{\Lambda}_{1,j}(m\lambda) + \frac{\log C}{m}. \tag{3.20}$$

We can apply this bound to get

$$\begin{aligned}
\max_{1\le j\le m} \frac{1}{m}\hat{\Lambda}^*_{1,j}(D) &= \max_{1\le j\le m} \sup_{\lambda\le 0}\left[\frac{\lambda}{m}D - \frac{1}{m}\hat{\Lambda}_{1,j}(\lambda)\right] \\
&= \max_{1\le j\le m} \sup_{\lambda\le 0}\left[\lambda D - \frac{1}{m}\hat{\Lambda}_{1,j}(m\lambda)\right] = \sup_{\lambda\le 0}\left[\lambda D - \min_{1\le j\le m}\frac{1}{m}\hat{\Lambda}_{1,j}(m\lambda)\right] \\
&\le \sup_{\lambda\le 0}\left[\lambda D - \Lambda_\infty(\lambda)\right] + \frac{\log C}{m} = \Lambda^*_\infty(D) + \frac{\log C}{m}.
\end{aligned}$$

This gives (3.17) and completes the proof of the upper bound when $D_{\min} < D \le D_{\text{ave}}$.

3.4 Proof: Upper bound, $\mathbf{D > D_{\text{ave}}}$

In this section we assume that $D > D_{\text{ave}}$, which means that we must have $D_{\text{ave}} < \infty$. Chebyshev's inequality gives

$$Q\left\{y_1^n : \frac{1}{n}\sum_{k=1}^n \rho(X_k, y_k) > D\right\} \le \frac{1}{nD}\sum_{k=1}^n E_Q\rho(X_k, Y_1) \overset{\text{a.s.}}{\to} \frac{D_{\text{ave}}}{D} < 1$$

as $n\to\infty$ by the ergodic theorem. So

$$\begin{aligned}
-\frac{1}{n}\log Q(B(X_1^n, D)) &= -\frac{1}{n}\log\left[1 - Q\left\{y_1^n : \frac{1}{n}\sum_{k=1}^n \rho(X_k, y_k) > D\right\}\right] \\
&\le -\frac{1}{n}\log\left[1 - \frac{1}{nD}\sum_{k=1}^n E_Q\rho(X_k, Y_1)\right] \overset{\text{a.s.}}{\to} 0.
\end{aligned}$$

Thus, for $D > D_{\text{ave}}$ we have

$$\limsup_{n\to\infty} -\frac{1}{n}\log Q\left(B(X_1^n, D)\right) \overset{\text{a.s.}}{\le} 0 \le \Lambda^*_\infty(D)$$

and this completes the proof of the upper bound when $D > D_{\text{ave}}$.

3.5 Proof: $\mathbf{D = D_{\min}}$

So far we have established the lower bound in all cases and the upper bound in all cases except for the situation where $D = D_{\min}$ and $\Lambda^*_\infty(D_{\min})$ is finite. In this section and the next two subsections we assume that $D = D_{\min} < \infty$ and $\Lambda^*_\infty(D_{\min}) < \infty$. Define $A(x_1^n)$ as in (2.6). The mixing properties of Q show that

$$\operatorname*{ess\,inf}_Q \frac{1}{n}\sum_{k=1}^n \rho(x_k, Y_k) = \frac{1}{n}\sum_{k=1}^n \rho_Q(x_k), \tag{3.21}$$

so the ergodic theorem gives

$$\lim_{n\to\infty} \operatorname*{ess\,inf}_{Q} \frac{1}{n}\sum_{k=1}^{n} \rho(X_k, Y_k) \overset{\text{a.s.}}{=} D_{\min}.$$

The convergence also holds in expectation. (3.21) allows us to compute

$$\begin{aligned}
&-\log Q(A(x_1^{n+m})) + \log C \\
&\quad = -\log Q\left\{y_1^{n+m} : \rho(x_k, y_k) = \rho_Q(x_k), 1 \le k \le n+m\right\} + \log C \\
&\quad \le -\log Q(A(x_1^n)) + \log C - \log Q(A(x_{n+1}^{n+m})) + \log C.
\end{aligned}$$

The appendix shows that $E[-\log Q(A(X_1^n))] = n\Lambda_n^*(D_{\min})$ which is finite since $\Lambda_\infty^*(D_{\min})$ is finite. The subadditive ergodic theorem and (2.1) give

$$\lim_{n\to\infty} -\frac{1}{n}\log Q(A(x_1^n)) \overset{\text{a.s.}}{=} \Lambda_\infty^*(D_{\min}). \tag{3.22}$$

The convergence also holds in expectation.

3.5.1 Proof: $\mathbf{D = D_{\min}}$, constant $\rho_{\mathbf{Q}}$

If $\rho_Q(X_1)$ is a.s. constant, then $Q(A(X_1^n)) \overset{\text{a.s.}}{=} Q(B(X_1^n, D_{\min}))$ and (3.22) gives (2.4). Notice that we have now completed the proof of (2.4) in each of the cases $D \neq D_{\min}$, $\Lambda_\infty^*(D) = \infty$ and $\rho_Q(X_1)$ a.s. constant. As we will see in the next section, if all of these conditions fail simultaneously, then (2.4) fails as well.

3.5.2 Proof: $\mathbf{D = D_{\min}}$, non-constant $\rho_{\mathbf{Q}}$

Here we investigate the behavior of $\log Q(B(X_1^n, D))$ when $D = D_{\min} < \infty$, $\Lambda_\infty^*(D_{\min}) < \infty$ and $\rho_Q(X_1)$ is not a.s. constant. This makes use of recurrence properties for random walks with stationary and ergodic increments.[6] What we need is summarized in the following:

Lemma 3.1. *Let $(X_n)_{n\ge1}$ be a real-valued stationary and ergodic process and define $Z_n := \sum_{k=1}^n X_k$, $n \ge 1$. If $EX_1 = 0$ and $\operatorname{Prob}\{X_1 \neq 0\} > 0$, then $\operatorname{Prob}\{Z_n > 0 \text{ i.o.}\} > 0$ and $\operatorname{Prob}\{Z_n \ge 0 \text{ i.o.}\} = 1$.*

Proof. Define $Z_0 := 0$. $(Z_n)_{n\ge0}$ is a random walk with stationary and ergodic increments. Kesten (1975) [10] shows that $\{\liminf_n n^{-1}Z_n > 0\}$ and $\{Z_n \to \infty\}$ differ by a null set. The ergodic theorem gives $\operatorname{Prob}\{n^{-1}Z_n \to 0\} = 1$, so $\operatorname{Prob}\{Z_n \to \infty\} = 0$. Similarly, by considering the process $-Z_n$, we see that $\operatorname{Prob}\{Z_n \to -\infty\} = 0$.

Now $\{|Z_n| \to \infty\}$ is invariant and must have probability 0 or 1. If it has probability 1, then since we cannot have $Z_n \to \infty$ or $Z_n \to -\infty$ we must have Z_n oscillating between increasingly larger positive and negative values, which means $\operatorname{Prob}\{Z_n > 0 \text{ i.o.}\} = 1$ and completes the proof.

Suppose $\operatorname{Prob}\{|Z_n| \to \infty\} = 0$. Define

$$N(A) := \sum_{n\ge0} I_A(Z_n), \quad A \subset \mathbb{R},$$

[6] $(Z_n)_{n\ge0}$ is a random walk with stationary and ergodic increments [1] if $Z_0 := 0$ and $Z_n := \sum_{k=1}^n X_k$, $n \ge 1$, for some stationary and ergodic sequence $(X_n)_{n\ge1}$.

to be the number of times the random walk visits the set A. Berbee (1979) [1][Corollary 2.3.4] shows that either $N(J) < \infty$ a.s. for all bounded intervals J or $\{N(J) = 0\} \cup \{N(J) = \infty\}$ has probability 1 for all intervals J (open or closed, bounded or unbounded, but not a single point). By assumption $|Z_n| \not\to \infty$, so we can rule out the first possibility. Since $\mathrm{Prob}\{Z_0 = 0\} = 1$, we see that for any interval J containing $\{0\}$ we must have $\mathrm{Prob}\{N(J) = \infty\} = 1$. In particular, taking $J := [0, \infty)$ shows that $\mathrm{Prob}\{Z_n \geq 0 \text{ i.o.}\} = 1$. Similarly, taking $J := (0, \infty)$ shows that $\mathrm{Prob}\{Z_n > 0 \text{ i.o.}\} = \mathrm{Prob}\{N(J) = \infty\} = \mathrm{Prob}\{N(J) > 0\} \geq \mathrm{Prob}\{X_1 > 0\} > 0$. □

Returning to the main argument,

$$\begin{aligned} -\log Q(B(X_1^n, D_{\min})) &\geq -\log Q\left\{y_1^n : \sum_{k=1}^n \rho_Q(X_k) \leq nD_{\min}\right\} \\ &= \begin{cases} 0 & \text{if } \sum_{k=1}^n \rho_Q(X_k) \leq nD_{\min} \\ \infty & \text{if } \sum_{k=1}^n \rho_Q(X_k) > nD_{\min} \end{cases} = \begin{cases} 0 & \text{if } Z_n \leq 0 \\ \infty & \text{if } Z_n > 0 \end{cases}, \end{aligned} \tag{3.23}$$

where $Z_n := \sum_{k=1}^n (\rho_Q(X_k) - D_{\min})$. Lemma 3.1 shows that $\mathrm{Prob}\{Z_n > 0 \text{ i.o.}\} > 0$. This and (3.23) prove (2.5a).

Lemma 3.1 also shows that $\mathrm{Prob}\{Z_n \leq 0 \text{ i.o.}\} = 1$. Let $(n_m)_{m\geq 1}$ be the (a.s.) infinite, random subsequence of $(n)_{n\geq 1}$ such that $Z_n \leq 0$. Note that

$$\sum_{k=1}^{n_m} \rho_Q(X_k) \leq n_m D_{\min}$$

so

$$\begin{aligned} -\log Q(B(X_1^{n_m}, D_{\min})) &\leq -\log Q\left\{y_1^{n_m} : \sum_{k=1}^{n_m} \rho(X_k, y_k) \leq \sum_{k=1}^{n_m} \rho_Q(X_k)\right\} \\ &= -\log Q(A(X_1^{n_m})). \end{aligned} \tag{3.24}$$

Now, the final expression in (3.24) is a.s. finite because $E[-\log Q(A(X_1^n))] = n\Lambda_n^*(D_{\min}) < \infty$. This proves (2.5b) and shows that $(n_m)_{m\geq 1}$ satisfies the claims of the theorem. (3.22) and (3.24) also show that

$$\begin{aligned} \limsup_{m\to\infty} -\frac{1}{n_m} \log Q(B(X_1^{n_m}, D_{\min})) &\leq \limsup_{m\to\infty} -\frac{1}{n_m} \log Q(A(X_1^{n_m})) \\ &\leq \limsup_{n\to\infty} -\frac{1}{n} \log Q(A(X_1^n)) \overset{\text{a.s.}}{=} \Lambda_\infty^*(D_{\min}). \end{aligned}$$

Combining this with the lower bound (3.5) proves (2.5c) and completes the proof of all parts of (2.5) and Theorem 2.1.

A Appendix

A common assumption in the literature is that ρ is either bounded or satisfies some moment conditions. Since we do not assume these things here, we need to reverify many properties of Λ and Λ^* that can be found elsewhere under these stronger conditions. We also neglected any measurability issues in the main text, but we deal with them here. Let us begin with the following Lemma which comes mostly from Dembo and Zeitouni (1998) [5].

Lemma A.1. *[5] Let Z be a real-valued, nonnegative random variable. Define*

$$\Lambda(\lambda) := \log Ee^{\lambda Z}.$$

Λ is nondecreasing and convex. Λ is finite, nonpositive and C^∞ on $(-\infty, 0)$ with

$$\lim_{\lambda \uparrow 0} \Lambda(\lambda) = \Lambda(0) = 0 \quad \textit{and} \quad \Lambda'(\lambda) = \frac{EZe^{\lambda Z}}{Ee^{\lambda Z}}, \quad \lambda < 0.$$

Λ' is finite, nonnegative, nondecreasing and C^∞ on $(-\infty, 0)$ with

$$\lim_{\lambda \downarrow -\infty} \Lambda'(\lambda) = \operatorname{ess\,inf} Z \qquad \textit{and} \qquad \lim_{\lambda \uparrow 0} \Lambda'(\lambda) = EZ.$$

If $\operatorname{ess\,inf} Z < EZ$, then Λ is strictly convex on $(-\infty, 0)$.

Proof. Since Z is nonnegative and real-valued, Λ is nondecreasing everywhere and Λ is finite and nonpositive on $(-\infty, 0]$ with $\Lambda(0) = 0$. Dembo and Zeitouni (1998) [5][Lemma 2.2.5, Example 2.2.24] show that Λ is convex everywhere and C^∞ on $(-\infty, 0)$ with $\Lambda'(\lambda)$ as stated. This implies that Λ' is nondecreasing and C^∞ (and thus finite) on $(-\infty, 0)$. The dominated convergence theorem shows that $\Lambda(\lambda) \uparrow 0$ as $\lambda \uparrow 0$.

Clearly Λ' is nonnegative. The monotone convergence theorem applied to the numerator and denominator separately in the expression for Λ' establishes that $\lim_{\lambda \uparrow 0} \Lambda'(\lambda) = EZ$.

We have

$$\Lambda'(\lambda) = \frac{EZe^{\lambda Z}}{Ee^{\lambda Z}} \geq \frac{E(\operatorname{ess\,inf} Z)e^{\lambda Z}}{Ee^{\lambda Z}} = (\operatorname{ess\,inf} Z)\frac{Ee^{\lambda Z}}{Ee^{\lambda Z}} = \operatorname{ess\,inf} Z. \tag{A.1}$$

Since Λ is convex, differentiable and nondecreasing, we also have (for $\lambda < 0$)

$$\Lambda'(\lambda) \leq \frac{\Lambda(0) - \Lambda(\lambda)}{0 - \lambda} = \frac{\Lambda(\lambda)}{\lambda} = \log \left[Ee^{\lambda Z}\right]^{1/\lambda} = -\log \left[E(e^{-Z})^{|\lambda|}\right]^{1/|\lambda|} = -\log \left\|e^{-Z}\right\|_{|\lambda|},$$

where $\|\cdot\|_p$ denotes the L^p norm. Taking limits gives

$$\begin{aligned}\lim_{\lambda \to -\infty} \Lambda'(\lambda) &\leq -\log \left[\lim_{\lambda \to -\infty} \left\|e^{-Z}\right\|_{|\lambda|}\right] = -\log \left\|e^{-Z}\right\|_\infty = -\log \operatorname{ess\,sup} e^{-Z} \\ &= -\log e^{-\operatorname{ess\,inf} Z} = \operatorname{ess\,inf} Z.\end{aligned}$$

Combining this with (A.1) establishes that $\lim_{\lambda \to -\infty} \Lambda'(\lambda) = \operatorname{ess\,inf} Z$.

An easy application of the dominated convergence theorem shows that $\frac{d}{d\lambda} EZ^n e^{\lambda Z} = EZ^{n+1} e^{\lambda Z}$ for $\lambda < 0$ and $n \geq 0$. So for $\lambda < 0$

$$\Lambda''(\lambda) = \frac{Ee^{\lambda Z} EZ^2 e^{\lambda Z} - EZe^{\lambda Z} EZe^{\lambda Z}}{[Ee^{\lambda Z}]^2} = \frac{EZ^2 e^{\lambda Z}}{Ee^{\lambda Z}} - \left(\frac{EZe^{\lambda Z}}{Ee^{\lambda Z}}\right)^2.$$

The Cauchy-Schwarz inequality shows that $\Lambda'' \geq 0$ with equality if and only if Z is (a.s.) constant. So $\Lambda'' > 0$ on $(-\infty, 0)$ whenever $\operatorname{ess\,inf} Z < EZ$. $\square$

A.1 Measurability issues

Halmos (1966) [9] shows that we can integrate out one variable in a product measurable function and still obtain a measurable function. It is important that this is the smallest product σ-algebra and not the completion of it w.r.t. some product measure. This immediately establishes the measurability (in x) of

$$Q(B(x,D)), \qquad E_Q e^{\lambda\rho(x,Y)}, \qquad E_Q \rho(x,Y)e^{\lambda\rho(x,Y)}$$

and any nice functions of them, where we denote $Y := Y_1$ to clean up the notation.

In particular,

$$\rho_Q^\lambda(x) := \frac{E_Q \rho(x,Y)e^{\lambda\rho(x,Y)}}{E_Q e^{\lambda\rho(x,Y)}}$$

is measurable (for $\lambda \le 0$). Lemma A.1 (with $Z := \rho(x,Y)$ for fixed x so that $\rho_Q^\lambda(x) = \Lambda'(\lambda)$) shows that $\rho_Q^\lambda(x) \downarrow \rho_Q(x)$ as $\lambda \downarrow -\infty$ for each x. This implies that ρ_Q is measurable.

Noting that $\sum_{k=1}^n \rho(x_1^n, y_1^n)$ is product measurable on $S^n \times T^n$ lets us repeat the above steps for any n. This clears up the measurability issues that we avoided in the main text.

A.2 Properties of Λ_n

Here we list some properties of Λ_n that hold for any $1 \le n < \infty$. Clearly Λ_n is nondecreasing with $\Lambda_n(0) = 0$. Λ_n is also convex. From this we see that Λ_n is either everywhere $-\infty$ on $(-\infty, 0)$ or it is finite on $(-\infty, 0]$. In the rest of this section we will only be considering the latter case. Note that if $\Lambda_n^*(D) < \infty$ for any D, then Λ_n must be finite on $(-\infty, 0]$.

Λ_n is a proper ($> -\infty$) closed (l.sc.) convex function [13] and continuous from the left. It is finite and C^1 on $(-\infty, 0)$. Λ_n', the derivative with w.r.t. λ, is nondecreasing and

$$\Lambda_n'(\lambda) = \frac{1}{n} E_P \left[\frac{E_Q \sum_{j=1}^n \rho(X_j, Y_j) e^{\lambda \sum_{k=1}^n \rho(X_k, Y_k)}}{E_Q e^{\lambda \sum_{k=1}^n \rho(X_k, Y_k)}} \right], \quad \lambda < 0,$$

$$\sup_{\lambda<0} \Lambda_n'(\lambda) = \lim_{\lambda \uparrow 0} \Lambda_n'(\lambda) = \sup_{\lambda<0} \frac{\Lambda_n(\lambda)}{\lambda} = \lim_{\lambda\uparrow 0} \frac{\Lambda_n(\lambda)}{\lambda} = D_{\text{ave}},$$

$$\inf_{\lambda<0} \Lambda_n'(\lambda) = \lim_{\lambda \to -\infty} \Lambda_n'(\lambda) = \inf_{\lambda<0} \frac{\Lambda_n(\lambda)}{\lambda} = \lim_{\lambda \to -\infty} \frac{\Lambda_n(\lambda)}{\lambda} = D_{\min}.$$

Furthermore, if $D_{\min} < D_{\text{ave}}$, then $\Lambda_n(\cdot)$ is strictly convex on $(-\infty, 0)$.

From the definition it is easy to see that Λ_∞ will also be nondecreasing and convex with $\Lambda_\infty(0) = 0$. Whenever $\Lambda_\infty^*(D) < \infty$ for some D, we must have Λ_∞ finite on $(-\infty, 0]$, which means Λ_∞ is continuous on $(-\infty, 0)$.

A.2.1 Proofs

We will now prove these claims using Lemma A.1. For a fixed $x_1^n \in S^n$, define the random variable $Z(x_1^n) := \sum_{k=1}^n \rho(x_k, Y_k)$. Let

$$\Lambda(x_1^n, \lambda) := \log E e^{\lambda Z(x_1^n)}$$

so that $\Lambda_n(\lambda) = \frac{1}{n} E_P \Lambda(X_1^n, \lambda)$. Lemma A.1 shows that $\Lambda(x_1^n, \cdot)$ has all of the properties that we want Λ to have. For example, $\Lambda(x_1^n, \cdot)$ is convex for each x_1^n, so Λ_n is convex also.

The rest of this proof is justification that these properties continue to hold after taking expectations (as long Λ_n is never $-\infty$).

Suppose that $\Lambda_n(\lambda)$ is finite for some $\lambda < 0$. Choose $\lambda' < \lambda$. The convexity of $\Lambda(x_1^n, \cdot)$ gives

$$\Lambda(x_1^n, \lambda) \leq \frac{\lambda}{\lambda'}\Lambda(x_1^n, \lambda') + \left(1 - \frac{\lambda}{\lambda'}\right)\Lambda(x_1^n, 0) = \frac{\lambda}{\lambda'}\Lambda(x_1^n, \lambda') \leq 0.$$

Taking expectations shows that $\Lambda_n(\lambda') > -\infty$. Since $\lambda' < \lambda$ was arbitrary and since Λ_n is increasing, we see that $\Lambda_n > -\infty$ everywhere and finite on $(-\infty, 0]$. Thus Λ_n is a proper convex function and Λ_n is continuous on $(-\infty, 0)$.

Since $0 \geq \Lambda(x_1^n, \lambda) \geq \Lambda(x_1^n, \delta) > -\infty$ for $\delta \leq \lambda \leq 0$, since $0 \geq E_P\Lambda(X_1^n, \delta) = n\Lambda_n(\delta) > -\infty$ and since $\Lambda(x_1^n, \lambda) \uparrow 0$ as $\lambda \uparrow 0$, the dominated convergence theorem gives $\Lambda_n(\lambda) \uparrow \Lambda_n(0) = 0$ as $\lambda \uparrow 0$. Since $\Lambda(x_1^n, \cdot)$ is nonnegative and increasing on $(0, \infty)$ for each x_1^n, the monotone convergence theorem shows that Λ_n is continuous from the left on $(0, \infty)$. So it is continuous from the left everywhere. Since it is nondecreasing, it is l.sc. and thus closed.

Since Λ_n is finite and convex on $(-\infty, 0)$, it has finite and nondecreasing right hand and left hand derivatives, $\overline{\Lambda}_n'$ and $\underline{\Lambda}_n'$, respectively, with the property that $\overline{\Lambda}_n' \geq \underline{\Lambda}_n'$ and $\underline{\Lambda}_n'(\lambda + \epsilon) \geq \overline{\Lambda}_n'(\lambda)$ for $\lambda < \lambda + \epsilon < 0$. When $\lambda - \epsilon < \lambda < 0$ we have

$$0 \leq \frac{\Lambda(x_1^n, \lambda) - \Lambda(x_1^n, \lambda - \epsilon)}{\epsilon} \uparrow \Lambda'(x_1^n, \lambda), \quad \text{as } \epsilon \downarrow 0,$$

so the monotone convergence theorem gives

$$\underline{\Lambda}_n'(\lambda) = \frac{1}{n}E_P\Lambda'(X_1^n, \lambda), \qquad \lambda < 0,$$

which is finite. When $\lambda < \lambda + \epsilon < 0$ we have

$$0 \leq \frac{\Lambda(x_1^n, \lambda + \epsilon) - \Lambda(x_1^n, \lambda)}{\epsilon} \leq \Lambda'(x_1^n, \lambda + \epsilon).$$

Since the right hand side has finite expectation, the dominated convergence theorem gives

$$\overline{\Lambda}_n'(\lambda) = \frac{1}{n}E_P\Lambda'(X_1^n, \lambda) = \underline{\Lambda}_n'(\lambda), \qquad \lambda < 0.$$

This shows that Λ_n is differentiable on $(-\infty, 0)$ and confirms the stated expression for Λ_n'. Since Λ_n is convex, the derivative Λ_n' is nondecreasing and continuous.

The monotone convergence theorem gives $\Lambda_n'(\lambda) \uparrow \frac{1}{n}E_P Z(X_1^n) = D_{\text{ave}}$ as $\lambda \uparrow 0$. The dominated convergence theorem gives $\Lambda_n'(\lambda) \downarrow \frac{1}{n}E_P \operatorname{ess\,inf} Z(X_1^n) = D_{\min}$ (because of the mixing properties of Q) as $\lambda \downarrow 0$. Suppose $D_{\min} < D_{\text{ave}}$. Then with positive probability $Z(X_1^n)$ is not a.s. constant and $\Lambda'(X_1^n, \cdot) > 0$ on $(-\infty, 0)$. Taking expectations shows that $\Lambda_n' > 0$ there also, so Λ_n is strictly convex on $(-\infty, 0)$.

Since Λ_n' is nondecreasing, $\sup_{\lambda<0} \Lambda_n'(\lambda) = \lim_{\lambda\uparrow 0} \Lambda_n'(\lambda)$ and similarly for the infimum. Noting that

$$\frac{\Lambda_n(\lambda) - \Lambda_n(0)}{\lambda - 0} = \frac{\Lambda_n(\lambda)}{\lambda}$$

is the slope of the cord above Λ_n from 0 to λ, we see that this slope is also nondecreasing in λ so we can interchange supremums and infimums with the appropriate limits as before. As $\lambda \uparrow 0$, this slope converges to the derivative at 0 (which we know is D_{ave}). As $\lambda \downarrow -\infty$ this slope converges to the limiting derivative (which we know is $D_{\min}$).

A.3 Properties of Λ_n^*

Suppose $1 \le n \le \infty$. Λ_n^* is convex, l.sc., nonnegative, nonincreasing and continuous from the right. $\Lambda_n^*(D) = \infty$ whenever $D < D_{\min}$ and $\Lambda_n^*(D) = 0$ whenever $D \ge D_{\text{ave}}$. If $D \le D_{\text{ave}}$, then $\Lambda_n^*(D) = \sup_{\lambda\in\mathbb{R}}[\lambda D - \Lambda_n(\lambda)]$. If $\Lambda_n^*(D) < \infty$ for some D, then Λ_n^* is finite and continuous on $(D_{\min}, \infty)$.

Suppose $1 \le n < \infty$. If $D_{\min} < \infty$, then $\Lambda_n^*(D_{\min}) = n^{-1}E[-\log Q(A(X_1^n))]$, where $A(x_1^n)$ is defined in (2.6). If $\Lambda_n^*(D) < \infty$ for some D, then Λ_n^* is C^1 on $(D_{\min}, \infty)$. Furthermore, if $D_{\min} < D_{\text{ave}}$, then Λ_n^* is strictly convex (and thus strictly decreasing) on $(D_{\min}, D_{\text{ave}})$ and for each $D \in (D_{\min}, D_{\text{ave}})$ there exists a unique $\lambda_D < 0$ such that $\Lambda_n^*(D) = \lambda_D D - \Lambda_n(\lambda_D)$.

A.3.1 Proofs

Now we will prove these claims. If $D < 0$, $\Lambda_n^*(D) = \infty$. Otherwise, for $D \ge 0$ and for $\lambda \le 0$, $D \mapsto \lambda D - \Lambda_n(\lambda)$ is a nonincreasing linear (and thus convex and l.sc.) function, so Λ_n^* is a supremum of nonincreasing, convex and l.sc. functions which is always nonincreasing, convex and l.sc. $\Lambda_n^*(D) \ge 0D - \Lambda_n(0) = 0D - 0 = 0$, so it is nonnegative. Since it is l.sc. and nonincreasing, it must be continuous from the right.

Define $\tilde\rho(x, y) := (\rho(x, y) - \rho_Q(x)) \vee 0$. For $n < \infty$, notice that

$$\sum_{k=1}^n \tilde\rho(x_k, Y_k) = \sum_{k=1}^n \rho(x_k, Y_k) - \sum_{k=1}^n \rho_Q(x_k) \quad Q \text{ a.s.}$$

We can write

$$\begin{aligned}\Lambda_n(\lambda) &= \frac{1}{n}E_P \log E_Q e^{\lambda\sum_{k=1}^n \tilde\rho(X_k,Y_k)+\lambda\sum_{k=1}^n \rho_Q(X_k)}\\ &= \frac{1}{n}E_P \log e^{\lambda\sum_{k=1}^n \rho_Q(X_k)} + \frac{1}{n}E_P \log E_Q e^{\lambda\sum_{k=1}^n \tilde\rho(X_k,Y_k)} = \lambda D_{\min} + \tilde\Lambda_n(\lambda),\end{aligned}$$

where $\tilde\Lambda_n$ is defined like Λ_n except with $\tilde\rho$ instead of ρ. This gives

$$\Lambda_n^*(D) = \sup_{\lambda\le 0}\left[\lambda(D - D_{\min}) - \tilde\Lambda_n(\lambda)\right]. \tag{A.2}$$

Since $\tilde\Lambda_n(\lambda) \le 0$ for $\lambda \le 0$, (A.2) gives the bound

$$\Lambda_n^*(D) \ge \sup_{\lambda\le 0} \lambda(D - D_{\min}).$$

When $D < D_{\min}$, this supremum is infinite. If $n = \infty$, then we get the same result by using (2.1). This shows that $D < D_{\min}$ implies $\Lambda_n^*(D) = \infty$.

Jensen's inequality gives

$$\Lambda_n(\lambda) = \frac{1}{n}E_P \log E_Q e^{\lambda\sum_{k=1}^n \rho(X_k,Y_k)} \ge \frac{1}{n}E_P E_Q \log e^{\lambda\sum_{k=1}^n \rho(X_k,Y_k)} = \lambda D_{\text{ave}} \tag{A.3}$$

for $n < \infty$ and the same bound holds for $n = \infty$ by applying (2.1). If $D \ge D_{\text{ave}}$, then necessarily $D_{\text{ave}} < \infty$ and (A.3) gives

$$\Lambda_n^*(D) = \sup_{\lambda\le 0}[\lambda D - \Lambda_n(\lambda)] \le \sup_{\lambda\le 0}[\lambda D - \lambda D_{\text{ave}}] = 0.$$

Since $\Lambda_n^* \geq 0$, this shows that $D \geq D_{\text{ave}}$ implies $\Lambda_n^*(D) = 0$.

If $D_{\text{ave}} = \infty$, then (A.3) shows that $\Lambda_n(\lambda) = \infty$ for $\lambda > 0$, so $\lambda D - \Lambda_n(\lambda) = -\infty < 0$ for all $\lambda > 0$ and all D. If $D \leq D_{\text{ave}} < \infty$ and $\lambda > 0$, then (A.3) gives

$$\lambda D - \Lambda_n(\lambda) \leq \lambda D_{\text{ave}} - \Lambda_n(\lambda) \leq \lambda D_{\text{ave}} - \lambda D_{\text{ave}} = 0.$$

So in both cases, when $D \leq D_{\text{ave}}$,

$$\sup_{\lambda > 0} [\lambda D - \Lambda_n(\lambda)] \leq 0 \leq \Lambda_n^*(D).$$

This shows that we can take the supremum over all of $\mathbb{R}$ in the definition of Λ_n^* whenever $D \leq D_{\text{ave}}$. This proof essentially comes from Dembo and Zeitouni (1998) [5][Lemma 2.2.5]. Notice that this means $\Lambda_n^*(D)$ is the conjugate of Λ_n at D as long as $D \leq D_{\text{ave}}$.

Now suppose that $\Lambda_n^*(D) < \infty$ for some D and that $n < \infty$. If we can show that Λ_n^* is finite on $(D_{\min}, D_{\text{ave}})$ then we will know that Λ_n^* is finite and continuous on $(D_{\min}, \infty)$, because it is convex everywhere and finite on $[D_{\text{ave}}, \infty)$. We can deal with the case $n = \infty$ by using (2.1) and the bounds in (3.6). So let us show that Λ_n^* is finite on $(D_{\min}, D_{\text{ave}})$. We can assume that $D_{\min} < D < D_{\text{ave}}$. Notice that the assumption $\Lambda_n^*(D) < \infty$ means that Λ_n has all of the nice properties detailed in Section A.2.

In particular, the strict convexity of Λ_n implies that there is a unique $\lambda_D < 0$ with $\Lambda_n'(\lambda_D) = D$. We have just seen that $\Lambda_n^*(D)$ is the conjugate of Λ_n at D, so Rockafellar (1970) [13][Theorem 23.5, Corollary 23.5.1, Theorem 25.1] gives

$$\Lambda_n^*(D) = \lambda_D D - \Lambda_n(\lambda_D) \quad \text{and} \quad \frac{d}{dD}\Lambda_n^*(D) = \lambda_D < 0.$$

This shows that Λ_n^* is finite, strictly convex and C^1 on $(D_{\min}, D_{\text{ave}})$. Since $\lambda_D \downarrow 0$ as $D \uparrow D_{\text{ave}}$, Λ_n^* is differentiable at 0 and so it is C^1 on $(D_{\min}, \infty)$.

The last thing we have to prove is the claim about $\Lambda_n^*(D_{\min})$ for $D_{\min} < \infty$ and $n < \infty$. (A.2) gives

$$\Lambda_n^*(D_{\min}) = \sup_{\lambda \leq 0} -\tilde{\Lambda}_n(\lambda) = \lim_{\lambda \downarrow -\infty} -\tilde{\Lambda}_n(\lambda),$$

because $\tilde{\Lambda}_n$ is nondecreasing. Applying the monotone convergence theorem and then the dominated convergence theorem and using the mixing properties of Q gives

$$\begin{aligned}
\Lambda_n^*(D_{\min}) &= \lim_{\lambda \downarrow -\infty} \frac{1}{n} E_P \left[-\log E_Q e^{\lambda \sum_{k=1}^n \tilde{\rho}(X_k, Y_k)} \right] \\
&= \frac{1}{n} E_P \left[-\log \left(\lim_{\lambda \downarrow -\infty} E_Q e^{\lambda \sum_{k=1}^n \tilde{\rho}(X_k, Y_k)} \right) \right] \\
&= \frac{1}{n} E_P \left[-\log E_Q \left(\lim_{\lambda \downarrow -\infty} e^{\lambda \sum_{k=1}^n \tilde{\rho}(X_k, Y_k)} \right) \right] \\
&= \frac{1}{n} E_P \left[-\log E_Q I_{\{y_1^n : \sum_{k=1}^n \tilde{\rho}(X_k, y_k) = 0\}}(Y_1^n) \right] = \frac{1}{n} E_P[-\log Q(A(X_1^n))].
\end{aligned}$$

A.4 Proof of Proposition 1.1

Suppose $Q(A_n) = 0$. Then by stationarity $\text{Prob}\left\{Y_{k+1}^{k+n} \in A_n, \text{any } k \geq 1\right\} = 0$ and

$$\text{Prob}\left\{\log W_n = \infty = -\log Q(A_n)\right\} = 1.$$

Similarly, if $Q(A_n) = 1$, then

$$\text{Prob}\{\log W_n = 0 = -\log Q(A_n)\} = 1.$$

So whenever A_n is trivial, we have $\text{Prob}\{\log W_n = -\log Q(A_n)\} = 1$.

The rest of this proof comes from Kontoyiannis (1998) [11]. Fix $(c_n)_{n\geq 1}$, $c_n \geq 0$, $\sum_n e^{-c_n} < \infty$. For any $K \geq 0$ and A_n not trivial, we have

$$\text{Prob}\{W_n < K\} \leq \sum_{1\leq k<K} \text{Prob}\{W_n = k\} \leq \sum_{1\leq k<K} \text{Prob}\left\{Y_k^{n+k-1} \in A_n\right\} \leq KQ(A_n).$$

With $K := e^{-\log Q(A_n)-c_n}$ this gives

$$\text{Prob}\{\log W_n < -\log Q(A_n) - c_n\} \leq \begin{cases} 0 & \text{if } Q(A_n) \text{ is trivial} \\ e^{-\log Q(A_n)-c_n}Q(A_n) & \text{otherwise} \end{cases} \leq e^{-c_n}.$$

Since this is summable, the Borel-Cantelli Lemma shows that

$$\text{Prob}\{\log W_n < -\log Q(A_n) - c_n \text{ i.o.}\} = 0$$

which implies that

$$\text{Prob}\{\log W_n \geq -\log Q(A_n) - c_n \text{ eventually}\} = 1.$$

Notice that this only uses the stationarity of Q.

Using the ψ^--mixing properties of Q to choose $d \geq 1$ and $C > 1$ large enough that $Q(A)Q(B) \leq CQ(A\cap B)$ whenever $A \in \sigma(Y_1^n)$ and $B \in \sigma(Y_{n+d}^\infty)$ for some n. Suppose A_n is not trivial. Define $\tilde{K} := (K-1)/(n+d)$ and $B_k := \left\{Y_k^{k+n-1} \notin A_n\right\}$ to get

$$\begin{aligned}
\text{Prob}\{W_n > K\} &= \text{Prob}\left\{\bigcap_{1\leq k\leq K} B_k\right\} \leq \text{Prob}\left\{\bigcap_{0\leq j\leq \tilde{K}} B_{j(n+d)+1}\right\} \\
&= \text{Prob}\{B_1\} \prod_{1\leq j\leq \tilde{K}} \text{Prob}\left\{B_{j(n+d)+1} \,\middle|\, B_{i(n+d)+1}, 0 \leq i < j\right\} \\
&= [1 - \text{Prob}\{B_1^c\}] \prod_{1\leq j\leq \tilde{K}} \left[1 - \text{Prob}\left\{B_{j(n+d)+1}^c \,\middle|\, B_{i(n+d)+1}, 0 \leq i < j\right\}\right] \\
&\leq [1 - Q(A_n)] \prod_{1\leq j\leq \tilde{K}} \left[1 - C^{-1}Q(A_n)\right] \leq \left[1 - C^{-1}Q(A_n)\right]^{\tilde{K}}.
\end{aligned}$$

With $K := e^{-\log Q(A_n)+c_n+\log n}$ this gives

$$\begin{aligned}
&\text{Prob}\{\log W_n > -\log Q(A_n) + c_n + \log n\} \\
&\quad \leq \begin{cases} 0 & \text{if } Q(A_n) \text{ is trivial} \\ [1 - C^{-1}Q(A_n)]^{(Q(A_n)^{-1}ne^{c_n}-1)/(n+d)} & \text{otherwise} \end{cases} \leq \alpha^{((n-1)e^{c_n})/(C(n+d))} \\
&\quad \leq e^{-c_n} \qquad \text{for all } n \text{ large enough,}
\end{aligned}$$

where $\alpha := \sup_{0<x\leq C^{-1}}[1-x]^{1/x} < 1$. The final inequality is easy to see by taking logarithms and noting that $c_n \to \infty$.

Again, this is summable by assumption and we can see that

$$\text{Prob}\{\log W_n \leq -\log Q(A_n) + c_n + \log n \text{ eventually}\} = 1.$$

Acknowledgments

I want to thank I. Kontoyiannis and M. Madiman for many useful comments. I. Kontoyiannis, especially, for invaluable advice and for suggesting the problems that led to this paper.

References

[1] H.C.P. Berbee. *Random Walks with Stationary Increments and Renewal Theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1979.

[2] Toby Berger. *Rate Distortion Theory*. Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

[3] Zhiyi Chi. The first-order asymptotic of waiting times with distortion between stationary processes. *IEEE Transactions on Information Theory*, 47(1):338–347, January 2001.

[4] Amir Dembo and Ioannis Kontoyiannis. Source coding, large deviations, and approximate pattern matching. *IEEE Transactions on Information Theory*, 48(6):1590–1615, June 2002.

[5] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, second edition, 1998.

[6] Frank den Hollander. *Large Deviations*. American Mathematical Society, Providence, 2000.

[7] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[8] Robert M. Gray. *Probability, Random Processes, and Ergodic Properties*. Springer-Verlag, New York, 1988.

[9] Paul R. Halmos. *Measure Theory*. Van Nostrand, Princeton, 1966.

[10] Harry Kesten. Sums of stationary sequences cannot grow slower than linearly. *Proceedings of the American Mathematical Society*, 49(1):205–211, May 1975.

[11] I. Kontoyiannis. Asymptotic recurrence and waiting times for stationary processes. *Journal of Theoretical Probability*, 11:795–811, July 1998.

[12] Ioannis Kontoyiannis and Junshan Zhang. Arbitrary source models and Bayesian codebooks in rate-distortion theory. *IEEE Transactions on Information Theory*, 48(8):2276–2290, August 2002.

[13] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[14] En-hui Yang and Zhen Zhang. On the redundancy of lossy source coding with abstract alphabets. *IEEE Transactions on Information Theory*, 45(4):1092–1110, May 1999.