More On Recurrence and Waiting Times

by

Abraham J. Wyner *

Department of Statistics

Wharton School

University of Pennsylvania

Philadelphia, PA

USA

Abstract

Let $\mathbf{X} = \{X_n : \ n = 1, 2, \ldots\}$ be a discrete valued stationary ergodic process distributed according to probability P. Let $\mathbf{Z}_1^n = \{Z_1, Z_2, \ldots, Z_n\}$ be an independent realization of an n-block drawn with the same probability as $\mathbf{X}$. We consider the *waiting time* W_n defined as the first time the n-block $\mathbf{Z}_1^n$ appears in $\mathbf{X}$. There are many recent results concerning this waiting time that demonstrate asymptotic properties of this random variable. In this paper, we prove that for all n the random variable $W_n P(Z_1^n)$ is approximately distributed as an exponential random variable with mean 1. We use a Poisson heuristic to provide a very simple intuition for this result, which is then formalized using the Chen-Stein method. We then rederive, with remarkable brevity, most of the known asymptotic results concerning W_n and prove others as well. We further establish the surprising fact that for many sources $W_n P(\mathbf{Z}_1^n)$ is $\exp(1)$ even if the probability law for $\mathbf{Z}$ is not the same as that of $\mathbf{X}$. We also consider the d-dimensional analog of the waiting time and prove a similar result in that setting. Nearly identical results are then derived for the *recurrence time* R_n defined as the first time the initial n-block $\mathbf{X}_1^n$ *reappears* in $\mathbf{X}$.

We conclude by developing applications of these results to provide concise solutions to problems that stem from the analysis of the Lempel-Ziv data compression algorithm. We also consider possible applications to DNA sequence analysis.

*Research supported by the National Science Foundation grant number DMS-9508933.

KEYWORDS: Recurrence times, entropy, pattern matching, Chen-Stein method, data compression, Lempel-Ziv algorithm. **AMS subject classifications** 60-G07.

I. Introduction

Let $\mathbf{X} = \{X_1, X_2, \ldots\}$ be a finite alphabet stationary ergodic process on the space of infinite sequences endowed with probability P. Let $X_i \in \mathcal{A}$ and denote any finite contiguous substring of $\mathbf{X}$ with the notation $\mathbf{X}_i^j = \{X_i, \ldots, X_j\}$. Let the sequence $\mathbf{Z}_1^n$ be an independent realization of the first n symbols of the process. Thus, for any n-block sequence of symbols $\mathbf{z}_1^n$ with $z_i \in \mathcal{A}$, we define

$$P(\mathbf{z}_1^n) = \Pr\{\mathbf{Z}_1^n = \mathbf{z}_1^n\}.$$

Since $\mathbf{X}$ and $\mathbf{Z}$ have the same distribution the random variable $P(\mathbf{X}_1^n)$ and $P(\mathbf{Z}_1^n)$ are also identically distributed as well as independent. Our main interests are the *waiting time* W_n, defined to be the first time the sequence $\mathbf{Z}_1^n$ appears in $\mathbf{X}$, and the *recurrence time* R_n, defined to be the first time the sequence $\mathbf{X}_1^n$ *reappears* in $\mathbf{X}$:

$$W_n = \inf\{k \geq 1 : \ \mathbf{X}_k^{k+n-1} = \mathbf{Z}_1^n\} \qquad R_n = \inf\{k \geq 1 : \ \mathbf{X}_{k+1}^{k+n} = \mathbf{X}_1^n\}.$$

These random variables have been the subject of a growing body of work that crosses over probability theory to computer science and information theory. Originally, A.D. Wyner and J. Ziv, motivated by interest in the Ziv-Lempel data compression algorithm, discovered (in [10]) that for all stationary, finite-alphabet, ergodic processes:

$$\lim_{n\to\infty} \frac{\log R_n}{n} = H \text{ in probability,} \tag{1}$$

where H is the process entropy. They found the same result to be true of W_n, if the process was also assumed to be Markov of arbitrary order. They conjectured that the same limit held almost surely (a fact that was later established in [4]). Subsequently, A.J. Wyner discovered in [11] that for Markov sources, W_n (and R_n) satisfied a second order limit law:

$$\lim_{n\to\infty} \frac{\log W_n - nH}{\sigma\sqrt{n}} \Rightarrow N(0,1) \tag{2}$$

where $\sigma^2 = \lim_{n\to\infty} \frac{Var(-\log P(\mathbf{X}_1^n))}{n}$. Interesting work on the "Wyner-Ziv" problem was presented by Shields who discovered (in [6]) that the limit theorem of (1) was not true of all stationary, ergodic processes especially if the process memory vanished sufficiently slowly. In 1994, Ornstein and Weiss showed that W_n and R_n had simple analogs for stationary, ergodic, d-dimensional random fields and that a limit law similar to (1) also held in that setting.

In this paper, we consider the quantities $W_n P(\mathbf{Z}_1^n)$ and $R_n P(\mathbf{X}_1^n)$. We establish that the respective products each are approximately distributed exponentially with mean 1 (for all n) providing an explicit error term. Since $P(\mathbf{Z}_1^n)$ is a very familiar object (since it is easily transformed into a random walk by taking logarithms), this result transforms W_n and R_n into equally familiar objects. We then use this non-asymptotic result to prove several new asymptotic results and reprise some older ones.

The accuracy of the approximation is contingent upon two quantities that characterize the length of the process memory. For $-\infty \leq i \leq j \leq \infty$ let $\mathcal{B}_i^j$ denote the σ-field generated by $\mathbf{X}_i^j$. Define for $k \geq 1$:

$$\psi(k) = \sup_{A \in \mathcal{B}_{-\infty}^0, B \in \mathcal{B}_k^\infty} \frac{|P(A \cap B) - P(B)P(A)|}{P(B)P(A)}$$

where $0/0$ is defined to be 0. We say $\mathbf{X}$ is *ψ-mixing* if $\psi(k) \to 0$ as $k \to \infty$. Furthermore, we define the quantity

$$\gamma(k) = \max_{x_{-k},\ldots,x_{2k}} [P(\mathbf{x}_0^{k-1}),\ P(\mathbf{x}_0^{k-1}|\mathbf{x}_k^{2k}),\ P(\mathbf{x}_0^{k-1}|\mathbf{x}_{-k}^{-1})].$$

Since $|\mathcal{A}|$ is finite there there must exist a k-block whose probability or conditional probability is maximal. Therefore, $\gamma(k)$ always exists.

The relationship between waiting times and probability functions has been pursued before. It is well known (to information theorists who study data compression) that $-\log P(\mathbf{X}_1^n)$ represents (in some sense) an ideal codeword length. Furthermore, a Lempel-Ziv codeword is very closely related to a quantity L_m * , defined to be the largest k such that a copy of $\mathbf{Z}_1^k$ can be found as a contiguous substring of $\mathbf{X}_1^m$:

$$L_m = \max\{k \geq 1 : \mathbf{Z}_1^k = \mathbf{X}_j^{j+k-1} \text{ for some } j \in [1, m-k+1]\}. \tag{3}$$

The relationship of W_n and L_m is through the observation that

$$\{L_m < n\} = \{W_n > m\}.$$

Thus W_n and L_m are "duals" of one another. In [12], it was discovered that if

$$T_m = \inf\{k : -\log P(\mathbf{Z}_1^k) > \log m\}, \tag{4}$$

*The quantity L_m is exactly the length of the first phrase in the Fixed-Database version of the LZ(77) algorithm; see [12].

then $-\log P(\mathbf{Z}_1^{T_m})$ and L_m are asymptotically equivalent. This was only proved if the process was assumed to be Markov. Finally, no history of the waiting time problem, however brief, would be complete without a reference to the ground breaking work of Szpankowski (see [9]) who has considered L_m and related match lengths in quite literally dozens of papers with almost as many collaborators.

The first explicit connection between $\log W_n$ and $-\log P(\mathbf{Z}_1^n)$ (as well as $\log R_n$ and $\log P(\mathbf{X}_1^n)$) was made by Kontoyiannis in [2]. He showed, with methods similar to those found in [10], that $-\log P(\mathbf{Z}_1^n)$ provides a strong asymptotic approximation to $\log W_n$ for mixing processes and that $\log R_n$ is approximated by $-\log P(\mathbf{X}_1^n)$ for all stationary ergodic processes.

II. Waiting Times

THEOREM A: *Let* $\mathbf{X}$ *be a stationary, finite valued, ergodic process. For* $t > 0$*, we have*

$$|\Pr\{W_n P(\mathbf{Z}_1^n) \geq t\} - \exp(-t)| \leq (1 \wedge t)[8n\gamma(n) + \psi(n)]$$

Thus, for processes with memory that vanishes sufficiently fast, the quantity $W_n P(Z_1^n)$ *has an exponential distribution with mean* 1.

Corollary A1: *If* $\mathbf{X}$ *is an ergodic Markov chain then there exists a positive constant* $\beta < 1$ *such that*

$$|\Pr\{W_n P(\mathbf{Z}_1^n) > t\} - \exp(-t)| \leq (1 \wedge t)9n\beta^n.$$

Corollary A2: *If* $n\gamma(n) \to 0$ *and* $\psi(n) \to 0$ *then*

$$\Pr\{\log[W_n P(\mathbf{Z}_1^n)] > \log\log n\} \to 0$$

Corollary A3: *If* $\psi(n) = o(n)$ *then for any* $\epsilon > 0$*:*

$$\Pr\{\log[W_n P(\mathbf{Z}_1^n)] < -(1+\epsilon)\log n\} \to 0$$

Corollary A4: *Suppose* $\mathbf{X}$ *is i.i.d. uniform over* $\mathcal{A}$ *with* $s = |\mathcal{A}|$*. Then for any* $a \in \mathbb{R}$ *and all integers* $n > 0$*:*

$$|\Pr\{\log_s W_n > a + n\} - \exp(-s^a)| \leq (1 \wedge s^a)8ns^{-n}.$$

Corollary A5: *If $\mathbf{X}$ is an ergodic Markov Chain then for any $\epsilon > 0$:*

$$-(1+\epsilon)\log n \leq \log[W_n P(\mathbf{Z}_1^n)] \leq \log\log n \ \text{ eventually, almost surely.}$$

Proofs of Corollaries A1-A5: It is easily shown that for Markov chains there exists $p < 1$ such that for all sequences $\mathbf{x}_{-n}^{2n}$: $P(\mathbf{x}_0^{n-1}) \leq p^n$, $P(\mathbf{x}_0^{n-1}|\mathbf{x}_{-n}^{-1}) \leq p^n$ and $P(\mathbf{x}_0^{n-1}|\mathbf{x}_n^{2n}) \leq p^n$. Typically p might be the largest transition probability of the forward or backward chain if less than 1. If not, it is possible, without loss of generality, to extend the alphabet so that the largest transition is less than 1. This implies that $\gamma(n) \leq p^n$. Now recall that for Markov chains there exists a positive constant $q < 1$ such that $\psi(n) \leq q^n$. Corollary A1 follows from Theorem A with $\beta = p \vee q$. Corollary A2, follows from Theorem A with $t = \log n$. To prove Corollary A3, observe that with $t = n^{-(1+\epsilon)}$, we have that

$$\begin{aligned}
\Pr\{W_n P(\mathbf{Z}_1^n) < n^{-(1+\epsilon)}\} &\leq 1 - \exp(-n^{-(1+\epsilon)}) + (1 \wedge n^{-(1+\epsilon)})[8n\gamma(n) + \psi(n)] \\
\Pr\{\log W_n P(\mathbf{Z}_1^n)) < -(1+\epsilon)\log n\} &\leq n^{-(1+\epsilon)} + 8\gamma(n)n^{-\epsilon} + \psi(n)n^{-(1+\epsilon)} \\
&= o(1).
\end{aligned}$$

Lastly, Corollary A4 is proved by noting that for $\mathbf{X}$ i.i.d. uniform, $P(\mathbf{X}_1^n) = s^{-n}$. This implies that $\gamma(n) = s^{-n}$ and it is obvious that $\psi(n) = 0$. Thus, Corollary A4 will follow from Theorem A with $t = s^a$. We point out the following: a stronger version of Corollary A3 can be found in [2]; Corollary A2 is a stronger version in a less general setting than another theorem in [2]; and Corollaries A1 and A4 are essentially new versions of theorems that have appeared only in [11]. Corollary A5 is proved by observing that for $\mathbf{X}$ ergodic Markov, $\sum_{n=1}^{\infty}\gamma(n) < \infty$ and $\sum_{n=1}^{\infty}\psi(n) < \infty$, which implies, by Borel-Cantelli, that Corollaries A2 and A3 hold almost surely.

We will need some new notation to provide a framework for Poisson approximation. To that end, we fix $\mathbf{z}_1^n \in \mathcal{A}^n$. Let $I = \{1, 2, \ldots m = \lceil \frac{t}{P(\mathbf{z}_1^n)} \rceil\}$. Consider the doubly infinite extension of $\mathbf{X}$ to $\mathbf{X}_{-\infty}^{\infty}$ and define

$$Y_i = \begin{cases} 1 & \text{if } \mathbf{X}_i^{i+n-1} = \mathbf{z}_1^n \\ 0 & \text{otherwise.} \end{cases}$$

Thus $Y_i = 1$ if the n-block at position i in $\mathbf{X}$ is equal to $\mathbf{z}_1^n$. We count the number of occurrences of $\mathbf{z}_1^n$ over the index set I with the random variable

$$W(\mathbf{z}_1^n) = \sum_{i \in I} Y_i.$$

For each $i \in I$ we denote the $2n$ neighborhood of i by $\mathcal{B}_i = \{i-2n, i-2n+1, \ldots, i+2n\}$. We now define three expectations over $\mathbf{X}$:

$$
\begin{aligned}
B_1(\mathbf{z}_1^n) &= \sum_{i \in I} \sum_{k \in \mathcal{B}_i} EY_i EY_k \\
B_2(\mathbf{z}_1^n) &= \sum_{i \in I} \sum_{i \neq k \in \mathcal{B}_i} EY_i Y_k \\
B_3(\mathbf{z}_1^n) &= \sum_{i \in I} s_i \quad \text{where } s_i = |E\{Y_i - EY_i | \sigma(Y_k : k \in I - \mathcal{B}_i)\}| \ .
\end{aligned}
$$

Finally, we let

$$\lambda(\mathbf{z}_1^n) \triangleq EW(\mathbf{z}_1^n).$$

Heuristically, if $P(\mathbf{z}_1^n)$ is small and matches of $\mathbf{z}_1^n$ are not likely to appear in clumps (this will be true if $\mathbf{z}_1^n$ exhibits little self-symmetry) then the waiting time until an occurrence of $\mathbf{z}_1^n$ (call this conditional waiting time $W_n(\mathbf{z}_1^n)$) should be approximately geometric with mean $\lambda = \frac{1}{P(\mathbf{z}_1^n)}$. Thus if we close our eyes to the dependence one might believe that

$$\Pr\{W_n(\mathbf{z}_1^n) P(\mathbf{z}_1^n) > t\} = \Pr\{W_n(\mathbf{z}_1^n) > \lceil t\lambda \rceil\} \approx (1 - \frac{1}{\lambda})^{\lceil t\lambda \rceil} \approx \exp(-t).$$

This statement we can make rigorous using the Chen-Stein method (see Barbour et al. [1]) which shows that:

$$|\Pr\{W(\mathbf{z}_1^n) = 0\} - \exp(-\lambda(\mathbf{z}_1^n))| \leq (1 \wedge \lambda^{-1}(\mathbf{z}_1^n)) \sum_{i=1}^{3} B_i(\mathbf{z}_1^n). \tag{5}$$

By the stationarity of $\mathbf{X}$ and the linearity of expectations we have that

$$\lambda(\mathbf{z}_1^n) = mP(\mathbf{z}_1^n) = t.$$

Furthermore, it is easy to see that

$$\Pr\{W_n P(\mathbf{z}_1^n) > t | \mathbf{Z}_1^n = \mathbf{z}_1^n\} = \Pr\{W(\mathbf{z}_1^n) = 0\}.$$

Thus we have

$$|\Pr\{W_n P(\mathbf{z}_1^n) > t | \mathbf{Z}_1^n = \mathbf{z}_1^n\} - \exp(-t)| \leq (1 \wedge t^{-1}) \sum_{i=1}^{3} B_i(\mathbf{z}_1^n). \tag{6}$$

We now state our key technical Lemma, leaving the proof for the appendix.

LEMMA A: *Let $B_i = E_{\mathbf{Z}_1^n} B_i(\mathbf{Z}_1^n)$. Then, $B_1 \leq 4nt\gamma(n)$, $B_2 \leq 4nt\gamma(n)$, and $B_3 \leq t\psi(2n)$.*

Proof of THEOREM A: Taking expectations with respect to $\mathbf{Z}_1^n$ in (6),

$$\Pr\{W_n P(\mathbf{Z}_1^n) > t\} \leq \exp(-t) + (1 \wedge t^{-1}) \sum_{i=1}^{3} E_{\mathbf{Z}_1^n} B_i(\mathbf{Z}_1^n).$$

Applying Lemma A, we have

$$\Pr\{W_n P(\mathbf{Z}_1^n) > t\} \leq \exp(-t) + (1 \wedge t)[8n\gamma(n) + \psi(2n)].$$

In the same way, we can prove reverse inequalities and thus,

$$|\Pr\{W_n P(\mathbf{Z}_1^n) > t\} - \exp(-t)| \leq (1 \wedge t)[8n\gamma(n) + \psi(2n)],$$

which proves Theorem A. •

III. Recurrence Times

THEOREM B: *Let* $\mathbf{X}$ *be a stationary, finite valued, ergodic process. Let* $t > 0$. *Then*

$$|P(R_n P(\mathbf{Z}_1^n) \geq t) - \exp(-t)| \leq (1 \wedge t)[8n\gamma(n)[1 + \psi(n)] + \psi(n)] + 8n\gamma(n)$$

Corollary B1: *If* $\mathbf{X}$ *is an ergodic Markov chain then there exists a positive constant* $\beta < 1$ *such that*

$$|\Pr\{R_n P(\mathbf{X}_1^n) > t\} - \exp(-t)| \leq (1 \wedge t) 16 n \beta^n.$$

Corollary B2: *If* $n\gamma(n) \to 0$ *and* $\psi(n) \to 0$ *then*

$$\Pr\{\log[R_n P(\mathbf{Z}_1^n)] > \log\log n\} \to 0$$

Corollary B3: *If* $\psi(n) = o(n)$ *and* $n\gamma(n) = o(1)$ *then for any* $\epsilon > 0$*:*

$$\Pr\{\log[R_n P(\mathbf{X}_1^n)] < -(1+\epsilon)\log n\} \to 0$$

Corollary B4: *Suppose* $\mathbf{X}$ *is i.i.d. uniform over* $\mathcal{A}$ *with* $s = |\mathcal{A}|$. *Then for any* a *and all integers* $n > 0$*:*

$$|\Pr\{\log_s R_n > a + n\} - \exp(-s^a)| \leq (1 \wedge s^a) 16 n s^{-n}.$$

Corollary B5: *If* $\mathbf{X}$ *is an ergodic Markov Chain then for any* $\epsilon > 0$*:*

$$-(1+\epsilon)\log n \leq \log[R_n P(\mathbf{Z}_1^n)] \leq \log\log n \ \text{ eventually, almost surely.}$$

Proof: We only need to make a simple modification in the definitions. First we fix $\mathbf{x}_1^n \in \mathcal{A}^n$ and let $I = \{4n+1, 4n+2, \dots m = \lceil \frac{t}{P(\mathbf{x}_1^n)} \rceil\}$. For $i \in I$, define

$$Y_i = \begin{cases} 1 & \text{if } \mathbf{X}_{i+1}^{i+n} = \mathbf{x}_1^n \\ 0 & \text{otherwise.} \end{cases}$$

The definitions of $\mathcal{B}_i$, $B_1(\mathbf{x}_1^n)$, $B_2(\mathbf{x}_1^n)$, $B_3(\mathbf{x}_1^n)$, and $W(\mathbf{x}_1^n)$ are the same as in Section 2, replacing every occurrence of $\mathbf{z}_1^n$ with $\mathbf{x}_1^n$.

In the recurrence time setting, there are two kinds of occurrences of the initial patterns to consider: short range occurrences, where $R_n < 4n$, and long range occurrences where $R_n > 4n$. We argue that for processes with vanishing memory, it is very likely that $R_n > 4n$ (short range matches occur with small probability). We formalize this by proving (in the appendix):

LEMMA B: $\Pr\{R_n < 4n\} \le 4n\gamma(n)$.

Observe that in the recurrence setting the quantities EY_i and EY_iY_k are conditional events, conditioned on the outcome of the initial n-block. It is easy to see that $B_1(\mathbf{x}_1^n) \le B_1(\mathbf{z}_1^n)[1+\psi(n)]$ follows from the ψ-mixing property (since we are conditioning on the initial $\mathbf{X}_1^n$ block). Similarly $B_2(\mathbf{x}_1^n) \le B_2(\mathbf{z}_1^n)[1+\psi(n)]$. Finally, it follows that

$$\Pr\{W(\mathbf{x}_1^n) = 0\} \le \Pr\{R_nP(\mathbf{x}_1^n) > t | \mathbf{X}_1^n = \mathbf{x}_1^n\} \le \Pr\{W(\mathbf{x}_1^n) = 0\} + \Pr\{R_n < 4n | \mathbf{X}_1^n = \mathbf{x}_1^n\}.$$

To apply the Chen-Stein equation to approximate $\Pr\{W(\mathbf{x}_1^n) = 0\}$ we need to find $\lambda(\mathbf{x}_1^n) = EW(\mathbf{x}_1^n)$. This is no more difficult than in the waiting time case, but slightly more complicated:

$$\begin{aligned} \lambda(\mathbf{x}_1^n) &= \sum_{i \in I} EY_i \\ &= [\frac{t}{P(\mathbf{x}_1^n)} - 4n]P(\mathbf{x}_1^n) \\ &= t - 4nP(\mathbf{x}_1^n) \end{aligned}$$

Since, for every $\mathbf{x}_1^n \in \mathcal{A}^n$, $P(\mathbf{x}_1^n) \le \gamma(n)$ it must be that

$$t - 4n\gamma(n) \le \lambda(\mathbf{x}_1^n) \le t.$$

Thus, it follows from (5) (and the fact that $e^{4n\gamma(n)} = 1 + 4n\gamma(n) + o(4n\gamma(n))$) that

$$
\begin{aligned}
|\Pr\{R_n P(\mathbf{X}_1^n) > t | \mathbf{X}_1^n = \mathbf{x}_1^n\} - \exp(-t)| &\le (1 \wedge \lambda(\mathbf{x}_1^n)^{-1}))[\sum_{i=1}^{3} B_i(\mathbf{x}_1^n)] \\
&\quad + \Pr\{R_n < 4n | \mathbf{X}_1^n = \mathbf{x}_1^n\} + 4n\gamma(n).
\end{aligned}
$$

Following the steps of the proof of Theorem A and using Lemma A and Lemma B proves Theorem B. •

IV. Generalization to Random Fields

A remarkable feature of the Poisson approximation approach is the ability to find extensions, with only minor modification of the proof used in the one-dimensional case, to random fields on the integer lattice $\mathbf{Z}^d$. We consider a family of random variables $\{X_{\mathbf{u}} : \mathbf{u} \in \mathbf{Z}^d\}$ indexed by d-dimensional vectors $\mathbf{u} = (u_1, \ldots, u_d)$. We assume that for all vectors $\mathbf{u}$ the random variable $X_{\mathbf{u}}$ takes values in a finite set $\mathcal{A}$. The distribution on the process is assumed to be a stationary and ergodic probability P on the product space $\mathbf{\Pi} = \prod\{\mathcal{A}_{\mathbf{u}} : \mathbf{u} \in \mathbf{Z}^d\}$ (with each $\mathcal{A}_{\mathbf{u}}$ a copy of $\mathcal{A}$). Suppose $\mathbf{x}$ is a realization of the process (and thus $\mathbf{x} \in \mathbf{\Pi}$) then $X_{\mathbf{u}} = x_{\mathbf{u}}$ is the coordinate at position $\mathbf{u}$. For any subset $U \subseteq \mathbf{Z}^d$ we let $\mathbf{X}_U(\mathbf{x}) = \{x_{\mathbf{u}}\}_{\mathbf{u} \in U}$. For any vector $\mathbf{v} \in \mathbf{Z}^d$, let $T_{\mathbf{v}}\mathbf{x}$ denote the realization with coordinates $\mathbf{X}_{\mathbf{u}}(T_{\mathbf{v}}\mathbf{x}) = \mathbf{X}_{\mathbf{u}+\mathbf{v}}(\mathbf{x}) = \mathbf{x}_{\mathbf{u}+\mathbf{v}}$. Thus, for any subset $U \in \mathbf{Z}^d$, we say that the d-dimensional pattern $\mathbf{X}_U$ occurs at position $\mathbf{v}$ if $\mathbf{X}_U = \mathbf{X}_{\mathbf{v}+U}$.

The d-dimensional analog of a sequence of length n in one dimension is the d-dimensional cube. For $\mathbf{u}, \mathbf{w} \in \mathbf{Z}^d$ let $[\mathbf{u}, \mathbf{w}) = \{\mathbf{v} \in \mathbf{Z}^d : u_j \le v_j < w_j \text{ for all } j\}$. We can define the d-dimensional cube with side $k > 0$, to be the cartesian product of $[0, k)^d$:

$$C(k) = \{\mathbf{u} \in \mathbf{Z}^d : 0 \le u_j < k \text{ for all } j\}.$$

We define the recurrence time R_n as the smallest value of k such that the initial n-cube pattern $\mathbf{X}_{C(n)}$ occurs at some coordinate inside $C(k)$ other than at coordinate $\mathbf{0}$:

$$R_n = \inf\{k \ge 1 : \mathbf{X}_{C(n)} = \mathbf{X}_{\mathbf{u}+C(n)} \text{ for some } \mathbf{u} \in C(k), \mathbf{u} \ne \mathbf{0}\}.$$

The waiting time analog of the recurrence time R_n, considers an independent process $\{\mathbf{Z_u}\}$ also distributed with probability measure P. We define the waiting time W_n as the smallest value of k such that the n-cube pattern $\mathbf{Z}_{C(n)}$ occurs in the process $\mathbf{X}$ at some coordinate inside $C(k)$:

$$W_n = \inf\{k \geq 1 : \mathbf{Z}_{C(n)} = \mathbf{X}_{\mathbf{u}+C(n)} \text{ for some } \mathbf{u} \in C(k)\}.$$

Before we state any theorems we need to introduce d-dimensional analogs of $\gamma(n)$ and $\psi(n)$. To that end, for any coordinate $\mathbf{u} \in I$, define the $2n$-neighborhood of by $\mathcal{B}_\mathbf{u} = \{\mathbf{v} : |u_j - v_j| \leq 2n\}$. Thus, for any $\mathbf{v} \in C(n) + \mathbf{u}$ and any $\mathbf{w} \in \{Z^d - \mathcal{B}_\mathbf{u}\}$ it must be that $|v_j - w_j| \geq n$ for all j. This means that every coordinate outside of $\mathcal{B}_\mathbf{u}$ is separated by at least n (in all directions) from every coordinate inside the n-cube at $\mathbf{u}$. We now are able to define

$$\gamma_0(n) = \max_{\mathbf{x_u}:\mathbf{u}\in\mathcal{B}_\mathbf{0}} P(\mathbf{x}_{C(n)}) \qquad \gamma_1(n) = \max_{\mathbf{x_u}:\mathbf{u}\in\mathcal{B}_\mathbf{0}} P(\mathbf{x}_{C(n)}|\mathbf{x_v} : \mathbf{v} \in \mathcal{B}_\mathbf{0} - \mathbf{C(n)}).$$

Thus $\gamma_0(n)$ is the probability of the n-cube with the highest probability with respect to P and $\gamma_1(n)$ is the conditional probability of the n-cube with the highest conditional probability given all its neighbors that are within a distance n of any coordinate. Since, we are assuming that $\mathcal{A}$ is finite alphabet, the n-cube with the largest probability or conditional probability must exist and therefore $\gamma(n) = \gamma_0(n) \vee \gamma_1(n)$ is well defined. Now let $\tau(n)$ denote the σ-field generated by $\{\mathbf{X_u} : \mathbf{u} \in C(n)\}$ and let $\tau(-n)$ denote the σ-field generated by $\{\mathbf{X_u} : \mathbf{u} \in Z^d - \mathcal{B}_\mathbf{0}\}$. Let

$$\psi(n) = \sup_{A\in\tau(n), B\in\tau(-n)} \frac{|P(A \cap B) - P(A)P(B)|}{P(A)P(B)}$$

THEOREM C: *Let* $\mathbf{X}$ *be a stationary, ergodic, d-dimensional random field (defined above). Let* $t > 0$*:*

$$|\Pr\{W_n P(\mathbf{Z}_{C(n)})^{\frac{1}{d}} \geq t\} - \exp(-t)| \leq (1 \wedge t)[(4n)^d \gamma(n) + \psi(n)].$$

Thus, for many processes with vanishing memory the quantity $W_n P(\mathbf{Z}_{C(n)})^{\frac{1}{d}}$ *has an exponential distribution with mean 1.*

Corollary C1: *If* $\mathbf{X}$ *is an i.i.d. random field then*

$$\frac{d \log W_n}{n^{\frac{d}{2}}} \Rightarrow N(H, \sigma^2)$$

where

$$H = \lim_{n\to\infty} \frac{E - \log P(\mathbf{Z}_{c(n)})}{n^d} \quad \textit{and} \quad \sigma^2 = \lim_{n\to\infty} \frac{Var(-\log P(\mathbf{Z}_{c(n)}))}{n^d}.$$

Proof of Corollary C1: It follows from Theorem C that $n^{-\frac{d}{2}}d\log W_n$ and $-n^{-\frac{d}{2}}\log P(\mathbf{Z}_{C(n)})$ are almost surely equal in the limit as $n \to \infty$. By the independence assumption

$$-\log P(\mathbf{Z}_{C(n)}) = \sum_{\mathbf{u}\in C(n)} \log P(\mathbf{Z}_{\mathbf{u}}),$$

Since there are exactly n^d vectors in $C(n)$ the Corollary follows from the Central Limit Theorem. We remark that Corollary C1 will also be true if $\mathbf{Z}$ is a Markov random field.

Proof of Theorem C: We use the Chen-Stein method in the d-dimensional setting. Given $\mathbf{Z}_{C(n)} = \mathbf{z}_{C(n)}$ we let $m = \lceil (\frac{t}{P(\mathbf{z}_{C(n)})})^{\frac{1}{d}} \rceil$ and $I = \{\mathbf{v} \in C(m)\}$. Then define

$$Y_{\mathbf{u}} = \begin{cases} 1 & \text{if } \mathbf{X}_{\mathbf{u}+C(n)} = \mathbf{z}_{C(n)} \\ 0 & \text{otherwise.} \end{cases}$$

Thus for any coordinate $\mathbf{u}$, $Y_{\mathbf{u}} = 1$ if the n-cube at $\mathbf{u}$ is equal to the n-cube $\mathbf{z}_{C(n)}$. The total number of matches of $\mathbf{z}_{C(n)}$ in $C(m)$ is given by

$$W(\mathbf{z}_{C(n)}) = \sum_{\mathbf{u}\in I} Y_{\mathbf{u}}.$$

The expected number of matches is easily derived from the stationarity and is given by

$$\begin{aligned} \lambda(\mathbf{z}_{C(n)}) &\triangleq EW(\mathbf{z}_{C(n)}) \\ &= m^d P(\mathbf{z}_{C(n)}) \\ &= t. \end{aligned}$$

The d-dimensional analogs B_1, B_2 and B_3 of section II are

$$\begin{aligned} B_1(\mathbf{z}_{C(n)}) &= \sum_{\mathbf{u}\in I}\sum_{\mathbf{v}\in\mathcal{B}_{\mathbf{u}}} EY_{\mathbf{u}}EY_{\mathbf{v}} \\ B_2(\mathbf{z}_{C(n)}) &= \sum_{\mathbf{u}\in I}\sum_{\mathbf{u}\neq\mathbf{v}\in\mathcal{B}_{\mathbf{u}}} EY_{\mathbf{u}}Y_{\mathbf{v}} \\ B_3(\mathbf{z}_{C(n)}) &= \sum_{\mathbf{u}\in I} s_{\mathbf{u}} \quad \text{where } s_{\mathbf{u}} = |E\{Y_{\mathbf{u}} - EY_{\mathbf{u}}|\sigma(Y_{\mathbf{v}} : \mathbf{v} \in I - \mathcal{B}_{\mathbf{u}})\}| \end{aligned}$$

In the appendix we prove the following:

LEMMA C: *The following bounds hold for all n:* $EB_1(\mathbf{Z}_{C(n)}) \leq (4n)^d t\gamma(n)$, *and* $EB_2(\mathbf{Z}_{C(n)}) \leq (4n)^d t\gamma(n)$, *and* $EB_3(\mathbf{Z}_{C(n)}) \leq t\psi(n)$.

From the Chen-Stein equation we have that

$$|\Pr\{W(\mathbf{z}_{C(n)}) = 0\} - \exp(-\lambda(\mathbf{z}_{C(n)}))| \leq (1 \wedge \lambda^{-1}(\mathbf{z}_{C(n)})) \sum_{i=1}^{3} B_i(\mathbf{z}_{C(n)}).$$

The theorem follows taking expectations and applying Lemma C $\bullet$.

V. Extensions

Suppose $\mathbf{Z}$ does not have the same distribution as $\mathbf{X}$. Let $\mathbf{Z}$ be distributed according to the stationary distribution Q and define

$$Q(\mathbf{z}_1^n) = \Pr\{\mathbf{Z}_1^n = \mathbf{z}_1^n\}.$$

We will need to define

$$\gamma_Q(n) = \max_{\mathbf{z}_1^n} Q(\mathbf{z}_1^n) \qquad \gamma_P(n) = \max_{\mathbf{x}_1^n} P(\mathbf{x}_1^n).$$

We will also need to bound the ratio of the conditional probability of an n-vector and the unconditional probability

$$\alpha(n) = \max_{\mathbf{x},\mathbf{y}\in\mathcal{A}^n} \frac{\Pr\{\mathbf{X}_0^{n-1} = \mathbf{x} | \mathbf{X}_n^{2n} = \mathbf{y}\}}{\Pr\{\mathbf{X}_0^{n-1} = \mathbf{x}\}},$$

Let

$$\alpha = \limsup_{n\to\infty} \alpha(n).$$

If we let

$$\rho(n) = \max_{x\in\mathcal{A}^{2n}} \sum_{i=1}^{n} [\log P(X_0 = x_0 | \mathbf{X}_{-n-i}^{-1} = \mathbf{x}_{-n-i}^{-1}) - \log P(X_0 = x_0 | \mathbf{X}_{-i}^{-1} = \mathbf{x}_{-i}^{-1})]$$

then $\alpha < \infty$ if $\rho(n) < \infty$ which is akin to a vanishing memory condition. Clearly, α is finite if $\mathbf{X}$ is Markov.

THEOREM D *(Mismatch Theorem): Let $\psi(n)$ be the ψ-mixing coefficents of $\mathbf{X}$. Then for any $t > 0$ and any n*

$$|\Pr\{W_n P(\mathbf{Z}_1^n) > t\} - \exp(-t)| \le (1 \wedge t)[4n\gamma_P(n) + 4n\alpha(n)\gamma_Q(n) + \psi(n)].$$

Corollary D1 *If P and Q are ergodic Markov chains then there exists a constant $\beta < 1$ such that*

$$|\Pr\{W_n P(\mathbf{Z}_1^n) > t\} - \exp(-t)| \le (1 \wedge t)[(5n + 4n\alpha)\beta^n.$$

Corollary D2 *If $\mathbf{Z}$ is Markov and $\mathbf{X}$ is ergodic, mixing and α is finite and $n\gamma_P(n) = o(1)$ then*

$$\lim_{n\to\infty} \frac{\log W_n}{n} = H(Q) + D(Q||P) \quad \text{in probability.}$$

Proofs of the Corollaries: If X is Markov then α is finite since $\sum_{n=1}^{\infty}\rho(n) < \infty$. Corollary D1 follows using the same steps used to prove Corollary A1. Corollary D2 follows by evaluating $\lim_{n\to\infty}\frac{\log W_n P(\mathbf{Z}_1^n)}{n}$: the Markovity of $\mathbf{Z}$ and the ergodic theorem implies that

$$\lim_{n\to\infty}\frac{-\log P(\mathbf{Z}_1^n)}{n} = -E\log P(\mathbf{Z}_1^n) \triangleq H(Q) + D(Q||P).$$

We remark, that if the conditions of Corollary D1 are satisfied then the convergence is almost sure by Borel Cantelli.

Proof of Theorem D: The proof rests on a modification of Lemma A, (proved in the Appendix)
LEMMA D: $B_2 = E_{\mathbf{Z}_1^n}B_2(\mathbf{Z}_1^n) \leq 4nt\alpha(n)\gamma_Q(n)$.
The proof of Theorem D follows the steps of the proof of Theorem A. The bounds on B_1 and B_3 of Lemma A apply to the mismatch case unchanged. If Lemma D is applied to bound B_2 then the theorem is proved.

To conclude, we remark that the mismatch case was first considered in [11] where it is remarked that the proof used to prove (2) may be carried over to the mismatch case with only slight changes. There are mismatch extensions of the main theorems in [2] as well. Furthermore, there are perhaps uncounted corollaries and extensions of Theorem A; we have remarked on only a few. It is, however, this author's belief that Theorem A provides a thorough settlement of the waiting time problem.

VI. Applications

It is clear that pattern matching theory, as developed in the preceding sections, has a natural connection to the Lempel-Ziv data compression algorithm (see section I). The relevant quantity is the match length L_m defined in (3). Since the algorithm's asymptotic performance is determined by this match length, if the distribution and moments of the match length can be found then performance properties can be derived. Our main theorem here provides a precise bound on the difference between the match length L_m and the random variable T_m, defined in (4). Before stating and proving that theorem, we establish an extension of Theorem A.

Let N be a positive, integer valued random variable, that is independent of $\mathbf{X}$ (but not necessarily independent of $\mathbf{Z}$). Consider the random variable $W_N P(\mathbf{Z}_1^N)$. The following theorem establishes criterion for when we can expect the distribution of this random variable to also be distributed $\exp(1)$.

THEOREM E: *Assume that there exist constants η_1 and η_2 so that $\eta_1 \leq N \leq \eta_2$. Then for $t > 0$.*

$$|\Pr\{W_N P(\mathbf{Z}_1^N) \geq t\} - \exp(-t)| \leq (1 \wedge t)[8\eta_2\gamma(\eta_1) + \psi(\eta_1)]$$

Proof of Theorem E: The proof of the theorem is follows closely the steps of the proof of Theorem A. Since we are assuming bounds on N, the proof is basically identical to the non-random case; we will only sketch the proof. A more general theorem, for unbounded N, surely holds, but for our applications we only need to establish the inequality of Theorem E for bounded N. The first step is to condition on $N = n$ and $\mathbf{Z}_1^n = \mathbf{z}_1^n$. Since N is independent of $\mathbf{X}$ (as is $\mathbf{Z}$), the inequality of (6) holds unchanged. We cannot apply Lemma A directly, however, because of the randomness of N. Instead we apply the following simple extension:

LEMMA E: *Let $B_i = E_{\mathbf{Z}_1^N} B_i(\mathbf{Z}_1^N)$. Then, $B_1 \leq 4\eta_2 t\gamma(\eta_1)$, $B_2 \leq 4\eta_2 t\gamma(\eta_1)$, and $B_3 \leq t\psi(2\eta_1)$.*

Applying Lemma E completes the proof of the Theorem.

The most useful application of this Theorem is with $N = T_m(\mathbf{Z})$. In this case, the quantity $P(\mathbf{Z}_1^N)$ is nearly $\frac{1}{m}$ so that $\Pr\{W_N > m\} \approx \exp(-1)$. It is then easy to prove the following extension of Theorem 3 of [12]:

THEOREM F: Let $\Delta = L_m - T_m(\mathbf{Z})$. If $\mathbf{Z}$ satisfies the conditions of Theorem E, then there exists a positive constant δ such that

$$\Pr\{\Delta \geq j\} \leq \alpha^j + O(\frac{\log m + j}{m^\delta})\alpha^j.$$

$$\Pr\{\Delta \leq -j\} \leq \exp(-\alpha^{-j}) + O(\frac{(\log m)^2}{m^\delta})\alpha^{-j}$$

A very useful application of Theorem F makes use of the elementary fact that $ET_m(\mathbf{Z}) = \frac{\log m}{H} + O(1)$ for $\mathbf{Z}$ both ergodic and Markov (see the appendix of [12] for a short proof). Since Theorem F implies that the tails of Δ are at least exponential we have

Corollary F: If $\mathbf{Z}$ is Markov and Ergodic then

$$E\Delta = O(1)$$

$$EL_m = \frac{\log m}{H} + O(1).$$

In conclusion, it follows easily from Corollary F and Theorem F that the Lempel-Ziv parsing algorithm is simply a partition of a sequence into equally probable segments of random length, with an approximately Normal distribution (since $T_m(\mathbf{Z})$ is Normal) and mean equal to $\log n/H$. From this fact it is easy to establish convergence rates (see [12], [11] and [10] for details).

Another application of Theorem F is in entropy estimation. Most approaches to entropy estimation proceed by first estimating a model and then forming a plug-in estimate by computing the entropy of the estimated model. From corollary F it follows that consistent estimates of the entropy can be constructed using match lengths. Algorithms that are based upon match lengths have found useful application in linguistics and DNA sequence analysis (see [13], [14] and [3]).

VIII Appendix

Proof of Lemma A: The proof of the technical lemma is straightforward; we need only evaluate B_1, B_2, and B_3. Recall that Y_i is the indicator of the event $\{\mathbf{X}_i^{i+n-1} = \mathbf{Z}_1^n\}$, the index set $I = \{1, 2, \ldots, m = \lceil \frac{t}{P(\mathbf{Z}_1^n)} \rceil\}$, and $\mathcal{B}_i = \{i - 2n, i - 2n + 1, \ldots, \imath + 2n\}$.

$$\begin{aligned} B_1 &= E_{\mathbf{Z}_1^n} \sum_{i \in I} \sum_{k \in \mathcal{B}_i} E_{\mathbf{X}}[Y_i | \mathbf{Z}_1^n] E_{\mathbf{X}}[Y_k | \mathbf{Z}_1^n] \\ &\stackrel{(a)}{=} E_{\mathbf{Z}_1^n} \frac{t}{P(\mathbf{Z}_1^n)} \sum_{k \in \mathcal{B}_0} E_{\mathbf{X}}[Y_i | \mathbf{Z}_1^n] E_{\mathbf{X}}[Y_k | \mathbf{Z}_1^n] \\ &\stackrel{(b)}{=} E_{\mathbf{Z}_1^n} \frac{t4n}{P(\mathbf{Z}_1^n)} P^2(\mathbf{Z}_1^n) \\ &\stackrel{(c)}{=} 4nt E_{\mathbf{Z}_1^n} P(\mathbf{Z}_1^n) \\ &\stackrel{(d)}{\leq} 4nt\gamma(n). \end{aligned}$$

Where (a) follows from the definition on B_1 using the stationarity of $\mathbf{X}$ and the fact that there are $t/P(\mathbf{Z}_1^n)$ summands. Step (b) follows since Y_k is the indicator of the event $\{\mathbf{X}_k^{k+n-1} = \mathbf{Z}_1^n\}$ whose expectation with respect to $\mathbf{X}$ is $P(\mathbf{Z}_1^n)$. Step (d) follows from the definition of $\gamma(n)$. To bound B_2 start with the definition:

$$
\begin{aligned}
B_2 &= E_{\mathbf{Z}_1^n} \sum_{i\in I} \sum_{k\in\mathcal{B}_i} E_{\mathbf{X}}[Y_i Y_k|\mathbf{Z}_1^n] \\
&\stackrel{(a)}{=} E_{\mathbf{Z}_1^n} \frac{t}{P(\mathbf{Z}_1^n)} \sum_{0\neq k\in\mathcal{B}_0} E_{\mathbf{X}}[Y_0 Y_k|\mathbf{Z}_1^n] \\
&\stackrel{(b)}{=} E_{\mathbf{Z}_1^n} \frac{t}{P(\mathbf{Z}_1^n)} \sum_{k=-2n,k\neq 0}^{2n} \Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{Z}_1^n|\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\} \Pr\{\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\} \\
&\stackrel{(c)}{=} \sum_{0\neq k=-2n}^{2n} t E_{\mathbf{Z}_1^n} E_{\mathbf{X}} 1\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}|\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\} \\
&\stackrel{(d)}{=} \sum_{0\neq k=-2n}^{2n} t \Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}\}.
\end{aligned}
$$

Step (a) follows from the stationarity and the definition of B_2 and the fact that there are $\frac{t}{P(\mathbf{Z}_1^n)}$ summands. Step (b) follows by conditioning on the event $Y_0 = 1$ which is identical to the event $\{\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\}$. Step (c) follows by canceling the $P(\mathbf{Z}_1^n)$ term in the denominator with the $\Pr\{\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\}$ term. Step (d) follows from the definitions of conditional probability and the fact that $\mathbf{Z}_1^n$ has the same distribution as $\mathbf{X}_0^{n-1}$. To complete the bound on B_2 we must evaluate $\Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}\}$ for integers k with $0 < |k| \leq 2n$. First consider $0 < k \leq n$.

$$
\Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}\} = E_{\mathbf{X}_{n+1}^{2n}} E_{\mathbf{X}_0^{n-1}}[1\{\mathbf{X}_0^{n-1} = \mathbf{X}_k^{k+n-1}\}|\mathbf{X}_{n+1}^{2n}].
$$

We will show that for every outcome of the "future" sequence x_{1+n}^{2n} there exists a unique sequence $U_k(x_n^{2n})$ such that

$$
\Pr\{\mathbf{X}_0^{n-1} = \mathbf{X}_k^{k+n-1}|\mathbf{X}_n^{2n} = x_{n+1}^{2n}\} = \Pr\{\mathbf{X}_0^{n-1} = U_i(x_n^{2n})|\mathbf{X}_n^{2n} = x_n^{2n}\}.
$$

In other words, if we condition first on the the outcome of the second n block from time n to $2n$, label it x, there exists only a single string $U_k(x)$ which will admit a match at position k. We prove by construction: write $U_k(x) = \{u_1, \ldots, u_n\}$. Clearly, the last k positions of $U_k(x)$ must be the first k positions of x: $u_{n-j+1} = x_{k+1-j}$ for $j = 1, 2, \ldots k$. The rest of the sequence is defined recursively so that the matching condition $\mathbf{X}_0^{n-1} = \mathbf{X}_k^{k+n-1}$ is satisfied, that is we let $u_{n-j+1} = u_{n-j+k+1}$ for $j = k, \ldots, n-1$. Therefore,

$$
\Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}\} \quad = \quad E[1\{\mathbf{X}_0^{n-1} = \mathbf{X}_k^{k+n-1}\}|\mathbf{X}_n^{2n}]
$$

$$
\begin{aligned}
&= E[1\{\mathbf{X}_1^n = U_i(\mathbf{X}_k^{k+n-1})\}|\mathbf{X}_n^{2n}] \\
&\leq \gamma(n).
\end{aligned}
$$

The same proof applies for $0 < -k \leq n$ and for $k > n$ we have by the definition of $\gamma(n)$ that $\Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}\} \leq \gamma(n)$. We now finish the bound on B_2 by observing that

$$B_2 \leq \sum_{0 \neq k = -2n}^{2n} \Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1}\} \leq 4n\gamma(n).$$

The final task is to evaluate B_3. Recall that $s_i = |E\{Y_i - EY_i|\sigma(Y_k : k \in I - \mathcal{B}_i)\}|$ and that

$$B_3 = E_{\mathbf{Z}_1^n} \sum_{i \in I} [s_i|\mathbf{Z}_1^n].$$

Given any set $A \in \sigma(X_\alpha : \alpha \notin \mathcal{B}_0)$ with $P(Z) > 0$ it follows from the ψ-mixing property of $\mathbf{X}$:

$$P(\{Y_0 = 1\} \cap A) - P(A)P(Y_0 = 1) \leq \psi(2n)P(A)P(Y_0 = 1).$$

Thus

$$E[Y_0|A] \leq [1 + \psi(2n)]P(\mathbf{Z}_1^n).$$

This implies that for any $i \in I$,

$$s_i \leq P(\mathbf{Z}_1^n)[1 + \psi(2n)] - P(\mathbf{Z}_1^n).$$

Hence,

$$
\begin{aligned}
B_3 &= E_{\mathbf{Z}_1^n} \sum_{i \in I} [s_i|\mathbf{Z}_1^n] \\
&\leq E_{\mathbf{Z}_1^n} \frac{t}{P(\mathbf{Z}_1^n)} [P(\mathbf{Z}_1^n)[1 + \psi(2n)] - P(\mathbf{Z}_1^n)] \\
&= \psi(2n)t
\end{aligned}
$$

Proof of Lemma B: Using the same procedure used to bound B_2 it follows that

$$
\begin{aligned}
\Pr\{R_n < 4n\} &\leq \sum_{k=1}^{4n-1} \Pr\{\mathbf{X}_{k+1}^{k+n} = \mathbf{X}_1^n\} \\
&\leq 4n\gamma(n).
\end{aligned}
$$

Proof of Lemma C: This is the d-dimensional version of Lemma A and the proof carries over with only changes in the index set. We leave the details to the reader.

Proof of Lemma D: We start with the definition of B_2:

$$
\begin{aligned}
B_2 &= E_{\mathbf{Z}_1^n} \sum_{i \in I} \sum_{k \in \mathcal{B}_i} E_{\mathbf{X}}[Y_i Y_k | \mathbf{Z}_1^n] \\
&= E_{\mathbf{Z}_1^n} \frac{t}{P(\mathbf{Z}_1^n)} \sum_{0 \neq k \in \mathcal{B}_0} E_{\mathbf{X}}[Y_0 Y_k | \mathbf{Z}_1^n] \\
&= E_{\mathbf{Z}_1^n} \frac{t}{P(\mathbf{z}_1^n)} \sum_{k=-2n, k\neq 0}^{2n} \Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{Z}_1^n | \mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\} \Pr\{\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\} \\
&= \sum_{0 \neq k = -2n}^{2n} t E_{\mathbf{Z}_1^n} E_{\mathbf{X}} 1\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1} | \mathbf{X}_0^{n-1} = \mathbf{Z}_1^n\} \\
&= \sum_{0 \neq k = -2n}^{2n} t \Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1} | \mathbf{Z}_1^n = \mathbf{X}_0^{n-1}\}
\end{aligned}
$$

As in the proof of Lemma A we will need a bound, for all $0 \leq |k| \leq 2n$, on $\Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1} | \mathbf{Z}_1^n = \mathbf{X}_0^{n-1}\}$. Define a sequence of random variables $\hat{\mathbf{X}}_i$ from $i = n, \ldots, 2n$ so that the unconditional distribution of $\hat{\mathbf{X}}_i$ is the conditional distribution of X_i given that $\mathbf{X}_0^{n-1} = \mathbf{Z}_1^n$. As before (in the proof of Lemma A), if $\mathbf{X}_n^{2n} = \mathbf{x}_n^{2n}$ then there exists a unique sequence $U_k(\mathbf{x}_n^{2n})$ that will admit a match of $\mathbf{X}_0^{n-1}$ at position k. Thus

$$
\begin{aligned}
\Pr\{\mathbf{X}_k^{k+n-1} = \mathbf{X}_0^{n-1} | \mathbf{Z}_1^n = \mathbf{X}_0^{n-1}\} &= E_{\mathbf{Z}_1^n} E_{\hat{\mathbf{X}}_n^{2n}} 1\{\mathbf{Z}_1^n = U_k(\hat{\mathbf{X}}_n^{2n})\} \\
&= E_{\hat{\mathbf{X}}_n^{2n}} E_{\mathbf{Z}_1^n} 1\{\mathbf{Z}_1^n = U_k(\hat{\mathbf{X}}_n^{2n})\} \\
&= \sum_{x \in \mathcal{A}^n} \Pr\{\mathbf{Z}_1^n = U_k(x) | \hat{\mathbf{X}}_n^{2n} = x\} \Pr\{\hat{\mathbf{X}}_n^{2n} = x\} \\
&\leq \max_{x,y \in \mathcal{A}^n} \Pr\{Z_1^n = y | \hat{\mathbf{X}}_n^{2n} = x\} \\
&= \max_{x,y \in \mathcal{A}^n} \Pr\{\hat{\mathbf{X}}_n^{2n} = x | \mathbf{Z}_1^n = y\} \frac{\Pr\{\mathbf{Z}_1^n = y\}}{\Pr\{\hat{\mathbf{X}}_n^{2n} = x\}} \\
&\leq \max_{x,y \in \mathcal{A}^n} \frac{\Pr\{\hat{\mathbf{X}}_n^{2n} = x | \mathbf{X}_0^{n-1} = y\}}{\Pr\{\mathbf{X}_n^{2n} = x\}} \gamma_Q(n) \\
&\leq \alpha(n) P \gamma_Q(n).
\end{aligned}
$$

Since this holds for every k, it must be that

$$B_2 \leq 4nt\alpha(n)\gamma_Q(n).$$

Proof of Lemma E: By conditioning first on M and applying the bounds as well as the independence of M and $\mathbf{X}$, we can follow the steps of the proof of Lemma A. The details are left to the reader. A detailed description of random length patterns can be found in [11].

References

[1] Barbour A., Holst, and Janson, *Poisson Approximation*, Oxford Press, 1992.

[2] Kontoyiannis, I., "Asymptotic Recurrence and Waiting Times For Stationary Processes" *Technical Report No. 92*, Stanford University, July 1996.

[3] Farach M., Noordeweir M. , Savari S., Shepp L., Wyner A.J., Ziv J. , "The Entropy of DNA: Algorithms and Measurements Based on Memory and Rapid Convergence", *SODA 1995*, Jan. 1995.

[4] Ornstein, D.S, and Weiss, B. "Entropy and data compression schemes". *IEEE Trans. on Information Theory*, **39**, pp. 78-83, 1993.

[5] Ornstein, D.S, and Weiss, B. "Entropy and Recurrence Rates for Stationary Random Fields". Presented at IEEE Int. Symp. on Information Theory, Trondheim, Norway, Jun27-July 1, 1994.

[6] Shields, P.C., "Waiting times: positive and negative results on the Wyner-Ziv Problem," *Journal of Theoretical Probability*, Vol. 6, nr. 3, pp.499-519 , 1993.

[7] Shields, P.C., "Entropy and Prefixes", *Annals of Probability*, vol. 20, pp 403-409, 1992.

[8] Shields, P.C., *The Ergodic Theory of Discrete Sample Paths,* American Mathematical Society, Providence, 1996.

[9] Jacquet P. and Szpankowski, W., " Autocorrelation on Words and Its Applications: Analysis of Suffix trees by String-Ruler Approach" *Journal of Combinatorial Theory* Vol. 66, NO. 2, May 1994.

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

[11] Wyner, A. J., *String Matching Theorems and Applications to Data Compression and Statistics*, Ph.D. Thesis, Stanford University, 1993.

[12] Wyner, A.J. "The redundancy and distribution of phrase lengths of the FDLZ algorithm" *IEEE Trans. on Information Theory*, September, 1997.

[13] Wyner, A.J. "Entropy Estimation and Patterns" Proceedings of the 1996 ISIT workshop , Haifa, Israel. 1996.

[14] Wyner, A.D., Ziv J, and Wyner A.J. "The Role of Pattern Matching in Information Theory", IEEE Trans. on Information Theory, 50th Anniversary Edition, October, 1999.

Permanent Address: Department of Statistics, The Wharton School, University of Pennsylvania 3000 Steinberg-Dietrich Hall, Philadelphia, PA 19104.6302

email: wyner@stat.wharton.upenn.edu