The Annals of Probability
1992, Vol. 20, No. 3, 1174–1198

RANDOM STATIONARY PROCESSES

BY KENNETH S. ALEXANDER[1] AND STEVEN A. KALIKOW[2]

University of Southern California

Given a finite alphabet, there is an inductive method for constructing a stationary measure on doubly infinite words from this alphabet. This construction can be randomized; the main focus here is on a particular uniform randomization which intuitively corresponds to the idea of choosing a generic stationary process. It is shown that with probability 1, the random stationary process has zero entropy and gives positive probability to every periodic infinite word.

1. Introduction. Stationary processes in which each random variable takes values in a fixed finite set (or *alphabet*) have long been an object of study in ergodic theory. In analogy to random walk in a random environment, we wish to consider here what happens when randomness is used in constructing the law of such a process. We will mainly focus on a randomized construction which intuitively corresponds to the idea of obtaining a generic stationary measure, selected uniformly over all the possibilities, though "uniformly" makes no formal sense here. Of particular interest are entropy and periodicity properties of our generic stationary process.

Let $A = \{x_1, \dots, x_d\}$ be an alphabet. We begin by describing a method, later to be randomized, of constructing an A-valued stationary process $X = \{X_n, n \in \mathbb{Z}\}$. This method has a more probability–theoretic flavor than the standard ergodic–theoretic method of cutting and stacking (see, e.g., [2]). It suffices to define the stationary measure ν on all finite words. We first choose the measure ν_1 on length-one words, by choosing the vector of probabilities $(\nu_1(x_1), \dots, \nu_1(x_d))$ from the set of all d-vectors having nonnegative entries which sum to 1. We then choose the measure ν_2 on length-two words, which is specified by the $d \times d$ matrix $(\nu_2(x_i x_j))$. This matrix is constrained to have nonnegative entries and to have marginals (i.e., row and column sums) given by the measure ν_1 on length-one words. The d^2 entries must thus satisfy $2d - 1$ linearly independent equality constraints, besides being nonnegative, so our choice of a matrix is equivalently specified by a point from a convex polyhedral region in a $(d-1)^2$-dimensional subspace of $\mathbb{R}^{d^2}$.

Then inductively, having specified the measure ν_n on length-n words, we proceed as follows: Fix a length-$(n-1)$ word w_{n-1}. Let W_{jk} denote the word

Received September 1990; revised June 1991.

[1]Research partially supported by NSF Grant DMS-90-06395.

[2]Author now at Cornell University.

AMS 1980 *subject classifications*. Primary 60G10; secondary 28D05, 60G57.

Key words and phrases. Stationary process, entropy, random measure.

$X_j \cdots X_k$. Given $W_{1,n-1} = w_{n-1}$, we know the marginal (conditional) distributions of X_0 and of X_n. The joint (conditional) distribution of X_0 and X_n is then specified by a $d \times d$ matrix, constrained to have nonnegative entries and to have row and column sums given by the marginal (conditional) distributions of X_0 and X_n, respectively. We denote this matrix by $\mathscr{P}(w_{n-1})$. Specifying such a matrix (or equivalently, a point from a polyhedral region, as discussed above) for each of the d^{n-1} possible choices of w_{n-1} determines the measure ν_{n+1} on length-$(n + 1)$ words. Note that these matrices are free parameters in the sense that the selection of any one does not further constrain the others.

This construction is well defined, in the sense that the set of allowed matrices is never empty, because one can always select each matrix entry to be the product of the corresponding marginals, which corresponds to making X_0 and X_n conditionally independent given $W_{1,n-1} = w_{n-1}$. Further, it is easy to see that the sequence of measures $\{\nu_n\}$ is consistent, in that ν_n is the image of ν_{n+1} under either of the projections of A^{n+1} onto A^n given by deleting the first or last letter of an $(n + 1)$-letter word. Since n is arbitrary, this also shows that the finite-dimensional distributions are shift-invariant; thus the measure on infinite words resulting from our construction is indeed always stationary. Further, every A-valued stationary process can be constructed in this way.

This construction is simplest in the case of a two-letter alphabet, say $A = \{0, 1\}$. There, given ν_n and given $W_{1,n-1} = w_{n-1}$, the choice of a joint conditional distribution for X_0 and X_n reduces to the choice of a single parameter from some allowed interval. If we let $a = \nu_n(X_0 = 1 | W_{1,n-1} = w_{n-1})$ and $b = \nu_n(X_n = 1 | W_{1,n-1} = w_{n-1})$, then the joint conditional distribution is determined by the choice of $\nu_n(X_0 = 1, X_n = 1 | W_{1,n-1} = w_{n-1})$ from the allowed interval

$$(1.1) \quad [0, \min(a, b)] \quad \text{if } a + b \le 1, \quad [a + b - 1, \min(a, b)] \quad \text{if } a + b > 1.$$

We can obtain a random stationary measure from this construction by choosing at random at each stage, from the allowed interval or polyhedron, the point which specifies each matrix $\mathscr{P}(w)$. Of course the result depends on the joint distribution of the matrices $\mathscr{P}(w)$, which we have not yet specified. Let $\mathscr{M}_1(A^\infty)$ denote the space of all probability measures on the set of doubly infinite words. The randomized construction gives a probability measure $\hat{\mu}$ on $\mathscr{M}_1(A^\infty)$, concentrated on stationary measures. Any probability measure on $\mathscr{M}_1(A^\infty)$ induces a probability measure on $\mathscr{M}_1(A^\infty) \times A^\infty$; one can choose a measure ν from $\mathscr{M}_1(A^\infty)$ using $\hat{\mu}$ and then use ν to select a specific word X from A^∞. Thus our randomized construction corresponds to a probability measure μ on the space $\mathscr{M}_1(A^\infty) \times A^\infty$, which is the joint law of (ν, X). μ has the property that given ν, X is a stationary process with law ν. We will call any probability measure μ on $\mathscr{M}_1(A^\infty) \times A^\infty$ which has this property a *construction law*.

The (unconditioned) measure ν_n may be viewed as a $d \times \cdots \times d$, n-dimensional array of nonnegative numbers which due to stationarity are subject to certain linear constraints, or equivalently may be viewed as a point in a polyhedral region in a subspace of $\mathbb{R}^{d^n}$. Let J_n denote this polyhedral region, whose points correspond one to one with stationary measures ν_n. In order that μ select a truly generic stationary measure, ideally μ would be such that ν_n were uniform over J_n for all n. Unfortunately the uniform laws on the sets J_n are not compatible in the necessary way as n varies, so there is no such μ.

The most natural option instead, perhaps, is to select each ν_{n+1} uniformly given ν_n, which seems as generic as possible. Since the $\mathscr{P}(w_{n-1})$ are free parameters, this can be accomplished by selecting, given ν_n, the matrices $\{\mathscr{P}(w_{n-1})\colon w_{n-1} \in A^{n-1}\}$ independently, each uniformly over its allowed polyhedral region. We will refer to this as *independent uniform selection*. For clarity of exposition we will first give our main results for this case, further restricting in the proof to $A = \{0, 1\}$. Then in Section 3 we will cover the general case.

Since independent uniform selection seems like a very random way of choosing a stationary process, one might expect that typically the stationary process selected would exhibit very random behavior, but quite the opposite seems to be true.

THEOREM 1.1. *Under independent uniform selection, with μ-probability 1, every doubly infinite periodic word S has $\nu(S) > 0$.*

THEOREM 1.2. *Under independent uniform selection, with μ-probability 1, the stationary process X has zero entropy.*

For intuition purposes, one may consider $A = \{0, 1\}$ and suppose that for some $(n-1)$-letter word w_{n-1}, given $W_{1,n-1} = w_{n-1}$, the letters X_0 and X_n are nearly determined in the sense that for each, one of the two possible values occurs only with small probability, say $\nu_n(X_0 = 1 | W_{1,n-1} = w_{n-1}) = \delta$ and $\nu_n(X_n = 0 | W_{1,n-1} = w_{n-1}) = \varepsilon$ with $0 < \delta < \varepsilon$. We can think of the events $X_0 = 1$ and $X_n = 0$ as errors. If the two errors are nearly independent, then the probability of both errors occurring is of order $\varepsilon\delta$. Under independent uniform selection, however, the ν_{n+1} probability of both errors is chosen uniformly from $[0, \delta]$ so is usually of much larger order δ. Thus errors are typically quite correlated even at long distances, which underlies our two main results.

Given our results, it is natural to ask whether the set of all doubly infinite periodic words has ν-probability 1, μ-a.s. We are unable to resolve this question. Some of our results (see the remarks following Lemma 3.7) make it plausible that only periodic words might occur, but simulations (see Section 5) seem to give some aperiodic words.

A completely different way of introducing randomness into the construction of stationary processes has been studied in the context of time series, with the coefficients in an autoregressive model being chosen randomly. See, for example, [1].

2. Periodic words: Two-letter alphabets, independent uniform selection. In this section we will prove Theorem 1.1 for $A = \{0,1\}$. As a preliminary, let us show that the word of all zeroes has positive probability. Let z_n, $n \le \infty$, be the n-letter word of all zeroes and let $Z_n := \nu(z_n)$, $H_n := Z_{n-1} - Z_n$ and $C_n := \max(Z_n - H_n, 0)$. Then Z_{n+1} is chosen uniformly from $[C_n, Z_n]$: $Z_{n+1} = Z_n - V_n(Z_n - C_n)$ for some iid sequence V_n, $n \ge 1$, uniform in $[0,1]$. Note that if $H_n \le Z_n$, then $H_{n+1} = V_n H_n$.

Fix $k \ge 1$ and let N be the least n such that $\max(V_{n-k}, \dots, V_{n-1}) < 1/3$. Then $H_{N-k+j} \le 3^{-j} Z_{N-k}$ for $j = 1, \dots, k$. Hence

$$Z_N \ge Z_{N-k}/2 \quad \text{and} \quad H_N \le 3^{-k} Z_{N-k} < 3^{-(k-1)} Z_N.$$

Now let T be the first $m > N$ such that $H_m > Z_m$, with $T = \infty$ if there is no such m. For $N \le j < T$ we have $H_{j+1} = V_j H_j$, so letting $M_n := \sum_{j=n}^{\infty} V_j V_{j-1} \cdots V_n$, we obtain

$$Z_N - Z_T = \sum_{j=N+1}^{T} H_j \le M_N H_N < 3^{-(k-1)} M_N Z_N.$$

If $T = \infty$ and $Z_\infty = 0$, this shows $M_N > 3^{k-1}$; if $T < \infty$, then

$$Z_T < H_T \le H_N < 3^{-(k-1)} Z_N,$$

so $Z_N < 3^{-(k-1)}(1 + M_N) Z_N$, which gives $M_N > 3^{k-1} - 1$. The distribution of M_N does not depend on the value of N, so we have

$$\mu(T < \infty) + \mu(T = \infty, Z_\infty = 0) \le \mu\big(M_1 > 3^{k-1} - 1\big)$$

which can be made arbitrarily small by taking k to be large. Thus $\mu(Z_\infty > 0) = 1$.

For general words, let us fix a doubly infinite periodic word $S = \cdots s_{-1} s_0 s_1 \cdots$ and let t be the length of the period. Many of the quantities we deal with in this section will be functions of S, a fact which wll be suppressed in the notation. Define *left and right error probabilities*

$$\alpha_{ni}^{L} := \nu_n\big(X_i \ne s_i | W_{i+1, i+n-1} = s_{i+1} \cdots s_{i+n-1}\big),$$

$$\alpha_{ni}^{R} := \nu_n\big(X_i \ne s_i | W_{i-n+1, i-1} = s_{i-n+1} \cdots s_{i-1}\big)$$

for $n \ge 1$; by stationarity these quantities depend only on the value of $i \bmod t$, so we will think of i as an equivalence class of integers. Note the error probabilities are functions of ν. For each n there are $2t$ error probabilities. These are naturally broken down into t pairs; we say indices i and j are *paired at stage* n if (some representatives of) i and j are the left and right endpoints, respectively, of an n-letter word; that is, if $j - i = n - 1 \bmod t$. We thereby obtain a set of t points in $[0,1]^2$:

$$Q_n := \Big\{\big(\alpha_{ni}^{L}, \alpha_{nj}^{R}\big) \colon 1 \le i \le t,\ (i,j) \text{ paired at stage } n\Big\}.$$

Since every finite word has positive ν-probability μ-a.s., to show $\nu(S) > 0$ μ-a.s., it is enough to show

$$\sum_{n \geq 1} \sum_{i=1}^{t} \left(\alpha_{ni}^L + \alpha_{ni}^R\right) < \infty \quad \mu\text{-a.s.} \tag{2.1}$$

This will be true if the points of Q_n move toward the origin sufficiently rapidly as $n \to \infty$. That in turn will be established in two stages. Roughly, first it will be shown that uniformly in all n and in all possible "pasts" ν_n, the time until the next index $n+m$ for which all points of Q_{n+m} are near the origin is typically not too large. Then supermartingale arguments will be used to show that once all points of Q_n are near the origin for some n, there is a positive probability that from that n onward these points will be drawn toward the origin so rapidly that (2.1) holds.

Let

$$Q'_n := \left\{\left(\alpha_{ni}^L, \alpha_{nj}^R\right) : 1 \leq i \leq t, \, (i,j) \text{ paired at stage } n+1\right\}.$$

The set of first coordinates of the t points in Q_n is the same as for Q'_n, and likewise for the second coordinates, but these coordinates are paired differently. For (i,j) paired at stage n, let

$$\beta_{nij} := \nu_n\left(X_i \neq s_i, X_j \neq s_j | W_{i+1, j-1} = s_{i+1} \cdots s_{j-1}\right)$$

be the probability of errors at both ends of an n-letter subword of S, where the representatives i and j are of course chosen so that $j - i = n - 1$. Now let i and j be paired at stage $n+1$ and let $(\alpha_{ni}^L, \alpha_{nj}^R)$ be the corresponding point of Q'_n. Then, given $W_{i+1, j-1} = s_{s+1} \cdots s_{j-1}$, $\alpha_{n+1,i}^L$ represents the probability of a left-end error at i given there is not a right-end error at j, where $j - i = n$, and similarly for $\alpha_{n+1,j}^R$. So $\alpha_{n+1,i}^L$ and $\alpha_{n+1,j}^R$ have the form

$$\begin{aligned} \alpha_{n+1,i}^L &= \left(\alpha_{ni}^L - \beta_{(n+1)ij}\right) / \left(1 - \alpha_{nj}^R\right), \\ \alpha_{n+1,j}^R &= \left(\alpha_{nj}^R - \beta_{(n+1)ij}\right) / \left(1 - \alpha_{ni}^L\right). \end{aligned} \tag{2.2}$$

Under independent uniform selection, we have four cases; here and throughout this paper we let U denote a generic uniform [0, 1] random variable.

CASE 1. $\alpha_{ni}^L + \alpha_{nj}^R \geq 1$ and $\alpha_{ni}^L \geq \alpha_{nj}^R$. Then [cf. (1.1)] $\beta_{(n+1)ij}$ has the form $\alpha_{ni}^L + \alpha_{nj}^R - 1 + U(1 - \alpha_{ni}^L)$, so

$$\begin{aligned} \alpha_{n+1,i}^L &= \left(1 - \alpha_{nj}^R - U\left(1 - \alpha_{ni}^L\right)\right) / \left(1 - \alpha_{nj}^R\right), \\ \alpha_{n+1,j}^R &= 1 - U. \end{aligned}$$

CASE 2. $\alpha_{ni}^L + \alpha_{nj}^R \geq 1$ and $\alpha_{ni}^L \leq \alpha_{nj}^R$. Then similarly to Case 1, we have

$$\begin{aligned} \alpha_{n+1,i}^L &= 1 - U, \\ \alpha_{n+1,j}^R &= \left(1 - \alpha_{ni}^L - U\left(1 - \alpha_{nj}^R\right)\right) / \left(1 - \alpha_{ni}^L\right). \end{aligned}$$

CASE 3. $\alpha_{ni}^L + \alpha_{nj}^R < 1$ and $\alpha_{ni}^L \ge \alpha_{nj}^R$. Then $\beta_{(n+1)ij}$ has the form $U\alpha_{nj}^R$, so

$$\alpha_{n+1,i}^L = \left(\alpha_{ni}^L - U\alpha_{nj}^R\right)/\left(1-\alpha_{nj}^R\right),$$
$$\alpha_{n+1,j}^R = (1-U)\alpha_{nj}^R/\left(1-\alpha_{ni}^L\right).$$

CASE 4. $\alpha_{ni}^L + \alpha_{nj}^R < 1$ and $\alpha_{ni}^L \le \alpha_{nj}^R$. Then similarly to Case 3, we have

$$\alpha_{n+1,i}^L = (1-U)\alpha_{ni}^L/\left(1-\alpha_{nj}^R\right),$$
$$\alpha_{n+1,j}^R = \left(\alpha_{nj}^R - U\alpha_{ni}^L\right)/\left(1-\alpha_{ni}^L\right).$$

In all four cases, it is easy to verify the following geometric description, valid unless $(1,1) \in Q_n$, which μ-a.s. never occurs:

(2.3) $(\alpha_{n+1,i}^L, \alpha_{n+1,j}^R)$ lies on the line l which passes through $(\alpha_{ni}^L, \alpha_{nj}^R)$ and $(1,1)$. If κ denotes the point on the bottom or left side of $[0,1]^2$ where l meets the boundary, then more precisely, $(\alpha_{n+1,i}^L, \alpha_{n+1,j}^R)$ is chosen uniformly over a segment $\bar{l}$ of l, which has one endpoint at κ, and the second endpoint between $(\alpha_{ni}^L, \alpha_{nj}^R)$ and $(1,1)$, inclusive. This second endpoint is equal to $(1,1)$ if and only if $\alpha_{ni}^L + \alpha_{nj}^R \ge 1$.

Note that $U = 1$ in Cases 1 through 4 gives maximum correlation of the left and right errors and corresponds to selecting the point κ from $\bar{l}$, or to making the maximum possible choice of $\beta_{(n+1)ij}$.

We may think of the progression from Q_n to Q_{n+1} as occurring in two steps: In the first step, the coordinates are re-paired to form Q'_n; in the second step, for each point $(\alpha_{ni}^L, \alpha_{nj}^R)$ of Q'_n, the corresponding point $(\alpha_{n+1,i}^L, \alpha_{n+1,j}^R)$ of Q_{n+1} is chosen from the corresponding line segment $\bar{l}$ as described in (2.3).

We wish to show the points of Q_n rapidly approach the origin. Define the event

$$A_n(\theta) := \left[\max\{\alpha_{ni}^*: 1 \le i \le t,\ ^* = L \text{ or } R\} \le \theta\right],$$

which says that all points of Q_n are near the origin, define the left and lower boundary strip

$$C_\delta := \left\{(x,y) \in [0,1]^2: \min(x,y) \le \delta\right\}$$

and define the event

$$B_n(\delta) := [Q_n \subset C_\delta].$$

From the geometric description (2.3), it is clear that regardless of the location of $(\alpha_{ni}^L, \alpha_{nj}^R)$, the fraction of the line segment $\bar{l}$ which lies in C_δ is always at least δ. Since t points are selected for each n, this establishes the following result.

LEMMA 2.1. *For a two-letter alphabet, under independent uniform selection,*

$$\mu(B_{n+1}(\delta)|\nu_n) \geq \delta^t \quad a.s. \text{ for all } n \geq 1 \text{ and } \delta \in (0,1).$$

We next show that we typically need not wait long in our inductive construction of ν for all the points of Q_n to be near the origin. Define the sum of the $2t$ errors at stage n:

$$\xi_n := \sum_{i=1}^{t} \left(\alpha_{ni}^L + \alpha_{ni}^R\right).$$

LEMMA 2.2. *Let $\theta \in (0,1)$ and for $n \geq 1$ let $T_n := \min\{m \geq 0$: $A_{n+m}(\theta)$ occurs$\}$. For a two-letter alphabet, under independent uniform selection, there exist $r = r(\theta, t) \geq 0$ and $\lambda = \lambda(\theta, t) > 0$ such that $\mu(T_n \leq r|\nu_n) \geq \lambda$ a.s. for all n. Consequently, $\mu(A_n(\theta)$ i.o.$) = 1$ for all $\theta \in (0,1)$.*

PROOF. By Lemma 2.1, it is enough to show that sufficient consecutive occurrences of the events $B_{n+j}(\delta)$ force the occurrence of an event $A_{n+m}(\theta)$; that is, for some r and δ,

$$B_{n+1}(\delta) \cap \cdots \cap B_{n+r}(\delta) \subset [T_n \leq r].$$

Roughly speaking this is because the sum ξ_{n+k} tends to decrease by a bounded-below amount as long as the $B_{n+k}(\delta)$ ($k = 1, 2, \dots$) continue to occur and $A_{n+k}(\theta)$ continues not to occur, and ξ_{n+k} is nonnegative so cannot decrease indefinitely. Let

$$\delta := \theta/2t^2(4t+2).$$

First we show that no error probability can increase much when an event $B_{n+1}(\delta)$ occurs. Let i and j be paired at stage $n+1$, so $(\alpha_{ni}^L, \alpha_{nj}^R)$ is a point of Q_n'. Suppose for simplicity that $\alpha_{ni}^L \leq \alpha_{nj}^R$. Now $(\alpha_{ni}^L, \alpha_{nj}^R)$ and $(\alpha_{n+1,i}^L, \alpha_{n+1,j}^R)$ are both on the line l of (2.3), which a.s. has a slope $\gamma \in (0,1)$ and a y-intercept $b \in [0,1)$. If $B_{n+1}(\delta)$ occurs, we have $\alpha_{n+1,i}^L \leq \delta$ and therefore

$$\alpha_{n+1,j}^R = b + \gamma\alpha_{n+1,i}^L < b + \delta < \alpha_{nj}^R + \delta.$$

A similar argument applies if $\alpha_{ni}^L > \alpha_{nj}^R$, and we obtain

$$B_{n+1}(\delta) \Rightarrow \alpha_{n+1,i}^* < \alpha_{ni}^* + \delta \tag{2.4}$$
$$\text{for all } n \geq 1,\ 1 \leq i \leq t \text{ and } * = L \text{ or } R.$$

Iterating (2.4) gives

$$B_{n+1}(\delta) \cap \cdots \cap B_{n+m}(\delta) \Rightarrow \xi_{n+k} \leq \xi_{n+k-1} + 2t\delta \tag{2.5}$$
$$\text{for all } n, m \geq 1 \text{ and all } 1 \leq k \leq m.$$

Now suppose $B_{n+1}(\delta) \cap \cdots \cap B_{n+2t}(\delta)$ occurs but $T_n > 2t$; that is, none of $A_n(\theta), \dots, A_{n+2t}(\theta)$ occurs. We will show that at some stage between $n+1$ and $n+t$, some error probability has a significant decrease. Specifically, we know there is an $i \leq t$ such that either $\alpha_{n+2t,i}^L > \theta$ or $\alpha_{n+2t,i}^R > \theta$. Let us assume for the moment (Case 1) that $\alpha^L_{n+2t,i} > \theta$; that is, at stage $n + 2t$ at least one of the t left error probabilities is large. We will show that at some stage between $n + t$ and $n + 2t$, at least one of the t right error probabilities must also be large. Essentially this is because, by stationarity and periodicity, if we start from an $(n + t)$-letter errorless word running from location $i + t$ to location $i + 2t + n - 1$, the probability of making no errors when adding t letters onto the left end of this string must be the same as the probability of making no errors when adding t letters onto the right end of the same string. More precisely,

$$
\begin{aligned}
1 - \theta \geq 1 - \alpha^L_{n+2t,i} &\geq \prod_{k=1}^{t} \left(1 - \alpha^L_{n+t+k,i+t-k}\right) \\
&= \frac{\nu(W_{i,i+2t+n-1} = s_i \cdots s_{i,i+2t+n-1})}{\nu(W_{i+t,i+2t+n-1} = s_{i+t} \cdots s_{i+2t+n-1})} \\
&= \frac{\nu(W_{i+t,i+3t+n-1} = s_{i+t} \cdots s_{i,i+3t+n-1})}{\nu(W_{i+t,i+2t+n-1} = s_{i+t} \cdots s_{i+2t+n-1})} \\
&= \prod_{k=1}^{t} \left(1 - \alpha^R_{n+t+k,i+2t+n-1+k}\right),
\end{aligned}
\tag{2.6}
$$

which implies that for some $1 \leq k \leq t$ and (since the second subscript of α is defined mod t) $1 \leq j \leq t$,

$$\alpha^R_{n+t+k,j} \geq \theta/t. \tag{2.7}$$

There must exist an m, $0 \leq m \leq t - 1$, such that i and j are paired at stage $n + m$. By (2.4), since $\alpha^L_{n+2t,i} > \theta$, we have

$$\alpha^L_{n+m,i} \geq \theta - 2t\delta \geq \theta/t - 2t\delta \quad \text{and} \quad \alpha^R_{n+m,j} \geq \theta/t - 2t\delta.$$

Therefore if (Case 1a) $\alpha^L_{n+m,i} \leq \alpha^R_{n+m,j}$, since by assumption $B_{n+m+1}(\delta)$ occurs, as in the argument preceding (2.4) we have

$$\alpha^L_{n+m+1,i} \leq \delta \leq \alpha^L_{n+m,i} - (\theta/t - (2t + 1)\delta).$$

Similarly, if (Case 1b) $\alpha^L_{n+m,i} > \alpha^R_{n+m,j}$, then

$$\alpha^R_{n+m+1,j} \leq \delta \leq \alpha^R_{n+m,j} - (\theta/t - (2t + 1)\delta).$$

Either way, with (2.4) we obtain

$$\xi_{n+m+1} \leq \xi_{n+m} + 4t\delta - \theta/t$$

and therefore by (2.5),

$$\xi_{n+2t} \leq \xi_n + (4t + 2)t\delta - \theta/t \leq \xi_n - \theta/2t. \tag{2.8}$$

A similar argument establishes (2.8) if (Case 2) $\alpha^R_{n+2t,i} > \theta$. Thus we have

$$B_{n+1}(\delta) \cap \cdots \cap B_{n+2t}(\delta) \quad \text{and} \quad T_n > 2t \Rightarrow \xi_{n+2t} \leq \xi_n - \theta/2t. \tag{2.9}$$

Applying (2.9) repeatedly shows that for $u \geq 1$,

$$B_{n+1}(\delta) \cap \cdots \cap B_{n+2ut}(\delta) \quad \text{and} \quad T_n > 2ut \Rightarrow \xi_{n+2ut} \leq \xi_n - u\theta/2t.$$

Since $\xi_n \in [0,1]$ for all n, this shows that for $u > 2t/\theta$, $B_{n+1}(\delta) \cap \cdots \cap B_{n+2ut}(\delta)$ implies $T_n \leq 2ut$, and the lemma follows easily. □

Define the class of functions

$$\mathscr{G} := \{g\colon (0,1) \to \mathbb{R}\colon g \text{ nonincreasing}, 0 \leq g(x) \leq 1/2x, x^2 g(x) \text{ convex}\}.$$

This class will be used when we generalize from independent uniform selection. As we will use the following minor lemma for both that special case and its generalization, we will give it in greater generality than presently needed.

LEMMA 2.3. *Let $g \in \mathscr{G}$, let $0 < x \leq y$ and let X and Y be random variables satisfying $EX \leq x$, $EY \leq y$, $0 \leq X \leq 5x/4$ and $0 \leq Y \leq 5y/4$. Then*

$$E(X + Y - g(\max(X,Y))XY) \leq x + y - g(y)xy/20.$$

In particular, for $0 < c \leq \frac{1}{2}$,

$$E(X + Y - c\min(X,Y)) \leq x + y - cx/20.$$

PROOF. Since $g(z)$ is nonincreasing and bounded by $1/z$, the function $f(x,y) = x + y - g(\max(x,y))xy$ is a nondecreasing function of x or of y when the other variable is held fixed. Therefore it is enough to prove the lemma when $EX = x$ and $EY = y$. But then we have $P[X \geq x/2] \geq 2/3$ and $P[Y \geq y/2] \geq 2/3$, so $P[X \geq x/2, Y \geq y/2] \geq 1/3$. It follows that

$$Eg(\max(X,Y)XY) \geq (1/3)g(5y/4)(x/2)(y/2) \geq g(y)xy/20,$$

where we have used the fact that $z^2 g(z)$ is nondecreasing for $g \in \mathscr{G}$. □

We will next show that, when all error probabilities are near 0, they tend to shrink like an exponentially decreasing supermartingale.

LEMMA 2.4. *For a two-letter alphabet, under independent uniform selection, there exist $\theta_0 = \theta_0(t) > 0$ and $h = h(t) \in (0,1)$ for which*

$$E(\xi_{(k+1)t} | \nu_{kt}) \leq (1-h)\xi_{kt} \quad a.s. \text{ on } A_{kt}(\theta_0) \tag{2.10}$$

for all $k \geq 1$.

PROOF. From Cases 3 and 4 preceding (2.3), we have a supermartingale-like property: For each i and j paired at some stage $n+1$,

$$\begin{gathered} \alpha_{ni}^L < 1/2 \text{ and } \alpha_{nj}^R < 1/2 \Rightarrow \\ E(\alpha_{n+1,i}^L | \nu_n) < \alpha_{ni}^L,\ E(\alpha_{n+1,j}^R | \nu_n) < \alpha_{nj}^R. \end{gathered} \tag{2.11}$$

Further, by (2.2),

$$\alpha^L_{n+1,i} < \alpha^L_{ni}/\left(1-\alpha^R_{nj}\right) \quad \text{and} \quad \alpha^R_{n+1,j} < \alpha^R_{nj}/\left(1-\alpha^L_{ni}\right). \tag{2.12}$$

From (2.12) it is easily seen that once they become small, the error probabilities can increase only very slowly; more precisely, for some $\theta_1 = \theta_1(t) > 0$,

$$\begin{aligned} A_{kt}(\theta_1) \ \Rightarrow \ \alpha^*_{kt+m,i} &\le 5\alpha^*_{kt,i}/4 \\ &\text{for all } 1 \le m \le t, 1 \le i \le t \text{ and } * = R \text{ or } L. \end{aligned} \tag{2.13}$$

Now consider the largest error probability at stage kt; that is, choose $1 \le i \le t$ and $M = R$ or L so that

$$\alpha^M_{kt,i} = \max\{\alpha^*_{kt,u} \colon 1 \le u \le t, \ * = R \text{ or } L\}.$$

Suppose for concreteness that $M = L$; that is, the largest error probability is a left one. Then

$$\alpha^L_{kt,i} \ge \xi_{kt}/2t. \tag{2.14}$$

We claim that for some $1 \le j \le t$,

$$\alpha^R_{kt,j} \ge \xi_{kt}/3t^2. \tag{2.15}$$

In fact, similarly to (2.6) and (2.7) there must be a right error probability $\alpha^R_{kt+v,j}$, with $0 \le v \le t-1$ and $1 \le j \le t$, satisfying $\alpha^R_{kt+v,j} \ge \alpha^L_{kt,i}/t \ge \xi_{kt}/2t^2$. But then our claim (2.15) follows from (2.13).

Now let $0 \le r \le t-1$ be such that (the equivalence classes of) i and j are paired at stage $kt + r + 1$. Note that

$$0 \le b \le a \le 1/6 \ \Rightarrow \ (a - b/2)/(1-b) \le a - b/3;$$

this and (2.13), together with the formulas from Cases 3 and 4, show that for $\theta_0 < \min(\theta_1, 1/8)$ we have

$$\begin{aligned} &E\left(\alpha^L_{kt+r+1,i} + \alpha^R_{kt+r+1,j} | \nu_{kt+r}\right) \\ &\qquad \le \alpha^L_{kt+r,i} + \alpha^R_{kt+r,j} - \min\left(\alpha^L_{kt+r,i}, \alpha^R_{kt+r,j}\right)/3 \quad \mu\text{-a.s. on } A_{kt}(\theta_0). \end{aligned} \tag{2.16}$$

With (2.11), (2.13), (2.14), (2.15) and Lemma 2.3, this shows that

$$\begin{aligned} E\left(\xi_{(k+1)t} | \nu_{kt}\right) &\le E\left(E\left(\xi_{kt+r+1} | \nu_{kt+r}\right) | \nu_{kt}\right) \\ &\le E\left(\xi_{kt+r} - \min\left(\alpha^L_{kt+r,i}, \alpha^R_{kt+r,j}\right)/3 \middle| \nu_{kt}\right) \\ &\le \xi_{kt} - \min\left(\alpha^L_{kt,i}, \alpha^R_{kt,j}\right)/60 \\ &\le (1 - 1/180t^2)\xi_{kt}, \quad \mu\text{-a.s. on } A_{kt}(\theta_0). \end{aligned} \tag{2.17}$$

□

PROOF OF THEOREM 1.1 FOR TWO-LETTER ALPHABETS. From Lemma 2.4 [or just from (2.11)] we know that $\{\min(\xi_{kt}, \theta_0) \colon k \ge 1\}$ is a nonnegative supermartingale, hence convergent a.s. From Lemma 2.2 and (2.2) [cf. (2.12)] we obtain that

$$\liminf_{k \to \infty} \xi_{kt} = 0 \quad \mu\text{-a.s.},$$

so $\xi_{kt} \to 0$ μ-a.s., and $E\xi_{kt} \to 0$.

To obtain summability, we need to alter the process so that (2.10) is valid a.s., not just on $A_{kt}(\theta_0)$. This is accomplished by removing most of the finite set of terms ξ_{kt} which are larger than θ_0. Specifically, define stopping times $N(n)$ inductively by

$$N(n+1) = \begin{cases} N(n) + 1, & \text{if } \xi_{N(n)t} < \theta_0, \\ \min\{m > N(n): \xi_{mt} \le (1-h)\theta_0\}, & \text{if } \xi_{N(n)t} \ge \theta_0. \end{cases}$$

Let $\mathscr{F}_n$ denote the σ-field $\sigma(\nu_{N(n)t}, N(n))$ and $Y_n := \min(\xi_{N(n)t}, \theta_0)$. If $Y_n = \theta_0$, then $Y_{n+1} \le (1-h)\theta_0$, so

$$E(Y_{n+1} | \mathscr{F}_n) \le (1-h)\theta_0 = (1-h)Y_n. \tag{2.18}$$

If $Y_n < \theta_0$, then by Lemma 2.4,

$$E(Y_{n+1} | \mathscr{F}_n) \le E(\xi_{(N(n)+1)t} | \mathscr{F}_n) \le (1-h)\xi_{N(n)t} = (1-h)Y_n. \tag{2.19}$$

Therefore,

$$EY_{n+1} \le (1-h)EY_n \quad \text{for all } n, \tag{2.20}$$

so that $\Sigma_n EY_n < \infty$ and then $\Sigma_n Y_n < \infty$ a.s. Since $\xi_{kt} \to 0$ μ-a.s., the sequence $\{Y_n, n \ge 1\}$ incorporates all but finitely many of the terms in the sequence $\{\xi_{kt}, k \ge 1\}$, and it follows that

$$\sum_{k=1}^{\infty} \xi_{kt} < \infty \quad \mu\text{-a.s.}$$

Because of (2.13), then

$$\sum_{n=1}^{\infty} \xi_n < \infty \quad \mu\text{-a.s.}$$

This [cf. (2.1)] proves the theorem. □

3. Periodic words: The general case. When the alphabet has more than two letters, there is more than one letter choice which constitutes an error, so $\beta_{(n+1)ij}$ is the sum of several entries of the matrix giving the conditional joint distribution of X_0 and X_n; as a result, under uniform selection of this matrix, $\beta_{(n+1)ij}$ will not be uniform over its allowed interval, given ν_n. Thus we need a generalization of the method in Section 2, to cover nonuniform $\beta_{(n+1)ij}$. Our results, specifically Theorem 3.2, will be stronger than the minimum necessary to handle the case of independent uniform selection of matrices, in order to focus attention on the robustness of our methods under deviations from uniformity.

There are two main ingredients that make the proof of Theorem 1.1, for two-letter alphabets with independent uniform selection, work. The first is that the lim inf of the sum of error probabilities is 0. The second is that, when large error possibilities are omitted, the expected values of the error probabilities are summable, which follows from (2.20). But (2.20) is much stronger than needed. This extra strength essentially comes from the fact discussed after the statement of Theorem 1.2: The probability $\beta_{(n+1)ij}$ of errors at both ends of an $(n+1)$-letter word, when the middle $n-1$ letters match those of S, is chosen uniformly over $[0, \min(\alpha_{ni}^L, \alpha_{nj}^R)]$ when both α_{ni}^L and α_{nj}^R are small, so that $E(\beta_{(n+1)ij}|\nu_n) = \min(\alpha_{ni}^L, \alpha_{nj}^R)/2$, which is much more than the value $\alpha_{ni}^L \alpha_{nj}^R$ of $\beta_{(n+1)ij}$ under independence of the two errors. In the general case, it is still necessary to bound the joint distribution of error probabilities away, on average, from the independent-errors case, by requiring that

$$E\left(\beta_{(n+1)ij}|\nu_n\right) \geq g\left(\max\left(\alpha_{ni}^L, \alpha_{nj}^R\right)\right)\alpha_{ni}^L \alpha_{nj}^R$$

for some function g which becomes large as its argument approaches 0. Note this condition is weaker than if "max" were replaced by "min," and it holds for independent uniform selection with $g(x) = 1/2x$. It says that the amount of overlap between the two error events must on the average be a large multiple of what it would be for independent errors.

To find the exact condition to impose on g, we can make use of the following lemma. Recall that

$$\mathscr{G} = \left\{g\colon (0,1) \to \mathbb{R}\colon g \text{ nonincreasing}, 0 < g(x) \leq 1/2x, x^2 g(x) \text{ convex}\right\}.$$

Note that if $g \in \mathscr{G}$, then $x^2 g(x) \to 0$ as $x \to 0$, so $x^2 g(x)$ is increasing.

Lemma 3.1. *Let $g \in \mathscr{G}$, let $a_1 \in (0,1)$ and define $a_{n+1} = a_n - g(a_n)a_n^2$, $n \geq 1$. Then*

$$\sum_{n=1}^{\infty} a_n < \infty$$

if and only if

$$\int_0^1 1/xg(x)\,dx < \infty.$$

Proof. Let $J_k := (a_1/2^{k+1}, a_1/2^k]$ and $I_k := \{n\colon a_n \in J_k\}$. Then I_k is an interval of integers since $\{a_n\}$ is decreasing. The gaps between successive points a_n in J_k are of length between $(a_1/2^{k+1})^2 g(a_1/2^{k+1})$ and $(a_1/2^k)^2 g(a_1/2^k)$. Therefore

$$1/2(a_1/2^k)g(a_1/2^k) - 1 < \left[(a_1/2^{k+1})/(a_1/2^k)^2 g(a_1/2^k)\right] \leq |I_k|,$$

where $[\cdot]$ denotes the integer part and

$$|I_k| \leq (a_1/2^{k+1})/(a_1/2^{k+1})^2 g(a_1/2^{k+1}) = 1/(a_1/2^{k+1})g(a_1/2^{k+1}).$$

Hence

$$\sum_{n \in I_k} a_n \geq 1/4g(a_1/2^k) - a_1/2^{k+1} \geq (1/4)\int_{J_k} 1/xg(x)\,dx - a_1/2^{k+1}$$

and

$$\sum_{n \in I_k} a_n \leq 2/g(a_1/2^{k+1}) \leq 4\int_{J_k} 1/xg(x)\,dx,$$

from which the lemma follows easily. □

THEOREM 3.2. *For some finite alphabet A, doubly infinite periodic word S and construction law* μ, *suppose that*:

(i) $\liminf_n \xi_n = 0$ μ-*a.*, *or equivalently* $\mu(A_n(\theta)\ i.o.) = 1$ *for all* $\theta > 0$.
(ii) *For some* $\theta_0 > 0$, $E(\beta_{(n+1)ij} | \nu_n) \geq g(\max(\alpha_{ni}^L, \alpha_{nj}^R))\alpha_{ni}^L \alpha_{nj}^R$ μ-*a.s. on* $A_n(\theta_0)$ *for all n and all* (i, j) *paired at stage* $n+1$, *where* $g \in \mathscr{G}$ *satisfies*

$$\int_0^1 1/xg(x)\,dx < \infty.$$

Then $\nu(S) > 0$ μ-*a.s.*

Obviously the integral condition in (ii) can be satisfied with functions $g(x)$ much smaller than the choice $g(x) = 1/2x$ which, as previously noted, appears for independent uniform selection. The integral condition is in a sense necessary—see Proposition 3.6.

Before proving Theorem 3.2, note that the proof of Lemma 2.2 uses the assumption of a two-letter alphabet with independent uniform selection only to permit the application of Lemma 2.1. As the lower bound δ^t from Lemma 2.1 can be replaced with any positive constant $c(\delta, t)$, we have the following.

LEMMA 3.3. *For a finite alphabet A, doubly infinite periodic word S and construction law* μ, *suppose that for each* $\delta \in (0,1)$ *there exists* $c(\delta, S) > 0$ *such that* $\mu(B_{n+1}(\delta) | \nu_n) \geq c(\delta, S)$ *a.s. for all* $n \geq 1$. *Then* $\mu(A_n(\theta)\ i.o.) = 1$ *for all* $\theta > 0$.

Substituting for Lemma 2.4 we have the following.

LEMMA 3.4. *For an arbitrary finite alphabet A, doubly infinite periodic word S and construction law* μ, *under Theorem* 3.2(ii), *there exists* $\theta_2 = \theta_2(t) \in (0, 1/2)$ *and* $\gamma = \gamma(t) \in (0,1)$ *for which*

$$E(\xi_{(k+1)t} | \nu_{kt}) \leq \xi_{kt} - \gamma g(\xi_{kt})\xi_{kt}^2 \quad a.s.\ on\ A_{kt}(\theta_2).$$

PROOF. By (2.2) and Theorem 3.2(ii), since $g(x) \to \infty$ as $x \to 0$, for some $\theta_1 \le \theta_0$ of (ii) we have for all n, i, j as in (ii),

$$\alpha_{ni}^L < \theta_1 \quad \text{and} \quad \alpha_{nj}^R < \theta_1 \Rightarrow$$

$$(3.1) \qquad \begin{aligned} &E\left(\alpha_{n+1,i}^L | \nu_n\right) \le \alpha_{ni}^L - g\left(\max\left(\alpha_{ni}^L, \alpha_{nj}^R\right)\right)\alpha_{ni}^L \alpha_{nj}^R/2, \\ &E\left(\alpha_{n+1,j}^R | \nu_n\right) \le \alpha_{nj}^R - g\left(\max\left(\alpha_{ni}^L, \alpha_{nj}^R\right)\right)\alpha_{ni}^L \alpha_{nj}^R/2 \end{aligned}$$

and (2.13) is valid.

Fix $k \ge 1$. Note that (2.14) and (2.15) are valid for some i and j, and let $0 \le r \le t-1$ be such that i and j are paired at stage $kt + r + 1$. By (2.13) and (3.1), parallel to (2.16) we have for $\theta_2 = \theta_1/2$,

$$\begin{aligned} &E\left(\alpha_{kt+r+1,i}^L + \alpha_{kt+r+1,j}^R | \nu_{kt+r}\right) \\ &\quad \le \alpha_{kt+r,i}^L + \alpha_{kt+r,j}^R - g(\max\left(\alpha_{kt+r,i}^L, \alpha_{kt+r,j}^R\right)\alpha_{kt+r,i}^L \alpha_{kt+r,j}^R \\ &\qquad \mu\text{-a.s. on } A_{kt}(\theta_2). \end{aligned}$$

With (3.1), (2.13), (2.14), (2.15) and Lemma 2.3, parallel to (2.17) this shows that

$$\begin{aligned} E\left(\xi_{(k+1)t} | \nu_{kt}\right) &\le E\left(E\left(\xi_{kt+r+1} | \nu_{kt+r}\right) | \nu_{kt}\right) \\ &\le E\left(\xi_{kt+r} - g(\max\left(\alpha_{kt+r,i}^L, \alpha_{kt+r,j}^R\right)\alpha_{kt+r,i}^L \alpha_{kt+r,j}^R | \nu_{kt}\right) \\ &\le \xi_{kt} - g(\max\left(\alpha_{kt,i}^L, \alpha_{kt,j}^R\right)\alpha_{kt,i}^L \alpha_{kt,j}^R \\ &\le \xi_{kt} - g(\xi_{kt}/2t)(\xi_{kt}/2t)\left(\xi_{kt}/3t^2\right) \\ &\le \xi_{kt} - g(\xi_{kt})\xi_{kt}^2/6t^3 \quad \mu\text{-a.s. on } A_{kt}(\theta_2). \end{aligned}$$ □

PROOF OF THEOREM 3.2. Just as in the proof of Theorem 1.1 for two-letter alphabets, from Theorem 3.2(i) and Lemma 3.4 we obtain $\xi_{kt} \to 0$ μ-a.s. and $E\xi_{kt} \to 0$. Continuing to parallel that proof, we define stopping times $N(n)$ by

$$N(n+1) = \begin{cases} N(n) + 1, & \text{if } \xi_{N(n)t} < \theta_2, \\ \min\left\{m > N(n) \colon \xi_{mt} \le \theta_2 - \gamma g(\theta_2)\theta_2^2\right\}, & \text{if } \xi_{N(n)t} \ge \theta_2, \end{cases}$$

where γ and θ_2 are from Lemma 3.4, and let $Y_n := \min(\xi_{N(n)t}, \theta_2)$. By (2.2) [cf. (2.12)], since $\theta_2 < 1/2$ we have $\xi_{N(n)t} < 2\theta_2$ for all n. For $g \in \mathscr{G}$ it is easily verified that $x - \gamma g(x)x^2$ is nondecreasing in a neighborhood of 0, which we may assume includes $(0, 2\theta_2)$. As in (2.18)–(2.20), by Lemma 3.4 and convexity of $g(x)x^2$ we have

$$EY_{n+1} \le E\left(Y_n - \gamma g(Y_n)Y_n^2\right) \le EY_n - \gamma g(EY_n)(EY_n)^2.$$

Letting $a_1 := EY_1$ and $a_{n+1} := a_n - \gamma g(a_n)a_n^2$, $n \ge 1$, we then have by obvious induction using monotonicity of $x - \gamma g(x)x^2$ that $EY_n \le a_n$ for all n. Since $\gamma g \in \mathscr{G}$, Theorem 3.2(ii) and Lemma 3.1 then show that $\sum_n EY_n < \infty$. As with Theorem 1.1, this proves the result. □

To prove Theorem 1.1 for general finite alphabets, we will show that the hypotheses of Theorem 3.2 are satisfied. The main tool is the next lemma on random joint distributions with specified marginals. Let $\mathbb{Q}_d$ denote the set of $d \times d$ matrices with nonnegative entries which sum to 1. Define $B\colon \mathbb{Q}_d \to [0,1]$ by

$$B(q) := \sum_{j=2}^{d} \sum_{k=2}^{d} q_{jk}.$$

If $\sum_{k=1}^{d} q_{1k} = u_1$ and $\sum_{j=1}^{d} q_{j1} = v_1$, then

$$B(q) = 1 - u_1 - v_1 + q_{11} \tag{3.2}$$

and

$$B(q) \in \left[\max(1 - u_1 - v_1, 0), \min(1 - u_1, 1 - v_1)\right]. \tag{3.3}$$

LEMMA 3.5. *Suppose* $u_1, \dots, u_d, v_1, \dots, v_d > 0$ *with* $d \geq 2$, $\sum_{i=1}^{d} u_i = 1$, $\sum_{i=1}^{d} v_i = 1$ *and* $u_1 \leq v_1$. *Let* Q *be a random matrix uniformly distributed over the set*

$$D := \left\{ q \in \mathbb{Q}_d\colon \sum_{k=1}^{d} q_{jk} = u_j \text{ for all } j \leq d,\ \sum_{j=1}^{d} q_{jk} = v_k \text{ for all } k \leq d \right\}.$$

Let Z *be a random variable with density*

$$f(\beta) := c(1 - v_1 - \beta)^{2d-4}, \qquad \beta \in \left[\max(1 - u_1 - v_1, 0), 1 - v_1\right],$$

where c *is such that* f *is a density. Then* $B(Q)$ *is stochastically larger than* Z.

Before proving this lemma, let us see how it is used to prove Theorem 1.1.

PROOF OF THEOREM 1.1, GENERAL CASE. Let A be the alphabet and $d := |A|$, and fix a doubly infinite periodic word S. By Theorem 3.2 and Lemma 3.3 it is sufficient to prove the following:

(3.4) for each $\delta \in (0,1)$ there exists $c(\delta, S) > 0$ such that $\mu(B_{n+1}(\delta) | \nu_n) \geq c(\delta, S)$ a.s. for all $n \geq 1$,

and

(3.5) for some $\gamma > 0$, $E(\beta_{(n+1)ij} | \nu_n) \geq \gamma \min(\alpha_{ni}^L, \alpha_{nj}^R)$ μ-a.s. on $A_n(1/2)$ for all n and all (i,j) paired at stage $n+1$.

Fix n, i and j as in (3.5). We may assume that: (i) the representatives i and j are chosen so $j - i = n$; (ii) $\alpha_{ni}^L \geq \alpha_{nj}^R$; and (iii) the correct letters at i and j are $s_i = s_j = a_1$. (The proof does not depend on what the correct letters are.) Let $Q := \mathscr{P}(s_{i+1} \cdots s_{j-1})$, $u_1 := 1 - \alpha_{ni}^L$ and $v_1 := 1 - \alpha_{nj}^R$. Then the probability of errors at both i and j is

$$\beta_{(n+1)ij} = B(Q).$$

Now the geometric description (2.3) and the note immediately following remain valid in the present context except that the choice of $(\alpha^L_{n+1,i}, \alpha^R_{n+1,j})$ from $\bar{I}$, or equivalently the variable U, is no longer necessarily uniform. It is therefore clear that the point $(\alpha^L_{n+1,i}, \alpha^R_{n+1,j})$ will fall in C_δ provided that $\beta_{(n+1)ij}$ falls in the upper δ fraction of its possible interval of values $[\max(1 - u_1 - v_1, 0), 1 - v_1]$ [cf. (3.3)]. (By the upper δ fraction of an interval $[x, y]$ we mean the interval $[\delta x + (1-\delta)y, y]$.) By Lemma 3.5 the probability of such a value for $\beta_{(n+1)ij}$ is at least as great as for the random variable Z of that lemma; the probability that Z falls in the upper δ fraction is readily shown to be at least δ^{2d-3}. Therefore, as there are t choices of (i, j) paired at stage n,

$$\mu(B_{n+1}(\delta)|\nu_n) \geq \delta^{(2d-3)t} \quad \text{a.s. for all } n \geq 1,$$

and (3.4) is proved.

Turning to (3.5), note that if $\max(\alpha^L_{ni}, \alpha^R_{nj}) \leq 1/2$, then $1 - u_1 - v_1 \leq 0$, so that in Lemma 3.5 we have $EZ = (1 - v_1)/(2d-2)$. Therefore by Lemma 3.5, on $A_n(1/2)$ we have

$$E\big(\beta_{(n+1)ij}|\nu_n\big) \geq (1 - \nu_1)/(2d-2) = \min\big(\alpha^L_{ni}, \alpha^R_{nj}\big)/(2d-2) \quad \mu\text{-a.s.},$$

which proves (3.5). □

Proof of Lemma 3.5. Let $\varphi(\cdot)$ denote the density of $B(Q)$. It is sufficient to show that

$$\text{(3.6)} \qquad \begin{array}{l} \varphi(\beta)/(1 - \nu_1 - \beta)^{2d-4} \text{ is a nondecreasing function of } \beta \text{ on} \\ I := \big[\max(1 - u_1 - v_1, 0), 1 - v_1\big]. \end{array}$$

Let $\tilde{\mathbb{Q}}_d$ denote the set of $d \times d$ matrices with positive entries which sum to at most 1, let π: $\mathbb{R}^{d\times d} \to \mathbb{R}^{(d-1)\times(d-1)}$ denote projection onto the upper left $(d-1) \times (d-1)$ submatrix, π_D the restriction of π to D and

$$\begin{aligned} \tilde{D} := \pi(D) = \Bigg\{ \tilde{q} \in \tilde{\mathbb{Q}}_{d-1} &: \sum_{k=1}^{d-1} \tilde{q}_{jk} \leq u_j \text{ for all } j \leq d-1, \\ &\sum_{j=1}^{d-1} \tilde{q}_{jk} \leq v_k \text{ for all } k \leq d-1, \sum_{j=1}^{d-1} \sum_{k=1}^{d-1} \tilde{q}_{jk} \geq 1 - u_d - v_d \Bigg\}. \end{aligned}$$

Then π_D is a one-to-one linear map of D onto $\tilde{D}$ which carries uniform measure to uniform measure. Define [see (3.2)]

$$D_\beta := \{q \in D: B(q) = \beta\} = \{q \in D: q_{11} = u_1 + v_1 + \beta - 1\},$$

$$\tilde{D}_\beta := \pi(D_\beta) = \big\{\tilde{q} \in \tilde{D}: \tilde{q}_{11} = u_1 + v_1 + \beta - 1\big\}.$$

Then for some constant c,

$$\text{(3.7)} \qquad \varphi(\beta) = c \operatorname{vol}\big(\tilde{D}_\beta\big) \quad \text{for all } \beta,$$

where vol$(\cdot)$ denotes $(d-1)^2$-dimensional volume.

Fix β and $\varepsilon > 0$ such that both β and $\beta + \varepsilon$ are in I. In view of (3.6) and (3.7), we wish to define a one-to-one map $\tilde{\eta}_\varepsilon$ of $\tilde{D}_\beta$ into $\tilde{D}_{\beta+\varepsilon}$, which shrinks volumes by at most a restricted amount. There is of course a corresponding map $\eta_\varepsilon = \pi_D^{-1} \circ \tilde{\eta}_\varepsilon \circ \pi_D$ of D_β into $D_{\beta+\varepsilon}$, which is what we actually define first. What we would like η_ε to do is as follows. A matrix $q \in D$ can be divided into four disjoint parts: the upper left element $R_{11}(q) = q_{11}$, the lower right $(d-1) \times (d-1)$ submatrix $R_{22}(q)$, the lower left $(d-1) \times 1$ submatrix $R_{21}(q)$ and the upper right $1 \times (d-1)$ submatrix $R_{12}(q)$. Thinking of q as a joint distribution, we will define η_ε so that it increases the mass in each of R_{11} and R_{22} by ε, decreases the mass in each of R_{12} and R_{21} by ε, keeps the relative masses the same within R_{12} and within R_{21} and keeps the marginal distributions the same (so that η_ε maps into D).

To construct η_ε, we first replace each entry in R_{12} and R_{21} with the corresponding relative proportion. Thus define χ: $D \to \mathbb{R}^{d \times d}$ by

$$\chi(q)_{jk} := \begin{cases} q_{jk}, & \text{if } j = k = 1 \text{ or } j \geq 2,\ k \geq 2, \\ q_{1k}/(u_1 - q_{11}), & \text{if } j = 1,\ k \geq 2, \\ q_{j1}/(v_1 - q_{11}), & \text{if } j \geq 2,\ k = 1. \end{cases}$$

Note that χ is a one-to-one map, and for $\zeta \in \chi(D)$ the inverse image $\chi^{-1}(\zeta)$ is given by multiplying the entries in $R_{12}(\zeta)$ and $R_{21}(\zeta)$ by $u_1 - \zeta_{11}$ and by $v_1 - \zeta_{11}$, respectively. Next we add on to the coordinates in R_{11} and R_{22}, defining ζ_ε: $\chi(D) \to \mathbb{R}^{d \times d}$ by

$$\zeta_\varepsilon(\chi)_{jk} := \begin{cases} \chi_{11} + \varepsilon, & \text{if } j = k = 1, \\ \chi_{jk} + \varepsilon \chi_{j1} \chi_{1k}, & \text{if } j \geq 2,\ k \geq 2, \\ \chi_{jk}, & \text{if } j = 1,\ k \geq 2 \text{ or } j \geq 2,\ k = 1. \end{cases}$$

Finally define η_ε on D_β by

$$\eta_\varepsilon := \chi^{-1} \circ \zeta_\varepsilon \circ \chi.$$

It is straightforward to check ζ_ε maps $\chi(D_\beta)$ into $\chi(D_{\beta+\varepsilon})$, so that η_ε maps D_β into $D_{\beta+\varepsilon}$. Furthermore all of these maps are determined by looking at the upper left $(d-1) \times (d-1)$ submatrix. More formally, the maps $\tilde{\chi}^{-1}, \tilde{\zeta}_\varepsilon, \tilde{\chi}$ given on appropriate subsets of $\mathbb{R}^{(d-1)\times(d-1)}$ by the preceding formulas form a commuting diagram with π and $\chi^{-1}, \zeta_\varepsilon, \chi$. Thus we have

$$\tilde{\eta}_\varepsilon := \pi_D \circ \eta_\varepsilon \circ \pi_D^{-1} = \tilde{\chi}^{-1} \circ \tilde{\zeta}_\varepsilon \circ \tilde{\chi}.$$

To compute the volume change, note that the Jacobian matrices of $\tilde{\chi}$ and $\tilde{\zeta}_\varepsilon$ are lower triangular when the coordinates are ordered using R_{11} then R_{12} then R_{21} then R_{22}. From this we easily compute the Jacobian at a point $\tilde{q} \in \tilde{D}_\beta$:

$$\begin{aligned} \partial\tilde{\chi}/\partial\tilde{q} &= 1/(u_1 - \tilde{q}_{11})^{d-2}(v_1 - \tilde{q}_{11})^{d-2} \\ &= 1/(1 - u_1 - \beta)^{d-2}(1 - v_1 - \beta)^{d-2}, \end{aligned}$$

using (3.2) and the Jacobian

$$\partial\tilde{\zeta}_\varepsilon/\partial\tilde{\chi} = 1.$$

Therefore, for $\tilde{q} \in \tilde{D}_\beta$,

$$\begin{aligned}\partial\tilde{\eta}_\varepsilon/\partial\tilde{q} &= \left(1 - u_1 - (\beta + \varepsilon)\right)^{d-2}\left(1 - v_1 - (\beta + \varepsilon)\right)^{d-2}/ \\ &\quad (1 - u_1 - \beta)^{d-2}(1 - v_1 - \beta)^{d-2} \\ &\geq \left(1 - v_1 - (\beta + \varepsilon)\right)^{2d-4}/(1 - v_1 - \beta)^{2d-4}.\end{aligned}$$

It follows that

$$\mathrm{vol}\left(\tilde{D}_{\beta+\varepsilon}\right) \geq \mathrm{vol}\left(\tilde{D}_\beta\right)\left(1 - v_1 - (\beta + \varepsilon)\right)^{2d-4}/(1 - v_1 - \beta)^{2d-4},$$

which with (3.6) and (3.7) proves the lemma. □

Let us now show that the integral condition in Theorem 3.2(ii) cannot be weakened.

PROPOSITION 3.6. *Let A be a finite alphabet and let $g \in \mathscr{G}$ with*

$$\int_0^1 1/xg(x)\,dx = \infty.$$

Then there exists a construction law μ such that for every doubly infinite periodic word S, all hypotheses of Theorem 3.2 other than the integral condition are satisfied, but $\nu(S) = 0$ μ-a.s.

PROOF. We may assume A is a two-letter alphabet, say $\{0, 1\}$, by putting no mass on other letters. We may also assume $g(x) \to \infty$ as $x \to 0$. Let $\delta > 0$ be such that $g(x) > 1$ on $(0, \delta)$.

Given i, j, a finite word $x = x_i \cdots x_j$ and a doubly infinite periodic word S with period t, we say S is a periodic extension of x if x is a subword of S and $|x| \geq 2t$. The latter condition ensures that a given x has at most one periodic extension.

Pick $\nu_1(0)$ uniformly in $(0, 1)$. Given ν_n and $x_1 \cdots x_{n-1} \in A^{n-1}$, we define the joint distribution of X_0 and X_n given $W_{1,n-1} = x_1 \cdots x_{n-1}$ as follows:

Specify

$$\begin{aligned}&\nu_{n+1}\left(X_0 \neq s_0, X_n \neq s_n | W_{1,n-1} = x_1 \cdots x_{n-1}\right) \\ &\qquad = g\left(\max\left(\alpha_{n0}^L, \alpha_{nn}^R\right)\right)\alpha_{n0}^L\alpha_{nn}^R \end{aligned} \tag{3.8}$$

if $x_1 \cdots x_{n-1}$ has a periodic extension S, and $A_n(\delta)$ occurs for S. (Here the error probabilities α_{ni}^* are defined for S).

(3.9) Otherwise, choose $\nu_{n+1}(X_0 = 1,\ X_n = 1 | W_{1,n-1} = x_1 \cdots x_{n-1})$ uniformly over its allowed interval.

This determines ν_{n+1}. In this way we obtain the law μ.

Now fix a periodic S with period t and let T be the least $n > 2t$ for which $A_n(\delta)$ occurs for S, with $T = \infty$ if there is no such n. Note that T is a function of ν. Since independent uniform selection is used to determine probabilities for all length-n subwords of S when $2t < n < T$, it follows from Lemma 2.2 that $T < \infty$ μ-a.s.

The first inequality in Theorem 3.2(ii), with $\theta_0 = \delta$, is immediate under (3.8), and under (3.9) follows from the fact $g(x) \le 1/2x$.

Since $g > 1$, it is easily verified that if $n > 2t$ and $A_n(\delta)$ occurs, then the errors for S satisfy

$$\alpha^*_{n+1,i} \le \alpha^*_{ni} \quad \text{for } i = 1, \dots, t \text{ and } * = R \text{ or } L$$

so $A_{n+1}(\delta)$ also occurs. It follows that (3.8), and not (3.9), will be used to define ν_{n+1} for all $n \ge T$. It also follows that each $\lim_n \alpha^*_{ni}$ exists; by an argument analogous to the proof of Lemma 3.4 it can be shown that no such limit can be strictly positive. Thus

$$\lim_n \xi_n = 0 \quad \mu\text{-a.s.},$$

so Theorem 3.2(i) holds.

For all $n \ge T$ and (i, j) paired at stage $n + 1$, by (3.8) and monotonicity of g,

$$\begin{aligned} \alpha^L_{n+1,i} + \alpha^R_{n+1,j} &\ge \alpha^L_{ni} + \alpha^R_{nj} - 2g\big(\max\big(\alpha^L_{ni}, \alpha^R_{nj}\big)\big)\alpha^L_{ni}\alpha^R_{nj} \\ &\ge \alpha^L_{ni} + \alpha^R_{nj} - g\big(\alpha^L_{ni}\big)\big(\alpha^L_{ni}\big)^2 - g\big(\alpha^R_{nj}\big)\big(\alpha^R_{nj}\big)^2 \end{aligned}$$

so that, since $x^2 g(x)$ is convex and approaches 0 as $x \to 0$,

$$\xi_{n+1} \ge \xi_n - g(\xi_n)\xi_n^2.$$

By Lemma 3.1 [trivially modified to allow for functions defined on $(0, \delta)$] this shows

$$\sum_n \xi_n = \infty \quad \mu\text{-a.s.},$$

which implies $\nu(S) = 0$ μ-a.s. □

For our remaining results we need a new concept of "error" that is not relative to any fixed word S. Given a stationary measure ρ and $w_{n-1} \in A^{n-1}$, there exist *preferred letters* $p_L(\rho_n, w_{n-1})$ and $p_R(\rho_n, w_{n-1})$ in A, with corresponding error probabilities $\gamma_L(\rho_n, w_{1,n-1})$ and $\gamma_R(\rho_n, w_{1,n-1})$, together defined by

$$\begin{aligned} 1 - \gamma_L(\rho_n, w_{1,n-1}) &:= \rho_n(X_0 = p_L | W_{1,n-1} = w_{n-1}) \\ &= \max\{\rho_n(X_0 = x_i | W_{1,n-1} = w_{n-1}) \colon 1 \le i \le d\} \end{aligned}$$

and

$$1 - \gamma_R(\rho_n, w_{1,n-1}) := \rho_n(X_n = p_L | W_{1,n-1} = w_{n-1})$$
$$= \max\{\rho_n(X_n = x_i | W_{1,n-1} = w_{n-1}) \colon 1 \le i \le d\},$$

with the convention that the preferred letter x_i has the minimal index i for which the maximum is achieved, if a tie must be broken; thus each preferred letter is unique. Let

$$L_n := \gamma_L(\nu_n, W_{1,n-1}), \qquad R_n := \gamma_R(\nu_n, W_{1,n-1}).$$

Let $\mathscr{S}$ denote the set of all doubly infinite periodic words. As has been mentioned, it is possible, but we are unable to prove, that $\nu(\mathscr{S}) = 1$ μ-a.s. The next two results may be considered (weak) evidence for this possibility; they will also be used in the proof of Theorem 1.2. The first result may be interpreted as saying that, regardless of μ, if we construct a semiinfinite word by choosing ν and then building toward the right, choosing X_1 then X_2 then X_3 and so forth, and the sum of the right error probabilities encountered as we build is finite, then the word built essentially must be periodic. This is not immediately obvious because the preferred letter at stage n need not be X_n in general.

LEMMA 3.7. *For any finite alphabet A and construction law μ,*

$$\mu\left(\sum_{n=1}^{\infty} R_n < \infty,\ X \text{ not periodic}\right) = 0.$$

PROOF. Define events

$$G := \left[\sum_{n=1}^{\infty} R_n < \infty\right],$$

$$H := [X_n \ne p_R(\nu_n, W_{1,n-1}) \text{ i.o.}].$$

Fix $\varepsilon > 0$ and choose N large enough so that the event

$$G_N := \left[\sum_{n=N}^{\infty} R_n \le \varepsilon\right]$$

has $\mu(G_N) \ge \mu(G) - \varepsilon$. Define

$$G_N^k := \left[\sum_{n=N}^{k} R_n \le \varepsilon\right], \qquad k \ge N,$$

and define T to be the least $n \ge N$ for which $X_n \ne p_R(\nu_n, W_{1,n-1})$, with $T = \infty$

if there is no such n. Then

$$\begin{aligned}\mu(G_N \cap H) &= \sum_{n=N}^{\infty} \mu(G_N, T = n)\\ &\le E\left(\sum_{n=N}^{\infty} \mu(G_N^n, T = n \mid \nu_n, W_{1,n-1})\right)\\ &= E\left(\sum_{n=N}^{\infty} R_n 1_{G_N^n}\right)\\ &= E\left(\sum_{n=N}^{\infty} R_n 1_{G_N}\right) + E\left(\sum_{n=N}^{\infty} R_n \sum_{k=n}^{\infty} 1_{G_N^k \setminus G_N^{k+1}}\right)\\ &\le \varepsilon + E\left(\sum_{k=N}^{\infty}\left(\sum_{n=N}^{k} R_n\right) 1_{G_N^k \setminus G_N^{k+1}}\right)\\ &\le \varepsilon + \varepsilon E\left(\sum_{k=N}^{\infty} 1_{G_N^k \setminus G_N^{k+1}}\right)\\ &\le 2\varepsilon,\end{aligned}$$

so that $\mu(G \cap H) \le 3\varepsilon$. Since ε is arbitrary this means $\mu(G \cap H) = 0$.

Now suppose ρ is a stationary measure and y a doubly infinite word with $(\rho, y) \in G \setminus H$ and $\rho_n(y_1 \cdots y_n) > 0$ for all n. Then for $n \ge$ some N we have $p_R(\rho_n, y_1 \cdots y_{n-1}) = y_n$. Therefore, $\mu(X = y \mid W_{1,N-1} = y_1 \cdots y_{N-1}) > 0$ since $(\rho, y) \in G$. It follows that $\mu(X = y) > 0$, which is possible only for periodic y. □

The next proposition will show that for independent uniform selection, a result somewhat similar to the summability condition in Lemma 3.7 holds a.s. The similarity is enhanced by the observation that the random variables L_n, R_n are identically distributed, in fact exchangable. On the other hand, the similarity is reduced by the fact that there are stationary measures ρ for which

$$\sum_{n=1}^{\infty} \min(\gamma_L(\rho_n, W_{1,n-1}), \gamma_R(\rho_n, W_{1,n-1})) < \infty \quad \rho\text{-a.s.},$$

but

$$\sum_{n=1}^{\infty} \gamma_R(\rho_n, W_{1,n-1}) = \infty \quad \rho\text{-a.s.}$$

We will not give the details here, but in particular this is true for the process with two-letter alphabet obtained by dividing the unit circle in half and iterating an irrational rotation, for certain irrationals.

We will need some definitions: We say a stationary measure ρ is *nondegenerate* if $\rho(w) > 0$ for every finite word w. Given a stationary measure ρ and $w_{n-1} \in A^{n-1}$, let

$$\beta(\rho_{n+1}, w_{n-1}) := \rho_{n+1}\big(X_0 \neq p_L(\rho_n, w_{n-1}),\ X_n \neq p_R(\rho_n, w_{n-1}) | W_{1,n-1} = w_{n-1}\big).$$

The possible values of

$$\beta_{n+1} := \beta(\nu_{n+1}, W_{1,n-1})$$

form the interval $[\beta_{n,\min}, \beta_{n,\max}]$ where

$$\beta_{n,\min} := \max[L_n + R_n - 1, 0], \qquad \beta_{n,\max} := \min[L_n, R_n]$$

[cf. (1.1)]. These endpoints are unequal provided ν is nondegenerate, in which case we define

$$U_{n+1} := (\beta_{n+1} - \beta_{n,\min})/(\beta_{n,\max} - \beta_{n,\min}) \in [0,1].$$

Note that U_{n+1} is a function of ν_{n+1} and $W_{1,n-1}$. If we think of $W_{1,n-1}$ as a portion of some periodic word S, then this U_{n+1} is just the variable U defined (for that S) implicitly by Cases 1–4. Of course U_{n+1} here is not necessarily uniform.

Note that the main hypothesis (3.10) of Proposition 3.8, like that of Lemma 3.3 and like Theorem 3.2(ii), just says that the distribution of the double-error probability β_{n+1} has sufficient mass toward the upper end of the allowed interval $[\beta_{n,\min}, \beta_{n,\max}]$.

PROPOSITION 3.8. *Suppose μ is invariant under time reversal of ν, suppose ν is nondegenerate a.s. and suppose that for some $\delta > 0$,*

$$E\big((U_{n+1} - (d-1)/d)1_{[U_{n+1} \geq (d-1)/d]} | \nu_n, W_{1,n-1}\big) \geq \delta \quad \mu\text{-a.s. for all } n \geq 1, \tag{3.10}$$

where $d = |A|$. Then

$$\sum_{n=1}^{\infty} E \min(L_n, R_n) < \infty.$$

In particular, for independent uniform selection,

$$\sum_{n=1}^{\infty} \min(L_n, R_n) < \infty \quad \mu\text{-a.s.}$$

PROOF. It is enough to show that

$$EL_{n+1} \leq EL_n - (\delta/2) E \min(L_n, R_n). \tag{3.11}$$

We wish to define a variation of L_{n+1} in which the preferred letter is the same as for L_n. Thus let

$$\tilde{L}_{n+1} := \nu_{n+1}\big(X_0 \neq p_L(\nu_n, W_{1,n-1}) | W_{1,n}\big).$$

It is immediate from the definition of preferred letter that

$$L_{n+1} \le \tilde{L}_{n+1}.$$

Further,

$$E\big(\tilde{L}_{n+1} | \nu_n, W_{1,n-1}, U_{n+1}\big) = L_n, \tag{3.12}$$

so that (L_n, L_{n+1}) forms a supermartingale:

$$E\big(L_{n+1} | \nu_n, W_{1,n-1}\big) \le L_n. \tag{3.13}$$

The idea is to show that, if $L_n \ge R_n$ and we further condition on X_n being an error and on $U_{n+1} \ge (d-1)/d$, the inequality (3.13) is strict enough so that (3.11) holds. Note that the two further conditions are conditionally independent given $\nu_n, W_{1,n-1}$.

Suppose first that $L_n \ge R_n$ and $L_n + R_n > 1$. If $X_n = p_R(\nu_n, W_{1,n-1})$, then [cf. (2.2)]

$$\tilde{L}_{n+1} = (L_n - \beta_{n+1})/(1 - R_n) = (1 - R_n - U_{n+1}(1 - L_n))/(1 - R_n);$$

with (3.12) this implies that

$$\begin{aligned} &E\big(\tilde{L}_{n+1} | \nu_n, W_{1,n-1}, U_{n+1}, X_n \ne p_R(\nu_n, W_{1,n-1})\big) \\ &\qquad = (L_n + R_n - 1 + U_{n+1}(1 - L_n))/R_n \ge U_{n+1}. \end{aligned}$$

Suppose next that $L_n \ge R_n$ and $L_n + R_n \le 1$. Then similarly, if $X_n = p_R(\nu_n, W_{1,n-1})$, then

$$\tilde{L}_{n+1} = (L_n - \beta_{n+1})/(1 - R_n) = (L_n - U_{n+1}R_n)/(1 - R_n);$$

with (3.12) this implies that

$$E\big(\tilde{L}_{n+1} | \nu_n, W_{1,n-1}, U_{n+1}, X_n \ne p_R(\nu_n, W_{1,n-1})\big) = U_{n+1}.$$

Combining these cases shows that

$$\begin{aligned} L_n \ge R_n \Rightarrow\ &E\big(\tilde{L}_{n+1} | \nu_n, W_{1,n-1}, U_{n+1}, X_n \ne p_R(\nu_n, W_{1,n-1})\big) \\ &\ge U_{n+1}. \end{aligned} \tag{3.14}$$

From the definition of preferred letter it is clear that $L_{n+1} \le (d-1)/d$. Therefore, from (3.14), (3.10) and (3.12) we obtain that if $L_n \ge R_n$, then

$$\begin{aligned} E\big(L_{n+1} | \nu_n, W_{1,n-1}\big) \le\ &E\big(\tilde{L}_{n+1} | \nu_n, W_{1,n-1}\big) \\ &- E\Big(\big(\tilde{L}_{n+1} - (d-1)/d\big) \\ &\qquad \times 1_{[U_{n+1} \ge (d-1)/d]} 1_{[X_n \ne p_R(\nu_n, W_{1,n-1})]} | \nu_n, W_{1,n-1}\Big) \\ \le\ &L_n - \delta R_n. \end{aligned} \tag{3.15}$$

The assumed invariance under time reversal implies that L_n and R_n are exchangable, so (3.15) and (3.13) show that

$$E(L_{n+1}) \le E\big(L_n - \delta \min(L_n, R_n) 1_{[L_n \ge R_n]}\big)$$
$$\le E(L_n) - (\delta/2) E \min(L_n, R_n).$$

For independent uniform selection, (3.10) follows from Lemma 3.5, analogously to the proof of Theorem 1.1. □

4. Zero entropy. The statement that a stationary process X with law ν has zero entropy is equivalent to the statement that "the past (or the future) determines the present," or more precisely to the statement that

$$\gamma_L(\nu_n, W_{1,n-1}) \to 0 \quad \nu\text{-a.s.},$$

so that X has zero entropy μ-a.s. provided $L_n \to 0$ μ-a.s. With this in mind, the proof of the following Theorem 4.1, which by Proposition 3.8 implies Theorem 1.2, becomes easy.

THEOREM 4.1. *Suppose μ is a construction law with the property that*

$$\sum_{n=1}^{\infty} \min(L_n, R_n) < \infty \quad \mu\text{-}a.s. \tag{3.16}$$

Then $\mu(X$ has zero entropy$) = 1$.

PROOF. Since $\{L_n\}$ is a nonnegative supermartingale [see (3.13)], we know that $L_n \to$ (some $L \ge 0$) μ-a.s. By (3.16), on the event $[L > 0]$ we have

$$\sum_{n=1}^{\infty} R_n < \infty \quad \mu\text{-a.s.}$$

Thus by Lemma 3.7,

$$X \text{ is periodic } \mu\text{-a.s. on the event } [L > 0].$$

If $\nu(X) > 0$ for some point (ν, X) of the probability space, necessarily with X periodic, then certainly $L_n \to 0$ at that point. Therefore, on the event $[L > 0]$ we have X periodic and $\nu(X) = 0$, both μ-a.s. As there are only countably many periodic words, this shows that $\mu(L > 0) = 0$. □

5. Simulations. We have simulated the construction of a realization X of the random stationary process with independent uniform selection and alphabet $A = \{0, 1\}$. Two hundred trials were executed. In each trial $X_1, \dots, X_k$ were constructed, with $k = 64$ in most trials; some trials were extended to larger k. Each trial used a separate independent realization of the random law ν. These were the results:

Apparent period	Trials
1	126
2	18
3	5
4	3
5	1
6	1
13	1
14	1
17	1
aperiodic	43

The behavior of the sequences L_n and R_n of error probabilities is relevant to the question of periodicity, as discussed in Section 3. In the apparently aperiodic trials, the sequence R_n for n roughly 40 to 120 is extremely irregular, with many small values (often 10^{-10} or less) interspersed in some trials with occasional much larger values (0.1 or more). In contrast, the sequence L_n, which is a supermartingale, usually seems to decrease relatively regularly, exponentially rapidly toward 0, even in most of the apparently aperiodic trials. But in a few of the apparently aperiodic trials, L_n decreases extremely slowly, remaining above 0.1 even for n near 100. Summability of EL_n would imply periodicity a.s., but it is possible that a small probability of slow decrease allows $\sum_{n=1}^{\infty} EL_n = \sum_{n=1}^{\infty} ER_n = \infty$, in spite of the first conclusion of Proposition 3.8. This would be compatible with a positive probability for aperiodic words. For now, such questions remain open.

REFERENCES

[1] NICHOLLS, D. F. and QUINN, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. *Lecture Notes in Statist.* **11**. Springer, New York.

[2] SHIELDS, P. (1973). *The Theory of Bernoulli Shifts*. Univ. Chicago Press.

DEPARTMENT OF MATHEMATICS DRB 155
UNIVERSITY OF SOUTHERN CALIFORNIA
1042 WEST 36TH PLACE
LOS ANGELES, CALIFORNIA 90089-1113