The Annals of Probability
1983, Vol. 11, No. 3, 678–684

STATIONARITY ON FINITE STRINGS AND SHIFT REGISTER SEQUENCES[1]

BY ARIF ZAMAN

Florida State University

Stationarity is a property of infinite sequences of random variables. An appropriate extension of this definition is made, to cover finite sequences. The set of finite stationary sequences is shown to be a convex set and its extreme points are related to shift register sequences (which are paths on a graph known as the shift net, or the de Bruijn graph). The set of finite stationary sequences as defined here is simply the set of finite dimensional projections of infinite stationary sequences.

1. Introduction. A definition of stationarity for finite length sequences of random variables is proposed in Section 2. A finite sequence will be called "stationary" if the joint probability of every consecutive block of elements is shift invariant. It is shown that the set of finite stationary distributions is exactly the set of finite dimensional projections of infinite stationary distributions. With this equivalence, some of the material here is simply a careful treatment, in a different language, of material that already exists or is part of the "folklore" of probability.

Using the concept of finite stationarity, one can define m-stationarity, analogously to m-dependence. A distribution on an infinite sequence will be called m-stationary iff the marginal distributions of every m consecutive elements is stationary in the finite sense. This is a larger class than the set of stationary distributions, but shrinks to the set of stationary distributions as m approaches infinity.

Section 3 is a brief review of a problem from combinatorial theory. Considering shift registers on computers (even though the problem predates computers by a half a century), one is led to a particular directed graph variously known as the de Bruijn graph, the shift net, and a host of other names. The counting and generation of closed paths on this graph have been open problems in the combinatorial literature for some time.

Section 4 shows that the set of finite stationary distributions forms a convex set. The extreme points of this set are in a 1–1 correspondence with prime cycles (closed paths which have no closed subpaths) on shift nets. This explicit representation of the extreme points is the central result. In a handwritten manuscript made a short time before his untimely death at age 34, Walter Weissblum (1966) had identified these extreme points. The manuscript contains further thoughts on finite stationarity, as well as proofs. The graph theoretic identification and the proofs here are independently derived. An application of this extreme point representation is found in Zaman (1983).

The final section examines the similarities and differences between the finite and the infinite cases. The ergodic theory results from the infinite case do not seem to be applicable to the finite case. The extreme point representation of the finite case does not seem to carry over to the infinite situation. Yet proofs of some results in the case of ordinary stationarity are still possible as extensions of the finite stationary results.

2. Stationarity in finite sequences. Let c^n denote the set of all sequences of length n, taking values in the set $\{0, 1, \cdots, c-1\}$. Confusion with the number c^n can always be eliminated by context. Henceforth n is assumed finite unless explicitly mentioned otherwise. Let $S(c)$ denote the set of stationary distributions on c^∞. The distribution of a

Received December 1981; revised September 1982.

[1] This work is part of the Author's Ph.D. thesis done at Stanford under Professor Persi Diaconis.

AMS 1980 *subject classification.* Primary 60G10; secondary 05C38.
Key words and phrases. Shift invariance, extreme points, de Bruijn graphs, shift registers.

sequence $X \in c^\infty$ is in $S(c)$ if and only if $(X_i, X_{i+1}, \cdots) \sim (X_j, X_{j+1}, \cdots)$ for every i and j. The symbol "$\sim$" denotes "equal in distribution."

For finite sequences, let $S^n(c)$ denote all distributions of $X \in c^n$ for which $(X_i, \cdots, X_{i+k}) \sim (X_j, \cdots, X_{j+k})$ for all i, j, k for which the equation is defined (i.e., for $1 \le i \le j \le n - k$). The distributions in $S^n(c)$ will be called the "stationary distributions on c^n" or simply stationary.

For a sequence $X \in c^n$ with distribution P, the symbol P^m for $m \le n$, will denote the marginal distribution of $(X_1, \cdots, X_m) \in c^m$. The relationship between finite and infinite stationarity is given by

THEOREM 1. $P \in S(c) \Leftrightarrow \forall n\ P^n \in S^n(c)$.

PROOF. The forward implication is true by definition. For the reverse, a measure on an infinite sequence is determined by its values on the cylinder sets. □

This theorem shows the desirable result that $S^\infty(c) = S(c)$, and hence a uniform definition of stationarity is possible. In the final section, it will be further shown that any $P \in S^n(c)$ can be extended to a probability $Q \in S(c)$ for which $P = Q^n$. So $S^n(c)$ is simply all the projections (marginals) of finite stationary distributions, or equivalently $S^n(c)$ is all the distributions which can be extended to infinite stationary distributions.

The definition of $S^n(c)$ can be immediately reduced to the following equivalents.

THEOREM 2. *The following statements are equivalent*:

(i) $P \in S^n(c)$

(ii) $(X_1, \cdots, X_{n-1}) \sim (X_2, \cdots, X_n)$

(iii) *For all* $a \in c^{n-1}$

$$\sum_{i=0}^{c-1} P\{X = (a_1, \cdots, a_{n-1}, i)\} = \sum_{i=0}^{c-1} P\{X = (i, a_1, \cdots, a_{n-1})\}.$$

PROOF. By the definition of $S^n(c)$, (i) $\Rightarrow$ (ii). Also, (ii) $\Leftrightarrow$ (iii) is true because (iii) is just (ii) written out explicitly. To show (ii) $\Rightarrow$ (i), assume $(X_1, \cdots, X_{n-1}) \sim (X_2, \cdots, X_n)$. Then for any $k < n$, by successive shifts,

$$(X_1, \cdots, X_k) \sim (X_2, \cdots, X_{k+1}) \sim (X_3, \cdots, X_{k+2}) \cdots \sim (X_{n-k+1}, \cdots, X_n).$$

So X has a stationary distribution. □

An m-stationary measure requires that the distribution of each $(X_{i+1}, \cdots, X_{i+m})$ be in $S^m(c)$. By Theorem 2, this is true iff

$$(X_{i+1}, \cdots, X_{i+m-1}) \sim (X_{j+1}, \cdots, X_{j+m-1})$$

for every i and j. Thus every sequence is 1-stationary, and identically distributed, though possibly dependent sequences are 2-stationary. In general, m-stationarity guarantees identical distribution of the joint $m - 1$ dimensional distributions of consecutive random variables.

3. The shift net T_{c^n}. Define the shift relationship $x \to y$, for points x, y in the set c^n by

$$x \to y \text{ iff } (x_2, \cdots, x_n) = (y_1, \cdots, y_{n-1}).$$

Define the directed graph T_{c^n} with c^n as the set of vertices and with arcs from x to y if and only if $x \to y$. The notation here follows that used by Lempel (1971) where T_{c^n} is called the shift net. Counting the number of cycles on this graph, as well as algorithms for generating cycles, have been open problems in the literature for a long time. An excellent recent survey is given by Fredricksen (1981), and a detailed discussion of shift register sequences can be found in the book by Golomb (1967).

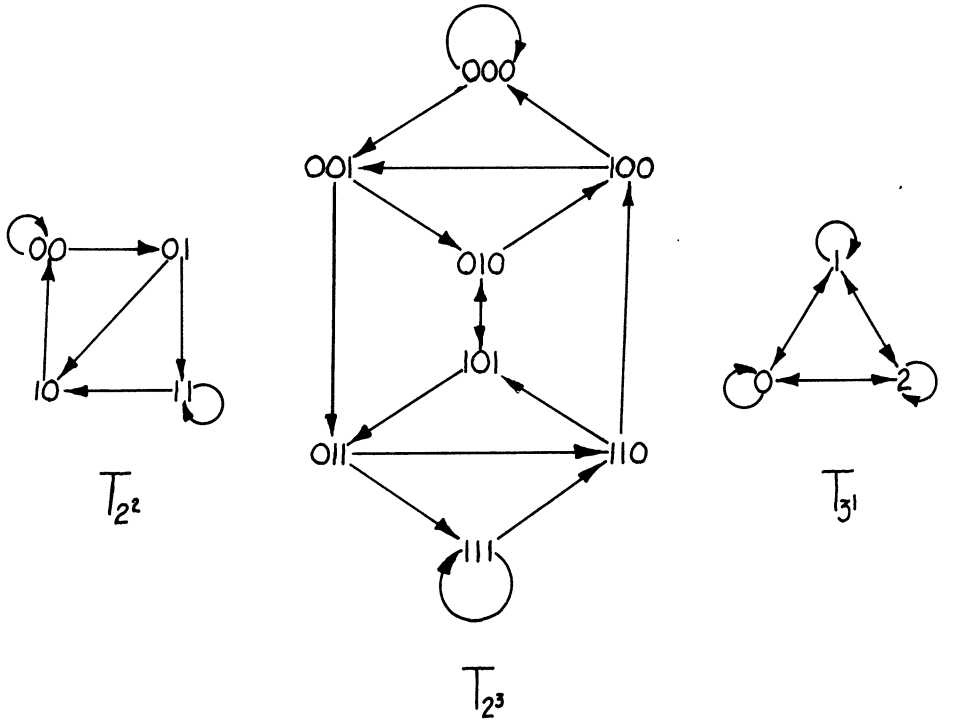

FIG. 1. *Examples of shift nets.*

A path on T_{c^n} is a sequence of points $a_i \in c^n$ such that $a_i \to a_{i+1}$. A loop is a finite path $(a_1, \cdots, a_e)$ for which $a_e \to a_1$. For loops we will identify a_{e+1} with a_1, a_{e+2} with a_2, and so on so that we may write $a_i \to a_{i+1}$ for $i = 1, \cdots, e$. A cycle is a loop with no repetitions, i.e. $a_i = a_j \Leftrightarrow i = j$.

A loop is called divisible if it contains a proper subset which is a loop. The term divisible is justified, because the remainder after removing the subset will also be a loop. For example, the cycle (001, 011, 110, 101, 010, 100) has subcycles (001, 011, 110, 100) and (101, 010), as well as (001, 010, 100) and (011, 110, 101).

A loop which is not divisible is called prime. It is clear that every loop has a prime factorization, even though, as the previous example demonstrated, the factorization is not unique. Clearly only cycles can be prime, because if for a loop, $a_i = a_j$ then $(a_i, \cdots, a_{j-1})$ is a proper subloop.

The following lemmas in this section are well known results, but proofs are given here because their proofs are simple and their ideas useful.

LEMMA 3. *A cycle on T_{c^n} is prime* iff $a_i \to a_j \Leftrightarrow j = i + 1$.

PROOF. For the forward implication, let a be a cycle in T_{c^n} and assume $a_i \to a_j$ for some $i + 1 \neq j$. If $i + 1 < j$ then $(a_j, \cdots, a_e, a_1, \cdots, a_i)$ is a subcycle; otherwise $(a_j, \cdots, a_i)$ is a subcycle, hence a is divisible.

For the reverse implication, assume that $a_i \to a_j \Leftrightarrow j = i + 1$, and that a has a proper subcycle $a' = (a'_1, \cdots, a'_{e'})$. Then for some i, $a'_1 = a_i \to a'_2$. But a'_2 is a vertex in a, so by hypothesis $a'_2 = a_{i+2} \to a'_3$. Continuing this by induction, a' must contain all the vertices of a, and hence is not a proper subset. □

Clearly there is a large amount of redundancy in the loop representation. If $(a_1, \cdots, a_e)$

is a loop in T_{c^n}, then the first elements $(a_{1,1}, \cdots, a_{e,1})$ form a loop in T_{c^1}, and this mapping is invertible. For example the loop (001, 011, 110, 101, 010, 100) maps to (0, 0, 1, 1, 0, 1). As this is true for all T_{c^n}, it establishes a 1–1 correspondence between loops on T_{c^n} and loops on T_{c^m} for all m and n. Explicitly $f_{n,m}$: loops on $T_{c^n} \to$ loops on T_{c^m} is given by

$$f_{n,m}(a_1, \cdots, a_e) = (a'_1, \cdots, a'_e) \text{ where } a'_i = (a_{i,1}, a_{i+1,1}, \cdots, a_{i+m-1,1}) \in c^m.$$

LEMMA 4. *$f_{n,n+1}$ establishes a 1–1 correspondence between the cycles of T_{c^n} and the prime cycles of $T_{c^{n+1}}$.*

PROOF. For x, y in c^n, and $x \to y$ define $\langle x, y\rangle \in c^{n+1}$ *by*

$$\langle x, y\rangle = (x_1, x_2, \cdots, x_n, y_n) = (x_1, y_1, \cdots, y_{n-1}, y_n).$$

For a loop a in T_{c^n}, clearly

$$f_{n,n+1}(a) = (\langle a_1, a_2\rangle, \langle a_2, a_3\rangle, \cdots, \langle a_e, a_1\rangle).$$

Using the previous lemma:

$$\begin{aligned} f_{n,n+1}(a) \text{ is a prime cycle iff } & \langle a_i, a_{i+1}\rangle \to \langle a_j, a_{j+1}\rangle \Leftrightarrow i+1 = j \\ \text{iff } & a_{i+1} = a_j \Leftrightarrow i+1 = j \\ \text{iff } & a \text{ is a cycle.} \square \end{aligned}$$

4. Loop measures in $S^n(c)$. With each loop a on T_{c^n} associate the loop measure P_a on c^n which chooses a vertex a_i at random. Specifically, if a has length e, and I is uniformly distributed on $\{1, 2, \cdots, e\}$, then P_a is the distribution of a_I.

LEMMA 5. *If a is a loop then P_a is stationary.*

PROOF. By the identification of a_1 and a_{e+1}, we have $a_I \sim a_{I+1}$. Since $a_I \to a_{I+1}$, $(X_2, \cdots, X_n) \sim (a_{I,2}, \cdots, a_{I,n}) = (a_{I+1,1}, \cdots, a_{I+1,n-1}) \sim (X_1, \cdots, X_{n-1})$, so by Theorem 2, $P_a \in S^n(c)$. $\square$

When the loop a is a prime cycle, the measure P_a is called a prime cycle measure. Since the constraints for stationarity given in Theorem 2 iii are linear, any probability measure which is a linear combination of measures in $S^n(c)$ is also in $S^n(c)$. This shows that $S^n(c)$ is a convex set.

THEOREM 6. *The set of extreme points of $S^n(c)$ is the set of all prime cycle measures.*

PROOF. As an induction hypothesis, assume all measures in $S^n(c)$ with less than k support points are mixtures of prime cycles measures. To start the induction, it is enough that the hypothesis is vacuously true when $k = 1$.

Let $P \in S^n(c)$ have k points in its support. By Theorem 2, for any $x \in \text{supp}(P)$ we have

$$0 < P\{x\} \le \sum_{i=0}^{c-1} P\{(i, x_2, \cdots, x_n)\} = \sum_{i=0}^{c-1} P\{(x_2, \cdots, x_n, i)\} = \sum_{x \to y} P\{y\}.$$

This shows that for every $x \in \text{supp}(P)$, there exists a $y \in \text{supp}(P)$ for which $x \to y$. Because supp(P) is a finite set, this implies the existence of a loop and hence a prime cycle in supp(P).

Let a be one such cycle, so that $a_i \in \text{supp}(P)$ for $i = 1, 2, \cdots, e$. Define

$$\alpha = \min_i eP\{a_i\}.$$

By definition $0 < \alpha \le 1$. If $\alpha = 1$ then $P = P_a$ and the induction is over. If $\alpha < 1$, define the measure

$$Q = \frac{P - \alpha P_a}{1 - \alpha}.$$

One can verify that Q is a probability. Q has at least one less than the k support points of P, because

$$(1-\alpha)Q\{a_i\} = P\{a_i\} - \alpha P_a\{a_i\} = P\{a_i\} - \min_i eP\{a_i\}\frac{1}{e}.$$

Finally, $Q \in S^n(c)$ because Q is a linear combination of the stationary measures P and P_a.

By the induction hypothesis, Q is a mixture of prime cycle measures, and

$$P = \alpha P_a + (1-\alpha)Q.$$

This shows that any $P \in S^n(c)$ is a mixture of prime cycle measures. It only remains to show that all prime cycle measures are extreme points.

Let P_a be a prime cycle measure. If P_a is a mixture of stationary measures, by what has just been shown, P_a is a mixture of prime cycle measures. For any P_b in that mixture, $\text{supp}(P_b) \subset \text{supp}(P_a)$. But a is a prime cycle, so the only cycle contained in it is itself, and hence $b = a$. This shows that P_a is an extreme point. □

THEOREM 7. *Loop measures are dense in* $S^n(c)$.

PROOF. Let $P \in S^n(c)$. We will find a sequence of loops $b(j)$, for which $P_{b(j)} \to P$ in distribution as $j \to \infty$. By Theorem 6, P is a mixture of prime cycle measures, i.e.

$$P = \textstyle\sum_{i=1}^{k} w_i P_{a(i)}$$

for prime cycles $a(i)$ and positive weights w_i summing to one.

On the graph T_{c^n}, any point $x \in c^n$ can traverse to any other point $y \in c^n$ in a path of at most length n by

$$x \to (x_2, \cdots, x_n, y_1) \to (x_3, \cdots, x_n, y_1, y_2) \to \cdots \to (x_n, y_1, \cdots, y_{n-1}) \to y.$$

Construct the loop $b(j)$ as follows. Start at the point $a(1)_1$. Take $[jw_1]$ steps in the cycle $a(1)$ (since $b(j)$ is a loop, it is allowed to go back over points). Now take the shortest path to the point $a(2)_1$, and take $[jw_2]$ steps in the cycle $a(2)$. Continue like this for $a(3), \cdots, a(k)$. Finally, after taking $[jw_k]$ steps in $a(k)$, take a path to $a(1)_1$ to complete the loop.

As $j \to \infty$ consider the measures $P_{b(j)}$. Firstly, at the most, $k(n-1)$ points were used for paths connecting the end of one cycle to the start of the next. Since this does not increase with j, these points may be ignored. Asymptotically, the measure will pick a point from the block $a(i)_1, \cdots, a(i)_{[jw_i]}$ with probability w_i. As j approaches infinity, this will tend to a uniform choice from the loop $a(i)$ with probability w_i.

So as $j \to \infty$

$$P_{b(j)} \to \textstyle\sum_{i=1}^{k} w_i P_{a(i)} = P. \square$$

5. Ergodic theory and infinite stationarity. The starting point of ergodic theory is a shift operator τ: $c^\infty \to c^\infty$, defined by $\tau(x) = (x_2, x_3, \cdots)$. A measure is stationary iff $P(A) = P(\tau(A))$ for all $A \subset c^\infty$. For such measures the ergodic theorem (e.g. Phelps, 1966) states that $S(c)$ is a simplex with extreme points given by the ergodic measures, i.e. stationary measures which assign either zero or one as the probability of shift invariant sets.

Unfortunately it seems difficult to apply this machinery in the finite case. There is no appropriate definition of $\tau(x)$ for $x \in c^n$. A parallel definition for the shift can be defined for sets in a restricted class as follows. If a set $A = c^1 \times B$ for some $B \subset c^{n-1}$, then $\tau(A) = B \times c^1$. With this definition, a measure is in $S^n(c)$ iff $P(A) = P(\tau(A))$ for sets A on which τ is defined (see Theorem 2).

Because of the difference between the finite and infinite cases, the ergodic theorem is actually false for $S^n(c)$. The set of measures $S^n(c)$ is not a simplex, and so the decomposition of a measure as a mixture of the extremal measures given in Theorem 6 is not unique. This is evident from the fact that loops do not have a unique prime factorization.

Consider the infinite graph T_{c^∞} on the set c^∞ determined by the relationship $x \to y$ iff $y = (x_2, x_3, \cdots)$. This is a different graph from the finite ones, because every $x \in c^\infty$ has a unique successor. A path on the graph is thus completely specified by its first vertex, and will loop only if that is a repeating sequence. A prime cycle on T_{c^∞} thus consists of starting from a repeating $x \in c^\infty$, and successively shifting until after one period it returns to x. All loops are simply further repetitions of prime cycles. This means that on T_{c^∞} the set of loop measures is identical to the set of prime cycle measures.

The definitions of the functions $f_{n,m}$ given before Lemma 4 can be extended to apply when n or m is infinite using these infinite graphs. It is clear that $f_{n,\infty}$ and $f_{\infty,n}$ are inverses of each other, and hence are 1–1.

THEOREM 8. *$Q \in S^n(c) \Rightarrow \exists\, P \in S(c)$ such that $P^n = Q$.*

PROOF. For a prime cycle, a, on T_{c^n} the loop $f_{n,\infty}(a)$ on T_{c^∞} is given by

$$a = ((a_{1,1}, \cdots, a_{n,1}), \cdots, (a_{e,1}, \cdots, a_{e+n-1,1}))$$

$$f_{n,\infty}(a) = ((a_{1,1}, a_{2,1}, \cdots), \cdots, (a_{e,1}, a_{e,2}, \cdots))$$

so that $P_a = P^n_{f_{n,\infty}}(a)$. By Theorem 6, any $Q \in S^n(c)$ can be represented as

$$Q = \sum w_a P_a$$

where the sum ranges over prime cycles a in T_{c^n} and w_a are mixing weights. The choice

$$P = \sum w_a P_{f_{n,\infty}}(a)$$

satisfies the conditions of the theorem.

Because the decomposition of Q is not unique, there may be several possible extensions P of the same measure Q. Hobby and Ylvisaker (1964) consider properties of such extensions. □

The differences between the finite and the infinite graphs results in the following modifications to the theorems in Section 4.

THEOREM 9. *Prime cycles measures on T_{c^∞}*
(i) *are extreme points of $S(c)$*
(ii) *are not all of the extreme points of $S(c)$*
(iii) *are dense in $S(c)$.*

PROOF. (i) Let P_a be a prime cycle measure. For any measure $Q \in S(c)$, if $\text{supp}(Q) \subset \text{supp}(P_a)$ then some $a_i \in \text{supp}(Q)$. But because $Q \in S(c)$

$$Q\{a_i\} = Q\{a_{i+1}\} = \cdots = Q\{a_e\} = Q\{a_1\} = \cdots = Q\{a_{i-1}\}.$$

So Q is the uniform measure supported on the vertices of the loop a, i.e. $Q = p$.

(ii) A simple counter example is given by an iid sequence on c^∞, generated by c sided dice. It is clearly in $S(c)$, yet almost surely it is non-repeating. Any loop measure, or mixture of loop measures will almost surely repeat.

(iii) Let X be a random sequence with distribution $P \in S(c)$. Then $P^n \in S^n(c)$ so by Theorem 7, there is some loop $a(n)$ on T_{c^n} for which

$$\| P_{a(n)} - P^n \| < 1/n$$

where the norm is the variation norm. As in the proof of the previous theorem $P^k_{f_{n,\infty}(a(n))} = P^k_{a(n)}$ for $k \le n$. So for all k

$$P^k_{f_{n,\infty}(a(n))} \to P^k \text{ as } n \to \infty.$$

By Billingsley (1968, page 19) this is enough to show that

$$P_{f_{n,\infty}(a(n))} \to P \text{ as } n \to \infty.$$

As this is true for all measures $P \in S(c)$, the measures $P_{f_{n,\infty}(a(n))}$, i.e. the prime cycle measures are dense in $S(c)$. □

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

DE BRUIJN, N. G. (1946). A combinatorial problem. *Nederl. Akad. Wetensch. Proc. Indag. Math.* **49** 758–764.

FREDRICKSEN, H. M. (1982). A survey of full length non-linear shift register cycle algorithms. *SIAM Rev.* **24** 195–221.

GOLOMB, S. W. (1967). *Shift Register Sequences.* Holden-Day, San Francisco.

HOBBY, C. and YLVISAKER, D. (1964). Some structure theorems for stationary probability measures on finite state sequences. *Ann. Math. Statist.* **35** 550–556.

LEMPEL, A. (1971). *m*-ary closed sequences. *J. Comb. Theory Ser. A* **10** 253–258.

PHELPS, R. (1966). *Lectures on Choquets Theorem.* Van Nostrand, New York-Toronto.

WEISSBLUM, W. (1966). Unpublished notes on finite stationary sequences.

ZAMAN, A. (1983). A non-clustering property of stationary sequences. Technical Report, Florida State University.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306