

AD-A129 628

A NON-CLUSTERING PROPERTY OF STATIONARY SEQUENCES(U)
FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS
A ZAMAN MAY 83 ARO-19367.5-MA DAAG29-82-K-0168

1/1

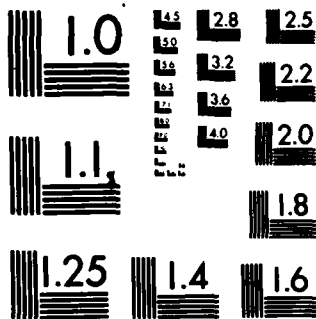
UNCLASSIFIED

F/G 12/1

NL



END
DATE
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ARO 19367.5-MA

12

ADA 1 29628

A NON-CLUSTERING PROPERTY OF STATIONARY SEQUENCES

by

Arif Zaman

FSU Statistics Report No. M655
USARO Technical Report No. D-60

May, 1983

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

DTIC
ELECTE
JUN 22 1983

This document has been approved
for public release and sale; its
distribution is unlimited.

AMS 1980 Classification: Primary 60G10, 26D15

Key Words and Phrases: Clustering, Stationary Sequences, Cyclic Sums

Research supported by the U.S. Army Research Office Grant DAAG 29-82-K-0168.
The U.S. Government is authorized to reproduce and distribute reprints for
Governmental purposes notwithstanding any copyright notation thereon.

83 06 22 004

DTIC FILE COPY

A Non-clustering Property of Stationary Sequences

by

Arif Zaman
Florida State University

Revision For THIS GRA&I THIS TAB THIS ISSUE Distribution	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Distribution Availability Codes *1 and/or Special	A



Abstract

For a random sequence of events, with indicator variables X_i , the behavior of the expectation $E\left(\frac{X_k + \dots + X_{k+m-1}}{X_1 + \dots + X_n}\right)$ for $1 \leq k \leq k+m-1 \leq n$ can be taken as a measure of clustering of the events. When the measure on the X 's is i.i.d., or even exchangeable, a symmetry argument shows that the expectation can be no more than m/n . When the X 's are constrained only to be a stationary sequence, the bound deteriorates, and depends on k as well. When m/n is small, the bound is roughly $2m/n$ for k near $n/2$ and is like $(m/n) \log n$ for k near 1 or n . The proof given is partly constructive, so these bounds are nearly achieved, even though there is room for improvement for other values of k .

AMS 1980 Classification: Primary 60G10, 26D15

Key Words and Phrases: Clustering, Stationary Sequences, Cyclic Sums

1. Introduction.

In considering portions of larger, but still finite strings of random variables, the following problem arose. If X_1, \dots, X_n is part of a stationary sequence of zeros and ones, one would not expect the ones within that portion to clump together, intuitively because each X_i is as likely as any other to have the value one. Based on that intuitive argument, one could expect the expression $\sup_{P \in S} E_P \left\{ \frac{X_k + \dots + X_{k+m-1}}{X_1 + \dots + X_n} \right\}$ (note: $0/0 = 0$) where $1 \leq k \leq k+m-1 \leq n$, and S is the set of stationary probability measures on binary sequences, to behave roughly like m/n . Indeed, if the probability P is restricted to be i.i.d. or even exchangeable, a simple symmetry argument yields a supremum of m/n , achieved when the X_i are identically 1. For the case of stationarity, the upper bounds on the supremum for m/n small are roughly $2m/n$ when k is near $n/2$, and like $(m/n) \log n$ for k closer to 1 or n (thm. 7). The key result is a constructive proof that finds the P which achieves the supremum for the two cases of $m = 1, k = 1$, and $m = 1, k = (n+1)/2$ (thm. 2).

I would like to thank Professor Michael Steele for insisting that this could be done, and Professor Larry Shepp for an improvement in the proof. I would also like to acknowledge the many simplifications and improvements suggested by the referee.

2. Results.

We shall immediately narrow our concern to the simpler problem of finding bounds for

$$R_{k,n} = \sup_{P \in S} E_P \left\{ \frac{X_k}{X_1 + \dots + X_n} \right\} \quad \text{for } 1 \leq k \leq n. \quad (1)$$

Notice that the variables X_{n+1}, X_{n+2}, \dots do not appear in the above expression, so only the marginal distribution of (X_1, \dots, X_n) affects the values of $R_{k,n}$. A small amount of notation is needed for the next theorem, which makes use of this observation.

A loop is a finite sequence a_1, \dots, a_m of zeros and ones. Subscripts out of range will be taken circularly, so that $a_0 = a_m$, and $a_{m+1} = a_1$. For a loop a and any positive integer n , the measure $P_{a,n}$ gives mass $1/m$ to each of $(a_1, \dots, a_n), (a_2, \dots, a_{n+1}), \dots, (a_m, \dots, a_{m+n-1})$.

Theorem 1.

If a binary sequence X has a stationary distribution, then the marginal distribution of (X_1, \dots, X_n) can be written as a convex combination of measures $P_{a,n}$ for $a \in A_n$, where A_n is a finite set of loops. Moreover, every $P_{a,n}$ is the marginal of some infinite stationary distribution.

More details, and a proof of this can be found in Zaman (1983) or Hobby and Ylvasaker (1964). Since expectation is a linear functional, thm. 1 allows replacing the maximization over S in eq. 1 by maximization over $P_{a,n}$ for $a \in A_n$, yielding

$$R_{k,n} = \max_{a \in A_n} E_{P_{a,n}} (X_k / \sum_{j=1}^n X_j). \quad (2)$$

Using the definition of $P_{a,n}$, the expectation can be further decomposed into

$$E_{P_{a,n}} (X_k / \sum_{j=1}^n X_j) = \frac{1}{m} \sum_{i=1}^m (a_{i+k} / \sum_{j=1}^n a_{i+j}) \quad (3)$$

where m is the length of the loop a . In a completely unrelated problem, sums of the form given in the right side of eq. 3 have been given the name cyclic sums, e.g. Daykin (1970).

Equations 2 and 3 convert the original probability problem of eq. 1 into a finite maximization of a function over a set of loops. This maximization is performed for chosen values of k in the appendix to prove the following key theorem.

Theorem 2.

(a) When $k = 1$ or n , the maximum in eq. 2 is achieved for $a = 0^{n-1}1^\beta$ (the notation 0^{n-1} refers to a block of $n-1$ zeros) for some number β depending on n .

(b) When $k = (n+1)/2$ for odd n , the maximum in eq. 2 is achieved for $a = 0^{k-1}1$.

Corollary 3.

Define

$$\alpha(n) = \sup_{\beta \geq 1} (n+\beta)^{-1} \sum_{i=1}^{\beta} 1/i. \quad (4)$$

Then,

$$R_{k,n} = \begin{cases} \alpha(n-1) & \text{if } k = 1 \text{ or } n & \text{(a)} \\ 2/(n+1) & \text{if } k = (n+1)/2 & \text{(b)} \end{cases}$$

The corollary is actually proved as a step in proving thm. 2, but can also be proved by writing out eq. 3 for the loops given in thm. 2.

Using these equalities for $R_{1,n}$ and $R_{(n+1)/2,n}$, a general bound for $R_{k,n}$ is easy to get. Theorems 4 and 5 do just that. The bounds of thm. 4 are depicted graphically in fig. 1.

Theorem 4.

Define

$$\alpha(k,n) = \sup_{n-k \leq \beta} (k+\beta)^{-1} [(n-k)/\beta + \sum_{i=n-k}^{\beta-1} 1/i].$$

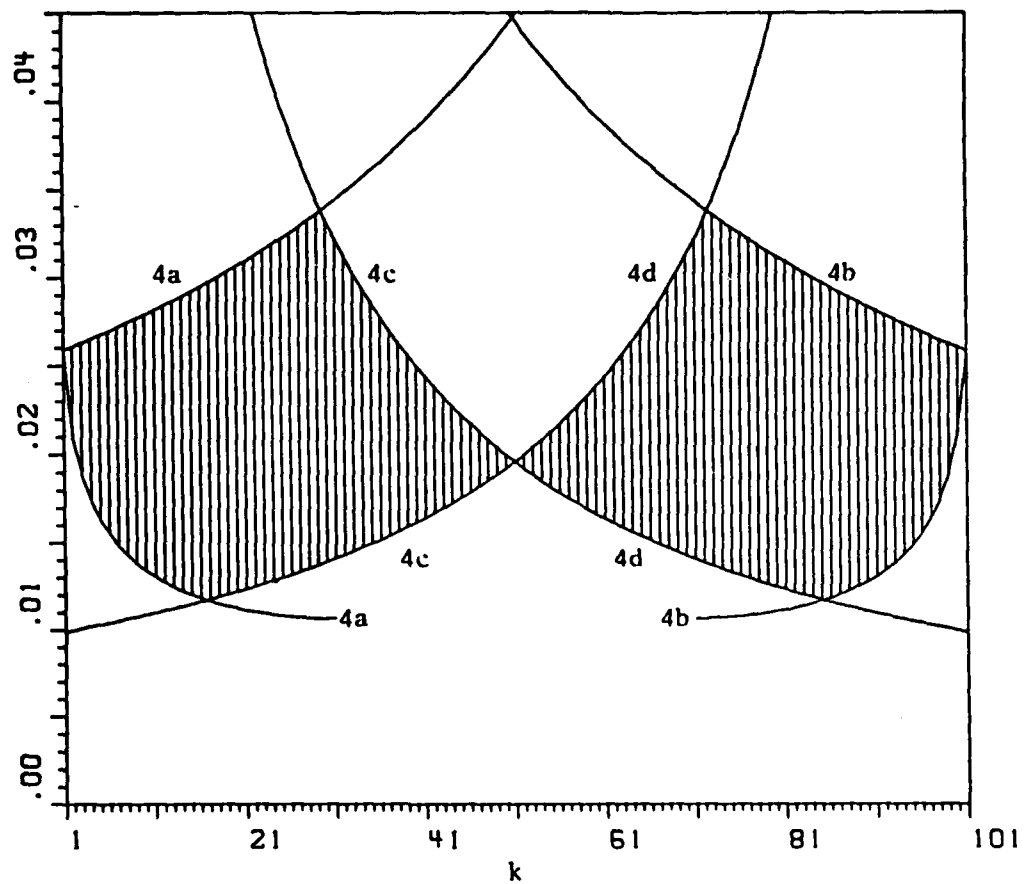


Fig. 1: Bounds on $R_{k,n}$ as a function of k , for $n = 101$.

The area between the upper and lower bounds of thm. 4 is shaded to indicate the possible region for $R_{k,n}$. The different bounds are labelled by the equation number in thm. 4.

Then

- (a) $\alpha(n-k, n) \leq R_{k,n} \leq \alpha(n-k)$ when $2k-1 \leq n$
 (b) $\alpha(k-1, n) \leq R_{k,n} \leq \alpha(k-1)$ when $2k-1 \geq n$
 (c) $1/(n+1-k) \leq R_{k,n} \leq 1/k$ when $2k-1 \leq n$
 (d) $1/k \leq R_{k,n} \leq 1/(n+1-k)$ when $2k-1 \geq n$.

Proof:

Parts (b) and (d) follow from (a) and (c) respectively, once the symmetry condition

$$R_{k,n} = R_{n-k+1,n} \quad (5)$$

is established. To prove this, note that if $P_{a,n}$ is the distribution of (X_1, \dots, X_n) then the distribution of (X_n, \dots, X_1) is given by $P_{a',n}$ for $a' = (a_n, \dots, a_1)$. Now for any loop a ,

$$E_{P_{a,n}} (X_k / \prod_{j=1}^n X_j) = E_{P_{a',n}} (X_{n+1-k} / \prod_{j=1}^n X_j)$$

from which eq. 5 follows.

The upper bound in (a) follows from Cor. 3a by

$$R_{k,n} \leq \sup_{P \in S} E_P (X_k / \prod_{j=k}^n X_j) = R_{1,n+1-k} = \alpha(n-k).$$

Similarly, for part (c), the result of Cor. 3b shows that for $2k-1 \leq n$

$$R_{k,n} \leq \sup_{P \in S} E_P (X_k / \prod_{j=1}^{2k-1} X_j) = R_{k,2k-1} = 1/k.$$

The lower bounds have been included in the theorem to get some idea on the room for improvement of these bounds. It is conjectured that the actual values of $R_{k,n}$ are much closer to the lower bounds than to the upper bounds. The lower bound (a) is obtained by using eq. 3 to get for $k \leq (n+1)/2$

$$R_{k,n} \geq \sup_{\substack{a=0^{n-k_1\beta} \\ k \leq \beta \leq n}} E_p (X_k / \sum_{j=1}^n X_j)$$

$$\geq \sup_{k \leq \beta \leq n} (n+\beta-k)^{-1} [(k-1)/\beta + \sum_{i=k}^{\beta} 1/i].$$

The lower bound in (c) is achieved by letting $a = 0^{n-k_1}$. For that value of a , if $2k-1 \leq n$ then by eq. 3

$$E_p (X_k / \sum_{j=1}^n X_j) = \frac{1}{n+1-k}.$$

It is not difficult to find loops which give even higher lower bounds, but that does not seem to be the more fruitful direction of moving the bounds. \square

Theorem 5.

$$R_{k,n} \leq \frac{1+\log(n-1)}{n} \quad \text{for } n \geq 3.$$

Before giving a proof, a logarithmic approximation for the function α will be established.

Lemma 6.

$$\frac{\log n - \log(\log n) - 1}{n} \leq \alpha(n) \leq \frac{\log n}{n} \quad \text{for } n \geq 3.$$

Proof:

Let β^* be a value of β which achieves the maximum in eq. 4, so that

$$\alpha(n) = (n+\beta^*)^{-1} \sum_{i=1}^{\beta^*} 1/i. \quad (6)$$

A crude bound to the harmonic series in eq. 6 gives

$$\alpha(n) \leq (1+\log \beta^*)/(n+\beta^*). \quad (7)$$

By calculus, the function $(1+\log x)/(n+x)$ for $x \geq 1$ reaches its maximum value of $(\log x^*)/n$ when $x^* \log x^* = n$. If $n > e$, $\log x^*$ can be bounded by

$$\log n - \log \log n \leq \log x^* \leq \log n. \quad (8)$$

Plugging this information about the maximum into eq. 7

$$\alpha(n) \leq (1+\log \beta^*)/(n+\beta^*) \leq (\log x^*)/n \leq (\log n)/n,$$

establishing the second inequality of the lemma.

For the first inequality, let x^* be as before, define $\beta = [x^*]$ (the integer part), and for notational convenience let $\ell = \log n - \log \log n$ which is the term on the left side of eq. 8. Then

$$\begin{aligned} \alpha(n) &\geq (n+\beta)^{-1} \sum_{i=1}^{\beta} 1/i \geq (n+x^*)^{-1} \log x^* \\ &\geq (n+n/\ell)^{-1} \ell = n^{-1} \ell^2 / (1+\ell) = n^{-1} [\ell - 1 + (\ell+1)^{-1}] \\ &> (\ell-1)/n. \end{aligned} \quad (9)$$

The last inequality substitutes a prettier expression at the cost of some precision. \square

The proof of thm. 5 then amounts to the following. By eq. 5

$$\begin{aligned} \max_k R_{k,n} &= \max_{k \leq (n+1)/2} R_{k,n} \\ \text{(by thm. 4a, c)} &\leq \max_{k \leq (n+1)/2} \{(1/k) \wedge \alpha(n-k)\} \\ \text{(by lem. 6)} &\leq \max_{k \leq (n+1)/2} \{(1/k) \wedge \log(n-k)/(n-k)\} \end{aligned} \quad (10)$$

Since $1/k$ is decreasing and the second function increasing as k increases, the maximum in eq. 10 is attained at some $k = k^*$ for which the two functions are equal. Thus

$$\begin{aligned} \max_k R_{k,n} &= 1/k^* = \log(n-k^*)/(n-k^*) \\ &= [1 + \log(n-k^*)]/n, \end{aligned}$$

where the last expression follows by some algebra. Since $k^* \geq 1$, replacing it by 1 gives the claimed result in thm. 5. \square

Returning to the original problem as stated in the introduction, one can state the following theorem based only on the definition of $R_{k,n}$.

Theorem 7.

$$\sup_{P \in S} E_P \left\{ \frac{\sum_{j=k}^{k+m-1} X_j}{\sum_{j=1}^n X_j} \right\} \leq \sum_{j=k}^{k+m-1} R_{j,n}.$$

For example, this proves that for any stationary measure P ,

$$E_P \left\{ \frac{X_k + \dots + X_{k+m-1}}{X_1 + \dots + X_n} \right\} \leq (m/n) [1 + \log(n-1)]$$

and for blocks near the middle

$$E_P \left\{ \frac{X_{-k} + \dots + X_k}{X_{-n} + \dots + X_n} \right\} \leq \frac{1}{n+1} + 2 \log\left(\frac{n}{n-k}\right) \leq (2k+1)/(n-k)$$

by using the values of $R_{k,n}$ given in theorems 5 and 4c, d.

APPENDIX

Proof of Theorem 2a.

The appendix will use eqns. 2, 3, 4, 5 and lem. 6 from the previous section. It is to be noted that these do not use thm. 2 in any way and are mainly definitional equations. To avoid repeating awkward summations, for the loop $a = a_1, \dots, a_m$ we define

$$\begin{aligned} S(j, k) &= \sum_{i=j}^k a_i \\ S_i &= S(i-n+1, i) \\ T_i &= a_i / S_i \\ T(j, k) &= \sum_{i=j}^k T_i. \end{aligned}$$

By eq. 5, $R_{1,n} = R_{n,n}$. We will choose to work with $R_{n,n}$ for which eq. 3 can be written as

$$E_{P_{a,n}} (X_n / \sum_{j=1}^n X_j) = T(1, m) / m. \quad (\text{A.1})$$

Consider the case when a is of the special form $0^{n-1}1^x$ for some integer x . Working out the sums involved in eq. A.1, for this a

$$\begin{aligned} E_{P_{a,n}} (X_n / \sum_{j=1}^n X_j) &= (n-1+x)^{-1} \sum_{i=1}^x 1/i \\ &\leq \alpha(n-1). \end{aligned} \quad (\text{A.2})$$

It is easy to see that in eq. A.2 equality is achieved for some value of $x \leq n$ which we shall denote by $\beta(n-1)$ (the argument $n-1$ will be assumed from now on). The proof that amongst the set of all loops, the given loop $0^{n-1}1^\beta$ maximizes the expectation will be done by contradiction. Assume there is some

$a = a_1, \dots, a_m$ and $\epsilon > 0$, for which

$$T(1, m)/m > \alpha(n-1) + \epsilon. \quad (\text{A.3})$$

The method of proof involves a stepwise modification of a . At each step the previous sequence will be denoted by a , and the modified one by a' . The variables m' , for the length of a' , as well as S' and T' will similarly be defined for a' . After each step, for the modified sequence the inequality

$$T'(1, m')/m' > \alpha(n-1) \quad (\text{A.4})$$

will be proved. Yet after a finite number of steps, the sequence a' will essentially look like $0^{n-1}1^\beta$, providing the contradiction.

Step 1.

Let m' be a multiple of m , large enough so that $n/m' < \epsilon$, for the ϵ in eq. A.3, and also $m' > 5n$ (this last restriction is not necessary, but allows the treatment of a loop as a long open string).

We have $a = a_1, \dots, a_m$.

Let $a' = 0^{n-1}a_n, \dots, a_{m'}$.

To prove eq. A.4 note that $a'_i \leq a_i$, so $S'_i \leq S_i$. So for $i = n, \dots, m'$ we have $T'_i \geq T_i$, and for $i = 1, \dots, n-1$, $T'_i \leq 1$. Hence

$$T(1, m') \leq (n-1) + T'(n, m').$$

Since m' is a multiple of m ,

$$\begin{aligned} \alpha(n-1) + \epsilon &< T(1, m)/m \\ &= T(1, m')/m' \\ &\leq [(n-1) + T'(1, m')]/m' \\ &\leq \epsilon + T'(1, m')/m' \end{aligned}$$

which proves eq. A.4.

Step 2.

Now $a = 0^{n-1} a_n, a_{n+1}, \dots, a_m$. Define $b = S(n, 2n-1)$.

Let $a' = 0^{n-1} 1^b 0^{n-b} a_{2n}, \dots, a_m$.

Note that a' is simply a , with the block a_n, \dots, a_{2n-1} rearranged so that all of its b ones are to the left of its zeros. We pause to prove the following lemma about switching the order of a neighboring pair of 0 and 1.

Lemma 8.

Let a and a' be two loops of the same length m , identical except that $a_{n+j} = a'_{n+j+1} = 0$ and $a'_{n+j} = a_{n+j+1} = 1$. If $a_{j+1} = 0$, then

$$T(1, m) \leq T'(1, m).$$

Proof:

The proof consists simply of noting that the only difference between T_i and T'_i is $T_{2n+j} \leq T'_{2n+j}$, $T_{n+j} = T'_{n+j+1}$ and $T'_{n+j+1} = T_{n+j}$. \square

Applying lem. 8 repeatedly over a large block yields

Corollary 9.

If a has a block of zeros $a_{j+1} = \dots = a_{j+b} = 0$ then construct a' by rearranging the block $a_{n+j}, \dots, a_{n+j+b}$ so that the ones are to the left of the zeros, but otherwise a and a' are identical. Then the conclusion of lem. 8 is still valid.

Returning to step 2 in the construction, $\alpha(n-1) < T(1, m)/m \leq T'(1, m')/m'$, where the first inequality was established in step 1, the second follows from cor. 9.

Step 3.

Now $a = 0^{n-1} 1^b 0^{n-b} a_{2n}, \dots, a_m$.

Let $a' = 0^{n-1} 1^\beta 0^{n-b} a_{2n}, \dots, a_m$

so that $m' = m + \beta - b$.

By the definition of β in eq. A.2,

$$T(1, n+b-1) = \sum_{i=1}^b 1/i \leq (n+b-1)\alpha(n-1) \quad (\text{A.5})$$

$$T'(1, n+\beta-1) = \sum_{i=1}^{\beta} 1/i = (n+\beta-1)\alpha(n-1).$$

For the remaining values $i = n+b, \dots, m$ we have $T_i = T'_{i+\beta-b}$ if $\beta \geq b-1$.

When $\beta < b-1$ the only difference is that $S_i > S'_{i+\beta-b}$ for $i = 2n, \dots, 2n+b-\beta-2$, so that in all cases

$$T(n+b, m) \leq T'(n+\beta, m'). \quad (\text{A.6})$$

Combining eqns. A.5 and A.6

$$T(1, m) - T'(1, m') \leq (b-\beta)\alpha(n-1).$$

This implies eq. A.4 as can be seen by this simple lemma.

Lemma 10.

If $T(1, m) - T(1, m') \leq (m-m')\alpha$ and $T(1, m)/m > \alpha$ then $T'(1, m')/m' > \alpha$.

Proof:

$$0 < T(1, m) - m\alpha < T'(1, m') - m'\alpha. \quad \square$$

Step 4.

If $b > \beta$, return to step 2; otherwise $n-b \geq n-\beta$, so the second block of zeros in a has at least $n-\beta$ elements. Let a_c be the first occurrence of a 1

in $a_{2n+\beta-1}, \dots, a_m$.

Now $a = 0^{n-1}1\beta 0^{n-\beta} a_{2n}, \dots, a_m$.

Let $a' = 0^{n-1}1\beta 0^{n-1} a_c, \dots, a_m$,

so that $m' = m+2n+\beta-c-1$.

Note that $T(1, 2n-1) = T'(1, 2n+\beta-2)$, $T_{2n+\beta-1}^c = 1$ and $T(c+1, m) = T'(2n+\beta, m')$ so that

$$T(1, m) - T'(1, m') \leq T(2n, 2n+\beta-2) + T_c - 1. \quad (\text{A.7})$$

Let $d = S(2n, 2n+\beta-2)$ so that there are $n-d-1$ zeros in $a_n, \dots, a_{2n+\beta-2}$. Then each S_i for $i = 2n, \dots, 2n+\beta-2$ sums at most $n-d-1$ zeros, and at least $d+1$ ones, i.e., each $S_i \geq d+1$. Since a_i and hence T_i is nonzero d times for $i = 2n, \dots, 2n+\beta-2$

$$T(2n, 2n+\beta-2) \leq d/(d+1). \quad (\text{A.8})$$

We will separate out three cases, and in each case establish

$$T(1, m) - T'(1, m') \leq (m-m')\alpha(n-1), \quad (\text{A.9})$$

which would imply eq. A.4 by lemma 10.

Case 1: $2n+\beta-1 \leq c < 3n$.

Here $(m-m') \geq 0$ and $d = S_k - 1$, so eqns. A.7 and A.8 imply

$$T(1, m) - T'(1, m') \leq (S_k - 1)/S_k + 1/S_k - 1 = 0,$$

establishing eq. A.9.

Case 2: $c \geq 3n$ and $n \neq 4, 6, 8$ or 10 .

Since $d \leq \beta-1$ and $m-m' \geq n+1-\beta$, using eqns. A.7, A.8, we need to show

$$(\beta-1)/\beta \leq (n+1-\beta)\alpha(n-1). \quad (\text{A.10})$$

to prove eqn. A.9. Looking at table 1, this holds for all given values of n except 4, 6, 8, 10. For values beyond the table, eq. A.7 was checked numerically up to $n = 100$, and the logarithmic approximations of lemma 6 will be used after that. Since β maximizes eq. A.2, we have

Table 1:

<u>n</u>	<u>$\beta(n-1)$</u>	<u>$\alpha(n-1)$</u>
1	1	1.000000
2	1	.500000
3	2	.375000
4	3	.305556
5	3	.261905
6	4	.231481
7	4	.208333
8	5	.190278
9	5	.175641
10	6	.163333
11	6	.153125
12	6	.144118
13	7	.136466
14	7	.129643
15	8	.123539

$$\begin{aligned} \alpha(n-1) &\geq [n-1+(\beta-1)]^{-1} \sum_{i=1}^{\beta-1} 1/i \\ &= \left(\frac{n+\beta-1}{n+\beta-2} \right) \alpha(n-1) - \left(\frac{1}{n+\beta-2} \right) (1/\beta) \end{aligned}$$

which gives $\alpha(n-1) \leq 1/\beta$. Since $\beta\alpha(n-1) \leq 1$ and $(\beta-1)/\beta \leq 1$,

$$\begin{aligned} (n+1-\beta)\alpha(n-1) - (\beta-1)/\beta &\geq (n+1)\alpha(n-1) - 2 \\ \text{(by lem. 6)} &\geq (n+1) \frac{\log(n-1) - \log \log(n-1) - 1}{n-1} - 2 \\ &\geq 0 \quad \text{for } n \geq 87. \end{aligned}$$

The final inequality can be calculated for $n = 87$, and since the penultimate expression is an increasing function of n , all larger n must also satisfy it. But this establishes eq. A.10 and hence A.9 for all $n \neq 4, 6, 8$ or 10 .

Case 3: $c \geq 3n$ and $n = 4, 6, 8$, or 10 .

This case is further broken into three subcases each involving a verification by table 1.

(3a) If $c = 3n$ and $S_c > 1$ then $T_c \leq 1/2$ so if

$$(\beta-1)/\beta - 1/2 \leq (n+1-\beta)\alpha(n-1)$$

then eq. A.9 is satisfied.

(3b) If $c = 3n$ and $S_c = 1$ then $S(2n+1, 2n+\beta-2) = 0$, and so

$T(2n, 2n+\beta-2) = T_{2n} \leq 1/\beta$. Using this in eq. A.7, we need to verify

$$1/\beta \leq (n+1-\beta)\alpha(n-1).$$

(3c) If $c > 3n$ then $m-m' \geq n+2-\beta$ and we need

$$(\beta-1)/\beta \leq (n+2-\beta)\alpha(n-1).$$

As these cases are exhaustive, and in each case eq. A.4 is true, Step 4 is complete.

Step 5.

Now $a = 0^{n-1} 1^\beta 0^{n-1} a_{2n+\beta-1}, \dots, a_m$

Let $a' = 0^{n-1} a_{2n+\beta-1}, \dots, a_m 0^{n-1} 1^\beta$.

Since a' is just a rotation of a , $T(1, m) = T'(1, m')$, so eq. A.4 will hold. Now, return to step 2 unless

$$a = 0^{n-1} 1^\beta 0^{n-1} 1^\beta \dots 0^{n-1} 1^\beta. \quad (\text{A.11})$$

At every return to step 2, some elements of the original sequence are deleted or reordered into blocks of $0^{n-1} 1^\beta$. Since no new disordered elements are created at any step, the procedure must stop after a finite number of steps. Since at each step eq. A.4 was verified, for the final a of eq. A.11 we must have

$$T(1, m)/m > \alpha(n-1)$$

yet simply computing,

$$T(1, m)/m = (n-1+\beta)^{-1} \sum_{i=1}^{\beta} 1/i = \alpha(n-1)$$

providing the contradiction which proves the theorem. \square

Proof of Theorem 2b.

Let n be odd, $k = (n+1)/2$, and $a = a_1, \dots, a_m$. As notation, define

$$S(j, j') = \sum_{i=j}^{j'} a_i$$

$$T_i = a_{i+k}/S(i+1, i+n)$$

$$T(j, j') = \sum_{i=j}^{j'} T_i$$

so that eq. 3 can be written as

$$E_{p, a, n} (X_k / \sum_{j=1}^n X_j) = T(1, m)/m.$$

For any loop a ,

$$T(1, k) = \sum_{i=1}^k a_{i+k}/S(i+1, i+n) \leq \sum_{i=1}^k a_{i+k}/S(k+1, n+1) = 1.$$

As this holds for all loops, it will also hold for the loop

$(a_{hk+1}, a_{hk+2}, \dots, a_{hk+n})$ for any integer h . Thus

$$T(hk+1, (h+1)k) \leq 1 \quad \text{for } h = 0, 1, 2, \dots$$

Adding these up for $h = 0, 1, \dots, m-1$,

$$m > \sum_{h=0}^{m-1} T(hk+1, (h+1)k) = T(1, mk) = kT(1, m), \quad (\text{A.12})$$

because a is periodic with period m . Rewriting A.12 gives

$$T(1, m)/m \leq 1/k = 2/(n+1) \quad (\text{A.13})$$

for any loop a . On the other hand, it is straightforward to verify that the loop $a = 0^{k-1}1$ achieves the upper bound in eq. A.13, thus proving thm. 2b and cor. 3b simultaneously. \square

Bibliography

Daykin, D.E. (1970) "Inequalities for certain cyclic sums", Proc. Edin. Math. Soc. (Ser. 2), Vol. 17, pp. 257-262.

Hobby, C. and Ylvasaker, D. (1964) "Some structure theorems for stationary probability measures on finite state sequences", Ann. Math. Stat., Vol. 35, pp. 550-556.

Zaman, A. (1983) "Stationarity on finite strings and shift register sequences", Ann. Prob., August, 1983 (To appear).

1. REPORT NUMBER FSU M655 USARO D-60	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) A non-clustering property of stationary sequences		5. TYPE OF REPORT & PERIOD COVERED Technical
7. AUTHOR(S) Arif Zaman		6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report No. M655
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Florida State University Tallahassee, FL 32306		8. CONTRACT OR GRANT NUMBER(S) USARO Grant DAAG29-82-K-0168
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office - Durham P.O. Box 12211 Research Triangle Park, NC 27709		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE May, 1983
		13. NUMBER OF PAGES 17
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

6. DISTRIBUTION STATEMENT (of this report)

Distribution unlimited.

7. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

8. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

9. KEY WORDS

Clustering, Stationary Sequences, Cyclic Sums

2 of 2

10. ABSTRACT (Continue on reverse side if necessary and identify by block number)

→ For a random sequence of events, with indicator variables X_i , the behavior of the expectation $E[(X_k + \dots + X_{k+m-1}) / (X_1 + \dots + X_n)]$ for $1 \leq k \leq k+m-1 \leq n$ can be taken as a measure of clustering of the events. When the measure on the X 's is i.i.d., or even exchangeable, a symmetry argument shows that the expectation can be no more than m/n . When the X 's are constrained only to be a stationary sequence, the bound deteriorates, and depends on k as well. When m/n is small, the bound is roughly $2m/n$ for k near $n/2$ and is like $(m/n) \log n$ for k near 1 or n . The proof given is partly constructive, so these bounds are nearly achieved, even though there is room for improvement for other values of k .

↙

FILME

7—8