A NON-CLUSTERING PROPERTY OF STATIONARY SEQUENCES

by

Arif Zaman
Purdue University

Technical Report #82-5

Statistics Department

February 1982

A Non-clustering Property of Stationary Sequences

by

Arif Zaman
Purdue University

Summary

For a random sequence of events, with indicator variables X_i, the behavior of the expectation $E\left(\frac{X_k+\ldots+X_{k+m-1}}{X_1+\ldots+X_n}\right)$ 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 and m. For k near n+1/2, the bound is like cm/(n-m) and k near 1 or n has a bound like (m/n)logn . The proof given is partly constructive, and so these bounds are achieved.

1. Introduction

In considering portions of larger, but still finite strings of random variables, the following problem arose. If $X_1,\ldots,X_n$ 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 would expect that the expression $\sup_{P\in\mathcal{S}} E_P\left\{\frac{X_k+\ldots+X_{k+m-1}}{X_1+\ldots+X_n}\right\}$ (note: $0/0 = 0$) where $1 \le k \le k+m-1 \le n$, and $\mathcal{S}$ is the set of stationary probability measures on binary sequences, should 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 like 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 which 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 Michael Steele for insisting that this could be done, and Larry Shepp for an improvement in the proof.

2. Results

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

$$R_{k,n} = \sup_{P\in\mathcal{S}} E_P\left\{\frac{X_k}{X_1+\ldots+X_n}\right\} \quad \text{for } 1 \le k \le n.$$

Notice that the variables $X_{n+1}, X_{n+2}, \ldots$ do not appear in the above expression, so only the marginal distribution of $(X_1,\ldots,X_n)$ 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 circular string is a finite sequence $a_1,\ldots,a_m$ of zeros and ones. Subscripts less than one, or greater than m will be taken circularly, so that $a_o = a_m$ and $a_{m+1} = a_1$. For a circular string a, the measure $P_{a,n}$ gives mass 1/m to each of $(a_1,\ldots,a_n),(a_2,\ldots,a_{n+1}),\ldots,(a_m,\ldots,a_{m+n-1})$. Note that n may be larger than m.

Theorem 1

If a binary sequence X has a stationary distribution, then the marginal distribution of $(X_1,\ldots,X_n)$ lies in a convex set of measures $\mathcal{S}^n$. The set of extreme points of $\mathcal{S}^n$ is of the form $\{P_{a,n}: a \in A_n\}$ for a finite set A_n of circular strings. Moreover, $P_{a,n} \in \mathcal{S}^n$ for every circular string a.

More details, and a proof of this can be found in Zaman (1981) or Hobby and Ylvasaker (1964). By this theorem the maximization over all stationary sequences $\mathcal{S}$, is the same as maximization over $\mathcal{S}^n$, for computing $R_{k,n}$. Further, since expectation is a linear functional, and $\mathcal{S}^n$ a convex set, any supremum must be attained at an extreme point. Thus

$$R_{k,n} = \max_{a \in A_n} E_{P_a} \left\{ \frac{X_k}{X_1+\ldots+X_n} \right\} \tag{1}$$

Doing an explicit maximization over these extreme points, the following key theorem is proved in the appendix.

Theorem 2

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

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

Corollary 3

Define

$$\alpha(n) = \sup_{\beta} \frac{1}{n+\beta} \sum_{i=1}^{\beta} 1/i \; . \tag{2}$$

Then,

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

The corollary is actually proved as a step in proving thm. 2, but can also be proved from thm. 2 using the explicit form of eq. 1 given in eq. 6 in the appendix.

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, and their results are summarized in the graphs in fig. 1.

Theorem 4

Define

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

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 $R_{k,n} = R_{n+k-1,n}$ is established. To prove this, note that

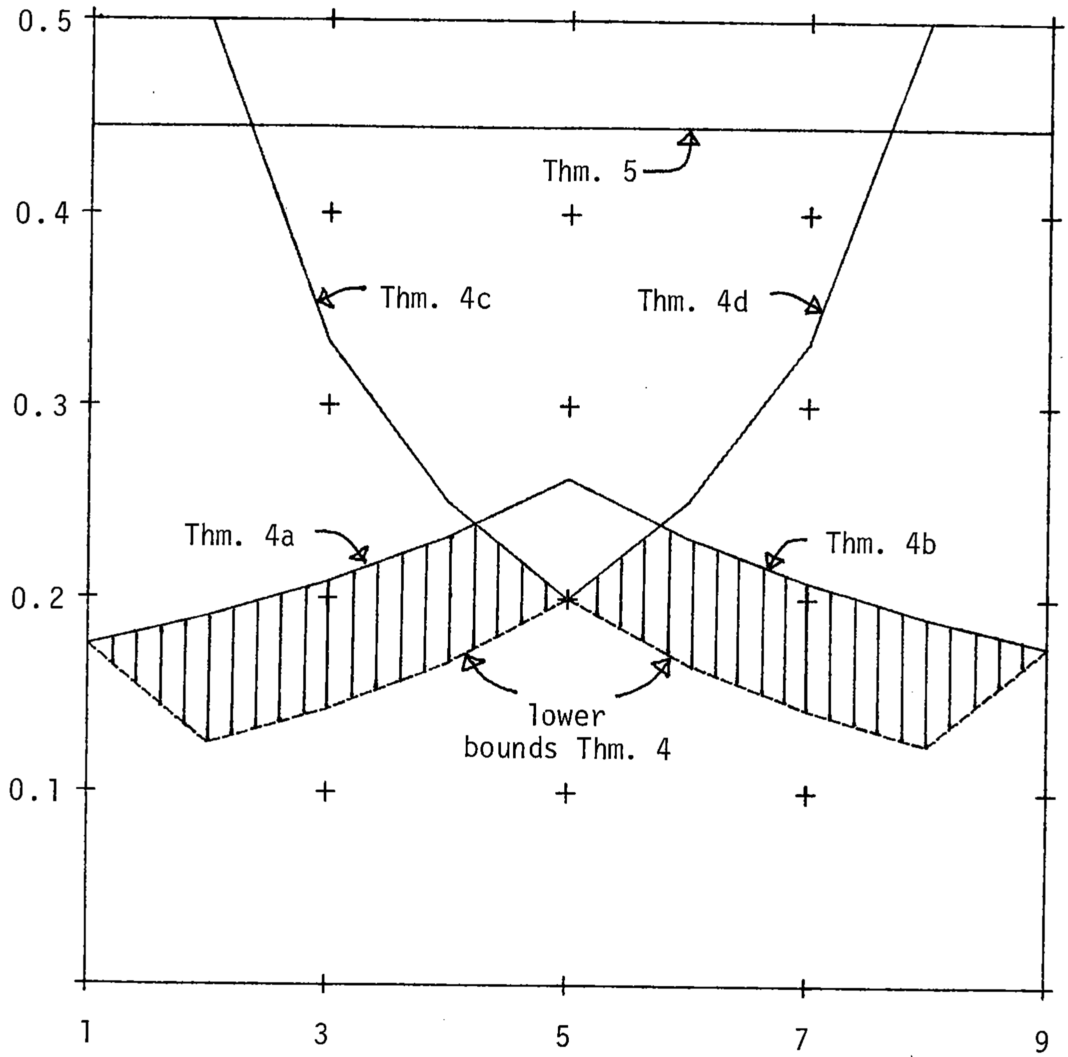

Figure 1a: Bounds on $R_{k,n}$ for $n = 9$

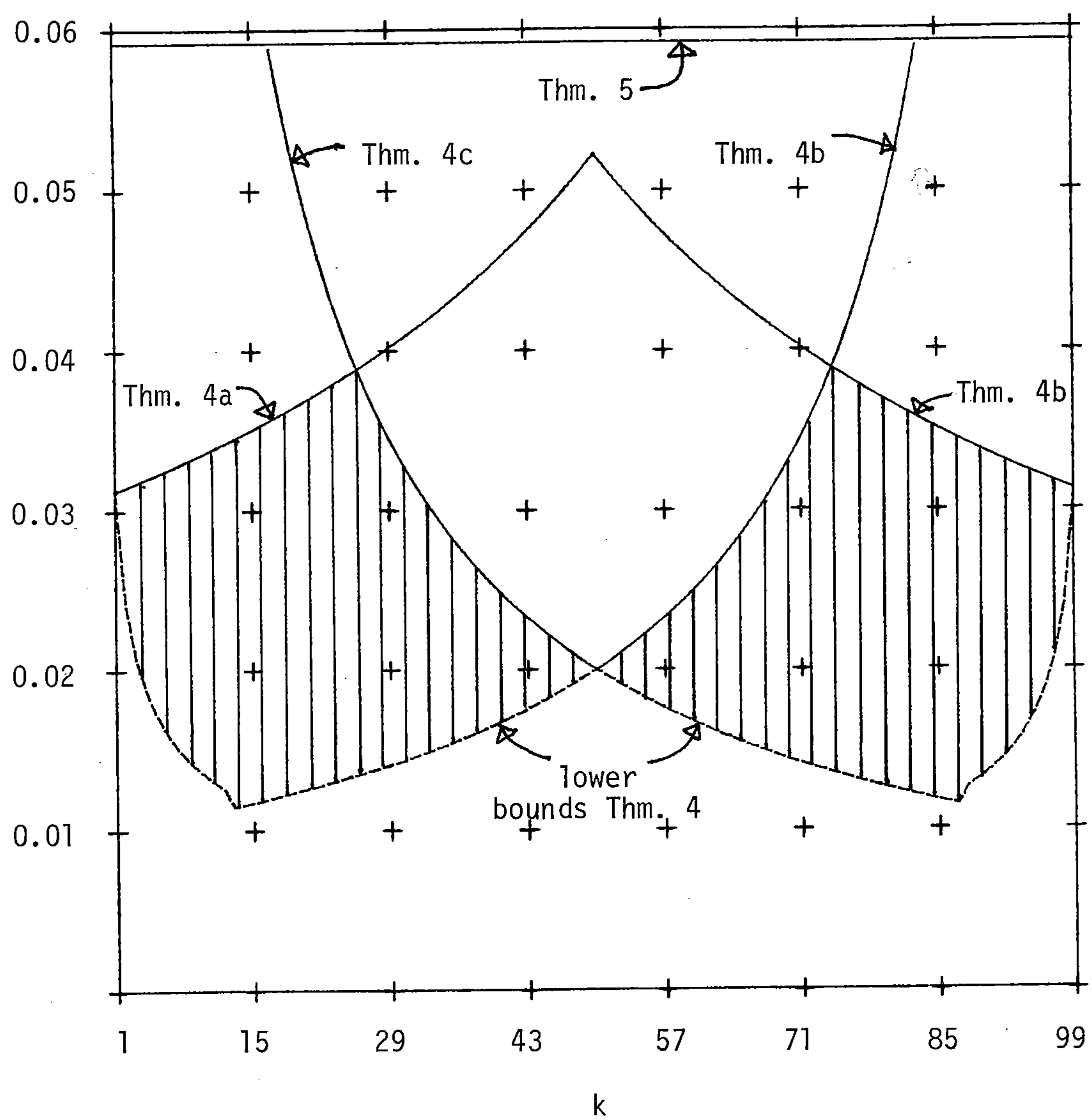

Figure 1b: Bounds on $R_{k,n}$ for n = 99

if $P_{a,n}$ is the distribution of $(X_1,\ldots,X_n)$ then the distribution of $(X_n,\ldots,X_1)$ is given by $P_{a',n}$ for $a' = (a_m,\ldots,a_1)$, so $P_{a',n} \in \mathcal{S}^n$. Now for any circular string a,

$$E_{P_{a,n}}\left\{\frac{X_k}{X_1+\ldots+X_n}\right\} = E_{P_{a',n}}\left\{\frac{X_{n+1-k}}{X_1+\ldots+X_n}\right\}$$

from which the symmetry condition follows directly.

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

$$R_{k,n} \le \sup_{P\in\mathcal{S}} E_P\left\{\frac{X_k}{X_k+\ldots+X_n}\right\} = R_{1,n+1-k} = \alpha(n-k)$$

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

$$R_{k,n} \le \sup_{P\in\mathcal{S}} E_P\left\{\frac{X_k}{X_1+\ldots+X_{2k+1}}\right\} = R_{k,2k+1} = \frac{1}{k} \quad .$$

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. 6 from the appendix to get for $k \le (n+1)/2$

$$R_{k,n} \ge \sup_{\substack{a=0^{n-k}1^{\beta} \\ k\le\beta\le n}} E_{P_a}\left\{\frac{X_k}{X_1+\ldots+X_n}\right\} = \sup_{k\le\beta\le n} \frac{1}{n+\beta-k}\left[\sum_{i=k}^{\beta} 1/i+(k-1)/\beta\right]$$

The lower bound in (c) is achieved by letting $a = 0^{n-k}1$. For that value of a, if $2k + 1 \le n$ then by eq. 6

$$E_{P_a}\left\{\frac{X_k}{X_1+\ldots+X_n}\right\} = \frac{1}{n+1-k}$$

It is not difficult to find sequences which give even higher lower bounds, but that doesn't seem to be the more fruitful direction of moving the bounds. □

Theorem 5

$$R_{k,n} \leq \frac{1+\log(n-1)}{n-\log(n-1)} \quad \text{for } n \geq 7 .$$

For a proof of this theorem, we first need a logarithmic approximation to the function α, given by the following lemma.

Lemma 6

$$\frac{\log(\frac{n-1}{2})-\log\log(\frac{n-1}{2})}{n-1+\log(\frac{n-1}{2})-\log\log(\frac{n-1}{2})} \leq \alpha(n) \leq \frac{\log n}{n-\log n} \quad \text{for } n \geq 7$$

Proof:

Let β_n be a value of β which achieves the maximum in eq. 2, so that

$$\alpha(n) = \frac{1}{n+\beta_n} \sum_{i=1}^{\beta_n} 1/i \quad .$$

This means

$$\alpha(n) \geq \frac{1}{n+\beta_n-1} \sum_{i=1}^{\beta_n-1} 1/i = \left(\frac{n+\beta_n}{n+\beta_n-1}\right)\alpha(n) - \frac{1}{n+\beta_n-1}(1/\beta_n)$$

so $\alpha(n) \leq 1/\beta_n$.

Similarly

$$\alpha(n) \geq \frac{1}{n+\beta_n+1} \sum_{i=1}^{\beta_n+1} 1/i = \left(\frac{n+\beta_n}{n+\beta_n+1}\right)\alpha(n) + \frac{1}{n+\beta_n+1}\left(\frac{1}{\beta_n+1}\right)$$

so $\alpha(n) \geq \frac{1}{\beta_n+1}$.

Combining these two results

$$\frac{1}{\beta_n+1} \leq \alpha(n) = \frac{1}{n+\beta_n} \sum_{i=1}^{\beta_n} 1/i \leq \frac{1}{\beta_n}$$

so

$$\frac{n-1}{\beta_n+1} \le \left(\sum_{i=1}^{\beta_n} 1/i \right) - 1 \le \frac{n}{\beta_n} \tag{3}$$

Using a logarithmic approximation for the center term,

$$\log(\beta_n/2) + 1/\beta_n \le \left(\sum_{i=1}^{\beta_n} 1/i \right) - 1 \le \log(\beta_n+1) - \frac{1}{\beta_n+1} \tag{4}$$

We shall use these equations coupled with the simple result that if $x \log x = y$, and $e \le y$,

$$\frac{y}{\log y} \le x \le \frac{y}{\log[\frac{y}{\log y}]} = \frac{y}{\log y - \log \log y} \ . \tag{5}$$

Let β^- be the smallest value of β which satisfies both eq. 3 and 4. By the left part of eq. 3, and the right part of eq. 4,

$$\frac{n-1}{\beta^-+1} = \log(\beta^-+1) - \frac{1}{\beta^-+1}$$

which by eq. 5 means that if $n \ge e$

$$\frac{n}{\log n} \le \beta^- + 1 \ .$$

Similarly, using the right part of eq. 3 with the left of eq. 4, the largest possible value β^+ must satisfy

$$\log(\beta^+/2) + 1/\beta^+ = n/\beta^+ \ .$$

Rewriting this as

$$(\beta^+/2)\log(\beta^+/2) = (n-1)/2$$

allows the use of eq. 5, when $(n-1)/2 > e$. Combining the results on β^- and β^+,

$$\frac{n}{\log n} - 1 \le \beta_n \le \frac{n-1}{\log(\frac{n-1}{2}) - \log \log(\frac{n-1}{2})} \ .$$

Reexpressing this in terms of α,

$$\frac{\log(\frac{n-1}{2})-\log\log(\frac{n-1}{2})}{n-1+\log(\frac{n-1}{2})-\log\log(\frac{n-1}{2})} \leq \frac{1}{\beta_n+1} \leq \alpha(n)$$

$$\alpha(n) \leq \frac{1}{\beta_n} \leq \frac{\log n}{n-\log n}$$

which proves the claimed result. □

The proof of theorem 5 then amounts to the following. By the symmetry mentioned in the proof of theorem 4,

$$\max_k R_{k,n} = \max_{k\leq(n+1)/2} R_{k,n}$$

(by thm. 4 a,c)
$$\leq \max_{k\leq(n+1)/2} \left\{ \frac{1}{k} \wedge \alpha(n-k) \right\}$$

$$\leq \max_{k\leq(n+1)/2} \left\{ \frac{1}{k} \wedge \frac{\log(n-k)}{n-k-\log(n-k)} \right\}$$

Consider maximizing this expression over all real values $1 \leq k \leq (n+1)/2$. SInce $1/k$ is decreasing, and the second function monotone increasing, there must be a unique crossover point k_n which attains this maximum, so that

$$\max_k R_{k,n} \leq \frac{1}{k_n} = \frac{\log(n-k_n)}{n-k_n-\log(n-k_n)}$$

$$= \frac{1+\log(n-k_n)}{n-\log(n-k_n)}$$

where the last expression follows by some algebra. Since $k_n \geq 1$, one can replace k_n by 1 to get the claimed result of the theorem. □

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

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

For example this proves that for any stationary measure P,

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

and for blocks near the middle

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

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

APPENDIX

Proof of Theorem 2b

The basic idea of the proof is to write out the expectation in eqn. 1 explicitly as

$$R_{k,n} = \max_{a \in A_n} \frac{1}{m(a)} \sum_{i=1}^{m(a)} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}} \tag{6}$$

where m(a) is the length of the cicular string a. For example, when n=4, k=2, m(a) = 6, and

$i =$	0	1	2	3	4	5	6	7	8	9
$a_i =$	$1(-a_5)$	0	1	1	0	1	0	$0(-a_1)$	1	1
$\sum_{j=1}^{n} a_{i+j} =$		3	2	1	2	2	3			

$$R_{k,n} = \frac{1}{6} \left\{ \frac{1}{3} + \frac{1}{2} + 0 + \frac{1}{2} + 0 + 0 \right\} .$$

First consider the case (b) where n is odd, and k = (n+1)/2. Let a be any circular string of length m. Then

$$\sum_{i=1}^{k} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}} \leq \sum_{i=1}^{k} \frac{a_{i+k}}{\sum_{j=k+1}^{n+1} a_j} = \sum_{i=k+1}^{n+1} a_i \Big/ \sum_{j=k+1}^{n+1} a_j = 1$$

Since the above claim is true for any a, it will also hold for the cicular strings $(a_{hk+1}, a_{hk+2}, \ldots, a_{hk+n})$ for any integer h. Thus

$$\sum_{i=hk+1}^{(h+1)k} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}} \leq 1 \qquad \text{for} \quad h=0,1,2,\ldots$$

Adding up these sums for h ranging from 0 to m - 1,

$$m \geq \sum_{h=0}^{m-1} \sum_{i=hk+1}^{(h+1)k} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}} = \sum_{i=1}^{mk} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}}$$

$$= \sum_{h=0}^{k-1} \left(\sum_{i=hm+1}^{(h+1)m} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}} \right) = k \sum_{i=1}^{m} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}}$$

The reason for the last equality is that the parenthesized expression is indendent of h, because a is circular. Rewriting the above result gives

$$E_{P_a}\left(\frac{X_k}{X_1+\ldots+X_n}\right) = \frac{1}{m} \sum_{i=1}^{m} \frac{a_{i+k}}{\sum_{j=1}^{n} a_{i+j}} \leq \frac{1}{k} = \frac{2}{n+1}$$

for any circular sequence a. On the other hand, it is straightforward to verify that the string $a = 0^{k-1}1$ achieves this upper bound, thus proving both thm. 2b, and cor. 3b simultaneously. □

Proof of thm 2b

By the symmetry condition shown in the proof of thm. 4, $R_{1,n} = R_{n,n}$. The computations here will be carried out for $R_{n,n}$ because they are notationally simpler. As further notation, let

$$S_j = \sum_{i=j-n+1}^{j} a_i$$

so that for any circular string $a = a_1,\ldots,a_m$,

$$E_{P_a}\left(\frac{X_n}{X_1+\ldots+X_n}\right) = \frac{1}{m} \sum_{i=1}^{m} a_i/S_i \ .$$

Consider the case where the string $a = 0^{n-1}1^\beta$ for some integer $\beta \leq n$. In this case

$$E_{P_a}\left(\frac{X_n}{X_1+\ldots+X_n}\right) = \frac{1}{n-1+\beta} \sum_{i=1}^{\beta} 1/i \leq \alpha(n-1)$$

with equality holding for some value of β which we shall call β_{n-1} in accordance with the notation used in the proof of lemma 6. The proof that the string $0^{n-1}1^{\beta_{n-1}}$ maximizes the above expectation of all sequences will be done by contradiction. Assume that there is some $a^0 = a_1^0,\ldots,a_m^0$ and $\varepsilon > 0$, for which

$$\frac{1}{m} \sum_{i=1}^{m} a_i^0/S_i > \alpha(n-1) + \varepsilon.$$

The method of proof involves a stepwise modification of a^0. 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', and S_j' for the partial sums of a' will also be used. After each step, it will be shown that for the modified sequence,

$$\frac{1}{m'} \sum_{i=1}^{m'} a_i'/S_i' > \alpha(n-1). \qquad (7)$$

Yet, after a finite number of steps, the sequence a' will essentially look like $0^{n-1}1^{\beta}$, providing the contradiction. A global view of this procedure is provided by the flowchart in Figure 2.

Step 1

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

$$a_i' = \begin{cases} 0 & \text{if} \quad i=1,\ldots,n-1 \\ a_i & \text{if} \quad i=n,\ldots,m' \end{cases}$$

To prove eq. 7 note that $a_i' \le a_i$ so $S_i' \le S_i$. So for $i = n,\ldots,m'$ we have $a_i/S_i \le a_i'/S_i'$, and for $i = 1,\ldots,n-1$, $a_i/S_i \le 1$ so

$$\sum_{i=1}^{m'} a_i/S_i \le (n-1) + \sum_{i=n}^{m'} a_i'/S_i' .$$

Since m' is a multiple of m, the length of a,

$$\alpha(n-1) + \varepsilon < \frac{1}{m} \sum_{i=1}^{m} a_i/S_i = \frac{1}{m'} \sum_{i=1}^{m'} a_i/S_i$$

$$\leq \varepsilon + \sum_{i=1}^{m'} a_i'/S_i'$$

which shows eq. 7.

Step 2

Now a looks like $0^{n-1}, a_n, a_{n+1}, \ldots, a_m$. Let $b = \sum_{i=n}^{2n-1} a_i$, and define a' by

$$a_i' = \begin{cases} 1 & \text{for } i=n,\ldots,n+b-1 \\ 0 & \text{for } i=n+b,\ldots,2n-1 \\ a_i & \text{otherwise} \end{cases}$$

Note that a' is simply the string a, with the zeros and ones in the block $a_n, \ldots, a_{2n-1}$ rearranged so that all the b ones are to the left of the zeros. Since a similar rearrangement of ones and zeros is done in step 4, it will be useful to establish the following general lemma about reorderings.

Lemma 8

Let a and a' be two strings of the same length m, which are identical except that

$$a_{n+j} = 0 \qquad a'_{n+j} = 1$$

$$a_{n+j+1} = 1 \qquad a'_{n+j+1} = 0 .$$

If $a_{j+1} = 0$ then

$$\sum_{i=1}^{m} a_i/S_i \leq \sum_{i=1}^{m} a_i'/S_i'.$$

The following corollary amounts to repeated applications of the lemma.

Corollary 9

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

Proof (of lemma)

S and S' differ only in the following two cases

$$S_{2n+j} - 1 = S'_{2n+j+1}$$

$$S_{n+j} + 1 = S'_{n+j} \ .$$

Hence the only differences in a_i/S_i and a'_i/S'_i are

$$a_{2n+j}/S_{2n+j} \le a'_{2n+j}/S'_{2n+j}$$

$$a_{n+j}/S_{n+j} = 0 = a'_{n+j+1}/S'_{n+j+1}$$

$$a_{n+j+1}/S_{n+j+1} = a'_{n+j}/S_{n+j} \ .$$

Thus proving the claim of the lemma. □

Returning to step 2 in the construction, we have

$$\alpha(n-1) < \frac{1}{m} \sum_{i=1}^{m} a_i/S_i \le \frac{1}{m'} \sum_{i=1}^{m'} a'_i/S'_i$$

where the first inequality was established in step 1, and the second follows directly form cor. 9.

Step 3

Now $a = 0^{n-1}1^b0^{n-b}a_{2n}a_{2n+1},\ldots,a_m$. Let $a' = 0^{n-1}1^{\beta_{n-1}}0^{n-b}a_{2n}a_{2n+1},$ $\ldots,a_m$ so that $m' = m + \beta_{n-1} - b$. From now on β without a subscript will refer to β_{n-1}. By the defining property of β_{n-1}, we get the inequality

$$\frac{1}{n-1+b} \sum_{i=1}^{n+b-1} a_i/S_i = \frac{1}{n-1+b} \sum_{i=1}^{b} 1/i$$

$$\leq \frac{1}{n-1+\beta} \sum_{i=1}^{\beta} 1/i = \sum_{i=1}^{n+\beta-1} a_i'/S_i' \ .$$

Also, for $i = n + b,\dots,m$ we have $a_i/S_i = a'_{i+\beta-b}/S'_{i+\beta-b}$ so

$$\frac{1}{m-n-b+1} \sum_{i=n+b}^{m} a_i/S_i = \frac{1}{m'-n-\beta-1} \sum_{i=n+\beta}^{m'} a_i'/S_i' \ .$$

The following equation then is simply a convex combination of the previous two,

$$\frac{1}{m} \sum_{i=1}^{m} a_i/S_i \leq \frac{1}{m'} \sum_{i=1}^{m'} a_i'/S_i' \ ,$$

Thus proving eq. 7.

If $\beta < b$, return to step 2, otherwise go on to

Step 4.

Now $a = 0^{n-1}1^{\beta}0^{n-\beta}a_{2n},\dots,a_m$. Define $c = \sum_{i=2n+\beta-1}^{3n-1} a_i$ and let

$$a_i' = \begin{cases} 1 & \text{for} \quad i=2n+\beta-1,\dots,2n+\beta+c-2 \\ 0 & \text{for} \quad i=2n+\beta+c-1,\dots,3n-1 \\ a_i & \text{otherwise} \ . \end{cases}$$

Again this is a rearrangement of zeros and ones, and so eq. 7 follows from a use of cor. 8.

Step 5

Now $a = 0^{n-1}1^{\beta}0^{n-\beta}a_{2n},\dots,a_{2n+\beta-2},1^{c},0^{n-\beta-c+1},a_{3n},\dots,a_m$. Before prescribing a', the claim

$$\sum_{i=2n}^{2n+\beta-1} a_i/S_i \leq 1$$

will be shown. For this, let $j_1, j_2, \ldots, j_d$ be the subscripts of the 1's in the block $a_{2n}, \ldots, a_{2n+\beta-1}$, so that a_{j_1} is the first 1, and a_{j_d} is the last occurance of a 1 in that block. Then $j_d \leq 2n + \beta - 1$, $j_{d-1} \leq 2n + \beta - 2$ and in general

$$j_k \leq 2n + \beta + k - d - 1.$$

Now

$$S_{j_k} = \sum_{i=2n}^{j_k} a_i + \sum_{i=j_k-n+1}^{2n-1} a_i$$

$$= k + \sum_{i=j_k-n+1}^{n+\beta-1} 1$$

$$= k + [(2n+\beta-j_k-1) \vee 0]$$

$$\geq k + [(d-k) \vee 0] \geq d \quad .$$

Thus

$$\sum_{i=2n}^{2n+\beta-1} a_i/S_i = \sum_{i=1}^{d} a_{j_i}/S_i \leq \sum_{i=1}^{d} 1/d = 1$$

proving the claim.

<u>Case 1</u>

Using this result, consider the case of $c > 0$. Let $a' = 0^{n-1}1^{\beta}0^{n-1}1^{c}0^{n-\beta-c+1}, a_{3n}, \ldots, a_m$. Then

$$\sum_{i=2n+1}^{2n+\beta-1} a_i'/S_i' \geq \frac{a'_{2n+\beta-1}}{S'_{2n+\beta-1}} = 1 \geq \sum_{i=2n+1}^{2n+\beta-1} a_i/S_i \quad .$$

Also $a_i' \leq a_i$ for all i, so $S_i' \leq S_i$. This means that for all i not in the range $2n + 1, \ldots, 2n + \beta - 1$, $a_i'/S_i' \geq a_i/S_i$, so

$$\sum_{i=1}^{m} a_i/S_i \leq \sum_{i=1}^{m'} a_i'/S_i'$$

thus proving eq. 7.

Case 2

$c = 0$ and $n \geq 16$. Let $a' = 0^{n-1}1^{\beta}0^{n-1}a_{3n},\ldots,a_m$. Then

$$m\alpha(n-1) < \left[\sum_{i=1}^{2n-1} + \sum_{i=2n}^{2n+\beta-1} + \sum_{i=2n+\beta}^{3n-1} + \sum_{i=3n}^{m}\right] a_i/S_i$$

$$\leq \sum_{i=1}^{2n-1} a_i/S_i + 1 + 0 + \sum_{i=3n}^{m} a_i/S_i$$

$$= \sum_{i=1}^{m'} a_i'/S_i' + 1 \quad .$$

Since $m' = m + \beta - n - 1$, this can be rewritten as

$$\frac{1}{m'} \sum_{i=1}^{m'} a_i'/S_i' > \frac{(m'+n+1-\beta)\alpha(n-1)-1}{m'}$$

$$= \alpha(n-1) + \frac{(n+1-\beta)\alpha(n-1)-1}{m'} \quad .$$

So to prove eq. 7 all that is needed is to show that the second term is positive, i.e. we need to prove

$$(n+1-\beta_{n-1})\alpha(n-1) \geq 1 \quad . \tag{8}$$

Using a lower bound for α and an upper bound for β, from Lemma 6, it is sufficient to show that

$$\left[n+1-\frac{n-2}{\ell}\right]\frac{\ell}{n-2+\ell} \geq 1$$

where $\ell = \log(\frac{n-2}{2}) - \log\log(\frac{n-2}{2})$. By algebra, the sufficient condition reduces to

$$\ell \geq \frac{2(n-2)}{n} \quad ,$$

which is true for $n \geq 44$. For $n = 16,\ldots,44$, an actual computation of the exact values of α and β can show eq. 8 directly.

Case 3

$c = 0$, $n \leq 15$. This is an annoying case. It can be verified that letting $a' = 0^{n-1}1^{\beta}0^{n-1}a_r, a_{r+1}, \ldots, a_m$ where a_r is the first 1 in $a_r, a_{r+1}, \ldots, a_m$, does satisfy eq. 7. This was verified on the computer by considering all possible values for $a_{2n}, \ldots, a_{2n+\beta-2}$, and a_{3n}, a_{3n+1}. For the purists, it has also been verified by another hand calculation which involves considering 8 different cases.

Step 6

The worst is over. We now have

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

Let $a' = 0^{n-1}a_{2n+\beta-1}, \ldots, a_m, 0^{n-1}1^{\beta}$, which is just a rotation of a, and hence doesn't affect any values. Now return to Step 2 unless $a = 0^{n-1}1^{\beta}0^{n-1}1^{\beta}, \ldots, 0^{n-1}1^{\beta}$.

The entire procedure is summarized by the flowchart in figure 2. For any return to step 2 (either from step 3 or 6) some elements of the original sequence are either deleted or reordered into blocks of $0^{n-1}1^{\beta}$. Since no new disordered elements are created at any step, in some finite number of steps the procedure must stop. So eventually

$$a = 0^{n-1}1^{\beta}0^{n-1}1^{\beta}, \ldots, 0^{n-1}1^{\beta}$$

and eq. 7 holds, so

$$\frac{1}{m} \sum_{i=1}^{m} a_i/S_i > \alpha(n-1).$$

But for this a,

$$\frac{1}{m} \sum_{i=1}^{m} a_i/S_i = \frac{1}{n-1+\beta} \sum_{i=1}^{\beta} 1/i = \alpha(n-1)$$

providing the contradiction which proves the theorem.

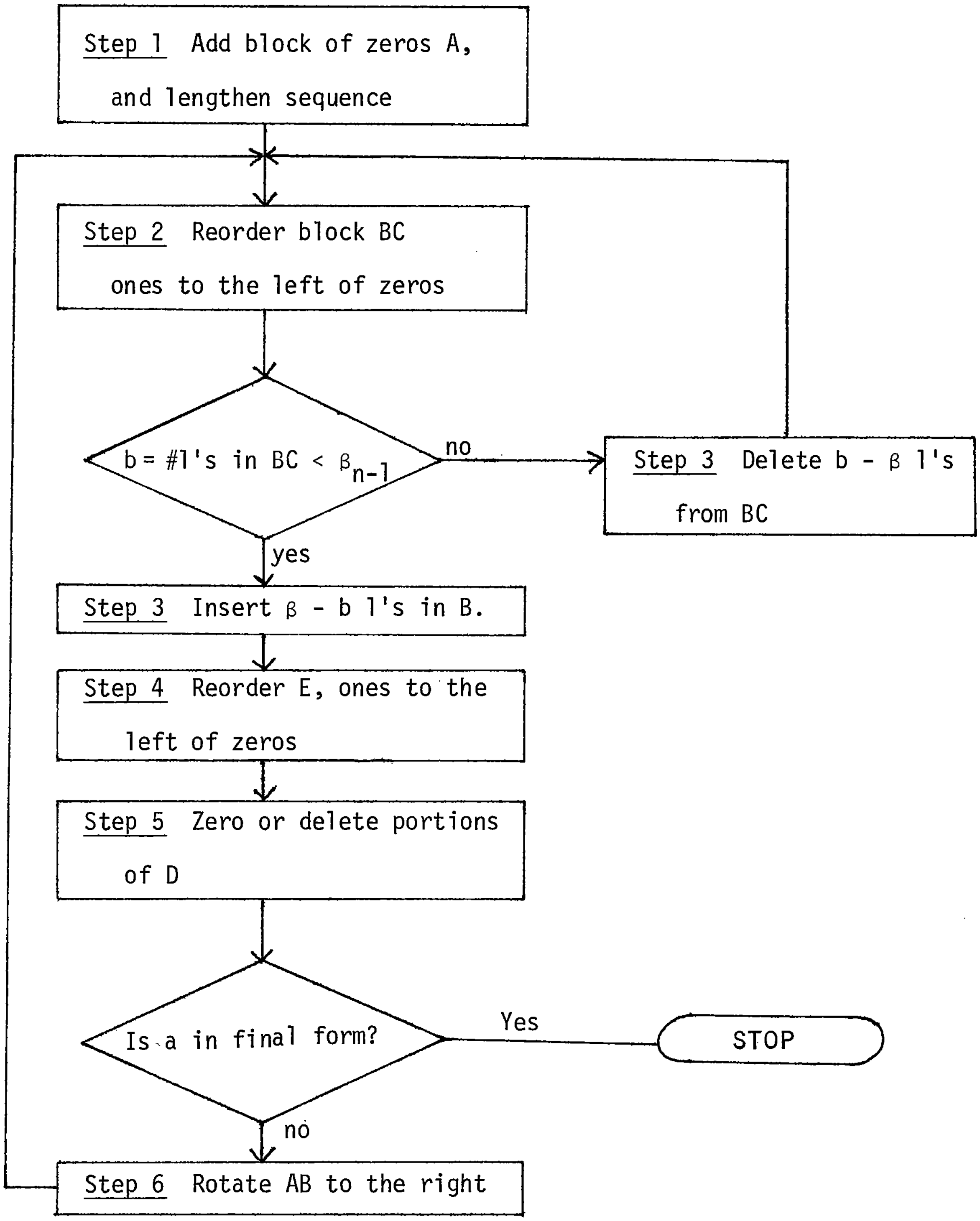

Figure 2 Flowchart: For the purposes of description, the blocks

$A = a_1, \ldots, a_{n-1}$ $D = a_{2n}, \ldots, a_{2n+\beta}$

$B = a_n, \ldots, a_{n+\beta-1}$ $E = a_{2n+\beta-1}, \ldots, a_{3n-1}$

$C = a_{n+\beta}, \ldots, a_{2n-1}$

Bibliography

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. (1981) "Stationarity on finite strings and shift register sequences", Technical report #81-33, Purdue University.