Statistical Span Property of Binary Run Sequences

based on recent publication with the same title on IEEE TIT

2023 Information Theory and Applications

Gangsan Kim and Hong-Yeop Song

Channel Coding Lab
School of Electrical and Electronic Engineering
YONSEI UNIVERSITY, Seoul, Korea

In this talk

- Motivation and
  the number of cyclically distinct run sequences

- Statistical span property of run sequences

- Three “open” problems and summary

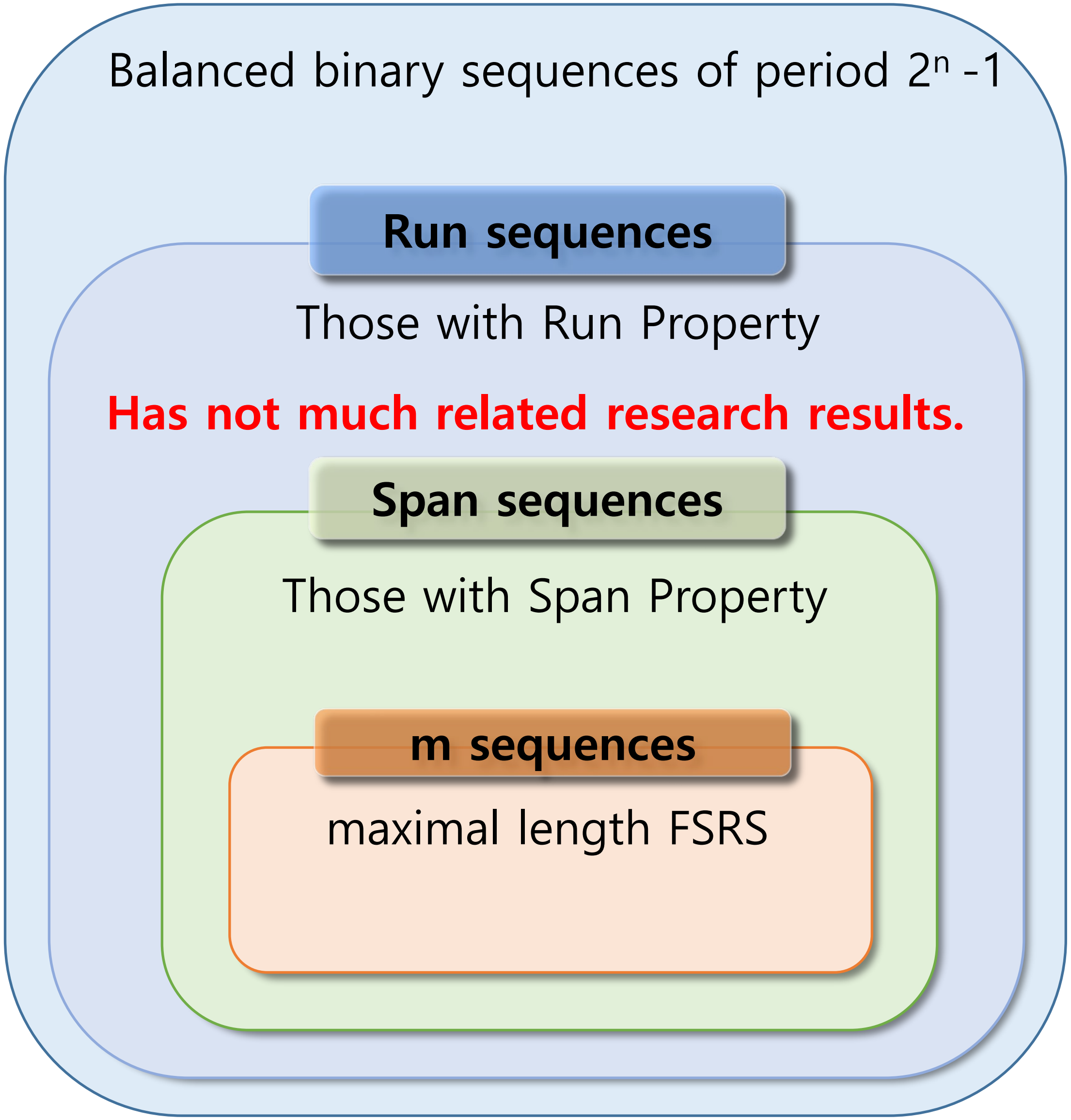
- **Motivation and the number of cyclically distinct run sequences**

- Statistical span property of run sequences

- Three “open” problems and summary

Motivation

In a world of balanced binary sequences of period $2^n - 1$



Which run sequences are span sequences?

or

Why some run sequences fail to be span sequences?

or

Does a run sequence have the span property at all?
In what sense?

Run sequences of length $15=2^4-1$

RUN property: 000, 1111, 00, 11, 0, 1, 0, 1

0 0 0 1 1 1 1 0 0 1 1 0 1 0 1

0 0 0 1 1 1 1 0 0 1 0 1 1 0 1

0 0 0 1 1 1 1 0 1 0 1 1 0 0 1

Run sequences of length 15=2^4-1

RUN property: 000, 1111, 00, 11, 0, 1, 0, 1

It is a run sequence but **not** a span sequence.

0 0 0 1 1 1 1 0 0 1 1 0 1 0 1

It is a **span sequence** and hence it is a run sequence.
It is **not** an m-sequence.

0 0 0 1 1 1 1 0 0 1 0 1 1 0 1

An **m-sequence** with a linear recursion $a_k = a_{k-1} + a_{k-4}$
It is a span sequence and hence a run sequence.

0 0 0 1 1 1 1 0 1 0 1 1 0 0 1

The number of run sequences

Theorem 1.

The number r_n of cyclically distinct (binary) run sequences of length $2^n - 1$ is

$$r_n = \frac{1}{2^{n-2}} \binom{2^{n-2}}{2^{n-3}, 2^{n-4}, \ldots, 2^1, 2^0, 1}^2$$

Idea of Proof

0 0 0 1 1 1 1 0 0 1 1 0 1 0 1

0 0 0 1 1 1 1 0 0 1 0 1 1 0 1

0 0 0 1 1 1 1 0 1 0 1 1 0 0 1

0 0 0 1 0 1 0 1 1 0 0 1 1 1 1

#of cyclically distinct examples

TABLE I: The number of cyclically distinct binary sequences with some randomness property

n	$2^n - 1$	# m-sequences $= \frac{\phi(2^n-1)}{n}$	# Span sequences ($\triangleq s_n$)	# Run sequences ($\triangleq r_n$)
3	7	2	2	2
4	15	2	16	36
5	31	6	2,048	88,200
6	63	6	67,108,864	7,304,587,290,000

Some analysis of the cardinality

Cor 1.

Let r_n be the number of cyclically distinct run sequences of period $2^n - 1$. Then,

$$r_n \rightarrow c2^{2^n - \frac{1}{2}n^2 + \left(\frac{3}{2} - \log_2 \pi\right)n}$$

for some constant c in the range $0.189 \leq c < 0.223$,

and hence,

$$\frac{r_n}{r_{n-1}} \rightarrow 2^{2^{n-1} - n + 2 - \log_2 \pi}.$$

Cor 2.

Let $s_n = 2^{2^{n-1} - n}$ be the number of cyclically distinct span sequences of period $2^n - 1$. Then,

$$\frac{r_n}{s_n} \rightarrow c2^{2^{n-1} - \frac{1}{2}n^2 + \left(\frac{5}{2} - \log_2 \pi\right)n}.$$

Statistical span property of run sequences

Average occurrence: n=4

All the cylically distinct $r_4(=36)$ run sequences of period 15

0 0 0 1 1 1 1 0 0 1 1 0 1 0 1	0 0 0 1 1 1 1 0 1 1 0 0 1 0 1	0 0 0 1 1 1 1 0 1 1 0 1 0 0 1
0 0 0 1 1 1 1 0 0 1 0 1 1 0 1	0 0 0 1 1 1 1 0 1 0 0 1 1 0 1	0 0 0 1 1 1 1 0 1 0 1 1 0 0 1
0 0 0 1 1 1 1 0 0 1 0 1 0 1 1	0 0 0 1 1 1 1 0 1 0 0 1 0 1 1	0 0 0 1 1 1 1 0 1 0 1 0 0 1 1
0 0 0 1 1 0 0 1 1 1 1 0 1 0 1	0 0 0 1 1 0 1 1 1 1 0 0 1 0 1	0 0 0 1 1 0 1 1 1 1 0 1 0 0 1
0 0 0 1 1 0 0 1 0 1 1 1 1 0 1	0 0 0 1 1 0 1 0 0 1 1 1 1 0 1	0 0 0 1 1 0 1 0 1 1 1 1 0 0 1
0 0 0 1 1 0 0 1 0 1 0 1 1 1 1	0 0 0 1 1 0 1 0 0 1 0 1 1 1 1	0 0 0 1 1 0 1 0 1 0 0 1 1 1 1
0 0 0 1 0 0 1 1 1 1 0 1 1 0 1	0 0 0 1 0 1 1 1 1 0 0 1 1 0 1	0 0 0 1 0 1 1 1 1 0 1 1 0 0 1
0 0 0 1 0 0 1 1 1 1 0 1 0 1 1	0 0 0 1 0 1 1 1 1 0 0 1 0 1 1	0 0 0 1 0 1 1 1 1 0 1 0 0 1 1
0 0 0 1 0 0 1 1 0 1 1 1 1 0 1	0 0 0 1 0 1 1 0 0 1 1 1 1 0 1	0 0 0 1 0 1 1 0 1 1 1 1 0 0 1
0 0 0 1 0 0 1 1 0 1 0 1 1 1 1	0 0 0 1 0 1 1 0 0 1 0 1 1 1 1	0 0 0 1 0 1 1 0 1 0 0 1 1 1 1
0 0 0 1 0 0 1 0 1 1 1 1 0 1 1	0 0 0 1 0 1 0 0 1 1 1 1 0 1 1	0 0 0 1 0 1 0 1 1 1 1 0 0 1 1
0 0 0 1 0 0 1 0 1 1 0 1 1 1 1	0 0 0 1 0 1 0 0 1 1 0 1 1 1 1	0 0 0 1 0 1 0 1 1 0 0 1 1 1 1

36 = # of occurrences of 1101 in all $r_4(=36)$ run sequences of length 15

Average occurrence: n=4

4-tuples	*Total #*	4-tuple	Total #
0000	0	1000	36
0001	36	1001	36
0010	36	1010	36
0011	36	1011	36
0100	36	1100	36
0101	36	**1101**	36
0110	36	1110	36
0111	36	1111	36

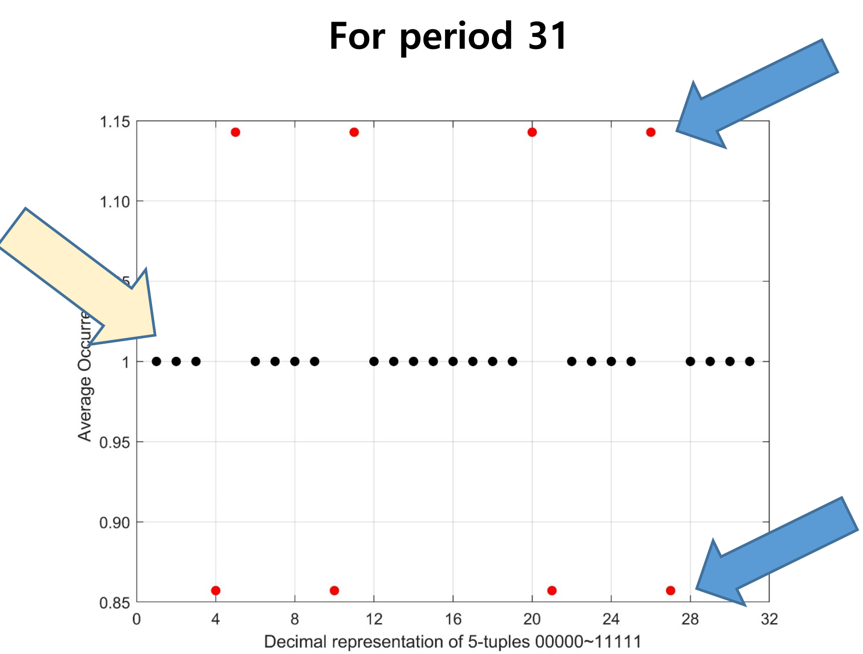
Normalizing

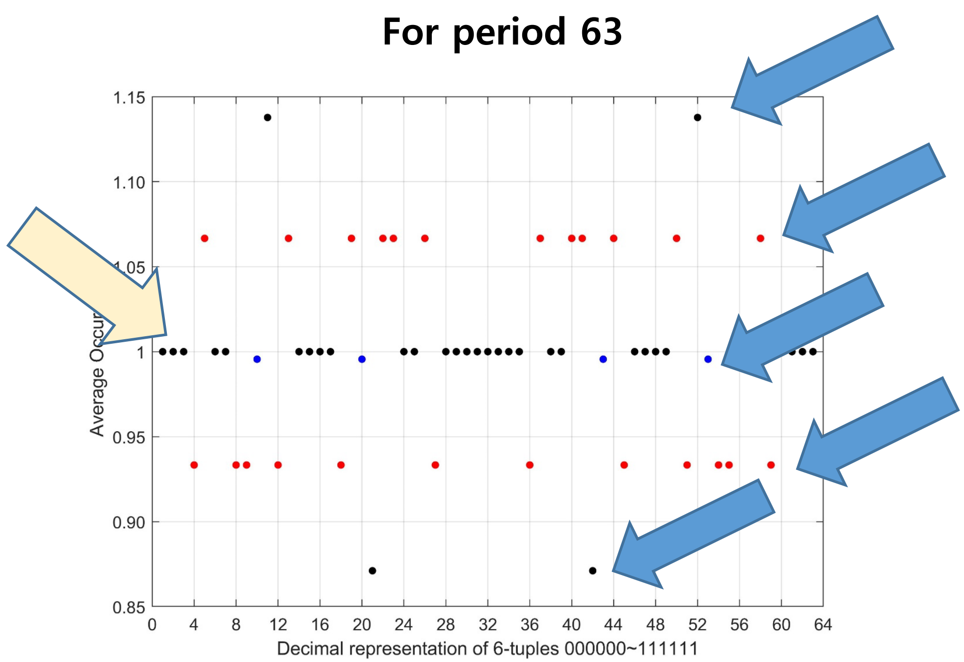
Divided by r_4

4-tuples	*Ave. #*	4-tuple	*Ave. #*
0000	0	1000	1
0001	1	1001	1
0010	1	1010	1
0011	1	1011	1
0100	1	1100	1
0101	1	**1101**	1
0110	1	1110	1
0111	1	1111	1

- Now, we call this normalizing # of occurrences
  Average occ. number or ***Ave. #***
- For the period 15, every nonzero 4-tuple has Ave# 1.
- **Is it always 1 for any other period $2^n - 1$ for $n > 4$?**

Not every n-tuple has the average number 1





Which n-tuples have the average number 1?

Lower Bound on the average number

Proposition 1

The average number of every nonzero n-tuple is at least

$$\frac{s_n}{r_n}$$

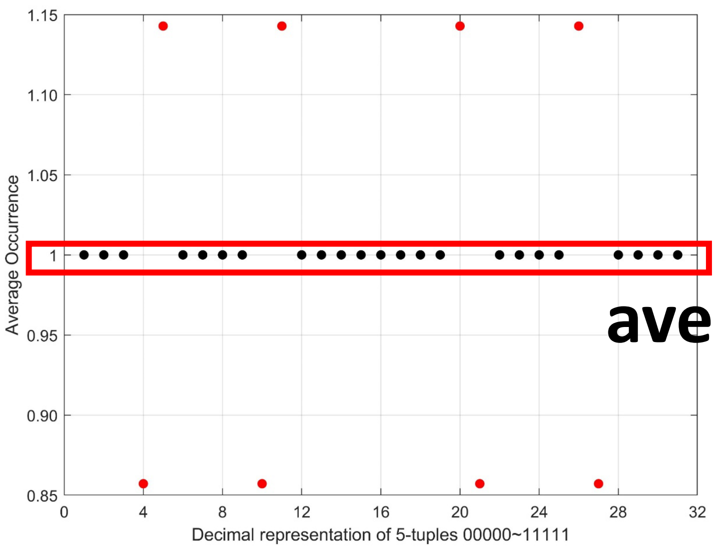
- Almost trivial. So small and tiny.

$\rightarrow$ 1 *over* $c2^{2^{n-1}-\frac{1}{2}n^2+\left(\frac{5}{2}-\log_2\pi\right)n}$

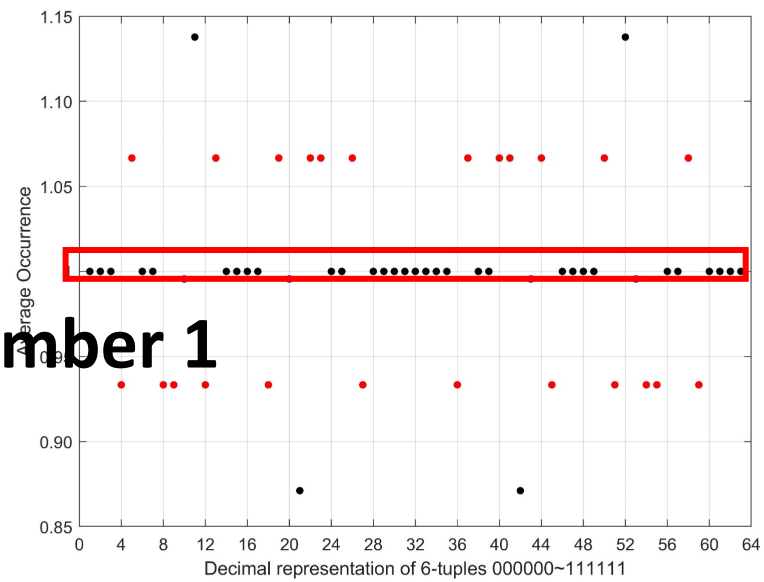
- It is NOT zero at least.

Average number 1

For period 31



For period 63



average number 1

Why and How do these n-tuples have the average number 1?

Two kinds of Average number 1

Recall the following $r_4(=36)$ run sequences of period 15

0 0 0 1 1 1 1 0 0 1 1 0 1 0 1	0 0 0 1 1 1 1 0 1 1 0 0 1 0 1	0 0 0 1 1 1 1 0 1 1 0 1 0 0 1
0 0 0 1 1 1 1 0 0 1 0 1 1 0 1	0 0 0 1 1 1 1 0 1 0 0 1 1 0 1	0 0 0 1 1 1 1 0 1 0 1 1 0 0 1
0 0 0 1 1 1 1 0 0 1 0 1 0 1 1	0 0 0 1 1 1 1 0 1 0 0 1 0 1 1	0 0 0 1 1 1 1 0 1 0 1 0 0 1 1
0 0 0 1 1 0 0 1 1 1 1 0 1 0 1	0 0 0 1 1 0 1 1 1 1 0 0 1 0 1	0 0 0 1 1 0 1 1 1 1 0 1 0 0 1
0 0 0 1 1 0 0 1 0 1 1 1 1 0 1	0 0 0 1 1 0 1 0 0 1 1 1 1 0 1	0 0 0 1 1 0 1 0 1 1 1 1 0 0 1
0 0 0 1 1 0 0 1 0 1 0 1 1 1 1	0 0 0 1 1 0 1 0 0 1 0 1 1 1 1	0 0 0 1 1 0 1 0 1 0 0 1 1 1 1
0 0 0 1 0 0 1 1 1 1 0 1 1 0 1	0 0 0 1 0 1 1 1 1 0 0 1 1 0 1	0 0 0 1 0 1 1 1 1 0 1 1 0 0 1
0 0 0 1 0 0 1 1 1 1 0 1 0 1 1	0 0 0 1 0 1 1 1 1 0 0 1 0 1 1	0 0 0 1 0 1 1 1 1 0 1 0 0 1 1
0 0 0 1 0 0 1 1 0 1 1 1 1 0 1	0 0 0 1 0 1 1 0 0 1 1 1 1 0 1	0 0 0 1 0 1 1 0 1 1 1 1 0 0 1
0 0 0 1 0 0 1 1 0 1 0 1 1 1 1	0 0 0 1 0 1 1 0 0 1 0 1 1 1 1	0 0 0 1 0 1 1 0 1 0 0 1 1 1 1
0 0 0 1 0 0 1 0 1 1 1 1 0 1 1	0 0 0 1 0 1 0 0 1 1 1 1 0 1 1	0 0 0 1 0 1 0 1 1 1 1 0 0 1 1
0 0 0 1 0 0 1 0 1 1 0 1 1 1 1	0 0 0 1 0 1 0 0 1 1 0 1 1 1 1	0 0 0 1 0 1 0 1 1 0 0 1 1 1 1

1101 : have average number 1 ⟶ *Average number 1, non-uniformly*

1001: exactly once in EVERY run sequence ⟶ *Average number 1, uniformly*

Average number 1 uniformly

Proposition 3.

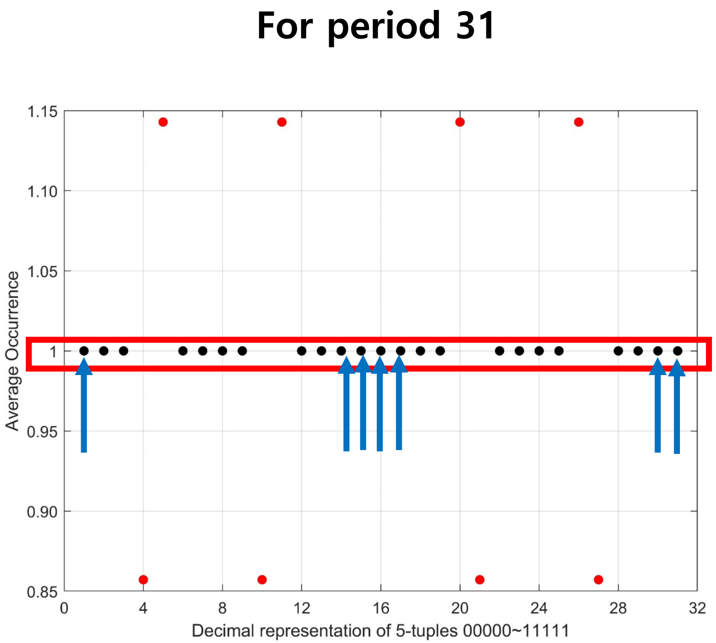
An n-tuple has the average number 1 **uniformly** if and only if it is one of the following **seven cases**:

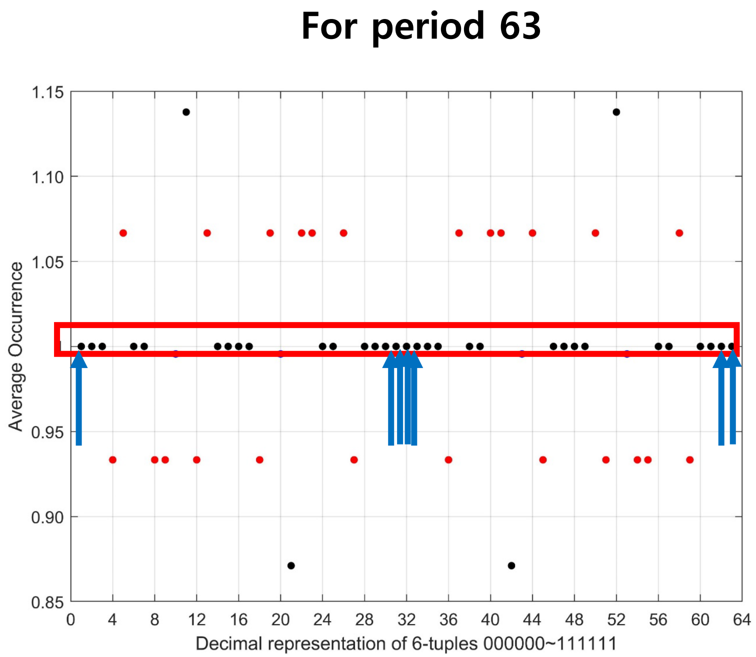
$$a0_{n-2}b \quad \text{and} \quad \text{a}1_{n-2}b,$$

where $a, b \in \{0,1\}$, **except for** the all-zero vector.

- **Easy:** Those n-tuples have the average number 1 uniformly.
- **Hard:** Only those n-tuples have the average number 1 non-uniformly.

Examples





There are only 7 n-tuples which have average number 1 uniformly.

Ex) 00001, 10000, 10001, 01110, 01111, 11110, 11111 for period 31

000001, 100000, 100001, 011110, 011111, 111110, 111111 for period 63

What happens for the remaining n-tuples with the average number 1?

We have to

and we are able to count

the average numbers

of ALL the n-tuples

Lemma 1: Formula of the average number

Classify all the n-tuples into FOUR cases as follows:

Case 1: Starting 0 … Ending 1

$$\mathbf{0}_x\mathbf{1}_{\alpha_1}\mathbf{0}_{\beta_1}\mathbf{1}_{\alpha_2}\mathbf{0}_{\beta_2}\cdots\mathbf{1}_{\alpha_m}\mathbf{0}_{\beta_m}\mathbf{1}_y$$

$$M\cdot\frac{2^{n-x-1}-\sum_{k=x}^{n-2}z_k}{2^{n-2}-\sum_{k=1}^{n-2}z_k}\cdot\frac{2^{n-y-1}-\sum_{k=y}^{n-2}p_k}{2^{n-2}-\sum_{k=1}^{n-2}p_k}\ \text{times}$$

* z_k: # of 0_k in

$$\cancel{\mathbf{0}_x}\ \boxed{\mathbf{1}_{\alpha_1}\mathbf{0}_{\beta_1}\mathbf{1}_{\alpha_2}\mathbf{0}_{\beta_2}\cdots\mathbf{1}_{\alpha_m}\mathbf{0}_{\beta_m}}\ \cancel{\mathbf{1}_y}$$

* p_k: # of 1_k in

$$\text{and}\quad M\triangleq\frac{1}{r_n}\times\frac{\left(2^{n-2}-\sum_{k=1}^{n-2}z_k\right)!}{\prod_{k=1}^{n-2}(2^{n-2-k}-z_k)!}\times\frac{\left(2^{n-2}-\sum_{k=1}^{n-2}p_k\right)!}{\prod_{k=1}^{n-2}(2^{n-2-k}-p_k)!}.$$

Case 2: Starting 1 … Ending 0

$$\mathbf{1}_y\mathbf{0}_{\beta_1}\mathbf{1}_{\alpha_1}\mathbf{0}_{\beta_2}\mathbf{1}_{\alpha_2}\cdots\mathbf{0}_{\beta_m}\mathbf{1}_{\alpha_m}\mathbf{0}_x$$

Case 3: Starting 0 … Ending 0

$$\mathbf{0}_x\mathbf{1}_{\alpha_1}\mathbf{0}_{\beta_1}\mathbf{1}_{\alpha_2}\mathbf{0}_{\beta_2}\cdots\mathbf{1}_{\alpha_m}\mathbf{0}_y$$

Case 4: Starting 1 … Ending 1

$$\underbrace{\mathbf{1}_y\mathbf{0}_{\beta_1}\mathbf{1}_{\alpha_1}\mathbf{0}_{\beta_2}\mathbf{1}_{\alpha_2}\cdots\mathbf{0}_{\beta_m}\mathbf{1}_x}_{n\text{-tuples}}$$

Example

$$r_5 = \frac{1}{2^3}\left(\frac{2^3!}{2^2!\,2^1!\,2^0!}\right)^2 = 88200$$

- **5-tuple 00011:**
  $p_k = z_k = 0$ for all k and $x = 3, y = 2.$
  The average number becomes

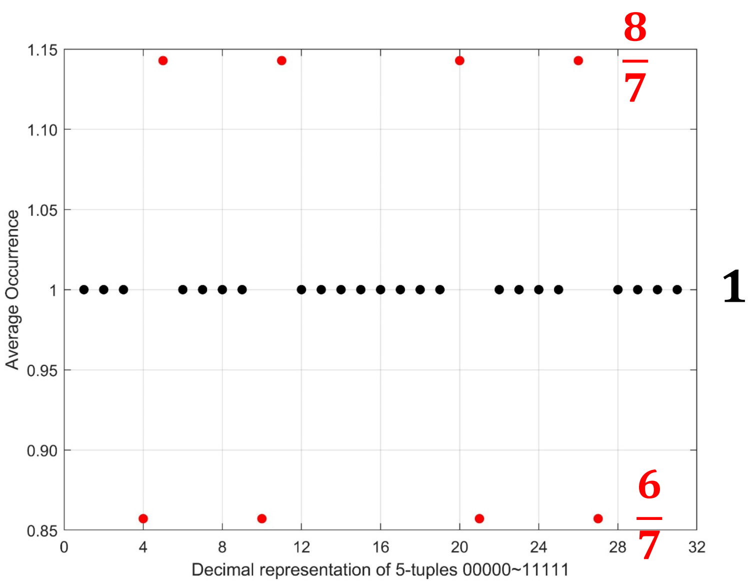
$$\frac{1}{8}M = \frac{1}{8r_5}\left(\frac{2^3!}{2^2!\,2^1!\,2^0!}\right)^2 = 1$$

- **5-tuple 01011:**
  $p_1 = z_1 = 1$ and $x = 1, y = 2.$
  The average number becomes

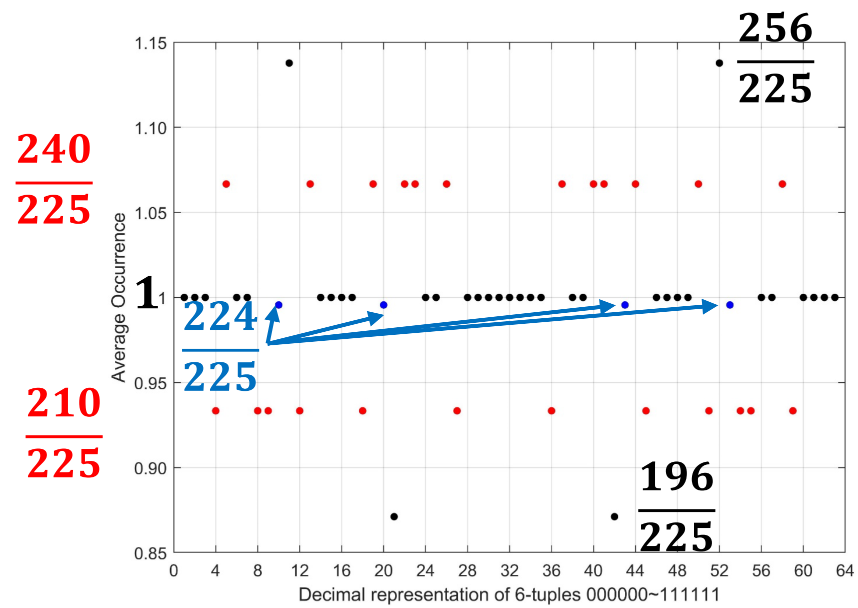
$$\frac{4}{7}M = \frac{4}{7r_5}\left(\frac{(2^3-1)!}{(2^2-1)!\,2^1!\,2^0!}\right)^2 = \frac{8}{7}$$

Examples

For period 31



For period 63



- We can calculate the average number of EVERY n-tuple
- We can identify all the n-tuples with average number 1

n-tuples with average number 1

Proposition 4.

The average number is 1 only for the following n-tuples:

$$a0_k1_{n-2-k}b \quad \text{and} \quad a1_{n-2-k}0_kb,$$

where $a, b \in \{0, 1\}$ and $k = 0, 1, \ldots, n-2$,

except for the all-zero n-tuple.

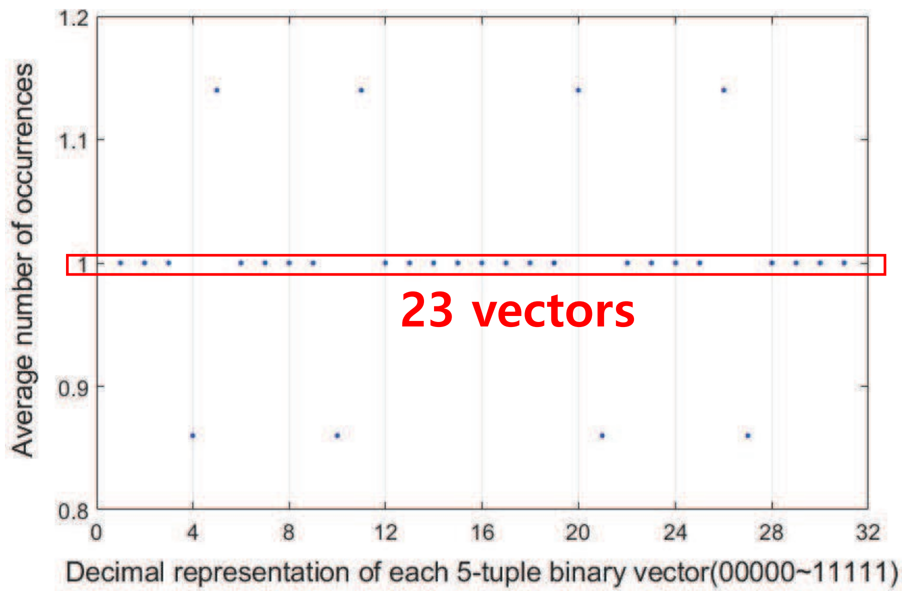
- **Easy:** Those n-tuples have the average number 1.
- **Hard:** Only those n-tuples have the average number 1.

Example: $n = 5$

The only 23 5-tuples of with average number 1

00011, 00111, 11100, 11000,
00010, 00110, 11101, 11001,
10011, 10111, 01100, 01000,
10010, 10110, 01101, 01001
non-uniform, **Proposition 4**.

00001, 10000, 10001, 01110,
01111, 11110, 11111
uniform, **Propositions 3, 4.**

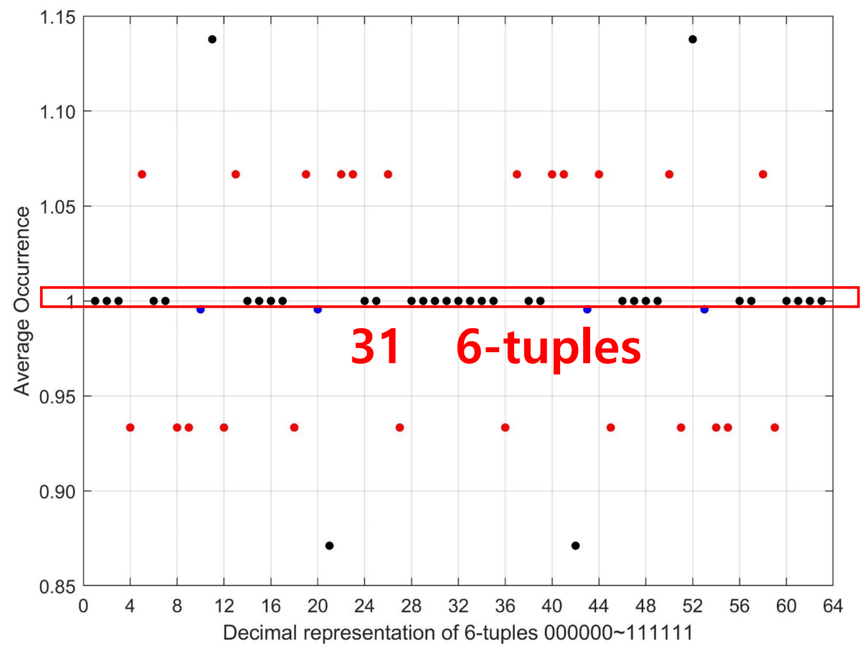


Example: $n = 6$

The only 31 6-tuples

000011, 000111, 001111, 111100, 111000, 110000, 000010, 000110, 001110, 111101, 111001, 110001, 100011, 100111, 101111, 011100, 011000, 010000, 100010, 100110, 100110, 011101, 011001, 010001
non-uniform, Proposition 4.

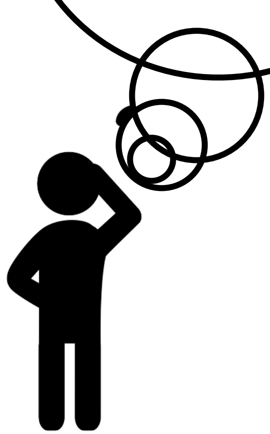
000001, 100000, 100001, 011110, 011111, 111110, 111111
uniform, Propositions 3, 4.



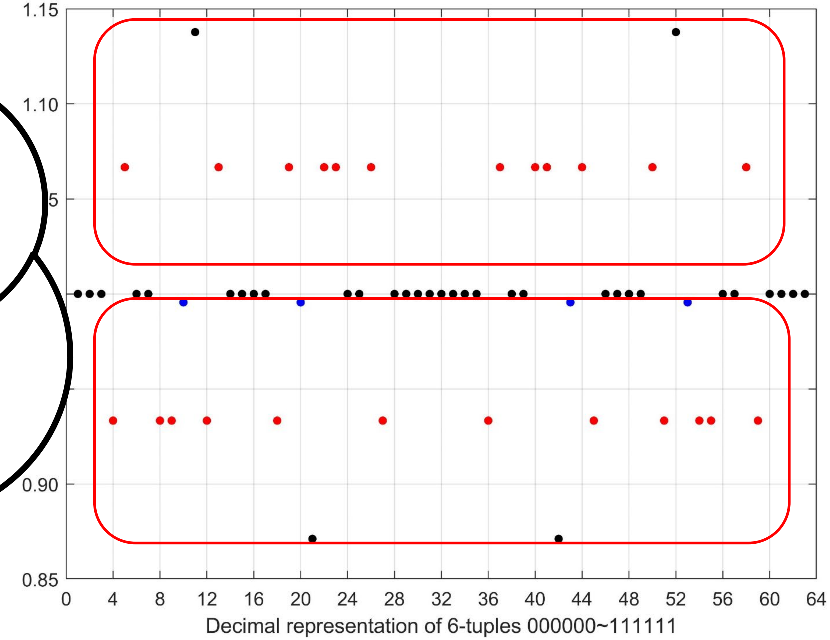
Statistical span property(1)

Does a run sequence have a *span property at all* ?

There are still many n-tuples whose average number is **not** 1.



$n = 6$

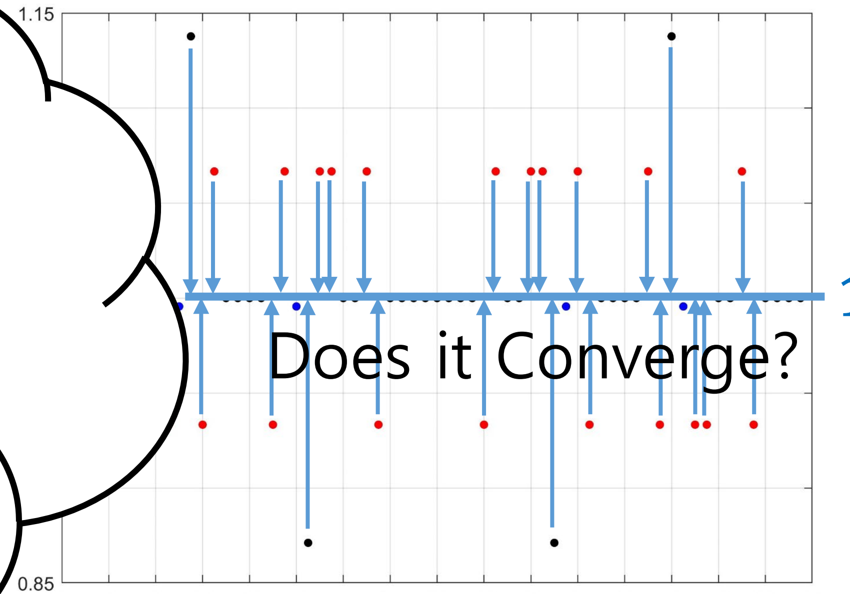


Statistical span property(2)

!

Perhaps all the average numbers converge to 1 as n increases?

as n increases



The answer is:

YES!

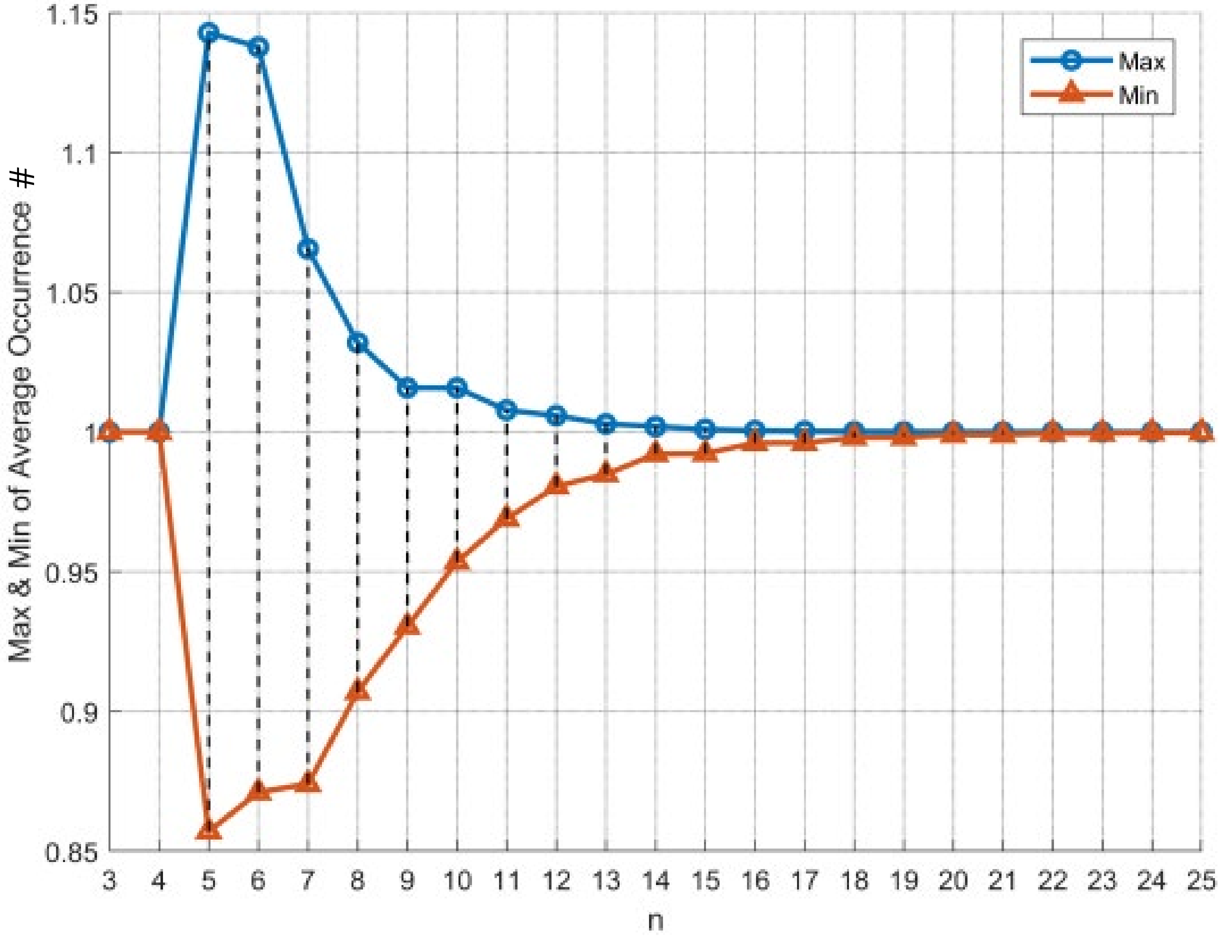
Statistical span property(3)

Theorem 2.

The average number of ANY n-tuple approaches to 1 as n increases.

statistical span property **of run sequences**

Statistical span property(4)



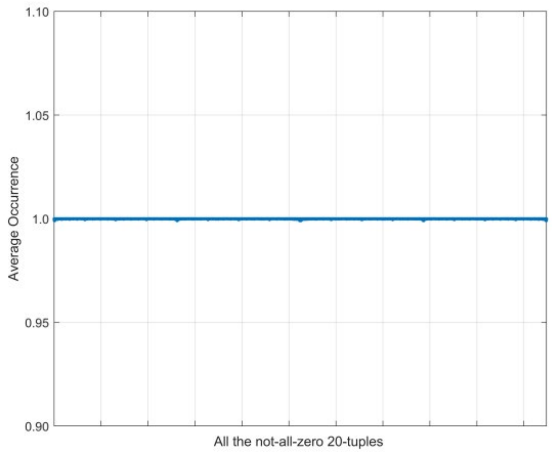
Three “open” problems

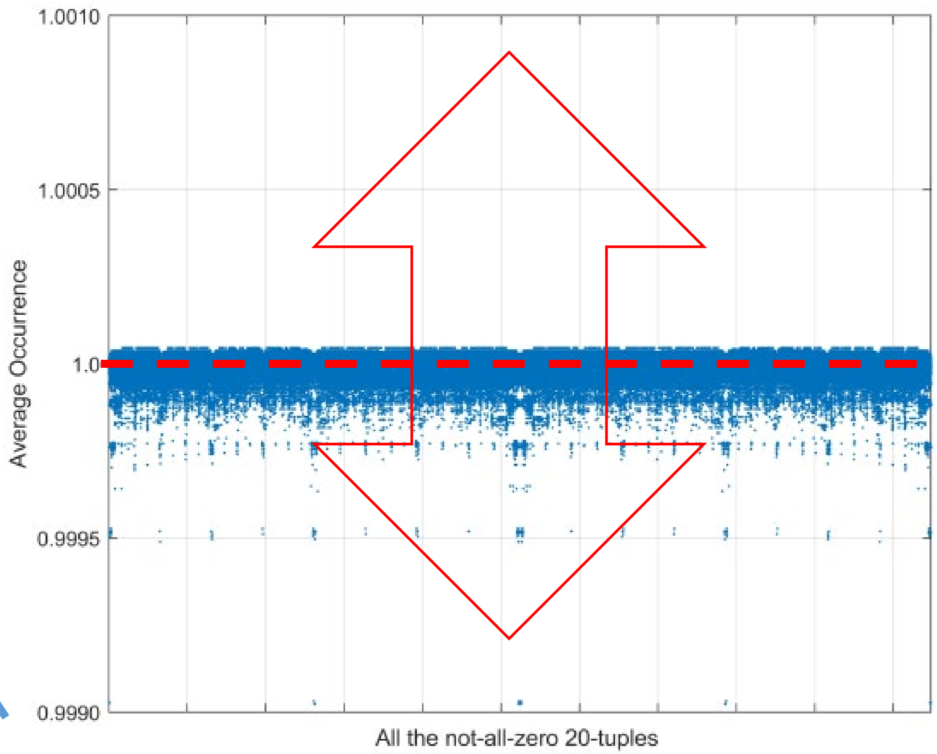
and

Summary

Open Problem #1

Example: n=20





Why are these so extremely non-symmetric?

Open Problem #2

Distribution of occurrences in a run sequence of period 31

	# occurrences							# sequences
	0	1	2	3	4	5	···	(total = 88200)
Distribution 0	0	31	0	0	0	0	···	2048 (span)
Distribution 1	2	27	2	0	0	0	···	6144
Distribution 2	4	23	4	0	0	0	···	26112
Distribution 3	6	19	6	0	0	0	···	33920
Distribution 4	7	18	5	1	0	0	···	6912
Distribution 5	8	17	4	2	0	0	···	1152
Distribution 6	8	15	8	0	0	0	···	7680
Distribution 7	10	13	6	2	0	0	···	2304
Distribution 8	10	11	10	0	0	0	···	1536
Distribution 9	12	7	12	0	0	0	···	320
Distribution 10	14	7	8	0	2	0	···	72
Average distribution	5.47	20.22	5.15	0.157	0.00163	0	···	sum = 31.00
Normalized average	0.176	0.652	0.166	0.005	0.00005	0	···	sum = 1.00

Is this distribution going to converge to something at all as n increases ?

In a random run sequence of period 31,

- **5.47** 5-tuples do not appear at all,
- **20.22** of them appear exactly 1 time,
- **5.15** of them appear exactly 2 times,
- etc.

$$5.47 = \frac{2 \cdot 6144 + 4 \cdot 26112 + \cdots + 14 \cdot 72}{88200}$$

Open Problem #3

0 0 0 1 1 1 1 0 0 1 1 0 1 0 1 Run but not span

0 0 0 1 1 1 1 0 0 1 0 1 1 0 1 Span but not m-sequence

0 0 0 1 1 1 1 0 1 0 1 1 0 0 1 m-sequence

Can we identify a series of such permutations (swapings) which transform a given run sequence into a span sequence or an m-sequence ?

Summary

- The number of cyclically distinct binary run sequences of period $2^n - 1$:

$$r_n = \frac{1}{2^{n-2}} \binom{2^{n-2}}{2^{n-3}, 2^{n-4}, \ldots, 2^0, 1}^2$$

- Derive formula for average number of every n-tuple
  - ✓ Only forms having **average number** 1 (uniform or non-uniform)

    $$a0_k 1_{n-2-k} b \text{ and } a1_{n-2-k} 0_k b,$$

  - ✓ Only forms having **average number** 1 **uniformly**

    $$a0_{n-2}b \quad \text{and} \quad a1_{n-2}b$$

- All the average numbers converge to 1 as n increases: ***statistical span property of run sequences***
- **Three open problems**
  1. **Why is this distribution of the average numbers so extremely non-symmetric about 1?**
  2. **Is the distribution of n-tuples in a random run sequence going to converge at all to any specific probability distribution?**
  3. **Is there a series of swapping that transforms a run sequence into a span sequence or an m-sequence?**