THE NUMBER OF RUNS OF WORDS ON A 2-LETTER ALPHABET

RICHARD J. MATHAR

Abstract. The number of runs of words with n letters equal to the first letter of a binary alphabet and m letters equal to the second letter of a binary alphabet is shown to be

$$R(n,m) = (1 + \frac{2mn}{n+m})\binom{n+m}{n}.$$

1. Notations

A word of a binary alphabet $\{0,1\}$ contains n letters of the first letter in the alphabet and m letters of the second letter of the alphabet. The words have length $n+m$. Distributing first the n letters in all possible ways over the places shows with a basic combinatorial argument that there are $\binom{n+m}{n} = \binom{n+m}{m}$ binary words.

Remark 1. *These are all shuffles of a word which contains n times the first letter of the alphabet by the word which contains m times the second letter of the alphabet.*

A *run* of letters in a word is a block of the same letters at adjacent positions in a word.

Example 1. *The word 00100011 has 4 runs, one of length 2, one of length 1, one of length 3 and one of length 2.*

Example 2. *The for $n = m = 1$ the words are 01 (2 runs) and 10 (2 runs) with a total of 4 runs.*

Example 3. *The for $n = 2$, $m = 1$ the words are 001 (2 runs), 010 (3 runs), and 100 (2 runs), with a total of 7 runs.*

Our aim is to construct the statistics $R(n,m)$, the total number of runs summed over all binary words of a given pair of $0 \le n, m$ [1, A349147].

2. Basics

If either $n = 0$ or $m = 0$, all words (actually: *the* word) have a single run of the same letter:

$$R(0,m) = 1. \tag{1}$$

If the role of the two letters in the alphabet is changed, the new words are also bijected to the old words, but this does not change the run lengths or the number or runs in the words, so

$$R(n,m) = R(m,n) \tag{2}$$

Date: November 9, 2021.

2020 *Mathematics Subject Classification.* Primary 68M15; Secondary 05A05.

and it suffices to tabulate results for the triangle of indices $0 \le m \le n$.

If $n = 1$, the statistics contains

- one word where the first letter of the alphabet is the first letter in the word (run length 2),
- contains $m-1$ words where the first letter of the alphabet is right to the k-th occurrence of the second letter of the alphabet, $1 \le k < m$ (run length 3)
- and one word where the last letter in the word is the first letter in the alphabet (run length 2),

so

$$R(1,m) = 2 + (m-1) \times 3 + 2 = 1 + 3m. \tag{3}$$

3. Recurrences

Theorem 1. *The number of compositions of n into p parts is $C_p^n \equiv \binom{n-1}{p-1}$.*

A binary word is an interlacing of blocks of the first and second letter of the alphabet. The run lengths of the first letter of the alphabet in a word are a composition of n in p_1 parts, $1 \le p_1 \le n$. The run lengths of the second letter of the alphabet in a word are a composition of m in p_2 parts, $1 \le p_2 \le m$. The number of runs in a word is $p_1 + p_2$.

In the following we shall not look at the cases where $n = 0$ or $m = 0$, so there are no "underflows" of some counts in the formulas.

There are 4 classes of binary words:

(1) The first letter and last letter of the word is the first letter of the alphabet, so $p_1 \ge 2$. The number of parts in the run lengths are $p_2 = p_1 - 1$.
(2) The first letter of the word is the first letter of the alphabet, so $p_1 \ge 1$, and the last letter of the word is the second letter of the alphabet; the number of parts in the run lengths are $p_2 = p_1$.
(3) The first letter of the word is the second of the alphabet and the last letter of the word is the first letter of the alphabet, so $p_1 \ge 1$; the number of parts in the run lengths are $p_2 = p_1$.
(4) The first letter and last letter of the word are the second letter of the alphabet, so $p_1 \ge 0$; the number of parts in the run lengths are $p_2 = p_1 + 1$.

Each binary word is entirely characterized by being in one of these four classes plus the two compositions of m and n. With respect to computation of R, each word is entirely characterized by being in one of these four classes plus the sum $p_1 + p_2$ of the number of parts.

In the four classes

(1) Each word contributes $p_1 + p_2 = 2p_1 - 1$ to R. There are $C_{p_1}^n$ compositions of n and $C_{p_2}^m = C_{p_1-1}^m$ compositions of m, so the for a given p_1 there are $(2p_1 - 1)C_{p_1}^n C_{p_1-1}^m$ runs. The runs in all words of this class are $\sum_{p_1=2}^n (2p_1 - 1)C_{p_1}^n C_{p_1-1}^m$.
(2) Each word contributes $p_1 + p_2 = 2p_1$ to R. There are $C_{p_1}^n$ compositions of n and $C_{p_2}^m = C_{p_1}^m$ compositions of m, so the for a given p_1 there are $2p_1 C_{p_1}^n C_{p_1}^m$ runs. The runs in all words of this class are $\sum_{p_1=1}^n 2p_1 C_{p_1}^n C_{p_1}^m$.

(3) Each word contributes $p_1+p_2=2p_1$ to R. There are $C^n_{p_1}$ compositions of n and $C^m_{p_2}=C^m_{p_1}$ compositions of m, so the for a given p_1 there are $2p_1C^n_{p_1}C^m_{p_1}$ runs. The runs in all words of this class are $\sum_{p_1=1}^n 2p_1C^n_{p_1}C^m_{p_1}$.

(4) Each word contributes $p_1+p_2=2p_1+1$ to R. There are $C^n_{p_1}$ compositions of n and $C^m_{p_2}=C^m_{p_1+1}$ compositions of m, so the for a given p_1 there are $(2p_1+1)C^n_{p_1}C^m_{p_1+1}$ runs. The runs in all words of this class are $\sum_{p_1=0}^n(2p_1+1)C^n_{p_1}C^m_{p_1+1}$.

Summing over all words of the four classes the total number of runs is

(4)
$$R(n,m)=\sum_{p_1=2}^{n}(2p_1-1)C^n_{p_1}C^m_{p_1-1}+4\sum_{p_1=1}^{n}p_1C^n_{p_1}C^m_{p_1}+\sum_{p_1=0}^{n}(2p_1+1)C^n_{p_1}C^m_{p_1+1}$$
$$=\sum_{p_1=2}^{n}(2p_1-1)\binom{n-1}{p_1-1}\binom{m-1}{p_1-2}+4\sum_{p_1=1}^{n}p_1\binom{n-1}{p_1-1}\binom{m-1}{p_1-1}+\sum_{p_1=0}^{n}(2p_1+1)\binom{n-1}{p_1-1}\binom{m-1}{p_1}$$
$$=\sum_{p_1=0}^{n-2}(2p_1+3)\binom{n-1}{p_1+1}\binom{m-1}{p_1}+4\sum_{p_1=0}^{n-1}(p_1+1)\binom{n-1}{p_1}\binom{m-1}{p_1}+\sum_{p_1=0}^{n}(2p_1+1)\binom{n-1}{p_1-1}\binom{m-1}{p_1}.$$

The terms are simplified with [2],

$$\sum_{p=0}^{n-2}\binom{n-1}{p+1}\binom{m-1}{p}=\binom{n+m-2}{m};\quad n+m>1; \tag{5}$$

$$\sum_{p=0}^{n}\binom{n-1}{p}\binom{m-1}{p}=\binom{n+m-2}{m-1};\quad n\geq 1; \tag{6}$$

$$\sum_{p=0}^{n}\binom{n-1}{p-1}\binom{m-1}{p}=\binom{n+m-2}{m-2};\quad n,m\geq 0. \tag{7}$$

$$\begin{aligned}\sum_{p=0}^{n-2}p\binom{n-1}{p+1}\binom{m-1}{p}&=\sum_{p=1}^{n-2}p\binom{n-1}{p+1}\frac{(m-1)!}{(m-1-p)!p!}\\&=\sum_{p=1}^{n-2}\binom{n-1}{p+1}\frac{(m-1)!}{(m-2-(p-1))!(p-1)!}\\=(m-1)\sum_{p=1}^{n-2}\binom{n-1}{p+1}\frac{(m-2)!}{(m-2-(p-1))!(p-1)!}&=(m-1)\sum_{p=1}^{n-2}\binom{n-1}{p+1}\binom{m-2}{p-1}\\&=(m-1)\sum_{p=0}^{n-3}\binom{n-1}{p+2}\binom{m-2}{p}=(m-1)\binom{n+m-3}{m};\quad n+m>2.\end{aligned} \tag{8}$$

$$\sum_{p=0}^{n-1} p\binom{n-1}{p}\binom{m-1}{p} = \sum_{p=1}^{n-1} p\binom{n-1}{p}\frac{(m-1)!}{(m-1-p)!p!}$$
$$= \sum_{p=1}^{n-1}\binom{n-1}{p}\frac{(m-1)!}{(m-2-(p-1))!(p-1)!}$$
$$= (m-1)\sum_{p=1}^{n-1}\binom{n-1}{p}\frac{(m-2)!}{(m-2-(p-1))!(p-1)!} = (m-1)\sum_{p=1}^{n-1}\binom{n-1}{p}\binom{m-2}{p-1}$$
$$= (m-1)\sum_{p=0}^{n-2}\binom{n-1}{p+1}\binom{m-2}{p} = (m-1)\binom{n+m-3}{m-1};\quad n+m>2. \tag{9}$$

$$\sum_{p=0}^{n} p\binom{n-1}{p-1}\binom{m-1}{p} = \sum_{p=1}^{n} p\binom{n-1}{p-1}\frac{(m-1)!}{(m-1-p)!p!}$$
$$= \sum_{p=1}^{n}\binom{n-1}{p-1}\frac{(m-1)!}{(m-2-(p-1))!(p-1)!}$$
$$= (m-1)\sum_{p=1}^{n}\binom{n-1}{p-1}\frac{(m-2)!}{(m-2-(p-1))!(p-1)!} = (m-1)\sum_{p=1}^{n}\binom{n-1}{p-1}\binom{m-2}{p-1}$$
$$= (m-1)\sum_{p=0}^{n-1}\binom{n-1}{p}\binom{m-2}{p} = (m-1)\binom{n+m-3}{m-2};\quad n+m>2. \tag{10}$$

$n\backslash m$	0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	4									
2	1	7	18								
3	1	10	34	80							
4	1	13	55	155	350						
5	1	16	81	266	686	1512					
6	1	19	112	420	1218	2982	6468				
7	1	22	148	624	2010	5412	12804	27456			
8	1	25	189	885	3135	9207	23595	54483	115830		
9	1	28	235	1210	4675	14872	41041	101530	230230	486200	
10	1	31	286	1606	6721	23023	68068	179608	432718	967538	2032316

Table 1. $R(n,m)$, the sum over the run lengths of the $\binom{n+m}{n}$ binary words with frequency n of the first and frequency m of the second letter of the alphabet in each word. [1, A349147]

Plugging these expressions into (4) yields

(11)
$$\begin{aligned}
R = {} & 2(m-1)\binom{n+m-3}{m}+3\binom{n+m-2}{m}+4(m-1)\binom{n+m-3}{m-1}+4\binom{n+m-2}{m-1} \\
& + 2(m-1)\binom{n+m-3}{m-2} + \binom{n+m-2}{m-2} \\
= {} & 2(m-1)\binom{n+m-3}{m}+2(m-1)\binom{n+m-3}{m-1}+3\binom{n+m-2}{m}+3\binom{n+m-2}{m-1}+\binom{n+m-2}{m-1} \\
& + 2(m-1)\binom{n+m-3}{m-1} + 2(m-1)\binom{n+m-3}{m-2} + \binom{n+m-2}{m-2} \\
= {} & 2(m-1)\binom{n+m-2}{m} + 3\binom{n+m-2}{m} + 3\binom{n+m-2}{m-1} + \binom{n+m-2}{m-1} \\
& + 2(m-1)\binom{n+m-2}{m-1} + \binom{n+m-2}{m-2} \\
= {} & 2(m-1)\binom{n+m-1}{m}+3\binom{n+m-2}{m}+3\binom{n+m-2}{m-1}+\binom{n+m-2}{m-1}+\binom{n+m-2}{m-2} \\
= {} & 2(m-1)\binom{n+m-1}{m} + 3\binom{n+m-1}{m} + \binom{n+m-1}{m-1} \\
= {} & 2m\binom{n+m-1}{m}+\binom{n+m-1}{m}+\binom{n+m-1}{m-1} = 2m\binom{n+m-1}{m}+\binom{n+m}{m}.
\end{aligned}$$

(12)
$$R(n,m) = \binom{n+m}{n}\frac{2nm+n+m}{m+n}, \quad n+m \ge 1.$$

By dividing $R(n,m)$ through the number of all binary words one obtains the expectation value $1+\frac{2nm}{m+n}$ of the number of runs.

The anti-diagonal sum of the d-th diagonal is [1, A057711]

$$\sum_{m=0}^{d} R(d-m,m) = (d+1)2^{d-1} = 1,2,6,16,40,\ldots \tag{13}$$

The row sums in the triangle up to the diagonal are

$$\begin{aligned}
\sum_{m=0}^{n} R(n,m) &= \sum_{m=0}^{n} \frac{(n+m)!}{n!m!}\frac{2nm+n+m}{m+n} = \sum_{m=0}^{n} \frac{(n+m-1)!}{n!m!}(2nm+n+m) \\
&= \sum_{m=0}^{n} \frac{\Gamma(n+m)}{\Gamma(n+1)\Gamma(m+1)}(2nm+n+m) \\
&= 2\sum_{m=0}^{n} \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} + \sum_{m=0}^{n} \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m+1)} + \sum_{m=0}^{n} \frac{\Gamma(n+m)}{\Gamma(n+1)\Gamma(m)} \\
&= 2\frac{1}{\Gamma(n)}\frac{n\Gamma(1+2n)}{(n+1)\Gamma(n+1)} + \frac{1}{\Gamma(n)}\frac{(n+1)\Gamma(1+2n)}{n\Gamma(2+n)} + \frac{1}{\Gamma(n+1)}\frac{n\Gamma(1+2n)}{(n+1)\Gamma(n+1)} \\
&= 2\frac{1}{\Gamma(n)}\frac{n\Gamma(1+2n)}{\Gamma(n+2)} + \frac{1}{\Gamma(n)}\frac{\Gamma(1+2n)}{n\Gamma(1+n)} + \frac{1}{\Gamma(n)}\frac{\Gamma(1+2n)}{\Gamma(n+2)} \\
&= \frac{\Gamma(2+2n)}{\Gamma(n)\Gamma(n+2)} + \frac{\Gamma(1+2n)}{\Gamma^2(n+1)} = \frac{(1+2n)!}{(n-1)!(n+1)!} + \frac{(2n)!}{(n!)^2} \\
&= n\binom{2n+1}{n} + \binom{2n}{n} = 1,5,26,125,574,\ldots
\end{aligned} \tag{14}$$

4. Summary

The sum of the number of runs in binary words of length $n+m$ composed of n letters of the first letter in the alphabet and m letters of the second letter in the alphabet is given by (12).

References

1. O. E. I. S. Foundation Inc., *The On-Line Encyclopedia Of Integer Sequences*, (2021), https://oeis.org/. MR 3822822
2. Ranjan Roy, *Binomial identities and hypergeometric series*, Amer. Math. Monthly **94** (1987), no. 1, 36–46. MR 0873603

URL: https://www.mpia.de/~mathar

Max-Planck Institute of Astronomy, Königstuhl 17, 69117 Heidelberg, Germany