Runs of Ones in Binary Strings

Félix Balado and Guénolé Silvestre

December 20th, 2022

1 Introduction

A run of ones in a binary string is an uninterrupted sequence of ones flanked either by a zero or by the start/end of the string. In the following, a run of ones will be simply referred to as a "run".

What is the expected number of runs of length i in a binary n-sequence drawn uniformly at random? It may appear at first that the analysis is straightforward with some clever use of the exponential distribution with parameter $1/2$. But after a few attempts we can convince ourselves that such an analysis is not trivial.

However, it is much easier to determine the total number of runs of length i over all binary n-strings, which, through the law of large numbers, will also allow us to solve the problem above asymptotically for large n as a corollary.

This problem was previously solved by Sinha and Sinha [1] —in fact, these authors also solved a harder problem in [2] from which the solution to the problem addressed here can be produced.

Here we want to show that the problem can be solved in a shorter and simpler way than in [2] or [1] by using an elementary counting argument.

2 Counting Runs

Let $r_n(i)$ be the total number of runs of length i over all binary n-strings.

First of all, let us get some visual intuition. In the diagrams below, for $n = 2, 3$ and 4 we list all binary n-strings and, right underneath, their runs "spectra" (i.e. the number of runs of lengths 1 to n found in each particular n-string). On the right we show $r_n(i)$.

- $n = 2$:

	0	0	1	1	
	0	1	0	1	
i					$r_2(i)$
1	0	1	1	0	2
2	0	0	0	1	1

- $n = 3$:

	0	0	0	0	1	1	1	1	
	0	0	1	1	0	0	1	1	
	0	1	0	1	0	1	0	1	
i									$r_3(i)$
1	0	1	1	0	1	2	0	0	5
2	0	0	0	1	0	0	1	0	2
3	0	0	0	0	0	0	0	1	1

- $n = 4$:

	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	
	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	
i																	$r_4(i)$
1	0	1	1	0	1	2	0	0	1	2	2	1	0	1	0	0	12
2	0	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	5
3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	2
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Assuming $n > 1$, it is trivial to see that

$$\begin{aligned} r_n(n) &= 1, \\ r_n(n-1) &= 2. \end{aligned} \tag{1}$$

Of course, this solves the problem for $n = 2$. Assuming $n > 2$, let us next see how, for $1 \leq i < n-1$, $r_n(i)$ can be recursively obtained from $r_{n-1}(i)$:

- On the one hand, consider the n-strings that start with a zero: these trivially contribute $r_{n-1}(i)$ runs to $r_n(i)$.
- On the other hand, consider the n-strings that start with a one. For $1 \leq i \leq n-1$, the 2^{n-i-1} n-strings that start with i ones followed by at least one zero add 2^{n-i-1} runs to $r_{n-1}(i)$ if $i < n-1$, and subtract 2^{n-i-1} runs from $r_{n-1}(i-1)$ if $i > 1$.

  Thus, the n-strings that start with a one contribute $r_{n-1}(i) + 2^{n-i-1} - 2^{n-i-2} = r_{n-1}(i) + 2^{n-i-2}$ runs to $r_n(i)$.

So, collecting these two contributions we have that

$$r_n(i) = 2\,r_{n-1}(i) + 2^{n-i-2}. \tag{2}$$

Now, by using the expression above recursively k times, with $k < n-1$, we get

$$r_n(i) = 2^k r_{n-k}(i) + k\,2^{n-i-2}. \tag{3}$$

When $k = n-i-1$, we have that $r_{n-k}(i) = r_{i+1}(i) = 2$ because of (1). Thus, inputting this value of k in (3) we get the explicit expression

$$r_n(i) = (n-i+3)\,2^{n-i-2} \tag{4}$$

for $1 \leq i < n-1$. Incidentally, (4) is also valid when $i = n-1$, i.e. it includes (1). We can now see that $r_{n-1}(i-1) = r_n(i)$, and so recurrence (2) can also be expressed as a recurrence on i (rather than as a recurrence on n) as $r_n(i) = 2\,r_n(i+1) + 2^{n-i-2}$.

Finally, for large n, the average number of runs of each length in a binary n-sequence drawn uniformly at random is simply $\bar{r}_n(i) = r_n(i)\ 2^{-n}$. Also, the total number of runs over all n-strings is

$$t(n) = \sum_{i=1}^{n} r_n(i) = (n+1)\,2^{n-2}. \tag{5}$$

Therefore the total frequency of runs of each length is

$$f_n(i) = \frac{r_n(i)}{t(n)} = \frac{n-i+3}{n+1}\,2^{-i} \tag{6}$$

for $1 \leq i \leq n-1$, whereas $f_n(n) = 1/t(n)$. Interestingly, $f_n(2) = 1/4$ independently of $n > 2$.

For large n, (6) should approximate the frequency of runs of each length in one single binary n-sequence drawn uniformly at random, in which case $f_n(i) \approx 2^{-i}$.

3 Relationship to Sequences from OEIS

Sequence A045623 from OEIS [3] (number of 1's in all compositions of $j+1$) is defined by $a(j) = (j+3)\,2^{j-2}$ for $j \geq 1$. Observing (4), $r_n(i) = a(n-i)$.

Also, sequence A001729 is defined by $b(j) = (j+2)\,2^{j-1}$. From (5), $t(n) = b(n-1)$.

After noticing this connection, we found that the gist of the counting argument we have used in Section 2 was already known in the problem of finding the number of 1's in all compositions of $j+1$.

Acknowledgements

Thanks to Kevin Ryde for hinting the connection with A001729.

References

[1] K. Sinha and B.P. Sinha. Energy-efficient communication: Understanding the distribution of runs in binary strings. In *1st International Conference on Recent Advances in Information Technology (RAIT)*, pages 177–181, Dhanbad, India, January 2012.

[2] K. Sinha and B.P. Sinha. On the distribution of runs of ones in binary strings. *Computers & Mathematics with Applications*, 58(9):1816–1829, 2009.

[3] N.J.A. Sloane and The OEIS Foundation Inc. The on-line encyclopedia of integer sequences, 2020.