Homework 8: Question and Answer

Maryam Bahrani (`mbahrani`)

Runs in random binary strings

A *run* in a binary string is a maximal sequence of consecutive identical bits. For example, the string 11010100001 has 7 runs, with lengths 2, 1, 1, 1, 1, 4, 1 respectively.

What is the average number of runs in a random binary string of length n? List the corresponding horizontal and vertical OBGFs. Is the distribution of the number of runs in binary strings concentrated?

Answer

In order to mark the runs or their bits, we use the observation that a bit string can be thought of as a sequence of alternating non-empty runs of zeros and ones, either starting with a run of zeros or starting with a run of ones.

$$\textsc{Seq}\left(\textsc{Seq}_{>0}\left(\mathcal{Z}_0\right)\right) + \textsc{Seq}\left(\textsc{Seq}_{>0}\left(\mathcal{Z}_1\right)\right) - E$$

Let u mark the runs in the OBGF $S(z,u)$:

$$\mathcal{S} = \textsc{Seq}\left(u\textsc{Seq}_{>0}\left(\mathcal{Z}_0\right)\right) + \textsc{Seq}\left(u\textsc{Seq}_{>0}\left(\mathcal{Z}_1\right)\right) - E$$

$$S(z,u) = \frac{2}{1-\frac{uz}{1-z}} - 1 = \frac{2(1-z)}{1-(u+1)z} - 1 = \frac{1+(u-1)z}{1-(u+1)z}.$$

Indeed, setting $u = 1$ retrieves the unmarked generating function for binary strings, from which enumerations can be extracted:

$$S(z,1) = \frac{1}{1-2z}$$

$$[z^n]S(z,1) = 2^n.$$

The horizontal OGF for the costs is

$$s_n(u) = [z^n]S(z,u) = [z^n]\frac{1}{1-(u+1)z} + [z^n]\frac{(u-1)z}{1-(u+1)z}$$

$$= (u+1)^n + (u-1)(u+1)^{n-1} = 2u(u+1)^{n-1}.$$

The cumulated cost is therefore

$$Q_N = s_n'(1) = 2(u+1)^{n-2}(nu+1)|_{u=1} = (n+1)2^{n-1},$$

and the average number of runs is

$$\frac{s_n'(1)}{s_n(1)} = \frac{(n+1)2^{n-1}}{2^n} = \frac{n+1}{2}.$$

Alternatively, we could have used the vertical OGF for the costs:

$$q_k(z) = [u^k]S(z,u) = [u^k]\frac{1}{1-\frac{z}{1-z}u} + \frac{\frac{z}{1-z}u}{1-\frac{z}{1-z}u} = 2\left(\frac{z}{1-z}\right)^k$$

$$Q_n = [z^n]\sum_k kq_k(z) = [z^n]\sum_k 2k\left(\frac{z}{1-z}\right)^k = 2[z^n](1-z)\sum_k k\frac{z^k}{(1-z)^{k+1}}$$

$$= n2^n - (n-1)2^{n-1} = (n+1)2^{n-1}$$

Furthermore, the variance is

$$\frac{s_n''(1)}{s_n(1)} + \frac{n}{2} - \left(\frac{n}{2}\right)^2 = \frac{(n-1)(n+2)2^{n-2}}{2^n} + \frac{n}{2} - \frac{n^2}{4} = \frac{3n-2}{4}.$$

The distribution is thus concentrated, since $\sigma = \theta(\sqrt{n}) = o(\mu)$.