Miranda Moore
COS 488/MAT 474
Problem Set 6, Q2

5/5

AofA Exercise 8.14 Suppose that a monkey types randomly at a 32-key keyboard. What is the expected number of characters typed before the monkey hits upon the phrase `THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG`?

Solution. Let p denote the given string, and note that the number of characters in p (including spaces) is 44. We follow the same steps as in the proof of Theorem 8.3, but using a 32-character alphabet rather than just 2 characters. Let $\mathcal{S}_p$ be the class of strings with no occurrence of the string p, and let $\mathcal{T}_p$ be the class of strings that end with p but have no other occurrence of p. We now set up two equations relating $\mathcal{S}_p$ and $\mathcal{T}_p$. First, we note that

$$\mathcal{S}_p + \mathcal{T}_p = \epsilon + \mathcal{S}_p \times (\mathcal{Z}_1 + \cdots + \mathcal{Z}_{32})$$

where the $\mathcal{Z}_i$ represent the 32 characters on the keyboard. What this means is that $\mathcal{S}_p$ and $\mathcal{Z}_p$ are disjoint, and every string in $\mathcal{S}_p + \mathcal{T}_p$ is either empty or it is an element of $\mathcal{S}_p$ with a character appended to the end. Translating this into a generating function equation, we have

$$S_p(z) + T_p(z) = 1 + 32zS_p(z).$$

Second, we note that by appending the string p onto the end of a string in $\mathcal{S}_p$, we get a string that is exactly in $\mathcal{T}_p$ (because p has no non-trivial autocorrelation, so the resulting string will only have one occurrence of p). In symbols,

$$\mathcal{S}_p \times p = \mathcal{T}_p.$$

The resulting generating function equation is

$$z^{44} S_p(z) = T_p(z),$$

where the power z^{44} comes from the fact that p has 44 characters. Solving these two equations simultaniously for $S_p(z)$, we obtain the generating function

$$S_p(z) = \frac{1}{z^{44} - 32z + 1}.$$

We now want to find the expected wait time before the first occurrence of p. For this, we note that

$$\begin{aligned}
S_p(z) &= \sum_{N=0}^{\infty} \{\#\text{ strings of length } N \text{ with no occurrence of } p\} z^N \\
S_p(\tfrac{1}{32}) &= \sum_{N=0}^{\infty} \{\#\text{ strings of length } N \text{ with no occurrence of } p\}/32^N \\
&= \sum_{N=0}^{\infty} \Pr\{\text{first } N \text{ characters of a random string have no occurrence of } p\} \\
&= \sum_{N=0}^{\infty} \Pr\{\text{position of end of first occurrence of } p \text{ is} > N\},
\end{aligned}$$

and this sum of cumulative probabilities is equal to the expectation. Therefore, the expected number of characters typed before the first occurrence of the string p is

$$S_p(\tfrac{1}{32}) = \frac{1}{(\frac{1}{32})^{44} - 32(\frac{1}{32}) + 1} = 32^{44} \approx 1.68 \times 10^{66}.$$