Miranda Moore
COS 488/MAT 474
Problem Set 7, Q1

5/5

AC Note I.23 *Alice, Bob, and coding bounds.* Alice wants to communicate n bits of information to Bob over a channel that transmits 0,1-bits but is such that any occurrence of 11 terminates the transmission. Thus, she can only send on the channel an encoded version of her message (where the code is of some length $\ell \geq n$) that does not contain the pattern 11.

Here is a first coding scheme: given the message $m = m_1 m_2 \cdots m_n$, where $m_j \in \{0, 1\}$, apply the substitution: $0 \mapsto 00$ and $1 \mapsto 10$; terminate the transmission by sending 11. This scheme has length $\ell = 2n + O(1)$, and we say that its *rate* is 2. Can one design codes with better rates? with rates arbitrarily close to 1, asymptotically?

Let $\mathcal{C}$ be the class of allowed code words. For words of length n, a code of length $L \equiv L(n)$ is achievable only if there exists a one-to-one mapping from $\{0, 1\}^n$ into $\cup_{j=0}^{L} \mathcal{C}_j$, i.e., only if $2^n \leq \sum_{j=0}^{L} C_j$. Find the OGF of $\mathcal{C}$ and use it to show that

$$L(n) \geq \lambda n + O(1), \qquad \text{where} \quad \lambda = \frac{1}{\log_2 \varphi} \doteq 1.440420, \quad \varphi = \frac{1+\sqrt{5}}{2}.$$

Thus no code can achieve a better rate than 1.44; i.e., a loss of at least 44% is unavoidable.

Solution. A better coding scheme is given by $0 \mapsto 0$ and $1 \mapsto 10$. It's clear by inspection that this coding scheme and its inverse are well-defined. For an arbitrary word of length n, the *expected* length of the encoded word is $1.5n + O(1)$, so the code has *expected* rate 1.5. However, the length of the encoded word can vary anywhere from n (if the word is all 0s) to $2n$ (if the word is all 1s).

We can't design codes with rates arbitrarily close to 1, as we will now prove. $\mathcal{C}$ is the class of bitstrings with no occurrence of 11. By symmetry, this class obviously has the same generating function as the class of bitstrings with no occurrence of 00, which we computed in part 1 of the course (see AofA p. 227). The OGF of $\mathcal{C}$ is thus

$$C(z) = \frac{1+z}{1-z-z^2} = F(z) + zF(z),$$

where $F(z)$ is the OGF of the Fibonacci numbers. The coefficients C_j are thus

$$C_j = F_j + F_{j+1} = F_{j+2}.$$

Nice

It is well-known (and easily verified by induction) that $\sum_{j=0}^{L} F_j = F_{j+2} - 1$. Summing the C_j from $j = 0$ to L, we thus have

$$\sum_{j=0}^{L} C_j = \sum_{j=0}^{L} F_{j+2} = F_{L+4} - 1 - F_0 - F_1 = F_{L+4} - 2.$$

Since $F_j \sim \frac{\varphi^j}{\sqrt{5}}$ as $j \to \infty$, this means that

$$\sum_{j=0}^{L} C_j = F_{j+4} - 2 \sim \frac{\varphi^{L+4}}{\sqrt{5}} - 2 \sim \frac{\varphi^{L+4}}{\sqrt{5}}.$$

We now solve for L as a function of n:

$$\begin{aligned} 2^n &\leq \sum_{j=0}^{L} C_j \sim \frac{\varphi^{L+4}}{\sqrt{5}} \\ \Rightarrow n &\leq \log_2(\varphi^L) + \log_2\left(\frac{\varphi^4}{\sqrt{5}}\right) \\ &= L\log_2(\varphi) + O(1) \\ \Rightarrow \frac{n}{\log_2(\varphi)} + O(1) &\leq L, \end{aligned}$$

as we wanted to show.