Exercise Sheet 3
Due 22.10.2020

Problem 1. We consider permutations on $[n]$, that is, bijective functions $\sigma\colon [n] \to [n]$. Recall that every permutation can be written as a product of cycles with disjoint support. E.g., for $n = 5$, the permutation $\sigma = (1\,2\,3)(4\,5)$ is the one with $f(1) = 2$, $f(2) = 3$, $f(3) = 1$ and swapping 4 and 5. (You should have encountered this notation before.)

A *derangement* is a permutation without fixed points, that is, $\sigma(i) \neq i$ for all $i \in [n]$. An *involution* is a permutation σ with $\sigma^2 = \mathrm{id}$.

(i) Let $\mathcal{D}$ be the (labeled) combinatorial class of derangements. Use the symbolic method to find the exponential generation function $D(z)$ of $\mathcal{D}$ and derive a formula for $[z^n]D(z)$. (*Hint:* You won't find a "closed form expression".)

(ii) Let $\mathcal{I}$ be the (labeled) combinatorial class of involutions. Use the symbolic method to find the exponential generation function $I(z)$ of $\mathcal{I}$ and derive a formula for $[z^n]I(z)$.

Problem 2. We consider words on the alphabet $X = \{a, b\}$. Recall that a word on X is an ordered, finite sequence of elements of X, e.g. $w = abaabaaaabbbaba$ is a word. A word contains a *run of length k* if it contains at least k consecutive occurrences of the letter a or the letter b. For example, the longest run in the word w above has length 4.

In the computer security community, humans are known to be notoriously bad at producing random numbers/words/password, ... (even if they try). For instance, humans tend to underestimate the probability of a long run in a random word; a human produced "random" word is therefore less likely to contain a long run than an actual random word of the same length.

(a) How likely is it that a word of length 250 contains a run of length 7 or more? Make a guess and write it down before continuing.

(b) Show: The OGF for words on $\{a, b\}$ whose longest run has length $\leq k$ is

$$W_{\leq k}(z) = \frac{1 - z^{k+1}}{1 - 2z + z^{k+1}} = \frac{1 + z + \cdots + z^k}{1 - z - \cdots - z^k}.$$

(*Hint:* analogously to $\mathsf{SEQ}(\mathcal{A})$ one can define $\mathsf{SEQ}_{\leq k}(\mathcal{A}) = \mathcal{E} + \mathcal{A} + \mathcal{A}^2 + \cdots + \mathcal{A}^k$, the class of sequences of length $\leq k$.)

(c) Use a computer algebra system (e.g., the free Sage, `http://www.sagemath.org`) to compute the probability that a random word of length 250 contains a run of length 7 or more. How does it compare to your guess?

Problem 3. Alice wants to communicate n bits of information to Bob over a channel (a wire, an optic fibre) 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 $l \geq n$) 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 $l = 2n + O(1)$, and we say that its rate is 2. Can one design codes with better rates? with rates arbitrarily close to 1, asymptotically?

Let C be the combinatorial class of allowed code words, and let $C(z) = \sum_{j\geq 0} c_j z^j$ be its generating function. For words of length n, a code of length $L = L(n)$ is achievable only if there exists a one-to-one mapping from $\{0, 1\}$ into $\bigcup_{j=0}^{L} C_j$, that is, $2^n \leq \sum_{j=0}^{L} c_j$. Determine $C(z)$ and use it to show that

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

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

Hint: We have done some groundwork for this already; cf. Example 1.3.5 in the notes.