Sarah Eugene - Exercices of Analytic Combinatorics

I - Smirnov words

Over an alphabet $\mathcal{A} = \{a_1, \dots, a_r\}$ of at least 3 letters, a Smirnov word is such that no two consecutive letters are the same.

$$\text{Exemple: on } \mathcal{A} = \{a, b, c\} \begin{cases} ababcba \text{ is a Smirnov word} \\ abaacba \text{ is not a Smirnov word} \end{cases}$$

Question 1? Find a surjective application of the set of all words $\mathcal{A}^\star = (a_1 + \cdots + a_r)^\star$ on the set S of Smirnov words built on the same alphabet.

Question 2? Find a combinatorial construction generating $\mathcal{A}^\star = (a_1 + \cdots + a_r)^\star$ from the Smirnov set S built over $\mathcal{A}$. (Hint: you have to replace some elements by a combinatorial class, and it is a reciprocal of the preceding question.)

Question 3? Express the generating function $A(z)$ of the words of $\mathcal{A}^\star$ as an expression involving the generating function $S(z)$ of the Smirnov words, where

$$A(a_1, \dots, a_r) = \sum_{w \in \mathcal{A}^\star} a_1^{|w|_{a_1}} \times \cdots \times a_r^{|w|_{a_r}} = \frac{1}{1 - a_1 - \cdots - a_r}$$
$$S(a_1, \dots, a_r) = \sum_{w \in S} a_1^{|w|_{a_1}} \times \cdots \times a_r^{|w|_{a_r}}$$

and $|w|_{a_i}$ is the number of letters a_i in the word w. (Example: $|aabbbcb|_b = 4$, $|aabbbcb|_a = 2$).

Question 4? By a set of algebraic substitutions $\{?_1, \dots, ?_r\}$, where $?_i$ and $?_j$ are equivalent up to a letter substitution ($a_i \mapsto a_j \rightsquigarrow ?_i \mapsto ?_j$), express $S(a_1, \dots, a_r)$ as a function of $A(?_1, \dots, ?_r)$. What is the algebraic function $?_i$

Question 5? By setting $a_i = z$, we count the number of Smirnov words by their lengths,

$$S_L(z) = \sum_{w \in S} a_1^{|w|_{a_1}} \times \cdots \times a_r^{|w|_{a_r}} \Bigg|_{a_1 = z, \dots, a_r = z} = \sum_{w \in S} z^{|w|}.$$

Compute $S_L(z)$ and prove that the number of Smirnov words of length n is $[z^n] S_L(z) = r(r-1)^{n-1}$.

(Remark: this last result can be computed by elementary asymptotics and is a good check; however the constructions we used extends nicely to more difficult problems, as follows).

Question 6? What is the generating function $S_{\leq m}(z)$ counting the number of words with at most m consecutive identical letters over an alphabet of r letters?
Remark: A run of one letter is a continuous sequence of this letter (that is not interrupted by any other letter). The solution of Question 6 provides the generating function of texts with maximal run equal to m.

For further lecture, you find in Flajolet-Sedgewick book "Analysis Combinatorics" (p. 308-312) the proof of the following beautiful result:

Proposition. *The longest run parameter L taken over the set of binary words of length n (endowed with the uniform distribution) satisfies the following estimate*

$$\mathbf{P}_n(L < \lfloor \log_2(n) \rfloor + h) = e^{-\alpha(n)2^{-h-1}} + \mathcal{O}\left(\frac{\log n}{\sqrt{n}}\right), \quad \alpha(n) := 2^{\{\log_2 n\}},$$

where $\{x\}$ is the fractional part of x. The mean satisfies

$$\mathbf{E}_n(L) = \log_1 n + \frac{\gamma}{\log 2} - \frac{3}{2} + P(\log_2 n) + o\left(\frac{\log^2 n}{\sqrt{n}}\right) \quad (P(.)\textit{periodic}.$$

II - Enumeration in free groups

We consider the alphabet $\mathcal{B} = \mathcal{A} \cup \overline{\mathcal{A}}$, where $\mathcal{A} = \{a_1, \ldots, a_r\}$ and $\overline{\mathcal{A}} = \{\overline{a_1}, \ldots, \overline{a_r}\}$.

A word over the alphabet $\mathcal{B}$ is said to be reduced if it arises from a word over $\mathcal{B}$ by the maximal application of the reductions $a_j\overline{a_j} \mapsto \epsilon$ and $\overline{a_j}a_j \mapsto \epsilon$ (with ϵ the empty word).

Therefore a reduced word has no factor of the form $a_j\overline{a_j}$ or $\overline{a_j}a_j$. The reduced words serve as canonical representations of elements of the free group $\mathbf{F}_r$ generated by $\mathcal{A}$.

Question 1? Let u_i and $\overline{u_i}$ respectively mark the number of occurrences of letters a_i and $\overline{a_i}$, respectively.

The generating function $R(u_1, \ldots, u_r, \overline{u_1}, \ldots, \overline{u_r})$ of the class $\mathcal{R}$ of reduced words may be obtained by a substitution within the generating $S(z)$ of Smirnov words over an alphabet of r letters. What is this substitution?

Question 2? Check that the OGF of reduced words with z marking length is

$$R(z) = \frac{1+z}{1-(2r-1)z},$$

which implies $R_n = 2r(2r-1)^{n-1}$.

Question 3? The Abelian image $\lambda(w)$ of an element of the free group $\mathbf{F}_r$ is obtained by letting all letters commute and applying the reductions $a_j.a_j^{-1} = 1$. Therefore $\lambda(w)$ can be put under the form $a_1^{m_1} \ldots {}_r^{m_r}$, with each $m_i \in \mathbb{Z}$. Let $(x_1, \ldots, x_r)$ be a vector of inderteminates and define $\mathbf{x}^{\lambda(w)} := x_1^{m_1} \ldots x_r^{m_r}$

Compute $Q(z, \mathbf{x}) = \sum_{w \in \mathcal{R}} z^{|w|}\mathbf{x}^{\lambda(w)}$.

This last quantity is related to the growth of the free group $\mathbf{F}_r$.