DEFINITIONS & NOTATION

JAMIE SIMPSON

Below are some common definitions associated with combinatorics on words. If you're composing a problem, you can assume your readers know them already.

A *word* or *string* is a sequence of symbols taken from a set of symbols called an *alphabet* A. The set of all such words is A^*. The number of occurrences of symbols in the word w is its *length*, written $|w|$. The number of occurrences of a letter a in a word is written $|w|_a$. If the alphabet A is an ordered set $\{a_1, a_2, \ldots, a_k\}$, then the *Parikh vector* of w is $[|w|_{a_1}|w|_{a_2} \ldots |w|_{a_k}]$. A word containing no symbols is the *empty word*, written ε, and we write $A^+ = A^* - \{\varepsilon\}$. If $w = xyz$ for words x, y and z then x is a *prefix* of w, z is a *suffix* of w and x, y and z are all *factors* (*subwords, substrings*) of w. A factor is *proper* if it is not the whole word. *Proper prefix* and *proper suffix* are defined in the same way. A factor which is both a prefix and a suffix of w is a *border* of w. We write $x[i]$ for the ith letter in w and $x[i..j]$ for the factor beginning with the ith letter and ending with the jth. A word of the form $w = xxx \cdots x$, with x a factor appearing n times, is written $w = x^n$ and is called a *power* of x. If $n = 2$ then x^n is a *square* and if $n = 3$ it's a *cube*. A word that is not a power is *primitive*. A positive integer p is a *period* of a word w if $w[i] = w[i+p]$ for all $1 \leq i \leq |w| - p$. The shortest period of a word is sometimes called *the* period. A periodic factor of a word is a *periodicity*. If a periodicity has period p and length n then its *exponent* is n/p. The *reverse* of a word is the word written backwards, thus the reverse of $x[1..n]$ is $x[n]x[n-1] \cdots x[2]x[1]$. A word that equals its reverse is a *palindrome*. If $w = uv$ then vu is a *conjugate* or *rotation* of w. If $|u| = j$ then vu is the *jth rotation* of w. A word, necessarily primitive, that is lexicographically less than any of its conjugates is a *Lyndon word*.

An infinite word w is *periodic* if $w[i + p] = w[i]$ for all $i \geq 1$. It is *eventually periodic* if $w[i + p] = w[i]$ holds for all i greater than some number k (which may be 0). An infinite word that is not eventually periodic is *aperiodic*. The *complexity* of a word w is a function $C(w, n)$

where $C(w,n)$ is the number of distinct factors of length n in the word. If $C(w,n) = n+1$ for all positive integers n then the (necessarily binary) word is *Sturmian*. Sturmian words may be defined in many equivalent ways. A factor u of a word w is called *right (respectively left) special* if there exist two distinct letters a and b such that ua and ub (respectively au and bu) are factors of w. An infinite word w is *recurrent* if every factor occurring in the word occurs infinitely often. The word is *uniformly recurrent* if for every factor u there exists an integer k such that every factor of w of length k contains at least one occurrence of u.