EXAMPLE PROJECT: ENUMERATING WORDS AVOIDING PATTERNS

ERIC ROWLAND

A *word* on a set Σ is a finite sequence of elements from Σ. For example, there are 5 length-3 words on $\{0, 1\}$ that contain no consecutive 0s:

$$010,\ 011,\ 101,\ 110,\ 111.$$

We say that these words *avoid the pattern* 00. The enumeration of words avoiding a pattern is a classical combinatorial question that has been studied in a number of contexts. For the pattern 00, it turns out that the number of length-n words on $\{0, 1\}$ that avoid 00 is the Fibonacci number $F(n + 2)$. If we change the pattern to 11, it will not surprise you to learn that the result is the same, since flipping all 0s to 1s and 1s to 0s puts words avoiding 00 into correspondence with words avoiding 11. We say that the patterns 00 and 11 are *avoiding-equivalent* since they are avoided by the same number of words.

It is more surprising that the patterns 0010 and 0110 are avoiding-equivalent, since a bijection from words avoiding 0010 to words avoiding 0110 is not as obvious. However, this equivalence occurs because these patterns have the same set of *self-overlaps*. The pattern 0010 can only overlap itself in 1 letter, namely as $001\underline{0}010$. The same is true for 0110, since it can only overlap itself as $011\underline{0}110$. In general, two patterns are avoiding-equivalent if and only if they have the same set of self-overlaps [1].

The generating series for the number of words avoiding a (contiguous) pattern is always rational. For example, if $s(n)$ is the number of length-n words on $\{0, 1\}$ that avoid 00, then

$$\sum_{n\geq 0} s(n)x^n = \sum_{n\geq 0} F(n+2)x^n = \frac{x+1}{1-x-x^2}.$$

This rational expression encodes the entire sequence $s(n)_{n\geq 0}$. Alternatively, if $s(n)$ is the number of length-n words on $\{0, 1\}$ that avoid 0010, then

$$\sum_{n\geq 0} s(n)x^n = \frac{x^3+1}{1-2x+x^3-x^4}.$$

Amazingly, these rational expressions can be written out directly by computing the set of self-overlaps [2]. Namely, let Σ be a finite set, and let p be a nonempty word on Σ. A *border* of p is a nonempty word that is both a prefix and suffix of p. Let $B(p)$ be the set consisting of the lengths of the borders of p. Let $k = |\Sigma|$, let $\ell = |p|$, and let $s(n)$ be the number of length-n words on Σ that avoid p. Then

$$\sum_{n \geq 0} s(n) x^n = \frac{\displaystyle\sum_{i \in B(p)} x^{\ell-i}}{\displaystyle\sum_{i \in B(p)} \left(x^{\ell-i} - k x^{\ell-i+1}\right) + x^{\ell}}.$$

What other generating series can be written down directly in this way? For example, if we count words containing exactly m instances of p, what is the generating series? If we count words avoiding multiple patterns simultaneously, what is the generating series? Do the coefficients in the numerators and denominators carry combinatorial meaning? What if we look at more general patterns?

References

[1] L. J. Guibas and A. M. Odlyzko, String overlaps, pattern matching, and nontransitive games, *Journal of Combinatorial Theory, Series A* **30** (1981) 183–208.

[2] A. D. Solov'ev, A combinatorial identity and its application to the problem concerning the first occurrence of a rare event (translated from Russian by N. A. Brunswick), *Theory of Probability and Its Applications* **11** (1966) 276–282.