Consecutive Pattern-Avoidance in Words

Mingjia Yang and Doron ZEILBERGER

Abstract. We use an extension of the celebrated Goulden-Jackson Cluster method, aided by symbolic computation, to enumerate *words* avoiding a given set of consecutive patterns. Our approach also lends itself to the enumeration of *permutations*, replicating, and giving a new approach to, results in this research area pioneered, in 2003, by Sergi Elizalde and Marc Noy, but it also enables the efficient enumeration, for example, of words in $1^s \dots n^s$, avoiding the consecutive pattern $1 \dots r$, for *any* s, and *any* r. All the sequences for $s = 1$, and $3 \leq r \leq 9$ are in the On-Line Encyclopedia of Integer Sequences, with many terms, using the explicit generating function of Elizalde and Noy. Also, quite a few of theses sequences for $s > 1$ are already there, but with very few terms. Our implied algorithms are $O(n^s)$ and hence yield many more terms, and, of course, new sequences. This article is accompanied by three Maple packages implementing our algorithms.

Introduction. Rodica Simion and Herbert Wilf initiated the study of enumerating *classical* pattern-avoidance. This is a very dynamic area with its own annual conference and Wikipedia page. Recall that a permutation $\pi = \pi_1 \dots \pi_n$ avoids a pattern $\sigma = \sigma_1 \dots \sigma_k$ if none of the $\binom{n}{k}$ length-k subsequence of π, reduces to σ.

Alex Burstein([Bu]), in a 1998 PhD thesis, under the direction of Herb Wilf, pioneered the enumeration of *words* avoiding a set of patterns. This field is also fairly active today, with notable contributions by, *inter alia*, Toufik Mansour (e.g [BuM]), Lara Pudwell([P]).

The enumeration of permutations avoiding a given (classical) pattern, or a set of patterns, is notoriously difficult, and it is widely believed to be intractable for most patterns, hence it would be nice to have other notions for which the enumeration is more feasible. Such an analog was given, in 2003, by Sergi Elizalde and Marc Noy, in a seminal paper ([EN]), that introduces the study of the enumeration of permutations avoiding *consecutive* patterns. A permutation $\pi = \pi_1 \dots \pi_n$ avoids a pattern $\sigma = \sigma_1 \dots \sigma_k$ if none of the $n - k + 1$ length-k *factors*, $\pi_i \pi_{i+1} \dots \pi_{i+k-1}$ of π, reduces to σ.

Algorithmic approaches to the enumeration of *permutations* avoiding sets of consecutive patterns were given by Brian Nakamura, Andrew Baxter, and Doron Zeilberger ([Na], [BaNaZ]). Our present approach may be viewed as an extension, from permutations to words, of Nakamura's paper, who was also inspired by the Goulden-Jackson , but in a sense, is more straightforward, and closer in spirit to the original Goulden-Jackson method ([GJ],that is beautifully exposited in [NZ]).

Guessing General Generating Functions in $x_1, \dots, x_n$ (for arbitrary n) for Words avoiding a pattern (or set of patterns)

Recall that the original Goulden-Jackson method ([GJ][NZ]) inputs a *finite* alphabet, A, that may be taken to be $\{1, ..., n\}$, and a finite set of 'bad words', B.

It outputs a certain **rational function**, let's call it $F(x_1, \dots, x_n)$, that is the multi-variable generating function, in $x_1, \dots, x_n$, for the disclerte n-variable function

$$G(m_1, \dots, m_n) \quad ,$$

the counts the words in $1^{m_1} \dots n^{m_n}$ (there are totally $(m_1 + \dots + m_n)!/(m_1! \cdots m_n!)$ of them) that **never** contain, as *consecutive* subwords (aka *factors* in linguistics) any member of B. In other words:

$$F(x_1, \dots, x_n) = \sum_{(m_1,\dots,m_n) \in N^n} G(m_1, \dots, m_n)\, x_1^{m_1} \cdots x_n^{m_n} \quad .$$

This is nicely implemented in the Maple package `DavidIan.txt`, that accompanies [NZ], and is freely available from

`http://sites.math.rutgers.edu/ zeilberg/tokhniot/DavidIan.txt` .

for example if $n = 4$, so the alphabet if $\{1, 2, 3, 4\}$ and the set of 'bad words' to avoid is $\{1234, 1432\}$, then, starting a Maple session, and typing:

```
read `DavidIan.txt`:  lprint(subs(t=0,GJgf(1,2,3,4,[1,2,3,4],[1,4,3,2],x,t)));
```

immediately returns

```
1/(1-x[1]-x[2]-x[3]-x[4]+ 2*x[1]*x[2]*x[3]*x[4])
```
,

that in Humanese reads

$$\frac{1}{1 - x_1 - x_2 - x_3 - x_4 \,+\, 2\, x_1 x_2 x_3 x_4} \quad .$$

Enumerating Words Avoiding Consecutive Patterns: Let the Computer Do the Guessing

Now we are interested in words in an *arbitrarily large* alphabet $\{1, \dots n\}$ avoiding a set of patterns. But each pattern, e.g. 123, entails *arbitrarily large* set of forbidden words, in this case the $\binom{n}{3}$ members of the set

$$\{i_1\, i_2\, i_3 \,|\, 1 \le i_1 < i_2 < i_3 \le n\} \quad .$$

We can ask `DavidIan.txt` to find the generating function for each specific n, and then hope to conjecture a **general** formula in terms of $x_1, \dots, x_n$.

This is accomplished by the Maple package `GJpats.txt` available from the webpage of this article. It used the original `DavidIan.txt` to produce the corresponding generating functions for increasing values for n. For example for words avoiding the consecutive pattern 123 for $n = 3$

`lprint(GFpats([1,2,3],x,3,0));` yields

```
1/(x[1]*x[2]*x[3]-x[1]-x[2]-x[3]+1)
```
.

This is simple enough. Moving right along `lprint(GFpats([1,2,3],x,4,0));` yields ,

```
1/(x[1]*x[2]*x[3]*x[4]-x[1]*x[2]*x[3]-x[1]*x[2]*x[4]
-x[1]*x[3]*x[4]-x[2]*x[3]*x[4]+x[1] + x[2]+x[3]+x[4]-1)
```

while `lprint(GFpats([1,2,3],x,4,0));` yields ,

```
-1/(x[1]*x[2]*x[3]*x[4]+x[1]*x[2]*x[3]*x[5]+x[1]*x[2]*x[4]*x[5]+x[1]*x[3]*x[4]*x[5]

+x[2]*x[3]*x[4]*x[5] -x[1]*x[2]*x[3]-x[1]*x[2]*x[4]-x[1]*x[2]*x[5]-x[1]*x[3]*x[4]-


x[1]*x[3]*x[5]-x[1]*x[4]*x[5]-x[2]*x[3]*x[4]-x[2]*x[3]*x[5]-x[2]*x[4]*x[5]
-x[3]*x[4]*x[5]+x[1]+x[2]+x[3]+x[4]+x[5]-1)     .
```

It looks like these are *symmetric* functions. Procedure `SPtoM(P,x,n,m)` expresses a polynomial, P, in the indexed variables $x[1], \ldots, x[n]$ in terms of the *monomial symmetric polynomials* m_λ. Applying this procedure we have

`SPtoM(denom(GFpats([1,2,3],x,5,0)),x,5,m);` yields

```
m[1, 1, 1, 1] - m[1, 1, 1] + m[1] - m[]   ,
```

`SPtoM(denom(GFpats([1,2,3],x,6,0)),x,6,m);` yields

```
m[1,1,1,1,1,1]-m[1,1,1,1]+m[1,1,1]-m[1]+m[]
```

`SPtoM(denom(GFpats([1,2,3],x,7,0)),x,7,m);` yields

m[1,1,1,1,1,1,1]-m[1,1,1,1,1,1]+m[1,1,1,1]-m[1,1,1]+m[1]-m[]

Now you don't have to be a Ramanujan to conjecture the following Theorem

Fact The generating function for words in $\{1, 2, \ldots, n\}$ avoiding the consecutive pattern 123, let's call it $F_3(x_1, \ldots, x_n)$ is

$$F_3(x_1, \ldots, x_n) = \frac{1}{1 - e_1 + e_3 - e_4 + e_6 - e_7 + e_9 - e_{10} \ldots} \quad ,$$

where e_i stands for the *elementary symmetric function* of degree i in $x_1, \ldots, x_i$, i.e. the coefficient of z^i in $(1 + x_1 z) \ldots (1_+ x_n z)$.

(Note that $e_i = m_{1^i}$).

Doing the analogous guessing for the consecutive patterns 1234 and 12345, a *meta-pattern* emerges, and we are safe in formulating the following

Theorem For $n \geq 1, r \geq 2$, the generating function for words in $\{1, 2, \ldots, n\}$ avoiding the consecutive pattern $12\ldots r$, let's call it $F_r(x_1,\ldots,x_n)$ is

$$F_r(x_1,\ldots,x_n)=\frac{1}{1-e_1+e_r-e_{r+1}+e_{2r}-e_{2r+1}+e_{3r}-e_{3r+1}+\ldots} \quad .$$

Of course, so far, these are 'only' guesses, but once we know them, the human can prove them, applying the cluster method to a *general* alphabet. At this time of writing this still has to be done by humans and will be given later.

Efficient Computations

The Theorem immediately implies the following partial recurrence equation for the actual coefficients.

Fundamental Recurrence: Let $f_r(\mathbf{m})$ be the number of words in the alphabet $\{1,\ldots,n\}$ with m_1 1-s, m_2 2-s, $\ldots m_n$ n-s (where we abbreviate $\mathbf{m}=(m_1,\ldots,m_n)$). Also let V_i be the set of $0-1$ vectors of length n with i ones, then

$$\begin{aligned} f_r(\mathbf{m}) = &\sum_{\mathbf{v}\in V_1} f(\mathbf{m}-\mathbf{v}) - \sum_{\mathbf{v}\in V_r} f(\mathbf{m}-\mathbf{v}) \\ &+ \sum_{\mathbf{v}\in V_{r+1}} f(\mathbf{m}-\mathbf{v}) - \sum_{\mathbf{v}\in V_{2r}} f(\mathbf{m}-\mathbf{v}) \\ &+ \sum_{\mathbf{v}\in V_{2r+1}} f(\mathbf{m}-\mathbf{v}) - \sum_{\mathbf{v}\in V_{3r}} f(\mathbf{m}-\mathbf{v}) + \ldots \quad . \end{aligned}$$

Suppose that we want to compute $f_3(1^{100})$, i.e. the number of permutations of length 100 that avoid the consecutive pattern 123. If we use the above recurrence literally, we would need about 2^{100} computations, but there is a shortcut!

Enter Symmetry

It follows from the generating function that $f_r(m_1,\ldots,m_n)$ is symmetric, hence for any $0-1$ vector v $f_r(v)$ only depends on the number of 1's. Hence the above fundamental recurrence implies the following

Fast Recurrence For Enumerating Permutations avoiding the consecutive pattern $1\ldots r$: Let $a_r(n)$ be the number of permutations of length n that avoid the consecutive pattern $1\ldots r$, then

$$a_r(n)=na_r(n-1)-\binom{n}{r}a_r(n-r)+\binom{n}{r+1}a_r(n-r-1)-\binom{n}{2r}a_r(n-2r)+\binom{n}{2r+1}a_r(n-2r-1)$$
$$-\binom{n}{3r}a_r(n-3r)+\binom{n}{3r+1}a_r(n-3r-1)+\ldots \quad .$$

This recurrence immediately implies the Elizalde-Noy exponential generating functions

$$\sum_{n=0}^{\infty} a_r(n) \frac{x^n}{n!} = \frac{1}{1 - x + \frac{x^r}{r!} - \frac{x^{r+1}}{(r+1)!} + \frac{x^{2r}}{(2r)!} - \frac{x^{2r+1}}{(2r+1)!} + \frac{x^{3r}}{(3r)!} - \frac{x^{3r+1}}{/} 3r+1)! + \dots} \quad .$$

While this 'explicit' (exponential) generating function is 'nice', it is more efficient to use the fast recurrence. And indeed, the OEIS has these sequences for $3 \leq 3 \leq 9$, with many terms (using the above recurrences, probably). These are (in order): `A049774, A117158, A177523, A177533, A177553, A230051, A230231`.

References

[BaNZ] Andrew Baxter, Brian Nakamura, and Doron Zeilberger, *Automatic Generation of Theorems and Proofs on Enumerating Consecutive-Wilf classes*, in: *"Advances in Combinatorics: Waterloo Workshop in Computer Algebra, W80, May 26-29, 2011"* [Volume in honor of Herbert S. Wilf], edited by Ilias Kotsireas and Eugene Zima, 121-138

[Bu] Alex Burstein, *"Enumeration of words with forbidden patterns"*, Ph.D. thesis, University of Pennsylvania, 1998.

[BuM] Alex Burstein and Toufik Mansour, *Words Restricted by Patterns with at Most 2 Distinct Letters*, The electronic journal of combinatorics 9(2) (2002), #R3

[EN] Sergi Elizalde and Marc Noy, *Consecutive patterns in permutations*, Advances in Applied Mathematics **30** (2003), 110-125 .

[GJ] Ian Goulden and David Jackson, *An inversion theorem for cluster decomposition of sequences with distinguished subsequences*, J. London Math. Soc.(2) **20** (1979), 567-576.

[Na] Brian Nakamura, *Computational Approaches to Consecutive Pattern Avoidance in Permutations*, Pure Math. Appl. (PU.M.A.) **22**(2011), 253268.
Also available from`https://arxiv.org/abs/1102.2480`

[NoZ] John Noonan and Doron Zeilberger, *The Goulden-Jackson Cluster Method: Extensions, Applications, and Implementations*, J. Difference Eq. Appl. **5**(1999), 355-377.
`http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/gj.html` .

[P] Lara Pudwell, *Enumeration schemes for pattern-avoiding words and permutations*, Rutgers University, May 2008,
`http://faculty.valpo.edu/lpudwell/papers/pudwell_thesis.pdf`

[Wi] The Wikipedia Foundation, *Permutation Pattern* [Initiated and largely written by Vince Vatter]
`https://en.wikipedia.org/wiki/Permutation_pattern` .

Mingjia Yang, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `my239 at math dot rutgers dot edu` .

Doron Zeilberger, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `DoronZeil at gmail dot com` .