The Carlitz-Scoville-Vaughan Theorem and its Generalizations

Ira M. Gessel

Department of Mathematics
Brandeis University

Joint Mathematics Meeting
San Diego
January 12, 2013

Counting pairs of sequences

In their 1976 paper *Enumeration of pairs of sequences by rises, falls and levels* in Manuscripta Mathematica, Leonard Carlitz, Richard Scoville, and Theresa Vaughan studied pairs of sequences of integers of the same length according to rises, falls and levels.

For example suppose the two sequences are

$$\begin{matrix} 1\,1\,2 \\ 2\,3\,1 \end{matrix}$$

In the first position the first sequence 112 has a level (11), and the second sequence has a rise (23). So for the pair of sequences, the specification of the first position is LR, and the specification of the second position is RF. They wanted to count pairs of sequences according to the number of RR, FR, LR, ..., LL. A general formula is very complicated so they considered special cases.

One of their results is the following: Let $\{A, B\}$ be a partition of $\{\mathrm{RR}, \ldots, \mathrm{LL}\}$. Then the reciprocal of the generating function for sequences in which every specification is in A is the generating function, with alternating signs, of the generating function for sequences in which every specification is in B.

One of their results is the following: Let $\{A, B\}$ be a partition of $\{\mathrm{RR}, \ldots, \mathrm{LL}\}$. Then the reciprocal of the generating function for sequences in which every specification is in A is the generating function, with alternating signs, of the generating function for sequences in which every specification is in B.

In the appendix to their paper, they proved a more general version of this result, which I will state a little differently.

The Carlitz-Scoville-Vaughan Theorem

Let A be an alphabet, and let R be a relation on A, that is, a subset of $A \times A = A^2$. Let $A^{(R)}$ be the set of words $a_1 \cdots a_n$ in A^* such that $a_1 \; R \; a_2 \; R \cdots R \; a_n$. Note that the empty word 1 and all words of length one are in $A^{(R)}$. Let $\bar{R} = A^2 - R$.

Then

$$\sum_{w \in A^{(\bar{R})}} w = \left(\sum_{w \in A^{(R)}} (-1)^{l(w)} w \right)^{-1}.$$

Here $l(w)$ is the length of w, and we are working in the ring of formal power series in noncommuting variables.

The Carlitz-Scoville-Vaughan Theorem

Let A be an alphabet, and let R be a relation on A, that is, a subset of $A \times A = A^2$. Let $A^{(R)}$ be the set of words $a_1 \cdots a_n$ in A^* such that $a_1 \; R \; a_2 \; R \cdots R \; a_n$. Note that the empty word 1 and all words of length one are in $A^{(R)}$. Let $\bar{R} = A^2 - R$.

Then
$$\sum_{w \in A^{(\bar{R})}} w = \left(\sum_{w \in A^{(R)}} (-1)^{l(w)} w \right)^{-1}.$$

Here $l(w)$ is the length of w, and we are working in the ring of formal power series in noncommuting variables.

Carlitz, Scoville, and Vaughan didn't do anything more with this result. But I believe that it should be considered one of the fundamental theorems of enumerative combinatorics.

A simple example

Let $A = \{x, y\}$, and let $R = \{xx\}$ and $\bar{R} = \{xy, yx, yy\}$. So $A^{(\bar{R})}$ is the set of words in the letters x and y with no consecutive xx, and $A^{(R)}$ is the set of words in which every consecutive pair is xx. Thus

$$\sum_{w \in A^{(R)}} (-1)^{l(w)} w = 1 - y - x + x^2 - x^3 + \cdots = (1 + x)^{-1} - y.$$

Therefore, by the CSV theorem,

$$\sum_{w \in A^{(\bar{R})}} w = \left((1 + x)^{-1} - y\right)^{-1}$$

A simple example

Let $A = \{x, y\}$, and let $R = \{xx\}$ and $\bar{R} = \{xy, yx, yy\}$. So $A^{(\bar{R})}$ is the set of words in the letters x and y with no consecutive xx, and $A^{(R)}$ is the set of words in which every consecutive pair is xx. Thus

$$\sum_{w \in A^{(R)}} (-1)^{l(w)} w = 1 - y - x + x^2 - x^3 + \cdots = (1+x)^{-1} - y.$$

Therefore, by the CSV theorem,

$$\sum_{w \in A^{(\bar{R})}} w = \big((1+x)^{-1} - y\big)^{-1} = (1+x)\big(1 - y(1+x)\big)^{-1}.$$

A simple example

Let $A = \{x, y\}$, and let $R = \{xx\}$ and $\bar{R} = \{xy, yx, yy\}$. So $A^{(\bar{R})}$ is the set of words in the letters x and y with no consecutive xx, and $A^{(R)}$ is the set of words in which every consecutive pair is xx. Thus

$$\sum_{w \in A^{(R)}} (-1)^{l(w)} w = 1 - y - x + x^2 - x^3 + \cdots = (1+x)^{-1} - y.$$

Therefore, by the CSV theorem,

$$\sum_{w \in A^{(\bar{R})}} w = \left((1+x)^{-1} - y\right)^{-1} = (1+x)\left(1 - y(1+x)\right)^{-1}.$$

Note that if we set $y = x$, we get a generating function for Fibonacci numbers.

Another simple example

Let $A = \{x_1, x_2, \dots\}$, let $R = \{x_i x_j : i \leq j\}$, so $\overline{R} = \{x_i x_j : i > j\}$.

Another simple example

Let $A = \{x_1, x_2, \dots\}$, let $R = \{x_i x_j : i \le j\}$, so $\bar{R} = \{x_i x_j : i > j\}$. Then the CSV theorem gives

$$\sum_{n=0}^{\infty} \mathbf{e}_n = \left(\sum_{n=0}^{\infty} (-1)^n \mathbf{h}_n \right)^{-1},$$

where

$$\mathbf{h}_n = \sum_{i_1 \le \cdots \le i_n} x_{i_1} \cdots x_{i_n}$$

is the noncommutative complete symmetric function and

$$\mathbf{e}_n = \sum_{i_1 > \cdots > i_n} x_{i_1} \cdots x_{i_n}$$

is the noncommutative elementary symmetric function.

Proof of the CSV theorem

I'll sketch three proofs.

Proof 1. (Essentially the same as Carlitz, Scoville, and Vaughan's proof.) We prove that

$$\sum_{v \in A^{(R)}} (-1)^{l(v)} v \cdot \sum_{w \in A^{(\bar{R})}} w = 1.$$

The left side is

$$\sum_{v \in A^{(R)}} \sum_{w \in A^{(\bar{R})}} (-1)^{l(v)} vw.$$

Every nonempty word that occurs in this sum appears twice, once with a plus sign and once with a minus sign.

Proof 2. Let us define an R-descent of a word $a_1 a_2 \cdots a_n$ in A^* to be in i such that $a_i \ \overline{R} \ a_{i+1}$. Let $\mathbf{h}_n^{(R)}$ be the sum of all words of length n with no R-descent, that is, the sum of all words $a_1 \cdots a_n$ for which $a_1 \ R \ a_2 \ R \cdots R \ a_n$. Then the set of words of length n with a given R-descent set can be expressed by inclusion-exclusion in terms of the $\mathbf{h}_n^{(R)}$.

Proof 2. Let us define an R-descent of a word $a_1 a_2 \cdots a_n$ in A^* to be in i such that $a_i \ \overline{R} \ a_{i+1}$. Let $\mathbf{h}_n^{(R)}$ be the sum of all words of length n with no R-descent, that is, the sum of all words $a_1 \cdots a_n$ for which $a_1 \ R \ a_2 \ R \cdots R \ a_n$. Then the set of words of length n with a given R-descent set can be expressed by inclusion-exclusion in terms of the $\mathbf{h}_n^{(R)}$.

For example the sum of the words of length 5 with R-descent set $\{3\}$ is $\mathbf{h}_3^{(R)}\mathbf{h}_2^{(R)} - \mathbf{h}_5^{(R)}$.

Proof 2. Let us define an R-descent of a word $a_1 a_2 \cdots a_n$ in A^* to be in i such that $a_i \; \overline{R} \; a_{i+1}$. Let $\mathbf{h}_n^{(R)}$ be the sum of all words of length n with no R-descent, that is, the sum of all words $a_1 \cdots a_n$ for which $a_1 \; R \; a_2 \; R \cdots R \; a_n$. Then the set of words of length n with a given R-descent set can be expressed by inclusion-exclusion in terms of the $\mathbf{h}_n^{(R)}$.

For example the sum of the words of length 5 with R-descent set $\{3\}$ is $\mathbf{h}_3^{(R)}\mathbf{h}_2^{(R)} - \mathbf{h}_5^{(R)}$.

In particular, inclusion-exclusion gives the sum of the words in which every position is an R-descent as

$$\sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty}(-1)^{n-1}\mathbf{h}_n^{(R)}\right)^k = \left(1 - \sum_{n=1}^{\infty}(-1)^{n-1}\mathbf{h}_n^{(R)}\right)^{-1}.$$

Proof 3. (At least for the commutative version.) Without loss of generality we may assume that A is finite. The coefficients on both sides can be expressed as matrix entries by the transfer matrix method. Then the result follows by matrix algebra. (See, e.g., Goulden and Jackson's *Combinatorial Enumeration*.)

An application: counting words by R-runs

An R-run in a word is a maximal nonempty subword in $A^{(R)}$, so the R-descents break up a word into R-runs. For a nonempty word, the number of R-runs is one more than the number of R-descents.

To count words in A^* by the number of R-runs, we define a new alphabet $\boxed{A^{(R)}}$ whose letters are $\boxed{a_1 a_2 \dots a_n}$ where $a_1 \cdots a_n \in A^{(R)}$. Now let $\mathcal{R} \subseteq \boxed{A^{(R)}}^2$ be the set of words of the form $\boxed{a_1 \cdots a_n}\boxed{a_{n+1} \cdots a_{n+r}}$ where $a_n a_{n+1} \in R$. In other words, $a_1 \cdots a_{n+r} \in A^{(R)}$. Then the CSV theorem allows us to count words of the form

$$\boxed{a_1 \cdots a_{n_1}}\boxed{a_{n_1+1} \cdots a_{n_2}} \cdots \boxed{a_{n_{k-1}+1} \cdots a_{n_k}}$$

in which the R-descent set of the word $a_1 \cdots a_{n_k}$ is $\{n_1, n_2, \dots, n_{k-1}\}$.

To get something useful from this, we apply the homomorphism that takes $\boxed{a_1 \cdots a_n}$ to $a_1 \cdots a_n t$, where t is a variable that commutes with all the letters. Then the image of $\mathbf{h}_n^{(\mathcal{R})}$ under this homomorphism will be a sum of words $a_1 \cdots a_n$ in $A^{(R)}$, each multiplying by a sum of powers of t corresponding to the ways in which this word can be cut into nonempty pieces. A word of length n can be cut in any of the $n-1$ spaces between its letters, so the total coefficient for a word of length n will be $t(1+t)^{n-1}$.

On the other side, we will be counting words in A^* where a word with j R-runs will be weighted t^j. So applying the CSV theorem gives

$$\sum_{w \in A^*} t^{R\text{-run}(w)} w = \left(1 - t \sum_{n=1}^{\infty} (1-t)^{n-1} \mathbf{h}_n^{(R)}\right)^{-1}.$$

More generally, we could assign a different weight to each possible R-run length. If we assign the weight t_i to a run of length i then the same argument gives

$$\sum_{w \in A^*} T(w)w = \left(1 + \sum_{n=1}^{\infty} u_n \mathbf{h}_n^{(R)}\right)^{-1},$$

where $T(w)$ is the weight of w and u_n counts compositions of n where each part i is weighted $-t_i$, so

$$\sum_{n=1}^{\infty} u_n z^n = \left(1 + \sum_{i=1}^{\infty} t_i z^i\right)^{-1}.$$

For example, to count words in which every R-run has length 3, we set $t_3 = 1$ and $t_i = 0$ for $i \neq 3$, so $u_{3k} = (-1)^k$ and $u_n = 0$ if 3 does not divide n, so the sum of all words in which every R-run has length 3 is

$$\left(\sum_{k=0}^{\infty}(-1)^k \mathbf{h}_{3k}^{(R)}\right)^{-1}.$$

For example, to count words in which every R-run has length 3, we set $t_3 = 1$ and $t_i = 0$ for $i \neq 3$, so $u_{3k} = (-1)^k$ and $u_n = 0$ if 3 does not divide n, so the sum of all words in which every R-run has length 3 is

$$\left(\sum_{k=0}^{\infty}(-1)^k \mathbf{h}_{3k}^{(R)}\right)^{-1}.$$

Similarly, the sum of all words in which every run length is odd is

$$\left(1 + \sum_{k=1}^{\infty}(-1)^k F_k \mathbf{h}_k^{(R)}\right)^{-1},$$

where F_k is the kth Fibonacci number.

Walks in digraphs

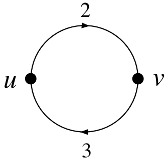
What if we we want to count words whose R-runs have the lengths 2, 3, 2, 3, ...? More generally, we can apply the CSV theorem to a very general situation: We are given a digraph in which each edge has a set of positive integers associated to it. Given two vertices u and v in the graph, we can count words in A^* whose sequence of R-run lengths corresponds to the numbers on a walk from u to v.

Walks in digraphs

What if we we want to count words whose R-runs have the lengths 2, 3, 2, 3, ...? More generally, we can apply the CSV theorem to a very general situation: We are given a digraph in which each edge has a set of positive integers associated to it. Given two vertices u and v in the graph, we can count words in A^* whose sequence of R-run lengths corresponds to the numbers on a walk from u to v.

In other words, we can count words whose R-run length sequences are in a regular language.

For example, the walks from u to v in



correspond to words in which the sequence of run lengths is 2, 3, 2, 3, . . . , 2.

They are counted by the $(1, 2)$ entry of the matrix

$$\begin{pmatrix} L_{5,0} & -L_{5,2} \\ -L_{5,3} & L_{5,0} \end{pmatrix}^{-1}$$

where $L_{m,i} = \sum_{n=0}^{\infty}(-1)^i \mathbf{h}^{(R)}_{mn+i}$.

Walks from u to v in the digraph

$$u \bullet \xrightarrow{2} \bullet \xrightarrow{1} \bullet \xrightarrow{3} \bullet \xrightarrow{2} \bullet\, v$$

correspond to words in which the sequence of run lengths is 2, 1, 3, 2. This is the $(1, 5)$ entry of the matrix

$$\begin{pmatrix} 1 & -\mathbf{h}_2^{(R)} & \mathbf{h}_3^{(R)} & -\mathbf{h}_6^{(R)} & \mathbf{h}_8^{(R)} \\ 0 & 1 & -\mathbf{h}_1^{(R)} & \mathbf{h}_4^{(R)} & -\mathbf{h}_6^{(R)} \\ 0 & 0 & 1 & -\mathbf{h}_3^{(R)} & \mathbf{h}_5^{(R)} \\ 0 & 0 & 0 & 1 & -\mathbf{h}_2^{(R)} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^{-1}$$

Walks from u to v in the digraph

$$u \bullet \xrightarrow{2} \bullet \xrightarrow{1} \bullet \xrightarrow{3} \bullet \xrightarrow{2} \bullet\, v$$

correspond to words in which the sequence of run lengths is 2, 1, 3, 2. This is the $(1,5)$ entry of the matrix

$$\begin{pmatrix}
1 & -\mathbf{h}_2^{(R)} & \mathbf{h}_3^{(R)} & -\mathbf{h}_6^{(R)} & \mathbf{h}_8^{(R)} \\
0 & 1 & -\mathbf{h}_1^{(R)} & \mathbf{h}_4^{(R)} & -\mathbf{h}_6^{(R)} \\
0 & 0 & 1 & -\mathbf{h}_3^{(R)} & \mathbf{h}_5^{(R)} \\
0 & 0 & 0 & 1 & -\mathbf{h}_2^{(R)} \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}^{-1}$$

(The formula we get here is the same as the inclusion-exclusion formula.)

Permutations

How can we count permutations by descents (or increasing runs) rather than arbitrary sequences?

We take $A = \{1, 2, \dots\}$ and $R = \leq = \{\, (i, j) : i \leq j \,\}$. (We could identify 1, 2, ..., with noncommuting variables $x_1, x_2, \dots$.)

Permutations

How can we count permutations by descents (or increasing runs) rather than arbitrary sequences?

We take $A = \{1, 2, \dots\}$ and $R = \leq = \{\,(i, j) : i \leq j\,\}$. (We could identify 1, 2, . . . , with noncommuting variables $x_1, x_2, \dots$.)

We apply the linear map that takes a sequence $\pi = a_1 \cdots a_n$ to $z^n/n!$ if π is a permutation of $[n] = \{1, 2, \dots, n\}$ and to 0 if π is not of this form.

Permutations

How can we count permutations by descents (or increasing runs) rather than arbitrary sequences?

We take $A = \{1, 2, \dots\}$ and $R = \leq = \{\, (i, j) : i \leq j \,\}$. (We could identify 1, 2, ..., with noncommuting variables $x_1, x_2, \dots$.)

We apply the linear map that takes a sequence $\pi = a_1 \cdots a_n$ to $z^n/n!$ if π is a permutation of $[n] = \{1, 2, \dots, n\}$ and to 0 if π is not of this form.

It is not hard to see that when restricted to the algebra generated by $\mathbf{h}_1^{(R)}, \mathbf{h}_2^{(R)}, \dots$, this is a homomorphism that takes $\mathbf{h}_n^{(R)}$ to $z^n/n!$.

Permutations

How can we count permutations by descents (or increasing runs) rather than arbitrary sequences?

We take $A = \{1, 2, \dots\}$ and $R = \leq = \{\, (i, j) : i \leq j \,\}$. (We could identify 1, 2, . . . , with noncommuting variables $x_1, x_2, \dots$.)

We apply the linear map that takes a sequence $\pi = a_1 \cdots a_n$ to $z^n/n!$ if π is a permutation of $[n] = \{1, 2, \dots, n\}$ and to 0 if π is not of this form.

It is not hard to see that when restricted to the algebra generated by $\mathbf{h}_1^{(R)}, \mathbf{h}_2^{(R)}, \dots$, this is a homomorphism that takes $\mathbf{h}_n^{(R)}$ to $z^n/n!$.

So we can count permutations by replacing $\mathbf{h}_n^{(R)}$ with $z^n/n!$ in our formulas.

For example, our previous formula for counting sequences by descents, which may be written

$$\sum_{w \in A^*} t^{R\text{-run}(w)} w = (1-t)\left(1 - t\sum_{n=0}^{\infty}(1-t)^n\, \mathbf{h}_n^{(R)}\right)^{-1}$$

gives a generating function for the Eulerian polynomials, which count permutations by descents:

$$\frac{1-t}{1-te^{(1-t)x}}$$

Similarly, we can see that the "doubly exponential" generating function for pairs (π, σ) of permutations of $[n]$ with no common ascents is

$$\left(\sum_{n=0}^{\infty}(-1)^n \frac{x^n}{n!^2}\right)^{-1}.$$

By applying other homomorphisms we can count permutations by the number of inversions, descent set, number of descents, major index, or number of peaks of π^{-1}.

Generalizations

Two of my students, Susan Parker and Brian Drake, generalized the CSV to give a combinatorial interpretation of the compositional inverse of power series.

Susan Parker's Theorem (1993)

(Rediscovered by Jean-Louis Loday, 2006)

Examples:

$$(x - x^2)^{\langle -1 \rangle} = \frac{1 - \sqrt{1 - 4x}}{2} = \sum_{k=0}^{\infty} \frac{1}{k+1} \binom{2k}{k} x^{k+1}$$

$$\left(\frac{x}{1+x} \right)^{\langle -1 \rangle} = \frac{x}{1-x}$$

Define Narayana polynomials by

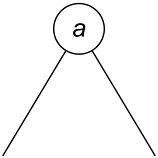
$$N_k(a,b) = \sum_{i=0}^{k-1} \frac{1}{k} \binom{k}{i} \binom{k}{i+1} a^i b^{k-i-1}.$$

Then

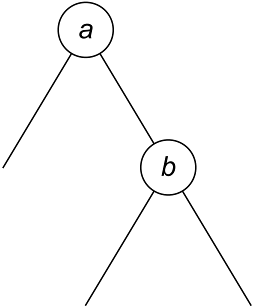
$$\left(x + (a+b) \sum_{k=1}^{\infty} (-1)^k N_k(a,b) x^{k+1} \right)^{\langle -1 \rangle}$$
$$= x + (a+b) \sum_{k=1}^{\infty} N_k(a,b) x^{k+1}$$

We work with ordered trees in which the leaves have a weight of x and the internal vertices are labeled a, b, c, ... I'll omit the leaves in most of my pictures.

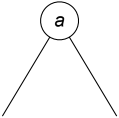
A letter is a single internal vertex with a fixed arity (number of children):

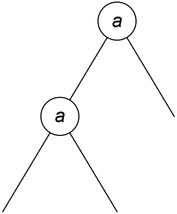


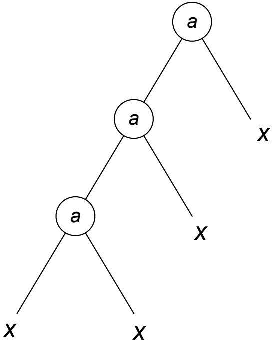
A link is a parent and child letter:



Parker's Theorem: Given a set of letters, the compositional inverse of the generating function for trees with a given set of links is the generating function for trees with the complementary set of links, with alternating signs.

As a simple example, take the single letter  . The trees

using only the link  look like



with generating function

$$\sum_{n=0}^{\infty} a^n x^{n+1} = \frac{x}{1 - ax}.$$

The complementary trees are the mirror images of these, with the same generating generating function, and thus

$$\left(\frac{x}{1+ax}\right)^{\langle -1\rangle} = \frac{x}{1-ax},$$

where the inverse is as power series in x.

The complementary trees are the mirror images of these, with the same generating generating function, and thus

$$\left(\frac{x}{1+ax}\right)^{\langle -1\rangle} = \frac{x}{1-ax},$$

where the inverse is as power series in x.

The CSV theorem is the special case of Parker's theorem for unary trees.

Brian Drake's Theorem (2008)

(Rediscovered by Vladimir Dotsenko, 2011)

Drake's theorem gives a similar interpretation for exponential generating functions corresponding to trees with labeled leaves and unlabeled internal vertices.

I'll just give an example of Drake's interpretation of

$$(e^x - 1)^{\langle -1 \rangle} = \log(1 + x).$$

Brian Drake's Theorem (2008)

(Rediscovered by Vladimir Dotsenko, 2011)

Drake's theorem gives a similar interpretation for exponential generating functions corresponding to trees with labeled leaves and unlabeled internal vertices.

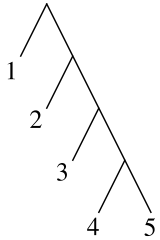
I'll just give an example of Drake's interpretation of

$$(e^x - 1)^{\langle -1 \rangle} = \log(1 + x).$$

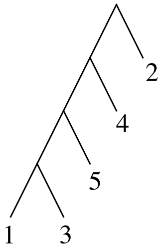
In other words,

$$\left(\sum_{n=1}^{\infty} \frac{x^n}{n!} \right)^{\langle -1 \rangle} = \sum_{n=1}^{\infty} (-1)^{n-1} (n-1)! \frac{x^n}{n!}.$$

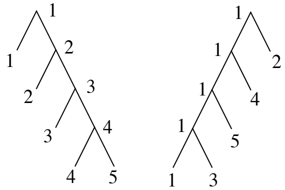
The interpretation is that $e^x - 1$ counts trees that look like

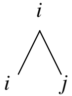


and $\log(1 + x)$ counts trees that look like



Explanation: we label each internal vertex with its least descendant:



Then all of the letters look like 

, with $i < j$.

The $e^x - 1$ trees have “right child” links and the $\log(1 + x)$ trees have “left child” links.

Forbidden subwords

Suppose we want to count words with forbidden subwords of length greater than 2. We can do this with the *Goulden-Jackson Cluster Theorem.*

Forbidden subwords

Suppose we want to count words with forbidden subwords of length greater than 2. We can do this with the *Goulden-Jackson Cluster Theorem.*

Let F be a set of "forbidden" words in A^* all of length at least 2. A *cluster* is a word in which an overlapping set of forbidden subwords is marked. For example, if $A = \{a\}$ and $F = \{aaa\}$ then the following are both clusters on the word a^6:

$$a\,a\,a\,a\,a\,a$$

$$a\,a\,a\,a\,a\,a$$

but $a\,a\,a\,a\,a\,a$ is not a cluster.

We define the cluster generating function to be

$$C(t) = \sum_{w \in A^*} w \sum_{\text{clusters } c \text{ on } w} t^{\#\text{ marked words in } c}$$

We define the cluster generating function to be

$$C(t) = \sum_{w \in A^*} w \sum_{\text{clusters } c \text{ on } w} t^{\#\text{ marked words in } c}$$

The Goulden-Jackson Cluster Theorem.

$$\sum_{w \in A^*} w t^{\#\text{ forbidden words in } w} = \left(1 - \sum_{a \in A} a - C(t-1)\right)^{-1}.$$

We define the cluster generating function to be

$$C(t) = \sum_{w \in A^*} w \sum_{\text{clusters } c \text{ on } w} t^{\#\text{ marked words in } c}$$

The Goulden-Jackson Cluster Theorem.

$$\sum_{w \in A^*} w t^{\#\text{ forbidden words in } w} = \left(1 - \sum_{a \in A} a - C(t-1)\right)^{-1}.$$

Sketch of the proof. Replace t by $t+1$. Then everything is positive and easy to interpret.

A simple example

Let $A = \{a, b, c\}$ and let $F = \{abc, bcc\}$. Then there are only three clusters:

$$a\,b\,c \qquad b\,c\,c \qquad a\,b\,c\,c$$

So

$$\sum_{w \in A^*} w t^{\#\text{ forbidden words in } w}$$

$$= \left(1 - a - b - c - abc(t-1) - bcc(t-1) - abcc(t-1)^2\right)^{-1}.$$

If we want to avoid all forbidden words, we set $t = 0$ in the Goulden-Jackson cluster theorem. With some work we obtain the following analogue of the CSV theorem.

If we want to avoid all forbidden words, we set $t = 0$ in the Goulden-Jackson cluster theorem. With some work we obtain the following analogue of the CSV theorem.

Theorem. The sum of all words in A^* that avoid the words in F may be written

$$\left(\sum_{w \in A^*} \mu(w) w \right)^{-1}$$

where for every word w, $\mu(w)$ is 0, 1, or -1.

An example

Take $A = \{a\}$ and take $F = \{aaaa\}$. Then

$$\sum_{n=0}^{\infty} \mu(a^n) a^n = (1 + a + a^2 + a^3)^{-1}$$

$$= \left(\frac{1 - a^4}{1 - a}\right)^{-1} = \frac{1 - a}{1 - a^4} \quad = 1 - a + a^4 - a^5 + \cdots$$

so

$$\mu(a^n) = \begin{cases} 1 & \text{if } n \equiv 0 \pmod 4 \\ -1 & \text{if } n \equiv 1 \pmod 4 \\ 0 & \text{otherwise} \end{cases}$$

Why 0, 1, or -1?

We can get a recurrence for computing

$$\sum_{\text{clusters } c \text{ on } w} t^{\#\text{ marked words in } c}.$$

We then set $t = -1$ and use the following lemma:

Why 0, 1, or -1?

We can get a recurrence for computing

$$\sum_{\text{clusters } c \text{ on } w} t^{\#\text{ marked words in } c}.$$

We then set $t = -1$ and use the following lemma:

Let s_n be a sequence of integers defined by $s_1 = 1$ and for $n > 1$,

$$s_n = -(s_{n-1} + s_{n-2} + \cdots + s_{n-\beta(n)}),$$

for some $\beta(n)$, where $1 \leq \beta(n) < n$. Then the nonzero entries of $s_1, s_2, s_3, \ldots$ are $1, -1, 1, -1, 1, \ldots$.

Why 0, 1, or -1?

We can get a recurrence for computing

$$\sum_{\text{clusters } c \text{ on } w} t^{\#\text{ marked words in } c}.$$

We then set $t = -1$ and use the following lemma:

Let s_n be a sequence of integers defined by $s_1 = 1$ and for $n > 1$,

$$s_n = -(s_{n-1} + s_{n-2} + \cdots + s_{n-\beta(n)}),$$

for some $\beta(n)$, where $1 \leq \beta(n) < n$. Then the nonzero entries of $s_1, s_2, s_3, \ldots$ are $1, -1, 1, -1, 1, \ldots$.

Example: Suppose s starts out: $1, 0, -1, 1, 0$. Then the next entry must be 0 or -1.

Note: This result is equivalent to a theorem of Curtis Greene: the Möbius function of a lattice of unions of intervals, under inclusion, is 0, 1, or -1. (C. Greene, A class of lattices with Möbius function $\pm 1, 0$, European J. Combin. 9 (1988), 225–240.)

Susan Parker's and Brian Drake's theses are not published, but they can be found at
`http://people.brandeis.edu/~gessel/homepage/students/`.

Further applications of the CSV theorem can be found in my Ph.D. thesis:
`http://people.brandeis.edu/~gessel/homepage/papers/thesis.pdf`