Discrete Mathematics 34 (1981) 89–91
North-Holland Publishing Company

NOTE

ENUMERATION OF WORDS BY THEIR NUMBER OF MISTAKES

Doron ZEILBERGER*

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received 21 December 1979
Revised 28 May 1980

Consider all words in $\{1, \ldots, n\}$. A fixed set of words is labeled as the set of "mistakes". A generating function for the number of words with $m_1 1$'s, $\ldots, m_n n$'s and k mistakes is given. This generalizes a result of Gessel who considered the case where all the mistakes are two-lettered. A similar result has been independently obtained by Goulden and Jackson.

1.

Fix an alphabet $\{1, \ldots, n\}$. To every word $w = \sigma_1 \cdots \sigma_l$ we associate the monomial $x^w = x_{\sigma_1} \cdots x_{\sigma_l}$ in the non-commuting indeterminates $x_1, \ldots, x_n$. A *subword of* $\sigma_1 \cdots \sigma_l$ is anything of the form $\sigma_i \sigma_{i+1} \cdots \sigma_j$, $1 \leqslant i \leqslant j \leqslant l$. Let L be a set of words to be labeled as "mistakes". We assume that no proper subword of a mistake is a mistake. The number of subwords of w which belong to L is the *number of mistakes of* w and will be denoted by $d(w)$. For example if $L = \{123, 231\}$, $d(1231) = 2$, because both 123 and 231 belong to L. A word w is said to be of type $(m_1, \ldots, m_n)$ if it has m_1 1's, m_2 2's, $\ldots$, m_n n's; e.g. the type of 12112331 is $(4, 2, 2)$. Let M be the set of words w such that every letter of w belongs to some mistake and every mistake, except the last, overlaps, on the right, with another mistake. For example if $L = \{123, 231, 312\}$, $M = \{123, 231, 312, 1231, 2312, 3123, 12312, 23123, 31231, \ldots, \text{etc.}\}$.

The following is a generalization of Theorem 7.2 in Gessel [2]; Gessel's theorem considers the case where L only contains two-lettered words.

Theorem

$$\sum_{w \in \text{all words}} t^{d(w)} x^w = \left[1 - x_1 - \cdots - x_n - \sum_{v \in M} (t-1)^{d(v)} x^v\right]^{-1}. \tag{1}$$

Proof. Let $s(w)$ denote the type of a word w. Let $C(\boldsymbol{m}) = C(m_1, \ldots, m_n)$ be the set of words of type $\boldsymbol{m} = (m_1, \ldots, m_n)$. Define

$$F(\boldsymbol{m}) = \sum_{w \in (\boldsymbol{m})} t^{d(w)} x^w. \tag{2}$$

*Current address: Dept. of Theoret. Math., Weizmann Insitute of Science, Rehovot, Israel.

We shall prove that for $\boldsymbol{m} \neq \boldsymbol{0}$

$$F(\boldsymbol{m}) = \sum_{i=1}^{n} F(\boldsymbol{m} - \boldsymbol{e}_i)x_i + \sum_{v \in M} (t-1)^{d(v)} F(\boldsymbol{m} - s(v))x^v, \tag{3}$$

where $\boldsymbol{e}_i = (0, \ldots, 1, 0, \ldots, 0)$ with the 1 on the ith place.

This will be accomplished by showing that for any $w \in C(\boldsymbol{m})$, the coefficient of x^w in the r.h.s. of (3) is $t^{d(w)}$. Indeed, let w_2 be the maximal tail of w which belongs to M; then $w = w_1 w_2$ for some word w_1, and $d(w) = d(w_1) + d(w_2)$. Note that w has $d(w_2)$ tails which belong to M and thus x^w appear $d(w_2)+1$ times in the r.h.s of (3). Since w loses a mistake by chopping off its last letter and loses $k+1$ mistakes by chopping off a tail which belongs to M and which has k mistakes, the coefficient of x^w in the r.h.s. of (3) is (Put $d(w_1) = d_1$, $d(w_2) = d_2$):

$$t^{d_1}[t^{d_2-1} + (t-1)t^{d_2-2} + (t-1)^2 t^{d_2-3} + \cdots + (t-1)^{d_2-1} + (t-1)^{d_2}] = t^{d_1} t^{d_2} = t^{d(w)}.$$

Here we used (2) with $\boldsymbol{m}$ replaced by $\boldsymbol{m} - \boldsymbol{e}_i$ and $\boldsymbol{m} - s(v)$.

Let $\delta(\boldsymbol{m})$ be the discrete delta function: $\delta(\boldsymbol{0}) = 1$; $\delta(\boldsymbol{m}) = 0$, $\boldsymbol{m} \neq \boldsymbol{0}$. Then, since $F(\boldsymbol{0}) = 1$ and by convention F is zero outside $\mathbf{N}^n$:

$$F(\boldsymbol{m}) - \sum_{i=1}^{n} F(\boldsymbol{m} - \boldsymbol{e}_i)x_i - \sum_{v \in M} (t-1)^{d(v)} F(\boldsymbol{m} - s(v))x^v = \delta(\boldsymbol{m}).$$

Summing both sides over all $\boldsymbol{m} \in \mathbf{Z}^n$ yields

$$\left[\sum_{w \in \text{all words}} t^{d(w)} x^w\right]\left[1 - x_1 - x_2 - \cdots - x_n - \sum_{v \in M} (t-1)^{d(v)} x^v\right] = 1,$$

from which (1) follows.

2. The commutative case

If we let $x_1, \ldots, x_n$ commute in (1) we obtain a generating function for $G(\boldsymbol{m}; k)$, the number of words of type $\boldsymbol{m}$ with exactly k mistakes:

$$\sum G(m_1, \ldots, m_n; k) x_1^{m_1} \cdots x_n^{m_n} t^k = \left[1 - x_1 - \cdots - x_n - \sum_{v \in M} (t-1)^{d(v)} x^v\right]^{-1}. \tag{4}$$

Example. $n = 3$, $L = \{123, 132\}$. Here $L = M$ and

$$\sum G(m_1, m_2, m_3, k) x_1^{m_1} x_2^{m_2} x_3^{m_3} t^k = [1 - x_1 - x_2 - x_3 + 2(1-t)x_1 x_2 x_3]^{-1}.$$

Putting $t = -1$ we get

coefficient of $x_1^{m_1} x_2^{m_2} x_3^{m_3}$ in $[1 - x_1 - x_2 - x_3 + 4x_1 x_2 x_3]^{-1} =$

#{words in $C(m)$ with an even number of mistakes}

−#{words in $C(m)$ with an odd number of mistakes}.

Askey and Gasper [1] proved that the l.h.s. is positive. It will be nice to give a direct proof that the r.h.s. is positive.

Finally let us mention that whenever L is finite but M is infinite it is still possible to evaluate the sum on the r.h.s. of (4) using the geometric series expansion of a certain matrix: $\sum A^k = (I-A)^{-1}$. Thus whenever L is finite the generating function $\sum G(\boldsymbol{m};k)\,\boldsymbol{x}^{\boldsymbol{m}}t^k$ is a rational function. The details are left to the sufficiently interested reader.

Remark. The results of this paper have been obtained independently by Goulden and Jackson [3]. We refer the reader to this very interesting paper for detailed applications and algorithms.

Acknowledgement

Many thanks are due to Ira Gessel for providing us with his thesis and for an illuminating correspondence.

References

[1] R. Askey and G. Gasper, Certain rational functions whose power series have positive coefficients, Amer. Math. Monthly 79 (1972) 337–341.

[2] I. Gessel, Generating Functions and Enumeration of Sequences, Ph.D. thesis, MIT, Cambridge, MA (1977).

[3] I.P. Goulden and D.M. Jackson, An inversion theorem for cluster decompositions of sequences with distinguished subsequences, J. London Math. Soc. (to appear).