for Jim Stonebraker Stockhausen

Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series A
All Rights Reserved by Academic Press, New York and London

Vol. 76, No. 1, October 1996
Printed in Belgium

A Note on the Stockhausen Problem*

Ronald C. Read and Lily Yen

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Communicated by the Managing Editors

Received June 28, 1995

We consider problems in the enumeration of sequences suggested by the problem of determining the number of ways of performing a piano composition (*Klavierstück XI*) by Karlheinz Stockhausen. © 1996 Academic Press, Inc.

1. INTRODUCTION

The score of the piano work *nr. 7 Klavierstück XI* by Karlheinz Stockhausen (1957) [4] consists of 19 fragments of music. The performer is instructed to choose at random one of these fragments, and play it; then choose another, different, fragment and play that, and so on. If a fragment is chosen that has already been played twice, the performance ends. It is natural to ask how many different ways there are of performing this piece, and what the expected length of a performance is.

In this paper we solve the more general problem in which the number of fragments is an integer n, and where the performance terminates when a fragment is chosen that has already been played some given number, r, of times.

There are two versions of the problem according as the identity of the fragment whose choice for the $(r+1)$-th time terminates the performance is considered to be important or irrelevant. The first represents the position of the performer, who knows which choice terminated the performance; the second corresponds to the listener, who does not have this information. We shall therefore refer to these two versions as Problem P and Problem L. If we denote the fragments by symbols from an alphabet of n symbols, we can restate these problems as follows:

* The research for this paper was supported by the Natural Sciences and Engineering Research Council of Canada under grant A8142 and a postdoctoral fellowship.

0097-3165/96 $18.00

Problem P. Determine the number and average length of strings for which

(1) adjacent symbols are different,

(2) the final symbol occurs $r+1$ times,

(3) no other symbol occurs more than r times.

Problem L. Determine the number and average length of strings for which

(1) adjacent symbols are different,

(2) no symbol occurs more than r times,

(3) it is possible to add to the string a symbol that has occurred r times already. In other words, there is an r-times repeated symbol which is not the last symbol in the string.

2. THE BASIC PROBLEM

In order to solve these problems we shall first solve a simpler problem, which we can call Problem B. This is to determine the number of strings of a given length with adjacent symbols different, and such that no symbol occurs more than r times.

First consider strings in which the symbols can occur *any* number of times, so that the only restriction is that adjacent symbols must be different. Let the generating function for these strings be

$$\Phi(x_1, x_2, \ldots, x_n)$$

in which the coefficient of $x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$ is the number of strings in which the ith symbol occurs a_i times ($i = 1, 2, \ldots, n$).

If, in such a string, we replace each symbol y by a non-empty block $yyy \cdots y$, we get strings in which repeated adjacent symbols are allowed. In fact we get all possible strings, since by replacing every block of repeated symbols by a single symbol we recover the original strings. This correspondence is one-to-one.

The replacement of single symbols by a block is reflected by the substitution

$$x_i = t_i + t_i^2 + t_i^3 + \cdots.$$

Hence the generating function for the resulting strings is

$$\Phi\left(\frac{t_1}{1-t_1}, \frac{t_2}{1-t_2}, \ldots, \frac{t_n}{1-t_n}\right). \tag{1}$$

But these strings are the general strings on n symbols, for which the generating function is easily seen to be

$$\frac{1}{1-(t_1+t_2+\cdots+t_n)}$$

which is therefore the same as (1).

Making the substitution $x_i = t_i/(1-t_i)$, that is, $t_i = x_i/(1+x_i)$, we get

$$\Phi(x_1, x_2, \ldots, x_n) = \frac{1}{1-\sum_{i=1}^{n} x_i/(1+x_i)} = \frac{1}{1-(p_1 - p_2 + p_3 - \cdots])},$$

where p_i denotes the i-th power sum symmetric function $x_1^i + x_2^i + \cdots + x_n^i$.

Our task is now to extract from this symmetric function those terms for which no power of an x_i exceeds r, and for this we use the following theorem:

THEOREM. *If $F(p_1, p_2, \ldots)$ is any function of the power sums p_i, then the coefficient of $x_1^{m_1} x_2^{m_2} x_3^{m_3} \cdots x_n^{m_n}$ in $F(p_1, p_2, \ldots,)$ is given by*

$$N(h_1^{m_1} h_2^{m_2} h_3^{m_3} \cdots h_n^{m_n} * F(p_1, p_2, \cdots)). \tag{2}$$

Here, $N(\cdot * \cdot)$ is defined, as in [2], by

$$N(p_1^{j_1} p_2^{j_2} * \cdots * p_1^{k_1} p_2^{k_2} \cdots) = \begin{cases} \prod_i i^{j_i} j_i! & \text{if } j_i = k_i \text{ for all } i, \\ 0 & \text{otherwise,} \end{cases}$$

extended by linearity to any two symmetric functions expressed in terms of the p_i's.

The proof of this theorem is implicit in [2].

To get the total number of strings of length k, say, we must sum the left-hand operand in (2) over all choices of $m_1, m_2, \ldots, m_n$ such that each m_i is less than r and $m_1 + 2m_2 + 3m_3 + \cdots + nm_n = k$. The sum is easily seen to be the coefficient of t^k in

$$(1 + h_1 t + h_2 t^2 + \cdots + h_r t^r)^n.$$

A general term in $F(p_1, p_2, \ldots) = 1/(1-(p_1 - p_2 + p_3 - \cdots))$ is

$$\frac{(\sum_i j_i)!\,(-1)^{j_2+j_4+\cdots}}{\prod_i j_i!} p_1^{j_1} p_2^{j_2} \cdots.$$

Now the $N(\cdot * \cdot)$ operation of this term with a similar term of the left-hand operand in (2) will multiply its coefficient by

$$\frac{(\sum_i j_i)!\,(-1)^{j_2+j_4+\cdots}}{\prod_i j_i!}\prod_i i^{j_i} j_i! = \left(\sum_i j_i\right)!\,(-1)^{j_2+j_4+\cdots}\prod_i i^{j_i}.$$

This is equivalent to the following two operations on the left-hand operand:

(1) replace every p_i by $(-1)^{i+1} i\lambda$,

(2) replace each power λ^m by $m!$.

The final result is then that the generating function, by length, for the Problem B sequences is obtained from

$$(1+h_1 t+h_2 t^2+\cdots+h_r t^r)^n$$

by means of the two operations just given. Denote the solution to this problem by $b_r(n, k)$. When $r=2$, we get

$$b_2(n, k)=n!\sum_{\alpha, \beta}(-1)^{k+\alpha}\binom{\alpha}{k-\alpha-2\beta}\frac{(\alpha+2\beta)!}{(n-\alpha-\beta)!\,\alpha!\,\beta!\,2^\beta}.$$

The total number $b_r(n)$ of strings, irrespective of length, is obtained by putting $t=1$. For $r=2$ we derive $(1+\frac{1}{2}\lambda^2)^n$, whence

$$b_2(n)=\sum_{\beta=0}^{n}\binom{n}{\beta}\frac{(2\beta)!}{2^\beta}.$$

The total length of all the strings can be found in the usual way by differentiating the generating function with respect to t and then setting $t=1$. The formula thus derived is

$$B_2(n)=n\sum_{\beta=0}^{n-1}\binom{n-1}{\beta}\frac{(2\beta+1)(2\beta+1)!}{2^\beta}.$$

3. PROBLEM P

To solve Problem P we first choose a terminating symbol; call it z. Then a typical string is of the form

$$AzB_1zB_2z\cdots zB_rz,$$

where $A, B_1, B_2, \ldots, B_r$ are strings with adjacent symbols different, and only A may be empty.

If we write $\sigma = \sum_{i=1}^{n} x_i/(1+x_i)$, then the generating function for A is $1/(1-\sigma)$ and for each of the B_i's it is $\sigma/(1-\sigma)$. Hence the generating function for the number of ways of choosing A and the B_i's is

$$\frac{\sigma^r}{(1-\sigma)^{r+1}}.$$

We can use similar methods to solve Problem P and obtain

$$c_2(n,k) = n \sum_{\alpha,\beta} \frac{(\alpha+2\beta)!\,(\alpha+2\beta)(\alpha+2\beta-1)(-1)^{k+\alpha+1}}{\beta!\,(n-1-\alpha-\beta)!\,2^{\beta+1}(k-\alpha-2\beta-3)!\,(2\alpha+2\beta+3-k)!},$$

and

$$c_2(n) = n \sum_{\alpha=1}^{n-1} \binom{n-1}{\alpha} \frac{(2\alpha)!\,\alpha(2\alpha-1)}{2^\alpha}.$$

A proof not using symmetric functions is given in Section 5.

For $n=2$, 3 and 4, the formula for $c_2(n)$ yields 2, 114, and 5844; the number of performances (in the present sense) of the Klavierstück (when $n=19$) is

1 74239 35148 33295 81673 10127 28286 29013 34594.

Similarly the total length of all performances is

$$C_2(n) = 3c_2(n) + n(n-1) \sum_{\alpha=0}^{n-2} \binom{n-2}{\alpha} \frac{(2\alpha+1)!\,(2\alpha+1)(2\alpha^2+3\alpha+2)}{2^\alpha}.$$

4. PROBLEM L

Problem L can be solved using the formulae that we have already derived. Clearly, any Problem B string with two or more symbols which occur r times can be terminated. Thus we need to subtract the number of strings which cannot be terminated, and these are of two kinds:

(1) those in which no symbol occurs r times. These are simply Problem B strings using the parameter $r-1$ in place of r.

(2) those in which exactly one symbol occurs r times and is the final symbol in the string. These strings are precisely the Problem P strings, again with $r-1$ in place of r.

Hence the numbers and lengths of the Problem L strings can be found from data already available.

$$d_r(n, k) = b_r(n, k) - b_{r-1}(n, k) - c_{r-1}(n, k),$$
$$d_r(n) = b_r(n) - b_{r-1}(n) - c_{r-1}(n).$$

In the case $r = 2$, this yields

$$d_2(n) = \sum_{\beta=1}^{n} \binom{n}{\beta} \left(\frac{(2\beta)!}{2^\beta} - \beta!\, \beta \right).$$

Similarly, the total length of these strings can be found and is

$$D_2(n) = B_2(n) - \sum_k \frac{n!\, k}{(n-k)!} - n! \sum_{i \geq 2} \frac{i(i-2)}{(n-i+1)!}.$$

With this interpretation, perhaps the more natural one, of what constitutes a "performance" the number of ways of performing the Klavierstück turns out to be

10249 37361 66664 45980 71114 32876 93179 82974.

The first three values for $n = 2$, 3, and 4 are 2, 78, and 2724.

5. AN ALTERNATIVE APPROACH TO PROBLEM P

We use symmetric functions differently in obtaining a short derivation of the generalized Stockhausen numbers. A combinatorial proof is given at the end of the section for the original Stockhausen strings.

Let m be the terminal symbol that occurs $r+1$ times, and each of the other symbols occur at most r times. As in Section 3, let $c_r(n, k)$ be the number of strings of length k on n symbols satisfying the conditions of Problem P.

Then the decomposition from Section 3 together with the coefficient extracting operator gives the generating series

$$\sum_{k \geq 0} c_r(n, k)\, z^k = \sum_{m=1}^{n} \sum_{\substack{0 \leq i_1, \ldots, i_n \leq r \\ i_m = r+1}} z^{r+1} [x_1^{i_1} \cdots x_n^{i_n}] \times \left(\sum_{j=1}^{n} x_j^{r+1} \right) \frac{\left(\sum_{j=1}^{n} zx_j/(1+zx_j)\right)^r}{\left(1 - \sum_{j=1}^{n} zx_j/(1+zx_j)\right)^{r+1}}. \tag{3}$$

Let the inner sum of (3) be denoted by $T_r(n, m)$, so

$$\sum_{k \geqslant 0} c_r(n, k)\, z^k = \sum_{m=1}^{n} T_r(n, m),$$

where

$$T_r(n, m) = z^{r+1}[x_1^r x_2^r \cdots x_{m-1}^r x_m^{r+1} x_{m+1}^r \cdots x_n^r] \left(\prod_{j=1}^{n} \frac{1}{1-x_j}\right) \times \left(\sum_{j=1}^{n} x_j^{r+1}\right) \frac{(\sum_{j=1}^{n} zx_j/(1+zx_j))^r}{(1-\sum_{j=1}^{n} zx_j/(1+zx_j))^{r+1}}.$$

We observe that

$$\left(\prod_{j=1}^{n} \frac{1}{1-x_j}\right)\left(\sum_{j=1}^{n} x_j^{r+1}\right) \frac{(\sum_{j=1}^{n} zx_j/(1+zx_j))^r}{(1-\sum_{j=1}^{n} zx_j/(1+zx_j))^{r+1}}$$

is a symmetric function of $x_1, ..., x_n$, so $T_r(n, m)$ is independent of m, for $1 \leqslant m \leqslant n$. Therefore

$$\sum_{k \geqslant 0} c_r(n, k)\, z^k = nT_r(n, n).$$

Now

$$\begin{aligned} T_r(n, n) &= z^{r+1}[x_1^r x_2^r \cdots x_{n-1}^r x_n^{r+1}] \left(\prod_{j=1}^{n} \frac{1}{1-x_j}\right) \\ &\quad \times \left(\sum_{j=1}^{n} x_j^{r+1}\right) \frac{(\sum_{j=1}^{n} zx_j/(1+zx_j))^r}{(1-\sum_{j=1}^{n} zx_j/(1+zx_j))^{r+1}} \\ &= z^{r+1}[x_1^r x_2^r \cdots x_{n-1}^r] \left(\prod_{j=1}^{n} \frac{1}{1-x_j}\right) \frac{(\sum_{j=1}^{n} zx_j/(1+zx_j))^r}{(1-\sum_{j=1}^{n} zx_j/(1+zx_j))^{r+1}}. \end{aligned}$$

We next capitalize on the multiplicative structure of $[x_1^r x_2^r \cdots x_{n-1}^r]$ and $\prod_{j=1}^{n-1} 1/(1-x_j)$ by transforming

$$\frac{(\sum_{j=1}^{n} zx_j/(1+zx_j))^r}{(1-\sum_{j=1}^{n} zx_j/(1+zx_j))^{r+1}}.$$

into a multiplicative expression, then isolate each x_i.

Let

$$L_w^{(r)} f = \frac{1}{r!} \frac{\partial^r f}{\partial w^r}\bigg|_{w=1} \qquad \text{and} \qquad \Theta_w\colon w^k \mapsto k!\, w^k.$$

Then

$$\begin{aligned}
T_r(n,n) &= z^{r+1}[x_1^r x_2^r \cdots x_{n-1}^r]\left(\prod_{j=1}^{n-1}\frac{1}{1-x_j}\right)L_w^{(r)}\left(\frac{1}{(1-w\sum_{j=1}^{n-1} zx_j/(1+zx_j))}\right)\\
&= z^{r+1}[x_1^r x_2^r \cdots x_{n-1}^r]\left(\prod_{j=1}^{n-1}\frac{1}{1-x_j}\right)L_w^{(r)}\Theta_w \exp\left(w\sum_{j=1}^{n-1}\frac{zx_j}{1+zx_j}\right)\\
&= z^{r+1}L_w^{(r)}\Theta_w[x_1^r x_2^r \cdots x_{n-1}^r]\left(\prod_{j=1}^{n-1}\frac{1}{1-x_j}\right)\left(\prod_{j=1}^{n-1}\exp\left(\frac{wzx_j}{1+zx_j}\right)\right)\\
&= z^{r+1}L_w^{(r)}\left([x^r]\frac{1}{1-x}\exp\left(\frac{wzx}{1+zx}\right)\right)^{n-1}
\end{aligned}$$

so

$$\sum_{k\geq 0} c_r(n,k)\,z^k = nz^{r+1}L_w^{(r)}\Theta_w\left([x^r]\frac{1}{1-x}\exp\left(\frac{wzx}{1+zx}\right)\right)^{n-1}.$$

For the original Stockhausen number we let $r=2$ and $z=1$ and obtain

$$\begin{aligned}
\sum_{k\geq 0} c_2(n,k) &= nL_w^{(2)}\Theta_w\left([x^2]\frac{1}{1-x}\exp\left(\frac{wx}{1+x}\right)\right)^{n-1}\\
&= n\sum_{j=0}\binom{n-1}{j}j(2j-1)\frac{(2j)!}{2^j}.
\end{aligned}$$

This simple expression seems to suggest a direct combinatorial proof which we show next. Given n symbols, there are n ways of choosing a symbol to be the terminal symbol that occurs three times. For the remaining $n-1$ symbols, there are $\binom{n-1}{j}$ ways of choosing j symbols to use in making $(2j)!/2^j$ strings such that each of the j symbols appears twice with no adjacency restrictions, and $\binom{2j}{2}$ ways of marking two positions on each such string. The terminal symbol is placed immediately before the two marked positions and of course at the terminal position of the string, thus making $n\sum_j\binom{n-1}{j}\binom{2j}{2}(2j)!/2^j$. By replacing any repeated symbols yy by a single symbol y, we obtain strings whose adjacent symbols are all distinct, and the correspondence is one-to-one.

6. ASYMPTOTIC ANALYSIS FOR PROBLEMS P AND L

We determine the average lengths of strings for Problems P and L as the number of symbols (n) tends to infinity. A theorem of Pólya and Szegő is our main tool.

and deduce that this term is asymptotic to

$$\frac{(2n-3)!\,(2n-3)\,n(n-1)(2n^2-5n+4)}{2^{n-2}}.$$

From these results we see that the average length of the P-sequences is asymptotic to

$$3+\frac{2n^2-5n+4}{n-1}=2n+\frac{1}{n-1}.$$

Thus the average length approaches $2n$ from above.

Similar analysis, which we do not give, shows that the average length of the L-sequences tends to $2n-1$ from below. Hence, in both cases, the expected lengths are one less than the maximum possible length. For the original Stockhausen problem when $n=19$, the average length of the P-sequences is about 38.0046857, and 36.9458622 for the L-sequences.

Notation. Readers wishing to have more details of the derivations of the formulae given above can consult the research report [3], which also gives all numerical values up to $n=20$.

ACKNOWLEDGMENTS

We thank D. M. Jackson for many helpful suggestions and M. Lemvig Hansen for unearthing Stockhausen's score from the library of Wilfrid Laurier University. We also thank the referee for many helpful recommendations.

REFERENCES

1. G. Pólya and G. Szegő, "Aufgaben und Lehrsätze aus der Analysis," Erster Band, I. Abschn., Kap. 4 §2, Nr. 178, p. 32, Berlin, 1925.
2. R. C. Read, The use of S-functions in combinatorial analysis, *Canad. J. Math.* **20** (1968), 808–841.
3. R. C. Read and L. Yen, "The Stockhausen Problem and Its Generalizations," Research Report CORR 95-06, pp. 13, University of Waterloo, Dept. of Combinatorics & Optimization, 1995.
4. K. Stockhausen, "nr. 7 Klavierstück XI," 12654 LW, Universal Edition, London, 1979.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium