ABSTRACT

The purpose of this paper is to present a direct and simpler combinatorial proof of a theorem of Yu. M. Vološin [4] on the enumeration of function compositions and to exhibit some of the consequences of this theorem. Many consequences are stated in the paper of Vološin, however, his methods are relatively intractable. Here, we obtain a generating function which facilitates enumeration. The methods and arguments employed here should be compared with Vološin. The combinatorial structure encountered here is a fairly general one with many applications, only a few of which are provided in the present paper.

AMS(MOS) Subject Classification: 05A15

Key Words: Enumeration of Function Composition

Work Unit No.  4 (Probability, Statistics, and Combinatorics)

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.
ON A THEOREM OF VOLOŠIN CONCERNING ENUMERATION OF FUNCTION COMPOSITIONS

Bernard Harris

1. Introduction and Summary. The purpose of this paper is to present a direct and simpler combinatorial proof of a theorem of Yu. M. Vološin [4] on the enumeration of function compositions and to exhibit some of the consequences of this theorem. Many consequences are stated in the paper of Vološin, however, his methods are relatively intractable. Here, we obtain a generating function which facilitates enumeration. The methods and arguments employed here should be compared with Vološin. The combinatorial structure encountered here is a fairly general one with many applications, only a few of which are provided in the present paper.

Let \( c_0 \) be the number of symbols denoting variables and let \( c_j \) be the number of symbols denoting functions of \( j \) variables, \( j = 1, 2, \ldots, n \). Let \( k_0 \) be the number of entries of variable symbols and for each \( j, j = 1, 2, \ldots, n \) let \( k_j \) be the number of entries of symbols denoting functions of \( j \) variables. Then Vološin established the following theorem.

**Theorem 1.** For given \( c_j \geq 0, k_j \geq 0, j = 0, 1, 2, \ldots, n \), the number of valid compositions that can be made with arbitrary insertions of commas and parentheses is

\[
R(k_0, k_1, \ldots, k_n; c_0, c_1, \ldots, c_n) = \frac{(k_0 + k_1 + \ldots + k_n)!}{k_0! k_1! \ldots k_n!} c_0^{k_0} c_1^{k_1} \cdots c_n^{k_n},
\]

provided that

\[
\sum_{j=1}^{n} (j-1)k_j = k_0 - 1.
\]

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.
The relation (2) is both necessary and sufficient for \( k_0, k_1, \ldots, k_n \) to permit a valid composition. Also note that \( c_j = 0 \) implies \( k_j = 0 \), since if \( k_j > 0 \), then (1) yields zero, which is appropriate.

The following example should aid the reader by clarifying the definitions and notation. Let \( x \) and \( y \) denote variables, let \( f \) be a symbol for a function of one variable and let \( g \) denote a function of two variables. Then we have \( c_0 = 2, c_1 = 1, c_2 = 1 \). Let \( k_0 = 2, k_1 = 1, k_2 = 1 \). Then the 12 solutions given by (1) are

\[
\begin{align*}
fgxx & \quad fgxy & \quad fgxy & \quad fgyy \\
gfxx & \quad gfixy & \quad gfxy & \quad gfy y \\
gx fx & \quad gfxy & \quad gyfx & \quad gy fy .
\end{align*}
\]

For example \( gxfy \) is a composition, since we can write \( g(x, f(y)) \). However \( xgfy \) is not a valid composition. It is easily seen that there is no way of inserting commas and parentheses so that this is a valid composition.

In [4], Vološin actually stated Theorem one in an equivalent, but substantially more complicated form. The differences are that he defined \( k_i, i = 1, 2, \ldots, n \) as the number of entries of functions of \( m_1 \) variables. The number of symbols \( c_i \) was correspondingly defined and he required \( c_i > 0, i = 0, 1, \ldots, n \). As will be seen in the sequel, letting \( m_1 = i \) and permitting \( c_i = 0 \) substantially simplifies all computations.

In section two, Theorem one is established. In section three, generating functions are derived, which are then used to exhibit other results of Vološin. In particular, the treatment given here provides simple explicit expressions for many special cases.
2. **Proof of Vološin's Theorem.** We first calculate the totality of possible arrangements and selections of the $N = k_0 + k_1 + \ldots + k_n$ symbols without regard to whether they constitute a valid composition. Then we subsequently determine the fraction of these which are valid compositions.

There are clearly $N!/(k_0!k_1!\ldots k_n!)$, ways to select the $k_j$ positions to be occupied by the symbols selected from the $c_j$ symbols, $j = 1, 2, \ldots, n$. After this allocation of positions, there are $c_j^k$ possible selections for each position and hence $c_j^k$ selections for each $j$. Thus the totality of arrangements and selections without regard to their validity as a function composition is

$$M(k_0, k_1, \ldots, k_n; c_0, c_1, \ldots, c_n) = \frac{(k_0 + k_1 + \ldots + k_n)!}{k_0!k_1!\ldots k_n!} \cdot \frac{k_0!k_1!\ldots k_n!}{c_0!c_1!\ldots c_n!}.$$ 

Now consider a sequence of $N$ symbols selected from the $M$ possibilities available, say $\tilde{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$. We characterize those sequences, which, by insertion of commas and parentheses, are valid compositions. Define $g(\sigma) = -(j-1)$ if $\sigma_j$ is a symbol denoting a function of $j$ variables; $j = 0$ means that $\sigma_j$ is a symbol denoting a variable. Accordingly we establish the following lemma.

**Lemma 1.** $\tilde{\sigma}$ is a valid function composition if and only if

$$h(m) = \sum_{r=m}^{N} g(\sigma_r) > 0, \quad m = 1, 2, \ldots, N$$

and

$$h(1) = 1.$$

**Proof.** Necessity. Assume that $(\sigma_1, \sigma_2, \ldots, \sigma_N)$ is a valid function composition. If $N = 1$, $\sigma_1$ must be a variable symbol. Hence $g(\sigma_1) = 1 = h(1)$. -3-
Now assume that the conclusion holds for $N = 1, 2, \ldots, k-1$, $k > 1$. Consider therefore any valid function composition with $N = k$ symbols. Then $\alpha_1$ must be a symbol for a function of $j$ variables for some $j$, $1 \leq j \leq k-1$. Hence, we must be able to decompose $(\alpha_2, \ldots, \alpha_k)$ into $j$ sequences of consecutive elements, namely, $(\alpha_2, \ldots, \alpha_{p_1})$, $(\alpha_{p_1+1}, \ldots, \alpha_{p_1+p_2})$, \ldots, $(\alpha_{p_1+p_2+\ldots+p_{j-1}+1}, \ldots, \alpha_{p_1+p_2+\ldots+p_j})$, where $p_1+p_2+\ldots+p_j = k$, such that each of these $j$ sequences is a valid composition. Each sequence is to be separated from the sequence which follows it by a comma and we place a parenthesis before $\alpha_2$ and after $\alpha_k$. As a consequence of the induction hypothesis, we have

$$
\begin{align*}
&\begin{cases} 
    h_1(m) = \sum_{r=m}^{p_1} g(\alpha_r) > 0, & m = 2, 3, \ldots, p_1, \\
    h_2(m) = \sum_{r=p_1+m}^{p_1+p_2} g(\alpha_r) > 0, & m = 1, 2, \ldots, p_2 \\
    \vdots \\
    h_j(m) = \sum_{r=p_1+p_2+\ldots+p_{j-1}+m}^{p_1+p_2+\ldots+p_j} g(\alpha_r) > 0, & m = 1, 2, \ldots, p_j 
\end{cases}
\end{align*}
$$

and

$$
\begin{align*}
&h_1(1) = h_2(1) = \ldots = h_j(1) = 1.
\end{align*}
$$

Consequently

$$
\begin{align*}
&h(m) = \sum_{r=m}^{k} g(\alpha_r) > 0, & m = 2, \ldots, k
\end{align*}
$$

and

$$
\begin{align*}
&h(1) = -(j-1) + \sum_{i=1}^{j} h_i(1) = -(j-1) + j = 1,
\end{align*}
$$

verifying (4) and (5).
Sufficiency. Assume that the sequence \( (a_1, a_2, \ldots, a_N) \) satisfies (4) and (5). Then \( h(N) = 1 \) and \( a_N \) must be a variable symbol. Each symbol for a function of \( j \) variables may be regarded as replacing \( j \) variables occurring to the right of the symbol by one variable, thus reducing the number of variables to the right by \( (j-1) = -g(\alpha) \). Hence, for each \( m, m = 1, 2, \ldots, N, \) \( h(m) \) is the number of variables available at \( m \). Hence, if (4) holds, then for each \( m, 1 \leq m \leq N, \) there are variables available at \( m \). If (5) holds as well, then the sequence \( \vec{a} \) may be regarded as a single variable and is therefore a valid composition.

A simple rule for insertion of commas and parentheses can easily be given.

We now obtain Volosin's condition (2).

Lemma 2. \( h(1) = 1 \) implies Volosin's condition holds. Further for any sequence \( \vec{a} \) satisfying Volosin's condition, there is exactly one cyclic permutation which is a valid function composition.

Proof. \( h(1) = \sum_{r=1}^{N} g(\alpha_r) = -\sum_{j=1}^{n} (j-1)k_j + k_0 = 1, \)
which is Volosin's condition.

For the converse, we can apply a well-known theorem of L. Takács [3]; if \( r_1, r_2, \ldots, r_N \) are non-negative integers with \( \sum_{i=1}^{N} r_i = r \leq N \), then among the \( N \) cyclic permutations, there are exactly \( N-r \) such that the partial sums \( \sum_{i=1}^{m} r_i \), are less than \( m \), \( m = 1, 2, \ldots, N. \)

To apply Takács' theorem, let \( r_i = 1 - g(\alpha_i), \) then since \( -(n-1) \leq g(\alpha_i) \leq 1, \) we have \( 0 \leq r_i \leq n. \) Also \( \sum_{i=1}^{N} g(\alpha_i) = 1 \) implies

-5-
\[ \sum_{i=1}^{N} r_i = N - \sum_{i=1}^{N} g(\alpha_i) = N - 1. \] 
Finally, \[ \sum_{i=m}^{N} g(\alpha_i) = (N-m+1) - \sum_{i=m}^{n} r_i > 0 \] 
is equivalent to \[ \sum_{i=m}^{N} r_i < N - m - 1, \quad m = 1, 2, \ldots, N. \] 
Thus (4) and (5) are precisely Takács' conditions with \( r = N-1. \)

Combining Lemmas one and two, we have established the theorem, since among the \( M(k_0, k_1, \ldots, k_n; c_0, c_1, \ldots, c_n) \) arrangements and selections, exactly \( R(k_0, k_1, \ldots, k_n; c_0, c_1, \ldots, c_n) = (N-1)M(k_0, k_1, \ldots, k_n; c_0, c_1, \ldots, c_n) \) of them constitute a valid composition.

**Remarks.** In [4], Vološin established the theorem by first obtaining a complicated recursion for \( R \), the number of compositions. This recursion was employed to deduce a functional equation for the generating function

\[
H_R(t_0, t_1, \ldots, t_n) = \sum_{k_0=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} R(k_0, k_1, \ldots, k_n; c_0, c_1, \ldots, c_n) t_0^k_0 t_1^k_1 \cdots t_n^k_n.
\]

Lagrange's formula was employed to expand \( H_R \) obtaining the desired result.

In the opinion of the present author, the proof given here provides more insight into the relevant combinatorial mechanism and clarifies the combinatorial significance of the Vološin condition (2).

R. L. Graham has pointed out to the author that he had previously seen results of the type of Lemma one. The author has investigated this comment and notes that a result very similar to Lemma one may be found in P. C. Rosenbloom [2], pp. 152-157.

In his paper, Volosin [4] discussed properties of the generating function (8), but never gave an explicit form. In this section, we show that by making a different choice of generating function, a useful explicit representation can be obtained. This facilitates computation and enables us to obtain other results from Volosin's paper as well as many additional results. Therefore, we define

$$F_\beta(x_0, x_1, \ldots, x_n; u; t) = e^{t \sum_{j=0}^{n} c_j x_j u^j},$$

where $\beta = (c_0, c_1, \ldots, c_n)$. Then we obtain Theorem 2.

**Theorem 2.** $N^{-1}$ times the coefficient of $\frac{N}{N'} x_0 x_1 \ldots x_n$ is the number of valid function compositions that can be made with $(c_0, c_1, \ldots, c_n; k_0, k_1, \ldots, k_n)$ specified.

**Proof.** By direct computation, we have

$$F_\beta(x_0, x_1, \ldots, x_n; u; t) = \sum_{N=0}^{\infty} \frac{N!}{k_0! k_1! \ldots k_n!} \sum_{k_0 k_1 \ldots k_n} \frac{N!}{N'} x_0 x_1 \ldots x_n u^{j(k)}.$$

Setting $k_1 + 2k_2 + nk_n = N - 1 = \sum_{j=0}^{n} k_j - 1$, we have Volosin's condition (2) and the theorem is verified.

With no loss of generality, we can remove the restriction imposed by the symbol $n$ and consider

$$F_\beta(x_0, x_1, \ldots, u; t) = e^{t \sum_{j=0}^{n} c_j x_j u^j},$$

-7-
where \( \tilde{c} = (c_0, c_1, \ldots) \). This reduces to (9) on setting \( c_j = 0 \) for all \( j > n \).

Volosi\'n has introduced a number of attributes of a function composition. They are: \((k_0, k)\), the characteristic of a composition, where \( k_0 \) is the number of entries of variable symbols and \( k = k_1 + k_2 + \ldots + k_n \) is the number of entries of function symbols; \( k_0 \) the valency of a composition; \( k \), the complexity of a composition. \( N = k_0 + k \) is called the length of a composition.

We denote the number of compositions with specified characteristic, valency, complexity and length by \( A_{k_0, k}(\tilde{c}) \), \( B_{k_0}(\tilde{c}) \), \( C_{k}(\tilde{c}) \) and \( N(\tilde{c}) \) respectively.

All of these can be readily enumerated using the generating functions (9) or (11), Specifically,

1. \( A_{k_0, k}(\tilde{c}) \) is \((k_0+k)^{-1} \) times the coefficient of

\[
\frac{1}{(k_0+k)!} \sum_{k=0}^{k_0+k} \binom{k_0+k-1}{k} \left( \prod_{i=0}^{k} \partial_{x_i} \right) \left( \prod_{i=0}^{k} x_i \right) \]

2. \( B_{k_0}(\tilde{c}) = \sum_{k=0}^{N} A_{k_0, k}(\tilde{c}), \quad C_{k}(\tilde{c}) = \sum_{k=0}^{\infty} A_{k_0, k}(\tilde{c}) \),

3. \( N(c) = \sum_{k=0}^{N} A_{k_0, k}(\tilde{c}) \).

However, in many special cases, we can obtain simple expressions for quantities of interest.

The following observations are immediate.

\( B_{k_0}(\tilde{c}) \) will be infinite whenever \( c_1 > 0 \) and finite whenever \( c_1 = 0 \). \( C_{k}(\tilde{c}) \) will be infinite unless there exists an \( n > 0 \) such that \( c_m = 0 \) for all \( m > n \).
To observe the first, note that if $f$ denotes a function of one variable, then for any $\tilde{a} = (a_1, \ldots, a_N)$ which is a valid function composition, $(f, a_1, \ldots, a_N)$ is also a valid function composition. Clearly, this can be continued indefinitely. However, if $c_1 = 0$, then the maximal length of any possible valid composition is bounded. This follows immediately from (2).

For the second, let $f_m$ denote a symbol for a function of $m$ variables. Then choose any subsequence $\{m_i\}$, with $c_{m_i} > 0$ for all $m_i$ and $m_i \to \infty$.

Then for every $k$, and every $k$-tuple $(m_1, m_2, \ldots, m_k), m_i > 0$ in the subsequence and any variable symbol $x$,

$$f_{m_1}(..., f_{m_2}(x, \ldots, x), f_{m_1}(x, \ldots, x))...$$

is a valid function composition with $k$ function symbols, $k \geq 1$.

We now show how the generating function (II) may be employed to obtain various combinatorial results and to obtain explicit representations for various quantities which arise in the study of functional compositions.

Let $c_m = c_{m+1} = \ldots = c_n = c$, $c_1 = 0$, $1 \leq i \leq m-1$, $n+1 \leq i$, $1 \leq m \leq n$, and $x_1 = x_2 = \ldots = x$. Then

$$F_G(x_0, x, x, \ldots ; u; t) = e^{\sum_{j=m}^{N} \frac{t^j}{j!}(c_0 x_0 + c x \sum_{j=m}^{n} u^j)^N}.$$

Hence, we have

$$F_G(x_0, x, x, \ldots ; u; t) = \sum_{N=0}^{\infty} \frac{t^N}{N!} \left(c_0 x_0 + c x \sum_{j=m}^{n} u^j\right)^N.$$

-9-
Thus, we get
\[ F_{\infty}(x, x, x, \ldots; u; t) = \sum_{N=0}^{\infty} \frac{t^N}{N!} \left( \sum_{r=0}^{N} \frac{N!}{r!} \frac{x^r}{x^0} c_0 \sum_{k=0}^{\infty} \left( \frac{r+k}{k} \right) u^k \sum_{j=0}^{\infty} (-1)^j u^{(n+1-m)j} \right). \]

To apply Theorem 2, we need to calculate the coefficient of \( \frac{t^N}{N!} u^{N-1} \).

Direct calculation shows that this is given by
\[
\begin{align*}
N = 0, & \quad 0 \\
N = 1, & \quad c_0 x_0 \\
N > 1, & \quad \sum_{r=1}^{\infty} \left( \sum_{r=1}^{\infty} \left( \sum_{j=0}^{\infty} (-1)^j \left( \frac{N-2+r-rm-nj-j+m}{j} \right)^{r-1} \right) \right) \frac{N!}{(N-r)!} \frac{x^r}{x^0} c_0.
\end{align*}
\]

Thus, we have established the following.

**Theorem 3.** \( A_{N-r,r}(c) \), the number of valid compositions of length \( N \) with characteristic \((N-r,r)\) when \( c_0 = c_{m+1} = \ldots = c_n = c \), \( c_j = 0 \), \( 1 \leq j < m \), \( j \geq n+1 \) and \( 1 \leq m \leq n \) is given by:
\[
c_0, \quad \text{when } N = 1, \ r = 0.
\]

For \( 1 \leq r \leq N-1, \ N > 1 \)
\[
A_{N-r,r}(c) = \begin{cases} 
\frac{1}{N} \left( \frac{N-r}{r} \right) \frac{N!}{(N-r)!} \frac{x^r}{x^0} c_0 \sum_{j=0}^{\infty} (-1)^j \left( \frac{N-2+r-rm-nj-j+m}{j} \right)^{r-1}, & \text{when } r \leq \frac{N-1}{n+1}, \\
0, & \text{otherwise},
\end{cases}
\]

where \( M = \left[ \frac{N-1-rm}{n+1-m} \right] \).
This theorem has many interesting consequences. For purpose of illustration, we cite some of these.

**Corollary 1.** The number of valid compositions of length $N$ with characteristic $(N-r, r)$ when $m = 1$, $c_1 = c_2 = \ldots = c_n = c$ and $n \geq N-1$ is given by

$$A_{N-r, r}(\tilde{c}) = \frac{1}{N} \binom{N}{r} \binom{N-2}{r-1} c^r, \quad N > 1, \quad 1 \leq r \leq N-1.$$  

**Proof.** Observe that $n \geq N-1$ implies $(N-1)/(n+1) \leq (N-1)/N < 1$ and $(N-1)/m = N-1$. Hence the upper limit of the summation in (12) is $M$. But in this case $M = [(N-1-rm)/(n+1-m)] \leq [(N-1-r)/(N-1)] = 0$. Thus $A_{N-r, r}(\tilde{c})$ is given by (13).

**Corollary 2.** When $c_0 = c_1 = \ldots = c_n = 1$, $n \geq N-1$, the number of valid compositions of length $N$ is given by the Catalan numbers: that is

$$N(\tilde{c}) = \frac{1}{N} \left( \frac{2(N-1)}{N-1} \right) \quad N = 1, 2, \ldots.$$  

**Proof.** For $N = 1$ the conclusion is trivial. Hence assume $N > 1$. Then, from Corollary 1, we have

$$N(\tilde{c}) = \frac{1}{N} \sum_{r=1}^{N-1} \binom{N}{r} \binom{N-2}{r-1} = \frac{1}{N} \sum_{s=0}^{N-2} \binom{N-2}{s+1} \binom{N}{s+1} = \frac{1}{N} \left( \frac{2N-2}{N-1} \right),$$  


An alternative method of obtaining this result is to exhibit a one-to-one correspondence between function compositions and a certain class of rooted trees with $N$ vertices. These trees are referred to by Vološin [4] as functional trees. The endpoints of such trees correspond to symbols for variables and vertices of degree $j$ refer to symbols for functions of $j$ variables.
To provide an additional illustration of the applications of the generating function (11), we now establish the following theorem.

**Theorem 4.** When \( c_0 = c_1 = \ldots = c_n = 1 \), the number of valid function compositions of length \( N \) is given by

\[
N(c) = \sum_{j=0}^{P} (-1)^j \begin{pmatrix} 2N-2-(n+1) \end{pmatrix}_{N-1} \quad 1 \leq n \leq N-1,
\]

where \( P = \left( \frac{(N-1)(n+1)}{n+1} \right) \).

**Proof.** From (11), we have that the number of compositions of length \( n \) is given by \( N^{-1} \) times the coefficient of \( \frac{t^N}{N!} u^{N-1} \) in

\[
\frac{1}{N} e^{t(1 + u + \ldots + u^n)} = \frac{1}{N} e^{t(1 - u^{n+1})(1-u)^{-1}}
\]

Hence the coefficient of \( \frac{t^N}{N!} \) is

\[
\frac{1}{N!} (1 - u^{n+1}) (1-u)^{-N}.
\]

Expanding this, we obtain

\[
\frac{1}{N} \sum_{j=0}^{N} (-1)^j \begin{pmatrix} N \end{pmatrix}_j u^{(n+1)j} \sum_{m=0}^{\infty} \begin{pmatrix} N+m-1 \end{pmatrix}_m u^m
\]

\[
= \frac{1}{N} \sum_{r=0}^{\infty} \sum_{j=0}^{N} u^r (-1)^j \begin{pmatrix} N \end{pmatrix}_j \begin{pmatrix} N+r-(n+1) \end{pmatrix}_{j-1}.
\]

Thus, the coefficient of \( u^{N-1} \) is

\[
\frac{1}{N} \sum_{j=0}^{N} (-1)^j \begin{pmatrix} N \end{pmatrix}_j \begin{pmatrix} 2N-2-(n+1)j \end{pmatrix}_{N-1}
\]

and the conclusion follows readily.

**Remark.** For \( n = N-1 \), this is again just the Catalan numbers, as it should be.

For \( n = 1 \), this provides a proof of another well-known combinatorial identity (H. W. Gould [1], formula 3.16). Then the only valid function
composition of length $N$ is of the form $f(f \ldots f(x))$, where $f$ is a function of a single variable. Obviously there is one such composition. Thus

$$\sum_{j=0}^{\frac{N-1}{2}} (-1)^j \binom{N}{j} \binom{2N-2-2j}{N-1} = N.$$ 

4. **Concluding Remarks.** The combinatorial structures given by functional composition are a very broad class of combinatorial structures and consequently there are many interesting specializations which can be obtained from these results. Some of these have been given in Theorem 3 and its corollaries and Theorem 4. In addition, these structures have many relations to other combinatorial problems. Some of these have already been alluded to by Vološin [4].

**References**


The purpose of this paper is to present a direct and simpler combinatorial proof of a theorem of Yu. M. Volosin [4] on the enumeration of function compositions and to exhibit some of the consequences of this theorem. Many consequences are stated in the paper of Volosin; however, his methods are relatively retractable. Here, we obtain a generating function which facilitates enumeration. The methods and arguments employed here should be compared with Volosin. The combinatorial structure enumerated here is a fairly general one with many applications, only a few of which