## SOME PROPERTIES OF COMPOSITIONS AND THEIR APPLICATION TO THE BALLOT PROBLEM

## S. G. Mohanty

(received July 27, 1964)

- 1. Introduction and summary. This paper is a continuation of two papers [4], [5] and brings out the solution of the ballot problem in its general form.
- In [5], Narayana has considered a generalised occupancy problem which can be viewed as a problem in compositions of integers. In what follows, we use the definitions of [6]. Furthermore, we say that an r-composition  $(t_1(m), \ldots, t_r(m))$  of m dominates an r-composition  $(t_1(n), \ldots, t_r(n))$  of n (m > n) if and only if

(1) 
$$\sum_{\alpha=1}^{i} t_{\alpha}(m) \geq \sum_{\alpha=1}^{i} t_{\alpha}(n), \quad \text{for } i = 1, \dots, r.$$

Evidently  $\sum_{\alpha=1}^{r} t_{\alpha}(m) = m$  and  $\sum_{\alpha=1}^{r} t_{\alpha}(n) = n$ . For integers  $\alpha=1$   $n_1,\ldots,n_k$  such that  $n_1 \geq \ldots \geq n_k$ , we are required in [5] to determine the number of r-compositions of  $n_1$  that dominate r-compositions of  $n_2$ , that in turn dominate r-compositions of  $n_3$ , and so on. In other words, we are looking for the number of elements in the set  $C = C(n_1,\ldots,n_k;r)$  =  $\{(t_1(n_1),\ldots,t_r(n_1)),\ldots,(t_1(n_k),\ldots,t_r(n_k)):$ 

Canad. Math. Bull. vol. 8, no. 3, April 1965

(2) 
$$\sum_{\alpha=1}^{i} t_{\alpha}(n_{j}) \geq \sum_{\alpha=1}^{i} t_{\alpha}(n_{j+1})$$

for i = 1, ..., r and j = 1, ..., k-1.

Now, corresponding to C, consider the set of lattice paths in a k-dimensional Euclidean space with axes  $X_{\ell}$ 's such that the  $(k(i-1)+\ell)$ th segment  $(i=1,\ldots,r)$  and  $\ell=1,\ldots,k$  of any path is the distance  $t_i(n_{\ell})$  on  $X_{\ell}$ . Suppose a step in a path represents k consecutive segments beginning from the segment on  $X_{\ell}$ . Then the set consists of paths from the origin to  $(n_{\ell},\ldots,n_{k})$  not crossing the region bounded by  $X_{\ell}=0$ ,  $X_{\ell}=X_{\ell}=1,\ldots,k-1$  and having exactly r components. Denote this set by  $L*(n_{\ell},\ldots,n_{k};r)$  or briefly L\*. Thus the above construction has established a 1:1 correspondence between C and L\*. Letting  $N\{.\}$  represent the number of elements in the set  $\{.\}$ , it is shown in [5] that,

(3) 
$$N\{C\} = N\{L^*\} = (n_1, \dots, n_k)$$

where

$$\begin{pmatrix} n_1 - 1 \\ r - 1 \end{pmatrix} \begin{pmatrix} n_2 - 1 \\ r \end{pmatrix} & \cdots & \begin{pmatrix} n_k - 1 \\ r + k - 2 \end{pmatrix}$$

$$\begin{pmatrix} n_1 - 1 \\ r - 2 \end{pmatrix} \begin{pmatrix} n_2 - 1 \\ r - 1 \end{pmatrix} & \cdots & \begin{pmatrix} n_k - 1 \\ r + k - 3 \end{pmatrix}$$

$$\vdots & \vdots & \vdots & \vdots \\ \begin{pmatrix} n_1 - 1 \\ r - k \end{pmatrix} \begin{pmatrix} n_2 - 1 \\ r - k + 1 \end{pmatrix} & \cdots & \begin{pmatrix} n_k - 1 \\ r - 1 \end{pmatrix}$$

$$\vdots & \vdots & \vdots & \vdots \\ \begin{pmatrix} n_1 - 1 \\ r - k \end{pmatrix} & \begin{pmatrix} n_2 - 1 \\ r - k + 1 \end{pmatrix} & \cdots & \begin{pmatrix} n_k - 1 \\ r - 1 \end{pmatrix}$$

The determinant (4) plays an important role in this paper. It is also proved that  $(n_1, \ldots, n_k)_r$  satisfies the following:

In section 2, we define a partial order on C and establish an isomorphism between C and a set of compositions of  $M > n_1 - r + k$ , which is specified later. We also show that for r = 2, it leads to an interesting correspondence among two different sets of lattice paths and a set of lattice points. As a special case, the number of r-compositions of n that are [p]-dominated by a given r-composition of m has been evaluated (for definition see [4] section 4). Section 3 deals with the application of results in section 2, in order to provide a solution to a generalised class of the ballot problem [2, p. 66]. Finally, some identities which arise as a natural consequence of the above are included in the last section.

2. <u>Isomorphism between two sets of compositions</u>. Recalling the definition of a composition vector [6] (that is, defining

$$T_i*(n_j) = \sum_{\alpha=1}^i t_{\alpha}(n_j),$$

it is remarked that the set C is trivially in 1:1 correspondence with the set  $T* = T*(n_1, \ldots, n_k; r) = \{(T_1*(n_1), \ldots, T_r*(n_1)), \ldots, (T_1*(n_k), \ldots, T_r*(n_k)):\}$ 

(6) 
$$T_i^*(n_j) \ge T_i^*(n_{j+1})$$

for  $i=1,\ldots,r$  and  $j=1,\ldots,k-1$ . Because of this correspondence and somewhat relative advantage of  $T^*$  over C, we refer frequently to  $T^*$  instead of C.

 of T\*, we say that  $\tau_1^*$  dominates  $\tau_2^*$  if and only if

$$T_{i1}^{*(n_j)} \ge T_{i2}^{*(n_j)}$$
 for  $i = 1, ..., r$  and  $j = 1, ..., k$ .

It can be shown that this relation is a partial order defined on elements of T\*. Proceeding in a manner as in [6], we can also prove the following lemma.

LEMMA. The elements in T\* form a distributive lattice.

Next, it is easy to verify that the inequalities (6) for all i and j are satisfied if and only if

(7) 
$$T_{i+1, j} \ge T_{ij}$$
 for  $i = 1, ..., r-1$  and  $j = 1, ..., k$ ,

where  $T_{ij} = T_i * (n_j) - i - j + k + 1$  and  $T_{rj} = n_j - r - j + k + 1$ . Since  $T_{ik} < \dots < T_{i1}$  follows from (7), we now consider k+1 composition vectors  $(T_{ik}, \dots, T_{i1}, M)$  for  $i = 1, \dots, r$  where  $M > n_1 - r + k$  is a constant. Because of the inequalities in (7), we notice that  $M > T_{i1}$  for  $i = 1, \dots, r$ . Let  $T = T(n_1, \dots, n_k; r)$  be the set

$$\{(T_{rk}, \dots, T_{r1}, M), \dots, (T_{1k}, \dots, T_{11}, M):$$

$$T_{i+1,j} \ge T_{ij}$$
 for  $i = 1, ..., r-1$  and  $j = 1, ..., k$ .

In terms of compositions, it may be seen that T is the set such that (k+1)-composition  $(n_k-r+1,n_{k-1}-n_k+1,\ldots,n_1-n_2+1,M-n_1+r-k)$  of M dominates (k+1)-compositions of M, each of which again dominates (k+1)-compositions of M and so on. Using the simple transform  $a_{ij} = T_i * (n_j) - i$ , define the set  $S = S(n_1,\ldots,n_k;r)$  to be  $\{(a_{rk} = n_k-r,\ldots,a_{r1} = n_1-r),\ldots,(a_{1k},\ldots,a_{1l}):$ 

(8) 
$$a_{i+1 \ j} \ge a_{ij}$$
 for  $i = 1, ..., r-1 \text{ and } j = 1, ..., k$ 

where (a, ..., a) is a vector of non-negative, nondecreasing it

integers. The relation of domination on T\* through 1:1 transformations used above is extended to T and S and therefore we have the theorem.

THEOREM 1. Sets T\*, T and S are isomorphic distributive lattices.

We have shown in section 1 that  $T*(n_1, \ldots, n_k; 2)$  through  $C(n_1, \ldots, n_k; 2)$  is 1:1 to  $L*(n_1, \ldots, n_k; 2)$ . Also  $T*(n_1, \ldots, n_k; 2)$  is 1:1 to  $S(n_1, \ldots, n_k; 2)$  by the theorem. But  $S(n_1, \ldots, n_k; 2)$  is the set  $\{(a_k, \ldots, a_1)\}$  of all vectors of non-negative and nondecreasing integers such that

(9) 
$$0 \le a_{j} \le n_{j} - 2$$
 for  $j = 1, ..., k$ .

Now, using the construction of lattice paths from non-negative nondecreasing vectors [4, p.253], we notice that  $S(n_1, \ldots, n_k; 2)$  is 1:1 to the set  $L(n_1, \ldots, n_k)$  of lattice paths from (0,0) to  $(n_1-1,k)$  not crossing the boundary given by the points (0,0), (1,1),  $(n_1-n_2+1,2)$ ,  $(n_1-n_3+1,3)$ ,...,  $(n_1-n_k+1,k)$ , Here we have two remarks to offer:

- (a) The above lattice paths are equivalent to paths from (0,0) to  $(n_i,k)$  not touching the same boundary;
- (b) The set of paths are, in general, also equivalent to paths from (0,0) to  $(n_1+e-2,k)$  [or  $(n_1+e-1,k)$ ] not crossing [or not touching] the boundary (0,0), (e,1),  $(n_1-n_2+e,2)$ ,  $(n_1-n_3+e;3)$ ,..., $(n_1-n_k+e,k)$ , where e is a positive integer. From section 1,

(10) 
$$N\{L*(n_1,...,n_k; 2)\} = (n_1,...,n_k)_2$$

We observe from (4) and (5) that

The expression on the left hand side of (11) represents the number of lattice points in the region  $R(n_1, \dots, n_k)$  of k-dimensional Euclidean space bounded by hyperplanes

$$X_{k} = \frac{1}{2}$$
,  $X_{k} = n_{k} - \frac{1}{2}$ ,  $X_{k-1} = X_{k} - \frac{1}{2}$ ,  $X_{k-1} = n_{k-1} - \frac{1}{2}$ , .

 $X_{1} = X_{2} - \frac{1}{2}$ ,  $X_{4} = n_{1} - \frac{1}{2}$ .

Thus, as a corollary of Theorem 1, we see that:

COROLLARY 1. 
$$N\{L*(n_1,...,n_k; 2)\} = N\{L(n_1,...,n_k) = N\{R(n_1,...,n_k)\} = (n_1,...,n_k)_2$$
.

It is not difficult to observe that  $T(n_1, \dots, n_k; 2)$  represent the set of (k+1)-composition vectors  $(T_1, \dots, T_k, M)$  which are dominated by the (k+1)-composition vector

$$(n_{k-1}, n_{k-1}, \dots, n_{1}+k-2, M), M > n_{1}+k-2.$$

Thus according to [6], the number  $N\{T(n_1,\ldots,n_k;2)\}$  is give by  $D_k$  in the recursive formula

(12) 
$$\begin{cases} D_0 = 1 \\ u \\ D_u = \sum_{\alpha=1}^{\infty} (-1)^{\alpha+1} \begin{pmatrix} n_{k-u+\alpha} + u - 2 \\ 0 \end{pmatrix} D_{u-\alpha} .$$

We know from Theorem 1 that

$$N\{T(n_1,...,n_k; 2)\} = (n_1,...,n_k)_2$$
.

Therefore, we have

COROLLARY 2. A solution of  $D_u$  in (12) is  $\binom{n_{k-u+1}, n_{k-u+2}, \dots, n_k}{2}$ .

A direct proof is also possible. We indicate it here. Using induction,  $\,D_{_{U}}\,$  can be written as the determinant

Subtraction of the i<sup>th</sup> row from the (i-1)st row (i = 2,...,u), and repetition of this process reduces (13) to  $(n_{k-u+1}, n_{k-u+2}, ..., n_k)_2$ .

We now consider a problem, the solution of which is obtained with the help of Corollary 1. It is required to determine the number of r-compositions of n that are [s]-dominated by the r-composition  $(t_1(m), \ldots, t_r(m))$  of  $m \ (m \ge sn)$  [4, page 254]. The r-composition  $(t_1(m), \ldots, t_r(m))$  of  $m \ [s]$ -dominates an r-composition  $(t_1(n), \ldots, t_r(n))$  of n if and only if

$$T_{i}(m) \geq sT_{i}(n)$$

for i = 1,...,r. Inequalities (14) are equivalent to

$$\left[\frac{T_{i}(m)}{s}\right] \geq T_{i}(n)$$

for i = 1,...,r, where [z] is the greatest integer less than or equal to z. Thus we are interested in the set of r-compositions of n that are dominated by the r-composition

$$\left(\left\lceil \frac{T_1(m)}{s}\right\rceil, \left\lceil \frac{T_2(m)}{s}\right\rceil - \left\lceil \frac{T_1(m)}{s}\right\rceil\right), \dots, \left\lceil \frac{m}{s}\right\rceil - \left\lceil \frac{T_{r-1}(m)}{s}\right\rceil\right)$$

of  $\left[\frac{m}{s}\right]$ . Transforming the set to the set of non-negative and non-decreasing vectors, as done earlier, we observe that the above set is 1:1 with the set of vectors  $(a_1, \ldots, a_{r-1})$  such that

- (i) a 's are non-negative integers,
- (ii)  $a_1 \leq \ldots \leq a_{r-1}$ ,

(iii) 
$$0 \le a_i \le \min \left( \left[ \frac{T_i(m)}{s} \right] - i, n-r \right)$$
 for  $i = 1, \dots, r-1$ .

From the discussion following Theorem 1, we can get the number of such vectors, which is stated as a theorem.

THEOREM 2. The number of r-compositions of n that are [s]-dominated by the r-composition  $(t_1(m), \ldots, t_r(m))$  of m (m > sn) is

$$\min\left(\left[\frac{T_{r-1}(m)}{s}\right] - r + 1, n - r\right) + 2,$$

$$\min\left(\left[\frac{T_{r-2}(m)}{s}\right]-r+2, n-r\right)+2, \ldots,$$

$$\min\left(\left[\frac{T_1(m)}{s}\right]-1, n-r+2\right)_2.$$

3. Generalised ballot problems. The ballot problem [2, p. 66] and its extension have been discussed by several authors [1], [4], [7]. We state it as follows:

If in a ballot, Candidate A scores a votes and Candidate B scores b votes, where  $a > b\mu$ ,  $\mu$  being a positive integer, what is the probability that at each instant A's vote exceeds  $\mu$  times B's vote?

Representing each vote for A by a unit horizontal step and each vote for B by a unit vertical step, one of the solutions suggested in [4] uses the correspondence between lattice paths and non-negative non-decreasing vectors. In fact, the ballot problem with two candidates, in a generalised form, involves counting of lattice paths not touching a certain boundary which lies to the left of the paths. Recalling remarks (a) and (b), we have obtained the solution to such a problem in Section 2. In this context, we present below two theorems, the proof of which obviously follows from the preceeding results.

THEOREM 3. Let x and y respectively represent votes for A and B at a particular instant. Suppose that A scores a votes and B scores b votes such that  $a > b\mu + \nu$ ,  $\mu$  and  $\nu$  being non-negative numbers. The number of ways in which  $x > y\mu + \nu$  happens is given by

$$(a-[\mu+\nu]+1, a-[2\mu+\nu]+1,...,a-[b\mu+\nu]+1)_2$$
.

At this point we note that Takacs [7] gives a solution for general  $\mu$  and  $\nu=0$ . When  $\mu$  is a positive integer and  $\nu=0$ , the ballot problem reduces to the case stated at the beginning of this section. Therefore, the required number is  $(a-\mu+1,\ldots,a-b\mu+1)_2=(a,b,\mu)$  say. We have to show that

$$(a,b,\mu) = \frac{a-b\mu}{a+b} \binom{a+b}{b}.$$

For b = 1, the result is true for all a and  $\mu$ . Adding

each row to the previous row in the determinant  $(a,b,\mu)$ , we obtain

$$(a,b,\mu) = (a+1,b,\mu) - (a+1,b-1,\mu)$$
.

Hence

(12) 
$$(a+1,b,\mu) = (a+1,b-1,\mu) + (a,b,\mu)$$
  

$$= (a+1,b-1,\mu) + (a,b-1,\mu) + (a-1,b,\mu)$$

$$= (a+1,b-1,\mu) + (a,b-1,\mu) + \dots + (b\mu+1,b-1,\mu),$$

because  $(b\mu, b, \mu) = 0$ . Applying induction, we get from (12) that

$$(a+1,b,\mu) = \sum_{\alpha=b\mu+1}^{a+1} (\alpha,b-1,\mu) = \sum_{\alpha=b\mu+1}^{a+1} \frac{\alpha-(b-1)\mu}{\alpha+b-1} {\alpha+b-1 \choose b-1}$$

$$= \sum_{\alpha=b\mu+1}^{a+1} {\alpha+b-2 \choose b-1} - \mu \sum_{\alpha=b\mu+1}^{a+1} {\alpha+b-2 \choose b-2}$$

$$= {a+b \choose b} - \mu {a+b \choose b-1} = \frac{a-b\mu+1}{a+b+1} {a+b+1 \choose b} ,$$

and the result follows.

Another variation of the ballot problem is given below, and the result will be used in the next section.

THEOREM 4. For A and B having a and b votes respectively, where  $a > b_1 \mu_1 + \nu_1 + (b - b_1) \mu_2 + \nu_2$ ,  $\mu_1$ ,  $\nu_1$ ,  $\mu_2$  and  $\nu_2$  being non-negative numbers and  $b_1 \le b$  a non-negative integer, the number of ways in which  $x > y\mu_1 + \nu_1$  when  $0 \le y \le b_1$ , and  $x > b_1 \mu_1 + \nu_1 + (y - b_1)\mu_2 + \nu_2$  when  $b_1 \le y \le b$  can happen is

$$(a - [\mu_1 + \nu_1] + 1, \dots, a - [b_1\mu_1 + \nu_1] + 1,$$

$$a - [b_1\mu_1 + \nu_1 + \mu_2 + \nu_2] + 1, \dots,$$

$$a - [b_1\mu_1 + \nu_1 + (b-b_1)\mu_2 + \nu_2] + 1)_2.$$

The above theorems illustrate the use of the results developed in Section 2, in some simple boundary cases.

4. Some combinatorial identities. The two A.P. case of [4, p. 256-258] is a special case of Theorem 4, with  $\mu_1$ ,  $\nu_1$ ,  $\mu_2$  and  $\nu_2$  as non-negative integers. Using the same notation as in [4], we therefore get

(13) 
$$N\{A_{p,q}(a+1,b+1;c+1,d+1)\} = N_{p,q}(a+1,b+1;c+1,d+1)$$
  

$$= (a+(p-1)b+c+(q-1)d+2, a+(p-1)b+c+(q-2)d+2; ...,$$

$$a+(p-1)b+c+2, a+(p-1)b+2, a+(p-2)b+2, ..., a+2)_{2}$$

$$= \sum_{k=0}^{q} (-1)^{k} \frac{a+1}{a+1+(p+q-k)(b+1)} {a+1+(p+q-k)(b+1) \choose p+q-k} ...$$

$$\frac{(q-k+1)b-c-(q-k)d}{(q-k+1)b-c-qd} {(q-k+1)b-c-qd \choose k}$$

by Theorem 4 and Theorem 3 of [4]. Put b = 1, c = 1, d = 0. Then A (a+1,2,2,1) is 1:1 with the set of paths from (0,0) p,q

to (p+a,p+q) not touching the line x + q + 1 = y, and the number of such paths is equal to

$$\binom{2p+q+a}{p+q}$$
 -  $\binom{2p+q+a}{p-1}$ 

by [3]. Therefore, we have an identity

(14) 
$$(a+p+2,...,a+p+2,a+p+1,a+p,...,a+2)_{2}$$

$$= \sum_{k=0}^{q} (-1)^{k} \frac{a+1}{a+1+2(p+q-k)} {a+1+2(p+q-k) \choose p+q-k} {q-k \choose k}$$

$$= {2p+q+a \choose p+q} - {2p+q+a \choose p-1} .$$

Consider (p+a,p+q) as the origin, x=p+a, y=p+q as x-axis and y-axis respectively, such that the old origin becomes (p+q,p+a). Thus the previous set of paths is the same as the set of paths from (0,0) to (p+q,p+a) not touching x+a+1=y. The number in the latter set gives rise to the identity

(15) 
$$(p+q+2,...,p+q+2, p+q+1,...,q+2)_2$$

$$= \sum_{k=0}^{a} (-1)^k \frac{q+1}{q+1+2(p+a-k)} {q+1+2(p+a-k) \choose p+a-k} {a-k \choose k}$$

$$= {2p+q+a \choose p+a} - {2p+q+a \choose p-1}.$$

Either from the remark preceding (15) or from the obvious identity  $\binom{2p+q+a}{p+q} = \binom{2p+q+a}{p+a}$ , we see that (14) equals (15). We can show that

(16) 
$$\sum_{k=0}^{p+q} (-1)^k \frac{a+1}{a+1+2(p+q-k)} {a+1+2(p+q-k) \choose p+q-k} {q-k \choose k}$$

$$= \sum_{k=0}^{p+a} (-1)^k \frac{q+1}{q+1+2(p+a-k)} {q+1+2(p+a-k) \choose p+a-k} {a-k \choose k}$$

$$= {2p+q+a \choose p+q} = {2p+q+a \choose p+a}$$

by formula (17) in [9]. Therefore

(17) 
$$\sum_{k=q+1}^{p+q} (-1)^k \frac{a+1}{a+1+2(p+q-k)} {a+1+2(p+q-k) \choose p+q-k} {q-k \choose k}$$

$$= \sum_{k=a+1}^{p+a} (-1)^k \frac{q+1}{q+1+2(p+a-k)} {q+1+2(p+a-k) \choose p+a-k} {a-k \choose k}$$

$$= {2p+q+a \choose p-1}.$$

Perhaps some of the identities might have been proved or can be proved directly. A less obvious identity arises as follows. In the ballot problem stated in Theorem 3, set  $\nu=0$  and  $\mu=\frac{a}{b+1} \ \, \text{where a and b+1 are relatively prime numbers.}$  Then an application of the result of Theorem 2 of [8] yields

$$(a - [\frac{a}{b+1}] + 1, a - [\frac{2a}{b+1}] + 1, \dots, a - [\frac{ba}{b+1}] + 1)_2$$

$$= \frac{1}{a+b+1} {a+b+1 \choose a}.$$

In conclusion, we remark that the solution in the form of a determinant might not reduce to a simpler expression, except in some special cases.

## REFERENCES

- 1. A. Dvoretzky and Th. Motzkin, A Problem of Arrangements, Duke Mathematical Journal, XIV, (1947), p. 305.
- 2. W. Feller, An Introduction to Probability Theory and Its Applications, Second Edition, New York, John Wiley and Sons, Inc., 1957.
- 3. H. D. Grossman, Fun with Lattice Points, Scripta Mathematica, (1946), p. 224.

- 4. S. G. Mohanty and T. V. Narayana, Some Properties of Compositions and their Applications to Probability and Statistics I, Biometrische Zeitschrift, Band 3, Heft 4, (1961), p. 252.
- 5. T. V. Narayana, A combinatorial problem and its application to probability theory I, Journal of the Indian Society of Agricultural Statistics VII, (1955), p. 169.
- 6. T.V. Narayana and G.E. Fulton, A Note on the Compositions of an Integer, Canadian Mathematical Bulletin, Vol. I, (1958), p. 169.
- 7. L. Takács, A Generalisation of the Ballot Problem and Its Application in the Theory of Queues, Journal of American Statistical Association, Vol. 57, (1962), p. 327.
- 8. L. Takács, Ballot Problems, Z. Wahrscheinlichkeitstheorie, Vol. 1, (1962), p. 154.
- 9. H. W. Gould, Some Generalisations of Vandermonde's Convolution, American Mathematical Monthly, Vol. 63, (1956), p. 84.

State University of New York at Buffalo and McMaster University