A MATHEMATICAL MODEL FOR SINGLE FUNCTION GROUP ORGANIZATION THEORY WITH APPLICATIONS TO SOCIOMETRIC INVESTIGATIONS

Thesis for the Degree of Ph. D.

MICHIGAN STATE COLLEGE

James Henry Powell

1954

[redacted]

This is to certify that the

thesis entitled

A Mathematical Model for Single Function Group Organization Theory with Applications to Sociometric Investigations

presented by

Mr. James Henry Powell

has been accepted towards fulfillment
of the requirements for

Doctor of Philosophy degree in Mathematical Statistics

Leo Katz
Major professor

Date November 5, 1954

O-169

James Henry Powell

candidate for the degree of

Doctor of Philosophy

Final examination, November 5, 1954, 3:00 P.M., Physics-Mathematics Building

Dissertation: A Mathematical Model for Single Function Group Organization Theory with Applications to Sociometric Investigations

Outline of Studies

Major subject: Mathematical Statistics
Minor subjects: Algebra, Analysis, Geometry

Biographical Items

Born, May 21, 1926, Columbus, Ohio

Undergraduate Studies, Michigan State College, 1943-44, cont. 1946-49

Graduate Studies, Michigan State College, 1949-52, cont. 1953-54, University of California at Berkeley, 1952-53

Experience: Graduate Assistant, Michigan State College, 1949-52, Special Graduate Research Assistant, Michigan State College, 1952-54, Member United States Navy, 1944-46

Member of Phi Kappa Phi, Pi Mu Epsilon, Sigma Xi

A MATHEMATICAL MODEL FOR SINGLE FUNCTION GROUP ORGANIZATION THEORY WITH APPLICATIONS TO SOCIOMETRIC INVESTIGATIONS

By

JAMES HENRY POWELL

A THESIS

Submitted to the School of Graduate Studies of Michigan
State College of Agriculture and Applied Science
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1954

4-8-55

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Dr. Leo Katz, his major professor, for suggesting the problem and for his continuous help in every phase of this work.

The writer also deeply appreciates the financial support of the Office of Naval Research which made it possible for him to complete this investigation.

A MATHEMATICAL MODEL FOR SINGLE FUNCTION GROUP ORGANIZATION THEORY WITH APPLICATIONS TO SOCIOMETRIC INVESTIGATIONS

By

James Henry Powell

AN ABSTRACT

Submitted to the School of Graduate Studies of Michigan State College of Agriculture and Applied Science in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

Year 1954

Approved ______________

JAMES HENRY POWELL ABSTRACT

This thesis is concerned with the one-dimensional theory of group organization as a complex of irreflexive binary relationships, taking values zero and one, between the pairs of individuals. The basic problems in connection with this theory are

(i) the investigation of the appropriate universe or universes of discourse,

(ii) the determination of the null distributions for certain proposed indices of the group structure, and

(iii) the development of simple, reasonably exact methods for use by field investigators.

In the first part of this thesis, a decomposition of the total sample space is given which clarifies the first kind of problem. Also, certain bipartitional functions tabulated by David and Kendall[1] and standard combinatorial methods augmented by a theorem developed in this part provide the machinery for counting the number of distinct points in each of the subsets in the decomposition of the total sample space. This machinery makes it possible to obtain the necessary probabilities for construction of certain null distributions. These probabilities are obtained by dividing the number of points in the disjoint subsets in the framework of the given decomposition by the total number of points in the appropriate

---

[1]See reference [7] in Bibliography.

universe of discourse.

The second part of this thesis shows how the theory developed in the first part gives the null distributions for certain indices employed in group organization theory. In particular, the probability distributions of (a) indices on group expansiveness, (b) the number of isolates and (c) maximum s_j are given in detail as illustrative examples of the manner in which the general theory is applied to produce probability distributions.

In the third part, simple, approximate distributions are suggested for certain of the variables treated in the second part of this thesis. In addition, a test criterion for one aspect of group structure is proposed along with a simple, approximate distribution for it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION[1]

0:1. General model for group organization.

An organization of a social group is a complex made up of many interactions among all the pairs of individuals comprising the group. One way of describing the organization of the group is by specifying the relationships between each pair of individuals. The relationship between a pair of individuals can be represented as a many-dimensional vector, in which each component is the strength of the relationship for a particular activity (category), measurable on some scale. The hope for simplicity and economy in a scientific description of the group rests on the possibility that most of the information is contained in relatively few facets of the group organization; perhaps, only one.

In particular, the simplest model of a group organization is a one-dimensional model with a binary relation between pairs of individuals which is not necessarily reflexive and takes only the values zero and one in each direction. Although the model is superficially simple, it is by no means empty, for it is the exact model used in sociometric investigations. The journal Sociometry, which has been in publication for some seventeen years, has contained papers by sociometrists and others on the investigation of group organization

---

[1]This work was sponsored by the Office of Naval Research.

using this simple model. We shall see shortly that this same model is also useful for representing communication networks and in other near-sociometric investigations.

0:2. Sociometric investigations.

Moreno [28], in 1934, invented, as a technique for the study of interpersonal relationships, the well-known sociometric test, which is still the basic tool used today. In the words of Moreno [28,p. 13], "The sociometric test consists in an individual choosing his associates for any group of which he is or might become a member." The test is applied to a well-defined social group (hereafter the word group shall mean social group) as follows:

> Each member of a group of n individuals is asked to choose a number, specified or not, of the n-1 others with whom he would prefer to be associated in a particular single activity. In addition, each member may be asked to name those with whom he does not wish to associate, and/or he may make selections, separately, for more than one activity.

We shall only be concerned with the test as applied to groups for the purpose of eliciting positive responses for one activity. Typical questions asked for eliciting positive choices are:

> With whom do you wish to sit?
>
> With whom do you wish to work?

There are two common forms employed for exhibiting the collected data. The first, the directed graph on n points (sociogram), has been explored extensively by many sociometrists, most of whose publications appear in the journal Sociometry. The sociogram is essentially descriptive in character. The second form, proposed by Forsyth and Katz [13], is that of a matrix of choices, $C = (c_{ij}): n \times n$, where c_{ij} is a representation of the response (choice) from individual i to individual j. The simplest representation is $c_{ij} = 1$ or 0 according as i does or does not choose j. The principal diagonal elements of C, c_{ii}, are usually taken to be zero. If rejection is also present, c_{ij} may equal -1 whenever individual i rejects individual j. If graduated choices are permitted, the possible values of c_{ij} are further complicated.

One of the problems in group organization theory is the construction of summary indices which attempt to measure various aspects of the organization of the whole group. In connection with this problem, it is essential to study, under some suitable null distribution, the probability distributions of the indices used. We shall investigate, under the simplest mathematical model for group organization theory, the appropriate sample spaces to which should be referred these probability distribution problems. This model corresponds exactly to the experimental technique proposed by Moreno in his sociometric test and will be defined in more detail in the preliminaries of Part I.

0:3. Review of work on related problems.

Luce and Perry [25], in 1949, used the same matrix formulation as proposed by Forsyth and Katz but restricted the entries to only zeros and ones. Luce and Perry were interested in more complex configurations. In particular, they were interested in two concepts, k-chain and clique. A k-chain from individual i_1 to individual i_{k+1} is an ordered sequence with k+1 members i_1, i_2, i_3, ..., i_{k-1}, i_k, i_{k+1} such that i_1 chooses i_2, i_2 chooses i_3, ..., i_{k-1} chooses i_k, and i_k chooses i_{k+1}. A clique is a subset (proper or not) of the group such that each individual in the subset chooses all the others in the subset. There is a fairly extensive literature on the problems involving more complex configurations. However, we will not be concerned with these problems.

A communication network can be represented by a matrix of zeros and ones. Shimbel [35] uses such a matrix representation with unit entries on the principal diagonal. He gives, correctly, the number of communication networks involving n points (individuals) as $2^{n(n-1)}$. We shall see later that this agrees with our expression for the number of directed graphs on n points. Primarily, he is concerned with powers of the matrix in order to answer questions on the connectivity of the network. Christie, Luce and Macy [3] have done a good deal of experimental work on communication problems in very small groups. A rather extensive bibliography on communication networks and its related problems appears at the end of their report.

The structure of animal societies has been studied by Rapoport [32, 33, 34] and Landau [21, 22, 23]. The problem of "peck order" or "peck right" in a flock of birds is represented by a matrix of ones and zeros. However, the "peck right" relation is dominant, i.e., if animal i pecks animal j then j cannot peck i. Thus, their results are of no use to us since we do not want to rule out the possibility of two individuals choosing each other.

Thrall [40] considered the ranking of social organizations on the basis of their common members. We will not be concerned with this type of investigation.

A monograph on Graph Theory as a Mathematical Model in Social Science, by Norman and Harary [15], gives a review of the previous work done on numbers of graphs. Pólya [31] gave explicitly the numbers of trees[2] and rooted trees[3], and, implicitly, the number of ordinary graphs on n points. Otter [30] gave simpler, purely combinatorial methods for counting trees and rooted trees and Harary and Uhlenbeck [16] gave the numbers of Husimi trees[4]. Also, Harary and Norman [14] have given the number of directed graphs on n points and t lines. Davis [8] defines and gives a method for counting the numbers of various subclasses of directed graphs on n points. The problem we shall face is essentially different from all of these and requires quite different methods of enumeration.

[2] A tree is an ordinary graph with a path between every pair of its points, i.e., a connected graph, and which has no cycles.

[3] A rooted tree is a tree in which there is a distinguished point.

[4] A Husimi tree is a connected graph in which no line lies on more than one cycle.

0:4. Previous work on the present problem.

The application of chance (probability) distributions to sociometric data was, first, suggested by Moreno and Jennings [29], in 1938. They took deviations from chance as a reference base for the measurements of social configurations. In their words, [29,p 9], "It appeared that the most logical ground for establishing such a reference could be secured by ascertaining the characteristics of typical configurations produced by chance balloting for a similar size population with a like number of choices." The theoretical computation of the desired probabilities was turned over to Lazarsfeld. In Part II on applications we shall give the specific history of attempts to obtain the distributions of the several sociometric variables whose exact distributions are obtained.

Bronfenbrenner [1 ,p. 9], in 1943, attempted "to develop an absolute criterion — a frame of reference against which the various phenomena of sociometric choice may be projected but which is itself independent of these phenomena." Thus, what Bronfenbrenner had in mind was a common universe of discourse for all the sociometric variables. This is the so-called "constant frame of reference" problem in sociometry. Bronfenbrenner elaborated on the general technique, proposed by Moreno and Jennings, of taking deviations from chance as a reference base. He developed some expressions and techniques for determining the probability of occurrence of the major sociometric variables. However, for the most part, the results obtained were incorrect.

Edwards [10], in 1948, surveyed the published work on the application of deviations from chance to the problems in sociometrics. She suggested some modifications and pointed out certain errors in Bronfenbrenner's work. However, Edwards did not question whether it was correct to assume that all sociometric variables should have the same universe of discourse.

In 1950, Criswell [6] pointed out that indices in different experimental settings do have different meanings, and that it was ridiculous to believe that all the variables considered in sociometric investigations would have the same frame of reference. Criswell [6, p. 107], further, pointed out the need for the "use of the appropriate chance distribution as a reference base for developing a score expressing a structural aspect of the group as a whole".

We note, finally, that all of these investigations considered probability distribution problems only for the case where all of the individuals make the same number of choices. In Part I, general methods are obtained which give the distributions of certain sociometric variables for the less restrictive case in which individuals are free to make any number of choices.

0:5. Statement of problem and plan of attack.

The basic problems of the one-dimensional theory of group organization as a complex of irreflexive binary relationships between the pairs of individuals are

(i) the investigation of the appropriate universe of discourse,

(ii) the determination of the null distribution for each proposed index of the group structure, and

(iii) the development of simple, reasonably exact methods for use by field investigators.

In Part I, we consider the simplest mathematical model for group organization and develop a decomposition of the total sample space which simplifies the first kind of problem. Also, in Part I, we develop the machinery of counting methods which makes it possible to obtain many of the null distributions exactly. In Part II, applications are given to classes of unsolved probability distribution problems of group organization theory. Finally, in Part III, simple approximate distributions are suggested for certain of the variables treated in Part II.

PART I

THEORY

1:1. Preliminaries.

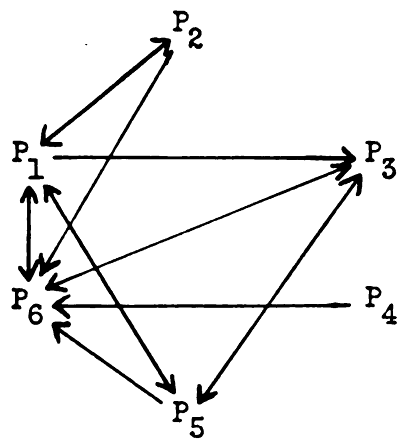
We shall be concerned with finite groups of n individuals. The organization of such a group for a single activity can be thought of as a configuration of the connections among all pairs of individuals. The connections between pairs are not necessarily reflexive and in the simplest case take in each direction only two values, 0 and 1. Such a configuration can be represented by a matrix C or by a linear directed graph G on n points (see Figure 1). C is the n x n matrix (c_{ij}) with $c_{ij} = 1$ if a connection exists from individual i to individual j, otherwise $c_{ij} = 0$. We adopt the usual convention that $c_{ii} = 0$ for all i. Let the points of the graph G be $P_1, P_2, \ldots, P_n$. Then, in the graph, a connection existing from individual i to individual j is represented by a directed line from P_i to P_j, $P_i \longrightarrow P_j$. The absence of a connection from individual i to individual j corresponds to no line from P_i to P_j. Obviously, $c_{ij} = 1$ (or 0) if and only if a directed line exists (or doesn't exist) from P_i to P_j. Hence, the graphical representation is isomorphic with the matrix representation.

We let $r_i = \sum_j c_{ij}$ be the ith row total of C and $s_j = \sum_i c_{ij}$ be the jth column total of C. Of course, r_i is exactly the number of lines issuing from the point P_i, and s_j is the number of lines terminating on the point P_j. Moreover, $\sum_i r_i = \sum_j s_j = t$, the total number of

directed lines. Finally, let the vectors $\underline{r}$ and $\underline{s}$, with elements r_i and s_i, respectively, be the two n-part, non-negative, ordered partitions of $\underline{t}$ which represent, respectively, the marginal row and column totals of C.

FIGURE 1

Two representations of organizational configurations on 6 points



(a)

graphical representation

	1	2	3	4	5	6
1	0	1	1	0	1	1
2	1	0	0	0	0	1
3	0	0	0	0	1	1
4	0	0	0	0	0	1
5	1	0	1	0	0	1
6	1	0	1	0	0	0

(b)

matrix representation

Unless otherwise noted, all graphs will be on $\underline{n}$ points and linearly directed (n-graphs), and all matrices will be n x n with 0's on the principal diagonal and composed of 1's and 0's elsewhere.

1:2. A decomposition of the space of all n-graphs (or n x n matrices).

We consider the decomposition of the space of all n-graphs in a way which seems best adapted to the problems which arise in the investigations in group organization theory.

Let us introduce the following notations:

Ω : Sample space, the space of all n-graphs (or n x n matrices of the type defined).

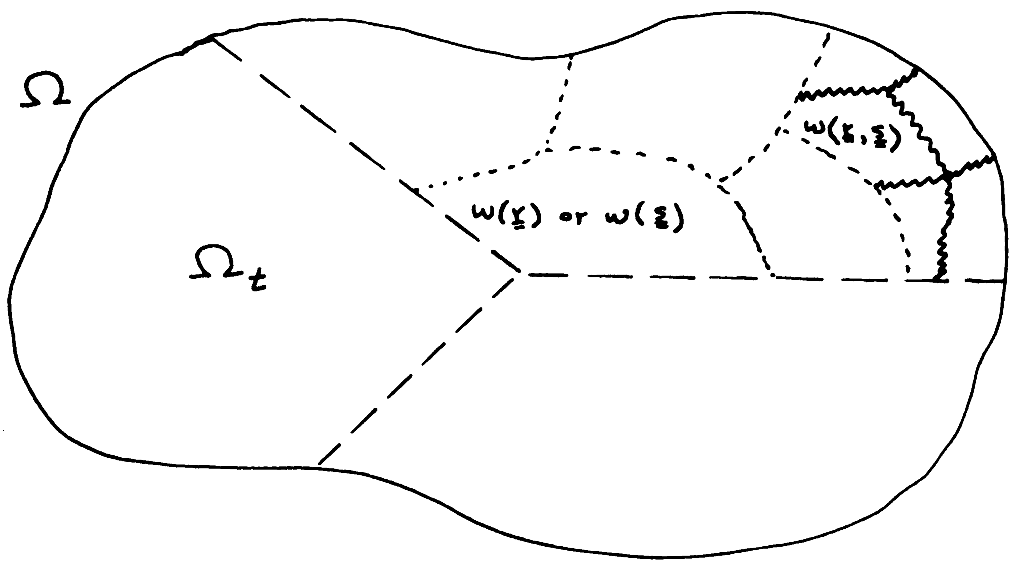
Ω_t : First-order subspace, the space of all n-graphs with fixed number of lines, t, $(t = 0, 1, 2, \ldots, n(n-1))$.

$\omega(\underline{r})$: Second-order subspace, the space of all n-graphs with fixed $\underline{r}$ (t is necessarily fixed).

$\omega(\underline{s})$: Second-order subspace, dual to $\omega(\underline{r})$.

$\omega(\underline{r}, \underline{s})$: Third-order subspace, the space of all n-graphs with fixed $\underline{r}$ and $\underline{s}$ (t is necessarily fixed).

A diagram of the relationships among these spaces would appear as follows:



We have a decomposition of Ω into subspaces of successively higher order. First, Ω is partitioned disjointly and exhaustively into a sum of first-order subspaces, Ω_t, according to

1:2.1 $$\Omega = \sum_{t=0}^{n(n-1)} \Omega_t \, .$$

Next, there is a decomposition of Ω_t by a disjunctive partitioning into a sum of second-order subspaces, $\omega(\mathbf{r})$ or $\omega(\mathbf{s})$, according to

1:2.2 $$\Omega_t = \sum_{(\mathbf{r})} \omega(\mathbf{r}) = \sum_{(\mathbf{s})} \omega(\mathbf{s}) \, ,$$

where the summation is over all n-part vectors $\mathbf{r}$ (or $\mathbf{s}$) with elements r_i (or s_j) subject to $0 \leq r_i \leq n-1$ (or $0 \leq s_j \leq n-1$) and $\sum_{i=1}^{n} r_i = \sum_{j=1}^{n} s_j = t$. Finally, we have a partitioning of $\omega(\mathbf{r})$ and $\omega(\mathbf{s})$ into sums of disjoint third-order subspaces $\omega(\mathbf{r}, \mathbf{s})$, according to

1:2.3(a) $$\omega(\mathbf{r}) = \sum_{(\mathbf{s})} \omega(\mathbf{r}, \mathbf{s}), \quad \text{and}$$

1:2.3(b) $$\omega(\mathbf{s}) = \sum_{(\mathbf{r})} \omega(\mathbf{r}, \mathbf{s}),$$

where the summation is the same as for 1:2.2. We summarize the foregoing in Table I.

TABLE I

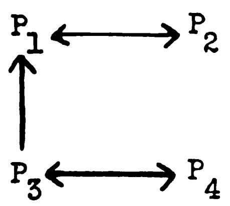
A DECOMPOSITION OF THE SPACE OF ALL n-GRAPHS

Space	Decomposition
Ω	$\Omega = \sum_t \Omega_t$
Ω_t	$\Omega_t = \sum_{(\underline{r})} \omega(\underline{r}) = \sum_{(\underline{s})} \omega(\underline{s})$
$\omega(\underline{r})$	$\omega(\underline{r}) = \sum_{(\underline{s})} \omega(\underline{r}, \underline{s})$
$\omega(\underline{s})$	$\omega(\underline{s}) = \sum_{(\underline{r})} \omega(\underline{r}, \underline{s})$
$\omega(\underline{r}, \underline{s})$	————————

Double and triple disjoint decompositions are also indicated. For example,

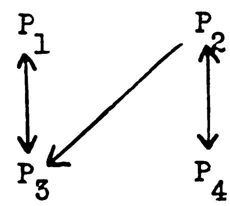
$$\Omega = \sum_t \sum_{(\underline{r})} \sum_{(\underline{s})} \omega(\underline{r}, \underline{s}) .$$

We conclude with a lemma and some remarks on the equivalence under permutation of certain subspaces. A matrix C_λ is equivalent under permutation to the matrix C_μ if and only if there exists a permutation matrix P such that $C_\lambda = PC_\mu P'$. Then, the graphs G_λ and G_μ, corresponding to the matrices C_λ and C_μ, are equivalent under alias, or renaming of the points, in the sense that



G_λ

is equivalent to



G_μ ,

when P_1 in G_λ is renamed P_3, P_2 in G_λ is renamed P_1, and P_3 in G_λ is renamed P_2. In the matrix representation,

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{C_\lambda} = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{C_\mu} \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{P'} .$$

We now proceed to the lemma showing the equivalence under permutation of the subspaces $\omega(\mathbf{r}, \mathbf{s})$ and $\omega(P\mathbf{r}, P\mathbf{s})$.

LEMMA 1.1. *If the vectors* $\mathbf{r}$ *and* $\mathbf{s}$ *are any two n-part, non-negative, ordered partitions of* t, *then the transformation* PCP' *is a one-to-one mapping of* $\omega(\mathbf{r}, \mathbf{s})$ *on* $\omega(P\mathbf{r}, P\mathbf{s})$.

PROOF: By definition $\omega(\mathbf{r}, \mathbf{s})$ contains all distinct matrices having $\mathbf{r}$ and $\mathbf{s}$ as marginal totals. Let $C_k \in \omega(\mathbf{r}, \mathbf{s})$, $k = 1, 2, \ldots, n(\mathbf{r}, \mathbf{s})$, and consider $PC_kP' = B_k$. Multiplication on the left of C_k by P permutes the rows of C_k and multiplication on the right of C_k by P' permutes the corresponding columns of C_k. Thus, multiplying C_k on the left by P and on the right by P' simultaneously permutes corresponding rows and columns. The effect on the marginal vectors $\underline{r}$ and $\underline{s}$ of C_k of the pre-multiplication by P and the post-multiplication by P' is that the vectors $\underline{r}$ and $\underline{s}$ are simultaneously permuted. Hence, B_k has row totals $P\underline{r}$ and column totals $P\underline{s}$, and

(1) $$B_k \varepsilon \omega(P\underline{r}, P\underline{s}).$$

Next, $C_k \neq C_h$ implies that $c_{ij}^{(k)} \neq c_{ij}^{(h)}$ for at least one pair (i, j). Now, the transformation PCP' = B takes $c_{ij}^{(k)}$ into $b_{i_1j_1}^{(k)}$ and $c_{ij}^{(h)}$ into $b_{i_1j_1}^{(h)}$. Thus, $b_{i_1j_1}^{(k)} \neq b_{i_1j_1}^{(h)}$ for at least one pair (i_1, j_1). Hence,

(2) $$C_k \neq C_h \Longrightarrow B_k \neq B_h ,$$

so that there are at least as many B's as C's.

Finally, similar arguments show that

(3) $$B_h \varepsilon \omega(P\underline{r}, P\underline{s}) \Longrightarrow C_h = P'B_hP ,$$

and $C_h \varepsilon \omega(\underline{r}, \underline{s})$ by definition. Combining (1), (2) and (3) completes the proof.

We give, next, a useful corollary. Consider I, any subset of $\{i\} = \{1, 2, \dots, n\}$, and $P^{(I)}$, any permutation matrix affecting only the components r_i and/or s_i of $(\underline{r}, \underline{s})$ which have suffixes in I. Thus,

if $r_i = d$ for $i \varepsilon I$, we have that $\underline{r}$ is identical to $P^{(I)}\underline{r}$. This gives us

COROLLARY 1.1. *Let $\underline{r}$ and $\underline{s}$ be the same as in the preceding lemma. Then, if $r_i = d$ for $i \varepsilon I$, the transformation $P^{(I)}CP^{(I)'}$ is a one-to-one mapping of $\omega(\underline{r}, \underline{s})$ on $\omega(\underline{r}, P^{(I)}\underline{s})$.*

We remark that this corollary together with standard methods of enumerating permutations will permit the calculations to be materially abridged in certain cases.

1:3. Enumeration of n-graphs (or n x n matrices).

For the enumeration of points in the various spaces we will find it more convenient to use the matrix formulation. Thus, the number of matrices (graphs) in Ω is the number of ways in which $n(n-1)$ positions may be specified as either zero or one. By elementary considerations, the number of distinct ways this can be done is

1:3.1 $$\eta = 2^{n(n-1)} .$$

For matrices in Ω_t we have $\underline{t}$ ones to distribute over $n(n-1)$ places and the number of ways this can be accomplished is the number of ways of specifying a particular $\underline{t}$ of the $n(n-1)$. Therefore, the number of matrices (graphs) in Ω_t is given by

1:3.2 $$\eta_t = \binom{n(n-1)}{t} ,$$

where $\binom{a}{b}$, $b \leq a$, is the binomial coefficient $a!/[b!(a-b)!]$. As is well-known, $\sum_t \binom{n(n-1)}{t} = 2^{n(n-1)}$.

In the enumeration of matrices in $\omega(\underline{r})$ we have, for each i, r_i ones to distribute over n-1 places. This can be done independently, for each i, in $\binom{n-1}{r_i}$ ways and thus the number of matrices (graphs) in $\omega(\underline{r})$ is given by

1:3.3 $$\eta(\underline{r}) = \prod_1^n \binom{n-1}{r_i} .$$

By a similar argument, the number of matrices in $\omega(\underline{s})$ is given by

1:3.4 $$\eta(\underline{s}) = \prod_1^n \binom{n-1}{s_j} .$$

The number of distinct matrices (graphs) in $\omega(\underline{r}, \underline{s})$, $\eta(\underline{r}, \underline{s})$, cannot be handled by any well-known counting methods. In the following section, we establish a theorem giving this number. This theorem will complete the enumerations needed for the sample space decomposition of the previous section.

1:4. The number of matrices (graphs) in $\omega(\underline{r}, \underline{s})$.

We first consider a relevant bipartitional function. P.V. Sukhatme [38] considered, among others, the problem of finding the number, $A(\underline{s}, \underline{r})$, of possible ways in which the cells of a ρ x σ

matrix can be filled with zeros and ones so that

i) the column totals from left to right form the parts of the partition $\underline{s}$ in some fixed order, and

ii) the row totals from top to bottom form the parts of the partition $\underline{r}$ in some fixed order.

He pointed out that $A(\underline{s}, \underline{r}) = A(\underline{r}, \underline{s})$. It is also evident that this number is not altered by addition of a sufficient number of rows or columns of zeros to make the matrix square. We may, therefore, take $n = \max(\rho, \sigma)$; then Sukhatme's matrices include those corresponding to our matrices in $\omega(\underline{r}, \underline{s})$ as well as those which violate the condition that only zeros appear on the principal diagonal. This led to the idea of deleting from $A(\underline{r}, \underline{s})$ the number of matrices having one or more 1's on the principal diagonal. The main theorem of this section expresses $\eta(\underline{r}, \underline{s})$ as a linear combination of the known $A(\underline{r}, \underline{s})$. This was first accomplished by a process of alternate inclusion and exclusion. However, the theorem and proof as given here is a considerable simplification of the original argument.

The definitions of $\eta(\underline{r}, \underline{s})$ and $A(\underline{r}, \underline{s})$ are completed with the conventions that both vanish identically if any component of the partitions is negative and that both are linearly additive over the partitions, e.g., $\eta[(\underline{r}, \underline{s}) + (\underline{r}', \underline{s}')] = \eta(\underline{r}, \underline{s}) + \eta(\underline{r}', \underline{s}')$.

Sukhatme [38] in 1938 gave tables of A(P,Q) for P,Q partitions of $\underline{t}$ for t = 1, 2, ..., 8. These are identical, for weights t = 1, 2, ..., 6, with tables given by MacMahon [26] in 1915. MacMahon was interested, among other things, in a different formulation of the same problem which Sukhatme later considered. He showed that the required solution is obtained from the coefficients in the expansion of products of elementary (unitary) symmetric functions in terms of monomial symmetric functions. The unitary symmetrics are defined by

$$a_k = \sum_{i<j<\ldots<k} x_i x_j \ldots x_k$$

whereas the monomial symmetrics are typified by

$$(p_1^{\pi_1} p_2^{\pi_2} \cdots p_s^{\pi_s}) = \sum (x_i^{p_1} x_j^{p_1} \cdots x_h^{p_2} x_k^{p_2} \cdots x_u^{p_s} x_v^{p_s})$$

where the summation takes place over all suffixes i, j, ..., v which are different. Recently, David and Kendall [7] recomputed and extended these tables through $t = 12$. Their tables were collated with those published by Cayley [2] in 1857 for weights $t = 1, 2, \ldots, 10$ and by Durfee [9] in 1882 for $t = 12$.

Sukhatme's paper [38, p. 387] gave, also, an algorithm for computation of a single $A(P,Q)$ of any weight and David and Kendall gave a systematic method for partially extending their tables to those of successive higher weights. Thus, implicitly, the problem of evaluating the number of graphs in $\omega(\underline{r}, \underline{s})$ is completely solved when we express $\eta(\underline{r}, \underline{s})$ in terms of $A(\underline{r}, \underline{s})$.

We shall require operators δ_i^j $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n)$ defined by

1:4.1 $$\delta_i^j(r_1, \ldots, r_i, \ldots, r_n; s_1, \ldots, s_j, \ldots, s_n)$$

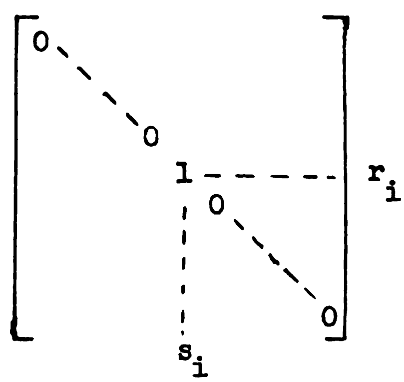
$$= (r_1, \ldots, r_i - 1, \ldots, r_n; s_1, \ldots, s_j - 1, \ldots, s_n).$$

The effect of the operation δ_i^j is to replace the double partition $(\underline{r}, \underline{s})$ of $\underline{t}$ by a double partition of $(t-1)$ with the ith part of $\underline{r}$ and the jth part of $\underline{s}$ each reduced by unity. If we take $(\delta_i^j + \delta_k^h)(\underline{r}, \underline{s}) = \delta_i^j(\underline{r}, \underline{s}) + \delta_k^h(\underline{r}, \underline{s})$, the obvious commutativity of the δ's serves to define every sum of monomials of the form $\sum_{(\alpha)} [\prod_{j,k} (\delta_j^k)^{\alpha_i}]$ as an operator on $(\underline{r}, \underline{s})$. We note that the number of non-trivial terms in the application of such an operator to $(\underline{r}, \underline{s})$ is necessarily finite for finite $\underline{t}$ since both $A[(\delta_i^j)^{m+k}(\underline{r}, \underline{s})] = 0$ and $\eta[(\delta_i^j)^{m+k}(\underline{r}, \underline{s})] = 0$ for $k > 0$ and $m = \min(r_i, s_j)$. Finally, we observe that any identity in the $A(\underline{r}', \underline{s}')$ and $\eta(\underline{r}'', \underline{s}'')$ is unaffected by application of these operators to the partitions involved. Any operation on an A or η is to be interpreted as an operation on partitions. Thus, the operator filters through A or η.

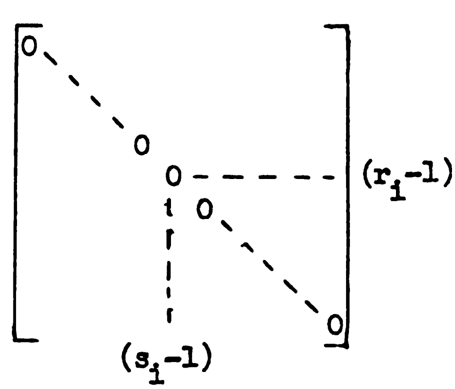
LEMMA 1.2. <u>If the vectors $\underline{r}$ and $\underline{s}$ are any two n-part, non-negative, ordered partitions of $\underline{t}$ then</u>

1:4.2 $$A(\underline{r}, \underline{s}) = \eta\left\{\prod_1^n (1 + \delta_i^i)(\underline{r}, \underline{s})\right\}.$$

PROOF: Every matrix belonging to the set enumerated by $A(\mathbf{r},\mathbf{s})$ has either no principal diagonal elements different from zero, or one specific principal diagonal element different from zero, or two specific such elements, etc. This exhaustive disjunction of the set gives subsets each isomorphic with a set enumerated by an $\eta(\mathbf{r}_\alpha, \mathbf{s}_\alpha)$ where the vectors $\mathbf{r}_\alpha$ and $\mathbf{s}_\alpha$ are formed from $\mathbf{r}$ and $\mathbf{s}$ by reduction by unity of each component corresponding to a specified principal diagonal element different from zero. Thus, the matrices of the type



are in one-to-one correspondence with matrices of type



.

Then,

$$A(\underline{r},\underline{s}) = \eta(\underline{r},\underline{s}) + \sum_i \eta[\delta_i^i(\underline{r},\underline{s})] + \sum_{i<j} \eta[\delta_i^i \delta_j^j(\underline{r},\underline{s})]$$

$$+ \cdots + \eta[\delta_1^1 \delta_2^2 \cdots \delta_n^n(\underline{r},\underline{s})] ,$$

and, rewriting the right-hand member,

$$A(\underline{r},\underline{s}) = \eta\left\{\prod_1^n (1 + \delta_i^i)(\underline{r},\underline{s})\right\} .$$

We next establish inverses for certain operators, as in

LEMMA 1.3. <u>The operator</u> $(1 + \delta_i^j)$ <u>has an inverse</u> (<u>right and left</u>) <u>given by</u>

1:4.3 $$(1 + \delta_i^j)^{-1} = 1 - \delta_i^j + (\delta_i^j)^2 - (\delta_i^j)^3 + \cdots .$$

PROOF: $(1 + \delta_i^j)^{-1}(1 + \delta_i^j)(\underline{r},\underline{s}) = (1 + \delta_i^j)^{-1}[(\underline{r},\underline{s}) + (\underline{r} - \underline{u}_i, \underline{s} - \underline{u}_j)]$, where $\underline{u}_i$ and $\underline{u}_j$ are vectors with unit $\underline{i}$th and $\underline{j}$th components, respectively, and all others zero. Next,

$$(1 + \delta_i^j)^{-1}[(\underline{r},\underline{s}) + (\underline{r} - \underline{u}_i, \underline{s} - \underline{u}_j)]$$

$$= (\underline{r},\underline{s}) - (\underline{r}-\underline{u}_i, \underline{s}-\underline{u}_j) + (\underline{r}-2\underline{u}_i, \underline{s}-2\underline{u}_j) - \cdots$$

$$+ (\underline{r}-\underline{u}_i, \underline{s}-\underline{u}_j) - (r-2\underline{u}_i, \underline{s}-2\underline{u}_j) + \cdots$$

$$= (\underline{r},\underline{s}) ,$$

and $(1+\delta_i^j)^{-1}$ is a left inverse. A similar proof shows that $(1+\delta_i^j)^{-1}$ is a right inverse.

Repeated application of 1:4.3 to both members of 1:4.2 gives immediately the main theorem.

THEOREM 1.1. *If the vectors $\underline{r}$ and $\underline{s}$ are any two n-part, non-negative, ordered partitions of t then the number of distinct matrices in $\omega(\underline{r}, \underline{s})$ having these vectors as marginal totals is given by*

1:4.4 $$\eta(\underline{r}, \underline{s}) = A\left\{\prod_{1}^{n}(1+\delta_i^i)^{-1}(\underline{r}, \underline{s})\right\} .$$

As a remark in application of Theorem 1.1, and to a lesser extent in application of Lemma 1.2, we note that the series in 1:4.3 with i = j may be terminated at the $\underline{m}_i$th power, for each i, where $\underline{m}_i = \min(r_i, s_i)$, since the functions $A(\underline{r}_\alpha, \underline{s}_\alpha)$ evaluated beyond this point all vanish. Thus, in every case, $\eta(\underline{r}, \underline{s})$ is evaluated by additions and subtractions of finitely many $A(\underline{r}_\alpha, \underline{s}_\alpha)$.

Some examples may be given to illustrate the application of Theorem 1.1 and to show the order of magnitude of the numbers, $\eta(\underline{r}, \underline{s})$. Suppose we wish to compute the number of directed graphs on four points where the numbers of lines in and out of the points are, respectively, (2,3), (1,2), (2,1) and (2,1); i.e., we wish to compute $\eta(2,1,2,2;3,2,1,1)$. The remark following the theorem gives us

$\eta(2,1,2,2;3,2,1,1)$

$$= A\left\{[1 - \delta_1^1 + (\delta_1^1)^2][1 - \delta_2^2][1 - \delta_3^3][1 - \delta_4^4](2,1,2,2;3,2,1,1)\right\}.$$

Operating as indicated on the right-hand member of the equation above, we have

$$\eta(2,1,2,2;3,2,1,1) = A(2^3,1;3,2,1^2)$$
$$-A(2^2,1^2;2^2,1^2) - A(2^3;3,1^3) - 2A(2^2,1^2;3,2,1)$$
$$+ 2A(2^2,1;2,1^3) + 2A(2,1^3;2^2,1) + 2A(2^2,1;3,1^2) + A(2,1^3;3,2)$$
$$- A(2,1^2;3,1) - A(1^4;2^2) - 2A(2,1^2;2,1^2) - A(2^2;1^4) - 2A(2,1^2;2,1^2)$$
$$+ A(1^3;2,1) + A(1^3;2,1) + 2A(2,1;1^3)$$
$$- A(1^2;1^2),^{*}$$

where the terms on the right are, by lines, of weight 7, 6, ..., 2, respectively. Note that the parts of the partitions in $\eta(\underline{r}, \underline{s})$ are pairwise ordered, as also they are (necessarily) in the right-hand member of the equation preceding the one above. However, in the evaluation of a single $A(\underline{r}_\alpha, \underline{s}_\alpha)$, this is not necessary and some combining of terms occurs. Sukhatme's tables for t = 7, 6, ..., 2

---

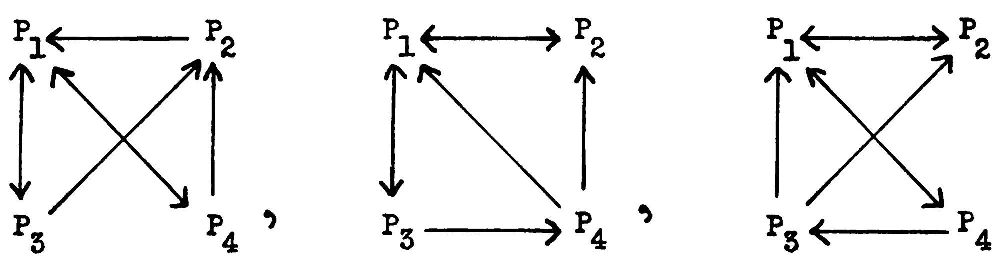
* $(n_1^{\pi_1}, n_2^{\pi_2}, \ldots)$ is conventional partition notation designating π_1 n_1's, π_2 n_2's, etc., in the partition.

give the required numbers. The sum of the 12 positive terms is 94, the 12 negative terms total 91 and $\eta(2,1,2,2;3,2,1,1) = 3$. The three graphs, in matrix representation and graph form are given in Figure 2, below.

FIGURE 2

Representations of points in $\omega(2,1,2,2,;3,2,1,1)$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \end{array}, \quad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \end{array}, \quad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$



Obviously, in this case, the three graphs or matrices could have been exhibited directly and the method seems cumbersome. By way of contrast, 16 positive and 16 negative terms give $\eta(1,1,1,1,1,1,1,1;3,2,1,1,1,0,0,0) = 4846 - 3703 = 1143$; it is no longer feasible to exhibit the distinct directed graphs on these eight points:

Table II summarizes the enumeration of n-graphs in the sample space decomposition of Section 1:2.

TABLE II

THE ENUMERATION IN THE SAMPLE SPACE DECOMPOSITION

Space	Enumeration
Ω	$\eta = 2^{n(n-1)}$
Ω_t	$\eta_t = \binom{n(n-1)}{t}$
$\omega(\underline{r})$	$\eta(\underline{r}) = \prod_1^n \binom{n-1}{r_i}$
$\omega(\underline{s})$	$\eta(\underline{s}) = \prod_1^n \binom{n-1}{s_j}$
$\omega(\underline{r},\underline{s})$	$\eta(\underline{r},\underline{s}) = A\left\{\prod_1^n (1 + \zeta_i^i)^{-1} (\underline{r},\underline{s})\right\}$

Obviously, the following relations, corresponding to 1:2.1, 1:2.2 and 1:2.3, hold:

1:4.5 $$\eta = \sum_{t} \eta_t$$

1:4.6 $$\eta_t = \sum_{(\underline{r})} \eta(\underline{r}) = \sum_{(\underline{s})} \eta(\underline{s}) \quad ,$$

1:4.7(a) $$\eta(\underline{r}) = \sum_{(\underline{s})} \eta(\underline{r},\underline{s}) \quad \text{, and}$$

1:4,7(b) $$\eta(\underline{s}) = \sum_{(\underline{r})} \eta(\underline{r},\underline{s}) \ .$$

Double and triple disjoint summations are also implied. For example,

$$\eta_t = \sum_{(\underline{r})} \sum_{(\underline{s})} \eta(\underline{r},\underline{s}) \quad .$$

According to Lemma 1.1 and the corollary following the lemma, certain of the subspaces $\omega(\underline{r}, \underline{s})$ fall into equivalence classes under permutation. In particular, for the case $r_i = d$ ($i = 1, 2, \ldots, n$), i.e., $\underline{r} = (d, d, \ldots, d) = (d^n)$, we have, according to the corollary to Lemma 1.1, that the structure for any $\underline{s}_\alpha = (s_{1\alpha}, s_{2\alpha}, \ldots, s_{n\alpha})$ is independent of the order of the components $s_{i\alpha}$ in $\underline{s}_\alpha$. Standard methods of enumerating permutations enable us to give the number of points (n-graphs) in $\omega(\underline{r})$ corresponding to $\underline{s}_\alpha = (0^{n_0}, 1^{n_1}, \ldots, k^{n_k})$,

where the superscripts n_i indicate the number of integers i in $\underline{s}_\alpha$ with $\sum_{i=0}^{k} n_i = n$ and $\sum_{i=0}^{k} i n_i = t$, as

$$1{:}4.8 \qquad \binom{n}{n_0, n_1, \ldots, n_k} \cdot \eta(d^n; 0^{n_0}, 1^{n_1}, \ldots, k^{n_k}) ,$$

where $\binom{n}{n_0, n_1, \ldots, n_k}$ is the multinomial coefficient equal to

$$\frac{n!}{\prod_{i=0}^{k} n_i!} .$$

In practice, the case $r_i = 1$ $(i = 1, 2, \ldots, n)$, does occur and for this reason it seems desirable to give a different and simpler method of computing $\eta(1^n; \underline{s})$ than by Theorem 1.1. For this method, we shall need a simple combinatorial property of multinomial coefficients.

LEMMA 1.4. *Given a set of* k *positive integers*, $s_1, s_2, \ldots, s_k$, *such that* $\sum_{j=1}^{k} s_j = n$, *then, the multinomial coefficient*

$$1{:}4.9 \qquad \binom{n}{s_1, s_2, \ldots, s_k} = \sum_{j=1}^{k} \binom{n-1}{\ldots, s_i - u_{ij}, \ldots} ,$$

where $u_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$.

PROOF:

$$\sum_{j=1}^{k} \binom{n-1}{\ldots, s_i - u_{ij}, \ldots} = \frac{(n-1)!}{(s_1-1)!s_2!\ldots s_k!} + \frac{(n-1)!}{s_1!(s_2-1)!s_3!\ldots s_k!}$$

$$+ \cdots + \frac{(n-1)!}{s_1!s_2!\ldots(s_k-1)!}$$

$$= \frac{s_1(n-1)!+s_2(n-1)!+\ldots+s_k(n-1)!}{s_1!s_2!\ldots s_k!}$$

$$= \frac{n!}{s_1!s_2!\ldots s_k!} = \binom{n}{s_1, s_2, \ldots, s_k} .$$

We now proceed to the counting theorem for the special case, $r_i = 1$.

THEOREM 1.2. *If* $r_i = 1$ $(i = 1, 2, \ldots, n)$ *and the vector* $\underline{s}$ *has components* s_j *such that* $s_j \begin{cases} \geq 1 \text{ for } j = 1,\ldots,k \\ = 0 \text{ for } j = k+1,\ldots,n \end{cases}$ *and* $\sum_{j=1}^{k} s_j = n$, *then,*

1:4.10 $$\eta(1^n; s_1, \ldots, s_n) = \frac{(n-2)!}{s_1!\ldots s_k!}\left[a_2 - \frac{a_3}{(n-2)} + \frac{a_4}{(n-2)^{(2)}} - \cdots (-1)^k \frac{a_k}{(n-2)^{(k-2)}}\right] ,$$

where $a_j = \sum_{i_1<i_2<\ldots<i_j} s_{i_1} s_{i_2} \ldots s_{i_j}$ *is the jth elementary*

symmetric function and $m^{(b)} = m(m-1)\ldots(m-b+1)$.

PROOF: First, according to the corollary following Lemma 1.1, we observe that arranging the components of the vector $\underline{s}$ in any specified order is no restriction. Second, according to Theorem 1.1 and the remark following the theorem, we have

$$\eta(1^n;\underline{s}) = A\left\{\prod_1^n (1+\delta_i^i)^{-1}(1^n,\underline{s})\right\} = A\left\{\prod_1^n (1-\delta_i^i)(1^n,\underline{s})\right\}$$

$$= A(1^n,\underline{s}) - \sum_{i_1=1}^{k} \delta_{i_1}^{i_1} A(1^n,\underline{s}) + \sum_{i_1<i_2}^{k} \delta_{i_1}^{i_1}\,\delta_{i_2}^{i_2} A(1^n,\underline{s})$$

$$- \ldots (-1)^k \sum_{i_1<i_2<\ldots<i_k} \delta_{i_1}^{i_1} \ldots \delta_{i_k}^{i_k} A(1^n,\underline{s}) .$$

Next, we observe that $A(1^n;s_1,\ldots,s_n)$ is just the multinomial coefficient $\binom{n}{s_1,\ldots,s_n}$, which may be written $\binom{n}{s_1,\ldots,s_k}$ since $s_{k+1} = \ldots = s_n = 0$. Therefore,

$$\eta(1^n;\underline{s}) = \binom{n}{s_1,\ldots,s_k} - \sum_{j_1=1}^{k} \binom{n-1}{\ldots,s_i-u_{ij_1},\ldots}$$

$$+\ldots(-1)^k \sum_{j_1<\ldots<j_k}^{k} \binom{n-k}{\ldots,s_i-u_{ij_1}-\ldots-u_{ij_k},\ldots} .$$

By Lemma 1.4, the first two terms of the right-hand member are equal, and, hence,

$$\eta(1^n;\underset{\approx}{s}) = \sum_{j_1<j_2}^{k} \binom{n-2}{\ldots,s_i-u_{ij_1}-u_{ij_2},\ldots}$$

$$-\ldots(-1)^k \sum_{j_1<\ldots<j_k}^{k} \binom{n-k}{\ldots,s_i-u_{ij_1}-\ldots-u_{ij_k},\ldots}.$$

This expression is factored to give

$$\eta(1^n;\underset{\approx}{s}) = \frac{(n-2)!}{s_1!\ldots s_k!}\left[\sum_{i_1<i_2}^{k} s_{i_1}s_{i_2} - \frac{\sum\limits_{i_1<i_2<i_3}^{k} s_{i_1}s_{i_2}s_{i_3}}{(n-2)}\right.$$

$$\left. +\ldots(-1)^k \frac{s_1 s_2\ldots s_k}{(n-2)^{(k-2)}}\right] \quad ,$$

or, finally,

$$\eta(1^n;\underset{\approx}{s}) = \frac{(n-2)!}{s_1!\ldots s_k!}\left[a_2 - \frac{a_3}{(n-2)} + \ldots(-1)^k \frac{a_k}{(n-2)^{(k-2)}}\right] .$$

We complete this section with an example of the decomposition in the special case, $r_i = 1$. Table III enumerates the directed graphs (matrices) in the decomposition of the second-order subspace,

$\omega(1^8)$, into third-order subspaces, $\omega(1^8, \underline{s}_\alpha)$, where the partitions $\underline{s}_\alpha$ are of the form $(0^{n_0}, 1^{n_1}, \ldots, k^{n_k})$. Column 2 of Table III gives the number of vectors equivalent under permutation to $\underline{s}_\alpha$, i.e., the number of permutations of 8 objects in $\underline{k}$ sets of which the first contains n_0, the second $n_1, \ldots,$ and the $\underline{k}$th n_k, where $\sum_{i=1}^{k} n_i = 8$. Column 3 gives the number of matrices in $\omega(1^8, \underline{s}_\alpha)$ for a fixed $\underline{s}_\alpha$ and is computed from 1:4.10 of Theorem 1:2. Finally, column 4 is the product of columns 2 and 3 and gives the number of matrices in $\omega(1^8)$ for which the column totals $\underline{s}$ are equivalent to $\underline{s}_\alpha$.

Note that

$$\eta(1^8) = \prod_{i=1}^{n} \binom{7}{1} = \binom{7}{1}^8 = 5,764,801$$

checks with

$$\eta(1^8) = \sum_{(\underline{s})} \eta(1^8, \underline{s}) = \sum_{(\underline{s}_\alpha)} \binom{n}{n_0, \ldots, n_k} \eta(1^8; \underline{s}_\alpha)$$

$$= 5,764,801 .$$

TABLE III

THE ENUMERATION OF POINTS IN THE DECOMPOSITION OF $\omega(1^8)$ INTO SUBSPACES, $\omega(1^8, \underline{s}_\alpha)$

$\underline{s}_\alpha = (0^{n_0}, \ldots, k^{n_k})$	$\binom{8}{n_0, \ldots, n_k}$	$\eta(1^8; \underline{s}_\alpha)$	$\binom{8}{n_0, \ldots, n_k} \eta(1^8; \underline{s}_\alpha)$
$(0^6, 1, 7)$	56	1	56
$(0^6, 2, 6)$	56	6	336
$(0^6, 3, 5)$	56	15	840
$(0^6, 4^2)$	28	20	560
$(0^5, 1^2, 6)$	168	12	2,016
$(0^5, 1, 2, 5)$	336	46	15,456
$(0^5, 1, 3, 4)$	336	85	28,560
$(0^5, 2^2, 4)$	168	130	21,840
$(0^5, 2, 3^2)$	168	180	30,240
$(0^4, 1^3, 5)$	280	93	26,040
$(0^4, 1^2, 2, 4)$	840	264	221,760
$(0^4, 1^2, 3^2)$	420	366	153,720
$(0^4, 1, 2^2, 3)$	840	562	472,080
$(0^4, 2^4)$	70	864	60,480
$(0^3, 1^4, 4)$	280	536	150,080
$(0^3, 1^3, 2, 3)$	1,120	1,143	1,280,160
$(0^3, 1^2, 2^3)$	560	1,758	984,480
$(0^2, 1^5, 3)$	168	2,325	390,600
$(0^2, 1^4, 2^2)$	420	3,578	1,502,760
$(0, 1^6, 2)$	56	7,284	407,904
(1^8)	1	14,833	14,833
			Total 5,764,801

1.5. Probability distributions.

We have discussed in some detail the structure and a decomposition of the space of all directed n-graphs, the sample space, Ω. Consider a probability measure, P, defined on Ω or a subspace of Ω. Together, Ω (or the subspace of Ω) and P form the probability space. In particular, we shall be concerned with uniform probability measures on Ω, the first-order subspaces Ω_t, and the second-order subspaces $\omega(\underline{r})[\omega(\underline{s})]$.

Next, consider a random variable X whose domain of definition is Ω or a subspace of Ω. The random variable is a single valued, measure preserving mapping from its domain to its range space. In particular, we shall be concerned with random variables whose range space is a subset of the real line. Then, the probability measure of a particular value of a random variable X is the probability measure of the inverse image under X. Hence, if the appropriate domain has a uniform probability measure on it, the probability measure of the inverse image under X is obtained by dividing the number of points in the disjoint subsets which were mapped under X into the particular value on the real line by the total number in the domain.

A large class of random variables treated in the social-psychological investigations of group organization have as their appropriate domain either Ω, the first-order subspaces Ω_t, or the second-order subspaces $\omega(\underline{r})[\omega(\underline{s})]$. The context of the particular investigation will determine the appropriate space.

Furthermore, many of these random variables have the property that their values are constant over all the points of each subset $\omega(\underline{r}_\alpha, \underline{s}_\alpha)$. Any random variable whose values depend on $\underline{r}$, $\underline{s}$ and $\underline{t}$ only has this property. A member of the class of random variables just described shall be said to have Structure L. Thus, we see that the disjoint and exhaustive decomposition of $\underline{\Omega}$ considered in Section 1:2 is directly related to Structure L.

The null distribution of a random variable X is the probability measure induced in the range space of X by the uniform probability measure over the appropriate domain of X. Then, if the random variable X has Structure L, we can construct (if not expediently) the probability distribution of X using the counting methods of Sections 1:3 and 1:4.

A number of indices (which we shall discuss in Part II) have been described in the literature, e.g., indices of group cleavage, expansiveness, and integration. These indices, being functions of random variables, are themselves random variables. The value of each index is a number on the real line and is uniquely determined by the configuration of the group, i.e., the index assigns to each point in the sample space a unique number on the real line. Thus, if the index has Structure L, the probability distribution of the index can be obtained.

1:6. Remarks.

The methods outlined give us the exact probability distributions arising in the null case for a broad class of random variables. It is unfortunate that, in some instances, the methods are computationally cumbersome. This is true, even, for moderate sized groups. In these cases, it is desirable to use simple approximations. However, it is important that the exact distribution be available in order to validate any proposed approximate distributions.

The theory developed above is based on the very simplest model of a group organization, i.e., a one-dimensional model with a binary relation between pairs of individuals which is not necessarily reflexive and takes only the values 0 and 1 in each direction. As yet no generalization has been made to models with relations which take more than two values nor to multidimensional problems.

PART II

APPLICATIONS

2:1. Preliminaries.

In this part we shall examine various distributions of random variables having Structure L and present examples of some of the indices from the literature. We hope to cover every possible situation, but, of course, not every index. It is not relevant nor is it within the scope of this thesis to consider the sociological and/or psychological aspects of the indices. For a list of the studies in the general area of index analysis see, for example, the bibliography in Leadership and Isolation by H. H. Jennings [17].

Through the remainder we shall consider only null cases in which each graph in the appropriate sample space is equally likely, i.e., a uniform probability distribution over the sample space.

In certain summations, to be encountered, we shall find it convenient to introduce the following indicator function: To a set A we assign a numerical single-valued function I_A of x, to be called the indicator of A and defined by $I_A(x) = 1$ or 0 according as $x \in A$ or $x \notin A$.

2:2. Distributions of indices on expansiveness of a social group.

One measure of "Group Expansiveness", equal to the total number of choices made by group divided by size of group, is given by Loomis and Proctor [24] in a contribution to Research Methods in

Social Relations. In our notation the index is $E = \frac{t}{n}$. A more appropriate name for this index might be Total or Gross Expansiveness.

The distribution problem, in the null case, is easily solved. Clearly, the appropriate sample space is Ω and our random variable, the number of distinct n-graphs with t lines, is constant over the subsets Ω_t in the disjoint and exhaustive decomposition of Ω. Thus, our random variable has Structure L and, according to Section 1:5 and the counting formulas of Section 1:3, the required probabilities are given by

2:2.1 $$P(t=k) = \frac{\eta_k}{\eta} = \frac{\binom{n(n-1)}{k}}{2^{n(n-1)}} .$$

This is an example of a random variable with constant values on Ω_t in Ω.

An obvious extension of the notion of Total or Gross Expansiveness is that of Relative Variability of Expansiveness, i.e., the "spread" of the numbers r_i $(i = 1, 2, \ldots, n)$ of outgoing connections for a fixed total number t. In this context the appropriate sample space is Ω_t.

As a measure of Relative Variability of Expansiveness, one might take as an index, the ratio of the standard deviation of the r_i's to their mean, $\bar{r}$.

We proceed to establish the formula giving the probabilities necessary to construct the probability distribution. First, $\omega(\underline{r})$ is "equivalent under permutation" with $\omega(\underline{Pr})$. Next, the number of

permutations in which $\underline{n}$ objects can be divided into $\underline{k}$ sets of which the first contains n_0, the second n_1, etc., is given by the multinomial coefficient

$$\binom{n}{n_0, n_1, \ldots, n_k} = \frac{n!}{n_0! n_1! \ldots n_k!}$$

Thus, the probability, in the null case, of obtaining a t-line graph with vector $\underline{r}$ equivalent under permutation to $r' = (0^{n_0}, 1^{n_1}, \ldots, k^{n_k})$, $\sum_{i=0}^{k} n_i = n$ and $\sum_{i=0}^{k} in_i = t$, is given by

$$2{:}2.2 \qquad P(r';t) = \binom{n}{n_0, n_1, \ldots, n_k} \frac{\eta(\underline{r}')}{\eta_t} = \frac{n!}{n_0! \ldots n_k!} \frac{\prod_{i=1}^{n} \binom{n-1}{r_i}}{\binom{n(n-1)}{t}} .$$

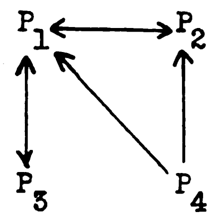
This is an example of a random variable having constant values on $\omega(\underline{r})$ in Ω_t .

2:3. Distribution of the number of isolates.

An isolate is an individual represented in the graph by a point P_i with no terminating lines and in the matrix by a column of all zeros, i.e., $s_i = 0$ in the vector $\underline{s}$. Individual 4 is such an isolate in Figure 3.

FIGURE 3

Two representations of organizational configuration on four points



(a)

$$
\begin{array}{c|cccc|}
 & 1 & 2 & 3 & 4 \\
1 & 0 & 1 & 1 & 0 \\
2 & 1 & 0 & 0 & 0 \\
3 & 1 & 0 & 0 & 0 \\
4 & 1 & 0 & 0 & 0
\end{array}
$$

(b)

Lazarsfeld, in a contribution to a paper by Moreno and Jennings [29], in 1938, gave the expected (mean) number of isolates for the case $r_i = d$ $(i = 1, 2, \ldots, n)$ as

$$n[(n-d-1)/(n-1)]^{n-1} ,$$

but he made no attempt to solve the distribution problem. Bronfenbrenner [1], in 1945, gave (without proof) an incorrect version of the distribution function for the case $r_i = d$ $(i = 1, 2, \ldots, n)$, which he claimed was "developed deductively and checked by empirical methods." Edwards [10], in 1948, gave correctly, for the same case, the probability of the maximum possible number of isolates,

2:3.1 $$P(n - 1 - d \text{ isolates}) = \binom{n}{n-1-d} \frac{(d+1)^{n-1-d}}{\binom{n-1}{d}^{n}} .$$

The exact probability distribution for this case was obtained by Katz [18], in 1950. The probability of exactly $\underline{k}$ isolates is given by

2:3.2 $$P_{[k]} = \sum_{j=k}^{n-d-1} (-1)^{k+j} \binom{j}{k} S_j \ ,$$

where $S_j = \binom{n}{j}\binom{n-j}{d}^{j} \binom{n-j-1}{d}^{n-j} \binom{n-1}{d}^{-n}$.

Using the methods developed in Part I we can now extend this result to the general case where the $\underline{i}$th individual has r_i outgoing connections, the r_i being not necessarily equal.

In most contexts, our sample space is $\omega(\underline{r})$ and, in the null case, we desire the number of graphs having a specified number of points with no terminating lines, i.e., a specified number of zeros in the vector $\underline{s}$. Our random variable X, the number of zero s_j's, is constant over the subsets $\omega(\underline{r}, \underline{s})$ in the decomposition of $\omega(\underline{r})$, i.e., has Structure L. Hence, according to Section 1:5 and the counting formulas of Sections 1:3 and 1:4, the probability of exactly $\underline{k}$ isolates is given by

2:3.3 $$P(X = k; \underline{r}) = \frac{\sum_{(\underline{s})} I_{A_k}(\underline{s})\, \eta(\underline{r},\underline{s})}{\eta(\underline{r})}$$

$$= \frac{\sum_{(\underline{s})} I_{A_k}(\underline{s}) A\left\{\prod_1^n (1+\delta_i^i)^{-1} (\underline{r},\underline{s})\right\}}{\prod_1^n \binom{n-1}{r_i}} \ ,$$

where A_k is the set of $\omega(\underline{\underline{r}},\underline{\underline{s}})$ such that the vectors $\underline{\underline{s}}$ have exactly $\underline{k}$ components equal to zero.

Thus, the chance (probability) distribution of any index (proposed for the study of group structure) which depends only on the number of isolates in the group can now be constructed. One such index, equal to the reciprocal of the number of isolates, is given by Loomis and Proctor [24] as a measure of "Group Integration".

Using Table III (p.33) and 2:3.3, we have obtained in Table IV the exact chance distribution for the number of isolates in a group of eight individuals, each making one choice (one outgoing connection). This was checked against the exact distribution obtained from Katz' formula, 2:3.2.

TABLE IV

PROBABILITY DISTRIBUTION FOR THE NUMBER OF ISOLATES IN A GROUP OF EIGHT INDIVIDUALS, EACH MAKING ONE CHOICE

k	P [exactly k isolates]
0	.002 573
1	.070 758
2	.328 435
3	.418 873
4	.162 032
5	.017 019
6	.000 311

Finally, we remark that in some contexts the appropriate sample space might be the larger space Ω_t. However, our counting methods of Sections 1:3 and 1:4 will still give us the required probabilities necessary to construct the probability distribution. In this case, the probability of exactly $\underline{k}$ isolates is given by

$$2:3.4 \qquad P(X=k;t) = \frac{\sum_{(\underline{r})} \sum_{(\underline{s})} I_{A_k}(\underline{s})\, \eta(\underline{r},\underline{s})}{\eta_t}$$

$$= \frac{1}{\binom{n(n-1)}{t}} \sum_{(\underline{r})} \sum_{(\underline{s})} I_{A_k}(\underline{s}) A\left\{\prod_1^n (1+\delta_i^i)^{-1} (\underline{r},\underline{s})\right\} ,$$

where the set A_k is the same as before.

These are examples of random variables with constant values on $\omega(\underline{r}, \underline{s})$ in $\omega(\underline{r})$ or Ω_t.

2:4. Distribution of maximum s_j.

In group organization theory, individuals who have the maximum s_j are sometimes called stars. The problems of determining whether a group contains a star or stars might be handled by consideration of the conditional distribution of the vector $\underline{s}$ of C, given the vector $\underline{r}$. The distributions of the maximum s_j, the two maximum s_j, etc., could then be used for determining stars.

The distribution problem, in the null case, for maximum s_j is analogous to the one for isolates. Our random variable, max s_j, is constant over $\omega(\underline{\underline{r}}, \underline{\underline{s}})$, i.e., has Structure L. The appropriate sample space is $\omega(\underline{\underline{r}})$ or Ω_t depending on the context of the particular investigation. Thus, the probabilities that max $s_j = k$ are given for the two situations by

2:4.1(a)
$$P(\max s_j = k; r) = \frac{\sum_{(\underline{\underline{s}})} I_{B_k}(\underline{\underline{s}})\, \eta(\underline{\underline{r}}, \underline{\underline{s}})}{\eta(\underline{\underline{r}})}$$

$$= \frac{\sum_{(\underline{\underline{s}})} I_{B_k}(\underline{\underline{s}}) A \left\{ \prod_1^n (1+\delta_i^i)^{-1} \quad (\underline{\underline{r}}, \underline{\underline{s}}) \right\}}{\prod_1^n \binom{n-1}{r_i}} ,$$

and

2:4.1(b)
$$P(\max s_j = k; t) = \frac{\sum_{(\underline{\underline{r}})} \sum_{(\underline{\underline{s}})} I_{B_k}(\underline{\underline{s}})\, \eta(\underline{\underline{r}}, \underline{\underline{s}})}{\eta_t}$$

$$= \frac{\sum_{(\underline{\underline{r}})} \sum_{(\underline{\underline{s}})} I_{B_k}(\underline{\underline{s}}) A \left\{ \prod_1^n (1+\delta_i^i)^{-1} \quad (\underline{\underline{r}}, \underline{\underline{s}}) \right\}}{\binom{n(n-1)}{t}} ,$$

respectively, where B_k is the set of $\omega(\underline{\underline{r}}, \underline{\underline{s}})$ such that the vectors $\underline{\underline{s}}$ have the maximum component equal to $\underline{k}$.

These are, again, examples of random variables with constant values on $\omega(\underline{\underline{r}}, \underline{\underline{s}})$ in $\omega(\underline{\underline{r}})$ or Ω_t.

Using Table III (p.33) and 2:4.1(a) we have obtained in Table V the cumulative probability distribution function for max s_j in a group of eight individuals, each making one choice.

TABLE V

CUMULATIVE PROBABILITY DISTRIBUTION FOR MAX s_j
IN A GROUP OF EIGHT INDIVIDUALS, EACH MAKING ONE CHOICE

k	P[max $s_j \leq$ k]
1	.002 573
2	.515 275
3	.918 897
4	.992 238
5	.999 582
6	.999 990
7	1.000 000

2:5. Remarks.

The major classical problems not having random variables with Structure L are

(i) the distribution of the number of mutuals, and

(ii) the problem of cleavage in a group.

A mutual is a pair of individuals i,j such that a connection exists from i to j as well as from j to i, i.e., symmetrically placed 1's in the matrix C. Cleavage is measured by the relative intensities of in-group and out-group connections in the situation where we have a group divided into subgroups by some external factor, e.g., by sex, or by nationality, or by management-labor. These problems will not be considered in this thesis.

We have presented methods for obtaining the exact distributions of a large class of random variables. In the cases of certain of these, e.g., the number of isolates and of the maximum s_j, the computations necessary to construct the distributions for even moderate sized groups are lengthy. For this reason we shall, in the next part, suggest simple approximate distributions for these variables. Even though the exact distribution may not be used it is important that it be known in order to validate any suggested approximations.

PART III

APPROXIMATE DISTRIBUTIONS

3:1. Preliminaries.

In Part II, two of the exact distributions considered, namely, distribution of number of isolates and distribution of maximum s_j, are extremely tedious to compute using the counting methods of Part I. From a field investigator's point of view, the practical problems of investigation require simple, easily-computed distributions. Therefore, with this in mind, we shall suggest for each of the above-mentioned variables a suitable approximate distribution. Also, we shall propose an index which seems reasonably suited for measuring the phenomenon of "concentration of choice" and suggest an easily-computed approximate distribution for it. Further, we shall make certain computational checks of each of the approximations against the exact distributions. These checks are not to be construed as answering all the field investigator's questions pertaining to the use of the approximations but, only, as giving an indication as to the relative orders of magnitude of the approximations. It will, of course, be necessary for field investigators to examine more closely and in greater detail the suggested approximation with reference to the size of the group, the numbers of choices made by the individuals, etc. It is beyond the scope of this thesis to consider in any detail these latter problems.

3:2. Approximate distribution of number of isolates.

Katz [18], for the case $r_i = d$ ($i = 1, 2, \ldots, n$) gives the limiting distribution as Poisson with P [exactly k isolates]

$= e^{-\lambda} \frac{\lambda^k}{k!}$, where $\lambda = n(1 - \frac{d}{n-1})^{n-1}$. However, for moderate values of n, Katz points out that the approximation is not too good. On the other hand, he shows that a binomial approximation is quite good. We shall modify the formulas given by Katz to include the general case where all the r_i are not necessarily equal.

First, we shall find it convenient to give a different form of the exact distribution than that obtained in Section 2:3. Let r_{i_j} be the number of choices made by individual i_j. Then, by a straightforward generalization of the exact distribution given by Katz for the case $r_i = d$ ($i = 1, 2, \ldots, n$), we have that the probability of exactly k isolates for the general case is given by

3:2.1 $$P_{[k]} = \sum_{h=k}^{\max_j(n-1-r_{i_j})} (-1)^{k+h} \binom{h}{k} S_h ,$$

where $S_h = \sum \left\{ \prod_{j=1}^{h} \frac{(n-h)^{(r_{i_j})}}{(n-1)^{(r_{i_j})}} \prod_{j=h+1}^{n} \frac{(n-h-1)^{(r_{i_j})}}{(n-1)^{(r_{i_j})}} \right\}$; the summation being over the $\binom{n}{h}$ distinct combinations of h i_j's at a time.

Katz points out that Fréchet has shown that the factorial moments are given by

$$\alpha_{(k)} = k!\, S_k \quad .$$

Hence, rewriting S_h as

$$S_h = \frac{(n-h)^h \prod_{j=1}^{n} (n-h-1)^{(r_{i_j})}}{\prod_{j=1}^{n} (n-1)^{(r_{i_j})}} \sum \prod_{j=1}^{h} \frac{1}{(n-h-r_{i_j})} \quad ,$$

the summation being the same as before, we have that the mean is

3:2.2 $$\text{mean} = \bar{x} = \frac{(n-1) \prod_i (n-2)^{(r_i)}}{\prod_i (n-1)^{(r_i)}} \sum_i \frac{1}{(n-1-r_i)} \quad .$$

For the case $r_i = d$ for all i, Katz gives the mean as $\text{mean} = \bar{\bar{x}} = n(1 - \frac{d}{n-1})^{n-1}$. We obtain from Katz' formula for the mean the following expression

3:2.3 $$\text{mean} = \bar{\bar{x}} = \frac{n}{(n-1)^{n-1}} \, (d')^{n-1} \quad ,$$

where $d' = n-1-d$. Rewriting 3:2.2, we have

3:2.4 $$\bar{x} = \frac{n}{(n-1)^{n-1}} [(\prod_i r_i')(\frac{1}{n} \sum_i \frac{1}{r_i'})] ,$$

where $r_i' = n-1-r_i$. Letting G and H stand for the geometric and harmonic means of the r_i''s, respectively, we have

3:2.5 $$\bar{x} = \frac{n}{(n-1)^{n-1}} \frac{G^n}{H} .$$

When the variation in the r_i''s is not too great the geometric and harmonic means are reasonably close together. Hence, the following approximation is suggested for $\bar{x}$:

3:2.6 $$\hat{\bar{x}} = \frac{n}{(n-1)^{n-1}} G^{n-1} .$$

Note that for the case $r_i = d$ for all $\underline{i}$, we have $G = H$ and 3:2.6 is exact.

For computing the binomial probabilities, $p_i = b(i;m,p)$, Katz uses the following two formulas to obtain m and p:

3:2.7 $$mp = n(1 - \frac{d}{n-1})^{n-1} , \text{ and}$$

3:2.8 $$\frac{1}{m} = 1 - (\frac{n-1}{n})[1 - \frac{d}{(n-2)(n-1-d)}]^{n-2} .$$

We remark that m is not necessarily an integer. Rewriting 3:2.7 and 3:2.8 in terms of d', we have

3:2.9 $$mp = \frac{n}{(n-1)^{n-1}} (d')^{n-1} \text{, and}$$

3:2.10 $$\frac{1}{m} = 1 - \frac{(n-1)^{n-1}}{n(n-2)^{n-2}} \left(\frac{d'-1}{d'}\right)^{n-2} .$$

Next, we equate 3:2.3 and 3:2.6 obtaining an estimate for d' as

3:2.11 $$\hat{d}' = G .$$

Note that $d' = G$ if $r_i = d$ for all i.

Finally, substituting 3:2.11 in 3:2.9 and in 3:2.10, we obtain the formulas necessary for computing the binomial probabilities required for the binomial approximation to the exact distribution. The formulas are

3:2.12 $$mp = \frac{n}{(n-1)^{n-1}} G^{n-1} \text{, and}$$

3:2.13 $$\frac{1}{m} = 1 - \frac{(n-1)^{n-1}}{n(n-2)^{n-2}} \left(\frac{G-1}{G}\right)^{n-2} .$$

As an example, we consider the distribution of the number of isolates over the sample space $\omega(1,1,2,2)$. For this small group it is easier to obtain the exact distribution by examining directly the 81 sample points (matrices) instead of by the method given in Section 2:3. The probabilities were checked by using formula 3:2.1. These values appear in the second column of Table VI.

For the computation of the approximate probabilities, we obtain $p = .35836$ from 3:2.12 and $m = 1.16927$ from 3:2.13. We then compute the binomial probabilities, $p_i = b(i;m,p)$, $i = 0,1,2,\ldots,([m]+1)$, where $[m]$ is the greatest integer in $\underline{m}$ (in this case, 1) using $p_0 = (1-p)^m$ and $p_{i+1}/p_i = (m-i)p/(i+1)(1-p)$ as used by Katz. These approximate probabilities appear in the third column. It will be seen that the fit is not too good. However, the size of the group is very small. Since the computations of the exact probabilities for a larger group are extremely lengthy, an example of larger magnitude was not undertaken.

TABLE VI

COMPARISON OF THE EXACT AND APPROXIMATE

DISTRIBUTIONS OF THE NUMBER OF ISOLATES

FOR $n = 4$, $\underline{r} = (1,1,2,2)$

k	$P_{[k]}$(exact)	p_k(approx.)
0	.56790	.59521
1	.41975	.38870
2	.01235	.01609

3:3. Approximate distribution of maximum s_j.

Since the variable under consideration in this section is expressed in terms of the column totals, s_j, of the matrix C, it will be desirable to examine the asymptotic distribution of the random variable s_j.

Let $r_i = d$ for $i = 1, 2, \ldots, n$. Then, assuming the d choices are distributed uniformly over the n-1 places, i.e., $P[c_{ij}=1] = \frac{d}{n-1}$, the random variable s_j follows, for each j ($j = 1, 2, \ldots, n$), the binomial distribution

$$P[s_j=k;\ n-1,p] = \binom{n-1}{k} p^k (1-p)^{n-1-k} \quad ,$$

where $p = \frac{d}{n-1}$.

Next, consider the limiting form of the binomial distribution as $p \longrightarrow 0$ subject to the condition that $(n-1)p = d$. Under these conditions

$$P[s_j=k;\ n-1,p] \sim e^{-d} \frac{d^k}{k!} \quad ;$$

see, for example, Cramér [5, p. 203]. This is known as the "Poisson approximation to the binomial distribution". Feller [11, p. 111], shows that the error is of the order of magnitude $d^2/(n-1)$.

Thus, the limiting form for the probability of $s_j \leq k$ ($j=1,2,\ldots,n$) is given by

3:3.1 $$P[s_j \leq k] = \sum_{s_j=0}^{k} e^{-d} \frac{d^{s_j}}{s_j!} \quad .$$

If the s_j ($j = 1, 2, \ldots, n$) are independent, the limiting form for the probability of the maximum $s_j \leq k$ is

3:3.2 $$P[\max s_j \leq k] = \left\{ \sum_{s_j=0}^{k} e^{-d} \frac{d^{s_j}}{s_j!} \right\}^n .$$

Of course, the s_j's are not independent since they have one constraint, namely, $\sum_{j=1}^{n} s_j = t$. However, 3:3.2 satisfies the requirement that we have a simple, easily-computed approximation to the exact distribution. Hence, we suggest the use of 3:3.2 to obtain the probabilities necessary to construct the approximate cumulative distribution of maximum s_j.

As an example, we consider the cumulative probability distribution over the sample space $\omega(1^8)$. The exact distribution is obtained from Table V (p. 45). These values appear in the second column of Table VII. The approximate probabilities are obtained from 3:3.2, using tables on Poisson's Exponential Binomial Limit by E.C. Molina [27]. These approximations appear in column three. The fit is not too good at the low end but improves at the high end in which we are mainly interested.

TABLE VII

COMPARISON OF THE EXACT AND APPROXIMATE

CUMULATIVE DISTRIBUTIONS OF MAXIMUM s_j

FOR $n = 8$, $\underline{\underline{r}} = (1^8)$

k	P[max $s_j \leq k$; (exact)]	P[max $s_j \leq k$; (approx.)]
1	.00257	.08588
2	.51527	.51188
3	.91890	.85782
4	.99224	.97109
5	.99958	.99526
6	.99999	.99934
7	1.00000	.99992

If the r_i's are not constant, we take, for each $\underline{i}$, $P[c_{ij}=1]$ = expected value = $\frac{r_i}{n-1}$. Then, for each $\underline{j}$, the Poisson approximation takes the following form:

3:3.3 $$P[s_j = k] = e^{-\lambda_j} \frac{\lambda_j^k}{k!} ,$$

where λ_j equals the total of the n-1 expected values of the c_{ij}'s $(i=1,\ldots,j-1,j+1,\ldots,n)$, i.e., $\lambda_j = \frac{(t-r_j)}{(n-1)}$. We, then, suggest the use of

$$3:3.4 \qquad P[\max s_j \leq k] = \prod_j \left\{ \sum_{s_j=0}^{k} e^{-\lambda_j} \frac{\lambda_j^{s_j}}{s_j!} \right\} ,$$

as an approximation to the exact distribution. Further, if the variability of the r_i's is not large, we might take $\lambda_j = \lambda = \frac{(t-\bar{r})}{(n-1)}$ for all j, where $\bar{r}$ = mean of the r_i's. Then 3:3.4 becomes

$$3:3.5 \qquad P[\max s_j \leq k] = \left\{ \sum_{s_j=0}^{k} e^{-\lambda} \frac{\lambda^{s_j}}{s_j!} \right\}^n .$$

We shall not consider any check of the suggested approximation for the case r_i not all constant.

3:4. Concentration of choice.

In studying the phenomenon of concentration of expressed choices on relatively few individuals and/or of sparsity of choices going to certain others, it seems reasonable to use as test criterion the variance of numbers of choices received, suitably standardized. Actually, for the case $r_i = d$ for all i with the one restriction, $\sum_{i=1}^{n} r_i = t$, the s_j's in their limiting form make up a Poisson series. Hence, we shall use Fisher's [12, p. 58] "index of dispersion" calculated by means of the formula

3:4.1 $$I_c = \frac{1}{\bar{s}} \sum_{j=1}^{n} (s_j - \bar{s})^2 ,$$

where $\bar{s}$ = mean = $\frac{1}{n} \sum_{j=1}^{n} s_j$. The sum, I_c, is large if either or both of the previously mentioned effects occur in an appreciable manner.

It has been shown by Sukhatme [36] and others that the index of dispersion, calculated as in 3:4.1, follows the ordinary χ^2 distribution with n-1 degrees of freedom. Tables for this distribution are readily available, for example, Cramér [5, p. 559].

Sukhatme [37] has done some empirical work on the fit of the χ^2 approximation for smallish n and small values of the Poisson parameter, λ. His results show that the agreement is "tolerably good", even when $\lambda = 1$, for n = 15. In fact the fit is not too bad for $\lambda = 1$, n = 10. However, in our situation we have a twofold approximation; first, the Poisson approximation and, second, the χ^2 approximation. Hence, it is necessary to make a further check of the approximation to the exact distribution.

Some empirical work indicates that a χ^2 approximation with n-1 degrees of freedom, corrected for continuity as suggested by Cochran [4, p. 332], gives a good fit in our case. As an example, consider a group of eight individuals, each making one choice. In column 2 of Table VIII we have the values of the index, computed from 3:4.1, for each g_α in column 1. The exact cumulative probabilities are found from the data compiled in Table III and are entered in column 3. The approximate probabilities appear in column 4 and were obtained from Tables for Statisticians and Biometricians [39, p. 26]. This check indicates that the approximation is reasonably good even for quite small groups making small numbers of choices. Thus, we are in a position to recommend the use of the above test criterion.

TABLE VIII

COMPARISON OF EXACT AND APPROXIMATE CUMULATIVE DISTRIBUTIONS OF INDEX OF CONCENTRATION OF CHOICE FOR $n = 8$, $\underline{r} = (1^8)$

$\underline{g}_\alpha$	I_c'	$P[I_c \leq I_c';(\text{exact})]$	$P[I_c \leq I_c';(\text{approx.})]$
(7,1)	42	1.00000	1.00000
(6,2)	32	.99999	1.00000
$(6,1^2)$	30	.99993	.99995
(5,3)	26	.99958	.99978
(4,4)	24	.99944	.99924
(5,2,1)	22	.99934	.99829
$(5,1^3)$	20	.99666	.99623
(4,3,1)	18	.99214	.99181
$(4,2^2)$	16	.98719	.98260
$(3^2,2),(4,2,1^2)$	14	.98340	.96400
$(3^2,1^2),(4,1^4)$	12	.93968	.92789
$(3,2^2,1)$	10	.88699	.86138
$(2^4),(3,2,1^3)$	8	.80510	.74734
$(2^3,1^2),(3,1^5)$	6	.57254	.57112
$(2^2,1^4)$	4	.33401	.34004
$(2,1^6)$	2	.07333	.11500
(1^8)	0	.00257	.00517

SUMMARY

We have been concerned with the one-dimensional theory of group organization as a complex of irreflexive binary relationships, taking values 0 and 1, between the pairs of individuals. The problems considered in connection with this theory were

(i) the investigation of the appropriate universe or universes of discourse,

(ii) the determination of the null distributions for certain proposed indices of the group structure, and

(iii) the development of simple, reasonably exact methods for use by field investigators.

The results obtained were:

1. a decomposition of the total sample space was given which clarifies the choosing of the appropriate universe of discourse,

2. the machinery for counting the number of distinct points in each of the subspaces in the decomposition of the total sample space was developed,

3. the structure of random variables whose null distributions are possible to obtain by using the developed counting methods was expounded,

4. applications were given to classes of unsolved probability distribution problems of group organization theory; in particular, the null distributions were given for (a) indices on expansiveness of a social group, (b) the number

of isolates in a group, and (c) the maximum s_j in a group,

5. simple, easily-computed, approximate distributions were obtained for the number of isolates and for the maximum s_j,

6. a test criterion for concentration of choice was proposed and a simple, approximate distribution suggested for it.

The first kind of problem was solved by the first result in the sense that it is now possible to examine the various sociometric variables in the general framework of the decomposition of the total sample space given here and, thereby, determine their appropriate universes of discourse.

A combination of results 2, 3 and 4 solved the second problem. The machinery of counting methods made it possible to obtain the necessary probabilities for construction of the null distributions. These probabilities were obtained by dividing the number of points in the disjoint subsets in the framework of the decomposition of the total sample space by the total number of points in the appropriate universe of discourse. The discussion on the structure of random variables enabled one to tell immediately whether a particular sociometric variable fell into one of the classes of random variables whose null distributions can be obtained using the theory developed here. The applications illustrated the above results.

The third kind of problem is partially solved by the fifth and sixth results. This particular problem covers such a wide range of possible investigations which might be undertaken that it was not feasible to consider it in any great detail. Thus, we have given only an indication of the type of solution one might expect. In particular, result 5 contains one aspect of this problem in the sense that approximate distributions were given and checked for two sociometric variables found in the literature. The other aspect of the third problem is indicated in the sixth result in that we proposed a test criterion not found in the literature and suggested and checked an approximate method for obtaining its null distribution.

Related problems considered in this research investigation which led to this thesis include the null distribution of the number of mutuals and the problem of cleavage in a group. These problems were not considered here since neither of them has the appropriate structure in the general framework of this thesis. However, it should be pointed out that an iterative method of application of the counting methods developed here enables one to solve both of these problems.

The unsolved problems in connection with this research are

(i) definitions of non-null cases,

(ii) multidimensional problems,

(iii) relaxation of scale restriction,

(iv) extension of existing tables of the bipartitional functions, $A(\underline{\underline{r}}, \underline{\underline{s}})$, and compilation of tables of the functions, $\eta(\underline{\underline{r}}, \underline{\underline{s}})$, and

(v) more work on the development of simple, approximate methods for use by field investigators.

In connection with the first of the unsolved problems, it would be necessary for the social-psychologists to define what they consider to be a suitable non-null case before any theoretical investigation should be undertaken.

As for the second problem, some exploratory work has been done but few results are as yet known.

The relaxation of the scale restriction means, for one thing, that we would have a matrix with elements not all 0's and 1's, and/or a graph with strengths. The graphical representation becomes very awkward and, for all practical purposes, the matrix representation would have to be used. Also, any relaxation of scale means that the distributions of random variables are virtually impossible to obtain.

Extension of existing tables of the bipartitional functions $A(\underline{\underline{r}}, \underline{\underline{s}})$, would be desirable since the method for counting the number of locally restricted directed graphs involves these functions. Some time has been spent on trying to arrange the computations necessary for extending these tables in such a way that actual extension might be carried out by some kind of large scale computational machinery. The results were not very successful. However, we are able to express in fairly compact form the numbers corresponding to sociometric situations in which each individual in the group makes precisely one choice. Even more important for our use is the compilation of tables of the functions, $\eta(\underline{\underline{r}}, \underline{\underline{s}})$. Of course, this problem depends to a large extent on the success or failure of the preceding one. The table of A's has a double entry for any weight, being entered for the non-ordered partitions whereas the η table would be the same kind except the order of the partitions is now important. Wherefore, the tabulation of the η's has the added complication that there are many more entries for each weight since there are many more ordered partitions than partitions for any given weight. Nothing tractable has been found for this latter problem except for the special case where each individual makes only one choice.

The developing and checking of simple, approximate distributions is, as we mentioned earlier, an extremely broad problem. It is one where close cooperation between the field investigators and the theorists would be desirable.

Before finishing this thesis it should be pointed out that the general methods developed here may be applied to the theory of communication networks and to other near-sociometric problems.

Finally, the following two published papers by Katz and Powell [19, 20]:

(a) "A proposed index of the conformity of one sociometric measurement to another,"

(b) "The number of locally restricted directed graphs,"

include some of the results in this thesis.

BIBLIOGRAPHY

1. Bronfenbrenner, U. "The measurement of sociometric status, structure and development," Sociometry Monographs, No. 6, Beacon House, New York, 1945.

2. Cayley, A. "A memoir on the symmetric functions of the roots of an equation," Philos. Trans. Roy. Soc. London. Ser. A, Vol. 147 (1857), pp. 489-499.

3. Christie, L.S., R.D. Luce, and J. Macy, Jr. "Communication and learning in task-oriented groups," Technical Report No. 231, Research Laboratory of Electronics, M.I.T., 1952.

4. Cochran, W.G. "The χ^2 test of goodness of fit," Ann. Math. Stat., Vol. 23 (1952), pp. 315-345.

5. Cramér, H. Mathematical Methods of Statistics, Princeton University Press, 1946.

6. Criswell, J.H. "Notes on the constant frame of reference problem," Sociometry, Vol. 13 (1950), pp. 93-107.

7. David, F.N. and M.G. Kendall. "Tables of symmetric functions. Parts II and III," Biometrika, Vol. 38 (1951), pp. 435-462.

8. Davis, R.L. "The number of structures of finite relations," Proc. Amer. Math. Soc., Vol. 4 (1953), pp. 486-495.

9. Durfee, W.P. "Tables of the symmetric functions of the twelfthic," Amer. J. Math., Vol. 5 (1882), pp. 46-61.

10. Edwards, D.S. "The constant frame of reference problem in sociometry," Sociometry, Vol. 11 (1948), pp. 372-379.

11. Feller, W. An Introduction to Probability Theory and Its Applications, John Wiley and Sons, New York, 1950.

12. Fisher, R.A. Statistical Methods for Research Workers, 2d ed. Oliver and Boyd, London, 1928.

13. Forsyth, E. and L. Katz. "A matrix approach to the analysis of sociometric data: Preliminary report," Sociometry, Vol. 9 (1946), pp. 340-347.

14. Harary, F. "The number of graphs and directed graphs," (Abstract), Bull. Amer. Math. Soc., Vol. 58 (1952), p. 664.

15. Harary, F. and R.Z. Norman. "Graph theory as a mathematical model in social science," Research Center for Group Dynamics Monograph, No. 2, Edwards Brothers, Inc., Ann Arbor, 1953.

16. Harary, F. and G.E. Uhlenbeck. "On the number of Husimi trees, I," Proc. Nat. Acad. Sci. U.S.A., Vol. 39 (1953), pp. 315-322.

17. Jennings, H.H. Leadership and Isolation, 2d ed., Longmans, Green and Co., New York, 1950.

18. Katz, L. "The distribution of the number of isolates in a social group," Ann. Math. Stat., Vol. 23 (1952), pp. 271-276.

19. Katz, L. and J.H. Powell. "A proposed index of the conformity of one sociometric measurement to another," Psychometrica, Vol. 18 (1953), pp. 249-256.

20. ____________. "The number of locally restricted directed graphs," Proc. Amer. Math. Soc., Vol. 5 (1954), pp. 621-626.

21. Landau, H.G. "On dominance relations and the structure of animal societies: I. Effect of inherent characteristics," Bull. Math. Biophysics, Vol. 13 (1951), pp. 1-19.

22. ______. "On dominance relations and the structure of animal societies: II. Some effects of possible social factors," Bull. Math. Biophysics, Vol. 13 (1951), pp. 245-262.

23. ______. "On dominance relations and the structure of animal societies: III. The condition for a score structure," Bull. Math. Biophysics, Vol. 15 (1953), pp. 143-148.

24. Loomis, C.P. and C.H. Proctor. "Analysis of Sociometric Data," Research Methods in Social Relations, Part 2, Chap. 17, The Dryden Press, New York, 1951.

25. Luce, R. Duncan and Albert D. Perry. "A method of matrix analysis of group structure," Psychometrika, Vol. 14 (1949), pp. 95-116.

26. MacMahon, P.A. Combinatory Analysis, Vol. I, Cambridge University Press, 1915.

27. Molina, E.C. Poisson's Exponential Binomial Limit, D. Van Nostrand Co., 1942.

28. Moreno, J.L. Who Shall Survive?, A New Approach to the Problem of Human Interrelations, Beacon House, New York, 1934.

29. Moreno, J.L. and H.H. Jennings. "Sociometric statistics of social configurations," Sociometry Monographs, No. 3, Beacon House, New York, 1945.

30. Otter, R. "The number of trees," Ann. of Math., Vol. 49 (1948), pp. 583-599.

31. Pólya, G. "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen," Acta Math., Vol. 68 (1937), pp. 145-254.

32. Rapoport, A. "Outline of a probabilistic approach to animal sociology: I," Bull. Math. Biophysics, Vol. 11 (1949), pp. 183-196.

33. ________. "Outline of a probabilistic approach to animal sociology: II," Bull. Math. Biophysics, Vol. 11 (1949), pp. 273-281.

34. ________. "Outline of a probabilistic approach to animal sociology: III," Bull. Math. Biophysics, Vol. 12 (1950), pp. 7-17.

35. Shimbel, A. "Applications of matrix algebra to communication nets," Bull. Math. Biophysics, Vol. 13 (1951), pp. 165-178.

36. Sukhatme, P.V. "The problem of k samples for Poisson population," Proc. Nat. Inst. Sci. India, Vol. 3 (1937), p. 297.

37. ________. "On the distribution of χ^2 in samples of the Poisson series," J. Roy. Stat. Soc. (Suppl.), Vol. 5 (1938), pp. 75-79.

38. ________. "On bipartitional functions," Philos. Trans. Roy. Soc. London. Ser. A, Vol. 237 (1938), pp. 375-409.

39. Tables for Statisticians and Biometricians, Part I, 2d ed. Cambridge University Press, 1924.

40. Thrall, R.M. "Ranking of community organizations," Report No. M-965, R-1, Engineering Research Institute, U. of Mich., 1953.