PROBABILITY OF INDECOMPOSABILITY OF A RANDOM MAPPING FUNCTION[1]

By Leo Katz[2]

Michigan State University

Summary. Consider a finite set Ω of N points and a single-valued function $f(x)$ on Ω into Ω. In case the mapping is one-to-one, it is a permutation of the points of Ω; we shall be concerned with more general mappings. Any mapping function effects a decomposition of the set into disjoint, *minimal*, non-null invariant subsets, as $\Omega = \omega_1 + \omega_2 + \cdots + \omega_k$, where $f(\omega_i) \subset \omega_i$ and $f^{-1}(\omega_i) \subset \omega_i$. These subsets have been referred to as trees and as components of the mapping; we shall say that f, as above, decomposes the set into k components.

Metropolis and Ulam [1] defined a *random* mapping by a uniform probability distribution over the Ω^Ω sample points of $f(x)$ and posed the problem of finding the expected number of components. Kruskal [2] subsequently solved this problem. In this paper, we consider a related problem, namely, what is the probability that a random mapping is indecomposable, i.e., that the minimal non-null set ω for which $f(\omega) = \omega$ and $f^{-1}(\omega) = \omega$, is the whole set $\omega = \Omega$?

This problem is solved in general, as is, also, an analogous problem for a specialized random mapping of some interest in social psychology. Finally, we examine the asymptotic behavior of these probabilities.

1. Indecomposability of a random mapping. A single-valued mapping specifies, for each point P_i, its image point P_{j_i}, $j_i = 1, 2, \cdots, N$ (a point may map into itself). A *random* mapping assigns, independently, to each P_i one of the image points P_j, $j = 1, 2, \cdots, N$, with equal probability $1/N$. The sample space consists of the N^N possible mappings, with uniform probability distribution. To each mapping is associated a value of the random variable k, $k = 1, 2, \cdots, N$, the number of components. Those for which $k = 1$ are indecomposable. We shall require, first, a characterization of the property of indecomposability, second, a disjunctive and exhaustive categorization of those mappings which possess this property, and, finally, an enumeration scheme within each category.

In order to obtain a suitable characterization of indecomposability, we consider that a single-valued mapping function takes any point of the (finite) set into a second, the second into a third, etc., until, at some stage, a point is taken into an earlier member of the sequence. At this stage, a cycle is formed; the length of the cycle is the number of repetitions of the mapping required to

Received November 8, 1954.

[1] Presented at the meeting of the Institute of Mathematical Statistics at Berkeley, California, December 27–30, 1954.

[2] Work done under the sponsorship of the Office of Naval Research.

take any point of the cycle into itself. No point of a cycle can be mapped on any point not of the cycle, but a point not of a cycle may map through a chained sequence into a point of the cycle. Thus, a *component* of a mapping consists of precisely one cycle, together with cycle-free chains terminating at points of the cycle. This provides the required characterization:

CHARACTERIZATION. A mapping function is indecomposable if and only if it generates only one cycle.

Next, we may categorize indecomposable mappings according to the length m of the cycle contained in them. Finally, we subcategorize m-cycle indecomposable mappings into sets according as the noncyclic elements are arranged with n_j requiring j stages to be mapped into the cycle, $j = 1, 2, 3, \cdots$. This subclassification corresponds to the nonzero, p-part, partitions of $(N - m)$, with p arbitrary.

We now view the indecomposable mapping as a directed graph, more precisely, as a tree rooted in an m-cycle. The directed joins, one emanating from each point, represent the mapping from point to image. In the following section, we shall consider that the graph of an indecomposable mapping consists of a central m-cycle, a first orbit of n_1 points connected by one-chains to the cycle, a second orbit of n_2 points connected by one-chains to the points of the first orbit and, hence, by two-chains to the points of the cycle, etc. An example of such an indecomposable mapping with $m = 6$, $n_1 = 4$, $n_2 = 3$ is given in Figure 1, below.

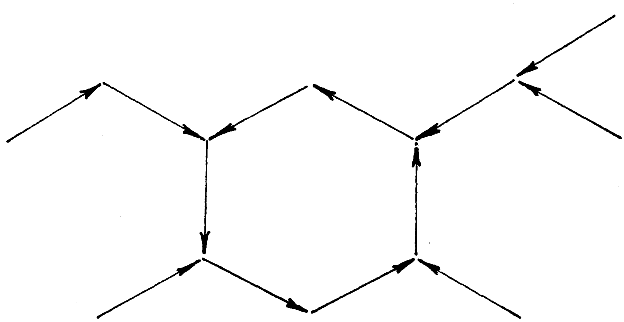

FIG. 1. Example of Mapping $m = 6$, $n_1 = 4$, $n_2 = 3$

2. Probability that mapping is indecomposable. We proceed formally, at first, to express the probability of indecomposability as the sum of compound probabilities that the mapping produces exactly one cycle *and* the cycle is of length m. These, in turn, are expressed as the sums of probabilities that the remaining $M = N - m$ are arranged in nonempty orbits of $n_1, n_2, \cdots, n_p$, respectively, for all possible such arrangements. It is convenient to give special treatment to the number m in the cycle itself. Consider the event $E_m(n_1, n_2, \cdots, n_p)$ that a random mapping is indecomposable with parameters $m, n_1, \cdots, n_p$. The probability of this event is the curiously linked expression

$$
\begin{aligned}
(1)\qquad & P\{E_m(n_1, n_2, \cdots, n_p)\} \\
& = \binom{N}{m, n_1, \cdots, n_p} \frac{(m-1)!}{N^m} \left(\frac{m}{N}\right)^{n_1} \left(\frac{n_1}{N}\right)^{n_2} \cdots \left(\frac{n_{p-1}}{N}\right)^{n_p},
\end{aligned}
$$

where $\binom{a}{b, c, d, \ldots}$ is the multinomial coefficient. The factorial in the second factor of the right member represents the number of distinct cyclical arrangements possible among the m points of the inner cycle; succeeding factors represent the possibilities of joins of points in an orbit to points of the next interior orbit.

With slight rearrangement of (1), the probability we seek is given by

$$
(2)\qquad \Pr\{\text{indecomposability}\} = \sum_{m=1}^{N} \frac{N!}{N^N} \left\{ \sum_p \sum_{[M]_p} \frac{1}{m} \frac{m^{n_1} n_1^{n_2} \cdots n_{p-1}^{n_p}}{n_1! n_2! \cdots n_p!} \right\},
$$

where $[M]_p$ stands for the collection of nonempty, p-part partitions

$$(n_1, n_2, \cdots, n_p)$$

of M.

We now evaluate the expression in braces in (2) by the following lemma.

LEMMA.

$$
\sum_p \sum_{[M]_p} \frac{1}{m} \frac{m^{n_1} n_1^{n_2} \cdots n_{p-1}^{n_p}}{n_1! n_2! \cdots n_p!} = \frac{N^{M-1}}{M!}
$$

where $N = M + m$ and $[M]$ is the class of nonzero distinct partitions $(n_1, \cdots, n_p)$ of M.

PROOF.[3] We proceed indirectly by expanding the binomial in the right member as

$$
(3)\qquad \frac{(M+m)^{M-1}}{M!} = \sum_{n_1=1}^{M} \binom{M-1}{n_1-1} \frac{1}{M!} m^{n_1-1} M^{M-n_1},
$$

or, letting $M_1 = M - n_1$ and simplifying,

$$
(3a)\qquad \frac{(M+m)^{M-1}}{M!} = \sum_{n_1=1}^{M} \frac{m^{n_1-1}}{(n_1-1)!} \frac{(M_1+n_1)^{M_1-1}}{M_1!}.
$$

We note that the second factor in the summand of (3a) is of the same type as the left member and may be similarly expanded. Letting $M_i = M_{i-1} - n_i$, $i = 2, 3, \cdots$, we obtain, by iteration of (3a),

$$
(4)\qquad \frac{(M+m)^{M-1}}{M!} = \sum_{n_1=1}^{M} \frac{m^{n_1-1}}{(n_1-1)!} \sum_{n_2=1}^{M_1} \frac{n_1^{n_2-1}}{(n_2-1)!} \cdots \sum_{n_p=1}^{M_{p-1}} \frac{n_{p-1}^{n_p-1}}{(n_p-1)!} n_p^{-1},
$$

with p arbitrary. But the summations in the right member are equivalent to the sum over all p and nonzero p-part partitions of M and the summand is that of the lemma, thus proving the lemma.

[3] This short proof is partly due to J. S. Frame.

The lemma and (2), upon changing the index of summation to $M = N - m$, gives immediately the principal theorem:

THEOREM. *The probability that a random mapping on N points is indecomposable is*

$$[(N - 1)!/N^N] \sum_{M=0}^{N-1} N^M/M!.$$

3. Hollow mapping. One realization of near-random mapping occurs in sociometric testing. When, for example, N individuals in a group are each asked to indicate which one of the others is his best source of information, the result is such a mapping except that, if no individual is permitted to name himself, the mapping is "hollow" in the sense that the matrix representation of the graph has diagonal elements identically vanishing. If, otherwise, selection is random, the probability of equation (1) is modified for this case by replacing each N in the denominator by (N-1) and taking the outer summation of equation (2) from $m = 2$ to $m = N$. With these adjustments, we have the following corollary.

COROLLARY. *The probability that a hollow random mapping on N points is indecomposable is*

$$[(N - 1)!/(N - 1)^N] \sum_{M=0}^{N-2} N^M/M!.$$

4. Computation and asymptotic probability. The probabilities of indecomposability, of the theorem and the corollary above, might be expressed in more compact form. However, as exhibited, it is apparent that the sum is a cumulative probability of a Poisson variable with parameter N, except for a constant. Molina's tables [3] are adequate for computation of the probability through $N = 100$. Thus,

$$\text{(5)} \qquad \Pr\{\text{indecomposability}\} = \frac{(N-1)!}{N^N} e^N P(N; N - 1),$$

where $P(N;\ N - 1) = \sum_{M=0}^{N-1} e^{-N}N^M/M!$. For $N > 100$, use of the Stirling approximation for the factorial and the facts that $(1 - 1/N)^{N-1/2} = e^{-1} + O(N^{-2})$ and that $P(N; N - 1) \rightarrow \frac{1}{2}$, we obtain

$$\text{(6)} \qquad \Pr\{\text{indecomposability}\} = \left(\frac{\pi}{2N}\right)^{1/2}, \qquad N \text{ large.}$$

Similarly, using the corollary, we have

$$\text{(5h)} \qquad \Pr\{\text{indecomposability} \mid \text{hollow}\} = \frac{(N-1)!}{(N-1)^N} e^N P(N; N - 2),$$

$$\text{(6h)} \qquad \Pr\{\text{indecomposability} \mid \text{hollow}\} = e\left(\frac{\pi}{2(N-1)}\right)^{1/2}, \qquad N \text{ large.}$$

TABLE I

Probabilities of Indecomposability of a random mapping function in the general and hollow cases

N	P{I|G}	P{I|H}	N	P{I|G}	P{I|H}
2	.75000	1.00000	26	.23372	.54135
3	.62963	1.00000	28	.22562	.52574
4	.55469	.96296	30	.21831	.51148
5	.50208	.92188	32	.21169	.49837
6	.46245	.88320	34	.20564	.48628
7	.43116	.84816	36	.20009	.47507
8	.40563	.81671	38	.19497	.46463
9	.38426	.78844	40	.19023	.45488
10	.36602	.76294	45	.17976	.43308
11	.35022	.73983	50	.17086	.41426
12	.33634	.71878	55	.16318	.39780
13	.32403	.69950	60	.15646	.38322
14	.31300	.68176	65	.15052	.37020
15	.30305	.66539	70	.14521	.35847
16	.29402	.65019	75	.14043	.34782
17	.28576	.63605	80	.13610	.33810
18	.27817	.62284	85	.13215	.32918
19	.27117	.61047	90	.12853	.32096
20	.26468	.59885	95	.12519	.31334
22	.25301	.57757	100	.12210	.30626
24	.24333	.55853	Large N	$(\pi/2N)^{1/2}$	$e(\pi/2(N-1))^{1/2}$

The most interesting feature of this last result is that the probability in the hollow case remains substantially larger than in the general case as N increases. This runs counter to standard sociometric folklore, which holds that the hollow model may be uniformly replaced by the general model with small error when N is large.

Both probabilities approach zero fairly slowly (as $N^{-1/2}$). Table I presents the exact probabilities as computed from (5) and (5a).[4]

5. Notes on related work. After the present paper had been prepared, David Blackwell called the attention of the author to an unpublished memorandum by Rubin and Sitgreaves [4]. In the memorandum, different methods are used to obtain the theorem of Section 2 of this paper; the hollow mapping case is not considered.

Using methods of this paper, Jay E. Folkert and the author have obtained and will publish the probability distributions of the numbers of components of single-valued and of multiple-valued mappings in the subcases in which mapping is arbitrary or hollow. The distribution for the single-valued, arbitrary case is given also in the memorandum cited above.

[4] The author is indebted to Mr. William L. Harkness for these computations.

REFERENCES

[1] N. Metropolis and S. Ulam, "A property of randomness of an arithmetical function," *Amer. Math. Monthly*, Vol. 60 (1953), pp. 252–253.

[2] M. D. Kruskal, "The expected number of components under a random mapping function," *Amer. Math. Monthly*, Vol. 61 (1954), pp. 392–397.

[3] E. C. Molina, *Poisson's Exponential Binomial Limit*, D. Van Nostrand, 1942.

[4] H. Rubin and R. Sitgreaves, "Probability Distributions Related to Random Transformations on a Finite Set," Tech. Rep. #19A, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.