JOURNAL OF RESEARCH of the National Bureau of Standards — B. Mathematics and Mathematical Physics
Vol. 71B, No. 1, January–March 1967

Minimum Number of Subsets to Distinguish Individual Elements

P. R. Meyers

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(December 9, 1966)

Given a set S of cardinality m, we determine the minimum cardinality $f(m)$ for a family F of subsets of S such that each $s \epsilon S$ can be expressed as the intersection of some subfamily of F. The problem is solved in the following inverse form. For a given number n of subsets of S, find $g(n)$: the maximum number of elements of S which can be written as the intersection of some of these subsets. We show that $g(n)$ is the largest binomial coefficient for combinations of n things.

Key Words: Classification design, combinatorics, set theory.

1. Introduction

Let S be a finite set of given cardinality $|S| = m$. An element $s \epsilon S$ will be said to be *distinguished* by a family $\mathscr{F}$ of subsets of S, if $\{s\}$ is the intersection of some subfamily of $\mathscr{F}$. In this note we solve the following combinatorial problem (conveyed by K. E. Kloss): What is the minimum possible cardinality $f(m)$ for a family which distinguishes *all* elements of S? (Trivially $f(m) \leq m$, since

$$\mathscr{F} = \{\{s\} : s \epsilon S\}$$

distinguishes all elements.)

The question may sound like another one which arises fairly naturally in a context of classification design or information retrieval: How many "categories" (subsets) must be established so that any item (element) in a collection can be uniquely specified by listing those categories under which it falls? The categories which uniquely specify some item may be a subcollection of those which specify another item, while any family of subsets with a one element intersection cannot be part of a larger family with a different nonempty intersection.

It will be more convenient to work with the following *inverse* form of the problem: to determine $g(n)$, the maximum cardinality of a set S of elements which are distinguished by some family $\mathscr{F} = \{F_1, \ldots, F_n\}$. The inversion is made precise by removing from each F_i those elements of S which are not distinguished by $\mathscr{F}$. (For this question n is fixed, but not $\mathscr{F}$.) It will be shown below that

$$g(n) = \binom{n}{[n/2]}, \tag{1}$$

where $(\cdot)$ is the binomial coefficient and $[n/2]$ is the largest integer not greater than $n/2$. This yields an implicit solution to the original problem, since

$$f(m) = \min \{n : g(n) \geq m\}, \tag{2}$$

follows when we observe that

$$m \leq g(n) \text{ iff } f(m) \leq n. \tag{3}$$

2. Proof

Let $h(n)$ be the binomial coefficient on the right-hand side of (1). We first show that $h(n) \leq g(n)$. For this purpose, let A be a set with cardinality $|A| = n$ and let

$$S^* = \{s_1, \ldots, s_{h(n)}\} \tag{4}$$

be the collection of all subsets of A which have cardinality $[n/2]$. For $1 \leq i \leq n$, let

$$H_i = \{s \epsilon S^* : i \epsilon s\},$$

and put $\mathscr{H} = \{H_1, \ldots, H_n\}$. The possibility

$$r \epsilon \cap \{H_i : s \epsilon H_i\}, (r \epsilon S^* - \{s\}),$$

is ruled out because r cannot be a subset of s, so that some $i \epsilon A$ must satisfy $i \epsilon s - r$ and thus $s \epsilon H_i$, $r \epsilon S^* - H_i$. It follows that

$$\{s\} = \cap \{H_i : s \epsilon H_i\},$$

i.e., each element of S^* is distinguished by $\mathscr{H}$. This implies $h(n) \leq g(n)$.

The proof of (1) will be completed by showing that $g(n) \leq h(n)$. Let $\mathscr{F} = \{F_1, \ldots, F_n\}$ be a family of finite sets with union S. For each $s \epsilon S$, let

$$F^{(s)}=\{F_i : s \epsilon F_i\},\ F_s=\cap\ \{F_i : s \epsilon F_i\},\ T=\{s \epsilon S : F_s=\{s\}\}.$$

Then T consists of those elements of S which are distinguished by $\mathscr{F}$ so that $|T| \leqslant h(n)$ is what must be proved.

A collection of sets will be called *independent* if no set-inclusions hold between any pair of members.

For example, the collection $\{F^{(s)} : s \epsilon T\}$ is an independent collection of subfamilies of an n member family. Since this collection has $|T|$ members, it suffices to show that any independent family of subsets of an n element set has at most $h(n)$ members. This can be shown using the well-known SDR theorem but we find it as easy to employ an elementary argument.

For an n element set A, let S_i denote the family of subsets of A which have cardinality i, $0 \leqslant i \leqslant n$. Each S_i, and in particular $S_{[n/2]}=S^*$, is an independent family. If $\{n/2\}$ is the smallest integer not less than $n/2$, then

$$|S_{[n/2]}|=|S_{\{n/2\}}|=h(n).$$

We shall show that any other independent family P of subsets of A can be mapped $1-1$ into $S_{[n/2]}$ and thus conclude that

$$|P| \leqslant |S_{[n/2]}|. \tag{5}$$

Suppose some member of P has cardinality less than $\{n/2\}$. Let P_j be the family of members of P which have smallest cardinality, say j. Let M be the family of members of S_{j+1} which contain a member of P_j. Since P is independent, $P'=M \cup (P-P_j)$ is also independent, and $P \cap M=\phi$. We will show below that

$$j<\left\{\frac{n}{2}\right\} \text{ implies } |P_j| \leqslant |M|, \tag{6}$$

and so $|P| \leqslant |P'|$.

Then by induction on the minimum cardinality of any member of P, P', etc., we obtain an independent family Q such that $|P| \leqslant |Q|$, $Q \cap S_i=\phi$ for $0 \leqslant i<\left\{\frac{n}{2}\right\}$, and $Q \cap S_i=P \cap S_i$ for $\{n/2\}<i \leqslant n$.

The structure of the family of all subsets of A is the same relative to the relationships "is a subset of" and "is a superset of." Hence a "mirror-image" of the preceding construction will produce from Q an independent family R such that $|Q| \leqslant |R|$, $R \cap S_i=\phi$ for $n \geqslant i>[n/2]$, and such that

$$R \cap S_i=Q \cap S_i=\phi \text{ for } [n/2]>i \geqslant 0.$$

In this fashion we arrive at the result $|P| \leqslant |R|$ and $R \subset S_{[n/2]}$.

It only remains to show (6).

Let K be the number of distinct pairs (p, m) where $p \epsilon P_j$, $m \epsilon M$, and $p \subset m$. We have

$$K=(n-j)|P_j| \tag{7}$$

since any $p \epsilon P_j$ can be extended in exactly $(n-j)$ ways to an $m \epsilon M$. Also however,

$$K \leqslant (j+1)|M|, \tag{8}$$

since any $m \epsilon M$ contains $j+1$ subsets of cardinality j and thus contains at most $j+1$ members of P_j.

Where $0 \leqslant j<\{n/2\}$,

$$(j+1)/(n-j) \leqslant 1$$

and therefore combining (7) and (8) we have (6), and the proof is complete.

(Paper 71B1–193)