DETERMINATION OF A SUBSET FROM CERTAIN COMBINATORIAL PROPERTIES

DAVID G. CANTOR AND W. H. MILLS

1. Let N be a finite set of n elements. A collection $\{S_1, S_2, \ldots, S_m\}$ of subsets of N is called a *determining* collection if an arbitrary subset T of N is uniquely determined by the cardinalities of the intersections $S_i \cap T$, $1 \leqslant i \leqslant m$. The purpose of this paper is to study the minimum value $D(n)$ of m for which a determining collection of m subsets exists.

This problem can be expressed as a coin-weighing problem **(1**; **7)**.

In a recent paper Cantor **(1)** showed that $D(n) = O(n/\log\log n)$, thus proving a conjecture of N. J. Fine **(3)** that $D(n) = o(n)$. More recently Erdös and Rényi **(2)**, Söderberg and Shapiro **(7)**, Berlekamp, Mills, and Leo Moser have independently found proofs that $D(n) = O(n/\log n)$.

In the present paper we show that a determining collection of $2^k - 1$ subsets exists for $n = 2^{k-1}k$. This implies that

$$D(n) \leqslant n \log 4/\log n + O(n(\log n)^{-2} \log\log n).$$

It follows from results of Erdös and Rényi **(2)** or Leo Moser **(5**, Addendum**)** on the lower bound of $D(n)$ that the constant $\log 4$ is best possible. More precisely, using Moser's result we obtain the estimate

$$D(n) = n \log 4/\log n + O(n(\log n)^{-2} \log\log n).$$

B. Lindström **(4**; **5)** has recently proved that $D(n)$ is asymptotic to $n \log 4/\log n$, which is a consequence of this estimate. His proof runs parallel to ours, but is quite independent. He gives a construction of a determining collection of $2^k - 1$ subsets for $n = 2^{k-1}k$ that is different from ours.

The authors would like to thank J. L. Selfridge for helpful conversations.

2. We now take N to be the set of the first n positive integers. Suppose $\epsilon_j = 0$ or 1 $(1 \leqslant j \leqslant n)$. Then a collection $\{S_1, S_2, \ldots, S_m\}$ of subsets of N is a determining collection if and only if the sums

$$g_i = \sum_{j \in S_i} \epsilon_j, \qquad 1 \leqslant i \leqslant m,$$

determine the ϵ_j uniquely. If

$$e_{ij} = \begin{cases} 0 & \text{if } j \notin S_i, \\ 1 & \text{if } j \in S_i, \end{cases}$$

Received March 26, 1964. Presented to the American Mathematical Society, November 29, 1963.

https://doi.org/10.4153/CJM-1966-007-2 Published online by Cambridge University Press

then

$$g_i = \sum_{j=1}^{n} e_{ij}\,\epsilon_j, \qquad 1 \leqslant i \leqslant m. \tag{1}$$

It follows that $D(n)$ is the minimum value of m such that there exists an m by n matrix (e_{ij}) of zeros and ones with the property that the sums (1) determine the ϵ_j uniquely.

It is convenient to weaken the condition that the unknowns ϵ_j and the matrix elements e_{ij} be zeros or ones. For $m \leqslant n$ we consider the m by n matrices (e_{ij}) with the property that if $x_1, x_2, \ldots, x_n$ are integers with $x_u = 0$ or 1 for $u > m$, then the sums

$$h_i = \sum_{j=1}^{n} e_{ij}\,x_j, \qquad 1 \leqslant i \leqslant m, \tag{2}$$

determine the x_j uniquely. Such matrices clearly exist because the n by n identity matrix is one. Let $D_0(n)$ denote the minimum value of m for which there exists such a matrix (e_{ij}) consisting entirely of zeros and ones. Let $D_1(n)$ denote the minimum value of m for which there exists such a matrix (e_{ij}) consisting entirely of zeros, ones, and minus ones. Clearly

$$D(n) \leqslant D_0(n) \leqslant n. \tag{3}$$

We know that $D(n)$ and $D_0(n)$ are equal for very small values of n, and we shall show that they are asymptotic for large n, but we have been unable to determine whether or not they are equal for all values of n.

3. Lower bounds for $D_0(n)$ and $D_1(n)$. Our lower bounds for $D_0(n)$ and $D_1(n)$ depend on the following lemma:

LEMMA 1. *Let m and t be positive integers. Let X be the additive group of all m-dimensional column vectors with integer elements, let Y be a finite set of t-dimensional column vectors with integer elements, and let c be the cardinality of Y. Suppose that A is an m by m matrix of integers, and that B is an m by t matrix of integers. If, for $x \in X$ and $y \in Y$, the column vector $Ax + By$ determines x and y uniquely, then $|\det A| \geqslant c$.*

Proof. Let G be the subgroup of X generated by the columns of A. Thus G is the set of all vectors Ax with $x \in X$. By hypothesis the column vectors of the form $Ax + By$, with $x \in X$ and $y \in Y$, are all distinct. Therefore as y ranges over the c elements of Y, By ranges over c distinct cosets of G in X. Hence the index $X : G$ of G in X is at least c. On the other hand $X : G$ is equal to the absolute value of the determinant of A. Thus

$$|\det A| = X : G \geqslant c,$$

and the proof is complete.

LEMMA 2. *If* $m = D_0(n)$, *then*

$$4^n \leqslant (m+1)^{(m+1)};$$

and if $m = D_1(n)$, *then*

$$4^n \leqslant (4m)^m.$$

Proof. Suppose that $x_1, x_2, \ldots, x_n$ are integers with $x_u = 0$ or 1 for $u > m$. Let (e_{ij}) be an m by n matrix such that the sums (2) determine the x_j uniquely. We apply Lemma 1 with A the matrix consisting of the first m columns of (e_{ij}), B the matrix consisting of the remaining $n - m$ columns of (e_{ij}), and Y the set of all $(n - m)$-dimensional vectors of zeros and ones. Then Y contains exactly 2^{n-m} elements. Hence

$$|\det A| \geqslant 2^{n-m}.$$

Suppose first that (e_{ij}) is a matrix of zeros and ones. It is well known **(6)** that the determinant of an m by m matrix of zeros and ones is at most $2^{-m}(m+1)^{(m+1)/2}$. Hence

$$2^{-m}(m+1)^{(m+1)/2} \geqslant |\det A| \geqslant 2^{n-m}.$$

Therefore if $m = D_0(n)$, then

$$(m+1)^{m+1} \geqslant 2^{2n} = 4^n.$$

Now suppose that (e_{ij}) is a matrix of zeros, ones, and minus ones. Since the determinant of an m by m matrix of zeros, ones, and minus ones is at most $m^{m/2}$, we have

$$m^{m/2} \geqslant |\det A| \geqslant 2^{n-m}.$$

Therefore, if $m = D_1(n)$, then

$$2^{2m}m^m \geqslant 2^{2n},$$

which completes the proof.

4. Explicit constructions.

LEMMA 3. *Let* k *be a non-negative integer,* $r = 2^k$, *and* $s = 2^{k-1}(k+2)$. *Then there exists an* r *by* s *matrix* $B = (b_{ij})$, *of zeros, ones, and minus ones, such that*

(i) *the bottom row of* B *contains only zeros and ones, and*

(ii) *if* $x_1, x_2, \ldots, x_s$ *are integers with* $x_u = 0$ *or* 1 *for* $u > r$, *then the sums*

$$\lambda_i = \sum_{j=1}^{s} b_{ij}x_j, \qquad 1 \leqslant i \leqslant r,$$

determine the x_j *uniquely.*

Proof by induction on k. For $k = 0$ we take B to be the 1 by 1 identity matrix (1). This matrix satisfies conditions (i) and (ii). Now suppose that

$B = B_k$ is a 2^k by $2^{k-1}(k+2)$ matrix of zeros, ones, and minus ones satisfying conditions (i) and (ii) for a given value of k. Set

$$B' = \begin{pmatrix} B & -B & I \\ B & B & O \end{pmatrix},$$

where O and I are the r by r zero and identity matrices respectively. Then B' is a 2^{k+1} by $2^k(k+3)$ matrix of zeros, ones, and minus ones, and the bottom row of B' contains only zeros and ones. Let

$$x_1, x_2, \ldots, x_s; \qquad y_1, y_2, \ldots, y_s; \qquad z_1, z_2, \ldots, z_r$$

be integers with $x_u = 0$ or 1 and $y_u = 0$ or 1, for $u > r$, and $z_j = 0$ or 1 for all j. The new values of λ_i corresponding to the matrix B' are the $2r$ sums

$$\lambda_i' = \sum_{j=1}^{s} b_{ij} x_j - \sum_{j=1}^{s} b_{ij} y_j + z_i, \qquad 1 \leqslant i \leqslant r,$$

and

$$\lambda_i'' = \sum_{j=1}^{s} b_{ij} x_j + \sum_{j=1}^{s} b_{ij} y_j, \qquad 1 \leqslant i \leqslant r.$$

We have

$$\lambda_i' + \lambda_i'' \equiv z_i \pmod{2}.$$

Hence λ_i' and λ_i'' determine z_i uniquely. Since

$$\lambda_i' + \lambda_i'' = 2\sum_{j=1}^{s} b_{ij} x_j + z_i, \qquad 1 \leqslant i \leqslant r,$$

it now follows from the induction hypothesis that the λ_i' and the λ_i'' determine the x_j uniquely. Finally, since

$$\lambda_i'' - \lambda_i' = 2\sum_{j=1}^{s} b_{ij} y_j - z_i, \qquad 1 \leqslant i \leqslant r,$$

it follows that the λ_i' and the λ_i'' also determine the y_j uniquely. Therefore, by a suitable permutation of the columns of B', we obtain a matrix B_{k+1} of the correct dimensions satisfying conditions (i) and (ii). This completes the proof.

COROLLARY. *If k is a non-negative integer, then*

$$D_1(2^{k-1}(k+2)) = 2^k.$$

Proof. Lemma 2 implies that $D_1(2^{k-1}(k+2)) \geqslant 2^k$. On the other hand, the matrix B of Lemma 3 is a 2^k by $2^{k-1}(k+2)$ matrix of zeros, ones, and minus ones with the appropriate properties. Therefore $D_1(2^{k-1}(k+2)) \leqslant 2^k$, which establishes the corollary.

THEOREM 1. *If k is a positive integer, then*

$$D_0(2^{k-1}k) = 2^k - 1.$$

Proof. It follows from Lemma 2 that $D_0(2^{k-1}k) \geqslant 2^k - 1$. We set $n = 2^{k-1}k$ and $m = 2^k - 1$. To complete the proof it is sufficient to show that there exists an m by n matrix $E = (e_{ij})$ of zeros and ones such that if $x_1, x_2, \ldots, x_n$ are integers with $x_u = 0$ or 1 for $u > m$, then the sums

$$h_i = \sum_{j=1}^{n} e_{ij} x_j, \qquad 1 \leqslant i \leqslant m,$$

determine the x_j uniquely. We proceed by induction on k. For $k = 1$ we take E to be the 1 by 1 identity matrix (1). We now suppose that $E = (e_{ij})$ is a matrix with the desired properties for a given value of k. Let $A = (a_{ij})$ be the $m + 1$ by n matrix of zeros and ones obtained by adding a row of zeros to the bottom of E. Thus

$$a_{ij} = \begin{cases} e_{ij} & \text{if } 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n, \\ 0 & \text{if } i = m + 1,\ 1 \leqslant j \leqslant n. \end{cases}$$

The matrix B of Lemma 3 can be written in the form $B = V - W$, where $V = (v_{ij})$ and $W = (w_{ij})$ are r by s matrices of zeros and ones,

$$r = 2^k = m + 1, \qquad s = 2^{k-1}(k + 2),$$

and the bottom row of W is identically zero. We set

$$A' = \begin{pmatrix} A & V \\ A & W \end{pmatrix}.$$

We note that A' is a 2^{k+1} by $2^k(k + 1)$ matrix of zeros and ones and that the bottom row of A' is identically zero. Let $x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_s$ be integers with $x_u = 0$ or 1 for $u > m$ and $y_u = 0$ or 1 for $u > r$. The sums h_i corresponding to the matrix A' are

$$h_i' = \sum_{j=1}^{n} a_{ij} x_j + \sum_{j=1}^{s} v_{ij} y_j, \qquad 1 \leqslant i \leqslant r,$$

and

$$h_i'' = \sum_{j=1}^{n} a_{ij} x_j + \sum_{j=1}^{s} w_{ij} y_j, \qquad 1 \leqslant i \leqslant r.$$

It follows from condition (ii) of Lemma 3 that the differences $h_i' - h_i''$ $(1 \leqslant i \leqslant r)$ determine the y_j uniquely. Hence, by the induction hypothesis, the h_i' and the h_i'' determine both the x_j and the y_j. Moreover, since $h_r'' = 0$, it follows that the x_j and the y_j are uniquely determined by the sums h_i' $(1 \leqslant i \leqslant r)$ and h_i'' $(1 \leqslant i \leqslant r - 1)$. Hence by permuting the columns of A' and removing the bottom row of zeros we obtain a matrix with the desired properties. This completes the proof of the theorem.

Theorem 1 enables us to obtain the following upper bound for $D_0(n)$:

THEOREM 2.

$$D_0(n) \leqslant \frac{n \log 4}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right).$$

Proof. We write $\operatorname{Log} x$ for $\log_2 x$. We assume that n is large enough so that $\operatorname{Log} n > 3 \operatorname{Log} \operatorname{Log} n$, and we set

$$k = [\operatorname{Log} n - 3 \operatorname{Log} \operatorname{Log} n] + 1.$$

We write $n = 2^{k-1}kQ + R$, where Q and R are integers and $0 \leqslant R < 2^{k-1}k$. We set $D_0(0) = 0$. It follows at once from the definition of $D_0(n)$ that $D_0(s + t) \leqslant D_0(s) + D_0(t)$ for all non-negative integers s and t. Hence

$$D_0(n) \leqslant QD_0(2^{k-1}k) + D_0(R) \leqslant (2^k - 1)Q + R.$$

Now

$$R < 2^{k-1}k \leqslant 2^{\operatorname{Log} n - 3 \operatorname{Log} \operatorname{Log} n}k = kn \operatorname{Log}^{-3} n < n \operatorname{Log}^{-2} n.$$

Moreover,

$$\begin{aligned}(2^k - 1)Q < 2^k Q \leqslant 2n/k &< 2n/(\operatorname{Log} n - 3 \operatorname{Log} \operatorname{Log} n) \\ &= \frac{n \log 4}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right).\end{aligned}$$

Combining the above inequalities, we obtain the desired result:

$$D_0(n) \leqslant (2^k - 1)Q + R \leqslant \frac{n \log 4}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right).$$

5. Asymptotic estimates for $D_0(n)$ and $D(n)$. Leo Moser has shown that

(4) $$D(n) \geqslant n \log 4/\log n + O(n \log^{-2} n).$$

Moser's proof can be found in a generalized form in Lindström's paper (**5**, pp. 488f.). We have already seen that $D(n) \leqslant D_0(n)$. Combining this with (4) and Theorem 2, we obtain asymptotic estimates for both $D(n)$ and $D_0(n)$:

THEOREM 3.

$$D_0(n) = \frac{n \log 4}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right)$$

and

$$D(n) = \frac{n \log 4}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right).$$

We note that the asymptotic estimate for $D_0(n)$ can be deduced directly from Theorem 2 and Lemma 2 without using Moser's result. Furthermore, from Lemma 2 and the corollary to Lemma 3 we can deduce the same asymptotic estimate for $D_1(n)$.

6. Modifications. In the original problem, one can use, instead of the intersections $S_i \cap T$, the unions $S_i \cup T$, the differences $S_i - T$, the differences $T - S_i$, or the symmetric difference $(S_i \cup T) - (S_i \cap T)$. However,

https://doi.org/10.4153/CJM-1966-007-2 Published online by Cambridge University Press

since the S_i are known sets, once the cardinality of T is known, the cardinality of $S_i \cap T$ can be deduced from the cardinality of any of these other sets and conversely. Hence replacing intersection by one of these other expressions changes the value of $D(n)$ by at most one and so preserves the asymptotic estimate.

References

1. David G. Cantor, *Determining a set from the cardinalities of its intersections with other sets*, Can. J. Math., *16* (1964), 94–97.
2. P. Erdös and A. Rényi, *On two problems of information theory*, Publ. Hung. Acad. Sci., *8* (1963), 241–254.
3. N. J. Fine, *Solution E1399*, Amer. Math. Monthly, *67* (1960), 697–698.
4. B. Lindström, *On a combinatory detection problem*, Publ. Hung. Acad. Sci., *9* (1964), 195–207.
5. ——— *On a combinatorial problem in number theory*, Can. Math. Bull., *4* (1965), 477–490.
6. H. J. Ryser, *Maximal determinants in combinatorial investigations*, Can. J. Math., *8* (1956), 245–249.
7. Staffan Söderberg and H. S. Shapiro, *A combinatory detection problem*, Amer. Math. Monthly, *70* (1963), 1066–1070.

Institute for Defense Analyses,
Princeton, New Jersey, and
University of Washington