DETERMINING A SET FROM THE CARDINALITIES OF ITS INTERSECTIONS WITH OTHER SETS

DAVID G. CANTOR

Let n be a positive integer and put $N = \{1, 2, \ldots, n\}$. A collection $\{S_1, S_2, \ldots, S_t\}$ of subsets of N is called *determining* if, for any $T \subset N$, the cardinalities of the t intersections $T \cap S_j$ determine T uniquely. Let $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ be n variables with range $\{0, 1\}$. It is clear that a determining collection $\{S_j\}$ has the property that the sums

$$\sum_{i \in S_j} \epsilon_i$$

determine the ϵ_i uniquely. We are interested in finding, as a function of n, the least integer $f(n)$ such that there exists a determining collection containing $f(n)$ subsets of N. This can be interpreted as a "coin-weighing" problem: given n coins known to weigh either α or β $(\alpha \neq \beta)$, $f(n)$ is the least number of weighings necessary, on a calibrated scale, to determine the weight of each of the n coins (one can always normalize so that $\alpha = 0$ and $\beta = 1$).

It is clear that the sets $\{1\}, \{2\}, \ldots, \{n\}$ form a determining collection, hence that $f(n) \leqslant n$. The purpose of this paper is to show that $f(n) = O(n/\log \log n)$, thus proving a conjecture of N. J. Fine **(1)**. The author would like to thank J. L. Selfridge for suggesting the problem, and for many helpful discussions. The case $n = 5$ comes from **(2)**.

Since there is no additional difficulty we allow the ϵ_i to have range $\{0, 1, 2, \ldots, k - 1\}$, where $k \geqslant 2$ is an integer fixed for the remainder of this paper. Then $f(n)$ is the least number of sets $S_j \subset N$ such that the sums

$$\sum_{i \in S_j} \epsilon_i$$

determine the ϵ_i uniquely.

We consider a more general problem where some of the variables range through the real numbers. A variable whose range is $\{0, 1, 2, \ldots, k - 1\}$ is called *restricted*; otherwise it is *unrestricted.*

Let $\epsilon_1, \epsilon_2, \ldots, \epsilon_r$ be restricted variables and $\sigma_1, \sigma_2, \ldots, \sigma_s$ unrestricted variables. By a *method* (r, s, t) we mean a collection of t subsets of the $r + s$ variables ϵ_i, σ_j, such that the t sums, obtained by summing the variables in each of the t subsets, determine the ϵ_i and σ_j uniquely. The existence of a method (r, s, t) means that there are t linear forms, with coefficients 0 or 1, in the $r + s$ variables ϵ_i, σ_j, such that the values of the linear forms determine the values of the ϵ_i and σ_j uniquely, hence that $f(r + s) \leqslant t$.

Received December 19, 1962.

https://doi.org/10.4153/CJM-1964-009-4 Published online by Cambridge University Press

Put $R(r, s, t) = (r + s)/t$. In Lemma 3, we define a "multiplication" of methods, which when applied to the methods constructed in Lemma 2 makes $R(r, s, t)$ arbitrarily large.

LEMMA 1. *Suppose there exists a method* (r, s, t). *Then, if* $p \geqslant 1$, $r' \leqslant r$, $s' \leqslant s$, $t' \geqslant t$, *there exist methods* (r', s', t'), $(r + s', s - s', t)$, *and* (pr, ps, pt). *For each* $a > r + s$, *there exists a method* $(a, 0, c)$ *with* $a/c \geqslant (r + s)/2t$.

Proof. The first part is obvious. Now set $a = q(r + s) - g$, where $0 \leqslant g < r + s$. By the first part there exists a method $(q(r + s), 0, qt)$; hence there exists a method $(a, 0, qt)$. But $a/qt \geqslant (q - 1)(r + s)/qt \geqslant (r + s)/2t$.

LEMMA 2. *For each* $m \geqslant 0$, *there exists a method* $(m + 1, k^m, k^m + 1)$.

Proof. Let $\epsilon_0, \epsilon_i, \ldots, \epsilon_m$ be restricted variables and $\sigma_1, \sigma_2, \ldots, \sigma_{k^m}$ be unrestricted variables. Put

$$L_0 = \sum_{i=1}^{k^m} \sigma_i$$

and

$$L_j = \sigma_j + \sum_{i=h}^{m} \epsilon_i,$$

where $1 \leqslant j \leqslant k^m$, and h is obtained from j by $k^{h-1} < j \leqslant k^h$. Then ϵ_i appears in those forms L_j for which $1 \leqslant j \leqslant k^i$. Hence,

$$-L_0 + \sum_{j=1}^{k^m} L_j = \sum_{i=0}^{m} \epsilon_i k^i.$$

By the uniqueness of the expansion of a number to the base k, this determines the ϵ_i uniquely, and then as

$$\sigma_j = L_j - \sum_{i=h}^{m} \epsilon_i,$$

the σ_j are determined.

Put

$$(r, s, t) * (u, v, w) = (rv + tu, sv, tw).$$

Under the map

$$(r, s, t) \to \begin{pmatrix} s & 0 \\ r & t \end{pmatrix},$$

the operation $*$ corresponds to matrix multiplication, hence is associative.

LEMMA 3. *Suppose there exist methods* (r, s, t) *and* (u, v, w). *Then there exists a method* $(r, s, t) * (u, v, w)$.

https://doi.org/10.4153/CJM-1964-009-4 Published online by Cambridge University Press

Proof. Let ϵ_{ij}, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant v$, be rv restricted variables; let δ_{mj}, $1 \leqslant m \leqslant t$, $1 \leqslant j \leqslant u$, be tu restricted variables; finally, let σ_{ij}, $1 \leqslant i \leqslant s$, $1 \leqslant j \leqslant v$, be sv unrestricted variables. For each fixed j, $1 \leqslant j \leqslant v$, the existence of method (r, s, t) implies that there exist t linear forms L_{mj}, $1 \leqslant m \leqslant t$, with coefficients 0 or 1, in the ϵ_{ij} and σ_{ij}, whose values determine the ϵ_{ij} and σ_{ij} uniquely. For each fixed m, $1 \leqslant m \leqslant t$, we apply method (u, v, w) to the L_{mj} and the δ_{mj}, treating the L_{mj} as the v unrestricted variables of the method (u, v, w). Thus there exist w linear forms K_n, $1 \leqslant n \leqslant w$, with coefficients 0 or 1 in the L_{mj} and the δ_{mj}, whose values determine the values of the δ_{mj} and L_{mj}, hence the values of the ϵ_{ij} and σ_{ij}. For different j, the L_{mj} are linear forms in distinct variables ϵ_{ij} and σ_{ij}. Hence the tw linear forms

$$J_{mn}(\epsilon_{ij}, \sigma_{ij}, \delta_{ij}) = K_n(L_{mj}(\epsilon_{ij}, \sigma_{ij}), \delta_{mj})$$

have coefficients 0 or 1, and determine the $rv + tu$ restricted variables ϵ_{ij}, δ_{mj}, and the sv unrestricted variables σ_{ij}.

Put $(r, s, t)^1 = (r, s, t)$, and inductively

$$(r, s, t)^{n+1} = (r, s, t)^n * (r, s, t).$$

LEMMA 4. *If* $0 < c - b < a$, *then*

$$R[(a, b, c)^m] = \frac{a}{c-b}\left[1 - \left(\frac{b}{c}\right)^m\right] + \left(\frac{b}{c}\right)^m.$$

Proof. An easy induction.

THEOREM. $f(n) = O(n/\log\log n)$. *More precisely,*

$$\limsup_{n\to\infty}(f(n)\log\log n/n) \leqslant 2\log k/(1 - 1/e).$$

Proof. By Lemmas 1, 2, and 3, there exists a method

$$(a_m, b_m, c_m) = (m, k^m, k^m + 1)^{k^m}.$$

Clearly,

$$c_m = (k^m + 1)^{k^m} = k^{mk^m}(1 + 1/k^m)^{k^m} \sim ek^{mk^m}.$$

By Lemma 4,

$$R(a_m, b_m, c_m) = m\left[1 - \left(1 - \frac{1}{k^m + 1}\right)^{k^m}\right] + \left(1 - \frac{1}{k^m+1}\right)^{k^m}$$
$$\sim m(1 - 1/e)$$

Put $d_m = a_m + b_m$; then $d_m \sim m(1 - 1/e)c_m$, and $\log\log d_m \sim m \log k$. By Lemma 1, there exist methods $(d_m, 0, c_m)$; hence

$$f(d_m) \leqslant c_m \sim d_m/m(1 - 1/e).$$

Thus

$$\limsup_{n\to\infty}\left[f(d_m)\,\frac{\log\log d_m}{d_m}\right] \leqslant \frac{\log k}{1 - 1/e}.$$

By Lemma 1.

$$f(n) \leqslant 2nf(d_m)/d_m, \qquad \text{where } d_m \leqslant n < d_{m+1}.$$

Hence

$$\limsup_{n\to\infty} f(n) \log\log n/n \leqslant 2 \log k/(1 - 1/e).$$

References

1. N. J. Fine, *Solution E1399*, Amer. Math. Monthly, *67* (1960), 697.
2. H. S. Shapiro, *Problem E1399*, Amer. Math. Monthly, *67* (1960), 82.

University of Washington,
Seattle, Washington