*This research was supported in part by an NSF Grant No. GP19568.

ON RANDOM SEARCH USING BINARY SYSTEMS DERIVED FROM FINITE PROJECTIVE PLANES*

May 1973

I.M. Chakravarti

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 869

ON RANDOM SEARCH USING BINARY SYSTEMS DERIVED FROM FINITE PROJECTIVE PLANES

by

I. M. Chakravarti
Department of Statistics
University of North Carolina at Chapel Hill

SUMMARY

If the points of a finite projective plane with $s+1$ points on a line are identified with the elements of the basic set S_n of search and the lines with the functions of a system F of search, then the incidence matrix of the plane defines a separating system of search on F. Bounds on probabilities of termination of search processes when functions are chosen by random sampling with replacement from the line set have been worked out using combinatorial properties of the point-line incidence in a finite projective plane, relevant to search. This extends the results on $PG(2,2)$ and $PG(2,3)$ search systems earlier obtained by Chakravarti and Manglik [1].

1. INTRODUCTION

1.1 *Rényi Model for Search.* In ([4], [5]), Rényi described a model for search as follows. Let S_n be a finite set of n distinguishable elements $a_1, a_2, \ldots, a_n$. It is required to determine the identity of an unknown element x belonging to S_n or equivalently, to separate it from the rest of the elements in S_n. It is assumed that it is not possible to observe x directly but one can choose a sequence of functions $f_1, f_2, \ldots, f_m$ from a given family or system F of functions and observe the values of these functions at x in order to determine the identity of x. We shall consider F to be a finite family of M functions, $F = \{f_1, f_2, \ldots, f_M\}$ and $S_n = \{a_1, a_2, \ldots, a_n\}$. If F contains a function which takes on different values for different elements of S_n, a single observation of this function at x will reveal its identity. In practice, the number of different values taken on by a member of F is much smaller than n. Here, we shall consider the special case where each function takes on only two values 0 and 1. Such a family of functions will be called a *binary search system.*

Problems of sorting, questionnaire, information retrieval, decoding a received message, search for a lost or hidden object, search for a mistake in a computer program, search for locating a failure in a complicated mechanism (a car, a computer, an airplane etc.) can be treated under this model of search.

1.2 *Separating Systems.* A system F of functions defined on the set S_n is a *separating system* if to every pair of distinct elements a_i, a_j, $a_i \neq a_j$ of S_n, there exists in F a function f such that $f(a_i) \neq f(a_j)$. If F is not a separating system, there will exist at least two distinct elements a_i and a_j in S_n, such that $f(a_i) = f(a_j)$ for all f in F.

On the other hand, if F is separating, then there exists at least one sequence of functions, namely the one consisting in observing the values of all the functions in F at x, which will identify all the elements of S_n.

A separating system F can be characterized in an alternative manner. Let

$$N = ((f_i(a_j))), \; i = 1, 2, \ldots, M, \; j = 1, 2, \ldots, n$$

denote the $M \times n$ matrix whose (i,j)-entry is $f_i(a_j)$. Then F is a separating system if and only if all the columns of the matrix N are distinct.

1.3 *Notions of Homogeneity.* A system F of functions defined on the set S_n, is called *completely homogeneous*, if for every f in F, the function g defined by $g(a_j) = f(a_{\pi(j)})$ belongs to F as well, where π is any permutation of the integers $\{1,2,\ldots,n\}$.

A system F of functions defined on the set S_n, is called *weakly homogeneous of order* k $(2 \le k \le n)$ if for every k-plet $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$ of distinct elements from S_n, the number R_k of functions f in F such that $f(a_{j_1}) = \ldots = f(a_{j_k})$, does not depend on the choice of the k-plet $a_{j_1}, \ldots, a_{j_k}$.

A system F is defined to be a *strongly homogeneous system of order* k, if for every k-plet of distinct elements $(a_{j_1}, \ldots, a_{j_k})$ of S_n and a sequence of k elements $y_{\ell_1}, \ldots, Y_{\ell_k}$ (not necessarily all distinct), the number $R_k(y_1, \ldots, y_{\ell_k})$ of functions f in F for which

$$f(a_{j_1}) = y_{\ell_1}, \ldots, f(a_{j_k}) = y_{\ell_k},$$

does not depend on the choice of the k-plet $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$, but it

may depend on the sequence $(y_{\ell_1},\ldots,y_{\ell_k})$.

The following properties hold for the different types of homogeneity defined above.

(a) If a system F is weakly homogeneous of order 2 and $R_2 < R_1$ (which is always true except when F consists of constant functions only), then it is also a separating system.

(b) A completely homogeneous system containing at least one constant function is a separating system.

(c) Strong homogeneity of order k implies weak homogeneity of all orders $\ell \leq k$.

(d) Strong homogeneity of order k imples strong homogeneity of every order $\ell < k$. However, for a weakly homogeneous system, the analogous property does not hold.

(e) If a binary system F is weakly homogeneous of all orders $\ell \leq 2k$, then it is also weakly homogeneous of order $2k+1$ $(k=1,2,\ldots)$. In particular, a binary system F which is weakly homogeneous of order 2 $(k=1)$ is also weakly homogeneous of order 3.

Many weakly and strongly homogeneous binary systems of search have been constructed by Manglik [2], from finite geometrical structures.

2. RANDOM SEARCH

2.1 *Random Search Using a Separating System.* A method for the successive choice of the functions $f_1,\ldots, f_m$ from a given family F, which leads in the end to the determination of the unknown x will be called a strategy of search. A strategy may be systematic or random. Let F be a binary system of R_1 functions, which is a separating system for the elements of S_n. A sequence of functions is selected from F by random sampling with replacement. Let $f_1, f_2,\ldots, f_m$ be

the sequence which happens to be chosen. Consider the matrix N having m rows and n columns, in which the k-th element of the j-th row is $f_j(a_k)$. Then a fixed unknown element x is separated from the rest of the elements of S_n by the functions $f_1, f_2, \ldots, f_m$ if and only if the x-th column of N, $(f_1(x), \ldots, f_m(x))^T$ is different from every other column of N.

On the other hand, the functions $f_1, f_2, \ldots, f_m$ separate each element of S_n from the rest if all the columns of N are distinct. The probability that a sequence of m functions chosen by random sampling with replacement from the system F, will separate a given unknown element x from the rest of the elements of S_n is denoted by $P_1(n,m,F,x)$.

The probability that a sequence of m functions chosen by random sampling with replacement from the system F, will separate all the elements of S_n from one another is denoted by $P_2(n,m,F)$.

Rényi [4] has shown that if F is a weakly homogeneous system of order 2,

$$P_1(n,m,F,x) \geq 1 - (n-1)\left(\frac{R_2}{R_1}\right)^m , \tag{2.1}$$

for all x in S_n and,

$$P_2(n,m,F) \geq 1 - \binom{n}{2}\left(\frac{R_2}{R_1}\right)^m . \tag{2.2}$$

Further, for a binary system F which is weakly homogeneous of order 2 and hence of order 3,

$$P_1(n,m,F,x) \leq 1 - (n-1)\left(\frac{R_2}{R_1}\right)^m + \binom{n-1}{2}\left(\frac{R_3}{R_1}\right)^m . \tag{2.3}$$

3. SEARCH SYSTEMS FROM FINITE PROJECTIVE PLANES

3.1 *Probabilities of Termination of Search Processes.* If the points of a finite projective plane $\Pi(s)$ with $s+1$ points on a line, are identified with the elements of S_n and the lines with the functions of F then the incidence matrix of the plane defines a strongly homogeneous system of order 2 (and hence a weakly homogeneous system of order 3) and thus a separating system on the line set $L(s)$ of the plane. We note that for the above correspondence, we have

$$(3.1) \qquad n = s^2 + s+1,\ M = s^2 + s+1 = R_1$$
$$R_2 = 1 + s(s-1) = s^2 - s + 1,\ R_3 = \tfrac{1}{2}(3R_2 - R_1) = (s-1)^2.$$

Hence using (2.1), (2.2) and (2.3) we have

$$(3.2) \qquad P_1(s^2 + s+1, m, L(s)x) \geq 1 - (s^2 + s)\left(\frac{s^2 - s+1}{s^2 + s+1}\right)^m .$$

$$(3.3) \qquad P_2(s^2 + s+1, m, L(s)) \geq 1 - \binom{s^2+s+1}{2}\left(\frac{s^2 - s+1}{s^2 + s+1}\right)^m .$$

The exact values of $P_1(s^2 + s+1, m, L(s), x)$ and $P_2(s^2 + s+1, m, L(s))$ for $s = 2,3$ obtained by Chakravarti and Manglik [1], are

$$(3.5) \qquad P_1(5, m, L(2), x) = 1 - 6\left(\frac{3}{7}\right)^m + \frac{11}{7^m} .$$

$$(3.6) \qquad P_2(5, m, L(2)) = 1 - 3\left(\frac{3}{7}\right)^{m-1} + 2\left(\frac{1}{7}\right)^{m-1} .$$

$$(3.7) \qquad P_1(13, m, L(3), x) = 13^m - 12.7^m + 66.4^m - 68.3^m - 12.2^m + 35)/13^m .$$

$$(3.8) \qquad P_2(13, m, L(3)) = (13^{m-1} - 6.7^m + 27.5^m + 20.4^m - 162.3^m + 216.2^m - 120)/13^{m-1} .$$

3.2 *Some Combinatorial Properties of Incidence in a Finite Projective Plane, Relevant to the Termination of Search Processes.*

In order to calculate the probabilities (3.5) through (3.8), it was necessary to enumerate all incidence patterns of points and lines in $\Pi(2)$ and $\Pi(3)$, leading to the termination of the search processes. The enumeration of all such configurations in $\Pi(s)$ where $s > 3$ is quite complicated and hence, it is difficult to get the exact values of the probabilities for termination of the search process for higher values of s. However, the following combinatorial properties (relevant to the termiantion of search) of finite projective planes can be used to improve considerably the bounds given in (2.1), (2.2) and (2.3).

Let $L(x;s)$ denote the set of lines in $\Pi(s)$, which are incident with a given point x, $L(\bar{x},s) = L(s) - L(x;s)$ the set of lines not incident with x, $L(\bar{x},y;s)$ the set of lines incident with y but not x.

Suppose that a set of r distinct lines have been selected in m steps of random sampling with replacement from $F = L(s)$.

A. For determining the identity of a given unknown element x we distinguish the following mutually exclusive cases.

(i) At least two distinct lines from $L(x;s)$ have been selected. In that case, the element x will be distinguished from the rest of the elements.

(ii) Exactly one line ℓ is selected from the set $L(x;s)$. In that case, the element x will be recognized only if the other points on ℓ are incident with at least one of the remaining lines selected. Hence, $s+1$ distinct lines of which one, ℓ, is from $L(x;s)$ and s others from $L(\bar{x};s)$ incident one each with the s points (other than x) on ℓ are needed.

(iii) None of the lines selected is incident with x. In that case, the lines will detect x only if each one of the other $s^2 + s$ points be incident with at least one of the selected lines. We prove the following.

Lemma 3.1 There exists a set of $(2s-1)$ lines, none incident with x, which covers all the $s(s+1)$ points other than x. Further, any set of $(s^2 - s+1)$ lines in $L(\bar{x};s)$ covers all the $s(s+1)$ points other than x. Proof: Let y be a point other than x and let ℓ be the line incident with y and x. Consider the set of $(2s-1)$ lines, which consists of the s lines (other than ℓ) incident with y and $(s-1)$ lines (distinct from ℓ) incident one each with the $(s-1)$ points, other than x and y, on ℓ. The s lines incident with y will cover s^2+1 points and the remaining $(s-1)$ points are covered by the other $(s-1)$ lines properly chosen.

The second part of the lemma is proved by noticing that there are exactly (s^2-s) lines in $L(\bar{x};s)$ which are incident neither with x nor with y where y is a given point other than x.

Thus, in m random drawings with replacement from $L(s)$ we consider three events: a) E_1 that the only distinct line selected is from $L(x;s)$; b) E_2 that exactly one distinct line ℓ from $L(x;s)$ and (s^2-s) or less lines from $L(\bar{x};s)$ where the latter lines do not cover all the s points (other than x) on ℓ; and c) E_3 that no line from $L(x;s)$ and s^2-s or less distinct lines from $L(\bar{x};s)$ are selected. Then we conclude that

$$P_1(s^2 + s+1,m,L(s),x) \geq 1 - P(E_1) - P(E_2) - P(E_3) \tag{3.9}$$

where

$$P(E_1) = (s+1)(1/M)^m$$

$$P(E_2) = (s+1) \sum_{t=1}^{s^2-s} \left[\left\{ \binom{s}{1}\binom{s^2-s}{t} - \binom{s}{2}\binom{s^2-2s}{t} + \ldots + (-1)^{s-2}\binom{s^2-s(s-1)}{t} \right\} \cdot \Delta^{t+1}0^m/M^m \right]$$

$$P(E_3) = \sum_{t=1}^{s^2-s} \binom{s^2}{t} \Delta^t 0^m/M^m, \quad \text{where } \Delta^t 0^m/M^m = \sum_{v=0}^{t} (-1)^v \binom{t}{v} (t-v)^m/M^m$$

is the probability that in m random drawings with replacement from M distinct units, exactly t given units appear. See Feller [3].

B. Consider the case of separating all the elements of S_n (which are the points of $\Pi(s)$) by lines selected by random sampling with replacement from $L(s)$. We prove the following:

Lemma 3.2 Any set of (s^2-s+2) distinct lines will separate all the points of the plane $\Pi(s)$.

Proof: Given a pair of points x and y in the plane, there is exactly one line incident with both x and y and there are exactly (s^2-s) lines incident neither with x nor with y. Hence, R_2, that is the number of lines (functions) such that $\ell(x) = \ell(y)$ is equal to $s^2 - s+1$. Hence any set of $s^2 \cdot s+2$ lines will separate all the points of $\Pi(s)$.

Thus if in m random drawings with replacement from $L(s)$, we have $(s^2 - s+2)$ distinct lines or more, the search will terminate with the separation of all the points of the plane. Hence,

$$P_2(s^2 + s+1), m, L(s)) \geq \sum_{r=R}^{M} \binom{M}{r} \Delta^r 0^m/M^m, \tag{3.10}$$

$$R = s^2 = s+2, \; M = s^2 + s+1.$$

REFERENCES

[1] I. M. Chakravarti and V. P. Manglik (1972). "On random search using binary systems derived from the incidence matrices of PG(2,2) and PG(2.3)." *Journal of Cybernetics, 2,* 2, 26-47.

[2] V. P. Manglik (1972). "On some strongly and weakly homogeneous binary search systems." *Journal of Cybernetics, 2,* 2, 61-77.

[3] W. Feller (1968). *An Introduction to Probability Theory and its Applications. Vol. 1,* John Wiley and Sons, New York.

[4] A. Rényi (1965). "On the theory of random search." *Bull. Amer. Math. Soc.,* 71, 809-828.

[5] A. Rényi (1969). "Lectures on the theory of search." *Institute of Statistics Mimeo Series No. 600.7, University of North Carolina at Chapel Hill, N.C. 27514.*