STATISTICAL PROGRAMMING[1]

BY D. F. VOTAW, JR.

Yale University[2]

1. Introduction. A "statistical programming" problem is encountered when the information about one or more constants in a programming problem is statistical. We shall first give examples of programming problems and then point out how certain statistical analogues of them arise. The results given in this paper pertain to these analogues.

Our first example is a transportation problem. Given a unit amount of a homogeneous product (e.g., oil) at each of n origins and required that a unit amount be received at each of n destinations, and given the cost, say c_{ij}, of shipping a unit amount from the ith origin to the jth destination $(i, j = 1, \cdots, n; n \geqq 2)$, find a most economical schedule of shipments of the product from origins to destinations. More specifically, find an $n \times n$ matrix (x_{ij}) of real numbers for which

$$\sum_{i,j=1}^{n} c_{ij} x_{ij} \tag{1}$$

assumes its *minimum* value, where

$$\begin{aligned} \sum_{i=1}^{n} x_{ij} &= 1, \qquad (j = 1, \cdots, n), \\ \sum_{j=1}^{n} x_{ij} &= 1, \qquad (i = 1, \cdots, n), \\ x_{ij} &\geqq 0. \end{aligned} \tag{2}$$

x_{ij} represents the amount shipped from the ith origin to the jth destination; and the matrix (x_{ij}) is called a "program." The expression in (1) is the total shipping cost. The condition (2) expresses the facts that at each origin the sum of all amounts shipped away must equal 1 and that at each destination the sum of all amounts received from the origins must equal 1. The problem stated above is a special case of the Hitchcock-Koopmans transportation problem, which is a well-known special case of a linear programming problem (see [1], [2, Part 1]).

The next example is the personnel assignment problem, which is closely related to the first example (see [4], [5, pp. 255–258], and [8]). Let us replace "origins" by "persons," "destinations" by "jobs," and regard c_{ij} as the productivity of the ith person if placed on the jth job. It is required that each person be

Received June 9, 1959; revised May 23, 1960.

[1] Research work sponsored in part by the Office of Naval Research. This paper includes some results in an unpublished paper of the same title presented at the Montreal meeting of the Institute of Mathematical Statistics, September, 1954 (see [6]).

[2] Now at MITRE Corporation.

placed *full time* on a job and that each job be filled: thus to the linear constraints in (2) on (x_{ij}) there must be added the following constraint, which is not linear:

$$\text{each } x_{ij} = 0 \text{ or } 1. \tag{3}$$

In view of (3) the admissible program matrices (x_{ij}) are simply the $n!$ permutation matrices of order $n \times n$. The problem is to find a permutation $(j_1', \cdots, j_n')$ such that

$$c_{1j_1'} + \cdots + c_{nj_n'} = \max_{(j_1, \cdots, j_n)} (c_{1j_1} + \cdots + c_{nj_n}) \tag{4}$$

where $(j_1, \cdots, j_n)$ denotes any permutation of $(1, \cdots, n)$. Having determined $(j_1', \cdots, j_n')$, we would assign person 1 to job j_1', $\cdots$, person n to job j_n' and thereby obtain maximum average productivity of the group of n persons relative to the jobs. Incidentally, when (3) does not hold, we may regard x_{ij} as the fraction of the ith person's time allocated to the jth job.

A sum of the form $c_{1j_1} + \cdots + c_{nj_n}$ will be termed a "permutation sum of c_{ij}'s." From [4, Lemma 2] we have that each program in the first example is a weighted mean of the $n \times n$ permutation matrices and that the sum in (1) is a weighted mean of permutation sums of c_{ij}'s. This implies that there is a permutation matrix (x_{ij}) for which the quantity in (1) assumes its minimum value. A similar result holds relative to the maximum value. Clearly, the transportation problem described in (1) and (2) and the assignment problem are nearly identical mathematically. For each problem the optimum sum is assumed when (x_{ij}) is a permutation matrix. The differences between the two are: (i) in one we seek a maximum sum and in the other a minimum sum; (ii) the integer constraint (3) is part of the assignment problem but not part of the transportation problem. It is noteworthy that both problems are related to a certain 0-sum, 2-person game (see [4], pp. 7–11).

The c_{ij}'s are constants in the two examples stated above, and it is presupposed that the numerical value of each is known to the "programmer." When the values are in fact not all known to the programmer, statistical information regarding them may nevertheless be available (e.g., in the form of aptitude indexes of personnel (see [7]) or in the form of records of unit costs of past shipments). Such situations lead to the problems treated in this paper. As regards the game discussed in [4] it should be noted that a "pseudogame" arises when the c_{ij}'s are not all known to both players. For a discussion of pseudo-games see [3, p. 357].

We shall set up a statistical analogue of the personnel assignment problem. By simply replacing "maximization" by "minimization," we can transform the analogue into an analogue of the transportation problem. Both these analogues are related to pseudo-games based on the game in [4].

A generalization of the assignment problem and of the statistical analogue will be considered in Section 4.

2. Statistical Programming. Consider an n^2-dimensional Euclidean sample space, W^*, and represent each point of W^* by $(w_{11}, w_{12}, \cdots, w_{nn})$. Let $H(w_{11},$

$w_{12}, \cdots, w_{nn}$) be the distribution function associated with W^* and let (W_{11}, $W_{12}, \cdots, W_{nn}$) be a random n^2-dimensional vector whose distribution is H.

We assume that the c_{ij}'s are known to be parameters of H; however, the *numerical values* of the c_{ij}'s are assumed to be *unknown* to the assigner. We also assume that an observed value, say $(w'_{11}, w'_{12}, \cdots, w'_{nn})$, of $(W_{11}, W_{12}, \cdots, W_{nn})$ can be obtained by him. In general the observed value supplies him with statistical information regarding $c_{11}, c_{12}, \cdots, c_{nn}$. Having obtained $(w'_{11}, w'_{12}, \cdots, w'_{nn})$ the assigner selects a permutation of $(1, \cdots, n)$.

With reference to the assignment problem we define statistical programming as partitioning the sample space W^* into $n!$ mutually exclusive and exhaustive subsets and establishing a one-to-one correspondence between the subsets and the $n!$ permutations of $(1, \cdots, n)$.[3] It is understood that when the observation $(w'_{11}, w'_{12}, \cdots, w'_{nn})$ is obtained one selects the permutation $j'_1, \cdots, j'_n$ corresponding to the subset, say $P_{j'_1,\dots,j'_n}$, in which $(w'_{11}, w'_{12}, \cdots, w'_{nn})$ lies. In advance of obtaining an observed value of $(W_{11}, W_{12}, \cdots, W_{nn})$ the permutation to be selected is a random variable, say $(J_1, \cdots, J_n)$, and so the permutation sum $c_{1J_1} + \cdots + c_{nJ_n}$ is a random variable. We can regard the distribution of this random sum as a "performance function" characterizing the statistical programming with which it is associated.

A natural kind of statistical programming is that in which one considers estimates of the $n!$ permutation sums and selects the permutation corresponding to the largest estimate. This will be termed "programming by estimation." A permutation sum of w'_{ij}'s would often be a suitable point estimate of the corresponding permutation sum of c_{ij}'s. This leads to a set of $n!$ regions $P_{j'_1,\dots,j'_n}$ such that any given $P_{j'_1,\dots,j'_n}$ would contain the set of points in W^* such that

$$(5) \qquad w_{1j'_1} + \cdots + w_{nj'_n} > \max_{(j_1,\cdots,j_n)\neq(j'_1,\cdots,j'_n)} (w_{1j_1} + \cdots + w_{nj_n}).$$

Incidentally, formula (5) does not assign points of W^* having two or more equal permutation sums. We shall assume that $H(w_{11}, w_{12}, \cdots, w_{nn})$ is continuous; consequently, such points can be assigned to the subsets arbitrarily.

3. Comparison of Purely Random Programming and Programming by Estimation. Let $(J_1^*, \cdots, J_n^*)$ be a purely random permutation (i.e., for any preassigned permutation $(j_1, \cdots, j_n)$ the probability that $(J_1^*, \cdots, J_n^*) = (j_1, \cdots, j_n)$ is $1/n!$). Let

$$(6) \qquad S = c_{1J_1^*} + \cdots + c_{nJ_n^*},$$

and let

$$(7) \qquad F(s) = \Pr(S \le s) \qquad (-\infty < s < \infty).$$

The possible values of S are the values of permutation sums of the c_{ij}'s; hence $F(s)$ is a purely discrete distribution having not more than $n!$ saltuses. Let d

[3] This definition can be generalized to provide a definition of statistical programming that is associated with general linear programming. In this paper, however, the generalization will not be carried out.

represent the number of distinct possible values of S (thus $1 \leqq d \leqq n!$). When $d = 1$, let s_1 be the one value. When $d > 1$, let the values be represented by $s_1, \cdots, s_d$ in increasing order of magnitude. The selection of a value of $(J_1^*, \cdots, J_n^*)$ will be termed "purely random programming."

For each (i, j) let

$$Y_{ij} = W_{ij} - c_{ij}, \tag{8}$$

and represent the sample space of Y_{ij}'s by Y^*. Let $(y_{11}, y_{12}, \cdots, y_{nn})$ be any point of Y^* and let the distribution function of $(Y_{11}, Y_{12}, \cdots, Y_{nn})$ be $K(y_{11}, y_{12}, \cdots, y_{nn})$. We shall assume that K is completely symmetric in its variables and continuous. These assumptions imply that (for $d > 1$)

$$F(s) = \Pr[(Y_{11}, Y_{12}, \cdots, Y_{nn}) \varepsilon R_s], \qquad (s_1 \leqq s \leqq s_d) \tag{9}$$

where R_s is the set of points in Y^* such that

$$\max_{\substack{(j_{1,q}, \cdots, j_{n,q})\\ q \leqq s}} (y_{1j_{1,q}} + \cdots + y_{nj_{n,q}}) > \max_{\substack{(j_{1,q'}, \cdots, j_{n,q'})\\ q' > s}} (y_{1j_{1,q'}} + \cdots + y_{nj_{n,q'}}) \tag{10}$$

where $(j_{1,q}, \cdots, j_{n,q})$ represents a permutation of $(1, \cdots, n)$ such that

$$c_{1j_{1,q}} + \cdots + c_{nj_{n,q}} = q (q = s_1, s_2, \cdots, s_d). \tag{11}$$

Let $(J_1', \cdots, J_n')$ be a permutation to be selected under programming by estimation with an observed value of $(W_{11}, W_{12}, \cdots, W_{nn})$ as in (5). Let

$$Z = c_{1J_1'} + \cdots + c_{nJ_n'} \tag{12}$$

and let $G(s)$ be the cumulative distribution function of Z, thus

$$G(s) = \Pr(Z \leqq s), \qquad (-\infty < s < +\infty). \tag{13}$$

$G(s)$ has the same saltus points as $F(s)$ — namely, $s = s_1, \cdots, s_d$. It follows from (12) and (13) that (for $d > 1$)

$$G(s) = \Pr\{(W_{11}, W_{12}, \cdots, W_{nn}) \varepsilon R_s'\}, \qquad (-\infty < s < +\infty) \tag{14}$$

where R_s' is a region in W^* such that

$$\max_{\substack{(j_{1,q}, \cdots, j_{n,q})\\ q \leqq s}} (w_{1j_{1,q}} + \cdots + w_{nj_{n,q}}) > \max_{\substack{(j_{1,q'}, \cdots, j_{n,q'})\\ q' > s}} (w_{1j_{1,q'}} + \cdots + w_{nj_{n,q'}}). \tag{15}$$

Clearly we have that

$$G(s) = \Pr\{(Y_{11}, Y_{12}, \cdots, Y_{nn}) \varepsilon R_s''\}, \tag{16}$$

where R_s'' is the region in Y^* such that

$$\max_{\substack{(j_{1,q}, \cdots, j_{n,q})\\ q \leqq s}} (y_{1j_{1,q}} + \cdots + y_{nj_{n,q}} + q) > \max_{\substack{(j_{1,q'}, \cdots, j_{n,q'})\\ q' > s}} (y_{1j_{1,q'}} + \cdots + y_{nj_{n,q'}} + q'). \tag{17}$$

$G(s)$ is the performance function for programming by estimation and $F(s)$ is the performance function for purely random programming. Under certain conditions $G(s)$ and $F(s)$ can be compared by means of the theorem below.

THEOREM: *If* $K(y_{11}, y_{12}, \cdots, y_{nn})$ *is completely symmetric in its variables and continuous, then for every* s *in* $(s_1 \leqq s \leqq s_d)$

$$G(s) \leqq F(s). \tag{18}$$

If, furthermore, every non-degenerate interval in Y^* *contains positive probability, and if* $d > 1$, *then for every* s *in the interval* $(s_1 \leqq s < s_d)$

$$G(s) < F(s). \tag{19}$$

PROOF: When $d = 1$, formula (18) is obviously correct. When $d > 1$, formula (18) is correct when $s = s_d$, since both G and F equal 1. When $d > 1$ and s is any point of $(s_1 \leqq s < s_d)$, we have from (9), (10), (16), and (17) that R_s'' is a subset of R_s since for each s the quantity $q' - q$ is non-negative. This completes the proof of (18). To prove (19) we shall show that for every s in $(s_1 \leqq s < s_d)$ there is an n^2-dimensional interval in $R_s - R_s''$. Let $t = \min\,(s_{\sigma+1} - s_\sigma)$ $(\sigma = 1, \cdots, d-1)$, and let r be in the interval $(0 < r < t/(n+1))$. For any given s in $(s_1 \leqq s < s_d)$ and for some $q \leqq s$ consider the following n^2-dimensional interval in Y^*

$$\begin{aligned} t/n - r/n < y_{ij_{i,q}} < t/n \qquad & (q \leqq s)(i = 1, \cdots, n), \\ 0 < y_{ij} < r/n \qquad & (i, j = 1, \cdots, n;\ j \neq j_{i,q}). \end{aligned} \tag{20}$$

It can be shown easily that throughout this interval the maximum permutation sum of y_{ij}'s is $y_{1j_{1,q}} + \cdots + y_{nj_{n,q}}$. Thus the interval lies in R_s (see (10)). It will now be shown that the interval does not lie in R_s''. Note that for each point of the interval

$$y_{1j_{1,s_d}} + \cdots + y_{nj_{n,s_d}} + s_d > s_d$$

but

$$\max_{\substack{(j_{1,q}, \cdots, j_{n,q}) \\ q \leqq s}} (y_{1j_{1,q}} + \cdots + y_{nj_{n,q}} + q) < t + (s_d - t) = s_d; \qquad (s_1 \leqq s < s_d)$$

hence no point of the interval lies in R_s'' (see (17)). We have thus shown that the interval lies in $R_s - R_s''$. It follows that the probability content of R_s exceeds that of R_s''; hence for each s in $(s_1 \leqq s < s_d)$ we have that $G(s) < F(s)$.

Let $E(S)$ and $E(Z)$ denote the expected value of S and of Z, respectively. It can be shown easily that

$$E(S) = \sum_{i,j} c_{ij}/n. \tag{21}$$

It should be noted that when (18) holds, $E(Z) \geqq E(S)$, and that when (19) holds, $E(Z) > E(S)$. An interesting feature of the *statistical analogue of the transportation problem* is that we can be *certain* of attaining the sum given in (21). This can be accomplished by setting every x_{ij} equal to $1/n$.

When the W_{ij}'s are mutually independent and each $(W_{ij} - c_{ij})$ has the same continuous distribution, $K(y_{11}, y_{12}, \cdots, y_{nn})$ satisfies the conditions that imply (18). When $W_{11}, W_{12}, \cdots, W_{nn}$ are normal and independent with means $c_{11}, c_{12}, \cdots, c_{nn}$, respectively, and a common variance, $K(y_{11}, Y_{12}, \cdots y_{nn})$ satisfies the conditions that imply (19). It should be noted that the theorem does not require that the W_{ij}'s be independent.

When (18) holds, programming by estimation is uniformly at least as good as purely random programming; when (19) holds, programming by estimation is uniformly better.

It is easy to find situations in which random programming is better than programming by estimation. For example, let W_{ij} have a negative exponential distribution, $(1/c_{ij})e^{-w_{ij}/c_{ij}}$, which has mean c_{ij}, and suppose that $n = 2$ and that W_{11}, W_{12}, W_{21}, and W_{22} are mutually independent. When $c_{11} = 20$, $c_{12} = c_{21} = 10$, and $c_{22} = 1$, we find that with programming by estimation the expected sum of c_{ij}'s is 20.467, approximately. This is less than 20.5, which is the expected sum of c_{ij}'s when random programming is used. We can also find situations in which the hypothesis of the above theorem is not fulfilled but the conclusion holds. An example of this arises when the four W_{ij}'s described above are associated with the following c_{ij}'s: $c_{11} = 9$, $c_{12} = 7$, $c_{21} = 8$, $c_{22} = 5$. Here the expected sum of c_{ij}'s is 14.531, approximately, when programming by estimation is used. This exceeds the expected sum, 14.5, when random programming is used.

4. A Statistical Analogue of the Generalized Optimum Assignment Problem. Consider a B-dimensional array ($B \geqq 2$) having n "layers" in each dimension and let $c_{i_1 i_2 \cdots i_B}$ be the element in "cell" $(i_1, i_2, \cdots, i_B)(i_b = 1, \cdots, n; b = 1, \cdots, B)$. As pointed out in [4, p. 11], this situation could be of interest when each of n jobs requires a team of $B - 1$ persons and $c_{i_1 i_2 \cdots i_B}$ represents, say, the productivity of the team consisting of persons $i_1, \cdots, i_{B-1}$ on job i_B. Other interpretations can be made easily. The problem here is to find an "assignment set" for which the "assignment sum" of c's equals its maximum value. This is the generalized optimum assignment problem. When $B = 2$, the problem is the personnel assignment problem stated in section 1.

By straightforward generalization of our statistical analogue of the personnel assignment problem we obtain a statistical analogue of the generalized assignment problem. The theorem stated in Section 3 generalizes to this statistical analogue.

With regard to the generalized assignment problem, it can be shown that under purely random assignment the expected total production equals

$$\sum_{i_1, i_2, \cdots, i_B} c_{i_1 i_2 \cdots i_B}/n^{B-1}. \tag{22}$$

REFERENCES

[1] G. B. Dantzig, "Application of the simplex method to a transportation problem" (chapter XXIII in [2]).

[2] Tjalling C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission Monograph No. 13, John Wiley and Sons, New York, 1951.

[3] J. C. C. McKinsey, *Introduction to the Theory of Games*, McGraw-Hill Book Company, New York, 1952.

[4] John von Neumann, "A certain zero-sum two-person game equivalent to the optimal assignment problem" (in *Contributions to the Theory of Games*, Vol. II edited by H. W. Kuhn and A. W. Tucker, Princeton University Press, 1953, pp. 5–12).

[5] D. F. Votaw, Jr., "Methods of solving some personnel classification problems," *Psychometrika*, Vol. 17 (1952), pp. 255–266.

[6] D. F. Votaw, Jr., "Statistical programming," (abstract), *Ann. Math. Stat.*, Vol. 25 (1954), p. 809.

[7] David F. Votaw, Jr., "Mathematical programming and personnel assignment," in *The RAND Symposium on Mathematical Programming*, R-351, The RAND Corporation, Santa Monica, California, 1960, pp. 119–120.

[8] D. F. Votaw, Jr., and A. Orden, "The personnel assignment problem" in *Symposium on Linear Inequalities and Programming* (Planning Research Division, Comptroller, Headquarters U. S. Air Force, Washington 25, D. C. 1 April 1952, pp. 155–163.