PROBABILISTIC PROGRAMMING

S. VAJDA

PROBABILISTIC PROGRAMMING

S. VAJDA

THE UNIVERSITY OF BIRMINGHAM
BIRMINGHAM, ENGLAND

1972

ACADEMIC PRESS New York and London

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DD

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 72-180796

AMS(MOS) 1970 Subject Classifications: 62-02, 62C25, 90-02, 90A15, 90B30, 90C15

PRINTED IN THE UNITED STATES OF AMERICA

Introduction

Mathematical programming (i.e. planning) is an application of techniques to the planning of industrial, administrative, or economic activities. Analytically, it consists of the optimization (maximization or minimization) of a function of variables (the "objective function") which describe the levels of activities (production of a commodity, distribution of facilities, etc.) and which are subject to constraints (e.g. restrictions in the availability of raw material, or on the capacities of communication channels). As a rule, the variables are also restricted to taking only nonnegative values.

The formulation of applied problems will incorporate "technological" coefficients (prices of goods, cost and capacity of production, etc.) on which a model of the situation to be analyzed can be based. In the classical situation these coefficients are assumed to be completely known. But if one wants to be more realistic, then this assumption must be relaxed. Tintner (1941) distinguishes between subjective risk, when "there exists a probability distribution of anticipation which is itself known with certainty," and subjective uncertainty, when "there is an a priori probability of the probability distributions themselves."

The former field leads to stochastic, or probabilistic, programming. We assume, then, that the (joint) distribution of the technological coefficients is given. This includes the case when they are independently distributed, and cases when some of them have given values, without stochastic deviations.

A number of possible attitudes to deal with such a situation have been proposed. For instance, we might be ready to wait until the actual values of the coefficients and constants become known—for instance the requirement for some commodities during the next selling period—but we find that we must choose now in which type of activity we shall invest funds which we have available.

Under these circumstances we shall remind ourselves that if the coefficients are random variables, then the best result which we can obtain, after their values have become known and are taken into account to find the most favorable activities, is also a random variable.

Assume, then, that various investment possibilities have been offered to us, all of them incurring some risk, but that we are confident that we shall be able to find the best procedure in all possible emerging situations. Which criteria shall we then apply in choosing among the offers?

This is a problem for economists not for mathematicians. Economists tell us that we might choose, perhaps, the investment which offers the largest expected value of the objective function, or the largest probability that the objective function will reach, at least, some critical value, or indeed some other criterion.

It is then the job of the mathematician to compute the relevant "preference functional," in most cases by first computing the distribution function of the optimum of the objective function.

This attitude, called the "wait-and-see" approach by Madansky (1960), is that which was originally called "stochastic programming" by Tintner (1955). These are not decision problems in the sense that a decision has to be made "here-and-now" about the activity levels. We wait until an observation is made on the random elements, and then solve the (deterministic) problem. Chapter I deals with such wait-and-see problems.

Chapter II contains the analysis of decision problems, in particular of so-called two-stage problems, which have been extensively studied. In these problems a decision concerning activity levels is made at once, in such a way that any emerging deviations from what would have been best, had one only known what values the stochastic elements were going to have, are in some way evaluated and affect the objective function.

In Chapter III we turn to "chance constraints," i.e. constraints which are not expected to be always satisfied, but only in a proportion of cases, or "with given probabilities." These are decision problems which reduce to problems treated in Chapter II, if the probabilities are equal to unity.

The reader should be familiar with the concepts and procedures of linear programming, and in particular with the simplex method of solving linear programs, a brief survey of which is given in Appendix I. Some concepts of nonlinear programming are also used sporadically, and for these the reader is referred to standard textbooks. All those concepts for which no special reference is given can be found, for instance, in Vajda (1961) and, more specifically, in Vajda (1967).

Features of importance for computation are mentioned, though alogorithms are not discussed in detail.

An extensive reference list is included, and in Appendix II a list is given of applications described in the literature. However, no completeness is claimed in this respect, or indeed in any other, since the subject is still being vigorously pursued by many workers.

The preparatory work, including a first draft, was carried out while the author held a Senior Research Fellowship, awarded by the Science Research Council (of Great Britain), in the Department of Mathematical Statistics of the University of Birmingham. The manuscript was completed while he was a David Brewster Cobb Senior Fellow in the Department of Transportation and Environmental Planning, in the same University.

Contents

II. Decision Problems

III. Chance Constraints

Appendix I

Appendix II

I

Stochastic Programming

Parameters

Consider the problem of minimizing $c'x$, subject to $Ax \geqslant b$, $x \geqslant 0$, where c is a n-vector, b is a m-vector, and A is an m by n matrix, while x is the unknown n-vector to be determined. Let $mn+m+n=M$.

(If not mentioned otherwise, vectors are column vectors. The transpose of vectors and matrices will be indicated by a prime.)

The components can be written, parametrically, as

$$c_j = c_{j0} + c_{j1} t_{(1)} + \cdots + c_{jr} t_{(r)} \qquad (j = 1, \ldots, n)$$

$$b_i = b_{i0} + b_{i1} t_{(1)} + \cdots + b_{ir} t_{(r)} \qquad (i = 1, \ldots, m)$$

$$a_{ij} = a_{ij0} + a_{ij1} t_{(1)} + \cdots + a_{ijr} t_{(r)} \qquad (i,j \text{ as above})$$

where $r \leqslant M$, and the parameters $t_{(k)}$ $(k = 1, \ldots, r)$ have a joint probability distribution. Hence results of the theory of parametric programming can be adapted to purposes of the study of the stochastic case.

We shall assume that the support T, say, of the parameters, i.e. the smallest closed set of values $t_{(k)}$ with probability measure unity, is convex and bounded in $t = (t_{(1)}, \ldots, t_{(r)})$.

Feasibility and Convexity

We shall be interested in the relationship between the feasibility of vectors x and the parameters. It will then be possible to obtain some insight into the distribution of the minimum of the objective function $c'x$. Such knowledge will also be relevant to decision problems, since we would not, for instance, choose a value x which has no or merely a small chance of being feasible when the actual value of t emerges.

When we express all constants as linear functions of the parameters, then a linear objective function, and linear constraints, will be linear in the $t_{(k)}$, and also in the components x_j of x, separately. It follows that $T(x)$, the set of all those t for which a given $x \geqslant 0$ is feasible, is convex, and it will be polyhedral;† if T is polyhedral. The set of those t for which every $x \geqslant 0$ is feasible is the intersection of an infinity of sets $T(x)$, and hence also convex, though possibly empty.

As a simple illustration (cf. Vajda, 1970), take the case when the

† Or polytopic, but we shall continue to use the more usual term.

constraints are

$$ax \geqslant b, \qquad x \geqslant 0$$

(a case where $m = n = 1$), and a as well as b take, independently, values in the closed interval $(-1, 1)$. The (convex) sets of (a, b) for which certain chosen values of x are feasible, are shown in Figure 1.

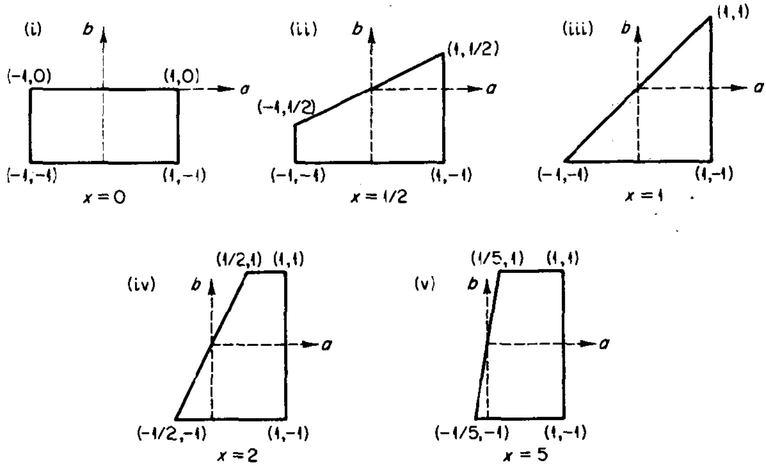

Figure 1

From the linear relationship between t and x it follows, also, that $X(t)$, the set of those $x \geqslant 0$ which are feasible for a given t is convex and polyhedral. The set of those $x \geqslant 0$ which are feasible for any t (*permanently feasible*) is the intersection of such sets and hence also convex (and possibly empty).

To take again the example above, for $a = -0.1, b = -\frac{1}{2}$, all x in [0,5] are feasible. In this example no permanently feasible x exists. If $0 < b \leqslant 1$ and $-1 \leqslant a \leqslant 0$, then $X(t) = X(a, b)$ is empty.

As far as feasibility is concerned, the objective function, and its possibly stochastic character, are irrelevant. But we must distinguish between the case when only $b' = (b_1, \ldots, b_m)$ is stochastic, and that when the elements of A (and perhaps those of b as well) are stochastic.

First, we deal with the case when only b is stochastic. Then the set of those t for which $X(t)$ is not empty is convex.

PROOF: If

$$Ax \geqslant b(t_1)$$

has a solution x_1, and

$$Ax \geqslant b(t_2)$$

has a solution x_2, then

$$Ax \geqslant b(\lambda t_1 + \mu t_2), \quad \text{with} \quad \lambda, \mu \geqslant 0, \qquad \lambda + \mu = 1$$

has also a solution, e.g. $\lambda x_1 + \mu x_2$. ■

The set of those $x \geqslant 0$ for which $T(x)$ is not empty is also convex.

PROOF: If

$$Ax_1 \geqslant b(t)$$

has a solution t_1, and

$$Ax_2 \geqslant b(t)$$

has a solution t_2, then

$$A(\lambda x_1 + \mu x_2) \geqslant b(t)$$

has also a solution, e.g. $b(\lambda t_1 + \mu t_2)$, since T is convex. ■

A theorem of Kall (1966) gives conditions for A to be such that some $x \geqslant 0$ exists whatever the values of the coefficients b, when the constraints are equations, $Ax = b$. Clearly the matrix A must then have rank m, and we assume that the first m columns, written $A_1, \ldots, A_m$, are independent. In fact, A must have more than m columns, because it is impossible that some b_0, and also $-b_0$ be represented as nonnegative linear combinations of the same m independent columns. Hence $k = n - m \geqslant 1$.

Kall's Theorem

For $Ax = b$ to have a nonnegative solution x for all b it is necessary and sufficient that there exist

$$\mu_j \geqslant 0 \qquad \text{and} \qquad \lambda_j < 0$$

such that

$$\sum_{j=m+1}^{m+k} \mu_j A_j = \sum_{j=1}^{m} \lambda_j A_j \tag{A}$$

where A_j is the jth column of the matrix A.

This condition is necessary: Let $b = \sum_{j=1}^{m} \beta_j A_j$. Such β_j exist, and for some b they will all be negative. Then, $Ax = \sum_{j=1}^{m+k} x_j A_j$ equals $\sum_{j=1}^{m} \beta_j A_j = b$, we have

$$\sum_{j=1}^{m} (\beta_j - x_j) A_j = \sum_{j=m+1}^{m+k} x_j A_j$$

and since $x_j \geqslant 0$ and $\beta_j < 0$, we can identify μ_j with x_j, and λ_j with $\beta_j - x_j$.

The condition is also sufficient: Let again

$$b = \sum_{j=1}^{m} \beta_j A_j.$$

The set of β_j is unique for a given b. If they are all nonnegative, then $Ax = b$ has a solution $x_j = \beta_j \geqslant 0$ for $j = 1, \ldots, m$, and $x_j = 0$ for $j = m+1, \ldots, m+k$.

On the other hand, if one at least of the β_j is negative, then we must find another representation of b in terms of A_j, with only nonnegative coefficients. Let, then, (A) hold. The largest of the ratios β_j/λ_j ($j \leqslant m$) is positive, if at least one of the β_j found above is negative. We may assume, without loss of generality, that the largest of these ratios is reached at $j = m$. Because $A_1, \ldots, A_m$ are independent, this is also true of $A_1, \ldots, A_{m-1}$ and $\sum_{j=1}^{m} \lambda_j A_j = A_0$, say. Thus the representation

$$b = \sum_{j=0}^{m-1} \gamma_j A_j$$

is also unique. We show, next, that all γ_j are nonnegative.

We have

$$b = \gamma_0 \sum_{j=1}^{m} \lambda_j A_j + \sum_{j=1}^{m-1} \gamma_j A_j$$

while we have, also $b = \sum_{j=1}^{m} \beta_j A_j$. Because both representations of b are unique, we must have

$$\beta_j = \gamma_j + \gamma_0 \lambda_j \qquad \text{for} \quad j = 1, \ldots, m-1$$

and

$$\beta_m = \gamma_0 \lambda_m.$$

Hence

$$\gamma_0 = \frac{\beta_m}{\lambda_m} > 0$$

and

$$\gamma_j = \beta_j - \lambda_j \gamma_0 = \lambda_j \left[\frac{\beta_j}{\lambda_j} - \frac{\beta_m}{\lambda_m} \right] \geqslant 0 \qquad \text{for} \quad j = 1, \ldots, m-1.$$

To complete the argument, we remember that also $A_0 = \sum_{j=m+1}^{m+k} \mu_j A_j$ and that the μ_j are $\geqslant 0$. Thus

$$b = \sum_{j=1}^{m-1} \gamma_j A_j + \sum_{j=m+1}^{m+k} \gamma_0 \mu_j A_j$$

and in this representation all coefficients are nonnegative.

For instance, if $A = [I, -I]$, where I is the identity matrix of order m, then $\mu_j = 1$, $\lambda_j = -1$ for all j for such a set of coefficients, and indeed the set

$$x_{i1} - x_{i2} = b_i \qquad (i = 1, \ldots, m)$$

has a nonnegative solution for any value of b_i.

We turn now to the case where A is stochastic. Simple examples show that then the sets of t, and x, respectively, for which $X(t)$ and $T(x)$ are not empty, are not always convex. For instance,

$$(t-3)x_1 + (1-t)x_2 \geqslant 1$$

cannot be satisfied, with nonnegative x_1 and x_2, if $1 \leqslant t \leqslant 3$. The set of those values, for which $X(t)$ is not empty, i.e. the set of values of t outside this interval, is not convex.

The set of those $x \geqslant 0$ for which $T(x)$ is not empty is not convex when the constraint is

$$(-1+2t)x_1 + (2-4t)x_2 \geqslant 1, \qquad 0 \leqslant t \leqslant 1.$$

The constraint can be written

$$(-1+2t)(x_1 - 2x_2) \geqslant 1$$

and because the first factor is in the interval $[-1, 1]$, the second must be $\geqslant 1$, or $\leqslant -1$. The set of such x_1 and x_2 is not convex.

Optimality and Convexity

We shall now consider optimality rather than feasibility. In this case we must also take into account whether or not c is stochastic. We assume again that $c'x$ is to be minimized.

The set $X^\circ(t)$ of those $x \geqslant 0$ which are optimal for a given t is known to be convex and polyhedral from elementary linear programming theory. Of course, it could be empty, and we start by studying the set of those t for which it is not empty.

If only c is stochastic, then this set is convex.

PROOF: If

$$c(t_1)'x_1 \leqslant c(t_1)'x$$

for all feasible x, and also

$$c(t_2)'x_2 \leqslant c(t_2)'x$$

for all feasible x, then when

$$\lambda, \mu \geqslant 0 \qquad \text{and} \qquad \lambda + \mu = 1$$

$$c(\lambda t_1 + \mu t_2)'x = \lambda c(t_1)'x + \mu c(t_2)'x$$

is bounded from below, for all feasible x, by

$$\lambda c(t_1)'x_1 + \mu c(t_2)'x_2.$$

Therefore a finite minimum exists for $t = \lambda t_1 + \mu t_2$ as well. ∎

The convexity of the set of those t for which an optimum exists, when only c is stochastic, was also proved by Simons (1962), who called this set *admissible.*

If only b is stochastic, then the set of those t which admit a finite minimum is the same as the set of those t which admit a finite maximum for the dual program, and is therefore also convex, by the above argument.

Now let c as well as b be stochastic. Then the following argument† can be used to show that the set of those t for which an optimal x exists is still convex.

Let t_1 lead to a finite minimum of $c(t_1)'x$, subject to $Ax \geqslant b(t_1)$, $x \geqslant 0$. Then there is also a finite maximum to the objective function $b(t_1)'y$, subject to $A'y \leqslant c(t_1)$, $y \geqslant 0$.

Let Q be the set of those t for which $b(t)'y$, subject to $A'y \leqslant c(t_1)$, $y \geqslant 0$, has a finite maximum. It is also the set of those t which make $Ax \geqslant b(t)$ consistent, and is thus independent of t_1.

Let P be the set of those t which make the minimum of $c(t)'x$, subject to $Ax \geqslant d$, finite. It is independent of d, because it is the set of those t for which $A'y \leqslant c(t)$ is consistent. In particular, $c(t)'x$ has a finite minimum if $d = b(t_1)$ with t_1 in Q.

The set we are looking for is the intersection of P and Q, which is convex.

If A is stochastic, then the set of those t for which $X^\circ(t)$ is not empty is not necessarily convex.

For instance, if we have

$$(t-3)x_1 + (1-t)x_2 \geqslant 1; \qquad x_1, x_2 \geqslant 0$$

then the set $X(t)$ is not empty only when $t < 1$, or $t > 3$, as we have seen. In these regions $X^\circ(t)$ is not empty either when, for instance, the objective function to be minimized is $x_1 + x_2$. Its minimum is obtained for $x_1 = 0$, $x_2 = 1/(1-t)$ when $t < 1$, and for $x_1 = 1/(t-3)$, $x_2 = 0$ when $t > 3$.

We call $T^\circ(x)$ the set of those t for which a given $x \geqslant 0$ is optimal. Simons (1962) calls this the region of validity. He deals with the case when only c is stochastic. In this case $T^\circ(x)$ is convex for any x.

† From the Ph.D. Thesis in the University of Birmingham of A. S. Gonçalves (1969).