STOCHASTIC PROGRAMMING
METHODS AND APPLICATIONS

A17

STOCHASTIC PROGRAMMING

Methods and Applications

by

Jati K. Sengupta

Iowa State University

and

Indian Institute of Management (Calcutta)

1972

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM

AMERICAN ELSEVIER PUBLISHING COMPANY, INC. - NEW YORK

ISBN North-Holland 07204 3071 2

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM
NORTH-HOLLAND PUBLISHING COMPANY, LTD. – LONDON

Sole distributors for the U.S.A. and Canada:

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017

PRINTED IN THE NETHERLANDS

PREFACE

This book presents some of the most important aspects of the current theory and methods of stochastic or probabilistic programming. The theory of stochastic programming is an active and growing field in current research and several disciplines like mathematics, statistics, operations research, and mathematical economics are involved in this field.

Emphasis in this book centers on methods and applications of stochastic programming in its various aspects, e.g., chance-constrained programming, two-stage programming under uncertainty, programming under recourse, reliability programming, and several other related aspects of stochastic programming.

Parts of the material covered in this book were developed in connection with my teaching a graduate course in operations research dealing with mathematical programming and other optimization techniques. The book includes, however, several recent results from the author's own research, which deal in particular with the following: (1) methods of reliability programming, (2) the implications of nonnormal distributions in the theory of chance-constrained and other types of stochastic programming, (3) the results of sensitivity and robustness analysis in geometric programming and other dynamic models of stochastic control, and (4) the use of computational techniques like SUMT (sequential unconstrained minimization techniques) in solving empirical problems of stochastic programming. Applied and computational aspects occupy a central place in our development of the current theory of stochastic programming. Hence the book is likely to be useful to applied research workers in various fields like applied mathematics and statistics, operations research, and mathematical economics. Whereas the technical aspects of computation and transformations in terms of nonlinear programming would be of primary interest to the operations researcher and the applied mathematician, the empirical and illustrative applications of the various methods would attract the attention of the economists and other social scientists.

The research materials contained in this book developed out of a grant from the U.S. National Science Foundation to which the author is deeply grateful. The author is also grateful to his former students who collaborated with him in joint research work in the field of stochastic programming. A special note of thanks is due to Mrs. Rita Bauman, who did an excellent job of taking care of this technical manuscript and seeing it through.

J.K.S.

TABLE OF CONTENTS

Page

CHAPTER 1

STOCHASTIC PROGRAMMING METHODS

Stochastic programming deals with the theory and methods of incorporating stochastic variations into a mathematical programming problem. For instance, in the usual linear programming (LP) model: $[\max z = c'x, Ax \leq b, x \geq 0]$ the parameters in the set (A, b, c) are given and known numbers and it is required to determine an optimal decision vector x subject to the constraints specified above. If the elements in the set (A, b, c) are stochastic, then stochastic variations are introduced into the programming problem through random variations in θ, where θ denotes the vector with elements (A, b, c). The sources of random variations may be several, depending on the type of problem and the type of decisions to be arrived at. For instance, the probability distribution of θ may be known (e.g., the random variations in weather in agricultural applications) and the problem is to choose a decision vector x which is optimal in some sense. In another case, only sample observations may be available and we have to estimate the unknown population parameters and incorporate these into the program in order to arrive at a decision vector x which is optimal in some sense.

Following Tintner (1955), we distinguish two basic types of stochastic programming, the passive and the active. The passive stochastic program arises when we follow the "wait-and-see" approach, i.e., we wait for the observations on the random vector θ and by utilizing these realized values in a suitable manner we derive either exactly or approximately the probability distribution of the maximand (i.e., the maximum value of the objective function) and of the optimal decision vector. This approach may be used to compare, for instance, two production (or resource allocation) situations through two probability distributions of profit say. The second method of stochastic programming called the active approach defines a "here-and-now" attitude about the decision vector x, i.e., in these problems decisions concerning the activity vector x is made at once without waiting for the realizations of the random vector θ and this is done by considering a set of possible decisions with their expected penalty costs due to the random variables deviating from their expectations. For instance, this approach may be used to compare two possible resource allocation patterns, e.g., we may have a production problem with x denoting different outputs and then consider the probability distribution of profits under two possible allocation of resources. The decision problem is to compare the probability distribution of optimal profits under these two resource allocations and then decide on a solution which is optimal in some sense (e.g., the expected penalty costs provide a criterion of optimality).

It is clear that in both the passive and active approaches of stochastic programming, the probability distribution of the vector $\theta = (A, b, c)$ occupies a central place. In the passive approach, it induces the probability distribution of the maximand and generates probabilistic variations around specified optimal basis, when the latter has to be appropriately defined in order to allow admissible perturbations (Sengupta, Tintner, and Millham, 1963; Tintner, Millham, and Sengupta, 1963). In the active approach, the additional restrictions on the decision-space interact with the probability distribution of the vector θ and help the decision maker to compare and evaluate alternative distributions of optimal profits. Emphasis and extension of this distribution approach are the main objectives of this book. The various probability distributions considered here include the normal, the truncated normal, the chi-square, the exponential, and the class of nonnegative distributions known as increasing hazard rate distributions. In particular, the implications of alternative distributions in terms of the nonlinearities they introduce into the transformed programming model are discussed in the case of chance-constraints in Chapter 2, where the stochastic linear constraints are interpreted according to the theory of chance-constrained programming. It is shown that in certain cases an optimal allocation of tolerance measures can be determined along with an optimal decision vector and this leads to the development of system reliability programming.

In Chapter 3 we discuss a few selected cases where stochastic variations can be incorporated into a nonlinear programming framework and yet the transformed problem would remain computable in principle by the existing algorithms like geometric programming, sequential unconstrained minimization technique (SUMT), and the methods of convex programming. Several cases of active approach of stochastic programming are discussed here, e.g., one-stage and two-stage linear programming under uncertainty, stochastic goal programming, and a nonlinear activity analysis model with a stochastic penalty cost. It is shown here that the penalty function approach plays an important role in computing optimal trajectories in a dynamic control model and the methods of simulated optimization are found useful in developing suitable approximations and suboptimal solutions.

In Chapter 4 we discuss several empirical applications of various aspects of stochastic programming. These applications are mainly illustrative of (i) various problems of reliability programming, (ii) the difficulties of developing suitable approximations, (iii) the problems of applying linear and other suboptimal decision rules in stochastic programming, and (iv) the simulation techniques for testing the stability of alternative feasible solutions in the form of feedback controls applied to a dynamic economic model characterized by the multiplier-accelerator principle.

1.1 The Passive Approaches to Stochastic Programming

In the passive approach to stochastic programming developed by Tintner (1955, 1960), it is assumed that the realizations of the random variables $\theta_k = (A_k, b_k, c_k)$, $k = 1, 2, \ldots, N$ in the LP problem [max $z_k = c_k'x$, $A_k x \leqq b_k$, $x \geqq 0$] are available, where $k = 1, 2, \ldots, N$ denotes a specific selection of the index set of observations. Now consider the parameter space spanned by the random elements of $\theta = (A, b, c)$; let V_k be the region in this parameter space where the objective function $z_k = c_k'x$ is both feasible (i.e., x is a feasible vector) and optimal and finite. Denote this optimal (i.e., maximal) value by z_k^*, for such a selection. It is clear that since some selections may not be feasible, therefore, we consider the index set $k = 1, 2, \ldots, K$ where $K \leq N$ and in this sense we are considering the probability distribution of optimal profits z^* in a truncat.d or restricted sense.

It appears to us that two basic questions are raised by Tintner (1955) in his passive approach. First, it is essential to understand that the randomness in the coefficients of (A, b, c) introduces into the stochastic program a new element unknown to the deterministic framework where the coefficients are fixed and known constants. The new element is that we have a probability distribution of optimal profits and a probability distribution of the optimal solution vector. The deterministic case becomes a special case of this probability distribution, if the point defined by the set (A, b, c) of deterministic coefficients is included in the index set $k = 1, 2, \ldots, K$ already mentioned above. It is clear, however, that because of truncations and restrictions in the parameter space, knowing the distribution of the elements of $\theta = (A, b, c)$ is not sufficient to determine the distribution of optimal solutions and also optimal profits. Indeed, Tintner (1955) has shown that even if the vector θ is normally distributed, the optimal profits and the solution vector, which have to be appropriately defined, are not normally distributed. Prekopa (1966) has derived the conditions and restrictions required for the objective function to follow asymptotically a normal distribution, provided the random variations are so structured around a specific set $(\bar{A}, \bar{b}, \bar{c})$ as to satisfy certain regularity conditions. These and other related aspects of the passive approach are discussed in some detail by Sengupta and Fox (1969) and by Tintner and Sengupta (1972).

A second question raised by the passive approach is how to estimate and numerically compute the distribution function of the optimal solution and the maximand, when the random variations are specially structured around say a specific set of values $\bar{\theta} = (\bar{A}, \bar{b}, \bar{c})$ which may, for instance, be the expected value of the random coefficients. In a number of empirical applications to agricultural and economic planning problems (Tintner, 1960; Sengupta, 1966; Tintner and Sengupta, 1972), it has been found that the empirical probability distribution of optimal profits is not usually normally distributed and even to determine numerically the shape of the probability distribution is a nontrivial matter.

Two of the recent developments in the passive approach may be briefly mentioned here. One is developed by Sengupta (1969[b], 1972[f]), to show the implications of the sampling distribution of optimal profits, as distinct from the parent distribution and to develop suitable probabilistic bounds on optimal profits which are random. An interesting application of this idea arises in what is known as fractile programming which is discussed in detail in later chapters. For instance, in the stochastic LP model $[\max z = c'x, x \in R]$ where the vector c alone is assumed normal with a mean vector m and dispersion matrix V, if the tolerance measure u $(0 < u < 1)$ can be suitably preassigned (e.g., $u > 0.50$), then a convenient deterministic program, known as the fractile program, can be easily derived from the stochastic program as follows:

$$\max t = m'x - q\cdot(x'Vx)^{1/2}, \quad x \in R \tag{1a}$$

where $q = F^{-1}(1 - u) > 0$, if $u > 0.50$ by maximizing the aspiration level t in the probability of profits

$$\text{Prob}\,[z \geq t] = 1 - F\left[\frac{t - m'x}{(x'Vx)^{1/2}}\right] = u \tag{1b}$$

The fractile programming problems of the type (1a) assume that the u-fractile of the cumulative distribution of profits z, i.e., $F(z_u) = u$ is known, since the distribution F is known (i.e., normal in this case).

However, when the sample observations on $\theta = (A, b, c)$ are to be used without the knowledge of the form of the parent distribution, this is impossible. However, the order-statistics $\theta_{(r)}$, $\theta_{(s)}$ $(r < s)$ where $\theta_{(i)}$ denotes the ith smallest observation provide a distribution-free confidence interval for the fractile θ (i.e., $F(\theta_u) = u$) for any continuous parent distribution.

$$\text{Prob } [\theta_{(r)} \leq \theta_u < \theta_{(s)}] = I_u\ (r, N - r + 1) - I_u\ (s, N - s + 1)$$

where $I_u\ (\cdot)$ is the incomplete beta function. Since the ordering on θ, i.e., $\theta_{(1)} \leq \theta_{(2)} \leq \ldots \leq \theta_{(r)} \leq \ldots \leq \theta_{(s)} \leq \ldots \leq \theta_{(N)}$ induces an ordering of profits z, this type of confidence interval may be used to compute the range of optimal profits $(z^*_{(r)} - z^*_{(s)})$ with their associated probabilities.

Another alternative may be used to approximate the parent cdf $F(\theta)$ of θ by its empirical distribution function $F_N(\theta)$ based on the ordered statistics $\theta_{(1)} \leq \theta_{(2)} \leq \ldots \leq \theta_{(N)}$; this is based on the Kolmogorov-Smirnov type distance statistic:

$$D_N = \sup_{-\infty < \theta < \infty} |F_N(\theta) - F(\theta)|$$

whose distribution is independent of the form of the parent distribution $F(\theta)$, except that it is continuous. It is known in nonparametric theory that $F_N(t)$ provides a point estimate of $F(t)$ in the sense for $\theta = t$, the empirical cdf $F_N(t)$ can be used as an estimate of $F(t)$, since the quantity $NF_N(t)$ defines a binomial variable with parameters N and $p = F(t)$. This approach can be extended to more than one parameter, provided $\sqrt{N}\left(F_N(\theta_i - F(\theta_i)\right)$ and $\sqrt{N}\left(F_N(\theta_j) - F(\theta_j)\right)$ are mutually independent for $i \neq j$, where the subscript $i = 1, \ldots, k$ identifies the ith parameter; cases of limited dependence are also investigated in the literature (Walsh, 1964).

The above idea of approximating the parent distribution of a random variable z whose distribution function $F(z|\theta)$ depends on the parameter θ may also be approached through empirical Bayes methods (Maritz, 1970). In the pure Bayesian approach, the parameter θ is itself regarded as a realization of another prior random variable with a prior density $g(\theta)$ say. Any decision d within a set D about θ is then evaluated by maximizing the expected utility function $\bar{U}(d)$:

$$U_g(d) = \int U(d, \theta)g(\theta)d\theta \tag{1c}$$

where a suitable scalar function, called the utility function $U(d, \theta)$ mentioned before, is presumed. This suitable scalar function may also be taken as a loss function $L(d, \theta)$, the expected value of which is minimized. Now with a set of observations t being available, the knowledge of θ is changed by Bayes theorem to a posterior distribution given t, $G(\theta|t)$ with a density $g(\theta|t)$. The new expected utility which is conditional on the observation t then becomes

$$U_g(d|t) = \int U(d, \theta)g(\theta|t)d\theta \tag{1d}$$

Intuitively, one expects that any decision based on maximizing the expected utility defined in (1d) which is based on the posterior distribution must be better than that based on maximizing the expected utility defined in (1c), since otherwise the observations t on θ have been of no value. This aspect has been discussed by Sengupta (1972[f]).

A second development in the passive approach, due to Bracken and

Soland (1966), considers a stochastic LP model where the vector c in the objective function is considered as the unknown mean vector of a multivariate stochastic process. A "prior" statistical distribution on the vector c is then conceived as a prior distribution on the mean of the stochastic process and sampling from this process (Tintner and Sengupta, 1972) enables one to arrive at a posterior distribution on c. The decision problem then is to choose an activity vector x which is feasible and optimal in some sense (e.g., expected value maximization) when a prior or posterior distribution on the vector c is given or estimated. It is clear that in both these cases the passive approach is very intimately connected with the final decisions taken.

1.2 The Active Approaches

In Tintner's version of the active approach (Tintner and Sengupta, 1972; Sengupta, 1970[c]), additional decision variables defined by the resource allocation matrix $U = [u_{ij}]$ where

$$b_i = \sum_{j=1}^{n} b_i u_{ij}; \quad u_{ij} \geq 0, \quad \sum_j u_{ij} \leq 1 \tag{1a}$$

are introduced, and the conditional distribution of optimal profits are determined. Thus we have to compare here the conditional probability distribution of profits under various resource allocation conditions. The implicit cost (penalty or premium) of preassigning a particular allocation matrix $U = [u_{ij}]$ of resources may be computed in this approach through the corresponding dual variables and their probability distributions.

It is clear that additional decision variables may be introduced in