

2

**STUDIES
ON
STATISTICAL APPROACHES
IN
STOCHASTIC PROGRAMMING**

by

Hiroshi MORITA

**STUDIES
ON
STATISTICAL APPROACHES
IN
STOCHASTIC PROGRAMMING**

by

Hiroshi MORITA

Submitted in partial fulfillment of the
requirement for the degree of
DOCTOR OF ENGINEERING
(Applied Mathematics and Physics)

at

Kyoto University

Kyoto, Japan

March, 1992

Preface

Many optimization problems arisen in practical decision making are formulated in the framework of mathematical programming. In the conventional deterministic models of mathematical programming, it is assumed that we are given precise information about the objective function and the constraints. However, when we make a decision in practice, it is necessary to take various constraints and assumptions as well as uncertainty into consideration. Since we seldom know such information for certain, we have to make a decision under uncertainty. Stochastic programming is a field of mathematical methods that deals with the optimization problems under uncertainty expressed by stochastic fluctuation. Optimal decisions are to be made on the basis of certain criteria, which take into account not only optimization of objective but also stability of an optimal decision against stochastic fluctuation. The application areas of stochastic programming include many fields, e.g., production, inventory, agriculture, finance and marketing, etc., all of which involve inevitable uncertainty in formulating the problem and estimating the component of the structured model based on given information.

The main theme of this dissertation is to explore statistical approaches in stochastic programming. There are three types of uncer-

tainty that stochastic programming deals with; the first type of uncertainty comes from the random variables with known probability distribution, the second type of uncertainty also comes from the random variables but with unknown probability distribution, and the third type of uncertainty comes from the coefficients that are unknown or can not be known for certain. It is important in case of the second and third uncertainties to estimate or to predict the unknown distribution and the uncertain coefficients from given information by using statistical techniques. Then the estimates of unknown parameters should be incorporated into the optimization problem together with a significance level of the estimation. In this sense, stochastic programming requires statistical techniques to estimate the unknown parameters with unspecified uncertainty. The role of statistical techniques is as essential as that of optimization techniques.

This dissertation begins with the confidence region method for stochastic programming. It provides a game theoretic minimax model. The constraint forces the unknown coefficients to exist in their confidence regions estimated under a certain significance level. A minimax solution is then constructed by optimizing the objective function under the assumption that the unknown parameters take the worst-case values among the estimated confidence region. As an important special case of this model, we consider two typical stochastic linear programming problems. One has a random right-hand side which are normally distributed with unknown distribution parameters, and the other has unknown cost coefficients which are estimated in terms of the confidence region by means of linear regression analysis. The former problem does

not become a convex programming problem but can be solved by decomposing it systematically into several cases. The later problem is solved by proposed algorithm, which alternately solves two problems, one is solved for making an optimal decision and the other is solved for computing the realizable worst-case values of the unknown coefficients.

Next, this dissertation investigates the minimax model of the confidence region method for stochastic linear programming problems that contain unknown parameters in the coefficients matrix. The linear constraints with unknown coefficients, which are restricted to the confidence regions, form reverse convex constraints. This type of linear programming problem is a reverse convex programming problem, which may have a disjunctive feasible region. We derive an efficient cutting plane algorithm that gives an optimal solution in a finite number of iterations, assuming that only one linear constraint has unknown coefficients. The extension to the problem where more than one linear constraint has unknown coefficients is also mentioned.

One of the typical problems, for which the minimax model due to the confidence region method can be effectively solved, is a generalized P-model of the stochastic linear knapsack problem that contains unknown distribution parameters. We propose a solution algorithm for the problem and show that a minimax solution is found in polynomial time.

Finance problem has recently become a hot field of operations research. As an important application of stochastic programming including the parameter estimation, a portfolio selection problem is discussed in this dissertation. A portfolio selection problem gives the investor an

optimal investment strategy to receive a large return under a small risk. The rate of return on investment is predicted from the historical data based on the market model. A single index model is the simplest market model among portfolio selection problems, but has been most frequently used in practice. The coefficients of a single index model are estimated by means of regression analysis. We formulate the portfolio selection problem as a stochastic linear knapsack programming problem. The probability maximizing model of stochastic linear knapsack problem is derived and we propose efficient algorithms to solve the problems with several types of random variables, e.g., block diagonally correlated random variables, a single index model and a multi index model.

Finally, this dissertation investigates a stochastic improvement method. Since, in the confidence region method, we solve the problem that is built by using the estimates obtained from the currently available information, we should rebuild the problem whenever new information becomes available. While, in the stochastic improvement method, we update the estimates of unknown parameters as well as the solution whenever new statistical data are delivered. We believe that this statistical approach is important in stochastic programming.

The statistical approaches proposed in this dissertation for stochastic programming extend the territory of mathematical programming. We hope for the development of the statistical approaches for stochastic programming to a dynamic technique and/or a sequential technique like a stochastic approximation procedure.

Hiroshi MORITA
March, 1992

Acknowledgement

The author would like to express his sincere appreciation to Professor Toshihide Ibaraki of Kyoto University for supervising this dissertation. His continuous encouragement and invaluable comments have helped to accomplish this dissertation.

The author is also heartily grateful to Professor Hiroaki Ishii of Okayama University for his enthusiastic discussions and persistent encouragement. He guided the author to the present study and have been giving the continuous and invaluable suggestions. Without his support, none of this work would have been possible.

Furthermore, the author is highly indebted to Professor Emeritus Toshio Nishida of Osaka University, Professor at Osaka International University, for his invaluable comments and encouragement.

Finally, the author wishes to thank to Associate Professor Shogo Shiode of Kobe University of Mercantile Marine for his helpful and stimulative discussions on stochastic programming.

Contents

Preface	i
Acknowledgement	v
Chapter 1. INTRODUCTION	1
1.1. Purpose of the Dissertation and Historical Background	1
1.2. Brief Review of Stochastic Programming	3
1.3. Outline of the Dissertation	15
Chapter 2. A CONFIDENCE REGION METHOD FOR STOCHASTIC LINEAR PROGRAMMING	21
2.1. Introduction	21
2.2. Stochastic Linear Programming with an Estimated Right- hand Side	23
2.3. A Solution Method	28
2.4. Asymptotic Properties	36
2.5. An Illustrative Example	38
2.6. Stochastic Linear Programming with an Estimated Objective Function	42
2.7. A Solution Algorithm for Problem P_{10} and an Illustrative Example	44
2.8. Conclusion	49

Chapter 3. STOCHASTIC LINEAR PROGRAMMING WITH ESTIMATED CONSTRAINTS	51
3.1. Introduction	51
3.2. Formulation of the Problem	53
3.3. Reverse Convex Programming	55
3.4. A Solution Algorithm	59
3.5. A Lower Bound of the Optimal Value	66
3.6. An Illustrative Example	70
3.7. Extension to the Linear Programming Problem with more than One Additional Constraint Including Unknown Coefficients	72
3.8. Conclusion	80
Chapter 4. A CONFIDENCE REGION METHOD FOR A STOCHASTIC LINEAR KNAPSACK PROBLEM	83
4.1. Introduction	83
4.2. A Stochastic Linear Knapsack Problem	85
4.3. A Confidence Region Method for a Stochastic Linear Knapsack Problem	89
4.4. Conclusion	95
Chapter 5. A PROBABILITY MAXIMIZING MODEL OF A STOCHASTIC LINEAR KNAPSACK PROBLEM	97
5.1. Introduction	97
5.2. A Probability Maximizing Model of a Stochastic Linear Knapsack Problem	100
5.3. An Optimal Solution on the Efficient Frontier	104
5.4. Conclusion	111

Chapter 6. AN APPLICATION TO PORTFOLIO SELECTION PROBLEMS	113
6.1. Introduction	113
6.2. A Block Diagonal Model	115
6.2.1. Decomposition into subproblems	115
6.2.2. A ranking method and a solution algorithm	120
6.2.3. A simple portfolio selection problem	124
6.3. Index Models	130
6.3.1. A single index model and a multi index model	130
6.3.2. Formulation of the problem of a single index model	132
6.3.3. A solution algorithm	134
6.3.4. Extension to a multi index model	140
6.4. Conclusion	142
Chapter 7. A STOCHASTIC IMPROVEMENT METHOD FOR STOCHASTIC PROGRAMMING	145
7.1. Introduction	145
7.2. Formulation of the Problem	147
7.3. A Stochastic Improvement Method	152
7.4. Consistency of the Estimated Region and Convergence of Algorithm SIM	157
7.5. Computational Results	160
7.6. Conclusion	161
Chapter 8. CONCLUSION	165
References	169

Chapter 1.

INTRODUCTION

1.1. Purpose of the Dissertation and Historical Background

Statistical approaches in stochastic programming are discussed in this dissertation. Stochastic programming has been developed as probabilistic generalization of mathematical programming and has become an important field of mathematical programming from the view point of theory as well as practice. For many problems encountered in practice, optimal decisions are usually made under some uncertainties, because practical problems are subject to stochastic fluctuation and necessary information for decision making are not always available for certain. In most of practical cases, the stochastic fluctuation has some probability distributions, but the exact forms of distributions are not available. Therefore, it is important to identify the distribution from historical and/or experimental observations. The optimization problems with unknown or uncertain parameters necessitate statistical approaches to evaluate them.

The main purpose of this dissertation is to provide a new class of statistical approaches in stochastic programming and to consider some

related applications. Some methods to be characterized as statistical approaches have recently been proposed [C3, D4]. In these methods, model formulation is affected by not only the optimization techniques to be solved but also how to treat the statistical data affect the model building. Many statistical methods have been studied [M2, R2] to process the statistical data. For example, a linear regression analysis is one of the most useful methods for estimating the unknown parameters of linear models and the results of analysis have desirable properties to be incorporated into optimization problems.

This dissertation focuses on a confidence region method and a stochastic improvement method, both of which are important statistical approaches in stochastic programming. The confidence region method provides a game theoretic minimax problem, which gives a minimax solution under the assumption that unknown parameters take the worst-case behavior among the confidence region under a certain significance level. The stochastic improvement method gives a point sequence that converges to an optimal solution of the stochastic programming problem and is improved when the unknown parameters in the problem are updated as a result of additional statistical data. As a typical example of stochastic programming including parameter estimation, we consider a stochastic linear knapsack problem with random cost coefficients, which arises as a mathematical formulation of a portfolio selection problem. In a portfolio selection problem, the rate of return on investment is viewed as a random variable and is expressed as a linear function of some indices based on the market model. This dissertation proposes solution algorithms for several types of portfolio

selection problems formulated as a probability maximizing model of a stochastic linear knapsack problem.

It is also emphasized here that these statistical approaches for stochastic programming with unknown parameters can be applied to many other practical optimization problems.

1.2. Brief Review of Stochastic Programming

This section reviews the development of stochastic programming related to the subjects of this dissertation. Stochastic programming deals with the optimization problem that involves stochastic fluctuations. Application areas of stochastic programming have been widened to a great extent, e.g., agriculture [F1, N1], water storage [P2], transportation [Q1], inventory [B5], finance and marketing [D5, M3, S1], production, engineering and others.

In 1941, Tintner [T2] made distinction between subjective risk and subjective uncertainty. The former implies the random variable which has a known probability distribution of anticipation, while the latter implies the random variable which has only a priori probability of the probability distributions themselves. Subjective risk led to the main body of stochastic programming, and recently subjective uncertainty has been treated in stochastic programming by statistical approach.

The epoch making paper by Dantzig [D1] in 1955 is motivated by the recognition that a method for the classical linear programs is not appropriate to formulate the problems in many practical situations. Since the quantities of activities in practical problems have uncertainty,

the classical programming models based on the fixed known data should be extended to the stochastic programming models. There are two situations in stochastic programming, which are called the “wait-and-see” situation and the “here-and-now” situation according to Madansky [M1]. The former situation requires to wait until an observation on the random variable is actually made and then to solve the nonstochastic problem. In contrast to this, the latter situation requires to solve the problem before observing the realization of the random variable. This dissertation belongs to the “here-and-now” situation.

Give the known probability space $(\Omega, \mathcal{A}, \mathcal{P})$, where Ω is a set of possible environments, \mathcal{A} is a set of possible events and \mathcal{P} is a probability measure. We consider the following optimization problem on $(\Omega, \mathcal{A}, \mathcal{P})$:

$$(1.1) \quad \begin{cases} \text{Minimize} & \gamma(\omega) = g_0(x, \omega), \\ \text{subject to} & g_i(x, \omega) \leq 0, \quad i = 1, 2, \dots, m, \\ & x \in X \subset R^n, \end{cases}$$

where X is a given subset of R^n and $g_i : R^n \times \Omega \rightarrow [-\infty, +\infty)$, $i = 0, 1, \dots, m$ are real-valued random functions. The environment variable ω is an element of Ω , which is introduced to determine the environment of the optimization problem prescribed under uncertainty.

The “wait-and-see” situation leads to the distribution problem to be described as follows. When the environment is completely known before we solve the problem, we can solve the deterministic problem by assigning the known values to the variables ω . However, different environments ω^1 and ω^2 may give different optimal decisions $x^1 = x^*(\omega^1)$ and $x^2 = x^*(\omega^2)$, respectively, and we want to know the behavior of

an optimal decision under all possible values of the environment variables. Typical questions in the wait-and-see situation are: “What is the expectation or 95% percentile of the optimal value?”, “What is the distribution of this optimal value?” and so on. These problems are called distribution problems. Due to the difficulty of an analysis of these problems for general stochastic programming problems, the following stochastic linear programming problem (1.2) has been studied intensively as a special case of stochastic programming.

$$(1.2) \quad \begin{cases} \text{Minimize} & \gamma = c(\omega)'x, \\ \text{subject to} & A(\omega)x = b(\omega), \\ & x \geq 0. \end{cases}$$

When the optimal basis does not change for all $\omega \in \Omega$, the distribution of the optima value γ has been considered by Babbar [B1] and Wagner [W1]. Bereanu [B6] derives the cumulative function and the mean value of the optimal value, when all coefficients are stochastic. Kall surveys the distribution problem for stochastic linear programming in his book [K1].

In the “here-and-now” situation, there are many probabilistic definitions of feasibility; $x \in X$ satisfies the constraints of problem (1.1) with a certain significance level $\alpha \in (0, 1)$:

$$(1.3) \quad \Pr\{\omega | g_i(x, \omega) \leq 0, \quad i = 1, 2, \dots, m\} \geq \alpha,$$

or in the average:

$$(1.4) \quad E\{g_i(x, \omega)\} \leq 0, \quad i = 1, 2, \dots, m,$$

and others. Similar definitions are also applied to optimality. Many concrete formulations of stochastic programming may be reduced to

the general version of the stochastic programming problem over the probability space $(\Omega, \mathcal{A}, \mathcal{P})$:

$$(1.5) \quad \left\{ \begin{array}{l} \text{Minimize} \quad F_0(x) = E\{f_0(x, \omega)\}, \\ \qquad \qquad \qquad = \int f_0(x, \omega) \mathcal{P}(d\omega), \\ \text{subject to} \quad F_i(x) = E\{f_i(x, \omega)\}, \\ \qquad \qquad \qquad = \int f_i(x, \omega) \mathcal{P}(d\omega) \leq 0, \quad i = 1, 2, \dots, m, \\ \qquad \qquad \qquad x \in X \subset R^n, \end{array} \right.$$

where

$$(1.6) \quad \begin{array}{l} f_0 : R^n \times \Omega \rightarrow R \cup \{-\infty, +\infty\}, \\ f_i : R^n \times \Omega \rightarrow R, \quad i = 1, 2, \dots, m, \end{array}$$

and the expectation operator E is assumed to be well-defined for every $x \in X$. The functions f_i , $i = 0, 1, \dots, m$, correspond to g_i , $i = 0, 1, \dots, m$, respectively in (1.1), and are introduced to express the several probabilistic definitions of feasibility and optimality, which are to be noted later.

Beale [B3] and Dantzig [D1] independently considered the two-stage problem in 1955. This is categorized as the "here-and-now" situation. In the first stage, we have to make a decision before the realizations of random variables take place. Since several constraints may be violated by the realized values of random variables, an adjustment of the violated constraints is made in the second stage. In the case of stochastic problem (1.1), the cost required for an adjustment of the violated constraints is considered to be proportional to the amount of violation. The adjustment cost is added to the objective function as a penalty, that is,

$$(1.7) \quad f_0(x, \omega) = g_0(x, \omega) + \sum_{i=1}^m q_i \cdot \max[0, g_i(x, \omega)],$$

where q_i , $i = 1, 2, \dots, m$, are positive weight coefficients.

For stochastic linear programming problem (1.2), Beale [B3] and Dantzig [D1] formulate the two-stage problem:

$$(1.8) \quad \left\{ \begin{array}{l} \text{Minimize} \quad c'x + E(\min_y q'y), \\ \text{subject to} \quad Ax + y = b(\omega), \\ \qquad \qquad \qquad x \geq 0, \quad y \geq 0, \end{array} \right.$$

where only b is assumed to be stochastic. The deterministic convex problem equivalent to (1.8) is derived by Wets [W3]. Everett and Ziemba [E6] and Walkup and Wets [W2] consider more general problem:

$$(1.9) \quad \left\{ \begin{array}{l} \text{Minimize} \quad c'x + E[Q(x, \omega)], \\ \text{subject to} \quad x \geq 0, \end{array} \right.$$

where

$$(1.10) \quad Q(x, \omega) = \min_{y \in R^l} \{q(\omega)'y \mid W(\omega)y = b(\omega) - A(\omega)x, y \geq 0\}.$$

This problem is characterized as "stochastic programming with recourse" by Walkup and Wets. We call the vector $y \in R^l$ and the matrix $W \in R^n \times R^l$ a recourse variable and a recourse matrix, respectively. The recourse action y is chosen so as to minimize the penalty cost with respect to a decision variable x and a realized environment ω . An optimal decision should minimize the total costs, which are the sum of the net cost $c'x$ and the expectation of the cost for an adjustment of violation $E[Q(x, \omega)]$. When q and W are nonstochastic, (1.9) is said to be a fixed recourse problem. When $\{y \mid Wy = z, y \geq 0\} \neq \phi$ for all $z \in R^n$, W is said to be a complete recourse matrix.

The simple recourse problem given as follows is surveyed by Ziemba [Z1] and Wets [W4].

$$(1.11) \quad \left| \begin{array}{l} \text{Minimize} \quad c'x + E[\min_{y^+, y^-} (p'y^+ + q'y^-)], \\ \text{subject to} \quad A(\omega)x + y^+ - y^- = b(\omega), \\ \quad \quad \quad x, y^+, y^- \geq 0, \end{array} \right.$$

where

$$(1.12) \quad \begin{aligned} y^+ &= \max[0, -A(\omega)x + b(\omega)], \\ y^- &= \max[0, A(\omega)x - b(\omega)]. \end{aligned}$$

The above two-stage recourse problems do not take into account the concept of time. However, the multi-stage recourse problems arisen in dynamic models are also considered. The L -shaped method for the two-stage stochastic linear problems given by Van Slyke and Wets [V2] is extended to the multi-stage recourse problems by Birge [B6]. The L -shaped method is an outer linearization decomposition approach for the L -shaped linear programs which are arisen in stochastic linear programming with recourse.

Charnes and Cooper [C1] introduce "chance constraints", which is an extension of the notion of constraints from the view point of the "here-and-now" situation. It is different from the two-stage problem in the manner of treatment of the violation of stochastic constraints. In the two-stage problem, the constraints should be always satisfied and therefore the amount of violation incurred in the first stage is adjusted in the second stage. On the other hand, it is not always required to satisfy the constraints for the chance constrained approach. The constraints should be satisfied with a given probability, that is, the chance

constraints watch only the probability that the constraints are violated, rather than the amount of violation. For stochastic problem (1.1), the chance constraints are given by

$$(1.13) \quad \Pr(\omega \mid g_l(x, \omega) \leq 0) \geq \alpha_l, \quad l = 1, 2, \dots, m,$$

or

$$(1.14) \quad \Pr(\omega \mid g_l(x, \omega) \leq 0, \quad l = 1, 2, \dots, m) \geq \alpha,$$

in the form of joint probability, where $\alpha, \alpha_1, \dots, \alpha_m \in (0, 1)$ are the satisficing levels of chance constraints, which are given in advance. In the formulation of (1.5), we have that (1.13) and (1.14) are expressed as

$$(1.15) \quad f_i(x, \omega) = \begin{cases} \alpha - 1 & \text{if } g_i(x, \omega) \leq 0 \\ \alpha & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \dots, m$, and

$$(1.16) \quad f_l(x, \omega) = \begin{cases} \alpha - 1 & \text{if } g_l(x, \omega) \leq 0, \quad l = 1, 2, \dots, m \\ \alpha & \text{otherwise,} \end{cases}$$

respectively.

The chance constraints can be transformed into the equivalent deterministic constraints. However, the transformation is usually difficult and the resulting nonlinear constraints are intractable in the sense of computation. Even if the original constraints are defined by linear functions, it is not easy to carry out the transformation to the deterministic problem unless the random variable has a certain special type of distribution like normal. Prékopa [P1] applies a logarithmic concave

probability measure to stochastic programming and shows that the feasible region of chance constraints is convex if the distribution is given by the logarithmic concave probability measure.

The stochastic objective function of the stochastic programming problem is handled by its certainty equivalent in the framework of deterministic models. For the following stochastic linear programming problem,

$$(1.17) \quad \left| \begin{array}{l} \text{Minimize } c(\omega)'x, \\ \text{subject to } A(\omega)x \leq b(\omega), \\ x \geq 0. \end{array} \right.$$

There are four types of certainty equivalent $F(x)$.

(i) E-model [C2]:

$$(1.18) \quad F(x) = E[c(\omega)'x].$$

The E-model minimizes the expectation of the objective value. This is a classical model that can handle the random cost coefficients [D1]. However, it is argued that the consideration on the variance may also be necessary, since it may not be desirable to optimize the expected value when its variance is very large. This consideration leads to the next model.

(ii) V-model [C2]:

$$(1.19) \quad F(x) = V[c(\omega)'x].$$

The V-model minimizes the variance of the objective value. It is also possible to consider both expectation and variance. The

EV-model has an objective function incorporating two conflicting objectives, such as

$$(1.20) \quad \frac{E[c(\omega)'x]}{V[c(\omega)'x]}, \quad E[c(\omega)'x] - p \cdot V[c(\omega)'x]$$

and so on, where p is a positive weight coefficient on the variance.

(iii) Probability maximizing model [C2]:

$$(1.21) \quad F(x) = \Pr(c(\omega)'x \leq f).$$

The probability maximizing model maximizes the probability that $c(\omega)'x$ does not exceed a given goal value f . If the joint distribution of the components $c_i(\omega)$ of c is normal with the mean μ and the variance covariance matrix V , the probability of $c(\omega)'x$ not exceeding f is

$$(1.22) \quad \Pr(c(\omega)'x \leq f) = \Phi\left(\frac{f - \mu'x}{\sqrt{x'Vx}}\right),$$

where $\Phi(\cdot)$ is the distribution function of $\mathcal{N}(0,1)$, i.e., standard normal distribution with the mean 0 and the variance 1. It is known that the maximization of this probability is equivalent to the maximization of the fractional function $(f - \mu'x)/\sqrt{x'Vx}$.

(iv) P-model [K3]:

$$(1.23) \quad F(x) = f, \quad \Pr(c(\omega)'x \leq f) = \alpha.$$

Kataoka [K3] considers to minimize f subject to $\Pr(c(\omega)'x \leq f) = \alpha$ for fixed α . Assuming again that c has the normal distribution, the chance constraint is transformed into

$$(1.24) \quad f = \mu'x + K_\alpha \sqrt{x'Vx},$$

where $K_\alpha = \Phi^{-1}(\alpha)$. If $\alpha > \frac{1}{2}$, then $K_\alpha > 0$ and f is convex. Ishii, Nishida and Nanbu [I5] generalize the P-model into the following model by considering the satisficing level α as a decision variable, too.

$$(1.25) \quad \left\{ \begin{array}{l} \text{Minimize} \quad f - \lambda\alpha, \\ \text{subject to} \quad \Pr(c(\omega)'x \leq f) = \alpha, \\ \quad \quad \quad A(\omega)x \leq b(\omega), \\ \quad \quad \quad x \geq 0, \end{array} \right.$$

where λ is a positive weight coefficient on α .

In the probability maximizing model, the goal value that a decision maker wants to attain is specified in advance and the probability that his goal value is attained is maximized. On the other hand, in the P-model, the probability that a decision maker's objective is attained is fixed and then his goal value is minimized.

Stochastic programming is closely connected with a game theory [S2] and an information theory [R1]. A game theoretic situation naturally arises in applying stochastic programming. A two-person zero-sum stochastic game can be constructed with problem (1.2). Player 1 is the decision maker who chooses a strategy vector $x \in \mathcal{M}$ where \mathcal{M} is a subset of R_+^n . Player 2 is nature, i.e., stochastic fluctuation, who chooses a strategy vector $z \in \mathcal{N}$, where $z = (A(\omega), b(\omega), c(\omega))$ and \mathcal{N} is a subset corresponding to the feasible region of variation in $\{(A(\omega), b(\omega), c(\omega)); \omega \in \Omega\}$ which is given in advance. The game is denoted by $G = (\mathcal{M}, \mathcal{N}; g)$ with the following payoff function g :

$$(1.26) \quad g(x, z) = c(\omega)'x + \sum_{i=1}^m p_i \cdot \max[0, A_i(\omega)'x - b_i(\omega)].$$

The mixed strategies of the two players are denoted by F and H , respectively, where F is a set of distributions F_x of vector x and H is a set of distributions H_z of vector z . By the minimax principle, the value of game is given by

$$(1.27) \quad \begin{aligned} \text{Maximize}_{F_x \in F} \text{Minimize}_{H_z \in H} \int_{\mathcal{M} \times \mathcal{N}} g(u, v) dF_x(u) dH_z(v) \\ = \int_{\mathcal{M} \times \mathcal{N}} g(u, v) dF_x^*(u) dH_z^*(v) \end{aligned}$$

for the optimum strategies $F_x^* \in F$ and $H_z^* \in \tilde{H} \subseteq H$, where \tilde{H} is a subset of H representing the restriction on the choice of the pure strategies of two players.

The statistical approach in stochastic programming is related to an information theory. If several parameters describing a model of mathematical programming are unknown, we are faced with the problem of estimating such parameters based on limited information and with the problem of updating them when additional information becomes available. The unknown parameters may have a priori distribution, and the observations provide posterior information. The values of information in this context, which is first considered by Raiffa and Schlaifer [R1], are introduced by Bracken and Soland [B8], to stochastic programming. They consider two types of values of information, i.e., the expected value of perfect information (EVPI) and the expected value of sample information (EVSI). The additional sample information reduces the uncertainty of the parameters, while the perfect information gives the complete knowledge of the parameters. The EVPI is considered as the upper bound of the value of information, representing what

would be paid for the perfect information. The EVSI is used to decide whether one more sampling of information is valuable or not, comparing it with the sampling cost. The EVSI approaches to the EVPI as sample size increases and the optimal sample size n^* is decided by

$$(1.28) \quad n^* = \min\{n \mid (\text{EVSI})_n - \gamma \cdot n \leq 0\},$$

where γ is the sampling cost for one observation [J1] and $(\text{EVSI})_n$ is the EVSI with n samples.

Another statistical approach is a game theoretic minimax one. Consider the case in which the distribution of some stochastic components is not known for certain, that is, the distribution parameters are unknown or it is only known that the distribution belongs to a certain class. The minimax approach defines an optimal decision under the most pessimistic situation, and then the resulting problem is formulated as follows.

$$(1.29) \quad \left| \begin{array}{l} \text{Maximize} \\ x \in X \end{array} \right. \left. \begin{array}{l} \text{Minimize} \\ F \in \mathfrak{F} \end{array} \right. E_F[f_0(x, \omega) - \phi(x, \omega)],$$

where F is a probability distribution of stochastic components, \mathfrak{F} is a given class of distributions, X is a nonempty closed subset of R^n and $\phi(x, \omega)$ is a penalty cost required to adjust the infeasibility with distribution F in class \mathfrak{F} . This problem is considered by Dupačová [D4]. When the unknown distribution parameters θ of the stochastic components are estimated and their confidence region S_α are obtained with a significance level α , the possible distribution F in problem (1.29) is restricted to $F \in \mathfrak{F}_0 \subseteq \mathfrak{F}$, where $\mathfrak{F}_0 = \mathfrak{F}(\cdot \mid \theta \in S_\alpha)$. Stability of

the optimal solutions of stochastic linear programming problems formulated along this line is studied by Dupačová [D3] for the simple recourse problem with a random right-hand side.

When stochastic components appear as statistical time series data in stochastic linear programming and the current components are predicted using the previous observations, Cibra [C3] has discussed the statistical approach in terms of prediction regions. Then the linear programming problem with a random right-hand side is solved using the method of parametric programming. He defines an α -optimal decision and gives the $(1 - \alpha)$ -prediction interval of objective value, in which the value of objective function located with probability at least $1 - \alpha$, and the ellipsoidal $(1 - \alpha)$ -prediction region of optimal solution, in which the optimal solution located with probability at least $1 - \alpha$.

The statistical approach in stochastic programming is useful and important for the cases in which we would like to use the available information to make an optimal decision for the problem under uncertainty and for the problem involving stochastic components with statistical properties, e.g., the estimator based on regression analysis and statistical time series data.

1.3. Outline of the Dissertation

This dissertation consists of eight chapters, all of which are concerned with statistical approach for the stochastic programming problems.

In Chapter 2, the confidence region method is considered for two types of stochastic linear programming problems. The unknown parameters in the problem are restricted to their confidence regions. The confidence region method gives a game theoretic minimax solution which optimizes the objective function under the assumption that the parameters may take the worst-case values in their confidence regions under a given significance level. First, we consider the stochastic linear programming problem which has a normally distributed random right-hand side that contains unknown distribution parameters. The confidence region are estimated from the information of sample data. An solution algorithm is given for this minimax problem due to the confidence region method. It is also discussed how an optimal solution behaves as the number of obtained samples increases. Next the stochastic linear programming problem that contains unknown cost coefficients is considered. The unknown cost coefficients are estimated that they locate on the ellipsoidal confidence region under a certain significance level. A proposed algorithm finds an optimal solution after a finite number of iterations by solving alternately two problems, one is to find an optimal decision and the other is to determine the realizable worst-case values of the unknown cost coefficients.

Chapter 3 discusses the confidence region method for the stochastic linear programming problem with a random coefficients matrix. The linear constraint whose unknown parameters are restricted to the confidence region forms the reverse convex feasible region. A Tuy's cutting plane method [H3,T1] is known as an effective procedure for reverse convex programming, but its finite convergence is not proved. This

dissertation gives a cutting plane algorithm for the problem that only one constraint has unknown parameters. It is a modification of the Tuy's algorithm, but finds an optimal solution after a finite number of iterations. The possibility of expansion of this algorithm to the problem that several constraints have unknown parameters is also discussed.

Chapter 4 treats the confidence region method for a generalized P-model of the stochastic linear knapsack problem. The cost coefficients are assumed to be normally distributed random variables with unknown means and variances, which are to be estimated by their confidence regions. It is shown that this problem can be transformed into the problem with a known distribution. A proposed algorithm finds an optimal solution in polynomial computational time.

Chapter 5 and Chapter 6 discuss a portfolio selection problem as an application of the stochastic linear knapsack problem. The portfolio selection problem is a mathematical finance problem that consider an optimal allocation of a given amount of money among several investments in order to receive a big return under a small risk. A probability maximizing model of the stochastic linear knapsack problem with random cost coefficients is considered in Chapter 5. This problem is transformed into the equivalent deterministic quadratic programming problem by introducing two positive parameters. An optimal portfolio on the efficient frontier is discussed in the mean variance framework, and several properties of an optimal portfolio are shown as the preparation for the next chapter.

Chapter 6 discusses two types of the portfolio selection problems. First, this dissertation considers the case in which the variance covariance matrix of random cost coefficients is block diagonal. The problem

is decomposed into subproblems corresponding to the blocks and an optimal allocation to each block is determined according to the generalized Zipkin's [Z2] ranking properties. A proposed algorithm finds an optimal solution in reasonable computational time. Next, this dissertation considers the case in which the random cost coefficients are expressed by the index models. In a single index model of a portfolio selection problem, the rate of return on investment is expressed as a linear function of the common index, where information on the correlation between possible investments are condensed to a linear regression model with one explanation factor. The coefficients in a linear regression model need be estimated from historical data. This dissertation gives an effective algorithm to find an optimal portfolio for the problem of a single index model. A multi index model introduces other common indices for capturing additional information. A modified algorithm is also developed to solve the problem of a multi index model by introducing several parameters corresponding to the indices.

Chapter 7 discusses a stochastic improvement method for the linear programming problem that the linear constraints have unknown coefficients. The unknown coefficients are estimated using a multi-variate regression analysis. Upon receiving new statistical data, this method improves the solution in real time by making use of a descent method in an adaptive way. It is shown that this iteration of improvements finds an optimal solution with probability one, based on the fact that the consistency of the estimator assures that the estimated problem converges to the problem with true parameters with probability one.

Finally, Chapter 8 summarizes the results obtained in this dissertation and discusses further directions of developments for statistical approaches in stochastic programming.

Chapter 2.

A CONFIDENCE REGION METHOD FOR STOCHASTIC LINEAR PROGRAMMING

2.1. Introduction

This chapter discusses the confidence region method for two types of stochastic linear programming problems that contain unknown parameters. One is the problem with a random right-hand side and the other is the problem with a random objective function. In both cases, random components are induced by noisy observations for estimation of the unknown parameters. The confidence region method gives a game theoretic minimax solution, where the unknown parameters are restricted to their confidence regions and a minimax solution optimizes the objective function considering the worst-case behavior of parameters with a given significance level.

Consider the following stochastic optimization problem on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$:

$$\text{SP: } \left\{ \begin{array}{l} \text{Maximize } E\{f_0(x, \omega)\}, \\ \text{subject to } E\{f_i(x, \omega)\} \leq 0, \quad i = 1, 2, \dots, m, \\ x \in X \subset R^n, \quad \omega \in \Omega, \end{array} \right.$$

where X is a subset of R^n and $f_0, f_i, i = 1, 2, \dots, m$, are extended real-valued functions. Let \mathfrak{F} denote a class of distribution and let the

parameters θ of distribution $F \in \mathfrak{F}$ be unknown. Then a minimax model due to the confidence region method is formulated as

$$(2.1) \quad \text{Maximize}_{x \in X} \text{Minimize}_{F \in \mathfrak{F}_0 \subseteq \mathfrak{F}} E_F[f_0(x, \omega) - \phi(x, \omega)]$$

where $\phi(x, \omega)$ is a penalty cost for adjusting the infeasibility. Given the confidence region S_α of unknown parameters θ , the probability distribution F of stochastic components is restricted to the subclass \mathfrak{F}_0 of \mathfrak{F} defined as follows.

$$(2.2) \quad F \in \mathfrak{F}_0 \equiv \mathfrak{F}(\cdot | \theta \in S_\alpha) \subseteq \mathfrak{F}$$

In practice, for example, sizes of industrial products are normally distributed but the mean and the variance are unknown, and their lifetimes are exponentially distributed but the mean is unknown. In the confidence region method, these unknown parameters are estimated and their confidence regions are obtained from observed data. This approach is also applicable when we need to get a desirable precision of estimates by additional information.

We consider the following stochastic linear programming problem.

$$\text{LP:} \quad \left\{ \begin{array}{l} \text{Maximize} \quad \sum_{j=1}^n c_j x_j, \\ \text{subject to} \quad \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ \quad \quad \quad x_j \geq 0, \quad j = 1, 2, \dots, n. \end{array} \right.$$

Section 2.2 through Section 2.5 discuss the case where only b_i , $i = 1, 2, \dots, m$, are random variables which have a certain class of distribution with unknown parameters. Section 2.2 formulates the problem

as a recourse problem in case that b_i , $i = 1, 2, \dots, m$, are normally distributed random variables. The details of solution method is described in Section 2.3. Section 2.4 clarifies asymptotic properties of the problem when sample size tends to infinity. An illustrative example is given in Section 2.5. The dual of this problem has random variables in the objective function, that is, b_i , $i = 1, 2, \dots, m$, are known but c_j , $j = 1, 2, \dots, n$, are unknown. Section 2.6 formulates the problem as a minimax model, when the unknown coefficients are observed including random noise. Section 2.7 describes the details of solution algorithm and an illustrative example. Finally, Section 2.8 gives a summary of Chapter 2.

2.2. Stochastic Linear Programming with an Estimated Right-hand Side

First, we consider the following stochastic linear programming problem P_1 :

$$P_1 : \quad \left\{ \begin{array}{l} \text{Maximize} \quad \sum_{j=1}^n c_j x_j, \\ \text{subject to} \quad \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ \quad \quad \quad x_j \geq 0, \quad j = 1, 2, \dots, n, \end{array} \right.$$

where a_{ij} and c_j for $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, are known but b_i , $i = 1, 2, \dots, m$, are random variables which have the distribution function $F(\cdot; \theta)$ with unknown parameters θ . Let A_i denote the i -th row of matrix $A = (a_{ij})$. Now we assume that $A = (A_1, A_2, \dots, A_m)'$ has a full column rank.

Consider the following two-stage problem P_2 with a "quadratic" recourse.

$$P_2 : \left| \begin{array}{l} \text{Maximize}_{x \geq 0} \sum_{j=1}^n c_j x_j - E_F \left[\sum_{i=1}^m w_i (A_i x - b_i)^2 \right], \end{array} \right.$$

where $w_i, i = 1, 2, \dots, m$ denote positive weight on the i -th constraint $A_i x = b_i$, respectively. The quadratic recourse in this model is more tractable than the corresponding simple recourse (1.11) and reflects the situation in which the constraint infeasibility has a critical meaning, since the quadratic recourse is greater than simple recourse when y_i is large. Problem P_2 is equivalent to the following problem P_3 :

$$P_3 : \left| \begin{array}{l} \text{Maximize}_{x \geq 0} \sum_{j=1}^n c_j x_j \\ - \int \cdots \int \sum_{i=1}^m w_i (A_i x - t_i)^2 dF(t_1, \dots, t_m; \theta), \end{array} \right.$$

Now unknown parameters θ are restricted to the confidence region S_α with a given significance level α , and we propose the following minimax problem P_4 due to the confidence region method:

$$P_4 : \left| \begin{array}{l} \text{Maximize}_{x \geq 0} \text{Minimize}_{\theta \in S_\alpha} L = \sum_{j=1}^n c_j x_j - \int \cdots \int \sum_{i=1}^m w_i (A_i x - t_i)^2 \times \\ dF(t_1, \dots, t_m; \theta), \end{array} \right.$$

Namely, we minimize the objective function added the penalty cost, assuming that the parameters θ take the worst-case values in the confidence region S_α . This model reflects the situation in which we should make a decision to minimize the maximal possible damage, if the values of parameters θ are not known perfectly.

We mainly consider the case of a class of normal distributions. Let $b_i, i = 1, 2, \dots, m$, be normally and independently distributed random variables with unknown distribution parameters $\theta = (\mu, \sigma^2)$. To construct the confidence region S_α of distribution parameters, we define the following notations.

μ_i : mean of $b_i, i = 1, 2, \dots, m$

σ_i^2 : variance of $b_i, i = 1, 2, \dots, m$

$\bar{\mu}_i$: sample mean of $b_i, i = 1, 2, \dots, m$

s_i^2 : sample variance of $b_i, i = 1, 2, \dots, m$

N : sample size

α, β : significance level (%)

$F_\alpha(m, n)$: α percentile of an F-distribution with (m, n) degrees of freedom

$\chi_\beta^2(n)$: β percentile of a χ^2 -distribution with n degrees of freedom

$\Phi(\cdot)$: the distribution function of an m -dimensional standard normal distribution

The confidence region of mean $\mu_i, i = 1, 2, \dots, m$, is obtained from the fact that the Hotelling statistics

$$(2.3) \quad F = \frac{N(N-m)}{m(N-1)} \sum_{i=1}^m \frac{(\mu_i - \bar{\mu}_i)^2}{s_i^2}$$

has an F-distribution with $(m, N-m)$ degrees of freedom. Then, under a significance level α , the confidence region of $\mu_i, i = 1, 2, \dots, m$, is given by

$$(2.4) \quad \sum_{i=1}^m \frac{(\mu_i - \bar{\mu}_i)^2}{s_i^2} \leq \frac{m(N-1)}{N(N-m)} F_\alpha(m, N-m).$$

Note that the confidence region of $\mu_i, i = 1, 2, \dots, m$, is an ellipsoidal region.

Due to the difficulty of solving the maximizing problem with respect to the variance in problem P_4 , we relax the confidence region of variances as a collection of the confidence interval of each variance assuming that random variables are mutually independent, instead of the confidence region of variance covariance matrix. They are constructed from the fact that the statistics

$$(2.5) \quad \chi^2 = (N-1) \frac{s_i^2}{\sigma_i^2}$$

has a χ^2 -distribution with $(N-1)$ degrees of freedom. Then the confidence interval of each σ_i^2 is given by

$$(2.6) \quad \frac{(N-1)s_i^2}{\chi_{\beta}^2(N-1)} \leq \sigma_i^2 \leq \frac{(N-1)s_i^2}{\chi_{1-\beta}^2(N-1)}, i = 1, 2, \dots, m,$$

where $\beta = \frac{1}{2}\alpha^{\frac{1}{m}}$. Note that the confidence region of $\sigma_i^2, i = 1, 2, \dots, m$, is a rectangular region.

From (2.4) and (2.6), the confidence region S_{α} of θ is given by

$$(2.7) \quad S_{\alpha} = \left\{ (\mu, \sigma^2) \left| \sum_{i=1}^m \frac{(\mu_i - \bar{\mu}_i)^2}{s_i^2} \leq \frac{m(N-1)}{N(N-m)} F_{\alpha}(m, N-m), \right. \right. \\ \left. \left. \frac{(N-1)s_i^2}{\chi_{\beta}^2(N-1)} \leq \sigma_i^2 \leq \frac{(N-1)s_i^2}{\chi_{1-\beta}^2(N-1)}, i = 1, 2, \dots, m \right\}.$$

The minimizing part of problem P_4 with respect to θ is given by

$$(2.8) \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \sum_{i=1}^m w_i \left(\sum_{j=1}^n a_{ij} x_j - t_i \right)^2 d\Phi(t_1, \dots, t_m; \mu, \sigma^2) \\ = \sum_{i=1}^m w_i \left\{ \left(\sum_{j=1}^n a_{ij} x_j - \mu_i \right)^2 + \sigma_i^2 \right\},$$

and then we obtain the following minimax problem.

$$P_5 : \left\{ \begin{array}{l} \text{Maximize}_{x \geq 0} \text{Minimize}_{(\mu, \sigma^2) \in S_{\alpha}} \sum_{j=1}^n c_j x_j \\ - \sum_{i=1}^m w_i \left\{ \left(\sum_{j=1}^n a_{ij} x_j - \mu_i \right)^2 + \sigma_i^2 \right\} \end{array} \right\}.$$

The minimizing part with respect to θ of problem P_5 is decomposed into the following two problems.

$$P'_5 : \left\{ \begin{array}{l} \text{Maximize}_{\sigma} L_1 = \sum_{i=1}^m w_i \sigma_i^2, \\ \text{subject to } \frac{(N-1)s_i^2}{\chi_{\beta}^2(N-1)} \leq \sigma_i^2 \leq \frac{(N-1)s_i^2}{\chi_{1-\beta}^2(N-1)}, i = 1, 2, \dots, m, \end{array} \right\}$$

$$P''_5 : \left\{ \begin{array}{l} \text{Maximize}_{\mu} L_2 = \sum_{i=1}^m w_i \left(\sum_{j=1}^n a_{ij} x_j - \mu_i \right)^2, \\ \text{subject to } \sum_{i=1}^m \frac{(\mu_i - \bar{\mu}_i)^2}{s_i^2} \leq K, \end{array} \right\}$$

where

$$(2.9) \quad K = \frac{m(N-1)}{N(N-m)} F_{\alpha}(m, N-m).$$

Problem P'_5 represents the maximization with respect to variances, and problem P''_5 represents the maximization with respect to means. Since the maximum of problem P'_5 is independent of a decision variable x and problem P_4 is a minimization problem with respect to x , we may regard L_1 as a constant in problem P_4 . Hence it is sufficient to consider only problem P''_5 .

2.3. A Solution Method

Here we will show several properties that are useful to solve problem P_4 induced from problem P_1 .

LEMMA 2.1. *The maximum of L_2 in problem P_5'' is attained on the boundary of feasible region, and the sign of $(\mu_i^* - \bar{\mu}_i)$ is opposite to that of $(A_i x - \bar{\mu}_i)$, where μ_i^* is an optimal solution of problem P_5'' .*

PROOF: It is easily shown that L_2 is a convex function of μ_i , $i = 1, 2, \dots, m$. Therefore the first part of Lemma 2.1 is clear. L_2 is maximized when it holds

$$(2.10) \quad \frac{(\mu_i^* - \bar{\mu}_i)^2}{s_i^2} = K_i^2, \quad i = 1, 2, \dots, m,$$

that is,

$$(2.11) \quad \mu_i^* = \bar{\mu}_i \pm s_i K_i, \quad i = 1, 2, \dots, m,$$

where $K_i \geq 0$ and $\sum_{i=1}^m K_i^2 = K$. Since $(A_i x - \mu_i^*)^2 = (A_i x - \bar{\mu}_i \mp s_i K_i)^2$, the maximand μ_i^* of L_2 is given as follows.

$$(2.12) \quad \mu_i^* = \begin{cases} \bar{\mu}_i - s_i K_i & \text{if } A_i x - \bar{\mu}_i > 0, \\ \bar{\mu}_i & \text{if } A_i x - \bar{\mu}_i = 0, \\ \bar{\mu}_i + s_i K_i & \text{if } A_i x - \bar{\mu}_i < 0, \end{cases}$$

for $i = 1, 2, \dots, m$. This proves the second part of Lemma 2.1. ■

By variable transformations $z_i = (\mu_i - \bar{\mu}_i)^2 / s_i^2$, $i = 1, 2, \dots, m$, problem P_5'' is transformed into the following problem.

$$P_6: \begin{cases} \text{Maximize}_z & L_2 = \sum_{i=1}^m w_i s_i^2 z_i + 2 \sum_{i=1}^m w_i s_i (A_i x - \bar{\mu}_i) \sqrt{z_i} \\ & + \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2, \\ \text{subject to} & \sum_{i=1}^m z_i \leq K, \\ & z_i \geq 0, \quad i = 1, 2, \dots, m. \end{cases}$$

The Kuhn-Tucker condition for problem P_6 is given by

$$(2.13) \quad \begin{aligned} & -w_i s_i^2 - w_i s_i (A_i x - \bar{\mu}_i) z_i^{-\frac{1}{2}} + \lambda + \xi_i = 0, \quad i = 1, 2, \dots, m, \\ & \sum_{i=1}^m z_i = K, \quad \xi_i z_i = 0, \quad \xi_i, z_i \geq 0, \quad i = 1, 2, \dots, m, \end{aligned}$$

where λ and ξ_i , $i = 1, 2, \dots, m$, are Lagrange multipliers. From Lemma 2.1, when $A_i x - \bar{\mu}_i = 0$, we obtain $z_i = 0$ and $\xi_i = w_i s_i^2 - \lambda$. While, when $A_i x - \bar{\mu}_i \neq 0$, we obtain

$$(2.14) \quad z_i = \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2}, \quad i = 1, 2, \dots, m,$$

and Lagrange multiplier λ satisfies

$$(2.15) \quad \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} = K.$$

Let λ^* denote the largest solution of equation (2.15). Then the following lemma shows that the optimand z_i^* , $i = 1, 2, \dots, m$, are given by

$$(2.16) \quad z_i^* = \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda^* - w_i s_i^2)^2}, \quad i = 1, 2, \dots, m.$$

LEMMA 2.2. *The objective function L_2 is maximized when λ in (2.14) is the largest solution of equation (2.15).*

PROOF: Substituting (2.14) into L_2 , we denote L_2 with $L_2(\lambda)$ as function of λ as follows.

$$(2.17) \quad L_2(\lambda) = \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2 \left(\frac{\lambda}{\lambda - w_i s_i^2} \right)^2.$$

Let λ_1 and $\lambda_2 (> \lambda_1)$ be two solutions of equation (2.15). Then

$$(2.18) \quad \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2 \frac{w_i s_i^2}{(\lambda_1 - w_i s_i^2)^2} = \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2 \frac{w_i s_i^2}{(\lambda_2 - w_i s_i^2)^2}.$$

Therefore,

$$(2.19) \quad (\lambda_1 + \lambda_2) \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda_1 - w_i s_i^2)^2 (\lambda_2 - w_i s_i^2)^2} = 2 \sum_{i=1}^m \frac{w_i^3 s_i^4 (A_i x - \bar{\mu}_i)^2}{(\lambda_1 - w_i s_i^2)^2 (\lambda_2 - w_i s_i^2)^2}.$$

Using (2.17) and (2.19), it is obtained that

$$(2.20) \quad L_2^*(\lambda_2) - L_2^*(\lambda_1) = \frac{(\lambda_2 - \lambda_1)^3}{2} \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda_1 - w_i s_i^2)^2 (\lambda_2 - w_i s_i^2)^2} > 0.$$

Hence the largest solution of equation (2.15) gives the maximum of $L_2(\lambda)$. ■

A lower bound of λ^* is given by the following lemma.

LEMMA 2.3. *It holds that*

$$(2.21) \quad \lambda^* > \lambda_0 \equiv \max_{i \in I} w_i s_i^2,$$

where

$$(2.22) \quad I = \{i \mid A_i x - \mu_i \neq 0, \quad i = 1, 2, \dots, m\}$$

and λ^* is the largest solution of equation (2.15).

PROOF: Related to (2.15), we define

$$(2.23) \quad G(\lambda) = \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} - K.$$

Since

$$(2.24) \quad \begin{aligned} \lim_{\lambda \rightarrow \lambda_0 + 0} G(\lambda) &= +\infty > 0, \\ \lim_{\lambda \rightarrow +\infty} G(\lambda) &= -K < 0, \end{aligned}$$

and

$$(2.25) \quad G'(\lambda) = -2 \sum_{i \in I} \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^3} > 0,$$

just one solution of equation (2.15) exists in an interval (λ_0, ∞) . So it is the largest solution of equation (2.15). ■

The optimal value L_2^* of problem P_5'' is given by

$$(2.26) \quad L_2^* = \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2 \left(\frac{\lambda}{\lambda - w_i s_i^2} \right)^2,$$

where it should be satisfied that

$$(2.27) \quad \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} = K,$$

and

$$(2.28) \quad \lambda > \lambda_0.$$

Problem P_5'' can also be solved by using a duality theorem of non-linear programming [I7] as shown below. We have that

$$(2.29) \quad \begin{aligned} \sup_{\mu} \left\{ L_2 = \sum_{i=1}^m w_i (A_i x - \mu_i)^2 \left| \sum_{i=1}^m \frac{(\mu_i - \bar{\mu}_i)^2}{s_i^2} \leq K \right. \right\} \\ = \inf_{\lambda} \left\{ \sup_{\mu} \{ L(\mu, \lambda) = \sum_{i=1}^m w_i (A_i x - \mu_i)^2 + \right. \\ \left. \lambda \left(K - \sum_{i=1}^m \frac{(\mu_i - \bar{\mu}_i)^2}{s_i^2} \right) \} \mid \lambda \geq 0 \right\}. \end{aligned}$$

Then, for $\lambda > \lambda_0$, the maximum of $L(\mu, \lambda)$ is attained at $\mu_i = \hat{\mu}_i$, $i = 1, 2, \dots, m$, where

$$(2.30) \quad \hat{\mu}_i = \frac{\lambda \bar{\mu}_i - w_i s_i^2 A_i x}{\lambda - w_i s_i^2}, \quad i = 1, 2, \dots, m,$$

and

$$(2.31) \quad L(\hat{\mu}, \lambda) = \lambda K + \sum_{i=1}^m \frac{\lambda w_i (A_i x - \bar{\mu}_i)^2}{\lambda - w_i s_i^2}.$$

From $dL(\hat{\mu}, \lambda)/d\lambda = 0$, we obtain that

$$(2.32) \quad K = \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2}$$

and

$$(2.33) \quad L(\hat{\mu}, \lambda) = \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2 \left(\frac{\lambda}{\lambda - w_i s_i^2} \right)^2.$$

Therefore, (2.26)–(2.28) are derived from (2.32), (2.33) and Lemma 2.3.

On the other hand, the optimal value L_1^* of problem P'_5 is given by

$$(2.34) \quad L_1^* = \sum_{i=1}^m \frac{(N-1)w_i s_i^2}{\chi_{1-\beta}^2(N-1)},$$

Then the objective function of problem P_5 is given by

$$(2.35) \quad L = \sum_{j=1}^n c_j x_j - L_1^* - L_2(\lambda^*),$$

and it should be maximized with respect to x . Since L_1^* is independent of a decision variable x , we neglect L_1^* . Problem P_5 is rewritten as follows.

$$P_7 : \begin{cases} \text{Maximize}_{x, \lambda} & L = \sum_{j=1}^n c_j x_j - \sum_{i=1}^m w_i (A_i x - \bar{\mu}_i)^2 \left(\frac{\lambda}{\lambda - w_i s_i^2} \right)^2, \\ \text{subject to} & \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} = K, \\ & \lambda > \lambda_0. \end{cases}$$

Note that $\lambda_0 = \max_{i \in I} w_i s_i^2$ depends on a decision variable x .

LEMMA 2.4. Problem P_7 is a convex programming problem if $\lambda \geq \frac{3}{2}\lambda_0$.

PROOF: It is sufficient to show that the Hessian matrix of the objective function is negative semi-definite for $\lambda \geq \frac{3}{2}\lambda_0$. The elements of Hessian matrix $H = (h_{ij})$, $i, j = 1, 2, \dots, n+1$, are given as follows.

$$(2.36a) \quad h_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j} = -2 \sum_{k=1}^m w_k a_{ki} a_{kj} \frac{\lambda^2}{(\lambda - w_k s_k^2)^2},$$

$i, j = 1, 2, \dots, n,$

$$(2.36b) \quad h_{n+1, i} = h_{i, n+1} = \frac{\partial^2 L}{\partial x_i \partial \lambda} = 4 \sum_{k=1}^m w_k a_{ki} (A_k x - \bar{\mu}_k) \frac{\lambda w_k s_k^2}{(\lambda - w_k s_k^2)^3}, \quad i = 1, 2, \dots, n,$$

$$(2.36c) \quad h_{n+1, n+1} = \frac{\partial^2 L}{\partial \lambda^2} = -2 \sum_{k=1}^m w_k (A_k x - \bar{\mu}_k)^2 \frac{w_k s_k^2 (2\lambda + w_k s_k^2)}{(\lambda - w_k s_k^2)^4}.$$

Then the matrix H is expressed as the sum of the following two matrices.

$$(2.37) \quad H = - \begin{pmatrix} U & v \end{pmatrix}' \begin{pmatrix} U & v \end{pmatrix} - \begin{pmatrix} 0 & \vdots & 0 \\ 0 & \dots & 0 & d \end{pmatrix},$$

where

$$(2.38) \quad \begin{aligned} U &= (u_{ij}) = \frac{\sqrt{2w_i \lambda}}{\lambda - w_i s_i^2} a_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n, \\ v &= (v_i) = -\frac{2\sqrt{2w_i} (A_i x - \bar{\mu}_i) w_i s_i^2}{(\lambda - w_i s_i^2)^2}, \quad i = 1, 2, \dots, m, \\ d &= 2 \sum_{k=1}^m \frac{(A_k x - \bar{\mu}_k)^2 w_k^2 s_k^2}{(\lambda - w_k s_k^2)^4} (2\lambda - 3w_k s_k^2). \end{aligned}$$

Since d is nonnegative for $\lambda \geq \frac{3}{2}\lambda_0$, $-H$ is expressed as the sum of two positive semi-definite matrices. Therefore, the Hessian matrix of the objective function of problem P_7 is positive semi-definite. ■

Because the value of λ depends on x , we should consider several cases according to the value of λ . We define

$$(2.39) \quad \lambda_{\max} = \max_{1 \leq i \leq m} w_i s_i^2.$$

Case A. $\lambda \geq \frac{3}{2}\lambda_{\max}$:

Since it is clear that $\lambda_0 \leq \lambda_{\max}$, problem P_7 is a convex programming problem from Lemma 2.4. Then we can use an appropriate solution method for a constrained nonlinear convex programming problem, e.g., the steepest descent method. ■

Case B. $\lambda_0 < \lambda < \frac{3}{2}\lambda_{\max}$:

For fixed λ , L is a convex function of x . Then we obtain optimal solution $x^*(\lambda)$ as a function of λ and choose the λ^* so as to minimize $x^*(\lambda)$. But we must divide λ into the following intervals, because the constraints differ on these intervals. For the sake of convenience, we arrange $w_i s_i^2$, $i = 1, 2, \dots, m$, in increasing order, together with the corresponding constraints. By definition of λ , we have

$$(2.40) \quad \lambda_0 = w_m s_m^2 \iff A_i x \neq \bar{\mu}_i, i = 1, 2, \dots, m,$$

and

$$(2.41) \quad \lambda_0 = w_p s_p^2 \iff \begin{cases} A_i x \neq \bar{\mu}_i, i = 1, 2, \dots, p, \\ A_i x = \bar{\mu}_i, i = p + 1, \dots, m, \end{cases}$$

for $1 \leq p < m$. Thus p denotes the largest index such that $A_i x \neq \bar{\mu}_i$. For $\lambda_0 = w_p s_p^2$, $1 \leq p < m$, the first constraint of problem P_7 is changed to

$$(2.42) \quad \sum_{i=1}^p \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} = K,$$

and equality constraints $A_i x = \bar{\mu}_i$, $i = p + 1, \dots, m$, are added. Then the constraints of problem P_7 for each p are shown as follows.

(i) $p < m - n$:

Since $\text{rank} A = n$, the constraints (2.41) are infeasible.

(ii) $m - n \leq p < m$:

$$(2.43) \quad \begin{aligned} \sum_{i=1}^p \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} &= K, \\ A_i x &= \bar{\mu}_i, \quad i = p + 1, \dots, m, \\ x_j &\geq 0, \quad j = 1, 2, \dots, n, \\ w_p s_p^2 &< \lambda < w_{p+1} s_{p+1}^2. \end{aligned}$$

(iii) $p = m$:

$$(2.44) \quad \begin{aligned} \sum_{i=1}^m \frac{w_i^2 s_i^2 (A_i x - \bar{\mu}_i)^2}{(\lambda - w_i s_i^2)^2} &= K, \\ x_j &\geq 0, \quad j = 1, 2, \dots, n, \\ w_m s_m^2 &< \lambda < \frac{3}{2}\lambda_{\max}. \end{aligned}$$

We need to obtain optimal solutions for all p . Choosing the best solution among them, we get a global optimal solution. ■

The following lemma shows the existence of the global optimal solution.

LEMMA 2.5. *The objective function L of problem P_7 is bounded.*

PROOF: Assume that an optimal solution is unbounded. Since the denominator of the first constraint of (2.43) or (2.44) is bounded, its left-hand side must be unbounded. This contradicts the boundedness of K . ■

We define problem $P_7(\lambda)$, which is minimized with respect to x , as problem P_7 for a fixed λ . Then the following lemma shows that an optimal solution of problem $P_7(\lambda)$ is continuous with respect to λ .

LEMMA 2.6. *An optimal solution of problem $P_7(\lambda)$ is continuous with respect to λ .*

PROOF: Let x_1 and x_2 denote optimal solutions of problems $P_7(\lambda_1)$ and $P_7(\lambda_2)$, respectively. It is easily shown that $L(x_1) \rightarrow L(x_2)$ as $\lambda_1 \rightarrow \lambda_2$ and that problem $P_7(\lambda)$ has a unique solution. Therefore, an optimal solution has the continuity with respect to λ . ■

Consequently, we can find an approximate optimal solution of problem P_7 by choosing the best solution among the optimal solutions of problem $P_7(\lambda)$ obtained for discretized values of λ between λ_0 and $\frac{3}{2}\lambda_{\max}$ in cases B. (ii) and B. (iii). Note that problem P_7 is a normal distribution model of problem P_4 , which is a minimax model due to the confidence region method for problem P_1 .

2.4. Asymptotic Properties

We discuss here how the problem behaves when sample size becomes sufficiently large. Since a sample mean and a sample variance

are consistent estimators for the mean and the variance of a normal distribution, the estimates of unknown parameters tend to the true values, if the number of independent samples tends to infinity. With true distribution parameters, simple quadratic recourse problem P_2 becomes the following problem.

$$P_8 : \quad \left| \begin{array}{l} \text{Maximize} \\ x \geq 0 \end{array} \right. \sum_{j=1}^n c_j x_j - \sum_{i=1}^m \{w_i (A_i x - \mu_i)^2 + w_i \sigma_i^2\},$$

Here we show that problem P_8 is obtained from our proposed problem when sufficiently large number of samples are available.

LEMMA 2.7. *Let $F_\alpha(M, N)$ and $\chi_\alpha^2(N)$ denote α percentiles of an F -distribution with (M, N) degrees of freedom and a χ^2 -distribution with N degrees of freedom, respectively. Then*

$$(2.45) \quad \lim_{N \rightarrow \infty} \frac{\chi_\alpha^2(N)}{N} = 1,$$

$$(2.46) \quad \lim_{N \rightarrow \infty} \frac{F_\alpha(M, N)}{N} = 0.$$

PROOF: As in [T1], using the Cornish-Fisher expansion, we have

$$(2.47) \quad \sqrt{2\chi_\alpha^2(N)} - \sqrt{2N-1} = u_\alpha + O\left(\frac{1}{\sqrt{N}}\right),$$

where u_α is an α percentile of a standard normal distribution. Since $|u_\alpha| < \infty$ for $0 < \alpha < 1$, (2.45) is derived.

The following relation holds between α percentiles of an F -distribution and a χ^2 -distribution.

$$(2.48) \quad M F_\alpha(M, \infty) = \chi_\alpha^2(M).$$

Since $F_\alpha(M, \infty)$ is finite, we obtain $\lim_{N \rightarrow \infty} F_\alpha(M, N)/N = 0$. ■

THEOREM 2.8. *Problem P_8 is obtained from problem P_5 as sample size tends to infinity.*

PROOF: From Lemma 2.7, optimal solutions of problems P'_5 and P''_5 tend to s_i^2 , $i = 1, 2, \dots, m$, and $\bar{\mu}_i$, $i = 1, 2, \dots, m$, respectively, as sample size tends to infinity. Therefore, the consistency of statistics $\bar{\mu}_i$ and s_i^2 implies that problem P_5 brings to problem P_8 as sample size tends to infinity. ■

Note that an optimal value λ^* in problem P_7 tends to infinity as sample size increases, when the linear programming problem

$$P_9 : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = \bar{\mu}, \\ & x \geq 0, \end{cases}$$

replacing the random variable b by the sample mean $\bar{\mu}$ is infeasible. It is considered that λ^* is an index of violation of constraints $A_i x = \bar{\mu}_i$, $i = 1, 2, \dots, m$. The estimates of variances are concerned only with the magnitude of recourse, which are independent of an optimal decision in both problems P_5 and P_8 .

2.5. An Illustrative Example

(Example 2.1.) Consider the following problem:

$$PE_1 : \begin{cases} \text{Minimize} & 2x_1 + x_2, \\ \text{subject to} & x_1 + x_2 = b_1, \\ & 2x_1 - x_2 = b_2, \\ & x_2 = b_3, \\ & x_1, x_2 \geq 0, \end{cases}$$

where b_1 , b_2 and b_3 are normally distributed random variables with unknown parameters, which are assumed to be independent each other.

Let the weight vector be given by $w = (10, 5, 10)$. As a result of sampling with 11 samples by computer simulation, the sample means $\bar{\mu}$ and the variances s^2 are obtained as given in Table 2.1, where the true means μ and the true variances σ^2 are also listed.

Table 2.1. Parameters of b_1 , b_2 and b_3

	$\bar{\mu}$	s^2	$w \cdot s^2$	μ	σ^2
b_1	2.979	0.007	0.070	3	0.01
b_2	0.056	0.360	1.800	0	0.36
b_3	1.020	0.043	0.430	1	0.04

The problem PE_1 is transformed into the following problem based on the confidence region method.

$$PE'_1 : \begin{cases} \text{Minimize}_{x_1, x_2, \lambda} & L = 2x_1 + x_2 + \frac{10(x_1 + x_2 - 2.979)^2 \lambda^2}{(\lambda - 0.070)^2} + \\ & \frac{5(2x_1 - x_2 - 0.056)^2 \lambda^2}{(\lambda - 1.800)^2} + \frac{10(x_2 - 1.020)^2 \lambda^2}{(\lambda - 0.430)^2}, \\ \text{subject to} & \frac{0.7(x_1 + x_2 - 2.979)^2}{(\lambda - 0.070)^2} + \frac{9(2x_1 - x_2 - 0.056)^2}{(\lambda - 1.800)^2} + \\ & \frac{4.3(x_2 - 1.020)^2}{(\lambda - 0.430)^2} = 1.388, \\ & x_1, x_2 \geq 0, \end{cases}$$

where a significance level is set to 5%.

Optimal solutions for all cases of p are given in Table 2.2. When $p = 3$, we solve problem PE'_1 for $1.80 < \lambda \leq 2.70$. When $p = 2$, the objective function L is minimized subject to

$$(2.49) \quad \begin{cases} \frac{0.7(x_1 + x_2 - 2.979)^2}{(\lambda - 0.070)^2} + \frac{4.3(x_2 - 1.020)^2}{(\lambda - 0.430)^2} = 1.388, \\ 2x_1 - x_2 = 0.056, \\ x_1 \geq 0, x_2 \geq 0, \end{cases}$$

Table 2.2. The optimal solutions for different p

Case	p	range of λ	λ^*	x_1^*	x_2^*	L^*
A.	—	$\lambda \geq 2.70$	2.700	0.736	1.710	16.809
B.	$p = 3$	$1.80 < \lambda \leq 2.70$	1.807	0.803	1.551	12.234
	$p = 2$	$0.43 < \lambda \leq 1.80$	1.471	0.821	1.587	13.232
	$p = 1$	$0.07 < \lambda \leq 0.43$	infeasible			

for $0.43 < \lambda \leq 1.80$. When $p = 1$, the objective function L is minimized subject to

$$(2.50) \quad \begin{aligned} \frac{0.7(x_1 + x_2 - 2.979)^2}{(\lambda - 0.070)^2} &= 1.388, \\ 2x_1 - x_2 &= 0.056, \\ x_2 &= 1.020, \\ x_1 \geq 0, x_2 &\geq 0, \end{aligned}$$

for $0.07 < \lambda \leq 0.43$, which have no infeasible point.

The best solution among these solutions is $(\lambda^*, x_1^*, x_2^*) = (1.807, 0.803, 1.551)$ and $L^* = 12.234$. Figure 2.1 illustrates an optimal solution of problem PE'_1 . Furthermore, in Table 2.3, the optimal solutions corresponding to several sample sizes are given, which illustrates the asymptotic properties discussed in Section 2.4. Note that $N = \infty$ implies that the estimates give the true values of parameters.

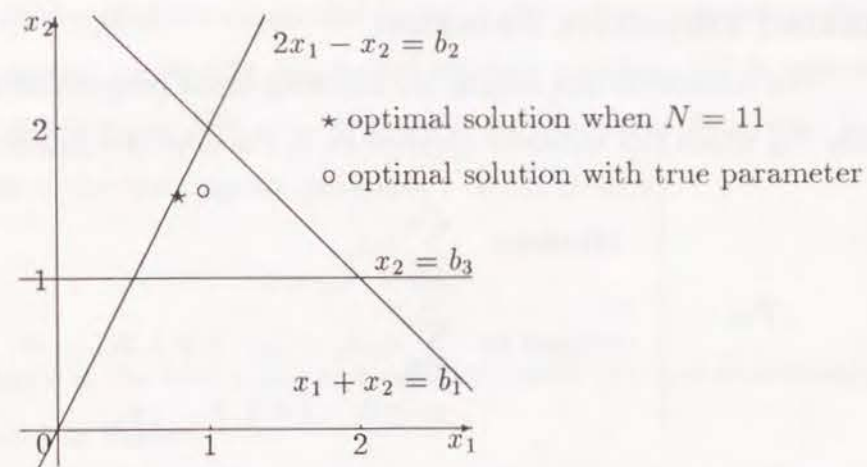


Figure 2.1. Optimal solutions of problem PE_1

Table 2.3. The best optimal solutions for several sample sizes

N	λ^*	x_1^*	x_2^*	L^*
11	1.807	0.803	1.551	12.234
100	5.870	0.825	1.710	11.355
300	8.642	0.820	1.649	10.640
1000	16.270	0.840	1.651	10.271
∞	—	0.967	1.580	9.557

2.6. Stochastic Linear Programming with an Estimated Objective Function

We consider in this section the following linear programming problem P_{10} which has unknown coefficients in the objective function:

$$P_{10} : \begin{cases} \text{Maximize} & \sum_{j=1}^n c_j x_j, \\ \text{subject to} & \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ & x_j \geq 0, \quad j = 1, 2, \dots, n, \end{cases}$$

where $A = (a_{ij})$ and $b = (b_i)$ are known coefficients, but $c = (c_j)$ are unknown coefficients which are estimated by noisy observations. We assume that $\text{rank } A = m$ and $\{x \mid Ax = b, x \geq 0\}$ is bounded and nonempty.

Problem P_{10} is viewed as the dual of problem P_1 . These problems treat two types of stochastic behaviors. In problem P_1 , there are normally distributed random variables with unknown parameters. On the other hand, in problem P_{10} , there are unknown coefficients which should be estimated only through the observations disturbed by normally distributed random errors.

We consider the situation in which the unknown coefficients c can not be directly observed, and so the objective function is estimated by means of linear regression analysis based on an output $y = c'x + \epsilon$ with a random disturbance ϵ for an input x . The normal linear regression model is expressed as follows.

$$(2.51) \quad y = Xc + \epsilon,$$

where X is an $N \times n$ matrix of N sample points, y is an N dimensional column vector of observations and ϵ is an N dimensional column vector of independent identically distributed random variables, which have a normal distribution $\mathcal{N}(0, \sigma^2)$. If there is no redundant data, that is, $\text{rank } X = n$, the least square estimator \hat{c} of c is given by

$$(2.52) \quad \hat{c} = (X'X)^{-1}X'y.$$

Note that \hat{c} is the best linear unbiased estimator of c and is normally distributed as follows.

$$(2.53) \quad \hat{c} \sim \mathcal{N}(c, \sigma^2(X'X)^{-1}).$$

Denoting the unbiased estimator of σ^2 with s^2 , we have

$$(2.54) \quad \begin{aligned} v_1 &= \frac{(c - \hat{c})'(X'X)(c - \hat{c})}{\sigma^2} \sim \chi^2(n), \\ v_2 &= \frac{(N - n)s^2}{\sigma^2} \sim \chi^2(N - n), \end{aligned}$$

which are mutually independent. So the following statistics has an F-distribution with $(n, N - n)$ degrees of freedom.

$$(2.55) \quad \frac{v_1/n}{v_2/(N - n)} = \frac{(c - \hat{c})'(X'X)(c - \hat{c})}{ns^2} \sim F(n, N - n)$$

Then we construct the joint confidence region of regression coefficients with a significance level α as follows.

$$(2.56) \quad (c - \hat{c})'(X'X)(c - \hat{c}) \leq ns^2 F_\alpha(n, N - n)$$

Note that this confidence region is an ellipsoid in R^n .

The unknown coefficients exist in the ellipsoidal confidence region with a significance level α . Then our confidence region method leads to a minimax problem that optimizes the objective function under the situation that estimated coefficients take the worst-case values in the confidence region. We obtain a minimax problem due to the confidence region method as follows.

$$P_{11} : \left\{ \begin{array}{l} \underset{c}{\text{Minimize}} \quad \underset{x}{\text{Maximize}} \quad c'x, \\ \text{subject to} \quad Ax = b, \\ \quad \quad \quad x \geq 0, \\ \quad \quad \quad (c - \hat{c})'(X'X)(c - \hat{c}) \leq K, \end{array} \right.$$

where $K = ns^2 F_\alpha(n, N - n)$.

2.7. A Solution Algorithm for Problem P_{10} and an Illustrative Example

Problem P_{11} is viewed as the following two-stage problem.

$$P_{12} : \left\{ \begin{array}{l} \underset{c}{\text{Minimize}} \quad \max\{c'x \mid Ax = b, \quad x \geq 0\}, \\ \text{subject to} \quad (c - \hat{c})'(X'X)(c - \hat{c}) \leq K. \end{array} \right.$$

Then we consider the following two problems $P'_{12}(c)$ and P''_{12} .

$$P'_{12}(c) : \left\{ \begin{array}{l} \text{Maximize} \quad c'x, \\ \text{subject to} \quad x \in \Omega_1 \equiv \{x \mid Ax = b, x \geq 0\}, \end{array} \right.$$

$$P''_{12} : \left\{ \begin{array}{l} \text{Minimize} \quad f(c), \\ \text{subject to} \quad c \in \Omega_2 \equiv \{c \mid (c - \hat{c})'(X'X)(c - \hat{c}) \leq K\}, \end{array} \right.$$

where $f(c)$ is the optimal value of problem $P'_{12}(c)$. Problem $P'_{12}(c)$ corresponds to the first stage problem to maximize with respect to x , where c is fixed, and problem P''_{12} corresponds to the second stage

problem to consider the worst-case values of the estimated coefficients. Since problem $P'_{12}(c)$ is a linear programming problem, it is sufficient to consider basic solutions for obtaining $f(c)$ for every c . We get the following lemmas concerning to these problems.

LEMMA 2.9. *The objective function $f(c)$ of problem P''_{12} is a piecewise linear convex function.*

PROOF: There are a finite number of basic feasible solutions for problem $P'_{12}(c)$. Denote them by $x_1^B, x_2^B, \dots, x_r^B$, where r is their number. Since the function $f(c)$ is expressed as the maximum of r linear functions of c , that is,

$$(2.57) \quad f(c) = \max\{(x_i^B)'c, \quad i = 1, 2, \dots, r\},$$

$f(c)$ is a piecewise linear convex function. ■

LEMMA 2.10. *Problem P''_{12} has a unique optimal solution.*

PROOF: From Lemma 2.9 and $x_i^B \geq 0$ for all i , the subgradient $\partial f(c)$ of $f(c)$ is nonnegative. Then any optimal solution of problem P''_{12} is located on the boundary of Ω_2 . Let c^1 and $c^2 (\neq c^1)$ denote optimal solutions of problem P''_{12} . Then

$$(2.58) \quad c^\lambda = \lambda c^1 + (1 - \lambda)c^2$$

for $0 < \lambda < 1$ is also feasible and is located in the interior of Ω_2 . The convexity of $f(c)$ implies that

$$(2.59) \quad f(c^\lambda) \leq \lambda f(c^1) + (1 - \lambda)f(c^2) = f(c^1) = f(c^2),$$

which means that c^λ is optimal and is not located on the boundary of Ω_2 . This contradicts that an optimal solution of problem P''_{12} is located on the boundary of Ω_2 . ■

An optimal solution of problem P_{12} is obtained from (2.57) by solving the following problem:

$$P_{13} : \begin{cases} \text{Minimize} & z, \\ \text{subject to} & (c - \hat{c})'(X'X)(c - \hat{c}) \leq K, \\ & c'x_i^B \leq z, \quad i = 1, 2, \dots, r. \end{cases}$$

However, r is a rather large number and it is not computationally feasible to get all basic feasible solutions in advance. To develop an effective solution algorithm, we define the following subproblem $P_{14}(x)$ with parameter x :

$$P_{14}(x) : \begin{cases} \text{Minimize} & x'c, \\ \text{subject to} & (c - \hat{c})'(X'X)(c - \hat{c}) \leq K. \end{cases}$$

Problem $P_{14}(x)$ is based on the fact that only one linear constraint of problem P_{13} is active when problem $P'_{12}(c)$ satisfies non-degeneracy assumption. An optimal solution of problem $P_{14}(x)$ is easily obtained by

$$(2.60) \quad c^*(x) = -\sqrt{\frac{K}{x'(X'X)^{-1}x}}(X'X)^{-1}x + \hat{c}$$

Given a basic feasible solution, we can decompose a matrix A into a basis B and a non-basis D , that is, $A = [B, D]$ and we denote correspondingly $x = [x_B, x_D]$ and $c = [c_B, c_D]$.

LEMMA 2.11. (x^*, c^*) is an optimal solution of problem P_{11} , if and only if, c^* is an optimal solution of problem $P_{14}(x^*)$ and satisfies the following optimality condition for linear programming problem $P'_{12}(c^*)$.

$$(2.61) \quad (c_D^*)' - (c_B^*)'B^{-1}D \leq 0$$

PROOF: The value of $(c^*)'x^*$ gives the minimum of $c'x^*$ over $c \in \Omega_2$ and the maximum of $(c^*)'x$ over $x \in \Omega_1$. Therefore, c^* is optimal to problem $P_{14}(x^*)$ and x^* is optimal to problem $P'_{12}(c^*)$. Then inequality (2.61) is satisfied from the nonnegativity of x . The 'only if' part is clear. ■

In the solution algorithm, problems $P'_{12}(c)$ and $P_{14}(x)$ are alternately solved until (2.61) is satisfied. To avoid cycling, we introduce a set J which stores optimal solutions of problems $P'_{12}(c)$ for every c . The following is a solution algorithm for problem P_{11} .

Algorithm 2.1.

- Step 1. Set $c^0 \leftarrow \hat{c}$, $k \leftarrow 0$ and $J \leftarrow \phi$.
- Step 2. Solve problem $P'_{12}(c^k)$ and let x^k be an optimal solution of problem $P'_{12}(c^k)$.
- Step 3. If $x^k \in J$, then go to Step 5. Otherwise, set $J \leftarrow J \cup \{x^k\}$ and go to Step 4.
- Step 4. Solve problem $P_{14}(x^k)$ and let c^{k+1} be an optimal solution of problem $P_{14}(x^k)$.

$$(2.62) \quad c^{k+1} = -\sqrt{\frac{K}{(x^k)'(X'X)^{-1}x^k}}(X'X)^{-1}x^k + \hat{c}$$

Let $k \leftarrow k + 1$ and return to Step 2.

Step 5. If $x^k = x^{k-1}$, then x^k is optimal to problem P_{11} and terminates. Otherwise, solve problem P''_{12} with $f(c) = \max\{(x^j)'c \mid x^j \in J\}$ and let c^{k+1} be an optimal solution of problem P''_{12} . Let $k \leftarrow k + 1$ and return to Step 2. ■

THEOREM 2.12. *If $x^k = x^{k-1}$ in Step 5 of Algorithm 2.1, then x^k is an optimal solution of problem P_{11} .*

PROOF: Since x^k is an optimal solution of problem $P'_{12}(c^k)$, we get

$$(2.63) \quad (c_D^k)' - (c_B^k)'B^{-1}D \leq 0,$$

from the optimality condition (2.61). If $x^k = x^{k-1}$, then problem $P_{14}(x^{k-1})$ is equivalent to problem $P_{14}(x^k)$. Therefore, c^k is optimal to problem $P_{14}(x^k)$. From Lemma 2.9, x^k is optimal to problem P_{11} . ■

THEOREM 2.13. *Algorithm 2.1 finds an optimal solution of problem P_{11} after a finite number of iterations.*

PROOF: Let $J_0 = \{x_i^B, i = 1, 2, \dots, r\}$ be a set of all basic feasible solutions of problem $P'_{12}(c)$. Note that J_0 is finite and $J \subseteq J_0$. In each iteration, J is augmented with a new solution. When $J = J_0$ holds, problem P''_{12} in Step 5 is equivalent to the original problem. Then from Lemma 2.8, Algorithm 2.1 gives an optimal solution in a finite number of iterations. ■

Now we show an illustrative example.

(Example 2.2.) Consider the following problem.

$$PE_2 : \begin{cases} \text{Maximize} & c_1 x_1 + c_2 x_2, \\ \text{subject to} & x_1 + 3x_2 \leq 15, \\ & x_1 + 2x_2 \leq 11, \\ & x_1 + x_2 \leq 8, \\ & 2x_1 + x_2 \leq 14, \\ & x_1, x_2 \geq 0. \end{cases}$$

Table 2.4 Computational result by Algorithm 2.1

k	c^k	x^k	J
0	(1.282, 1.694)'	(5, 3)'	ϕ
1	(0.977, 1.960)'	(3, 4)'	{(5, 3)'}
2	(1.514, 1.367)'	(6, 2)'	{(5, 3)', (3, 4)'}
3	(0.941, 2.032)'	(3, 4)'	{(5, 3)', (3, 4)', (6, 2)'}
4	(0.979, 1.958)'	(3, 4)', (5, 3)'	

The coefficients c_1 and c_2 of objective function are unknown and are estimated by noisy observations. By computer simulation with sample size $N = 20$ and a significance level $\alpha = 5\%$, we have obtained

$$(2.64) \quad \begin{aligned} X'X &= \begin{pmatrix} 190.0 & 165.0 \\ 165.0 & 157.5 \end{pmatrix}, & \hat{c} &= \begin{pmatrix} 1.282 \\ 1.694 \end{pmatrix}, \\ s^2 &= 0.2884, & F(2, 18; 0.05) &= 3.55, \\ \Omega_2 &= \{c \mid (c - \hat{c})'(X'X)(c - \hat{c}) \leq 2.048\}. \end{aligned}$$

Algorithm 2.1 applied to this problem PE_2 proceeds as shown in Table 2.4. It is concluded that optimal solutions are $c^* = (0.979, 1.958)'$ and $x^* = (5 - 2\lambda, 3 + \lambda)'$ for any $0 \leq \lambda \leq 1$, and the optimal value is 10.769.

2.8. Conclusion

The confidence region method for two types of stochastic programming problems are considered in this chapter. One is the problem with

a random right-hand side whose probability distribution includes unknown distribution parameters. The other is the problem with random cost coefficients whose stochastic behavior is introduced through the estimation of cost coefficients by randomly disturbed observations.

A proposed minimax model based on the confidence region method is useful and effective for the above two problems in the sense that the unknown parameters, which are estimated by historical or experimental data, are incorporated into the decision problem with an appropriate significance level. Note that the above two problems deal with the different stochastic aspects of parameters. These problems are viewed as dual problems to each other, since the dual of the problem with a random objective function becomes the problem with a random right-hand side and the dual of the problem with an estimated right hand side becomes the problem with estimated objective function. Therefore, for the problem with random cost coefficients whose distribution functions includes unknown parameters, we can solve its dual problem by using a proposed solution method in Section 2.3, and for the problem with unknown coefficients in the right-hand side of constraints, we can solve its dual problem by using Algorithm 2.1.

It is left for the further research to investigate the confidence region method for other problems, e.g., the problem with nonlinear constraint such as $f(x, \omega) \leq 0$ and the problem with distribution free stochasticity.

Chapter 3.

STOCHASTIC LINEAR PROGRAMMING WITH ESTIMATED CONSTRAINTS

3.1. Introduction

This chapter discusses the confidence region method for a stochastic linear programming problem that contains unknown coefficients in a coefficient matrix. Consider the following linear programming problem:

$$\text{LP: } \left\{ \begin{array}{l} \text{Maximize } \sum_{j=1}^n c_j x_j, \\ \text{subject to } \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ x_j \geq 0, \quad j = 1, 2, \dots, n, \end{array} \right.$$

where some of a_{ij} , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, are assumed to be unknown while others are known. The stochastic variation comes from estimation of unknown coefficients by means of the multiple regression analysis by noisy observations of $\sum_{j=1}^n a_{ij} x_j$, $i = 1, 2, \dots, m$. The confidence region of the left-hand side of constraints $\sum_{j=1}^n a_{ij} x_j$, $i = 1, 2, \dots, m$, becomes the intersection of reverse convex regions, which is the intersection of complement of convex regions and may

be disjunctive. Meyer [M4] have considered reverse convex programs and proposed an algorithm for finding a Kuhn-Tucker point when one feasible point is given in advance. Note that his iterative linearization algorithm gives a sequence of feasible points, but does not always converge to an optimal solution. The algorithms by a cutting plane method [T3] and by a combinatorial method [H1, U1] are also known for reverse convex programming. However, the convergence of a cutting plane method is not yet proved. For linear programming with one additional reverse convex constraint, the combinatorial algorithm, which finds an optimal solution in a finite number of iterations, is proposed by Hillestad and Jacobsen [H2], where they obtain all vertices of linear constraints by the lexicographic simplex algorithm and find an optimal solution by searching the edges of polytope defined by the linear constraints.

We mainly consider a linear programming problem, in which unknown coefficients appear in only one constraint. This problem is equivalently transformed into a linear programming problem with an additional reverse convex constraint. Our algorithm is based on a cutting plane method but finds an optimal solution of the reverse convex problem in a finite number of iterations. The proof is based on the fact that an optimal solution lies on the intersection of the boundary of reverse convex region and the edges of polytope constructed by known linear constraints.

Section 3.2 formulates the problem and estimates the unknown coefficients in the constraint. Several properties on an optimal solution of a reverse convex programming problem are summarized in Section 3.3.

Section 3.4 describes an iterative algorithm and proves its finite convergence to an optimal solution. This algorithm generates a point sequence converging to an optimal solution, but each point in the sequence before convergence is not feasible. A lower bound of the optimal value obtained at each iteration is discussed in Section 3.5. An illustrative example is given in Section 3.6. Section 3.7 extends the discussion to the problem, in which unknown coefficients appear in more than one constraint. Section 3.8 is a summary of this chapter.

3.2. Formulation of the Problem

We consider the following linear programming problem with one additional linear constraint that contains unknown coefficients.

$$P_1 : \left\{ \begin{array}{l} \text{Maximize} \quad \sum_{j=1}^n c_j x_j, \\ \text{subject to} \quad \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ \quad \quad \quad \sum_{j=1}^n \beta_j x_j = \eta, \\ \quad \quad \quad x_j \geq 0, \quad j = 1, 2, \dots, n, \end{array} \right.$$

where a_{ij} , b_i , c_j and η are assumed to be known coefficients, but some of β_j , $j = 1, 2, \dots, n$, are assumed to be unknown. By the confidence region method, constraint $\sum_{j=1}^n \beta_j x_j = \eta$ is interpreted as that η is restricted to the confidence interval of $\sum_{j=1}^n \beta_j x_j$ with a certain significance level. Let $CI(x; \alpha)$ denote a confidence interval of $\sum_{j=1}^n \beta_j x_j$ with a significance level α . Then problem P_1 is treated as the following problem with the constraint that η should be in the confidence

interval $CI(x; \alpha)$.

$$P_2 : \begin{cases} \text{Maximize} & \sum_{j=1}^n c_j x_j, \\ \text{subject to} & \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ & \eta \in CI(x; \alpha), \\ & x_j \geq 0, \quad j = 1, 2, \dots, n, \end{cases}$$

Multiple regression analysis is a useful technique to construct the confidence region of the linear constraint with unknown coefficients. We consider the following regression model

$$(3.1) \quad y = X\beta + \varepsilon,$$

where y is an N dimensional column vector of observations, X is an $N \times n$ matrix of n independent variables observed on each of N individuals, β is an n dimensional column vector of unknown regression parameters, and ε is an N dimensional column vector of unobserved random disturbances which are assumed to be mutually independent and normally distributed with the mean 0 and the variance σ^2 . When $\text{rank}(X) = n$, the least square estimator of the regression parameter β is given by

$$(3.2) \quad \hat{\beta} = (X'X)^{-1} X'y,$$

which is exactly the same as the maximum likelihood estimator of β and is the best linear unbiased estimator of β [M2]. Then the confidence region of the regression parameter is given by

$$(3.3) \quad (\beta - \hat{\beta})'(X'X)^{-1}(\beta - \hat{\beta}) \leq ns^2 F_\alpha(n, N - n),$$

where s^2 is the sample variance. The confidence interval of $\beta'x$ for a given x is

$$(3.4) \quad \hat{\beta}'x - \sqrt{nF_\alpha(n, N - n)}\sqrt{x'Vx} \leq \beta'x \leq \hat{\beta}'x + \sqrt{nF_\alpha(n, N - n)}\sqrt{x'Vx},$$

where $V = s^2(X'X)^{-1}$. The constraint $\eta \in CI(x; \alpha)$ is equivalent to the following two nonlinear inequalities.

$$(3.5) \quad \begin{aligned} -\hat{\beta}'x + \sqrt{nF_\alpha(n, N - n)}\sqrt{x'Vx} &\geq -\eta, \\ \hat{\beta}'x + \sqrt{nF_\alpha(n, N - n)}\sqrt{x'Vx} &\geq \eta. \end{aligned}$$

For a sake of simplicity, we denote them by

$$(3.6) \quad \begin{aligned} g^-(x) &\geq -\eta, \\ g^+(x) &\geq \eta, \end{aligned}$$

respectively. Since the regions $\{x|g^-(x) \leq -\eta\}$ and $\{x|g^+(x) \leq \eta\}$, which are defined by reversing the inequalities, are convex, the regions $\{x|g^-(x) \geq -\eta\}$ and $\{x|g^+(x) \geq \eta\}$ are called *reverse convex*.

3.3. Reverse Convex Programming

We summarize properties on reverse convex programming in this section. The reverse convex programming problem is defined by

$$P_3 : \begin{cases} \text{Minimize} & f(x), \\ \text{subject to} & g_i(x) \geq 0, \quad i = 1, 2, \dots, l, \end{cases}$$

where $f : R^n \rightarrow R$ is a differentiable and pseudo-concave function, and $g_i : R^n \rightarrow R^1, i = 1, 2, \dots, l$, are differentiable and pseudo-convex functions. It is assumed that

$$(3.7) \quad G = \{x|g_i(x) \geq 0, \quad i = 1, 2, \dots, l\}$$

is nonempty and compact. Let $\text{conv}G$ denote the convex hull of G . A vector $\bar{x} \in G$ is said to be a *vertex*, if \bar{x} cannot be written as a convex combination of other points of G , and let $V(G)$ denote the set of vertices of G . Then, since f is pseudo-concave, an optimal solution of the following problem P_4 belongs to $V(\text{conv}G)$,

$$P_4 : \begin{cases} \text{Minimize} & f(x), \\ \text{subject to} & x \in \text{conv}G, \end{cases}$$

and it holds that $V(\text{conv}G) = V(G) \subset G$, which is easily obtained by showing that $V(\text{conv}G) \subseteq V(G)$ and $V(\text{conv}G) \supseteq V(G)$. We see that it is sufficient to seek an optimal solution of problem P_3 among the vertices of G . Note that the number of vertices of G is finite and, as a consequence, $\text{conv}G$ is a convex polytope [H1].

As a special case of reverse convex programming, we also consider the following linear programming problem with one additional reverse convex constraint.

$$P_5 : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = b, \\ & g(x) \geq 0, \\ & x \geq 0, \end{cases}$$

where $g : R^n \rightarrow R$ is a continuous and pseudo-convex function. We define that

$$(3.8) \quad G \equiv \{x | g(x) \geq 0\},$$

$$(3.9) \quad D_0 \equiv \{x | Ax = b, x \geq 0\},$$

where D_0 is assumed to be nonempty and bounded. Let E_i , $i = 1, 2, \dots, K$, denote the edges of D_0 , where K is their number. Then

$\text{conv}(D_0 \cap G)$ is a convex polytope and an optimal solution of problem P_5 lies on E_i , $i = 1, 2, \dots, K$ [H2], since

$$(3.10) \quad V(\text{conv}(D_0 \cap G)) \subset \bigcup_i V(E_i \cap G) \subset \bigcup_i \{E_i \cap G\}.$$

Now we consider the problem

$$P_6 : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & x \in D_0, \end{cases}$$

which is obtained by dropping the reverse convex constraint. Let x_0 denote an optimal solution of problem P_6 . If $x_0 \in G$, x_0 is obviously an optimal solution of problem P_5 . Otherwise, x_0 satisfies the following inequalities:

$$(3.11) \quad \begin{aligned} x_0 &\in D_0, \\ g(x_0) &< 0, \\ c'x_0 &> c'\hat{x}, \end{aligned}$$

where \hat{x} is an optimal solution of problem P_5 . Let ∂G and $\text{int}G$ denote the boundary and the set of interior points of G , respectively. Let w denote the point where the line segment $[x, x_0]$ meets ∂G . Then w is given by

$$(3.12) \quad w = tx + (1-t)x_0$$

with $t \in (0, 1]$, which is uniquely determined by solving the equation

$$(3.13) \quad g(tx + (1-t)x_0) = 0.$$

Since

$$(3.14) \quad c'w = tc'x + (1-t)c'x_0 > c'x$$

for every $x \in D_0 \cap \text{int}G$, an optimal solution of problem P_5 lies on $D_0 \cap \partial G$ [T4].

Here we state problem P_2 again.

$$P_2 : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = b, \\ & g^-(x) \geq -\eta, \\ & g^+(x) \geq \eta, \\ & x \geq 0, \end{cases}$$

where $g^-(x)$ and $g^+(x)$ are convex functions defined by (3.5). The feasible region of problem P_2 is an intersection of the convex polytope D_0 and two reverse convex regions. Since $(g^-(x) + \eta) + (g^+(x) - \eta) \geq 0$ holds for any $x \in R^n$, at least one reverse convex constraint is always satisfied for any $x \in R^n$. Then it is sufficient to consider only one reverse convex constraint which is not satisfied by x_0 . Let x^* denote an optimal solution of problem P_2 . If $g^-(x_0) < -\eta$, it holds that $g^+(x^*) < \eta$, and if $g^+(x_0) < \eta$, it holds that $g^-(x^*) < -\eta$. Therefore, defining that

$$(3.15) \quad g(x) = \begin{cases} g^-(x) + \eta & \text{if } g^-(x_0) < -\eta, \\ g^+(x) - \eta & \text{if } g^+(x_0) < \eta, \end{cases}$$

problem P_2 is viewed as a linear programming problem with one additional reverse convex constraint.

3.4. A Solution Algorithm

We can get an optimal solution of problem P_2 , in principle, by computing all edges of D_0 and comparing the objective values of all points, at which edges E_i , $i = 1, 2, \dots, K$, of D_0 meet the boundary of G . In other words, an optimal solution of the following problem P_7 is also optimal to problem P_2 .

$$P_7 : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & x \in \text{conv}(D_0 \cap G). \end{cases}$$

The proposed algorithm is based on a cutting plane method, where the linear programming problems, generated by adding linear constraints to cut off some regions not in G from D_0 , are solved iteratively until an optimal solution is found. Here we assume that $x_0 \notin G$. An optimal solution is found at the point which is located on the intersection of some edges and boundary ∂G .

We construct sequences of cutting hyperplanes $H(x_k)$ and linear programming problems P^k given as follows:

$$P^k : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & x \in D_k, \end{cases}$$

where

$$(3.16) \quad D_k = D_{k-1} \cap H(x_{k-1})$$

for $k \geq 1$, when x_{k-1} is an optimal solution of problem P^{k-1} . Figure 3.1 illustrates these D_0 , D_k and G .

We show the way to construct the cutting hyperplane $H(x_k)$ as follows. For an optimal solution x_k of problem P^k , let $\text{adj}(x_k)$ denote a

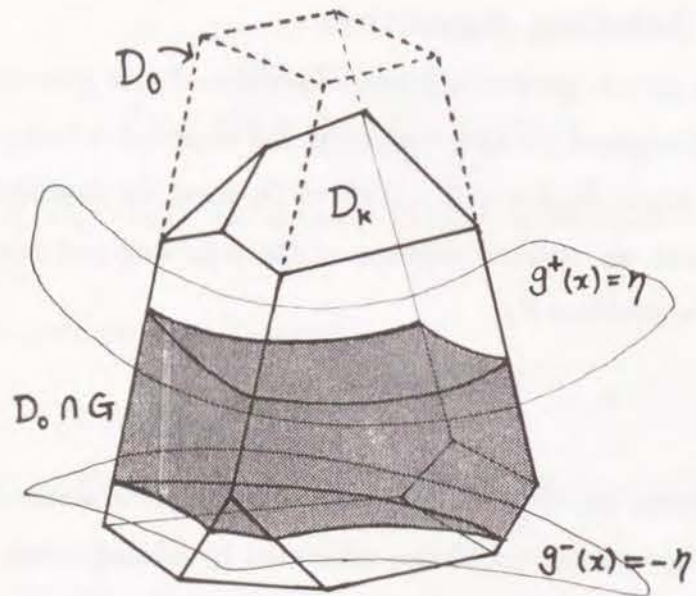


Figure 3.1. D_0, D_k and G

set of vertices of D_k adjacent to x_k . If x_k is a nondegenerate vertex, there are n affinely independent adjacent vertices of D_k . We denote the set of such vertices by

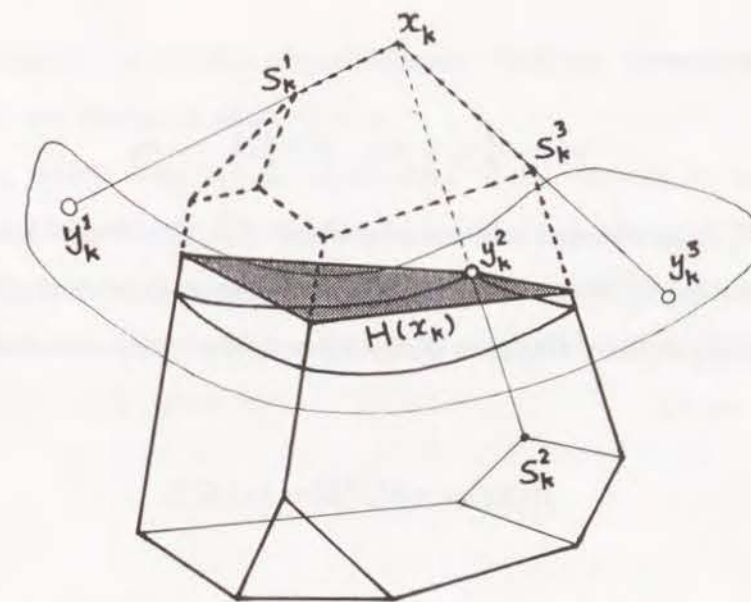
$$(3.17) \quad \text{adj}(x_k) = \{s_k^1, s_k^2, \dots, s_k^n\}.$$

See Figure 3.2. The n linearly independent vectors on D_k emanating from x_k are given by

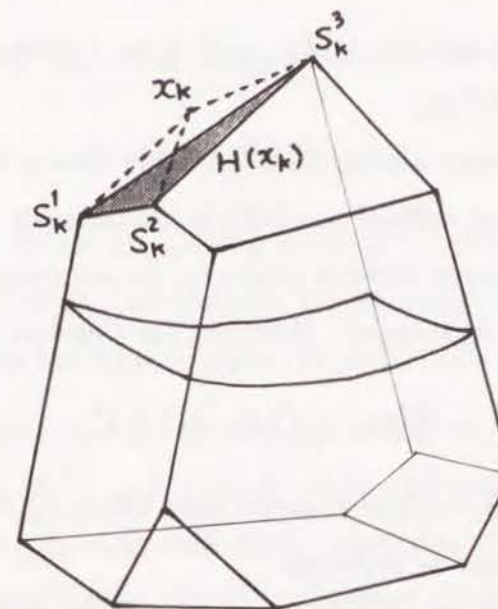
$$(3.18) \quad u_k^j = s_k^j - x_k, \quad j = 1, 2, \dots, n.$$

The point that the extension of edge $[x_k, s_k^j]$ meets to G , say y_k^j , is given by solving the following one-dimensional problems for each j and k :

$$P_8(j, k) : \begin{cases} \text{Minimize } \theta, \\ \text{subject to } g(\theta u_k^j + x_k) \geq 0. \end{cases}$$



(a) $H(x_k)$ of (3.20) when x_k lies on an edge of D_0



(b) $H(x_k)$ of (3.23) when x_k does not lie on an edge of D_0

Figure 3.2. The cutting hyperplane $H(x_k)$

Then we have

$$(3.19) \quad y_k^j = \bar{\theta}_k^j u_k^j + x_k, \quad j = 1, 2, \dots, n,$$

where $\bar{\theta}_k^j$ is an optimal solution of problem $P_8(j, k)$. If problem $P_8(j, k)$ is infeasible, $\bar{\theta}_k^j$ is set to an arbitrary number greater than 1. The cutting hyperplane $H(x_k)$ is defined from affinely independent $y_k^j, j = 1, 2, \dots, n$, by

$$(3.20) \quad H(x_k) : eM_k^{-1}(x - x_k) \leq 1,$$

where

$$(3.21) \quad M_k = [y_k^1 - x_k, y_k^2 - x_k, \dots, y_k^n - x_k]$$

is an $n \times n$ matrix and $e = (1, 1, \dots, 1)'$ is an n dimensional vector of ones. See Figure 3.2 (a).

If x_k is a degenerate vertex, there are more than n adjacent vertices of D_k . Then the cut defined by (3.20) is not uniquely determined and if we choose n adjacent vertices arbitrary, the resulting cut may delete portions of the feasible region. However, the Carvajal-Moreno cut

$$(3.22) \quad H(x_k) : z^*(x - x_k) \leq z^*$$

gives the cutting hyperplane in such a case, where (z_0^*, z^*) is an optimal solution of the following problem:

$$P_9 : \left\{ \begin{array}{l} \text{Minimize } z_0, \\ \text{subject to } \bar{\theta}_k^j u_k^j z \geq z_0, \quad j = 1, 2, \dots, n', \\ \sum_{j=1}^{n'} \bar{\theta}_k^j u_k^j z \leq 1, \end{array} \right.$$

where n' denote the number of such adjacent vertices. Note that (3.20) and (3.22) are the same when $n' = n$.

The x_k , which does not lie on an edge of D_0 , cannot be an optimal solution of problem P_2 . Since simplex $S(x_k, s_k^1, s_k^2, \dots, s_k^n)$ has no edge of D_0 and so has no optimal solution of problem P_2 , we construct the cutting hyperplane $H(x_k)$ by taking the base of simplex $S(x_k, s_k^1, s_k^2, \dots, s_k^n)$ given by

$$(3.23) \quad H(x_k) : e\bar{M}_k^{-1}(x - x_k) \leq 1,$$

where

$$(3.24) \quad \bar{M}_k = [s_k^1 - x_k, s_k^2 - x_k, \dots, s_k^n - x_k]$$

See Figure 3.2 (b). Since cut (3.23) deletes one vertex of D_k , we have

$$(3.25) \quad |V(D_k)| = |V(D_{k-1})| - 1.$$

THEOREM 3.1. *When an optimal solution x_k of problem P^k is not feasible to problem P_2 , the cutting hyperplane $H(x_k)$ defined above separates x_k from the feasible region of problem P_2 .*

PROOF: If x_k does not lie on an edge of D_k , $H(x_k)$ is the base of simplex $S(x_k, s_k^1, s_k^2, \dots, s_k^n)$ and all points of the simplex are located in $D_0 \setminus G$. Then it is clear that $H(x_k)$ separates x_k from $D_0 \cap G$. If x_k lies on an edge of D_k , $H(x_k)$ is defined by $y_k^j, j = 1, 2, \dots, n$, which is the base of simplex $S(x_k, y_k^1, y_k^2, \dots, y_k^n)$. Since $y_k^j \in \partial G$ for all j and G is a reverse convex region, any point that is expressed as a convex combination of $y_k^j, j = 1, 2, \dots, n$, is located outside of

```

procedure CPMEP1:
begin
  Set  $k \leftarrow 0$ ;
  Solve problem  $P^0$ ;
  while  $x_k \notin D$  do
    begin
      if  $x_k \in$  edges of  $D_0$ 
        then Construct  $H(x_k)$  of (3.20) or (3.22);
        else Construct  $H(x_k)$  of (3.23);
       $D_{k+1} \leftarrow D_k \cap H(x_k)$ ;
      Let  $k \leftarrow k + 1$ ;
      Solve problem  $P^k$ ;
    end
     $x_k$  is an optimal solution of problem  $P_2$ ;
  end.

```

Figure 3.3. Procedure CPMEP1

the region G . Therefore, simplex $S(x_k, y_k^1, y_k^2, \dots, y_k^n)$ is contained in $D_0 \setminus G$, and $H(x_k)$ separates x_k from $D_0 \cap G$. ■

Now we describe a solution procedure, named CPMEP1, to solve problem P_2 in Figure 3.3, which generates a point sequence $\{x_k\}$, $k = 1, 2, \dots$, of optimal solutions of linear programming problems P^k . Since $x_k \in D_k \setminus D_{k+1}$, the optimal value of problem P^k is non-increasing at each iteration, that is, $c'x_k \geq c'x_{k+1}$.

THEOREM 3.2. *The point sequence $\{x_k\}$ of optimal solutions has the subsequence $\{x_{k_q}\}$, $0 \leq k_1 < k_2 < \dots$, whose objective values strictly*

decrease, that is,

$$(3.26) \quad c'x_{k_q} > c'x_{k_{q+1}},$$

for any q .

PROOF: Let l denote the number of basic feasible solutions of problem P^k . Since the cutting hyperplane $H(x_k)$ cuts off at least one basic solution of problem P^k from D_k , at most l cutting hyperplanes cut off all such basic solutions. After adding l cutting hyperplanes to D_k , the optimal value of problem P^{k+l} is less than that of problem P^k , that is, $c'x_k > c'x_{k+l}$. This implies the existence of the strictly decreasing subsequence $\{x_{k_q}\}$ such that $c'x_{k_q} > c'x_{k_{q+1}}$. ■

From the fact of $(D_0 \cap G) \subseteq D_k$, the sequence $\{c'x_k\}$ of optimal values is bounded by $c'x^*$ from below. The following theorem shows the convergence of the sequence.

THEOREM 3.3. *The sequence $\{c'x_k\}$ of optimal values of problems P^k , $k = 1, 2, \dots$, generated by procedure CPMEP1 converges to the optimal value $c'x^*$ of problem P_2 after a finite number of iterations.*

PROOF: (Finiteness) There exists at most one point that is optimal to one of problems P^k , $k = 1, 2, \dots$, but is infeasible to problem P_2 on one edge of D_0 . Once the vertices, which are not located on an edge of D_0 , are selected as an optimal solution of problem P^k , they are cut off from D_k by $H(x_k)$ of (3.23). Therefore, we can obtain an optimal solution of problem P_2 on an edge of D_0 after a finite number of iterations. If the proposed algorithm iterates infinitely many times, there must be an

edge of D_0 which has more than one optimal solution of problem P^k . This is a contradiction.

(Convergence) The decreasing sequence $\{c'x_k\}$ of optimal values bounded from below converges to a certain value μ . Let μ be achieved at \tilde{x} , that is, $\mu = c'\tilde{x}$. When $\mu > c'x^*$, the point \tilde{x} is infeasible to problem P_2 . Theorem 3.2 implies that there exists the point whose objective value is less than μ as far as $\mu > c'x^*$ holds. Therefore we obtain $\mu \leq c'x^*$. It is clear that $\mu \geq c'x^*$ and hence $\mu = c'x^*$. ■

Problem P^{k+1} is effectively solved by a dual simplex method, since problem P^{k+1} is obtained by adding one linear constraint to problem P^k and an optimal solution x_k of problem P^k is available. The number of active constraints of D_0 in problem P^k tells us whether x_k lies on an edge of D_0 or not. In fact, denoting their number by m' , x_k lies on an edge of D_0 if and only if $m' \geq m - 1$.

3.5. A Lower Bound of the Optimal Value

Since x_k generated in our algorithm is not feasible to problem P_2 before converging to an optimal solution, we derive a lower bound of the optimal value by projecting x_k upon the feasible region $D_0 \cap G$. Let \underline{x}_k be a point that gives a lower bound of the optimal value at the k -th iteration.

First, we consider the straight line from x_k normal to the hyperplane $\hat{\beta}'x = \eta$, which denotes a regression line and is located in the center of confidence interval $CI(x; \alpha)$. Then we construct the hyperplane tangent to the surface $g(x) = 0$ as illustrated in Figure 3.4. This

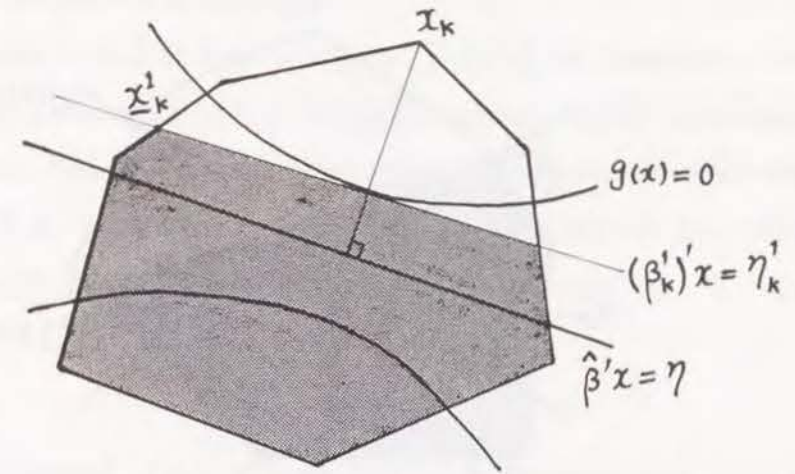


Figure 3.4. Tangent hyperplane (I)

hyperplane is defined by

$$(3.27) \quad (\beta_k^1)'x = \eta_k^1,$$

where

$$(3.28) \quad \begin{aligned} \beta_k^1 &= \nabla g(x_k + \alpha_k^1 \hat{\beta}), \\ \eta_k^1 &= (\beta_k^1)'(x_k + \alpha_k^1 \hat{\beta}), \\ \alpha_k^1 &= \{\alpha \mid g(x_k + \alpha \hat{\beta}) = 0\}. \end{aligned}$$

This is a linear approximation of the surface $g(x) = 0$ at the k -th iteration. Then we solve the following linear programming problem:

$$P_{10}^1(k) : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = b, \\ & (\beta_k^1)'x = \eta_k^1, \\ & x \geq 0, \end{cases}$$

and obtain an optimal solution \underline{x}_k^1 .

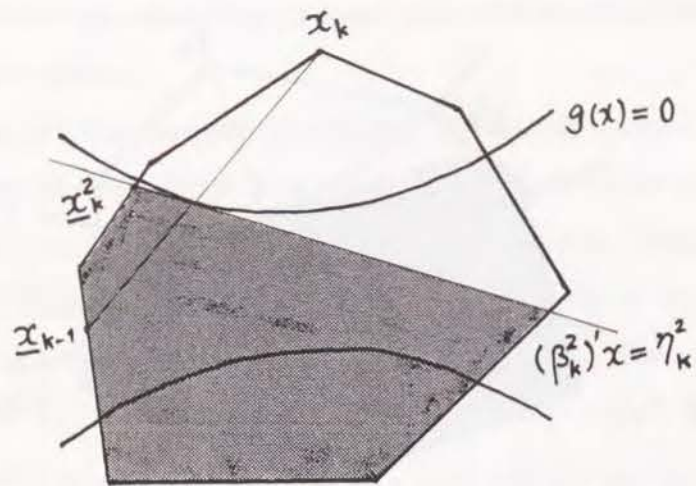


Figure 3.5. Tangent hyperplane (II)

Next, we construct the hyperplane tangent to the surface $g(x) = 0$ at an intersection point of the line segment $[x_k, \underline{x}_{k-1}]$ and the surface $g(x) = 0$ as illustrated in Figure 3.5. This hyperplane is defined by

$$(3.29) \quad (\beta_k^2)'x = \eta_k^2,$$

where

$$(3.30) \quad \begin{aligned} \beta_k^2 &= \nabla g(x_k + \alpha_k^2(\underline{x}_{k-1} - x_k)), \\ \eta_k^2 &= (\beta_k^2)'(x_k + \alpha_k^2(\underline{x}_{k-1} - x_k)), \\ \alpha_k^2 &= \{\alpha \mid g(x_k + \alpha(\underline{x}_{k-1} - x_k)) = 0\}. \end{aligned}$$

Then we solve the following linear programming problem.

$$P_{10}^2(k) : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = b, \\ & (\beta_k^2)'x = \eta_k^2, \\ & x \geq 0, \end{cases}$$

and obtain an optimal solution \underline{x}_k^2 .

Note that both of optimal solutions \underline{x}_k^1 and \underline{x}_k^2 are feasible to problem P_2 , since problems $P_{10}^1(k)$ and $P_{10}^2(k)$ are more strictly constrained than problem P_2 . The first point \underline{x}_k^1 reflects the estimated constraint $\hat{\beta}'x = \eta$ and the second point \underline{x}_k^2 reflects the previous point generating a lower bound. The updated lower bound is determined by

$$(3.31) \quad c'\underline{x}_k = \max[c'\underline{x}_{k-1}, c'\underline{x}_k^1, c'\underline{x}_k^2].$$

Then the sequence $\{c'\underline{x}_k\}$ of lower bounds is nondecreasing with an upper bound $c'x^*$. It is easily shown that $|c'x_k - c'x^*|$ tends to 0 for k sufficiently large, since x_k becomes identical to x^* after a finite number of iterations. If $x_k = x^*$, it holds $\underline{x}_k = x^*$.

When one feasible point is obtained, the following iterative scheme gives a point that satisfies the Kuhn-Tucker first order necessary conditions for problem P_2 [M4].

- (i) Choose $z_0 \in D_0 \cap G$.
- (ii) Given z_i , let z_{i+1} be an optimal solution of the following problem, which is obtained by the linearization of $g(x) = 0$ at $x = z_i$.

$$P(z_i) : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = b, \\ & g(z_i) + \nabla g(z_i)(x - z_i) \geq 0, \\ & x \geq 0. \end{cases}$$

By Meyer [M4, Theorem 2.1], the above procedure gives an accumulation point z^* of the monotonically increasing sequence $\{z_i\}$ and z^* is an optimal solution of problem $P(z^*)$. Then z^* satisfies the Kuhn-Tucker first order necessary conditions of problem P_2 . We shall say

that a feasible solution \underline{x}_k is an ϵ -optimal solution of problem P_2 when it satisfies $|c'x_k - c'\underline{x}_k| \leq \epsilon$ for a prescribed positive number ϵ . If we choose an ϵ -optimal solution for z_0 in the above scheme for a sufficiently small ϵ , it is considered that z^* is an optimal solution of problem P_2 . An optimal solution of problem P_2 is given by procedure CPMEP1 as well as by the above scheme. The former procedure generates a sequence of infeasible points and the later one generates a sequence of feasible ones.

3.6. An Illustrative Example

(Example 3.1.) Consider the following problem:

$$PE_1 : \begin{cases} \text{Maximize} & 4x_1 + 3x_2 + 13x_3, \\ \text{subject to} & 2x_1 + 3x_2 + 8x_3 \leq 4, \\ & 2x_1 + x_2 + 4x_3 \leq 3.2, \\ & 3x_1 + 8x_2 + 4x_3 \leq 3, \\ & \beta_1x_1 + \beta_2x_2 + \beta_3x_3 = 1, \\ & x_1, x_2, x_3 \geq 0, \end{cases}$$

where β_1 , β_2 and β_3 are unknown coefficients. Using the results of sampling by computer simulation with sample size $N = 10$ and a significance level $\alpha = 5\%$, we obtain the following estimates of β .

$$(3.32) \quad \hat{\beta} = (1.9959, 1.0193, 2.8677)$$

$$V = \begin{pmatrix} 1.0497 & -1.1185 & 0.3399 \\ -1.1185 & 2.2543 & -1.5573 \\ 0.3399 & -1.5573 & 2.4672 \end{pmatrix} \times 10^{-2}$$

$$F_{0.05}(3, 7) = 4.35$$

The computational results of procedure CPMEP1 in Section 3.4 are as follows.

(i) first iteration:

$$x_0 = (0.5000, 0.0000, 0.3750) : \text{infeasible}$$

$$s_0^1 = (0.0000, 0.1539, 0.4423), \quad y_0^1 = (-0.1266, 0.1928, 0.4594)$$

$$s_0^2 = (0.0000, 0.0000, 0.5000), \quad y_0^2 = (-0.0671, 0.0000, 0.5168)$$

$$s_0^3 = (1.0000, 0.0000, 0.0000), \quad y_0^3 = (1.8743, 0.0000, -0.6557)$$

$$H(x_0) : 1.2680x_1 + 1.0165x_2 + 2.0997x_3 \leq 1$$

(ii) second iteration:

$$x_1 = (0.0000, 0.0000, 0.4762) : \text{infeasible}$$

$$s_1^1 = (0.0000, 0.0000, 0.0000), \quad y_1^1 = (0.0000, 0.0000, 0.4578)$$

$$s_1^2 = (0.7886, 0.0000, 0.0000), \quad y_1^2 = (1.8744, 0.0000, -0.6558)$$

$$s_1^3 = (0.0000, 0.1806, 0.3888), \quad y_1^3 = (0.0000, 0.0000, 0.2703)$$

$$H(x_1) : 1.2977x_1 + 0.9629x_2 + 2.1843x_3 \leq 1$$

(iii) third iteration:

$$x_2 = (0.0000, 0.0000, 0.4578) : \text{feasible}$$

An optimal solution is, therefore, $x^* = x_2 = (0.0000, 0.0000, 0.4578)$ and the optimal value is 5.9514. This problem is solved by adding two cutting hyperplanes.

3.7. Extension to the Linear Programming Problem with more than One Additional Constraint Including Unknown Coefficients

We consider the following linear programming problem with more than one additional constraint including unknown coefficients.

$$P_{11} : \left\{ \begin{array}{l} \text{Maximize } \sum_{j=1}^n c_j x_j, \\ \text{subject to } \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ \sum_{j=1}^n \beta_{ij} x_j = \eta_i, \quad i = 1, 2, \dots, l, \\ x_j \geq 0, \quad j = 1, 2, \dots, n, \end{array} \right.$$

where matrix $B = (\beta_{ij})$ includes at least one unknown coefficient in each row. Let $CI_i(x; \alpha)$ denote a confidence interval of $\sum_{j=1}^n \beta_{ij} x_j$ under a significance level α . Then problem P_{11} is equivalent to the following problem with estimated constraints.

$$P_{12} : \left\{ \begin{array}{l} \text{Maximize } \sum_{j=1}^n c_j x_j, \\ \text{subject to } \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m, \\ \eta_i \in CI_i(x; \alpha), \quad i = 1, 2, \dots, l, \\ x_j \geq 0, \quad j = 1, 2, \dots, n, \end{array} \right.$$

Each constraint $\eta_i \in CI_i(x; \alpha)$ is equivalent to the following two reverse convex constraints

$$(3.33) \quad \begin{aligned} g_i^-(x) &\geq -\eta_i, \\ g_i^+(x) &\geq \eta_i, \end{aligned}$$

which are defined similarly to (3.6). We define

$$(3.34) \quad G = \bigcap_{i=1}^l G_i = \bigcap_{i=1}^l \{x \mid g_i(x) \geq 0\},$$

where

$$(3.35) \quad g_i(x) = \begin{cases} g_i^-(x) + \eta_i & \text{if } g_i^-(x_0) < -\eta_i, \\ g_i^+(x) - \eta_i & \text{if } g_i^+(x_0) < \eta_i. \end{cases}$$

An optimal solution of problem P_{12} is obtained by solving problem P_7 . The difficulty of this problem is that an optimal solution is not always attained on an edge of D_0 , but may be attained on an intersection point of nonlinear functions $g_i(x) = 0, i = 1, 2, \dots, l$.

Based on the Tuy's cuts, we construct the cutting hyperplanes as follows. Let I_k denote an index set of constraints violated by x_k , that is,

$$(3.36) \quad I_k = \{i \mid g_i(x_k) < 0, \quad i = 1, 2, \dots, l\}.$$

The point that the extension of edge $[x_k, s_k^j]$ meets to G_i for $i \in I_k$, say $y_{k,i}^j$, is given by solving the following one-dimensional problems for each i, j and k :

$$P_{13}(i, j, k) : \left\{ \begin{array}{l} \text{Minimize } \theta_{k,i}^j, \\ \text{subject to } g_i(\theta_{k,i}^j u_k^j + x_k) \geq 0, \end{array} \right.$$

where u_k^j is defined by (3.18). Then we have

$$(3.37) \quad y_{k,i}^j = x_k + \bar{\theta}_{k,i}^j u_k^j, \quad j = 1, 2, \dots, n,$$

where $\bar{\theta}_{k,i}^j$ denotes an optimal solution of problem $P_{13}(i, j, k)$. If x_k is a nondegenerate vertex of D_k , the cutting hyperplane $H(x_k; i)$, which cuts off the point x_k from the i -th reverse convex region G_i , is given by

$$(3.38) \quad H(x_k; i) : eM_{k,i}^{-1}(x - x_k) \leq 1,$$

where

$$(3.39) \quad M_{k,i} = [y_{k,i}^1 - x_k, y_{k,i}^2 - x_k, \dots, y_{k,i}^n - x_k]$$

is an $n \times n$ matrix. The degenerate vertex can be cut off by the Carvajal-Moreno cuts (3.22) for the violated constraints $g_i(x) \geq 0$, $i \in I_k$.

The cut $H(x_k)$ at the k -th iteration is defined by

$$(3.40) \quad H(x_k) = \bigcap_{i \in I_k} H(x_k; i),$$

which separates an optimal solution x_k of problem P^k from the feasible region D_k . Here we state problem P^k again.

$$P^k : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & x \in D_k, \end{cases}$$

The feasible region D_k of problem P^k is defined by (3.16) and it is shown that $D_{k+1} \supset (D_0 \cap G)$ for all k [H1]. As discussed in Section 3.4, we obtain a point sequence $\{x_k\}$ of optimal solutions of problems P^k , $k = 1, 2, \dots$. It is not, however, guaranteed that a point sequence $\{x_k\}$ of optimal solutions of problems P^k , $k = 1, 2, \dots$ converges to an optimal solution of problem P_{12} . When the cutting hyperplanes generated in these iterations become parallel to the contour lines of objective

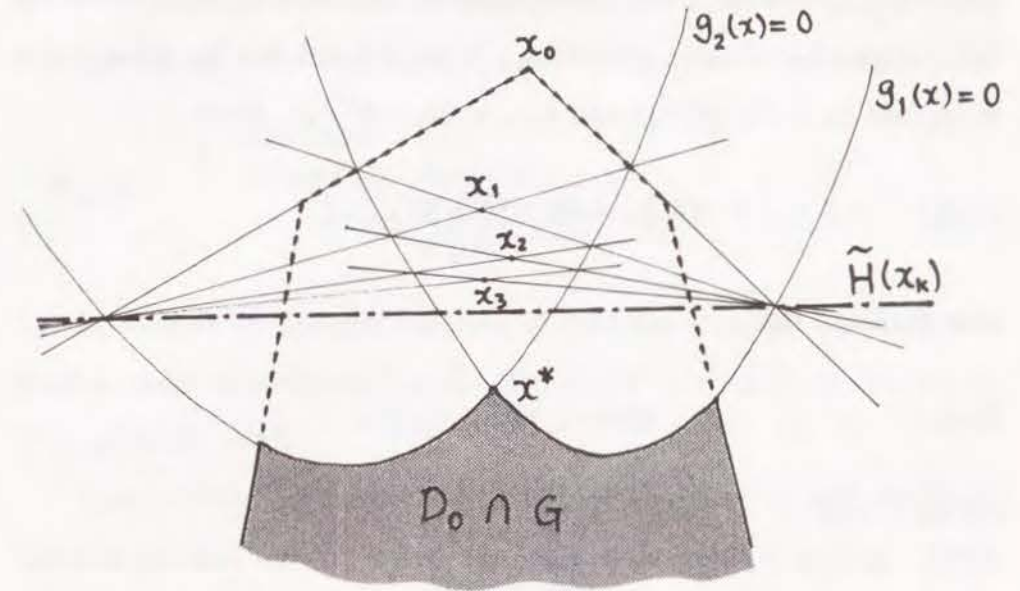


Figure 3.6. Convergence to an infeasible point

function, that is, $eM_{k,i}^{-1} \rightarrow c$ as $k \rightarrow \infty$, the amount of updated value $|c'x_k - c'x_{k+1}|$ tends to zero and the sequence $\{x_k\}$ may not reach a feasible point. Figure 3.6 illustrates the convergence to an infeasible point x^o . The cutting hyperplanes are becoming parallel to the contour of objective function. A point sequence $\{x_k\}$ of optimal solutions of problems P^k , $k = 1, 2, \dots$ converges to an infeasible point on such hyperplane.

THEOREM 3.4. *If a point sequence $\{x_k\}$ of optimal solutions of problems P^k , $k = 1, 2, \dots$ converges to an infeasible point x^o such that $c'x^o > c'x^*$, then the cutting hyperplane $H(x_k)$ becomes parallel to the contour of objective function.*

PROOF: Let $\epsilon > 0$. When a sequence $\{x_k\}$ converges to x^o , it holds $\|x_k - x_{k+1}\| < \epsilon$ for a sufficiently large k . Let $\text{dist}(x_k, H(x_k))$ denote the distance between x_k and $H(x_k)$. Then it holds $\text{dist}(x_k, H(x_k)) \leq \epsilon$ from that $x_k \in D_k \setminus H(x_k)$ and $x_{k+1} \in D_k \cap H(x_k)$. Since

$$(3.41) \quad D_{k+1} = D_k \cap \left[\bigcap_{i \in I_k} H(x_k; i) \right],$$

it is obtained that

$$(3.42) \quad \text{dist}(x_k, H(x_k; i)) \leq \epsilon$$

for all i , and

$$(3.43) \quad \begin{aligned} c'x_k &\geq c'y_{k,i}^j, \\ y_{k,i}^j &\in H(x_k; i) \end{aligned}$$

for all i and j . If the cutting hyperplane $H(x_k)$ does not become parallel to the contour of objective function, (3.42) and (3.43) imply that there exists at least one $y_{k,i}^j$ such that

$$(3.44) \quad \|x_k - y_{k,i}^j\| \leq \epsilon.$$

From $x_k \in D_0$ and $y_{k,i}^j \in G$, it is concluded that there is an accumulation point $x^o \in D_0 \cap G$. This contradicts that x^o is infeasible. ■

Now we define the distance between a point x_k and the feasible region by

$$(3.45) \quad d(x_k) = \sum_{i \in I_k} \|x_k - x_{k,i}^*\|,$$

where $x_{k,i}^*$, $i \in I_k$, is the point on $g_i(x) = 0$ nearest to x_k . Consider the following linear programming problem $P_{14}(k)$, which is the inner linearization of problem P_{12} at $x_{k,i}^*$, $i = 1, 2, \dots, l$.

$$P_{14}(k) : \begin{cases} \text{Minimize} & c'x, \\ \text{subject to} & Ax = b, \\ & \nabla g_i(x_{k,i}^*)'(x - x_{k,i}^*) \geq 0, \quad i = 1, 2, \dots, l, \\ & x \geq 0. \end{cases}$$

Let x_k^* denote an optimal solution of problem $P_{14}(k)$. Then x_k^* is a feasible point of problem P_{12} and $c'x_k^* \leq c'x^* \leq c'x_k$. Note that $x_k \rightarrow x^*$ implies $x_k^* \rightarrow x^*$.

The convergence is checked by $|c'x_k - c'x_{k+1}| \leq \epsilon_1$ and the feasibility is checked by $d(x_k) \leq \epsilon_2$ for sufficiently small ϵ_1 and ϵ_2 . Using these criteria, there are two cases in which this algorithm fails to obtain an optimal solution. One case is that the convergence speed is too slow even though the algorithm converges to an optimal solution. The other case is that the algorithm converges to an infeasible point and the cutting hyperplanes become parallel to objective function. Then $|c'x_k^* - c'x_k|$ tends to zero in the former situation but $|c'x_k^* - c'x_k|$ does not tend to zero in the latter situation.

When a point sequence $\{x_k\}$ of optimal solutions of problems P^k , $k = 1, 2, \dots$ does not go to an optimal solution of problem P_{12} , it is necessary to introduce an exceptional cutting hyperplane to restart the point sequence $\{x_k\}$ toward x^* . Assuming that the cutting hyperplanes become parallel to the contour of objective function, it holds

$$(3.46) \quad \left\| \max_{i,j} c'y_{k,i}^j - \min_{i,j} c'y_{k,i}^j \right\| \leq \epsilon,$$

and

$$(3.47) \quad \min_{i,j} c' y_{k,i}^j \leq c' x_{k+1} \leq \max_{i,j} c' y_{k,i}^j.$$

When $|c' x_k - c' x_{k+1}| < \epsilon_1$, we set the accumulation point x^o to x_k . When $d(x^o) < \epsilon_2$, we consider that x^o is feasible and the sequence of optimal solutions attains to an ϵ -optimal solution of problem P_{12} . If $d(x^o) \geq \epsilon_2$, we consider that x^o is infeasible. Then it holds that

$$(3.48) \quad c' x^o = \min_{i,j} c' y_{k,i}^j > c' x^*.$$

In this case, we construct an exceptional cutting hyperplane defined by

$$(3.49) \quad \tilde{H}(x_k) : c' x \leq c' x^o.$$

After adding the exceptional cutting hyperplane $\tilde{H}(x_k)$ to problem P^k , problem P^{k+1} with the feasible region $D_{k+1} = D_k \cap \tilde{H}(x_k)$ has an optimal solution x_{k+1} on the hyperplane $\tilde{H}(x_k)$. Note that

$$(3.50) \quad D_{k+1} = D_0 \cap \tilde{H}(x_k),$$

and at least one adjacent vertex s_{k+1}^j of D_{k+1} is a vertex of D_0 , since it holds $c' y_{k,i}^j > c' x^o$ for all i and j and $y_{k,i}^j$ are cut off by the hyperplane $\tilde{H}(x_k)$. Then we have that

$$(3.51) \quad \min_{i,j} c' y_{k+1,i}^j < c' x^o,$$

and the sequence of optimal values of problems P^k restarts from $c' x^o$. Figure 3.7 shows the resulting algorithm, named CPMEP2, to solve problem P_{12} .

procedure CPMEP2:

begin

Set $k \leftarrow 0$;

Give tolerance levels ϵ_1 and ϵ_2 ;

Let $\tilde{H}(x_0) = R^n$;

repeat

begin

Construct exceptional cut $\tilde{H}(x_k)$ of (3.49);

Let $D_{k+1} \leftarrow D_k \cap \tilde{H}(x_k)$ and solve problem P^k ;

Let $k \leftarrow k + 1$;

while $|c' x_k - c' x_{k-1}| > \epsilon_1$ **do**

begin

Construct $H(x_k)$ of (3.40);

Let $D_{k+1} \leftarrow D_k \cap H(x_k)$ and solve problem P^k ;

Let $k \leftarrow k + 1$;

end

end

until $d(x_k) \leq \epsilon_2$;

x_k is an optimal solution of problem P_{12} ;

end.

Figure 3.7. Procedure CPMEP2

Consider the sequence of exceptional cuts and let $\{x_i^o\}$ denote the sequence of corresponding accumulation points. If we assume that the exceptional cuts are generated by this algorithm infinitely many

times, it contradicts the finiteness of $V(D_0)$, since at least one vertex of D_0 is located between two any exceptional cuts. Therefore, the sequence $\{c'x_i^o\}$ is a monotonically decreasing sequence bounded by $c'x^*$ from below and this algorithm converges to an optimal solution of problem P_{12} .

Next we consider a lower bound of the optimal value and a sequence of feasible solutions of problem P_{12} . We find that an optimal solution x_k^* of the inner linearized problem P_{14} is a feasible solution of problem P_{12} and $c'x_k^*$ is a lower bound of the optimal value. Then consider the following scheme.

- (i) Let $z_0 = x_k^* \in D_0 \cap G$.
- (ii) Given z_i , let z_{i+1} be an optimal solution of the following linearized problem, which is obtained by linearizing $g_j(x) = 0, j = 1, 2, \dots, l$, at $x = z_i$.

$$P'(z_i) : \begin{cases} \text{Maximize} & c'x, \\ \text{subject to} & Ax = b, \\ & g_j(z_i) + \nabla g_j(z_i)(x - z_i) \geq 0, \\ & \qquad \qquad \qquad j = 1, 2, \dots, l, \\ & x \geq 0. \end{cases}$$

When we obtain x_k^* which is very close to an optimal solution, that is, $|c'x_k^* - c'x_k| \leq \epsilon$ for a sufficiently small ϵ , this iterative scheme is considered to construct a sequence of feasible solutions converging to an optimal solution of problem P_{12} .

3.8. Conclusion

The confidence region method for a stochastic linear programming problem that contains unknown coefficients in a coefficient matrix is

considered in this chapter. The unknown coefficients are estimated by means of regression analysis based on noisy observations. The constraints including unknown coefficients are replaced by confidence regions derived under a certain significance level. The confidence region of such constraints are reverse convex and our problem is regarded as a linear programming problem with additional reverse convex constraints.

We have mainly considered a linear programming problem in which only one constraint includes unknown coefficients. In such a problem, an optimal solution lies among the points that intersect edges of polytope constructed by the known linear constraints to the boundary of reverse convex region. A cutting plane method is proposed by utilizing the Tuy's cut. We show that the algorithm finds an optimal solution of the reverse convex programming problem in a finite number of iterations. Since the algorithm generates a sequence of infeasible solutions converging to an optimal solution of problem P_2 , a lower bound of the optimal value is also considered at each iteration. A lower bound of the optimal value gives a sequence of *feasible* solutions of problem P_2 according to optimal solutions of problems $P^k, k = 1, 2, \dots$.

For the linear programming problem in which more than one constraint includes unknown coefficients, an optimal solution may be located on the intersection of the boundaries of reverse convex regions. Therefore, the algorithm for the former problem may converge to an infeasible point. We discuss how to find the convergence to an infeasible point and propose a modified algorithm to restart the point sequence toward an optimal solution. A lower bound of the optimal value is also obtained at each iteration by solving the linearized problem.

Chapter 4.

A CONFIDENCE REGION METHOD FOR A STOCHASTIC LINEAR KNAPSACK PROBLEM

4.1. Introduction

The linear knapsack problem has been well studied and some efficient algorithms of $O(n)$ computational time are known, if all coefficients are known, where n is the number of items. This chapter discusses the confidence region method for a stochastic linear knapsack problem. A stochastic linear knapsack problem with random cost coefficients is one of the generalized versions of a linear knapsack problem. We consider the following P-model of a stochastic linear knapsack problem introduced by Kataoka [K3], which is to find an optimal solution minimizing the goal value subject to a chance constraint that the total cost is less than the goal value with probability more than α_0 .

$$P_1 : \left\{ \begin{array}{l} \text{Maximize } f, \\ \text{subject to } \Pr\left(\sum_{j=1}^n c_j x_j \geq f\right) \geq \alpha_0, \\ \sum_{j=1}^n a_j x_j = b, \\ 0 \leq x_j \leq b_j, \quad j = 1, 2, \dots, n, \end{array} \right.$$

where it is assumed that

- (i) $a_j > 0, b_j > 0$ for $j = 1, 2, \dots, n$, and $\sum_{j=1}^n a_j b_j \geq b > 0$,
- (ii) $1/2 < \alpha_0 \leq 1$,
- (iii) $c_j, j = 1, 2, \dots, n$, are mutually independent random variables which have a normal distributions $\mathcal{N}(\hat{\mu}_j, \hat{\sigma}_j)$, $j = 1, 2, \dots, n$, respectively.

Problem P_1 is solved parametrically in at most $O(n^2 \log n)$ computational time by the algorithm of Ishii and Nishida [I2].

In usual stochastic programming, the distribution parameters of random variable are known in advance. However, the confidence region method for stochastic programming discussed in the previous chapters assumes that the distribution parameters of random variable are unknown. We shall show that the game theoretic minimax model based on the confidence region method for a stochastic linear knapsack problem can be solved in a polynomial time by reducing it to a problem with known distribution parameters.

Section 4.2 summarizes the solution procedure for problem P_1 with known distribution parameters. By introducing the auxiliary problem with a positive parameter, problem P_1 is solved in at most $O(n^2 \log n)$ computational time [I2]. Section 4.3 discusses the minimax model based on the confidence region method for problem P_1 with unknown distribution parameters, where unknown distribution parameters are restricted to the confidence region. It is shown that this problem is also solved by the same procedure as proposed in Section 4.2. Section 4.4 summarizes this chapter.

4.2. A Stochastic Linear Knapsack Problem

We summarize the solution procedure for problem P_1 . Problem P_1 is equivalent to the following deterministic problem:

$$P_2 : \begin{cases} \text{Maximize} & \sum_{j=1}^n \hat{\mu}_j x_j - K_\alpha \sqrt{\sum_{j=1}^n \hat{\sigma}_j^2 x_j^2}, \\ \text{subject to} & \sum_{j=1}^n a_j x_j = b, \\ & 0 \leq x_j \leq b_j, \quad j = 1, 2, \dots, n, \end{cases}$$

where K_α is the α -percentile of a standard normal distribution. By variable transformations

$$(4.1) \quad y_j = a_j x_j, \quad \gamma_j = a_j b_j, \quad \mu_j = \hat{\mu}_j / a_j, \quad \sigma_j^2 = K_\alpha^2 \hat{\sigma}_j^2 / a_j^2,$$

for $j = 1, 2, \dots, n$, problem P_2 is further transformed into the following problem:

$$P_3 : \begin{cases} \text{Maximize} & \sum_{j=1}^n \mu_j y_j - \sqrt{\sum_{j=1}^n \sigma_j^2 y_j^2}, \\ \text{subject to} & \sum_{j=1}^n y_j = b, \\ & 0 \leq y_j \leq \gamma_j, \quad j = 1, 2, \dots, n. \end{cases}$$

Now we introduce the following auxiliary problem P_3^R with a positive parameter R .

$$P_3^R : \begin{cases} \text{Maximize} & R \sum_{j=1}^n \mu_j y_j - \frac{1}{2} \sum_{j=1}^n \sigma_j^2 y_j^2, \\ \text{subject to} & \sum_{j=1}^n y_j = b, \\ & 0 \leq y_j \leq \gamma_j, \quad j = 1, 2, \dots, n. \end{cases}$$

The following theorem shows a relation between problems P_3 and P_3^R .

THEOREM 4.1. Let y^* and y^R denote optimal solutions of problems P_3 and P_3^R , respectively. If

$$(4.2) \quad R = \sqrt{\sum_{j=1}^n \sigma_j^2 (y_j^R)^2},$$

then it holds that $y^* = y^R$.

PROOF: Since problems P_3 and P_3^R are convex programming problems with a single linear constraint and upper bounded variables, optimal solutions of problems P_3 and P_3^R satisfy the following Kuhn-Tucker conditions (4.3) and (4.4), respectively.

$$(4.3) \quad \begin{aligned} & \frac{\sigma_j^2 y_j}{\sqrt{\sum_{j=1}^n \sigma_j^2 y_j^2}} - \mu_j + u_j - v_j + \lambda = 0, \quad j = 1, 2, \dots, n, \\ & v_j y_j = 0, \quad u_j (y_j - \gamma_j) = 0, \quad j = 1, 2, \dots, n, \\ & \sum_{j=1}^n y_j = b, \quad 0 \leq y_j \leq \gamma_j, \quad u_j, v_j \geq 0, \quad j = 1, 2, \dots, n, \end{aligned}$$

$$(4.4) \quad \begin{aligned} & \sigma_j^2 y_j - R \mu_j + u_j - v_j + \lambda = 0, \quad j = 1, 2, \dots, n, \\ & v_j y_j = 0, \quad u_j (y_j - \gamma_j) = 0, \quad j = 1, 2, \dots, n, \\ & \sum_{j=1}^n y_j = b, \quad 0 \leq y_j \leq \gamma_j, \quad u_j, v_j \geq 0, \quad j = 1, 2, \dots, n, \end{aligned}$$

where $\lambda, u_j, v_j, j = 1, 2, \dots, n$, are Lagrange multipliers. Let $(y^*, u^*, v^*, \lambda^*)$ and $(y^R, u^R, v^R, \lambda^R)$ denote solutions of (4.3) and (4.4), respectively. Then if $R = \sqrt{\sum_{j=1}^n \sigma_j^2 (y_j^R)^2}$, it holds that

$$(4.5) \quad y_j^* = y_j^R, \quad u_j^* = u_j^R / R, \quad v_j^* = v_j^R / R, \quad \lambda^* = \lambda^R / R.$$

Therefore y^R is also an optimal solution of problem P_3 . ■

Let R^* denote the R satisfying $R = \sqrt{\sum_{j=1}^n \sigma_j^2 (y_j^R)^2}$. Theorem 4.1 implies that y^{R^*} is an optimal solution of problem P_3 . Then we define $T(R)$ as a function of R .

$$(4.6) \quad T(R) = R - \sqrt{\sum_{j=1}^n \sigma_j^2 (y_j^R)^2}$$

Since $T(R)$ is continuous and non-decreasing function of R , there exists the unique R^* that satisfies (4.2). Furthermore, it holds that

$$(4.7) \quad \begin{aligned} T(R) < 0 & \iff R < R^*, \\ T(R) = 0 & \iff R = R^*, \\ T(R) > 0 & \iff R > R^*. \end{aligned}$$

The solutions of (4.4) are given as functions of λ as follows.

$$(4.8) \quad \begin{aligned} u_j^R(\lambda) &= \max(0, R\mu_j - \lambda - \sigma_j^2 \gamma_j), \\ v_j^R(\lambda) &= \max(0, \lambda - R\mu_j), \\ y_j^R(\lambda) &= \begin{cases} \gamma_j & \text{if } \lambda \leq -\gamma_j \sigma_j^2 + R\mu_j, \\ \frac{R\mu_j - \lambda}{\sigma_j^2} & \text{if } -\gamma_j \sigma_j^2 + R\mu_j < \lambda \leq R\mu_j, \\ 0 & \text{if } \lambda > R\mu_j, \end{cases} \end{aligned}$$

for $j = 1, 2, \dots, n$, and

$$(4.9) \quad \sum_{j=1}^n y_j^R(\lambda) = b.$$

Since $y_j^R(\lambda)$ is a non-increasing function of λ , y_j^R is found by solving equation $\sum_{j=1}^n y_j^R(\lambda) = b$. We sort $-\gamma_j \sigma_j^2 + R\mu_j$ and $R\mu_j, j = 1, 2, \dots, n$, in non-decreasing order and denote their different values by

$$(4.10) \quad \Lambda_0^R \equiv -\infty < \Lambda_1^R < \Lambda_2^R < \dots < \Lambda_{N_1}^R < \Lambda_{N_1+1}^R \equiv +\infty,$$

```

procedure AUX( $R$ ):
begin
   $i \leftarrow 0$ ;
   $j \leftarrow N_1 + 1$ ;
  repeat
    begin
       $l \leftarrow \lceil (i + j)/2 \rceil$ ;
       $\lambda \leftarrow \Lambda_l^R$ ;
      if  $\sum_{j=1}^n y_j^R(\lambda) > b$  then  $i \leftarrow l$ ;
      if  $\sum_{j=1}^n y_j^R(\lambda) < b$  then  $j \leftarrow l$ 
    end
  until  $\sum_{j=1}^n y_j^R(\lambda) = b$  or  $j - i = 1$ ;
  find  $\lambda^R$  such that  $\sum_{j=1}^n y_j^R(\lambda) = b$ ;
   $y^R \leftarrow y^R(\lambda)$ ;
end.

```

Figure 4.1. Solution procedure for auxiliary problem P_3^R

where N_1 denotes their number which is at most $2n$. $y_j^R(\lambda)$ is determined by (4.8), for each interval $\lambda \in (\Lambda_k^R, \Lambda_{k+1}^R)$, $k = 0, 1, \dots, N_1$.

Now we give an algorithm $AUX(R)$ in Figure 4.1 to solve problem P_3^R , which finds an optimal solution y^R of problem P_3^R in at most $O(n \log n)$ computational time. To find R^* , we obtain the values of positive R such that $\Lambda_i^R = \Lambda_j^R$ for all pairs i and j with $i \neq j$. We sort such R 's in non-decreasing order and denote their different values by

$$(4.11) \quad R_0 \equiv 0 < R_1 < R_2 < \dots < R_{N_2} < R_{N_2+1} \equiv +\infty,$$

where N_2 denotes their number which is at most $\frac{1}{2}n(3n-1)$. For each interval of R , the order of Λ_j^R , $j = 1, 2, \dots, n$, is uniquely determined and problem P_3^R is solved by algorithm $AUX(R)$. We denote (R_{i^*}, R_{i^*+1}) the interval to which R^* belongs. For $\lambda \in (\Lambda_k^R, \Lambda_{k+1}^R)$, compute the λ satisfying $\sum_{j=1}^n y_j^R(\lambda) = b$, where R is contained as a parameter and we denote it by $\lambda(R)$. Note that the order of Λ_j^R , $j = 1, 2, \dots, n$, is the same for all $R \in (R_{i^*}, R_{i^*+1})$. Then we have

$$(4.12) \quad R^* \in \{R | \Lambda_k^R \leq \lambda(R) \leq \Lambda_{k+1}^R\} \cap (R_{i^*}, R_{i^*+1}),$$

and let $\bar{S}_k = (R_k^L, R_k^U)$ denote the closure of a set in (4.12). If $\bar{S}_k = \phi$, λ^* is not located in interval $(\Lambda_k^R, \Lambda_{k+1}^R)$ and we search other intervals of λ for λ^* . Otherwise, we have $\lambda^* \in (\Lambda_k^R, \Lambda_{k+1}^R)$ and $R^* \in (R_k^L, R_k^U)$, so R^* is obtained by solving

$$(4.13) \quad R^* - \sqrt{\sum_{j=1}^n \sigma_j^2 y_j^{R^*}(\lambda)} = 0.$$

Figure 4.2 is an algorithm to solve problem P_3 , which finds an optimal solution y^* of problem P_3 in at most $O(n^2 \log n)$ computational time.

4.3. A Confidence Region Method for a Stochastic Linear Knapsack Problem

We consider the problem in which normally distributed random cost coefficients have unknown distribution parameters. Based on estimation through noisy observations, we derive a confidence region in which the parameters are restricted under a certain significance level.

procedure SLKP:

begin

$i \leftarrow 0;$

$j \leftarrow N_2 + 1;$

repeat

begin

$l \leftarrow [(i + j)/2];$

AUX(R_l);

if $T(R_l) > 0$ then $i \leftarrow l;$

if $T(R_l) < 0$ then $j \leftarrow l$

end;

until $T(R_l) = 0$ or $j - i = 1;$

if $T(R_l) = 0$ then $y^* \leftarrow y^{R_l};$

else

for every $k = 0, 1, 2, \dots, N_1$ do

if $\bar{S}_k \neq \phi$ then

begin

AUX(R_k^L);

AUX(R_k^U);

if $T(R_k^L) = 0$ then $y^* \leftarrow y^{R_k^L};$

if $T(R_k^U) = 0$ then $y^* \leftarrow y^{R_k^U};$

if $T(R_k^L) < 0$ and $T(R_k^U) > 0$ then $y^* \leftarrow y^{R^*};$

end;

y^* is an optimal solution of problem $P_3;$

end.

Figure 4.2. Solution algorithm for problem P_3

The minimax model based on the confidence region method, discussed in this section, gives a minimax solution that optimizes the objective function under the possible worst-case values of estimated parameters.

Now consider problem P_3 , which is equivalent to problem P_1 by variable transformations (4.1). The confidence region of distribution parameters has been given in Section 2.2. Then the minimax problem P_4 is given as follows.

$$P_4 : \begin{cases} \underset{y}{\text{Maximize}} \underset{(\mu, \sigma) \in S_\alpha}{\text{Minimize}} & \sum_{j=1}^n \mu_j y_j - \sqrt{\sum_{j=1}^n \sigma_j^2 y_j^2}, \\ \text{subject to} & \sum_{j=1}^n y_j = b, \\ & 0 \leq y_j \leq \gamma_j, \quad j = 1, 2, \dots, n, \end{cases}$$

where

$$(4.14) \quad S_\alpha = \left\{ (\mu, \sigma) \left| \sum_{j=1}^n \frac{(\mu_j - \bar{\mu}_j)^2}{s_j^2} \leq \frac{n(N-1)}{N(N-n)} F_\alpha(n, N-n), \right. \right. \\ \left. \left. \frac{(N-1)s_j^2}{\chi_{\beta}^2(N-1)} \leq \sigma_j^2 \leq \frac{(N-1)s_j^2}{\chi_{1-\beta}^2(N-1)}, \quad j = 1, 2, \dots, n \right\}.$$

Optimal solutions σ_j^* , $j = 1, 2, \dots, n$, of the minimizing part of problem P_4 with respect to σ_j , $j = 1, 2, \dots, n$, are given by

$$(4.15) \quad \sigma_j^* = \sqrt{\frac{N-1}{\chi_{1-\beta}^2(N-1)}} s_j^2, \quad j = 1, 2, \dots, n,$$

from the nonnegativity of y_j . By substituting (4.15) into problem P_4 with variable transformations

$$(4.16) \quad \xi_j = \frac{\mu_j - \bar{\mu}_j}{s_j}, \quad j = 1, 2, \dots, n,$$

the minimizing part of problem P_4 with respect to $\mu_j, j = 1, 2, \dots, n$, becomes the following problem with a quadratic constraint.

$$P_5 : \begin{cases} \text{Maximize}_{\xi} & \sum_{j=1}^n s_j y_j \xi_j + \sum_{j=1}^n \bar{\mu}_j y_j + \sqrt{\frac{N-1}{\chi_{1-\beta}^2(N-1)}} \sqrt{\sum_{j=1}^n s_j^2 y_j^2}, \\ \text{subject to} & \sum_{j=1}^n \xi_j^2 \leq K, \\ & \xi_j \geq 0, j = 1, 2, \dots, n, \end{cases}$$

where

$$(4.17) \quad K = \frac{n(N-1)}{N(N-n)} F_{\alpha}(n, N-n).$$

Now we introduce the following subsidiary problem $P_5(\theta)$ with a parameter θ .

$$P_5(\theta) : \begin{cases} \text{Minimize} & \frac{1}{2} \sum_{j=1}^n \xi_j^2, \\ \text{subject to} & \sum_{j=1}^n s_j y_j \xi_j \geq \theta, \\ & \xi_j \geq 0, j = 1, 2, \dots, n. \end{cases}$$

Let $\xi(\theta) = (\xi_1(\theta), \xi_2(\theta), \dots, \xi_n(\theta))$ denote an optimal solution of problem $P_5(\theta)$. Then the following theorem describes a relation between problems P_5 and $P_5(\theta)$.

THEOREM 4.2. *If $\xi(\theta)$ satisfies*

$$(4.18) \quad \sum_{j=1}^n \xi_j^2(\theta) = K$$

and

$$(4.19) \quad \sum_{j=1}^n s_j y_j \xi_j(\theta) = \theta,$$

then $\xi(\theta)$ is an optimal solution of problem P_5 .

PROOF: Since problems P_5 and $P_5(\theta)$ are convex programming problems, optimal solutions of problems P_5 and $P_5(\theta)$ satisfy the following Kuhn-Tucker conditions (4.20) and (4.21), respectively.

$$(4.20) \quad \begin{aligned} \nu_j &= \frac{1}{2} \eta \xi_j - s_j y_j, \quad j = 1, 2, \dots, n, \\ \sum_{j=1}^n \xi_j^2 &\leq K, \quad \xi_j, \nu_j \geq 0, \quad \nu_j \xi_j = 0, \quad j = 1, 2, \dots, n, \end{aligned}$$

$$\eta \left(\sum_{j=1}^n \xi_j^2 - K \right) = 0.$$

$$(4.21) \quad \begin{aligned} \nu_j &= \xi_j - \eta s_j y_j, \quad j = 1, 2, \dots, n, \\ \sum_{j=1}^n s_j y_j \xi_j &\geq \theta, \quad \xi_j, \nu_j \geq 0, \quad \nu_j \xi_j = 0, \quad j = 1, 2, \dots, n, \end{aligned}$$

$$\eta \left(\theta - \sum_{j=1}^n s_j y_j \xi_j \right) = 0,$$

where $\eta, \nu_j, \xi_j, j = 1, 2, \dots, n$, are Lagrange multipliers. Let (ξ^*, ν^*, η^*) and $(\xi(\theta), \nu(\theta), \eta(\theta))$ denote solutions of (4.20) and (4.21), respectively. Then (ξ^*, ν^*, η^*) is constructed from $(\xi(\theta), \nu(\theta), \eta(\theta))$ as follows.

$$(4.22) \quad \xi^* = \xi(\theta), \quad \nu^* = \nu(\theta)/\eta(\theta), \quad \eta^* = 2/\eta(\theta),$$

where $\eta(\theta) > 0$ from (4.19) and (4.21). The conditions $\sum_{j=1}^n \xi_j^2 \leq K$ and $\eta(\sum_{j=1}^n \xi_j^2 - K) = 0$ are assured by (4.18). Therefore $\xi(\theta)$ is an optimal solution of problem P_5 . ■

From (4.18), (4.19) and (4.21), an optimal solution ξ^* of problem P_5 is given by

$$(4.23) \quad \xi_j^* = \sqrt{\frac{K}{\sum_{l=1}^n s_l^2 y_l^2}} s_j y_j, \quad j = 1, 2, \dots, n.$$

Then problem P_4 becomes the following problem.

$$P_6 : \begin{cases} \text{Maximize} & \sum_{j=1}^n \bar{\mu}_j y_j - K' \sqrt{\sum_{j=1}^n s_j^2 y_j^2}, \\ \text{subject to} & \sum_{j=1}^n y_j = b, \\ & 0 \leq y_j \leq \gamma_j, \quad j = 1, 2, \dots, n, \end{cases}$$

where

$$(4.24) \quad K' = \sqrt{K} + \sqrt{\frac{N-1}{\chi_{1-\beta}^2(N-1)}}.$$

Problem P_6 is essentially the same as problem P_3 and so we can solve problem P_6 by algorithm SLKP.

THEOREM 4.3. *Assuming that the square root of a number can be calculated in $O(1)$ computational time, problem P_4 can be solved in $O(n^2 \log n)$ computational time.*

PROOF: It is clear that K' of (4.24) can be calculated in $O(1)$ computational time. Once problem P_4 is transformed into problem P_6 , it can be solved in $O(n^2 \log n)$ computational time by algorithm SLKP. ■

4.4. Conclusion

We introduced a minimax model based on the confidence region method for a stochastic linear knapsack problem with unknown distribution parameters. In this model, unknown parameters are restricted in the confidence region and an optimal decision is sought to optimize the objective function by assuming the possible worst-case values of such parameters within the confidence region. The proposed minimax problem is transformed to the problem with known distribution parameters, which is then solved effectively by algorithm SLKP in at most $O(n^2 \log n)$ computational time.

This parametric type algorithm is applicable to other nonlinear programming problems, which have quadratic objective function or constraint. Such problems often occur in the E-V model and the P-model of stochastic programming. In this sense, it is worthwhile to investigate the case in which random variables are not independent each other. Some generalization of this problem, which is related to a portfolio selection problem as a probability maximizing model of a stochastic linear knapsack problem, are discussed in the following chapters.

Chapter 5.

A PROBABILITY MAXIMIZING MODEL OF A STOCHASTIC LINEAR KNAPSACK PROBLEM

5.1. Introduction

This chapter discusses a probability maximizing model of the stochastic linear knapsack problem P_1 .

$$P_1 : \left\{ \begin{array}{l} \text{Maximize } \Pr \left(\sum_{j=1}^n c_j x_j \geq d \right), \\ \text{subject to } \sum_{j=1}^n a_j x_j = b, \\ 0 \leq x_j \leq \gamma_j, \quad j = 1, 2, \dots, n, \end{array} \right.$$

where $c = (c_j)$ is assumed to be a normally distributed random vector, $x = (x_j)$ is an allocation vector and $a = (a_j)$, $\gamma = (\gamma_j)$, b and d are known coefficients. This problem finds an optimal solution which maximizes the probability that an objective value exceeds a prescribed goal value. Ishii and Nishida [I3] have considered the case in which the random cost coefficients c_1, c_2, \dots, c_n are assumed to be mutually independent. They derive an equivalent deterministic fractional problem with a single constraint and a lower and upper bounded allocation

vector, and propose a solution algorithm which finds an optimal solution in at most $O(n^4)$ computational time by introducing two auxiliary problems.

A portfolio selection problem is a mathematical formulation of a finance model to allocate a given amount of money among several investments for obtaining a large return under a small risk. Markowitz [M3] formulates the portfolio selection problem as maximizing the expected utility under uncertainty, where the portfolio is selected by taking into account both risk and return on investment. When we invest a money to some investments, we can obtain some return after a certain time period. However, the amount of return on investment is uncertain and the uncertainty of each investment is different from each other. There are several mathematical models to predict the return on investment, e.g., a capital asset pricing model (CAPM) and an arbitrage pricing theory (APT), which provide the distribution of the rate of return on investment. Therefore it necessitates the investigation of probabilistic and statistical approaches for the portfolio selection problem.

We consider the one-period portfolio selection problem. Since the return on investment is assumed to be a random variable, there are two conflicting objectives that the expectation of total return should be maximized and the variance of total return (i.e., risk) should be minimized. An efficient frontier of admissible portfolios, where the variance of total return for a fixed expected total return is minimized, is obtained in the mean variance framework. Szegö [S4] considers the properties of an efficient frontier with one riskless asset.

This chapter deals mainly with a probability maximizing model. When the investor is aiming for total return in advance, he manages to achieve his goal. Then, he wants to find the portfolio that maximizes the probability of achieving his goal rather than one that maximizes the total return.

A probability maximizing model of portfolio selection problem P_1 with a normality assumption of return is equivalent to a fractional programming problem with the squared root of a quadratic form in the mean variance framework. Since the returns on investment are not considered to be mutually independent, it is necessary to investigate the problem with correlated random coefficients, as a generalization of the problem with the independent random cost coefficients. Section 5.2 starts with a probability maximizing model of the stochastic linear knapsack problem as a main problem, and formulates an equivalent problem that maximizes the ratio of excess return to standard deviation. We introduce two auxiliary problems, one of which is a quadratic programming problem with positive parameters R and q . An optimality condition that an optimal solution of the auxiliary problem is an optimal solution of the main problem is then derived. Section 5.3 shows that an optimal solution of the main problem is geometrically found on the efficient frontier in the mean variance framework and that an optimal parameter value R^* uniquely exists. The relation between an EV model and a probability maximizing model is also noted. Section 5.4 gives a summary of Chapter 5.

5.2. A Probability Maximizing Model of a Stochastic Linear Knapsack Problem

We consider the following probability maximizing model of a stochastic linear knapsack problem.

$$P_1 : \begin{cases} \text{Maximize} & \Pr \left(\sum_{j=1}^n c_j x_j \geq d \right), \\ \text{subject to} & \sum_{j=1}^n a_j x_j = b, \\ & 0 \leq x_j \leq \gamma_j, \quad j = 1, 2, \dots, n, \end{cases}$$

where $c = (c_j)$ is a random vector with an n -dimensional normal distribution $\mathcal{N}(\hat{\mu}, \hat{V})$. We assume that $a = (a_j)$, $\hat{\gamma} = (\hat{\gamma}_j)$ and b are positive and satisfy $\sum_{j=1}^n a_j \hat{\gamma}_j \geq b$ for the sake of feasibility of problem P_1 . Moreover we assume at least one feasible solution x of problem P_1 satisfies $\hat{\mu}'x \geq d$, that is, $\Pr(c'x \geq d) \geq \frac{1}{2}$. This assumption is satisfied for $d \leq \max\{\hat{\mu}'x\}$. When $\Pr(c'x \geq d) < \frac{1}{2}$ for any feasible x , there is no practical meaning, because it holds that $\Pr(c'x^1 \geq d) \leq \Pr(c'x^2 \geq d)$ for $c'x^1 = c'x^2$ and $(x^1)'Vx^1 \leq (x^2)'Vx^2$, which implies a large variance is preferred to a small variance.

Then problem P_1 is equivalent to the following deterministic fractional problem.

$$P_2 : \begin{cases} \text{Maximize} & \frac{\hat{\mu}'x - d}{\sqrt{x'\hat{V}x}}, \\ \text{subject to} & a'x = b, \\ & 0 \leq x \leq \hat{\gamma}. \end{cases}$$

After the following variable transformations

$$(5.1) \quad y_j = \frac{a_j x_j}{b}, \quad \mu_j = \frac{\hat{\mu}_j b}{a_j}, \quad \sigma_{ij} = \frac{\hat{\sigma}_{ij} b^2}{a_i a_j}, \quad \gamma_j = \frac{\hat{\gamma}_j b}{a_j},$$

Table 5.1

Problem	Optimal solution	Optimal value
P_3	y^*	q^*
P^q	y^q	Z^q
$P^q(R)$	$y^q(R)$	$Z^q(R)$

problem P_2 is further transformed into

$$P_3 : \begin{cases} \text{Maximize} & \frac{\mu'y - d}{\sqrt{y'Vy}}, \\ \text{subject to} & e'y = 1, \\ & 0 \leq y \leq \gamma, \end{cases}$$

where $\hat{\mu} = (\hat{\mu}_j)$, $\mu = (\mu_j)$, $\hat{V} = (\hat{\sigma}_{ij})$, $V = (\sigma_{ij})$ and $e = (1, 1, \dots, 1)'$ with a suitable dimension.

We introduce two auxiliary problems P^q and $P^q(R)$ with positive parameters q and R :

$$P^q : \begin{cases} \text{Maximize} & \mu'y - q\sqrt{y'Vy}, \\ \text{subject to} & e'y = 1, \\ & 0 \leq y \leq \gamma, \end{cases}$$

$$P^q(R) : \begin{cases} \text{Maximize} & R\mu'y - \frac{q}{2}y'Vy, \\ \text{subject to} & e'y = 1, \\ & 0 \leq y \leq \gamma. \end{cases}$$

We shall denote an optimal solution and the optimal value of each problem as described in Table 5.1.

Since problems P^q and $P^q(R)$ are convex programming problems, the following Kuhn-Tucker conditions KT^q and $KT^q(R)$ for problems

P^q and $P^q(R)$ define their optimal solutions, respectively.

$$\begin{aligned} & \frac{q(Vy)_j}{\sqrt{y'Vy}} + \mu_j + u_j - v_j + \lambda = 0, \quad j = 1, 2, \dots, n, \\ (KT^q) \quad & v_j y_j = 0, \quad u_j (y_j - \gamma_j) = 0, \quad j = 1, 2, \dots, n, \\ & e'y = 1, \quad 0 \leq y_j \leq \gamma_j, \quad u_j, v_j \geq 0, \quad j = 1, 2, \dots, n, \end{aligned}$$

$$\begin{aligned} & q(Vy)_j + R\mu_j + u_j - v_j + \lambda = 0, \quad j = 1, 2, \dots, n, \\ (KT^q(R)) \quad & v_j y_j = 0, \quad u_j (y_j - \gamma_j) = 0, \quad j = 1, 2, \dots, n, \\ & e'y = 1, \quad 0 \leq y_j \leq \gamma_j, \quad u_j, v_j \geq 0, \quad j = 1, 2, \dots, n, \end{aligned}$$

where $\lambda, u_j, v_j, j = 1, 2, \dots, n$, are Lagrange multipliers.

The following lemmas show how problem P_3 is related to the auxiliary problems P^q and $P^q(R)$.

LEMMA 5.1. *The optimal values of problems P_3 and P^q satisfy*

$$\begin{aligned} & Z^q < -d \iff q < q^*, \\ (5.2) \quad & Z^q = -d \iff q = q^*, \\ & Z^q > -d \iff q > q^*. \end{aligned}$$

PROOF: See Dinkelbach [D2]. ■

LEMMA 5.2. (Ishii and Nishida [I3]) *If $R = \sqrt{y^q(R)'Vy^q(R)}$, then $y^q(R)$ is an optimal solution y^q of problem P^q .*

PROOF: Let $(y^q, u^q, v^q, \lambda^q)$ and $(y^q(R), u^q(R), v^q(R), \lambda^q(R))$ denote solutions of KT^q and $KT^q(R)$, respectively. Dividing the first equation in $KT^q(R)$ by $R = \sqrt{y^q(R)'Vy^q(R)}$, then $(y^q, u^q, v^q, \lambda^q)$ is constructed from $(y^q(R), u^q(R), v^q(R), \lambda^q(R))$ as follows.

$$(5.3) \quad y^q = y^q(R), \quad u^q = \frac{u^q(R)}{R}, \quad v^q = \frac{v^q(R)}{R}, \quad \lambda^q = \frac{\lambda^q(R)}{R}.$$

Therefore, $y^q(R)$ is an optimal solution of problem P^q . ■

LEMMA 5.3. (Ishii and Nishida [I3]) *The $y^1(R)$ satisfying*

$$(5.4) \quad qR = \sqrt{y^1(R)'Vy^1(R)}$$

is an optimal solution y^q of problem P^q .

PROOF: By the Kuhn-Tucker conditions $KT^1(R)$ and $KT^q(qR)$, we see that $(y^1(R), q u^1(R), q v^1(R), q \lambda^1(R))$ is a solution of $KT^q(qR)$. Therefore, from Lemma 5.2, the $y^1(R)$ satisfying (5.4) is an optimal solution y^q of problem P^q . ■

Lemma 5.1 and Lemma 5.3 assure that optimal solutions $y^1(R)$ of problems $P^1(R)$ with positive R contain an optimal solution y^* of problem P_3 . Thus problem P_3 can be solved by considering only problems $P^1(R)$ for positive R . We, therefore, denote $P^1(R)$, $y^1(R)$ and $Z^1(R)$ with $P(R)$, $y(R)$ and $Z(R)$, respectively, in the subsequent discussion.

For obtaining an optimal solution $y^* = y(R^*)$ of problem P_3 , we define function $T(R)$ of R by

$$(5.5) \quad T(R) = R\{\mu'y(R) - d\} - y(R)'Vy(R).$$

The following theorem shows the optimality condition for $y(R)$.

THEOREM 5.4. *Let R^* be a solution of equation $T(R) = 0$. Then $y(R^*)$ is an optimal solution of problem P_3 .*

PROOF: From Lemma 5.3, the $y(R)$ satisfying (5.4) is an optimal solution of problem $P(R)$. Then Lemma 5.1 implies that the $y(R)$ satisfying

both $Z^q = -d$ and (5.4) is an optimal solution of problem P_3 . Since $Z^q = -d$ and (5.4) are equivalent to

$$(5.6) \quad R\{\mu'y(R) - d\} - y(R)'Vy(R) = 0,$$

the $y(R^*)$ satisfying $T(R^*) = 0$ is an optimal solution of problem P_3 . ■

Note that the optimal value is given by

$$(5.7) \quad q^* = \frac{\sqrt{y(R^*)'Vy(R^*)}}{R^*}.$$

In the next section, we discuss the efficient frontier in the mean variance framework, and show the uniqueness of R^* .

5.3. An Optimal Solution on the Efficient Frontier

Consider the region of admissible portfolios in the space of ν and π , where $\nu = y'Vy$ is a variance and $\pi = \mu'y$ is a mean of the objective value. The region of admissible portfolios is defined by

$$(5.8) \quad S = \{(\nu, \pi) \mid y'Vy = \nu, \mu'y = \pi, e'y = 1, 0 \leq y \leq \gamma\}.$$

Since we would prefer to minimize the variance as far as we get same mean of the objective value, the efficient frontier in the space of ν and π is given by

$$(5.9) \quad B = \left\{ (\nu, \pi) \mid \nu = \min_y \{y'Vy \mid \mu'y = \pi, e'y = 1, 0 \leq y \leq \gamma\} \right\}.$$

Let J_L and J_U be subsets of $\{1, 2, \dots, n\}$ such that $J_L \cap J_U = \emptyset$ and $B_{J_L}^{J_U}$ denote the efficient frontier of feasible region when $y_j = 0$ for

$j \in J_L$, $y_j = \gamma_j$ for $j \in J_U$ and $0 < y_j < \gamma_j$ for $j \notin J_L \cup J_U$. Then $B_{J_L}^{J_U}$ is an arc of parabola and B is composed of a sequence of such arcs of parabolas $B_{J_L}^{J_U}$ [S4]. B is a continuous curve which has two extremal points $Q_M = (\nu_M, \pi_M)$ and $Q_m = (\nu_m, \pi_m)$, where π_M and π_m give the largest and the smallest means, respectively, and ν_M and ν_m are the corresponding variances. Note that π_M and π_m are easily calculated by solving linear programming problems as follows.

$$(5.10) \quad \begin{aligned} \pi_M &= \max\{\mu'y \mid e'y = 1, 0 \leq y \leq \gamma\}, \\ \pi_m &= \min\{\mu'y \mid e'y = 1, 0 \leq y \leq \gamma\}. \end{aligned}$$

Let B denote by

$$(5.11) \quad \nu = f(\pi) = \begin{cases} f_1(\pi), & \text{for } \pi_0 \leq \pi \leq \pi_1, \\ \vdots \\ f_i(\pi), & \text{for } \pi_{i-1} \leq \pi \leq \pi_i, \\ \vdots \\ f_K(\pi), & \text{for } \pi_{K-1} \leq \pi \leq \pi_K, \end{cases}$$

as a sequence of parabolas f_1, f_2, \dots, f_K , where $\pi_0 = \pi_m$ and $\pi_K = \pi_M$ and K is the number of parabolas of which the efficient frontier B consists. We define $f'(\pi) \in \partial f(\pi) = [f'(\pi - 0), f'(\pi + 0)]$ as the subdifferential of $f(\pi)$. Note that

$$(5.12) \quad f'_i(\pi_i - 0) \leq f'_{i+1}(\pi_i + 0), \quad i = 0, 1, \dots, K,$$

from the convexity of S .

Now we show some illustrative examples of efficient frontier.

$$(Example 5.1) \quad \mu = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.8 \end{pmatrix}.$$

The efficient frontier B corresponding to these data is given by

$$(5.13) \quad \nu = \begin{cases} 5\pi^2 - 17\pi + 15, & \text{for } \frac{3}{2} \leq \pi \leq \frac{12}{7}, \\ \frac{11}{12}\pi^2 - 3\pi + 3, & \text{for } \frac{12}{7} \leq \pi \leq \frac{18}{7}, \\ 5\pi^2 - 24\pi + 30, & \text{for } \frac{18}{7} \leq \pi \leq \frac{14}{5}, \end{cases}$$

which is composed of a sequence of arcs of $B_\phi^{(1)}$, B_ϕ^ϕ and $B_{\{1\}}^\phi$ as shown in Figure 5.1. Then we have

$$(5.14) \quad \begin{aligned} \pi_0 &= \frac{3}{2}, \pi_1 = \frac{12}{7}, \pi_2 = \frac{18}{7}, \pi_3 = \frac{14}{5}, K = 3, \\ f'(\pi_0) &= -2, f'(\pi_1) = \frac{1}{7}, f'(\pi_2) = \frac{12}{7}, f'(\pi_3) = 4, \\ Q_m &= \left(\frac{3}{4}, \frac{3}{2}\right), Q_M = \left(2, \frac{14}{5}\right). \quad \blacksquare \end{aligned}$$

$$(5.15) \quad \mu = \begin{pmatrix} 3 \\ 6 \\ 8 \end{pmatrix}, V = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 4 & 8 \\ 2 & 8 & 25 \end{pmatrix}, \gamma = \begin{pmatrix} 2/3 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

The efficient frontier B corresponding to these data is given by

$$(5.15) \quad \nu = \begin{cases} \frac{1}{3}\pi^2 - 2\pi + 4, & \text{for } 4 \leq \pi \leq 5, \\ \frac{22}{25}\pi^2 - \frac{34}{5}\pi + \frac{43}{3}, & \text{for } 5 \leq \pi \leq \frac{20}{3}, \\ \frac{13}{4}\pi^2 - 35\pi + 97, & \text{for } \frac{20}{3} \leq \pi \leq \frac{22}{3}, \end{cases}$$

which is composed of a sequence of arcs of $B_{\{3\}}^\phi$, $B_\phi^{(2)}$ and $B_{\{1\}}^\phi$ as shown in Figure 5.2. Then we have

$$(5.16) \quad \begin{aligned} \pi_0 &= 4, \pi_1 = 5, \pi_2 = \frac{20}{3}, \pi_3 = \frac{22}{3}, K = 3, \\ f'(\pi_0) &= \frac{2}{3}, f'(\pi_1) \in \left[\frac{4}{3}, 2\right], f'(\pi_2) \in \left[\frac{74}{15}, \frac{25}{3}\right], f'(\pi_3) = \frac{38}{3}, \\ Q_m &= \left(\frac{4}{3}, 4\right), Q_M = \left(\frac{136}{9}, \frac{22}{3}\right). \quad \blacksquare \end{aligned}$$

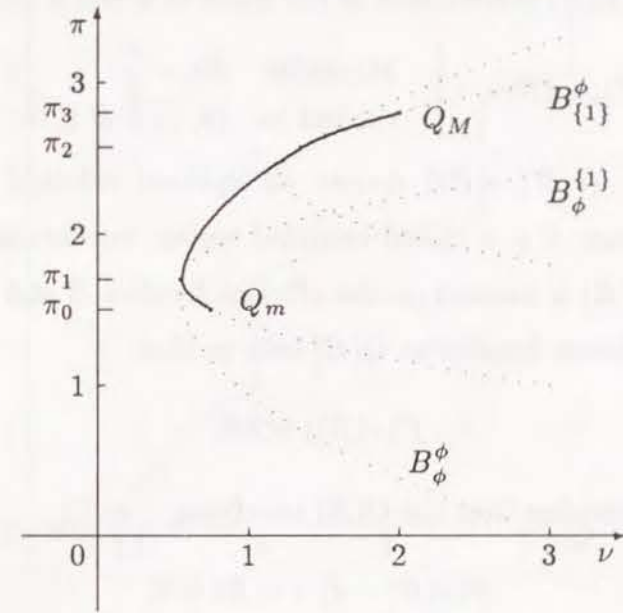


Figure 5.1. Efficient frontier for Example 5.1.

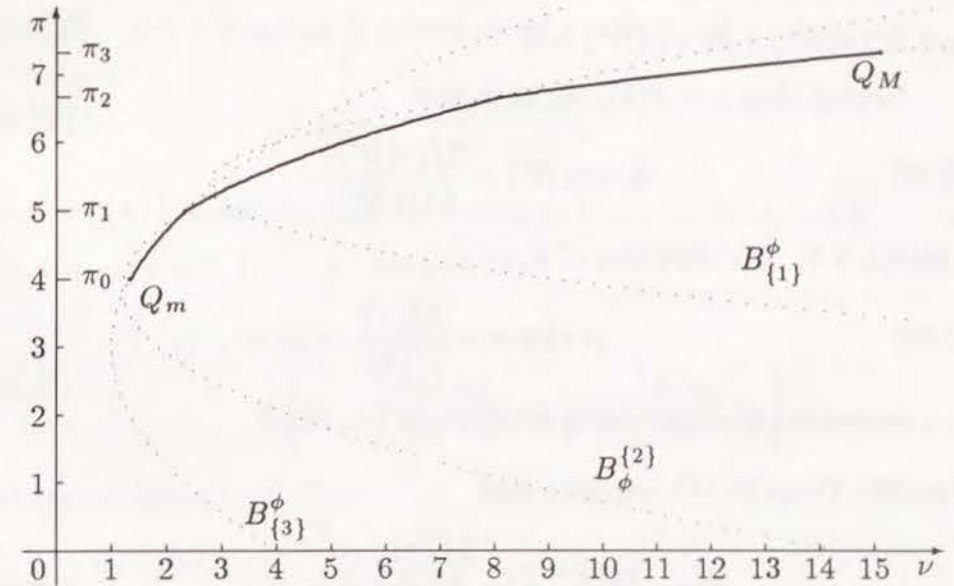


Figure 5.2. Efficient frontier for Example 5.2.

Problem $P(R)$ is rewritten in the space of ν and π as

$$P_{\langle \nu, \pi \rangle}(R) : \begin{cases} \text{Maximize} & R\pi - \frac{\nu}{2}, \\ \text{subject to} & (\nu, \pi) \in B \subset S. \end{cases}$$

Let $Q(R) = (\nu(R), \pi(R))$ denote an optimal solution of problem $P_{\langle \nu, \pi \rangle}(R)$. Since S is a closed bounded region surrounded by convex parabolas, $Q(R)$ is located on the efficient frontier B and the tangent line of the efficient frontier at $Q(R)$ tells us that

$$(5.17) \quad f'(\pi(R)) = 2R.$$

Theorem 5.4 implies that the $Q(R)$ satisfying

$$(5.18) \quad R(\pi(R) - d) - \nu(R) = 0,$$

is given by an optimal solution of problem P_3 . Since $Q(R^*)$ is an optimal solution of problem $P_{\langle \nu, \pi \rangle}(R^*)$, it holds that $f'(\pi(R^*)) = 2R^*$ and $R^*(\pi(R^*) - d) - \nu(R^*) = 0$.

Noting that $\nu = f(\pi)$, we have that

$$(5.19) \quad d = \pi(R^*) - \frac{2f(\pi(R^*))}{f'(\pi(R^*))}.$$

LEMMA 5.5. The function of π ,

$$(5.20) \quad g(\pi) = \pi - \frac{2f(\pi)}{f'(\pi)},$$

is a monotonically increasing function on $[\pi_m, \pi_M]$.

PROOF: From (5.11), we have that

$$(5.21) \quad \begin{aligned} g(\pi) &= \pi - \frac{2f_i(\pi)}{f'_i(\pi)}, \\ g'(\pi) &= \frac{2f_i(\pi)f''_i(\pi)}{\{f'_i(\pi)\}^2} - 1, \end{aligned}$$

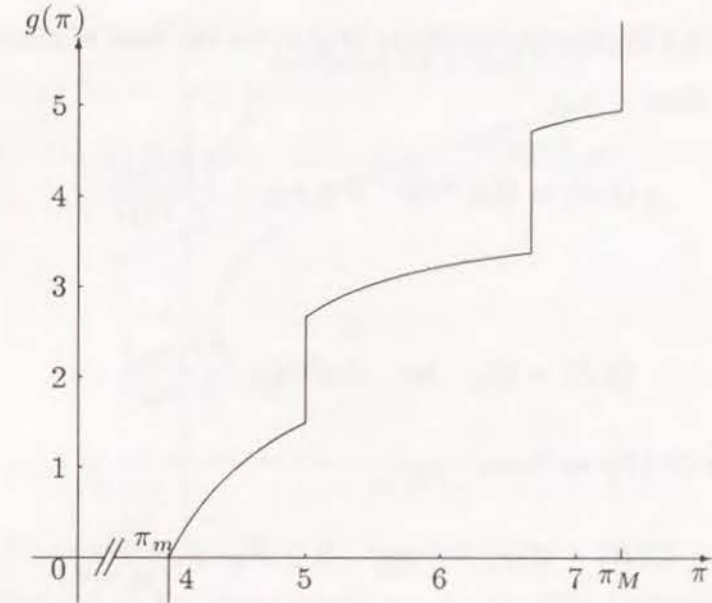


Figure 5.3. $g(\pi)$ for Example 5.2

on $\pi_{i-1} < \pi < \pi_i$. For a parabola $f_i(\pi) = \alpha_i \pi^2 - 2\beta_i \pi + \gamma_i$, it holds that $\beta_i^2 - \alpha_i \gamma_i < 0$ and $\alpha_i > 0$ because of $\nu > 0$. Then

$$(5.22) \quad g'(\pi) = \frac{\alpha_i \gamma_i - \beta_i^2}{(\alpha_i \pi - \beta_i)^2} > 0$$

and so $g(\pi)$ is monotonically increasing on (π_{i-1}, π_i) for $i = 0, 1, \dots, K$.

At $\pi = \pi_i$, $i = 0, 1, \dots, K$, from (5.12),

$$(5.23) \quad \begin{aligned} g(\pi_i) &\in [g(\pi_i - 0), g(\pi_i + 0)] \\ &\equiv \left[\pi_i - \frac{2f_i(\pi_i)}{f'_i(\pi_i - 0)}, \pi_i - \frac{2f_i(\pi_i)}{f'_{i+1}(\pi_i + 0)} \right], \end{aligned}$$

where we define

$$(5.24) \quad g(\pi_0 - 0) = -\infty, \quad g(\pi_K + 0) = +\infty$$

Therefore, $g(\pi)$ is an increasing function on $[\pi_m, \pi_M]$. ■

Figure 5.3 illustrates the shape of $g(\pi)$ for the data of Example 5.2.

It is found that

$$(5.25) \quad Q(R) = Q_M \quad \text{for} \quad d \geq \pi_M - \frac{2f(\pi_M)}{f'(\pi_M)}$$

and

$$(5.26) \quad Q(R) = Q_m \quad \text{for} \quad d \leq \pi_m - \frac{2f(\pi_m)}{f'(\pi_m)}.$$

Then, from (5.17), we have

$$(5.27) \quad Q(R) = Q_M \quad \text{for any} \quad R \geq R_M \equiv \frac{\nu_M}{\pi_M - d}$$

and

$$(5.28) \quad Q(R) = Q_m \quad \text{for any} \quad R \leq R_m \equiv \frac{\nu_m}{\pi_m - d}.$$

Lemma 5.5 implies that the $\pi(R^*)$ satisfying (5.21), as well as R^* , is unique, because $Q(R)$ moves from $Q(R_m)$ to $Q(R_M)$ on the efficient frontier as R increases from R_m to R_M .

THEOREM 5.6. *The parameter value $R = R^*$ such that $y(R^*)$ is an optimal solution of problem P_3 uniquely exists.*

Note that, in the (ν, π) -space, the tangent line of efficient frontier at the point, given by an optimal solution of problem $P_{(\nu, \pi)}(R^*)$, crosses the π -axis at the midpoint of d and $\pi(R^*)$ as shown in Figure 5.4.

Problem $P(R)$ is the same to the EV model

$$\text{EV : } \begin{cases} \text{Maximize} & \mu'y - w \cdot y'Vy, \\ \text{subject to} & e'y = 1, \\ & 0 \leq y \leq \gamma, \end{cases}$$

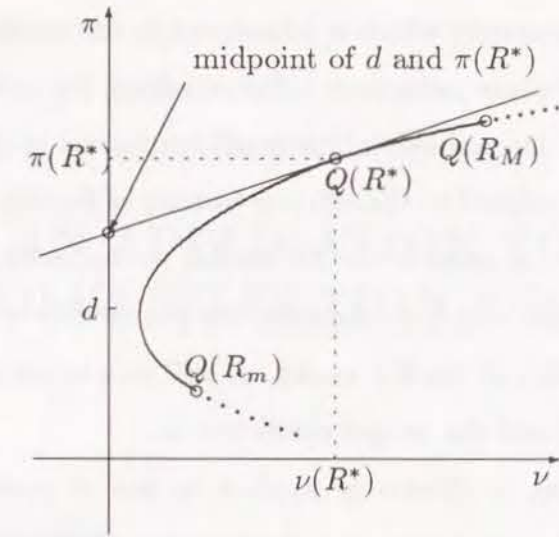


Figure 5.4. An optimal solution on the efficient frontier

which is maximized the expectation on which the variance is imposed as a penalty with weight coefficient $w = \frac{1}{2R}$. When we set a goal value d and decide the portfolio which maximizes the probability that attains the goal value, it can be found which value of weight coefficient is imposed on the variance for this portfolio.

A solution algorithm for problem P_3 is discussed in the next chapter, where the variance covariance matrix of random cost coefficients has special structures used in the portfolio selection problems.

5.4. Conclusion

We discuss a probability maximizing model of a stochastic linear knapsack problem in this chapter. This problem is equivalent to a deterministic fractional problem under the assumption of normally distributed random cost coefficients. We show the optimality condition

for a positive parameter which is introduced in the auxiliary quadratic problem. The positive parameter value satisfying the optimality condition is uniquely determined, which is also explained in the mean variance framework related to the efficient frontier of feasible region. Since problem $P(R)$ is the same to the EV model, we discussed a relationship between a solution which maximizing the probability of achieving the goal and a solution of the EV model, as well as a relationship between the goal value d and the weight coefficient w .

This problem is effectively applied to one of portfolio selection problems and it necessitates the development of efficient solution algorithms for the problems with special structures of variances used in portfolio selection problems, e.g., a block diagonal model and index models, etc.. We discuss solution algorithms for such problems in the next chapter.

Chapter 6.

AN APPLICATION TO PORTFOLIO SELECTION PROBLEMS

6.1. Introduction

Finance problems have recently become a hot field of operations research. Portfolio theory gives us an optimal investment strategy or an effective use of the funds taking into account the avoidance of risk, where we can not avoid making a decision under uncertainty. Stochastic programming models are necessary for investment analysis decision making. This chapter gives solution algorithms for two types of portfolio selection problems via the corresponding stochastic linear knapsack problems. Note that the notation of problems and parameters in this chapter is the same as those used in Chapter 5. As shown in Section 5.2, when random cost coefficients have a normal distribution, the probability maximizing model P_1 of a stochastic linear knapsack problem is transformed into the deterministic fractional problem P_3 and the auxiliary problem $P(R)$ with a positive parameter R is introduced.

To generalize the problem with independent random cost coefficients to the problem with correlated random cost coefficients, we introduce in this chapter two types of models of random variables, one

is a block diagonal model and the other is index models. In the block diagonal model, the variance covariance matrix of random cost coefficients is block diagonal, where it is assumed that the investments are classified into several groups such that the investments in the same group are correlated each other but the investments in the different groups are mutually independent.

Elton, Gruber and Padberg [E2, E3, E4] consider the problems with special structures of variances of random cost coefficients, e.g., a single index model, a multi index model and a constant correlation coefficients model. They solve such problems by using a ranking procedure, but the computational complexity is not mentioned clearly.

Section 6.2 discusses a generalized stochastic linear knapsack problem in which random cost coefficients have a block diagonal variance covariance matrix. This problem is transformed into an equivalent fractional problem, and then decomposed into subproblems corresponding to the blocks. We propose a parametric algorithm by extending the Zipkin's ranking method [Z2]. A portfolio selection problem is also discussed as an example of this model.

In Section 6.3, we discuss a portfolio selection problem of a single index model. A single index model is based on the CAPM (capital asset pricing model) and represents the rate of return on investment which is regressed to the market index by using many historical data about the risk and the return on investment. A single index model decomposes the risk on investment into the risk common to all investments caused by a market and the risk individual to each investment

caused by other factors. In this problem, all investments are mutually correlated through the market index. By introducing a parameter corresponding to the common market index, we can obtain an optimal portfolio using the proposed efficient algorithm. The APT (arbitrage pricing theory) provides a multi index model, where the rate of return on investment is assumed to be determined by several common indices besides the market index. An extension of the algorithm to the problem of a multi index model is also investigated in Section 6.3.4. Section 6.4 gives a summary of Chapter 6.

6.2. A Block Diagonal Model

6.2.1. Decomposition into subproblems

We consider the case in which variance covariance matrix of random cost coefficients is block diagonal. In this case, random cost coefficients are classified into several groups with the mutually correlated coefficients. It is assumed that there are m blocks and the k -th block has r_k random coefficients, where $\sum_{k=1}^m r_k = n$ and each r_k is sufficiently small compared with n . We shall now cite the main problem P_3 and its auxiliary problem $P(R)$ of Section 5.2 again, but they are specialized to the problem setting of this section.

$$\begin{array}{l}
 P_3 : \left\{ \begin{array}{l} \text{Maximize} \quad \frac{\mu'y - d}{\sqrt{y'Vy}}, \\ \text{subject to} \quad e'y = 1, \\ \quad \quad \quad y \geq 0, \end{array} \right. \\
 \\
 P(R) : \left\{ \begin{array}{l} \text{Minimize} \quad -R\mu'y + \frac{1}{2}y'Vy, \\ \text{subject to} \quad e'y = 1, \\ \quad \quad \quad y \geq 0, \end{array} \right.
 \end{array}$$

where the mean μ and the variance covariance matrix V are expressed as

$$(6.1) \quad V = \begin{pmatrix} V^1 & 0 & \dots & 0 \\ 0 & V^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V^m \end{pmatrix},$$

$$\mu = (\mu^1, \mu^2, \dots, \mu^m)',$$

with an $r_k \times r_k$ positive definite matrix $V^k = (\sigma_{ij}^k)$ and an r_k -vector μ^k for $k = 1, 2, \dots, m$.

Let $y^k = (y_1^k, y_2^k, \dots, y_{r_k}^k)$ denote decision variables in the k -th block. Then problem $P(R)$ is decomposed into m subproblems and we obtain the k -th subproblem as follows:

$$P^k(R, w_k) : \begin{cases} \text{Minimize} & -R(\mu^k)'y^k + \frac{1}{2}(y^k)'V^k y^k, \\ \text{subject to} & e'y^k = w_k, \\ & y^k \geq 0, \end{cases}$$

where w_k is a nonnegative parameter w_k such that

$$(6.2) \quad \sum_{k=1}^m w_k = 1,$$

which denotes the amount of allocation given to the k -th block. Note that problem $P^k(R, w_k)$ has two nonnegative parameters R and w_k . As far as no confusion occurs, we abbreviate $P^k(R, w_k)$ as P^k and suppress the superscript k on y^k , u^k and V^k .

Now we consider problem P^k for the k -th block. Since problem P^k is a convex programming problem, the following Kuhn-Tucker conditions give an optimal solution of problem P^k .

$$(6.3) \quad \begin{aligned} -R\mu_j + (Vy)_j - u_j + \lambda_k &= 0, \quad j = 1, 2, \dots, r_k, \\ e'y &= w_k, \\ u_j y_j &= 0, \quad y_j \geq 0, \quad u_j \geq 0, \quad j = 1, 2, \dots, r_k. \end{aligned}$$

The matrix form of these constraints other than the complementary condition is described by

$$(6.4) \quad \begin{pmatrix} V^k & -E_k & e \\ e' & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{r_k} \\ u_1 \\ \vdots \\ u_{r_k} \\ \lambda_k \end{pmatrix} = \begin{pmatrix} R\mu_1 \\ \vdots \\ R\mu_{r_k} \\ w_k \end{pmatrix},$$

where E_k is an $r_k \times r_k$ identity matrix. For a fixed w_k , we can solve (6.4) as a linear programming problem with complementary slackness conditions that both u_j and y_j are not simultaneously brought into basis. Since the rank of the matrix in (6.4) is $r_k + 1$, there exists an $r_k + 1$ dimensional basic vector and the corresponding basic matrix. The possible number of the basic vectors is 2^{r_k} . Here we denote one of the basic vectors and the corresponding basic matrix by $(y_1, \dots, y_{l_k}, u_{l_k+1}, \dots, u_{r_k}, \lambda_k)'$ and B_k , respectively, with appropriate arrangement of elements, if necessary. The basic solution is expressed as

$$(6.5) \quad \begin{aligned} (y_1, \dots, y_{l_k}, u_{l_k+1}, \dots, u_{r_k}, \lambda_k)' &= B_k^{-1}(R\mu_1, \dots, R\mu_{r_k}, w_k)', \\ u_1 = u_2 = \dots = u_{l_k} &= y_{l_k+1} = \dots = y_{r_k} = 0, \end{aligned}$$

where $1 \leq l_k \leq r_k$. Then an optimal solution of the k -th subproblem $P^k(R, w_k)$ is given as the following linear functions of nonnegative parameters R and w_k .

$$(6.6) \quad y_i^k(R, w_k) = \begin{cases} \beta_i^k R + b_{i, r_k+1}^k w_k, & i = 1, 2, \dots, l_k, \\ 0, & i = l_k + 1, \dots, r_k, \end{cases}$$

$$(6.7) \quad u_i^k(R, w_k) = \begin{cases} 0, & i = 1, 2, \dots, l_k, \\ \beta_i^k R + b_{i, r_k+1}^k w_k, & i = l_k + 1, \dots, r_k, \end{cases}$$

$$(6.8) \quad \lambda_k = \beta_{r_k+1}^k R + b_{r_k+1, r_k+1}^k w_k,$$

where

$$(6.9) \quad \beta_i^k = \sum_{j=1}^{r_k} b_{ij}^k \mu_j^k, \quad i = 1, 2, \dots, r_k + 1,$$

and b_{ij}^k , $i, j = 1, 2, \dots, r_k + 1$, denotes the (i, j) -th element of the inverse matrix of B_k .

As mentioned above, the allocation w_k to the k -th block determines the corresponding basic matrix. Conversely, by feasibility, the basic matrix determines the existence range of the corresponding allocation w_k as stated in the next lemma.

LEMMA 6.1. For a basic matrix B_k , the corresponding set of allocation w_k , if any, is a convex subset on $R_+^1 = \{w \mid w \geq 0\}$.

PROOF: Since optimal solutions $y^k(R, w_k)$ and $u^k(R, w_k)$ given by (6.6) and (6.7) should be nonnegative so that a matrix B_k is basic, the allocated value w_k for a basic matrix B_k must belong to

$$(6.10) \quad \Omega^k(w) = \bigcap_{i=1}^{r_k} \{w \mid \beta_i^k R + b_{i, r_k+1}^k w \geq 0\}.$$

Clearly, $\Omega^k(w)$ is a convex set on R_+^1 . ■

Then we return to problem $P(R)$. Problem $P(R)$ is expressed by using the values of objective functions of the subproblems as follows:

$$P_4 : \begin{cases} \text{Minimize} & \sum_{k=1}^m Z^k(R, w_k), \\ \text{subject to} & \sum_{k=1}^m w_k = 1, \\ & w_k \geq 0, \quad k = 1, 2, \dots, m, \end{cases}$$

where

$$(6.11) \quad Z^k(R, w_k) = -R(\mu^k)' y^k(R, w_k) + \frac{1}{2} y^k(R, w_k)' V^k y^k(R, w_k)$$

is the optimal value of subproblem $P^k(R, w_k)$. $Z^k(R, w)$ is a convex function of $w \in [0, 1]$. We abbreviate $Z^k(R, w_k)$ as $Z^k(w_k)$ and $y^k(R, w_k)$ as $y^k(w_k)$ or y^k . Possibly, $Z^k(w_k)$ has non-differentiable points if the basic matrix is degenerate, and their number is at most 2^{r_k} . Except for these non-differentiable points, let $DZ^k(w)$ denote a derivative of function $Z^k(w)$.

LEMMA 6.2. The derivative of $Z^k(w_k)$ is obtained from a Lagrange multiplier λ_k given by (6.8) as follows.

$$(6.12) \quad DZ^k(w_k) = -\lambda_k.$$

PROOF: From (6.3) and (6.8), we obtain that

$$(6.13) \quad \begin{aligned} DZ^k(w_k) &= (V^k y^k(w_k) - R\mu^k)' \frac{d}{dw_k} y^k(w_k) \\ &= (u^k - \lambda_k e)' \frac{d}{dw_k} y^k(w_k) \\ &= -\lambda_k \sum_{i=1}^{l_k} b_{i, r_k+1}^k \\ &= -\lambda_k. \end{aligned}$$

Note that $\sum_{i=1}^{l_k} b_{i, r_k+1}^k = 1$, since b_{i, r_k+1}^k is an $(r_k + 1, r_k + 1)$ -st element of matrix $B_k B_k^{-1}$. ■

6.2.2. A ranking method and a solution algorithm

We first review the Zipkin's simple ranking method [Z2], which treats a maximization problem with a continuously differentiable convex objective function and a single linear constraint. Based on the values of $DZ^k(0)$, $k = 1, 2, \dots, m$, we can obtain the number of blocks which have positive allocations. Strictly speaking, $DZ^k(0)$ denotes the right derivative of $Z^k(w)$ at $w = 0$. The Kuhn-Tucker conditions for problem P_4 show that $DZ^k(w_k)$, $k = 1, 2, \dots, m$, have a real number M satisfying

$$(6.14) \quad \begin{aligned} w_k > 0 &\implies DZ^k(w_k) = M, \\ w_k = 0 &\implies DZ^k(w_k) \geq M. \end{aligned}$$

The positive w_k appears in an optimal solution according to the values of $DZ^k(0)$, $k = 1, 2, \dots, m$. We arrange $DZ^k(0)$ in increasing order. Note that the order of $DZ^k(0)$ is independent of R , since $DZ^k(0) = -\beta_{r_k+1}^k R$. Let N denote the number of positive w_k . Then we have

$$(6.15) \quad \begin{aligned} w_k > 0 &\text{ for } k = 1, 2, \dots, N, \\ w_k = 0 &\text{ for } k = N + 1, \dots, m, \end{aligned}$$

and M and w_k , $k = 1, 2, \dots, m$, are given as functions of N from (6.2), (6.8), (6.12) and (6.14). Therefore, N^* is optimal, i.e., $w_k(N^*) = w_k^*$, $k = 1, 2, \dots, m$, if N^* satisfies

$$(6.16a) \quad DZ^{N^*}(0) < M(N^*),$$

$$(6.16b) \quad DZ^{N^*+1}(0) \geq M(N^*).$$

Note that only (6.16a) applies if $N^* = m$.

We now extend the ranking method in order to treat an optimization problem with a piecewise linear convex objective function. Let ξ_i^k , $i = 1, 2, \dots, p_k$, denote the non-differentiable points of $DZ^k(w)$, where p_k is their number which is at most 2^{r_k} . At these points, we define $DZ^k(\xi_i^k)$ as the subdifferential of $Z^k(w)$ at $w = \xi_i^k$, $i = 1, 2, \dots, p_k$, that is,

$$(6.17) \quad DZ^k(\xi_i^k) = [DZ_-^k(\xi_i^k), DZ_+^k(\xi_i^k)],$$

where $DZ_-^k(\xi_i^k)$ and $DZ_+^k(\xi_i^k)$ are left and right derivatives of $Z^k(w)$ at $w = \xi_i^k$, respectively. Note that $DZ^k(w)$ is strictly increasing, continuous and piecewise linear from the convexity of $Z^k(w)$ and Lemma 6.2.

The optimality of solution $y^k(R, w)$ of problem $P^k(R, w)$ can be checked by using the following function $T(R)$.

$$(6.18) \quad \begin{aligned} T(R) &= R\{\mu^1 y - d\} - y^1 V y \\ &= R\left\{\sum_{k=1}^m (\mu^k)' y^k - d\right\} - \sum_{k=1}^m (y^k)' V^k y^k \\ &= \sum_{k=1}^m \{R(\mu^k) - (V^k y^k)\}' y^k - R d \\ &= \sum_{k=1}^m w_k \lambda_k - R d. \end{aligned}$$

In the ranking method, we solve the following equations for the blocks with positive allocations.

$$(6.19) \quad DZ^k(w_k) = M, \quad k = 1, 2, \dots, N^*,$$

that is,

$$(6.20) \quad \lambda_k = -M, \quad k = 1, 2, \dots, N^*.$$

From (6.18), (6.20) and Theorem 5.4, the optimality condition is given by

$$(6.21) \quad M = -Rd,$$

and from (6.8) we obtain

$$(6.22) \quad w_k = \begin{cases} \frac{d - \beta_{r_k+1}^k}{b_{r_k+1, r_k+1}^k} R, & k = 1, 2, \dots, N^*, \\ 0, & k = N^* + 1, \dots, m. \end{cases}$$

Here it remains to check whether w_k belongs to $\Omega^k(w)$ or not. Since all w_k are also directly proportional to R , Lemma 6.1 implies that we can check it without knowing the value of R .

Let $\{s_1, s_2, \dots, s_p\}$ denote a set of all left and right derivatives, $DZ_+^k(\xi_i^k)$ and $DZ_-^k(\xi_i^k)$, $i = 1, 2, \dots, p_k$, $k = 1, 2, \dots, m$, at the non-differentiable points, where p is the number of different values of left and right derivatives and $p \leq \sum_{k=1}^m p_k$.

LEMMA 6.3. *The order of all left and right derivatives s_α , $\alpha = 1, 2, \dots, p$, is independent of the value of R .*

PROOF: Since at least one basic variable becomes zero for $w = \xi_i^k$, from (6.6) and (6.7), ξ_i^k is proportional to R and so is s_α . Thus the order of s_α , $\alpha = 1, 2, \dots, p$, is independent of R . ■

We denote $s_{\alpha_k, 1}, s_{\alpha_k, 2}, \dots, s_{\alpha_k, p(k)}$ the derivatives which are between $DZ^k(0)$ and $DZ^{k+1}(0)$, where $p(k)$ is their number and $p = \sum_{k=1}^m p(k)$. We sort them as follows.

$$(6.23) \quad DZ^k(0) \leq s_{\alpha_k, 1} \leq \dots \leq s_{\alpha_k, p(k)} \leq DZ^{k+1}(0), \quad k = 1, 2, \dots, m,$$

where $DZ^{m+1}(0)$ denotes a sufficiently large number.

Now we show a solution algorithm to solve problem P_3 .

Algorithm 6.1. (Block diagonal model)

Step 1. Calculate $DZ^k(w_k)$, $k = 1, 2, \dots, m$.

Step 2. Let $N \leftarrow 1$, $\alpha \leftarrow \alpha_{N,1}$.

Step 3. Calculate w_k , $k = 1, 2, \dots, m$, by (6.22).

Step 4. If $w_k \in \Omega^k(w)$, then let $N^* \leftarrow N$ and go to Step 7.

Step 5. If $\alpha = \alpha_{N, p(N)}$, then let $\alpha \leftarrow \alpha + 1$ and return to Step 3.

Step 6. If $N < m$, then let $N \leftarrow N + 1$, $\alpha \leftarrow \alpha_{N,1}$ and return to Step 4. Otherwise terminate by concluding that problem P_3 has no solution.

Step 7. Normalizing w_k to satisfy condition $\sum_{k=1}^m w_k = 1$, i.e.,

$$(6.24) \quad w_k^* = \begin{cases} \frac{d - \beta_{r_k+1}^k}{b_{r_k+1, r_k+1}^k} R^*, & k = 1, 2, \dots, N^*, \\ 0, & k = N^* + 1, \dots, m, \end{cases}$$

where

$$(6.25) \quad R^* = \left\{ \sum_{k=1}^{N^*} \frac{d - \beta_{r_k+1}^k}{b_{r_k+1, r_k+1}^k} \right\}^{-1},$$

an optimal solution and the optimal value of problem P_3 are given by

$$(6.26) \quad y_i^k = \begin{cases} \beta_i^k R^* + b_{r_k+1, r_k+1}^k w_k^*, & i = 1, 2, \dots, l_k, \\ 0, & i = l_k + 1, \dots, r_k, \end{cases}$$

for $k = 1, 2, \dots, m$, and

$$(6.27) \quad q^* = \frac{\sqrt{(y^*)' V y^*}}{R^*},$$

respectively, and then terminate. ■

The complexity of Algorithm 6.1 is dominated by the selection of a basic matrix and the calculation of its inverse matrix in Step 1. Therefore, in the worst case, it takes $O(\sum_{k=1}^m r_k^3 2^{r_k}) \leq O(n \cdot \max_k r_k^2 2^{r_k})$ computational time. The complexity for the problem with mutually correlated random cost coefficients is $O(n^3 2^n)$. Since

$$(6.28) \quad O\left(\sum_{k=1}^m r_k^3 2^{r_k}\right) \leq O\left(\prod_{k=1}^m r_k^3 2^{r_k}\right) \leq O(n^3 2^n),$$

our algorithm gives a reasonable complexity when each r_k is sufficiently small compared with n .

6.2.3. A simple portfolio selection problem

Our algorithm is applied to the following simple portfolio selection problem. We consider the case in which there are n infinitely divisible investments with rates of returns $r = (r_1, r_2, \dots, r_n)$. It is assumed that returns on investments are not independent, but they are correlated in such a way that the variance covariance matrix of the rate of return on investment is block diagonal. The investor would like to know the proportion to be invested into each investment in order to get a large return under a small risk. Let r_F denote the rate of return on riskless asset, i.e., the variance of the rate of return on riskless asset is zero. It is interpreted, in practice, as the riskless lending rate, e.g., bank deposits and saving bonds.

To achieve the above goal, we consider to find a solution that maximizes the ratio of excess return [E1], i.e., expected rate of return

minus riskless rate, to standard deviation under the constraint that a total sum of the invested proportion becomes equal to 1 and a short sale, which is a negative investment, is not allowed. Then the problem is formulated as

$$P_5 : \begin{cases} \text{Maximize} & \frac{r'x - r_F}{\sqrt{x'Vx}}, \\ \text{subject to} & e'x = 1, \\ & x \geq 0. \end{cases}$$

This is in the same form as problem P_3 and the proposed algorithm can be directly applied.

We consider the expected rate of return \bar{r} against the standard deviation σ . Let point $A(\sigma_A, \bar{r}_A)$ in the space of σ and \bar{r} denote portfolio A , which is obtained by solving problem P_5 with $\sigma_A = \sqrt{(x^*)'Vx^*}$ and $\bar{r}_A = r'x^*$, where x^* is an optimal solution of problem P_5 . When the investor invests all of his original fund to the riskless asset, this portfolio B corresponds to point $B(0, r_F)$ in the space of σ and \bar{r} . Let X , where $0 \leq X \leq 1$, denote the fraction of the original fund that the investor places in portfolio A . The expected rate of return of the combined portfolio with riskless asset and risky portfolio is given by

$$(6.29) \quad \bar{r} = (1 - X)r_F + X\bar{r}_A = r_F + \frac{\bar{r}_A - r_F}{\sigma_A}\sigma,$$

where $\sigma = X\sigma_A$. As illustrated in Figure 6.1, (6.29) means that the combined portfolio is expressed as a line segment AB in the space of σ and \bar{r} . The investor chooses a certain portfolio, i.e., a combination of portfolio A and B , according to his risk preference. For example, a risk-averse investor would place a small fraction of his fund on portfolio A ,

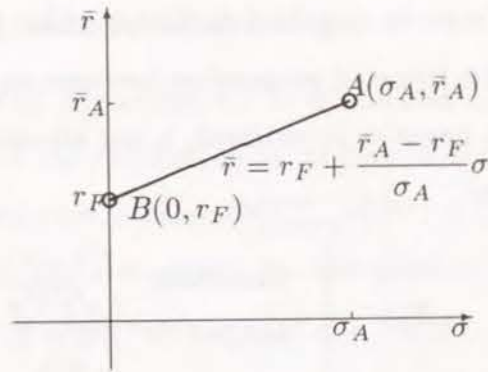


Figure 6.1. The combined portfolio of A and B

while a risk-prone investor would place a large fraction of his fund on portfolio A.

(Example 6.1) Consider a portfolio selection problem with six investments as given in Table 6.1. The rate of riskless asset is set to 4.5%. This model consists of three blocks. The derivatives $DZ^1(w)$, $DZ^2(w)$ and $DZ^3(w)$ are given as follows, where the variables in the parentheses denote the corresponding basic variables, which have positive allocations.

$$(6.30a) \quad DZ^1(w) = \begin{cases} 9w - 9R, & 0 \leq w < \xi_1^1, & (y_1) \\ \frac{108}{37}w - \frac{318}{37}R, & w \geq \xi_1^1, & (y_1, y_2) \end{cases}$$

$$(6.30b) \quad DZ^2(w) = \begin{cases} 25w - 8R, & 0 \leq w < \xi_1^2, & (y_5) \\ \frac{36}{13}w - \frac{70}{13}R, & \xi_1^2 \leq w < \xi_2^2, & (y_4, y_5) \\ 4w - 6R, & w \geq \xi_2^2, & (y_4) \end{cases}$$

$$(6.30c) \quad DZ^3(w) = w - 5R, \quad (y_6)$$

Table 6.1. Parameters of six assets

Asset number	Expected return	Standard deviation	Correlation coefficients					
			#1	#2	#3	#4	#5	#6
#1	8(%)	4(%)	1	-0.5	0	0	0	0
#2	9	3	-0.5	1	0	0	0	0
#3	3	1	0	0	1	0.5	0.4	0
#4	6	2	0	0	0.5	1	0.8	0
#5	8	5	0	0	0.4	0.8	1	0
#6	5	1	0	0	0	0	0	1

where

$$(6.31) \quad \xi_1^1 = \frac{1}{15}R, \quad \xi_1^2 = \frac{2}{17}R, \quad \xi_2^2 = \frac{1}{2}R.$$

Figure 6.2 illustrates the derivatives of objective functions of subproblems. The derivatives at $w = 0$ and $w = \xi_i^k$ are given by

$$(6.32) \quad \begin{aligned} DZ^1(0) &= -9R \\ DZ^1(\xi_1^1) &= -\frac{42}{5}R = s_{\alpha_{1,1}} \\ DZ^2(0) &= -8R \\ DZ^2(\xi_1^2) &= -\frac{86}{17}R = s_{\alpha_{2,1}} \\ DZ^3(0) &= -5R \\ DZ^2(\xi_2^2) &= -4R = s_{\alpha_{3,1}} \end{aligned}$$

Corresponding to the above values of α , we have the following results.

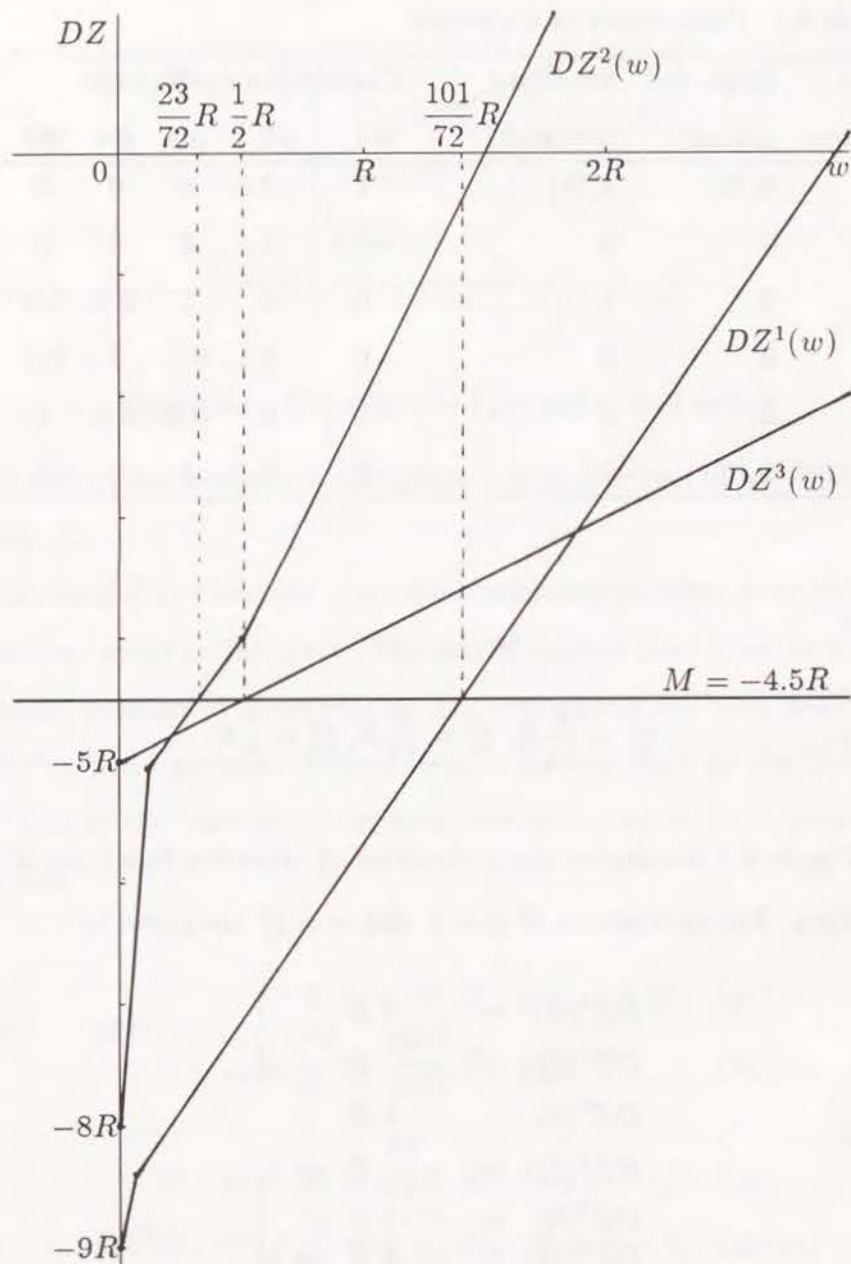


Figure 6.2. Derivatives of objective functions of subproblems

(i) $\alpha = \alpha_{1,1}$:

$$(6.33) \quad \begin{aligned} w_1 &= \frac{1}{2}R \notin \Omega^1(w) = \{w \mid 0 \leq w \leq \xi_1^1\} \\ w_2 &= 0 \\ w_3 &= 0 \end{aligned}$$

There is no optimal solution.

(ii) $\alpha = \alpha_{2,1}$:

$$(6.34) \quad \begin{aligned} w_1 &= \frac{101}{72}R \in \Omega^1(w) = \{w \mid 0 \leq w \leq \xi_1^1\} \\ w_2 &= \frac{7}{50}R \notin \Omega^2(w) = \{w \mid 0 \leq w \leq \xi_1^2\} \\ w_3 &= 0 \end{aligned}$$

There is no optimal solution.

(iii) $\alpha = \alpha_{3,1}$:

$$(6.35) \quad \begin{aligned} w_1 &= \frac{101}{72}R \in \Omega^1(w) = \{w \mid 0 \leq w \leq \xi_1^1\} \\ w_2 &= \frac{23}{72}R \in \Omega^2(w) = \{w \mid 0 \leq w \leq \xi_1^2\} \\ w_3 &= \frac{1}{2}R \in \Omega^3(w) = \{w \mid w \geq 0\} \end{aligned}$$

In this case, we have $R^* = \frac{9}{20}$ from (6.25) and an optimal portfolio obtained from (6.26) is shown in Table 6.2. The expected return is 7.47500% and the standard deviation of return is 1.15704%. The

probability that the total return exceeds 4.5% is 99.49%.

Table 6.2. Optimal portfolio of Example 6.1

Allocation to block	Optimal portfolio
$w_1^* = 101/160$	$y_1^* = 39/160$
	$y_2^* = 62/160$
$w_2^* = 23/160$	$y_3^* = 0$
	$y_4^* = 19/160$
	$y_5^* = 4/160$
$w_3^* = 36/160$	$y_6^* = 36/160$

6.3. Index Models

6.3.1. A single index model and a multi index model

A capital asset pricing model (CAPM) introduced by Sharpe [S3] explains the relation between risk and return on investment in a stock market. The risk of investment is assumed to be linearly decomposed into the risk common to all investments caused by a market and the risk individual to each investment caused by other factors. This model is called as a single index model, since it is expressed through the market index only.

$$(6.36) \quad c_i - r_f = \alpha_i + \beta_i(r_m - r_f) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

where

c_i : the rate of return on investment i .

r_m : the rate of return on market portfolio, which is the portfolio of all investments in the market. In the efficient market, r_m

is assumed to have a normal distribution $CalN(\bar{r}_m, \sigma_m^2)$. The distribution parameters should be estimated. Here we assume to be faced with the situation that they have been already estimated and we choose a suitable portfolio.

r_f : the rate of return on riskless asset.

β_i : beta-value of investment i , which is a measure of the responsiveness of investment i to changes in the market index.

α_i : alpha-value of investment i , which is the return on investment i that is independent of changes in the market index.

ϵ_i : a random noise of investment i with a normal distribution $CalN(0, \sigma_{\epsilon_i}^2)$, which is independent of the market index.

Replacing $\alpha_i + r_f(1 - \beta_i)$ by α_i , (6.36) is simply rewritten by

$$(6.37) \quad c_i = \alpha_i + \beta_i r_m + \epsilon_i, \quad i = 1, 2, \dots, n.$$

The values of α_i , β_i and $\sigma_{\epsilon_i}^2$ cannot be directly observed from historical data and might change over time. When α_i , β_i and $\sigma_{\epsilon_i}^2$ are viewed as constants through time, we can estimate them by using regression analysis.

Because a single index model is sometimes too simple to explain the practical stock market, a multi index model is introduced based on an arbitrage pricing theory (APT). In the APT, the rate of return on investment is explained by common indices besides the market index. A multi index model introduces extra indices in order to capture additional information, which are incorporated into ϵ_i in a single index model. Then the rate of return on investment is expressed as the following linear combination of indices I_1, I_2, \dots, I_L .

$$(6.38) \quad c_i = \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + \dots + \beta_{iL}I_L + \epsilon_i, \quad i = 1, 2, \dots, n,$$

where I_l is the l -th index with a normal distribution $CalN(\bar{I}_l, \sigma_l^2)$ which is independent each other and β_{il} is a measure of responsiveness of investment i to changes in the l -th index. It is also assumed that I_l is independent of ϵ_i .

When $L = 1$ and $I_1 = r_m$, a multi index model coincides with a single index model. A multi index model can describe the dependency on all indices, and has a higher explanatory power of risk and return on investment than a single index model. However, it is important to take notice of the multicollinearity between indices and stability of the coefficient estimates.

6.3.2. Formulation of the problem of a single index model

Let be n investments infinitely divisible. We consider the following probability maximizing model of a stochastic linear knapsack problem P_1 .

$$P_1 : \begin{cases} \text{Maximize} & \Pr \left(\sum_{j=1}^n c_j x_j \geq d \right), \\ \text{subject to} & \sum_{j=1}^n a_j x_j = b, \\ & 0 \leq x_j \leq \gamma_j, \quad j = 1, 2, \dots, n, \end{cases}$$

which is the same to the problem considered in Chapter 5 except that $c = (c_j)$ is a vector of the rate of return on investment expressed by a single index model (6.37). After appropriate variable transformations, we have an equivalent deterministic problem P_3 and the auxiliary prob-

lem $P(R)$ as follows.

$$P_3 : \begin{cases} \text{Maximize} & \frac{\mu'y - d}{\sqrt{y'Vy}}, \\ \text{subject to} & e'y = 1, \\ & 0 \leq y \leq \gamma, \end{cases}$$

$$P(R) : \begin{cases} \text{Maximize} & R\mu'y - \frac{1}{2}y'Vy, \\ \text{subject to} & e'y = 1, \\ & 0 \leq y \leq \gamma, \end{cases}$$

where

$$(6.39) \quad \begin{aligned} \mu &= E[c] = \alpha + \beta\bar{r}_m, \\ V &= V[c] = \beta\beta'\sigma_m^2 + \text{diag}(\sigma_{e1}^2, \sigma_{e2}^2, \dots, \sigma_{en}^2). \end{aligned}$$

Problem $P(R)$ has a unique solution, which is the y -part of a solution of the following Kuhn-Tucker conditions.

$$(6.40) \quad \begin{aligned} -R\mu + \sigma_m^2(\beta'y) \cdot \beta + \text{diag}(\sigma_{e1}^2, \sigma_{e2}^2, \dots, \sigma_{en}^2) \cdot y + \lambda \cdot e + \xi_1 - \xi_2 &= 0 \\ e'y = 1, \quad 0 \leq y \leq \gamma, \quad \xi_1, \xi_2 \geq 0 \\ \xi_{1j}(y_j - \gamma_j) = 0, \quad \xi_{2j}y_j = 0, \quad j = 1, 2, \dots, n. \end{aligned}$$

Introducing $t = \sigma_m(\beta'y)$ as a parameter, the solution of (6.40) can be expressed with three parameters R , λ and t as follows.

$$(6.41) \quad \begin{aligned} \xi_{1j} &= \max(0, R\mu_j - \sigma_m\beta_j t - \sigma_{ej}^2\gamma_j - \lambda), \\ \xi_{2j} &= \max(0, -R\mu_j + \sigma_m\beta_j t + \lambda), \end{aligned}$$

$$(6.42) \quad y_j = \begin{cases} \gamma_j, & \text{if } \lambda \leq \Lambda_j^L, \\ \frac{R\mu_j - \sigma_m\beta_j t - \lambda}{\sigma_{ej}^2}, & \text{if } \Lambda_j^L < \lambda < \Lambda_j^U, \\ 0, & \text{if } \lambda \geq \Lambda_j^U, \end{cases}$$

where

$$(6.43) \quad \begin{aligned} \Lambda_j^L &= -\sigma_{e_j}^2 \gamma_j + R\mu_j - \sigma_m \beta_j t, \\ \Lambda_j^U &= R\mu_j - \sigma_m \beta_j t. \end{aligned}$$

From Theorem 5.4, $y(R^*)$ is an optimal solution of problem P_3 , where R^* is a solution of $T(R) = 0$. The uniqueness of R^* has been shown in Section 5.3.

6.3.3. A solution algorithm

Recall that Λ_j^L and Λ_j^U , $j = 1, 2, \dots, n$, are linear functions of R and t . The R^* that gives an optimal solution of problem P_3 is unique and is a solution of equation $T(R) = 0$. A procedure to decide R^* by an extended binary search will be discussed later in this subsection. Here we consider a solution algorithm for a given R . Then Λ_j^L and Λ_j^U , $j = 1, 2, \dots, n$, are linear functions of parameter t . They are n pairs of parallel line segments in the space of t and Λ on interval (t_m, t_M) as shown in Figure 6.3.

$$(6.44) \quad \begin{aligned} t_m &= \min\{\sigma_m \beta' y | e'y = 1, 0 \leq y \leq \gamma\}, \\ t_M &= \max\{\sigma_m \beta' y | e'y = 1, 0 \leq y \leq \gamma\}. \end{aligned}$$

Note that the number of their intersection points is at most $2n(n-1) = O(n^2)$. We denote the different values of t at the intersection points as

$$(6.45) \quad t_1 < t_2 < \dots < t_N,$$

where N is their number. Note that, for any $t \in [t_i, t_{i+1}]$, the order of values of Λ_j^L and Λ_j^U , $j = 1, 2, \dots, n$, is independent of the value of t .

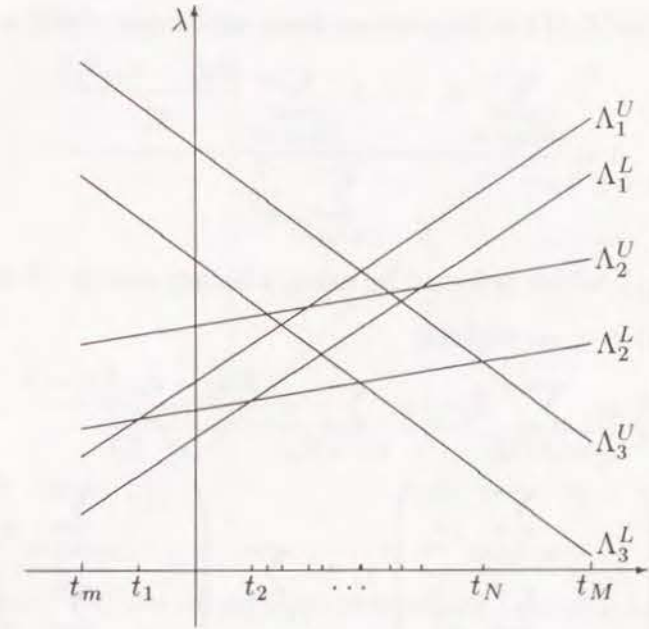


Figure 6.3. An illustrative example of t vs. Λ for $n = 3$

When the value of t is fixed, we can sort all Λ_j^L and Λ_j^U , $j = 1, 2, \dots, n$ and determine the form of y_j , $j = 1, 2, \dots, n$, in (6.42) for each λ . Now we decompose a set $\{1, 2, \dots, n\}$ into the following index sets for each λ .

$$(6.46) \quad \begin{aligned} J_1(\lambda) &= \{j | \lambda \leq \Lambda_j^L\}, \\ J_2(\lambda) &= \{j | \Lambda_j^L < \lambda < \Lambda_j^U\}, \\ J_3(\lambda) &= \{j | \lambda \geq \Lambda_j^U\}. \end{aligned}$$

Then we have

$$(6.47) \quad y_j(t) = \begin{cases} \gamma_j, & j \in J_1(\lambda), \\ \frac{R\mu_j - \sigma_m \beta_j t - \lambda}{\sigma_{e_j}^2}, & j \in J_2(\lambda), \\ 0, & j \in J_3(\lambda). \end{cases}$$

Parameter λ in (6.47) is determined from constraint $e'y(t) = 1$, that is,

$$(6.48) \quad \lambda = \frac{\sum_{j \in J_1(\lambda)} \gamma_j - 1 + \sum_{j \in J_2(\lambda)} \frac{R\mu_j - \sigma_m \beta_j t}{\sigma_{e_j}^2}}{\sum_{j \in J_2(\lambda)} \frac{1}{\sigma_{e_j}^2}}$$

for $t \in [t_i, t_{i+1}]$, which is found by using a binary search. It also satisfies that $\sigma_m \beta' y(t) = t$, so we have

$$(6.49) \quad \begin{aligned} \beta' y(t) &= \sum_{j \in J_1(\lambda)} \beta_j \gamma_j + \sum_{j \in J_2(\lambda)} \beta_j \frac{R\mu_j - \sigma_m \beta_j t - \lambda}{\sigma_{e_j}^2} \\ &= \sum_{j \in J_1(\lambda)} \left(\beta_j - \frac{\sum_{i \in J_2(\lambda)} \frac{\beta_i}{\sigma_{e_i}^2}}{\sum_{i \in J_2(\lambda)} \frac{1}{\sigma_{e_i}^2}} \right) \gamma_j + \sum_{j \in J_2(\lambda)} \left(\beta_j - \frac{\sum_{i \in J_2(\lambda)} \frac{\beta_i}{\sigma_{e_i}^2}}{\sum_{i \in J_2(\lambda)} \frac{1}{\sigma_{e_i}^2}} \right) \frac{R\mu_j}{\sigma_{e_j}^2} \\ &\quad - \sum_{j \in J_2(\lambda)} \left(\beta_j - \frac{\sum_{i \in J_2(\lambda)} \frac{\beta_i}{\sigma_{e_i}^2}}{\sum_{i \in J_2(\lambda)} \frac{1}{\sigma_{e_i}^2}} \right) \frac{\beta_j}{\sigma_{e_j}^2} t, \end{aligned}$$

and the coefficient of t is

$$(6.50) \quad \sum_{j \in J_2(\lambda)} \left(\beta_j - \frac{\sum_{i \in J_2(\lambda)} \frac{\beta_i}{\sigma_{e_i}^2}}{\sum_{i \in J_2(\lambda)} \frac{1}{\sigma_{e_i}^2}} \right) \frac{\beta_j}{\sigma_{e_j}^2} = \frac{\sum_{i,j \in J_2(\lambda)} \frac{(\beta_i - \beta_j)^2}{\sigma_{e_i}^2 \sigma_{e_j}^2}}{\sum_{j \in J_2(\lambda)} \frac{1}{\sigma_{e_j}^2}} \geq 0.$$

Therefore, $\beta' y(t)$ is a piecewise linear and nonincreasing function of t as illustrated in Figure 6.4. Let t^* denote the value of t such that $\sigma_m \beta' y(t^*) = t^*$. Then we have

$$(6.51) \quad \begin{aligned} \sigma_m \beta' y(t) < t &\iff t^* < t, \\ \sigma_m \beta' y(t) = t &\iff t^* = t, \\ \sigma_m \beta' y(t) > t &\iff t^* > t. \end{aligned}$$

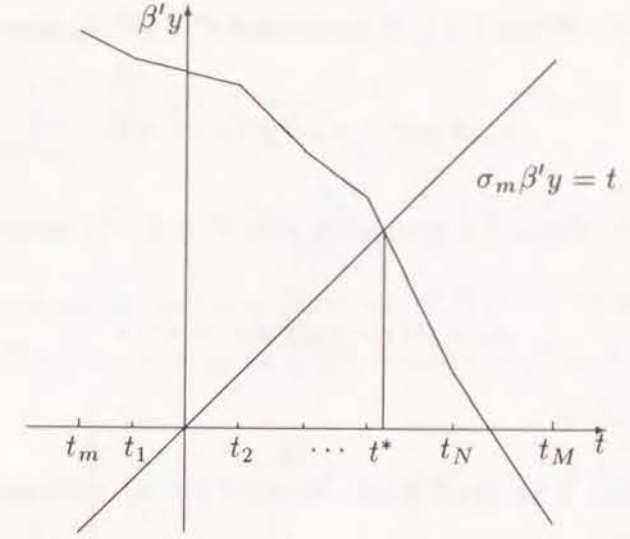


Figure 6.4. An illustrative example of t vs. $\beta' y$ for $n = 3$

Now we propose an algorithm to obtain t^* for a fixed R .

Algorithm 6.2.

- Step 1. Calculate t_m and t_M and let $t^L \leftarrow t_m$ and $t^U \leftarrow t_M$.
- Step 2. Let $t \leftarrow (t^L + t^U)/2$.
- Step 3. Sort Λ_j^L and Λ_j^U , $j = 1, 2, \dots, n$, for fixed t .
- Step 4. Obtain $y_j(t)$, $j = 1, 2, \dots, n$, and $\lambda(t)$.
- Step 5. Calculate \bar{t} such that $\sigma_m \beta' y(\bar{t}) = \bar{t}$.
- Step 6. If \bar{t} and t belong to the same interval, then terminate by concluding that $t^* = \bar{t}$. Otherwise, if $\bar{t} < t$, let $t^U \leftarrow t$ and if $\bar{t} > t$, let $t^L \leftarrow t$. Return to Step 2.

THEOREM 6.4. Algorithm 6.2 finds the t^* such that $\sigma_m \beta' y(t^*) = t^*$ in $O(n^3 \log n)$ computational time.

PROOF: First note that the form of $y_j(t)$ in (6.47) is the same for all t belongs to the same interval. Thus if \bar{t} and t belong to the same interval,

\bar{t} is equal to t^* . When $\bar{t} < t$, it holds that $t^* \in [t^L, t]$, since

$$(6.52) \quad \sigma_m \beta' y(t) < \sigma_m \beta' y(\bar{t}) = \bar{t} < t$$

implies $t^* < t$. When $\bar{t} > t$, it holds that $t^* \in [t, t^U]$, since

$$(6.53) \quad \sigma_m \beta' y(t) > \sigma_m \beta' y(\bar{t}) = \bar{t} > t$$

implies $t^* > t$.

From Step 2 to Step 6 are iterated for all intervals of t whose number is $O(n^2)$ in the worst case. The computational complexity for each step is as follows.

Step 1. $O(n \log n)$ to calculate t_m and t_M .

Step 2. $O(1)$

Step 3. $O(n \log n)$ to sort $O(n)$ Λ 's.

Step 4. $O(n \log n)$ to find λ by binary search.

Step 5. $O(n)$ to calculate \bar{t} .

Step 6. $O(n)$ to find whether \bar{t} and t belongs to the same interval or not.

Therefore, Algorithm 6.2 finds t^* in $O(n^3 \log n)$ computational time. ■

Since t^* and the corresponding λ^* satisfy linear equations $e'y = 1$ and $\sigma_m \beta' y = t$ simultaneously, that is,

$$(6.54) \quad \sigma_m \sum_{j \in J_2(\lambda^*)} \frac{\beta_j}{\sigma_{ej}^2} t^* + \sum_{j \in J_2(\lambda^*)} \frac{1}{\sigma_{ej}^2} \lambda^* = \sum_{j \in J_2(\lambda^*)} \frac{\mu_j}{\sigma_{ej}^2} R + \sum_{j \in J_1(\lambda^*)} \gamma_j - 1,$$

and

$$(6.55) \quad (\sigma_m^2 \sum_{j \in J_2(\lambda^*)} \frac{\beta_j^2}{\sigma_{ej}^2} + 1)t^* + \sigma_m \sum_{j \in J_2(\lambda^*)} \frac{\beta_j}{\sigma_{ej}^2} \lambda^* = \sigma_m \sum_{j \in J_2(\lambda^*)} \frac{\beta_j \mu_j}{\sigma_{ej}^2} R + \sigma_m \sum_{j \in J_1(\lambda^*)} \beta_j \gamma_j,$$

t^* and λ^* are expressed as linear functions of R in the neighborhood of the fixed R . Then we have y_j as linear function of R as follows.

$$(6.56) \quad y_j = \begin{cases} \gamma_j, & j \in J_1(\lambda^*), \\ \frac{\mu_j}{\sigma_{ej}^2} R - \frac{\sigma_m \beta_j}{\sigma_{ej}^2} t^* - \frac{1}{\sigma_{ej}^2} \lambda^*, & j \in J_2(\lambda^*), \\ 0, & j \in J_3(\lambda^*). \end{cases}$$

Note that

$$(6.57) \quad \bigcap_{j \in J_2(\lambda^*)} \left\{ R \mid 0 \leq \frac{\mu_j}{\sigma_{ej}^2} R - \frac{\sigma_m \beta_j}{\sigma_{ej}^2} t^* - \frac{1}{\sigma_{ej}^2} \lambda^* \leq \gamma_j \right\}$$

expresses an interval and we denote it by $[R^l, R^u]$. The forms of y_j in (6.56) are the same for all $R \in [R^l, R^u]$.

R^* is determined so as to satisfy $T(R) = 0$. $T(R)$ is given as

$$(6.58) \quad T(R) = \left(\sum_{j \in J_1(\lambda^*)} \mu_j \gamma_j - d \right) R - \sigma_m \sum_{j \in J_1(\lambda^*)} \beta_j \gamma_j t^* - \left(\sum_{j \in J_1(\lambda^*)} \gamma_j - 1 \right) \lambda^* - \sum_{j \in J_1(\lambda^*)} \sigma_{ej}^2 \gamma_j^2,$$

which is a linear function of R from taking into account (6.54) and (6.55). Since $T(R)$ is a continuous and piecewise linear function of R such that

$$(6.59) \quad \begin{aligned} \lim_{R \rightarrow 0} T(R) &< 0, \\ \lim_{R \rightarrow \infty} T(R) &> 0, \end{aligned}$$

R^* is found by the following algorithm using an extended binary search.

Algorithm 6.3.

Step 1. Calculate R_m and R_M and let $R^L \leftarrow R_m$ and $R^U \leftarrow R_M$.

Step 2. Let $R \leftarrow (R^L + R^U)/2$.

Step 3. Apply Algorithm 6.2 with the current value of R .

Step 4. Obtain R^l , R^u and \hat{R} such that $T(\hat{R}) = 0$.

Step 5. If $\hat{R} \in [R^l, R^u]$, then terminate by concluding that $R^* = \hat{R}$.

Otherwise, if $T(R) < 0$, let $R^L \leftarrow R^u$ and if $T(R) > 0$, let $R^U \leftarrow R^l$. Return to Step 2.

The computational complexity is dominated by the number of iteration from Step 2 to Step 5. Note that the number K of parabolas in the efficient frontier (5.11) is the same to the number of intervals of R defined by (6.57). Since it should be iterated K times in the worst case, Algorithm 6.3 takes $O(Kn^3 \log n)$ computational time.

6.3.4. Extension to a multi index model

This section discusses how to extend Algorithm 6.2 to the problem of the multi index model (6.38). The rate of return on investment expressed by a multi index model (6.38) is a normally distributed random vector with

$$(6.60) \quad \begin{aligned} \mu &= E[c] = \alpha + \beta_1 \bar{I}_1 + \beta_2 \bar{I}_2 + \cdots + \beta_L \bar{I}_L, \\ V &= V[c] = \beta_1 \beta_1' \sigma_1^2 + \beta_2 \beta_2' \sigma_2^2 + \cdots + \beta_L \beta_L' \sigma_L^2 \\ &\quad + \text{diag}(\sigma_{e1}^2, \sigma_{e2}^2, \dots, \sigma_{en}^2). \end{aligned}$$

Introducing L parameters $t_l = \sigma_l(\beta_l' y)$, $l = 1, 2, \dots, L$, the y -part of a solution of the Kuhn-Tucker conditions for the auxiliary problem $P(R)$

is given by

$$(6.61) \quad y_j = \begin{cases} \gamma_j, & \text{if } \lambda \leq \bar{\Lambda}_j^L, \\ \frac{R\mu_j - \sum_{l=1}^L \sigma_l \beta_{jl} t_l - \lambda}{\sigma_{e_j}^2}, & \text{if } \bar{\Lambda}_j^L < \lambda < \bar{\Lambda}_j^U, \\ 0, & \text{if } \lambda \geq \bar{\Lambda}_j^U, \end{cases}$$

where

$$(6.62) \quad \begin{aligned} \bar{\Lambda}_j^L &= -\sigma_{e_j}^2 \gamma_j + R\mu_j - \sum_{l=1}^L \sigma_l \beta_{jl} t_l, \\ \bar{\Lambda}_j^U &= R\mu_j - \sum_{l=1}^L \sigma_l \beta_{jl} t_l. \end{aligned}$$

When we give the values of t_1, t_2, \dots, t_L , we can sort $\bar{\Lambda}_j^L$ and $\bar{\Lambda}_j^U$, $j = 1, 2, \dots, n$ and determine the form of y_j , $j = 1, 2, \dots, n$ in (6.61) for each λ . Parameter λ is determined from constraint $e'y(t_1, t_2, \dots, t_L) = 1$ with parameters t_1, t_2, \dots, t_L in a manner analogous to (6.48). Let $t_1^*, t_2^*, \dots, t_L^*$ denote the values of t_1, t_2, \dots, t_L such that

$$(6.63) \quad \sigma_l \beta_l' y(t_1^*, t_2^*, \dots, t_L^*) = t_l^*, \quad l = 1, 2, \dots, L.$$

Since $\sigma_l \beta_l' y(t_1, t_2, \dots, t_L)$, $l = 1, 2, \dots, L$ are piecewise linear and non-increasing functions of t_1, t_2, \dots, t_L , we have

$$(6.64) \quad \begin{aligned} \sigma_l \beta_l' y(t_1, t_2, \dots, t_L) < t_l &\iff t_l^* < t_l, \\ \sigma_l \beta_l' y(t_1, t_2, \dots, t_L) = t_l &\iff t_l^* = t_l, \\ \sigma_l \beta_l' y(t_1, t_2, \dots, t_L) > t_l &\iff t_l^* > t_l, \end{aligned}$$

for $l = 1, 2, \dots, L$. We can solve problem $P(R)$ by Algorithm 6.2 using a binary search for each t_l . The L -dimensional space of t_1, t_2, \dots, t_L is

divided into several cells, in each of which the order of the values of $\bar{\Lambda}_j^L$ and $\bar{\Lambda}_j^U$, $j = 1, 2, \dots, n$, is uniquely determined. Since the number of such cells is $O(n^{2^L})$, we require $O(n^{2^L+1} \log n)$ computational time to solve problem $P(R)$ of a multi index model.

6.4. Conclusion

We discuss an application of a probability maximizing model of a stochastic linear knapsack problem to portfolio selection problems.

The first problem considered in this chapter assumes that the variance covariance matrix of random cost coefficients is block diagonal. This problem is decomposed into several subproblems, for which the auxiliary quadratic problems are defined and solved parametrically. Our algorithm finds an optimal solution effectively by using a ranking method to merge subproblems. This problem is applied to the simple portfolio selection problem, which finds a portfolio that maximizes the ratio of excess return to standard deviation.

The second problem considered in this chapter assumes that the random cost coefficient, i.e., the rate of return on investment, is expressed by index models. A single index model and a multi index model are used to explain risk and return on investment in the stock market. We mainly considered a single index model, which expresses the rate of return on investment as a linear regression model of the market index. By introducing a parameter corresponding to the market index, we solve the auxiliary quadratic problem parametrically and propose an efficient solution algorithm to find an optimal portfolio.

The worst case complexity of the proposed algorithm to solve problem P_2 is $O(Kn^3 \log n)$ and the number K may be at most $O(2^n)$. However, when the distribution of $\pi_0, \pi_1, \dots, \pi_K$ on the π -axis is uniform, Algorithm 6.3 finds an optimal R^* in $O(n)$ iterations by using a binary search. Moreover, when the distribution of t_1, t_2, \dots, t_N on (t_m, t_M) is uniform, Algorithm 6.2 finds an optimal t^* in $O(\log n)$ iterations. Therefore the average complexity is approximately $O(n^2(\log n)^2)$. The problem of a multi index model can be solved by the procedure similar to the problem of a single index model, but the computational complexity becomes higher order polynomial.

In the future research, it would be necessary to investigate the distribution of stock return, which is considered not to have exactly normal distribution [N2]. The coefficients in index models are estimated by means of regression analysis based on the historical data. In case we rely on an estimator of unknown coefficients in the index models, it is worthwhile to investigate the problem based on the confidence region of estimates as discussed in the previous chapters from the viewpoint of a game theoretic strategy. The average complexity of the proposed algorithms applied to the practical finance data for the problem of a single index model as well as a multi index model should be investigated, too.

Chapter 7.

A STOCHASTIC IMPROVEMENT METHOD FOR STOCHASTIC PROGRAMMING

7.1. Introduction

This chapter discusses a stochastic improvement method for stochastic programming. Many optimization problems arisen in practice are formulated as stochastic programming problems. However, in most of them, the parameters describing the problems are unknown or known with uncertainty. Such problems necessitate statistical approaches to estimate the unknown parameters. Many studies of statistical approaches for stochastic programming are proposed, for example, a Bayesian analysis [B8, J1], a minimax model [D4], a prediction of the regions of an optimal solution and an optimal value [C3] and a confidence region method [M7, M8].

The mathematical models show the relationships that exist among the system variables. We are interested in identifying the relationships and in estimating the parameters of the relationships. Since many models are assumed to have a structure expressed by a linear equation or by a system of simultaneous equations, a multivariate regression analysis is a useful statistical technique.

In practical applications, it is desirable that the estimates should be updated whenever we obtain new available information. When we would like to improve the accuracy of estimates, we observe new available data and update the estimates of parameters of the relationships. It is not, however, efficient to solve such a problem from scratch every time the estimates are updated. It is desirable to update the optimization problem as well as its optimal solution by modifying the previous data. For such purpose, we propose a stochastic improvement method for the linear programming problem that contains unknown coefficients in the constraints, which are iteratively improved by using newly available data obtained one after another. The stochastic improvement method solves the problem by an iterative algorithm alternating between the improvement phase of optimal solutions and the updating phase of the estimates of unknown coefficients.

Section 7.2 provides a formulation of the linear programming problem considered in this paper and describes a procedure for updating the estimates. A solution algorithm of the stochastic improvement method is discussed in Section 7.3, which consists of the feasibility step for obtaining a feasible point and the optimality step for obtaining an improved point. We update the estimates of unknown coefficients as well as solution by making use of an affine scaling method, whenever new statistical data are delivered. Section 7.4 shows the consistency of the estimated feasible region and proves that a point sequence generated by the proposed algorithm converges to a point that gives the optimal value with probability one. The results of the numerical experiments in Section 7.5 illustrate the convergence of proposed algorithm. Finally,

Section 7.6 gives a summary and discusses the direction of further research.

7.2. Formulation of the Problem

We consider the following problem.

$$P : \begin{cases} \text{Minimize} & \sum_{j=1}^n c_j x_j, \\ \text{subject to} & \sum_{j=1}^n a_{ij} x_j + b_i \leq 0, \quad i = 1, 2, \dots, m, \end{cases}$$

where both $A = (a_{ij})$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, and $b = (b_i)$, $i = 1, 2, \dots, m$, are fixed coefficients but known with uncertainty. The true values of A and b are only given by estimation with perfect information. We assume that problem P with the true values of coefficients has a bounded feasible region. Then we aim to obtain an optimal solution of problem P with the true values of coefficients after estimating them using additional sample information in an adaptive way.

We consider the situation that, for given sample point x , the value of $\sum_{j=1}^n a_{ij} x_j + b_i$ is available for the observation y_i including normally distributed random error u_i , i.e., $y_i = \sum_{j=1}^n a_{ij} x_j + b_i + u_i$. Then the multivariate regression model defined by (7.1) is useful to estimate the unknown coefficients [M2].

$$(7.1) \quad Y^N = X^N B + U^N,$$

where

$$\begin{aligned}
 Y^N &= \begin{pmatrix} y_1^1 & \cdots & y_m^1 \\ y_1^2 & \cdots & y_m^2 \\ \vdots & \ddots & \vdots \\ y_1^N & \cdots & y_m^N \end{pmatrix} && : \text{matrix of observations} \\
 X^N &= \begin{pmatrix} x_1^1 & \cdots & x_n^1 & 1 \\ x_1^2 & \cdots & x_n^2 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_1^N & \cdots & x_n^N & 1 \end{pmatrix} && : \text{matrix of sample points} \\
 B &= \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \\ b_1 & \cdots & b_m \end{pmatrix} && : \text{matrix of parameters to be estimated} \\
 U^N &= \begin{pmatrix} u_1^1 & \cdots & u_m^1 \\ u_1^2 & \cdots & u_m^2 \\ \vdots & \ddots & \vdots \\ u_1^N & \cdots & u_m^N \end{pmatrix} && : \text{matrix of errors}
 \end{aligned}$$

We assume that X^N has a full column rank and that $(u_1^k, u_2^k, \dots, u_m^k)'$ is normally distributed independent random errors with the mean zero and the variance matrix $\Sigma = (\sigma_{ij})$. These assumptions are essential for normal distribution of the maximum likelihood estimator. The maximum likelihood estimator of regression parameter matrix B and variance covariance matrix Σ of disturbance are given by

$$(7.2) \quad \hat{B} = (\hat{A}, \hat{b})' = ((X^N)'X^N)^{-1}(X^N)'Y^N,$$

$$(7.3) \quad \hat{\Sigma} = \frac{1}{N}(Y^N)'(I - X^N((X^N)'X^N)^{-1}(X^N)')Y^N.$$

From $E[U^N] = 0$ and

$$(7.4) \quad \begin{aligned} \hat{B} &= ((X^N)'X^N)^{-1}(X^N)'((X^N)B + U) \\ &= B + ((X^N)'X^N)^{-1}(X^N)'U^N, \end{aligned}$$

\hat{B} is unbiased for B and is a linear function of U^N . Therefore, $\hat{B} = (\hat{\beta}_{ij})$ is a random matrix which has multivariate normal distribution with

$$(7.5) \quad \begin{aligned} E[\hat{B}] &= B \\ \text{Cov}[\hat{\beta}_{ij}, \hat{\beta}_{kl}] &= \sigma_{jl}g_{ik}^N, \end{aligned}$$

where $G^N = (g_{ij}^N) = ((X^N)'X^N)^{-1}$. Let $\text{vec}B = (\beta_1', \beta_2', \dots, \beta_m)'$ denote the vector obtained by stacking the columns of B on top of one another. We obtain that

$$(7.6) \quad \begin{aligned} \text{vec}\hat{B} &= (I \otimes ((X^N)'X^N)^{-1}(X^N)') \text{vec}Y^N, \\ V[\text{vec}\hat{B}] &= ((X^N)'(\Sigma \otimes I)^{-1}X^N)^{-1}, \end{aligned}$$

where \otimes denotes the Kronecker product, which is an $(mk \times nl)$ matrix defined for an $m \times n$ matrix $P = (p_{ij})$ and a $k \times l$ matrix Q as follows.

$$(7.7) \quad P \otimes Q = \begin{pmatrix} p_{11}Q & p_{12}Q & \cdots & p_{1n}Q \\ p_{21}Q & p_{22}Q & \cdots & p_{2n}Q \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}Q & p_{m2}Q & \cdots & p_{mn}Q \end{pmatrix}$$

If each row of X^N is independent of each other, we have $V[\text{vec}\hat{B}] \rightarrow 0$ as sample size N tends to infinity. Then \hat{B} is consistent, that is, it holds that

$$(7.8) \quad \lim_{N \rightarrow \infty} \Pr(\|\hat{B} - B\| < \epsilon) = 1$$

for any positive ϵ .

Noticing that \hat{B} are normally distributed random variables, we have the following estimated problem \hat{P} :

$$\hat{P} : \begin{cases} \text{Minimize} & c'x, \\ \text{subject to} & \hat{A}x + \hat{b} \leq 0, \end{cases}$$

by replacing the unknown matrix B in problem P with the estimator \widehat{B} . When we obtain k observations $(x^1, y^1), (x^2, y^2), \dots, (x^k, y^k)$, we have the estimates $B^k = (A^k, b^k)'$ of unknown coefficients B as the realizations of \widehat{B} and the following resultant deterministic linear programming problem P^k :

$$P^k : \begin{cases} \text{Minimize} & c'x, \\ \text{subject to} & A^k x + b^k \leq 0. \end{cases}$$

This chapter discusses a solution method of problem P through the sequence of problems $\{P^k\}$ resulting from the sequential sample information.

When we obtain a new observation y^{k+1} at the sample point x^{k+1} , the estimate B^k is updated to B^{k+1} .

LEMMA 7.1. Assume that $G^k = ((X^k)'X^k)^{-1}$ exists. Then it holds

$$(7.9) \quad G^{k+1} = G^k - \frac{G^k x^{k+1} (x^{k+1})' G^k}{1 + (x^{k+1})' G^k x^{k+1}}.$$

PROOF: The proof is given by showing that $G^{k+1}(G^{k+1})^{-1} = I$. Noting that $(G^{k+1})^{-1} = (G^k)^{-1} + x^{k+1}(x^{k+1})'$ and G^k is symmetric and positive definite, we have

$$(7.10) \quad \begin{aligned} & G^{k+1}(G^{k+1})^{-1} \\ &= \left(G^k - \frac{G^k x^{k+1} (x^{k+1})' G^k}{1 + (x^{k+1})' G^k x^{k+1}} \right) \left((G^k)^{-1} + x^{k+1} (x^{k+1})' \right) \\ &= I + \frac{G^k x^{k+1} (x^{k+1})' \left((x^{k+1})' G^k x^{k+1} I - G^k x^{k+1} (x^{k+1})' \right)}{1 + (x^{k+1})' G^k x^{k+1}} \\ &= I, \end{aligned}$$

where

$$\begin{aligned} & G^k x^{k+1} (x^{k+1})' \left((x^{k+1})' G^k x^{k+1} I - G^k x^{k+1} (x^{k+1})' \right) \\ &= (x^{k+1})' G^k x^{k+1} G^k x^{k+1} (x^{k+1})' - G^k x^{k+1} (x^{k+1})' G^k x^{k+1} (x^{k+1})' \\ &= (x^{k+1})' G^k x^{k+1} G^k x^{k+1} (x^{k+1})' - (x^{k+1})' G^k x^{k+1} G^k x^{k+1} (x^{k+1})' \\ &= 0. \quad \blacksquare \end{aligned}$$

THEOREM 7.2. When we obtain a new observation $y^{k+1} = Bx^{k+1} + u^{k+1}$ at the sample point x^{k+1} with a random disturbance u^{k+1} , the estimate B^k is updated to B^{k+1} as follows.

$$(7.11) \quad B^{k+1} = \left(I - \frac{G^k x^{k+1} (x^{k+1})'}{1 + (x^{k+1})' G^k x^{k+1}} \right) (B^k + G^k x^{k+1} (y^{k+1})').$$

PROOF: We have

$$(7.12) \quad \begin{aligned} B^{k+1} &= G^{k+1} (X^{k+1})' Y^{k+1} \\ &= \left(G^k - \frac{G^k x^{k+1} (x^{k+1})' G^k}{1 + (x^{k+1})' G^k x^{k+1}} \right) \left((X^k)' Y^k + x^{k+1} (y^{k+1})' \right) \\ &= \left(I - \frac{G^k x^{k+1} (x^{k+1})'}{1 + (x^{k+1})' G^k x^{k+1}} \right) (B^k + G^k x^{k+1} (y^{k+1})') \end{aligned}$$

Notice that G^k is calculated by (7.9) recursively. \blacksquare

We consider point sequence $\{x^k\}$ which converges to a point that gives the optimal value of problem P. At the starting time, the initial estimates are calculated by more than $n+1$ samples and an initial point of the sequence is given arbitrarily, which is not necessary to be feasible to the estimated problem because it is projected to the feasible

region if not feasible. Each updated point x^{k+1} is calculated from A^k , b^k and x^k . Two types of steps are considered to get an improved point in the stochastic improvement method. One is the feasibility step that gives a point feasible to the updated feasible region, and the other is an optimality step that gives an improved point. In the following sections, we propose an algorithm that generates a point sequence $\{x^k\}$ and we will show several properties.

7.3. A Stochastic Improvement Method

The improved point x^{k+1} is given by the following two types of steps using the estimates A^k and b^k in problem P^k . Now let D and D^k denote feasible regions of problems P and P^k as follows:

$$(7.13) \quad D = \left\{ x \left| \sum_{j=1}^n a_{ij}x_j + b_i \leq 0, \quad i = 1, 2, \dots, m \right. \right\}$$

$$(7.14) \quad D^k = \left\{ x \left| \sum_{j=1}^n a_{ij}^k x_j + b_i^k \leq 0, \quad i = 1, 2, \dots, m \right. \right\}$$

Note that D is the region with true values of coefficients and D^k is an estimated region of D with the estimates of coefficients.

First, we introduce the feasibility step that constructs from x^k a feasible point to the updated feasible region of problem P^k . The projection of x^k to feasible region D^k is the main operation in the feasibility step. Let x_F^k denote the projected point of x^k . For $x^k \notin D^k$, it is assured that x^k satisfies the estimated constraints $\widehat{A}x + \widehat{b} \leq 0$ with some probability, that is, $\Pr(\widehat{A}x^k + \widehat{b} \leq 0) > 0$, since the estimators \widehat{A}

and \widehat{b} are normally distributed random variables. So we can select the values of \widetilde{A} and \widetilde{b} such that $\widetilde{A}x^k + \widetilde{b} \leq 0$ from their confidence regions under an appropriate significance level. For an infeasible point $x^k \notin D^k$, we consider the following auxiliary problem $P_{(B)}$ minimizing the object function with respect to $B = (A, b)'$.

$$P_{(B)} : \quad \left| \begin{array}{l} \text{Minimize}_{B} \quad \text{vec}(B - B^k)' V[\text{vec} \widehat{B}]^{-1} \text{vec}(B - B^k), \\ \text{subject to} \quad Ax^k + b \leq 0. \end{array} \right.$$

Problem $P_{(B)}$ has an optimal solution \widetilde{B}^* which gives the narrowest confidence region of unknown coefficients B , that is, \widetilde{B}^* belongs to the confidence region of B under the highest significance level. \widetilde{B}^* is also viewed as an optimal solution of the following problem:

$$P_{\alpha} : \quad \left| \begin{array}{l} \text{Maximize}_{B, \alpha} \quad \alpha, \\ \text{subject to} \quad B \in S_B(\alpha), \\ \quad \quad \quad Ax^k + b \leq 0, \end{array} \right.$$

where $S_B(\alpha)$ denotes the confidence region of B under a significance level α .

Since it is, however, too complicated to implement the selection of both \widetilde{A}^* and \widetilde{b}^* , we set $\widetilde{A} = A^k$ for the sake of simplicity and select only \widetilde{b}^* so as to maximize the significance level of the confidence region of b to which \widetilde{b}^* belongs. The resultant auxiliary problem minimizing with respect to b is given as follows.

$$P_{(b)} : \quad \left| \begin{array}{l} \text{Minimize}_{b} \quad (b - b^k)' V[\widehat{b}]^{-1} (b - b^k), \\ \text{subject to} \quad A^k x^k + b \leq 0, \end{array} \right.$$

where $V[\widehat{b}]$ is a submatrix of $V[\text{vec} \widehat{B}]$ corresponding to b . Then a feasible point x_F^k is given by the following problem.

$$P_F : \quad \left| \begin{array}{l} \text{Minimize}_x \quad \|x - x^k\|, \\ \text{subject to} \quad A^k x + b^k \leq 0. \end{array} \right.$$

Here we alter the feasible region D^k into \bar{D}^k as follows, because the optimality step, which will be discussed below, requires a strictly interior point of D^k .

$$(7.15) \quad \bar{D}^k = \left\{ x \mid \begin{array}{l} A_i^k x + \tilde{b}_i^* \leq 0 \quad \text{for } i \in I, \\ A_i^k x + b_i^k - \nu \leq 0 \quad \text{for } i \in I', \\ A_i^k x + b_i^k \leq 0 \quad \text{for } i \notin I \cup I' \end{array} \right\},$$

where

$$(7.16) \quad I = \{i \mid A_i^k x^k + b_i^k > 0, \tilde{b}_i^* \neq b_i^k\},$$

$$(7.17) \quad I' = \{i \mid A_i^k x^k + b_i^k \leq 0, A_i^k x_F^k + b_i^k = 0\}$$

and ν is a prescribed tolerance level. Then x_F^k is located in the interior of D^k .

Next, we consider the optimality step that gives an improved point. We utilize an affine scaling method [B2] for getting an improved point x^{k+1} . An interior point method for solving linear programming starts from a strictly interior point and generates a point sequence converging to an optimal solution through the interior of feasible region. The affine scaling variant of the Karmarkar's algorithm is one of the interior point methods, which solves the linear programming problem using an affine transformation. We consider the following iterative scheme based on the affine scaling method.

$$(7.18) \quad x^{k+1} = x_F^k + \alpha^k d^k,$$

where

$$(7.19) \quad \alpha^k = [(d^k)'(A^k)'(T^k)^2 A^k d^k]^{-\frac{1}{2}},$$

$$(7.20) \quad d^k = -((A^k)'(T^k)^2 A^k)^{-1} c,$$

$$(7.21) \quad T^k = \text{diag} \left(\frac{1}{A_1^k x_F^k + b_1^k}, \frac{1}{A_2^k x_F^k + b_2^k}, \dots, \frac{1}{A_n^k x_F^k + b_n^k} \right).$$

Note that x_F^k , which is obtained in the feasibility step, is a strictly interior point of D^k .

We repeat sampling, estimation and stochastic improvement by the feasibility step and the optimality step, until the step size of improvement becomes less than a prescribed tolerance level ν . Figure 7.1 shows the resulting algorithm SIM to solve problem P by the stochastic improvement method.

When we can obtain N_k samples $(x^{k+1}, y^{k+1}), (x^{k+2}, y^{k+2}), \dots, (x^{k+N_k}, y^{k+N_k})$ all together at one iteration, the estimated problem P^k is updated to problem P^{k+N_k} and the variance of estimator \hat{B} with $k + N_k$ samples is smaller than that of estimator \hat{B} with $k + 1$ samples. It is also desirable that the sample points are chosen independently of each other for the consistency of estimators. Since, however, the stochastic improvement method assumes that the sample point for the next observation is not independently generated but is provided by the x^{k+1} given by (7.18) based on the present sample (x^k, y^k) , the sample points are not completely independent. This dependency of the sample points may slow down the speed that the variance of estimator converges to zero. When the infinitely many sample points are restricted to a subspace of R^n , the estimator has no consistency.

procedure SIM:

begin

Give initial samples X^N and observations Y^N for $N > n$;

Let $G^N \leftarrow ((X^N)'X^N)^{-1}$ and $B^N \leftarrow G^N(X^N)'Y^N$;

Set a tolerance level ν ;

Let $k \leftarrow N$ and define initial point x^{k+1} arbitrarily;

repeat

begin

Let $k \leftarrow k + 1$;

Observe new data (x^k, y^k) ;

Let $G^k \leftarrow G^{k-1} - \frac{G^{k-1}x^k(x^k)'G^{k-1}}{1 + (x^k)'G^{k-1}x^k}$;

Let $B^k \leftarrow \left(I - \frac{G^{k-1}x^k(x^k)'}{1 + (x^k)'G^{k-1}x^k} \right) (B^{k-1} + G^{k-1}x^k(y^k)')$;

if $x^k \notin \text{int}D^k$

then Feasibility step: Obtain x_F^k and construct \bar{D}^k ;

else Let $x_F^k \leftarrow x^k$;

Optimality step: Let $x^{k+1} \leftarrow x_F^k + \alpha^k d^k$ by (7.18);

end

until $\|\alpha^k d^k\| < \nu$;

end.

Figure 7.1. Algorithm SIM

However, since the estimate B^k in (7.18)–(7.21) is one realization of normally distributed random matrix \hat{B} , that is, $\Pr(\hat{B} = B^k) = 0$ for any k , it is considered that the infinitely many sample points are re-

stricted to a proper subspace of R^n with probability zero and this model assures us of the consistency with probability one.

7.4. Consistency of the Estimated Region and Convergence of Algorithm SIM

The estimated feasible region D^k with the estimates of coefficients converges to the true feasible region D with the true values of coefficients as sample size k becomes sufficiently large. From the consistency of estimator \hat{B} , the following theorem holds.

THEOREM 7.3. *With probability one,*

$$(7.22) \quad \lim_{k \rightarrow \infty} D^k = D.$$

PROOF: It is sufficient to show that

$$(7.23) \quad \limsup_{k \rightarrow \infty} D^k \subset D \subset \liminf_{k \rightarrow \infty} D^k$$

with probability one.

If $\bar{x} \in \limsup_{k \rightarrow \infty} D^k$, then there exists a sequence $\{y^k\}_{k \in K} \rightarrow \bar{x}$ such that $y^k \in D^k$ for all $k \in K$, where K is an infinite subset of $\{1, 2, \dots\}$. Thus, $A^k y^k + b^k \leq 0$ for all $k \in K$. The consistency of estimators ensures that $A\bar{x} + b \leq 0$ with probability one, and hence $\bar{x} \in D$ with probability one. This proves $\limsup_{k \rightarrow \infty} D^k \subset D$.

If $\underline{x} \notin \liminf_{k \rightarrow \infty} D^k$, then $A^k \underline{x} + b^k \not\leq 0$ for k sufficiently large. The consistency of estimators ensures that $A\underline{x} + b \not\leq 0$ with probability one, and hence $\underline{x} \notin D$ with probability one. This proves $D \subset \liminf_{k \rightarrow \infty} D^k$. ■

To measure the distance between the estimated feasible region D^k and the true feasible region D , we consider the Hausdorff distance.

$$(7.24) \quad \delta(D^k, D) = \max(\max_{x \in D^k} \min_{y \in D} \|x - y\|, \max_{x \in D} \min_{y \in D^k} \|x - y\|),$$

where $\|\cdot\|$ denotes the Euclidean norm. From Theorem 7.3, for any positive ϵ , there exist a number n_1 and a region D_ϵ^+ , with probability one, such that

$$(7.25) \quad \delta(D, D_\epsilon^+) \leq \epsilon$$

and

$$(7.26) \quad D \subset D_\epsilon^+ = \bigcup_{k=n_1}^{\infty} D^k.$$

Moreover there also exist a number n_2 and a region D_ϵ^- , with probability one, such that

$$(7.27) \quad \delta(D, D_\epsilon^-) \leq \epsilon$$

and

$$(7.28) \quad D \supset D_\epsilon^- = \bigcap_{k=n_2}^{\infty} D^k.$$

Let x_ϵ^+ and x_ϵ^- denote optimal solutions of problems $\min\{c'x \mid x \in D_\epsilon^+\}$ and $\min\{c'x \mid x \in D_\epsilon^-\}$, respectively. The following theorem shows the convergence of algorithm SIM to a point that gives the optimal value of problem P with probability one.

THEOREM 7.4. *Let x^* denote an optimal solution of problem P with the true value of coefficients. Then $|c'x^k - c'x^*| \rightarrow 0$ as $k \rightarrow \infty$ with probability one.*

PROOF: For any positive ϵ , from $D^k \subset D_\epsilon^+$ and $\lim_{k \rightarrow \infty} D^k = D$ with probability one, we obtain $c'x_\epsilon^+ \leq c'x^k$ for k sufficiently large. Note that the sequence of objective values $\{c'x^k\}$ of problem P^k at $x = x^k$ has an accumulation point $\hat{\mu}$.

If $\hat{\mu} \leq c'x_\epsilon^-$, then $c'x^k \leq c'x_\epsilon^-$ for k sufficiently large. Because $\delta(D_\epsilon^+, D_\epsilon^-) \leq 2\epsilon$ holds with probability one, it is obvious that $|c'x_\epsilon^+ - c'x_\epsilon^-| \rightarrow 0$ as $\epsilon \rightarrow 0$. From $c'x_\epsilon^+ \leq c'x^k \leq c'x_\epsilon^-$ and $c'x_\epsilon^+ \leq c'x^* \leq c'x_\epsilon^-$, it is shown that $|c'x^k - c'x^*| \rightarrow 0$ with probability one as $k \rightarrow \infty$.

If $c'x_\epsilon^- \leq \hat{\mu}$, then $c'x_\epsilon^- \leq c'x^k$ for k sufficiently large. However, from $D_\epsilon^- \subset D^k$ and $\lim_{k \rightarrow \infty} D^k = D$ with probability one, we obtain $c'x_\epsilon^- \geq c'x^k$ for k sufficiently large. Therefore, $\lim_{k \rightarrow \infty} c'x^k = c'x_\epsilon^-$ and $c'x_\epsilon^- \rightarrow c'x^*$ as $\epsilon \rightarrow 0$. ■

Since $B^k \rightarrow B$ as $k \rightarrow \infty$, the algorithm SIM generates a point sequence $\{x^k\}$ which has an accumulation point \hat{x} . Then Theorem 7.4 implies that

$$(7.29) \quad \hat{x} \in \arg \min\{c'x \mid Ax + b \leq 0\}.$$

The stochastic improvement method solves problem P by alternating between the feasibility step and the optimality step. If we apply only the optimality step in problem P^k without updating the estimates, the sequence generated by the stochastic improvement method may converge to an optimal solution of the estimated problem P^k . Since the

sequence of problems $\{P^k\}$ converges to problem P as sample size tends to infinity, it is meaningful to apply the optimality steps several times at one iteration, as it may obtain a point close to an optimal solution of the estimated problem P^k at each iteration. Moreover, if we can obtain random samples besides the sample point x^k , the convergence speed of $\delta(D^k, D)$ with random samples is faster than the convergence speed of $\delta(D^k, D)$, obtained by using the proposed algorithm, since random samples assure the consistency of estimator \hat{B} .

7.5. Computational Results

We have conducted numerical experiments with the proposed algorithm for the test problem. The unknown coefficients in A and b are chosen at random from the sample space uniformly distributed on $\{0, 1, \dots, 49\}$ and $\{-50, -51, \dots, -99\}$, respectively. The objective function is $c'x = x_1 + x_2 + \dots + x_n$. Let an optimal solution of the problem with the true values of coefficients be previously calculated and the existence ascertained. Let the observations of $Ax^k + b$ include the random error which has an independent normal distribution $\mathcal{N}(0, \frac{1}{2})$.

We have solved the problem, where $m = 40$ and $n = 20$, twenty times by applying the proposed algorithm with $\nu = 10^{-6}$. Figure 7.2 shows the average difference between $c'x^k$ and $c'x^*$, which implies that the convergence mainly depends on the affine scaling method for k less than about 60 and, after that, x^k is considered to be close to a point that gives an optimal value of problem P^k and the convergence mainly depends on the accuracy of estimates. The convergence speed

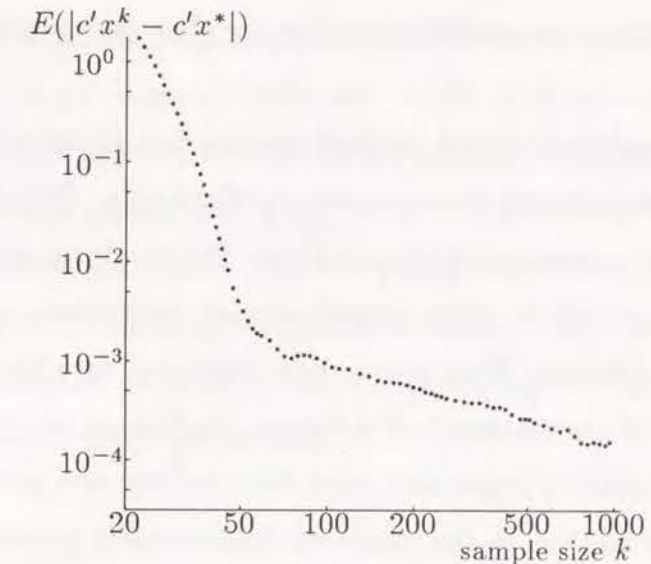


Figure 7.2. The average difference between $c'x^k$ and $c'x^*$.

for k sufficiently large is about $O(k^{-0.85})$. An optimal solution with a tolerance level ν necessitates the estimates with an accuracy ν , i.e., $\|B - B^k\| < \nu$. Therefore, when the affine scaling method (optimality step) gives a point sufficiently close to an optimal solution of the estimated problem, the number of iterations of the proposed algorithm mainly depends upon the consistency of the estimators.

7.6. Conclusion

We have introduced in this chapter a new stochastic approach based on the stochastic improvement method for a stochastic linear programming problem. This is a new statistical approach to stochastic programming. As an example of this approach, the linear programming problem with linear constraints that contain unknown coefficients is considered. The unknown coefficients are estimated by means of a

multivariate regression analysis based on the observations with random noise.

In the confidence region method discussed in Chapter 2 through Chapter 4, we estimate the unknown coefficients by their confidence regions under a certain significance level. The resulting deterministic problems are solved by using several solution procedures proposed in the previous chapters. If we receive new statistical data for improving the accuracy of the estimates of unknown coefficients, we must reconstruct the confidence region and solve the resulting new problem.

Contrary to this, in the stochastic improvement method, we can update the estimates of unknown coefficients whenever we receive new statistical data, and then improve the current solution in an adaptive way. The point sequence $\{x_k\}$ generated by the algorithm SIM converges to a point that gives the optimal value of problem P with the true values of coefficients with probability one, which is proven from the fact that the consistency of maximum likelihood estimator of unknown coefficients assures that the distance between the true feasible region and the estimated feasible region converges to zero with probability one as sample size tends to infinity. The stochastic improvement consists of the feasibility step that obtains a point feasible to the updated feasible region by projecting an infeasible point to the feasible region and the optimality step that obtains an improved point by applying the affine scaling method. The proposed algorithm alternates between these two steps.

The algorithm SIM utilizes only the mean of estimator of unknown coefficients. Therefore, it is a subject of further research to develop a

model and a solution algorithm that makes use of higher moments of estimators, e.g., variance. This will enable us to estimate the confidence region of the constraints that contain unknown coefficients with an appropriate significance level, which is also viewed as the chance constraint $\Pr(\hat{A}x + \hat{b} \leq 0) \geq \alpha$ with a significance level α . It is necessary and worthwhile to investigate the stochastic improvement method for the estimated problem by the confidence region, which is a generalization of our proposed method.

Chapter 8.

CONCLUSION

In this dissertation, we have discussed statistical approaches for stochastic programming. Various algorithms have been proposed for several types of stochastic programming problems. Here we summarize the obtained results and discuss some directions of future research.

In Chapter 2 through Chapter 4, we have proposed algorithms based on the confidence region method for stochastic linear programming problems. The unknown coefficients that describe the problem are estimated from available information and then their confidence regions are derived under a given significance level. The confidence region method provides a minimax solution that optimizes the objective function by assuming the worst case behavior of parameters in their confidence regions. In other words, a minimax solution by the confidence region method minimizes the maximal possible damage in decision making.

Chapter 2 has discussed the confidence region method for two types of stochastic linear programming problems, which are dual each other

but have different stochasticity. The first problem, which has a normally distributed random right-hand side with unknown distribution parameters, has been formulated as a minimax model with a quadratic recourse. An optimal solution of the problem with true parameters is obtained from the asymptotic optimal solution of the quadratic recourse model with sufficiently large sample size.

Chapter 3 has discussed the stochastic linear programming problem with estimated constraints. When the unknown coefficients in the linear constraints are estimated by means of multiple regression analysis, the confidence regions of linear constraints form the intersection of reverse convex regions. We have proposed solution algorithms that use a cutting plane method for reverse convex programming. We have shown the *finite* convergence of the cutting plane algorithm for the linear programming problem with *one* unknown linear constraint and the convergence of the modified cutting plane algorithm to an optimal solution for the linear programming problem with *several* unknown linear constraints.

Chapter 4 has discussed the confidence region method for the P-model of the stochastic linear knapsack problem that contains unknown distribution parameters. We have proposed a polynomial time algorithm by reducing it to the problem with known distribution parameters.

In the practical decision making under uncertainty, it is necessary and useful to reduce uncertainty by making use of statistical information. In this sense, the confidence region method is an interesting approach to stochastic programming. It is important for us to develop

confidence region methods for other types of stochastic programming problems.

Chapter 5 and Chapter 6 have dealt with a probability maximizing model of the stochastic linear knapsack problem and its application to a portfolio selection problem. In Chapter 5, we have discussed a probability maximizing model of the stochastic linear knapsack problem with normally distributed random cost coefficients. The optimality condition was derived for an optimal solution of the introduced auxiliary problem. The relationship between an optimal solution and an efficient frontier of the feasible region was clarified in the mean variance framework, as a result of the uniqueness of optimal parameter value introduced in the auxiliary problem.

Chapter 6 has developed efficient algorithms for the stochastic linear knapsack problem, which is found in the portfolio selection problems. The block diagonal model has been solved by decomposing it into subproblems and merging them by ranking method. The problem of index models, which are based on the CAPM or the APT, has been solved in polynomial time by introducing the parameters corresponding to the indices.

Since the stock market deals with more than 1000 investments, it necessitates the development of more efficient algorithms to get an optimal portfolio for the problem of multi index model. As it is in general difficult to predict with certainty the rate of return on investment, estimating the rate of return on investment by the confidence region under a given significance level is interesting. Moreover, since there are different models of portfolio selection problems, e.g. E-V model and

P-model, it is worthwhile to clarify their relation to the probability maximizing model.

Chapter 7 has discussed a stochastic improvement method for stochastic linear programming, which improves alternately the estimates of unknown parameters and the current solution of the estimated problem. We have proposed a solution algorithm, which alternates between the feasibility step and the optimality step. The point sequence generated by the proposed algorithm converges to a point that gives the optimal value with probability one. We consider that a future research direction of the stochastic improvement method is the development of a model and its solution algorithm that make use of higher moments of estimator, thereby enabling us to deal with the confidence region of unknown constraints. The stochastic improvement method may be developed in order to solve the stochastic model without transforming into the equivalent deterministic model. It is also interesting to apply the stochastic improvement method to the problem with ever-changing data like time series data.

Many of the real life problems have to face with uncertainty, for which prediction or estimation has to be done from enormous statistical data. It is therefore important to investigate statistical approaches for the decision problems under uncertainty. The author hopes that the works contained in this dissertation will contribute to the development of statistical approaches for stochastic programming.

References

- [B1] M. M. Babbar, *Distribution of Solutions of a Set of Linear Equations — With an Application to Linear Programming*, Journal of American Statistical Association **50** (1955), 854–869.
- [B2] E. R. Barnes, *A Variation on Karmarkar's Algorithm for Solving Linear Programming Problems*, Mathematical Programming **36** (1986), 174–182.
- [B3] E. M. L. Beale, *On Minimizing a Convex Function Subject to Linear Inequalities*, Journal of Royal Statistical Society **B17** (1955), 173–184.
- [B4] A. Ben-Tal, *The Entropic Penalty Approach to Stochastic Programming*, Mathematics of Operations Research **10** (1985), 263–279.
- [B5] A. Ben-Tal and E. Hochman, *Stochastic Programs with Incomplete information*, Operations Research **24** (1976), 336–347.
- [B6] B. Bereanu, *On Stochastic Linear Programming Distribution Problems, Stochastic Technology Matrix*, Z. Wahrscheinlichkeitstheorie verw. Geb. **8** (1967) 148–152.
- [B7] J. R. Birge, *Decomposition and Partitioning Methods for Multi-stage Stochastic Linear Programs*, Operations Research **33** (1985), 989–1007.

- [B8] J. Bracken and R. M. Soland, *Statistical Decision Analysis of Stochastic Linear Programming Problems*, Naval Research Logistics Quarterly **13** (1966), 205–226.
- [C1] A. Charnes and W. W. Cooper, *Chance-Constrained Programming*, Management Science **6** (1959), 73–79.
- [C2] A. Charnes and W. W. Cooper, *Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints*, Operations Research **11** (1963), 18–39.
- [C3] T. Cipra, *Prediction in Stochastic Linear Programming*, Kybernetika **23** (1987), 214–226.
- [D1] G. B. Dantzig, *Linear Programming under Uncertainty*, Management Science **1** (1955), 197–206.
- [D2] W. Dinkelbach, *On Nonlinear Fractional Programming*, Management Science **13** (1967), 492–498.
- [D3] J. Dupačová, *Stability in Stochastic Programming with Recourse – Estimated Parameters*, Mathematical Programming **28** (1984), 72–83.
- [D4] J. Dupačová, *The Minimax Approach to Stochastic Programming and an Illustrative Application*, Stochastics **20** (1987), 73–88.
- [D5] J. Dupačová and R. J-B Wets, *Asymptotic Behavior of Statistical Estimators and of Optimal Solutions for Stochastic Optimization Problems*, Annals of Statistics **16** (1988), 1517–1549.
- [E1] E. J. Elton and M. J. Gruber, “Modern Portfolio Theory and Investment Analysis,” Wiley, New York, 1981.
- [E2] E. J. Elton, M. J. Gruber and M. W. Padberg, *Simple Criteria for Optimal Portfolio Selection*, The Journal of Finance **31** (1976), 1341–1357.

- [E3] E. J. Elton, M. J. Gruber and M. W. Padberg, *Simple Criteria for Optimal Portfolio Selection with Upper Bounds*, Operations Research **25** (1977), 952–967.
- [E4] E. J. Elton, M. J. Gruber and M. W. Padberg, *Simple Criteria for Optimal Portfolio Selection: Tracing Out the Efficient Frontier*, The Journal of Finance **33** (1978), 296–302.
- [E5] Yu. Ermoliev and R. J-B Wets, “Numerical Techniques for Stochastic Optimization,” Springer-Verlag, Berlin, 1988.
- [E6] R. Everett and T. Ziemba, *Two Period Stochastic Programs with Simple Recourse*, Operations Research **27** (1979), 485–502.
- [F1] R. J. Freund, *The Introduction of Risk into a Programming Model*, Econometrica **24** (1956), 253–263.
- [H1] R. J. Hillestad and S. E. Jacobsen, *Reverse Convex Programming*, Applied Mathematics and Optimization **6** (1980), 63–78.
- [H2] R. J. Hillestad and S. E. Jacobsen, *Linear Programs with an Additional Reverse Convex Constraint*, Applied Mathematics and Optimization **6** (1980), 257–269.
- [H3] R. Horst and H. Tuy, “Global Optimization,” Springer - Verlag, Berlin, 1990.
- [I1] H. Ishii, *Stochastic Optimization*, in “M. Iri and H. Konno, ed. Application of Mathematical Programming: Theory,” Sangyo Tosho, Tokyo, 1982, pp. 1–40 (in Japanese).
- [I2] H. Ishii and T. Nishida, *Stochastic Linear Knapsack Problem*, Technology Reports of the Osaka University **32** (1982), 25–30.
- [I3] H. Ishii and T. Nishida, *Stochastic Linear Knapsack Problem : Probability Maximization Model*, Mathematica Japonica **29** (1984), 273–281.

- [I4] H. Ishii and T. Nishida, *The Stochastic Linear Continuous Type Knapsack Problem: A generalized P Model*, European Journal of Operational Research **19** (1985), 118–124.
- [I5] H. Ishii, T. Nishida and N. Nanbu, *A Generalized Chance Constraint Programming Problem*, Journal of Operations Research Society of Japan **21** (1978), 124–145.
- [I6] H. Ishii, S. Shiode, T. Nishida and K. Iguchi, *An Algorithm for a Partially Chance-Constrained E-Model*, Journal of Operations Research Society of Japan **22** (1979), 233–255.
- [I7] K. Isii, *Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming*, Annals of Institute Statistical Mathematics **16** (1964), 277–293.
- [J1] R. Jagannathan, *Use of Sample Information in Stochastic Recourse and Chance-Constrained Programming Models*, Management Science **31** (1985), 96–108.
- [K1] P. Kall, “Stochastic Linear Programming,” Springer - Verlag, Berlin, 1976.
- [K2] N. Karmarkar, *A New Polynomial-Time Algorithm for Linear Programming*, Combinatorica **4** (1984), 373–395.
- [K3] S. Kataoka, *A Stochastic Programming Model*, Econometrica **31** (1963), 181–196.
- [M1] A. Madansky, *Inequalities for Stochastic Linear Programming Problems*, Management Science **6** (1960), 197–204.
- [M2] K. V. Mardia, J. T. Kent and J. M. Bibby, “Multivariate Analysis,” Academic Press, London, 1979.

- [M3] H. M. Markowitz, “Portfolio Selection,” A Cowles Foundation Monograph, 1986.
- [M4] R. Meyer, *The Validity of a Family of Optimization Method*, SIAM Journal on Control **8** (1970), 41–54.
- [M5] H. Morita and H. Ishii, *A Stochastic Improvement Method for Stochastic Programming* (to appear in Computational Statistics and Data Analysis).
- [M6] H. Morita and H. Ishii, *An Efficient Algorithm for a Stochastic Linear Knapsack Problem with a Single index model* (to appear in Mathematica Japonica).
- [M7] H. Morita, H. Ishii and T. Nishida, *Confidence Region Method for Stochastic Programming Problem*, Journal of Operations Research Society of Japan **30** (1987), 218–230.
- [M8] H. Morita, H. Ishii and T. Nishida, *Confidence Region Method for a Stochastic Linear Knapsack Programming Problem*, Mathematica Japonica **33** (1988), 559–564.
- [M9] H. Morita, H. Ishii and T. Nishida, *Stochastic Programming with Estimated Objective*, Technology Reports of the Osaka University **39** (1989), 1–7.
- [M10] H. Morita, H. Ishii and T. Nishida, *Stochastic Linear Knapsack Programming Problem and its Application to a Portfolio Selection Problem*, European Journal of Operational Research **40** (1989), 329–336.
- [M11] H. Morita, H. Ishii and T. Nishida, *Stochastic Linear Programming Problem with Partially Estimated Constraint*, Mathematica Japonica **35** (1990), 551–559.

- [N1] T. Nanseki, "Stochastic Programming Models for Agricultural Planning under Uncertainty: Theory, Methods and Applications," Taimeido Publishing Co.,Ltd., Tokyo, 1991 (In Japanese).
- [N2] K. Nishina and Y. Tabata, *The distribution of stock returns*, Discussion paper 56, Faculty of Economics, Osaka University, 1986.
- [P1] A. Prékopa, *Logarithmic Measures with Application to Stochastic Programming*, Acta Scientiarum Mathematicarum **32** (1971), 301–316.
- [P2] A. Prékopa and T. Szántai, *Flood Control Reservoir System Design Using Stochastic Programming*, Mathematical Programming Study **9** (1978), 138–151.
- [Q1] L. Qi, *The A-forest Iteration Method for the Stochastic Generalized Transportation Problem*, Mathematics of Operations Research **12** (1987), 1–21.
- [R1] H. Raiffa and R. Schlaifer, "Applied Statistical Decision Theory," Harvard Business School, Boston, 1961.
- [R2] C. R. Rao, "Linear Statistical Inference and its Applications," Wiley, New York, 1971.
- [R3] R. T. Rockafellar, "Convex Analysis," Princeton University Press, 1970.
- [R4] J. B. Rosen, *Iterative Solution of Nonlinear Optimal Control Problems*, SIAM Journal on Control **4** (1966), 223–244.
- [S1] J. K. Sengupta, "Stochastic Programming: Methods and Applications," North-Holland, Amsterdam, 1972.
- [S2] J. K. Sengupta, "Decision Models in Stochastic Programming: Operational Methods of Decision Making under Uncertainty," North - Holland, New York, 1982.

- [S3] W. F. Sharpe, "Portfolio Theory and Capital Markets," McGraw-Hill, 1970, pp. 117–122.
- [S4] G. P. Szegö, "Portfolio Theory," Academic Press, New York, 1980.
- [T1] K. Takeuchi, "Approximation of Probability Distributions," Kyokushuppan, Tokyo, 1973 (in Japanese).
- [T2] G. Tintner, *The Theory of Choice under Subjective Risk and Uncertainty*, Econometrica **9** (1941), 298–304.
- [T3] H. Tuy, *Concave Programming Under Linear Constraints*, Soviet Mathematics **5** (1964), 1437–1440.
- [T4] H. Tuy, *Convex Programs with an Additional Reverse Convex Constraint*, Journal of Optimization Theory and Applications **52** (1987), 463–485.
- [T5] H. Tuy, *Global Minimization of a Difference of Two Convex Functions*, Mathematical Programming Study **30** (1987), 150–182.
- [U1] U. Ueimg, *A Combinatorial Method to Compute a Global Solution of Certain Non-Convex Optimization Problems*, in "F. A. Lootsma, ed., Numerical Methods for Non-Linear Optimization," Academic Press, New York, 1972, pp. 223–224.
- [V1] S. Vajda, "Probabilistic Programming," Academic Press, New York, 1972.
- [V2] R. M. Van Slyke and R. J-B Wets, *L-shaped Linear Programs with Applications to Optimal Control and Stochastic Programming*, SIAM Journal of Applied Mathematics **17** (1969), 638–663.
- [W1] H. M. Wagner, *On the Distribution of Solutions in Linear Programming Problems*, Journal of American Statistical Association **53** (1958), 161–163.

- [W2] D. W. Walkup and R. J-B Wets, *Stochastic Programs with Recourse*, SIAM Journal of Applied Mathematics **15** (1967), 1299–1314.
- [W3] R. J-B Wets, *Programming Under Uncertainty: The Equivalent Convex Program*, SIAM Journal of Applied Mathematics **14** (1966), 89–105.
- [W4] R. J-B Wets, *Solving Stochastic Programs with Simple Recourse*, Stochastics **10** (1983), 219–242.
- [Z1] W. Ziemba, *Stochastic Programs with Simple Recourse*, in “P. Hammer and G. Zoutendijk ed., Mathematical Programming in Theory and Practice,” North-Holland, Amsterdam, 1974, pp. 213–274.
- [Z2] P. H. Zipkin, *Simple Ranking Methods for Allocation of One Resource*, Management Science **26** (1980), 34–43.