On Stochastic Programming

STOCHASTIC PROGRAMMING AND ITS APPLICATION TO PRODUCTION HORIZON PROBLEM

By SHINJI KATAOKA
Assistant of Physics

In this paper, we consider statistical properties of an optimal solution of mathematical programming when constants are random variables. The reason why we call it stochastic programming is that our objects are not only to obtain an optimum expectation value of an objective function, but also its distribution function, by which a statistical reliability criterion of the optimal policy can be computed numerically. Furthermore, it will be possible to compare the optimal policy with others, for instance, an optimal solution of the deterministic problem obtained by inserting mean values into the random variables.

As a method of prediction of demand, a linear regression model is used by which estimates of amounts of future demand and their distribution functions are assumed. Since the confidence limits of demand become large with the time interval of prediction, the necessity of such a reliability test for the optimal solution would increase in order to make the plan reliable.

Introduction

Recently, many methods of mathematical programming have been developed and expanded to the stochastic cases, for instance, production horizon problems under random demand, or inventory planning under the same situation. Two comprehensive books, Studies in 'The Mathematical Theory of Inventory and Production and Dynamic Programming, mention mathematical treatments of these problems exhaustively.1,2 Certainly, the most available principle to lead to an optimal policy would be to optimize the expectation value of the objective function (cost, profit, etc.), because programming or scheduling would more or less depend on the prediction or conjecture which necessarily contains ambiguity. However, it seems that there are few theories discussing reliability of the obtained optimal policy as is usually seen in mathematical statistics.3

1 A part of this work was done at the Operations Research Center, Massachusetts Institute of Technology, being supported by the Office of Ordnance Research under Contract No. DA 19-020-ORD-2684.
2 G. Tintner proposed a method of this kind of a problem in his paper, "Stochastic Linear Programming with Applications to Agricultural Economics", Proceedings of the Symposium in Linear Programming, 1955. Following his notations, let G(z) be a linear function of a vector z, and A and c be a matrix and a vector respectively, containing random parameters. Then his problem was to obtain a distribution function of G(z) such that

\[
\text{max } G(z)
\]

under the conditions:

\[Az = c, \quad z \geq 0\]

Taking advantage of the property that an optimal solution was a basic feasible solution, he obtained the maximum value of G(z) at each point of the parameter space and constructed a distribution function of max G(z) with it.

His distribution function is interpreted as follows: what distribution would be expected, if we could be previously informed of realized values of the random variables and make an optimal scheduling with that information? This is clearly different from the usual situation. The expectation value, not the distribution function, of G(z) was discussed by Dantzig and others exactly.4,5
In order to define a criterion of reliability of a policy, we must know the distribution function of the value of the objective function which mainly means cost in this paper. Once the distribution is determined, it will be natural to define the reliability criterion as follows: for a given $\alpha$ ($100 > \alpha > 0$) to obtain a limit $\eta$ such that

$$P_r(\text{cost of the policy} \leq \eta) = \alpha/100$$

and we call $\eta$ "the $\alpha$ percent confidence limit of cost" of the policy. This means that when that policy is used, let $\alpha$ be 95, we could expect with 95% confidence that the cost would be under $\eta$, in other words, the chance that cost exceeds $\eta$ would be 5% in probability.

Next, suppose there are two policies, $A$ and $B$, of which expectation values of cost are 100 and 110, and let their 95% confidence limits be 120 and 115 respectively. Then how do we choose the optimal one from these two policies? From the minimum expectation point of view, $A$ is better than $B$. But, the 95% limit of $A$ is 120, while that of $B$ is 115. Hence, the policy $A$ has more undesirable chances for the cost to exceed over 120 than $B$, that is:

$$P_1(\text{cost of } A \geq 120) > P_1(\text{cost of } B \geq 120).$$

Then the decision making executives would select one of these two policies according to their own judgment. Some executive may choose $A$ owing to the merit of the minimum cost, and the other may select $B$ because of its good stability.

Before proceeding to the production horizon problem, some general properties of stochastic programming are discussed in Section (1). We define two policies:

(i) minimum expectation policy: $\hat{x}$

(ii) deterministic minimum policy: $\bar{x}$

The minimum expectation policy $\hat{x}$ is determined by solving

$$\min_{\hat{x}} E f(x, s)$$

where $s$ is a vector of random variables and $f(x, s)$ is a cost function. The deterministic minimum policy $\bar{x}$ is obtained by solving

$$\min_{\bar{x}} f(x, \bar{s})$$

where $\bar{s}$ is the vector of the expectation values of the random variables. Then we consider relationship between $\hat{x}$ and $\bar{x}$, and derive inequalities which are useful for estimating $\hat{x}$.

In Section (2), a production horizon problem is considered. Although it is important problem how to choose the distribution function of demand at each time period, most authors seemed to assume that it does not change over all periods. Essentially, however, programming more or less depends on prediction of which reliability usually decreases with the time interval of the prediction. In this sense, the asymptotic property of an optimal policy would not be able to be expressed by a simple functional equation.

As the first step, assuming that (i) data of demand in $m$ time periods are given, and the linear regression analysis can be applied to the data, and (ii) the sample size $m$ is large enough for a Student distribution to be well approximated by a normal distribution, we use the normal distribution as the distribution of demand. As a consequence, later computational procedures can be remarkably simplified. In the latter part of Section (2), we derive a general computational procedure for the production horizon problem on the base of the above mentioned distribution function. In Section (3), the approximate distribution function of the optimal value of the inventory problem and the reliability criterion are given.
I. General Stochastic Programming

(A) Definitions and Assumptions

$s$: vector of random variables (for instance, demand, price, and other production coefficients)

$\bar{s} = E_s$

$S$: set of all possible values of $s$

$x$: vector of controllable variables (amounts of production)

$X$: set of feasible $x$'s

$f(x, s)$: objective function (cost, profit, time, etc.)

Assumption (1): Let $f(x, s)$ be a cost function and convex one of $s$. Reliability criterion:

$Pr(f(x, s) \leq \eta) = \alpha$

$\eta$: a percent confidence limit of cost

Assumption (2): The set of feasible $x$'s is not dependent on $S$. In other words, the conditions which $x$'s should fulfill do not contain random variable $s$. Moreover, we define several kinds of policies.

$\min_{x \in X} \mathbb{E} f(x, s) = \mathbb{E} f(x_*, s)$: minimum value of expectation

$x_*$: minimum expectation policy (M.E.P.)

$\mathbb{E} \min_{x \in X} f(x, s)$: expectation of minimum value

$\min_{x \in X} f(x, s) = f(x_*, s)$: deterministic minimum value

$x_*$: deterministic minimum policy (D.M.P.)

(B) Inequalities in Stochastic Programming

Recently A. Madansky proved the following inequalities about those policies.\(^6\)

\[
\min_{x \in X} f(x, s) \leq f(x_*, s) \leq \mathbb{E} f(x, s) \leq \mathbb{E} \min_{x \in X} f(x, s).
\] (1.1)

For $f(x, s)$ is a convex function of $s$,

\[
\min_{x} f(x, s) \leq \mathbb{E} \min_{s} f(x, s).
\] (1.2)

On the other hand, as $x$ is a M.E.P.,

\[
\min_{x} \mathbb{E} f(x, s) = \mathbb{E} f(x_*, s) \leq \mathbb{E} f(x, s).
\] (1.3)

Then, from (1.1), (1.2), and (1.3), we obtain the following inequalities.

\[
f(x, s) \leq \min_{x} \mathbb{E} f(x, s) \leq \mathbb{E} f(x_*, s).
\] (1.4)

Thus we can say:

(i) (1.1) shows that minimum value of expectation is not smaller than expectation of minimum value.

(ii) If $f(x, s)$ is a convex function of $s$ (this condition usually holds, shown in the later section), from (1.4), the minimum value of expectation exists between the deterministic minimum value and the expectation of the cost of deterministic minimizing policy. These inequalities give us a simple method for estimating and evaluating $x$, that is,
then,
\[
\frac{E_f(\bar{x}, s) - f(\bar{x}, \bar{s})}{f(\bar{x}, \bar{s})} = \frac{E_f(\bar{x}, s) - f(\bar{x}, \bar{s})}{f(\bar{x}, \bar{s})}.
\]
(1.6)
The left-hand side of (1.6) is the relative error of \(f(\bar{x}, \bar{s})\), which is limited by the right-hand side of (1.6). If the right-hand side is small comparing with a unity, the D.M.P., \(\bar{x}\), would be a fairly good approximation for \(\bar{x}\).

II. Production Horizon for Random Demand

As an application of our theory, we consider a production horizon problem in the case when demand in each period is a random variable. Because many authors' papers and books on this subject have been published, our concerned model is not so fresh to discuss over again. However, it would be somewhat interesting to consider it from our point of view. For the sake of simplicity, we take the model of simple commodity production over \(n\) periods, no time lag between production and selling.

(A) Demand Prediction

As we mentioned in the introduction, the most important problem in the production horizon problem would be how to predict amounts of demand in future. According to the property and quantity of the concerned commodity, there would be a lot of methods for forecasting. Here we adopt the most popular one, linear regression, in order to make our discussion simple.

(a) Definitions and Assumptions

Assumption (3): We are given the amounts of demand, \(s^{(k)}\), at time \(t_k\) \((k=1, ..., m)\).

Assumption (4): The amount of demand at time \(t, s_t\), is a random variable which belongs to the normal distribution (mean value \(\beta_1 + \beta_2 t\), variance \(\sigma^2\)) or, in other words,
\[
s_t = \beta_1 + \beta_2 t + \varepsilon
\]
where \(\varepsilon\) is a normal random variable which has \(E\varepsilon = 0, E\varepsilon^2 = 1\), and \(\beta_1, \beta_2\) and \(\sigma^2\) are unknown parameters.

Definitions:
\[
\bar{s} = \frac{1}{m} \sum_{k=1}^{m} s^{(k)}: \text{ average demand (sample mean)}
\]

\[
\bar{t} = \frac{1}{m} \sum_{k=1}^{m} t_k: \text{ average time}
\]

\(\hat{\beta}_1\): least square estimate of \(\beta_1\)

\(\hat{\beta}_2\): least square estimate of \(\beta_2\)

\(\hat{\sigma}^2\): least square estimate of \(\sigma^2\)

(b) Prediction of Demand by Linear Regression

By the method of least square we have
\[
\hat{\beta}_2 = \frac{\sum (s^{(k)} - \bar{s})(t_k - \bar{t})}{\sum (t_k - \bar{t})^2}
\]
(2.2)
\[
\hat{\beta}_1 = \bar{s} - \hat{\beta}_2 \bar{t}
\]
(2.3)
\[ \sigma^2 = \frac{1}{m} \sum_{k=1}^{m} (s_{tk} - \beta_1 - \beta_2 t_k)^2. \] (2.4)

Then the least square estimate of \( s_t \) at \( t \), \( \hat{s}_t = \beta_1 + \beta_2 t \)

**Definition:**

\[ Z_t = s_t - \hat{s}_t \]

then

\[ \text{Var} Z_t = \sigma^2 \left[ \frac{m+1}{m} + \frac{(t-t_0)^2}{\sum (t_k-t)^2} \right]. \]

**Definition:**

\[ T = \frac{Z_t / \sqrt{\text{Var} Z_t}}{Z_t / \sqrt{m \sigma^2 (m-2) \sigma^2}} = \frac{Z_t \left[ \frac{m+1}{m} + \frac{(t-t_0)^2}{\sum (t_k-t)^2} \right]^{1/2}}{\sqrt{m-2}}, \] (2.5)

then the distribution of the random variable \( T \) will be a Student's \( t \)-distribution of which degree of freedom is \( m-2 \).

(c) **Distribution Function of \( s \)**

Since it is known that for a sufficiently large \( m \), the distribution function of a random variable

\[ u = T \left( \frac{m-4}{m-2} \right)^{1/2} \]

is well approximated by the normal distribution, \( E u = 0, E u^2 = 1 \).

Then the density distribution function of \( s \) will be approximated by

\[ \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left\{ -\frac{(s_t-m_t)^2}{2\sigma^2} \right\}, \] (2.6)

where

\[ m_t = \beta_1 + \beta_2 t, \] (2.7)

and

\[ \sigma^2 = \left\{ \sum (s_{tk} - \beta_1 - \beta_2 t_k)^2 \right\} \left[ \frac{m+1}{m} + \frac{(t-t_0)^2}{\sum (t_k-t)^2} \right] \left( m-4 \right). \]

Definition (2.7) tells us an interesting fact that the variance of the expected amount of demand is a monotone increasing function of a prediction time interval. Hereafter we use this distribution function as that of demand at a time period. Putting \( t = t_m + \tau_0 \), we define

\[ m_t = \beta_1 + \beta_2 (t_m + \tau_0 - t), \]

and

\[ \sigma^2 = \left\{ \sum (s_{tk} - \beta_1 - \beta_2 t_k)^2 \right\} \left[ \frac{m+1}{m} + \frac{(t_m + \tau_0 - t)^2}{\sum (t_k-t)^2} \right], \]

where \( \tau_0 \) is the time interval of one period, and \( i = 1, 2, 3, \ldots \).

(B) **Computational Procedures of Minimum Expectation Policy.**

(a) **Definitions and Assumptions**

**Definitions:**

- \( n \): total number of periods
- \( \varphi(s_t; m_t, \sigma^2) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(s_t-m_t)^2}{2\sigma^2}} \)
- \( s_t \): random demand for the commodity during \( i \)-th period
- \( m_t \): \( E s_t \)
- \( \sigma^2_t = E (s_t-m_t)^2 \)
- \( \varphi(s_t; m_t, \sigma^2_t) \): density distribution function of the amount of demand during \( i \)-th period. (= \( \varphi(s_t) \))
- \( S_t = \sum_{i=1}^{t} s_i \)

---

Assumption (5): The amounts of demand, \( s_1, s_2, \ldots, s_n \), are independent to each other. Then we have

\[
s_t^2 = E(S_t - M_t)^2 = E(\sum_{j=1}^{t} (s_j - m_j))^2 = \sum_{j=1}^{t} s_j^2.
\]

Definitions:

- \( x_i \): amount of the commodity produced during \( i \)-th period (controllable variable)
- \( c_i \): capacity of production
- \( y_0 \): initial inventory (constant)
- \( C_i = \sum_{j=1}^{i} c_i + y_0 \)
- \( y_i \): inventory at the end of \( i \)-th period
- \( X_i = \sum_{j=1}^{i} x_j + y_0 \)

(b) Cost Function \( f(x, s) \)

The cost function \( f(x, s) \) consists of two parts: the production cost and the inventory-penalty cost.

- \( p_t(x_i) \): production cost of \( x_i \)
  - linear cost \( p_t x_i \)
  - convex cost \( p_t x_i + q_t x_i^2 \) \( p_t \geq 0, q_t > 0 \)
  - concave cost \( p_t \log(1 + x_i) \)

The inventory-penalty cost contains two parts:
- inventory cost of \( i \)-th period for \( y_i \geq 0 \)
- penalty cost of \( i \)-th period for \( y_i < 0 \)

Suppose:

\[
G(y) = \begin{cases} ay & \text{for } y \geq 0 \\ -by & \text{for } y < 0 \end{cases}, \quad a, b > 0,
\]

then the cost of \( i \)-th period \( f_i \)

\[
f_i = p_t(x_i) + G(y_i)
\]

and we have the following relations

\[
\begin{align*}
\begin{array}{c}
y_1 = X_1 - S_1 \\
y_2 = X_2 - S_2 \\
\vdots \\
y_n = X_n - S_n, \\
x_1 = X_1 - y_0 \geq 0 \\
x_2 = X_2 - X_1 \geq 0 \\
\vdots \\
x_n = X_n - X_{n-1} \geq 0.
\end{array}
\end{align*}
\]

Inserting (2.10) and (2.11) into (2.9),

\[
f_i = f(X_i, X_{i-1}, S_i) = p_t(X_i - X_{i-1}) + G(X_i - S_i)
\]

Total cost:

\[
f(x, s) = \sum_{i=1}^{n} f_i(X_{i-1}, S_i) \quad (X_0 = y_0)
\]
From Appendix I

expectation of \( G_t \):

\[
Q_t(X_t) = \int_{-\infty}^{\infty} G_t(X_t - S_t) \varphi(S_t; \mu_t, \sigma_t^2) dS_t
\]

(2.13)

expectation of total cost:

\[
E(f, s) = \sum_{i=1}^{n} \{ \rho_i(X_i - X_{i-1}) + Q_i(X_i) \}.
\]

(2.14)

Further, define

\[
\begin{align*}
 f_1(y, X_1) &= \rho_1(X_1 - y) + Q_1(X_1) \\
 f_2(X_1, X_2) &= \rho_2(X_2 - X_1) + Q_2(X_2) \\
 f_n(X_{n-1}, X_n) &= \rho_n(X_n - X_{n-1}) + Q_n(X_n),
\end{align*}
\]

(2.15)

then, the problem will be

\[
\min \{ f_1(y, X_1) + f_2(X_1, X_2) + \ldots + f_n(X_{n-1}, X_n) \}
\]

under (2.8) and (2.11).

This is a typical dynamic programming which can be solved by successive minimization.

(c) Computational Procedures

Define:

\[
\begin{align*}
 F_n(X_{n-1}) &= \min f_n(X_{n-1}, X_n) \\
 & \quad \text{s.t. } X_{n-1} \leq X_n \leq C_n \\
 & \quad \text{and } 0 \leq X_{n-1} \leq C_{n-1} \\
 F_{n-1}(X_{n-2}) &= \min \{ f_n(X_{n-2}, X_{n-1}) + F_{n-1}(X_{n-1}) \} \\
 & \quad \text{s.t. } X_{n-2} \leq X_{n-1} \leq C_{n-1} \\
 & \quad \text{and } 0 \leq X_{n-2} \leq C_{n-2}
\end{align*}
\]

(2.16)

\[
\begin{align*}
 F_1(X_1) &= \min \{ f_2(X_1, X_2) + F_2(X_2) \} \\
 & \quad \text{s.t. } X_1 \leq X_2 \leq C_2 \\
 & \quad \text{and } 0 \leq X_1 \leq C_1
\end{align*}
\]

\[
\min Ef(x, s) = \min \{ F_1(X_1) + f_1(y, X_1) \}.
\]

(2.17)

Simultaneously, we can get the optimal policy at each step as a function of the predeces-sive variable,

\[
\begin{align*}
 \hat{X}_n &= X_n(X_{n-1}) \\
 \hat{X}_{n-1} &= X_{n-1}(X_{n-2}) \\
 \vdots \\
 \hat{X}_2 &= X_2(X_1) \\
 \hat{X}_1 &= \text{constant}
\end{align*}
\]

and from the last step of (2.16), optimal \( \hat{X}_1 \) can be obtained,

\[
y_0 + \hat{x}_1 = \hat{X}_1 = \text{const}.
\]

Then, by solving (2.17) backwards, \( X_1, X_2, \ldots, X_n \), the optimal policy \( \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n \) will be obtained,

\[
\begin{align*}
 \hat{x}_1 &= \hat{X}_1 - y_0 \\
 \hat{x}_2 &= \hat{X}_1 - \hat{X}_1 \\
 \vdots \\
 \hat{x}_n &= \hat{X}_n - \hat{X}_{n-1}
\end{align*}
\]

(2.18)

The deterministic solution is also obtained by inserting \( G_t(X_t - M_t) \) into \( Q_t(X_t) \) in (2.15), and we have the deterministic minimum policy as follows:
The most simple example of this production problem is the case of linear cost
\[ p(x_t) = x_t \]

Thus our problem will be simplified as follows:
\[ E f(x, s) = -y_n + Q_1(X_1) + Q_2(X_2) + \ldots + Q_n(X_n) + X_n \]
under
\[ 0 \leq X_i \leq \ldots \leq X_n \leq C_n \]
and
\[ 0 \leq X_i \leq C_i \quad i=1, 2, \ldots, n. \]

Because \( Q_i(x_i) \) and \( Q_n(X_n) + X_n \) are strictly convex, we can solve this minimizing problem by Modigliani-Holt and Klein's method\(^6,7\) but this successive minimization method has two merits: (i) it consists of simple iterations suitable to computer, and (ii) it is available not only to the convex case but also to the general nonlinear programming.

Thus, because \( Q_i(X_i) \) and \( Q_n(X_n) + X_n \) are strictly convex, they have only one minimum point, and accordingly the computational procedures become simpler than that of the general case mentioned above.

\[ F_{n-1}(X_{n-1}) = \min_{X_{n-1} \leq X_n \leq C_n} \{Q_n(X_n) + X_n\} \]

Let \( \hat{X}_n \) be a value such that
\[ \min_{X_n \leq C_n} \{Q_n(X_n) + X_n\} = Q(\hat{X}_n) + \hat{X}_n \]

I \hspace{1cm} if \( \hat{X}_n \leq C_{n-1} \)

<table>
<thead>
<tr>
<th>( F_{n-1}(X_{n-1}) = )</th>
<th>( Q_n(\hat{X}<em>{n-1}) + \hat{X}</em>{n-1} )</th>
<th>( Q_n(X_{n-1}) + X_{n-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{X}<em>n(X</em>{n-1}) = )</td>
<td>( \hat{X}_n )</td>
<td>( X_{n-1} )</td>
</tr>
</tbody>
</table>

II \hspace{1cm} if \( C_{n-1} \leq \hat{X}_n \leq C_n \)

<table>
<thead>
<tr>
<th>( F_{n-1}(X_{n-1}) = )</th>
<th>( Q_n(\hat{X}_n) + \hat{X}_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{X}<em>n(X</em>{n-1}) = )</td>
<td>( \hat{X}_n )</td>
</tr>
</tbody>
</table>

Both cases I and II can be joined together,

<table>
<thead>
<tr>
<th>( F_{n-1}(X_{n-1}) = )</th>
<th>( Q_n(\hat{X}_n) + \hat{X}_n )</th>
<th>( Q_n(X_{n-1}) + X_{n-1} ) ( \ldots (n, F) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{X}<em>n(X</em>{n-1}) = )</td>
<td>( \hat{X}_n )</td>
<td>( X_{n-1} ) ( \ldots (n, X) )</td>
</tr>
</tbody>
</table>

\(^4\) The deterministic minimization policy in this case is obtained by inserting \( M_i \) into \( \hat{S}_i \) in (2.19) and (2.20). However, we can transform this problem including non-linear \( G_i \) into a linear programming, which has a simple algorithm due to the simplicity of the matrix, just like "Method of Modified Distribution" in the transportation problem. \textit{Appendix II}.
Similarly,

\[ F_{n-1}(X_{n-2}) = \min_{0 \leq X_{n-2} \leq C_{n-2}} \{ Q_{n-1}(X_{n-1}) + F_{n-1}(X_{n-1}) \} \]

Solution: Find \( \tilde{X}_{n-1} \) such that

\[
\min_{0 \leq X_{n-1} \leq C_{n-1}} \{ Q_{n-1}(X_{n-1}) + F_{n-1}(X_{n-1}) \} = Q_{n-1}(\tilde{X}_{n-1}) + F_{n-1}(\tilde{X}_{n-1})
\]

Then,

\[
\begin{array}{c|c|c}
F_{n-2}(X_{n-3}) = & X_{n-2} \leq \tilde{X}_{n-1} & \tilde{X}_{n-1} < X_{n-2} \\
\hline
Q_{n-1}(\tilde{X}_{n-1}) + F_{n-1}(\tilde{X}_{n-1}) & Q_{n-1}(X_{n-2}) + F_{n-1}(X_{n-2}) & \ldots \ldots (n-1, F) \\
\hline
\tilde{X}_{n-1}(X_{n-2}) = & \tilde{X}_{n-1} & X_{n-2} \ldots \ldots (n-1, X)
\end{array}
\]

Solution: Find \( \tilde{X}_2 \) such that

\[
\min_{0 \leq X_2 \leq C_2} \{ Q_2(X_2) + F_2(X_2) \} = Q_2(\tilde{X}_2) + F_2(\tilde{X}_2)
\]

Then

\[
\begin{array}{c|c|c}
F_1(X_1) = & X_1 \leq \tilde{X}_2 & \tilde{X}_2 < X_1 \\
\hline
Q_2(\tilde{X}_2) + F_2(\tilde{X}_2) & Q_2(X_1) + F(X_1) & \ldots \ldots (2, F) \\
\hline
\tilde{X}_2(X_1) = & \tilde{X}_2 & X_1 \ldots \ldots (2, X)
\end{array}
\]

Thus, at the last step,

\[
\min Ef(x,s) = \min_{0 \leq X_1 \leq C_1} Q_1(X_1) + F_1(X_1)
\]

(2.21)

Then we can get \( \tilde{X}_2, \tilde{X}_3, \ldots, \tilde{X}_n \) by solving (2, X), (n, X) backwards,

\[
\begin{align*}
\tilde{X}_1 &= \tilde{X}_1 \\
\tilde{X}_2 &= \max \{ \tilde{X}_1, \tilde{X}_2 \} \\
\tilde{X}_3 &= \max \{ \tilde{X}_2, \tilde{X}_3 \} \\
& \quad \ldots \ldots \\
\tilde{X}_n &= \max \{ \tilde{X}_{n-1}, \tilde{X}_n \}
\end{align*}
\]

(2.22)

and the minimizing expectation policy \( \hat{x} \) is as follows:

\[
\begin{align*}
\hat{x}_n &= \tilde{X}_n - \tilde{X}_{n-1} \\
\hat{x}_{n-1} &= \tilde{X}_{n-1} - \tilde{X}_{n-2} \\
& \quad \ldots \ldots \\
\hat{x}_1 &= \tilde{X}_1 - \tilde{X}_2 \\
\hat{x}_0 &= \tilde{X}_2
\end{align*}
\]

(2.23)
III. Distribution Function of \( f(x, s) \)

(A) Variance and Third Moment of \( f(x, s) \)

Next stage of our problem is to compute the higher order moments of the total cost \( f(x, s) \) of a policy \( x \); variance and third order moment. Because the production cost does not include random variable \( s \), then \( f(x, s) - Ef(x, s) \) implies only the inventory-penalty cost part.

Define,

\[
M_{i,j} = \sum_{k=1}^{i} m_k, \sigma_{i,j}^2 = \sum_{k=1}^{i} \sigma_k^2, S_{i,j} = \sum_{k=1}^{i} s_k
\]

\[
Q_{i,j}^{(1,1)} = \int_{-\infty}^{\infty} G(X_i - S_i) \varphi(S_i; M_i, \sigma_i^2) \sum_{k=1}^{i} G(X_j - S_i - S_j) \varphi(S_j; M_j, \sigma_j^2) dS_i dS_j
\]

\[
Q_{i,j}^{(1,2)} = \int_{-\infty}^{\infty} G(X_i - S_i) \varphi(S_i; M_i, \sigma_i^2) \sum_{k=1}^{i} G(X_j - S_i - S_j) \varphi(S_j; M_j, \sigma_j^2) dS_i dS_j
\]

\[
Q_{i,j,k}^{(1,1,1)} = \int_{-\infty}^{\infty} G(X_i - S_i) \varphi(S_i; M_i, \sigma_i^2) \sum_{k=1}^{i} G(X_j - S_i - S_j) \varphi(S_j; M_j, \sigma_j^2) dS_i dS_j
\]

\[
H(S_i) = \int_{-\infty}^{\infty} G(X_i - S_i) \varphi(S_i; M_i, \sigma_i^2) dS_i
\]

See Appendix 1, B.

Further define,

\[
G(X_i - S_i) = G_i
\]

mean value:

\[
Q^{(1)} = \sum_i E G_i = \sum_i Q_i^{(1)}
\]

\[
Q^{(2)} = \sum_i E G_i^2 = \sum_i Q_i^{(2)}
\]

\[
Q^{(3)} = \sum_i E G_i^3 = \sum_i Q_i^{(3)}
\]

\[
Q^{(1,1)} = \sum_{i,j} E G_i G_j = \sum_{i,j} Q_{i,j}^{(1,1)}
\]

\[
Q^{(1,2)} = \sum_{i,j} E G_i G_j^2 = \sum_{i,j} Q_{i,j}^{(1,2)}
\]

\[
Q^{(1,1,1)} = \sum_{i,j,k} E G_i G_j G_k = \sum_{i,j,k} Q_{i,j,k}^{(1,1,1)}
\]

and

\[
T^{(1)} = E(\sum_i G_i)^2 = Q^{(2)} + 2Q^{(1,1)}
\]

\[
T^{(2)} = E(\sum_i G_i^3)^2 = Q^{(3)} + 3Q^{(1,2)} + 6Q^{(1,1,1)}
\]

then we have the variance and third moment of cost function as follows:

\[
\sigma_f^2 = Ef(f(x, s)) - Ef(x, s))^2
\]

\[
= E(\Sigma G_i)^2 - (\Sigma Q_i)^2
\]

\[
= T^{(2)} - Q^{(2)}
\]

and

\[
\xi_f^3 = Ef(f(x, s)) - Ef(x, s)^3
\]

\[
= E(\Sigma G_i) - (\Sigma Q_i)^3
\]

\[
= E(\Sigma G_i)^3 - 3E(\Sigma G_i)^2 Q^{(3)} - 3E(\Sigma G_i) Q^{(1,2)} - Q^{(1,1,1)}
\]

\[
= T^{(3)} - 3T^{(2)} Q^{(1,1)} + 2Q^{(1,2)}
\]

(B) Approximate Distribution Function and Reliability Criterion

Appendix III provides us the least square approximation of the distribution function by Hermite polynomial when we are given \( \sigma_f^2 \) and \( \xi_f^3 \). The results are as follows:

put \( x \) (not the vector used above)
let the distribution function of \( x \) be \( \phi(x) \), then

\[
\phi(x) = \phi_0(x) + \frac{\xi^3}{3\sqrt{\pi}} e^{-\frac{x^2}{2}} (1-x^2)
\]  

(3.8)

where

\[
\phi_0(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dy.
\]

And the reliability criterion, \( \alpha \) percent limit \( \eta \), will be obtained as follows:

Let \( x \) be a solution of

\[
\phi(x) = \alpha, \quad x = x(\alpha),
\]  

(3.9)

then we have

\[
\eta = \sum_{i=1}^{n} \phi_i(x_i) + Q_i(x_1) + \sigma_f x(\alpha).
\]  

(3.10)

### Appendix I

(A) \[
\int_{-\infty}^{\infty} \sum_{i=1}^{n} H(S_i) \phi_i(S_i) \, dS_i = \int_{-\infty}^{\infty} H(S_i) dS_i \phi_i(S_i) \tag{A 1, 1}
\]

where

\[
\phi_i(S_i) = \int_{-\infty}^{S_i} \phi(S_i-S_i) \, dS_i
\]

Put

\[
\Phi_i(t) = \int_{-\infty}^{\infty} e^{ist} \phi_i(S_i) \, dS_i,
\]

\[
\varphi_i(t) = \int_{-\infty}^{\infty} e^{ist} \phi_i(s_i) \, ds_i,
\]

then,

\[
\Phi_i(t) = \int_{0}^{\infty} \sum_{i=1}^{n} e^{ist} \phi_i(S_i) \, dS_i = \sum_{i=1}^{n} \varphi_i(t)
\]

then if

\[
\varphi_i(s_i) = \varphi(s_i; m_i, \sigma_i^2)
\]

\[
\varphi_i(t) = e^{mt - \frac{\sigma_i^2}{2} t^2}
\]

\[
\therefore \quad \Phi_i(t) \exp \left( \sum_{k=1}^{i} m_k t - \frac{\sigma_k^2}{2} t^2 \right) = e^{mt - \frac{\sigma_i^2}{2} t^2}
\]

\[
\therefore \quad \Phi_i(S_i) = \varphi(S_i; M_t, \sigma_i^2)
\]

then

\[
Q_i(X_i) = \int_{-\infty}^{\infty} G(X_i - S_i) \phi_i(S_i) \, dS_i
\]

(3.11)
\[ Q_{ij}(X_i, X_j) = \lim_{t \to \infty} \int_{s=0}^{\infty} \varphi(s) \cdots \varphi(s) G(x_i-s_i) \cdots \varphi(s_j) G(x_j-s_j) ds_{t+1} \cdots ds_i, \]

by Appendix I, A

\[ = \lim_{t \to \infty} \int_{s=0}^{\infty} \varphi(s) \cdots \varphi(s) G(x_i-s_i) ds_{i+1} \cdots G(x_j-s_j) \varphi(s_j) ds_j, \]

again from Appendix I, A

\[ = \int_{s=0}^{\infty} G(x_i-s_i) \varphi(s_i) ds_i, \]

\[ = \int_{s=0}^{\infty} G(x_j-s_j) \varphi(s_j) ds_j. \]

**Appendix II**

Define

\[ y_i = u_i - v_i, \quad u_i, v_i \geq 0 \]

Total cost

\[ f = p \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} G(y_i) = \sum (px_i + au_i + bv_i) \cdots \min \quad (A 2, 1) \]

under

\[ x_i + y_{i-1} - y_i = s_i, \]

or

\[ x_i + (u_{i-1} - v_{i-1}) - (u_i - v_i) = s_i \]

and

\[ x_i \geq 0 \]

\[ u_i, v_i \geq 0 \]

\[ u_i, v_i = 0. \]

**Simplex Tableau**

\[
\begin{array}{ccccccccc}
X_1 & X_2 & \ldots & X_n & u_1 & v_1 & u_2 & v_2 & \ldots & u_n & v_n \\
p & p & \ldots & p & a & b & a & b & \ldots & a & b \\
1 & -1 & 1 & & & & & & & \bar{s}_1 - y_0 \\
1 & -1 & -1 & 1 & 1 & -1 & & & & \bar{s}_2 \\
1 & 1 & -1 & -1 & 1 & & & & & \bar{s}_n \\
\end{array}
\]

Though the condition (A2, 3) seems to be nonlinear, owing to the linear dependency of process vectors \( U_i, V_i \), it turns out \( U_i \) and \( V_i \) never enter the optimal solution simultaneously, then (A2, 3) are satisfied automatically.

**Appendix III**

First we consider the following function \( g(z) \),

\[ g(z) = e^{-az} \sum_{r=1}^{n} b_r H_r(az) \quad (A 3, 1) \]
ON STOCHASTIC PROGRAMMING

where \( H_r(az) = \frac{(-1)^r}{a^r} e^{a x} \frac{d^r}{dx^r} e^{-a x} \) : Hermite polynomial

If \( b_r \) is chosen such that,
\[
b_r = \frac{\alpha}{\sqrt{2\pi} r! \sqrt{\pi}} \int_{-\infty}^{\infty} f(z) H_r(az) dz,
\]
then \( g(Z) \) will be the least square approximation of \( f(z) \), weight \( e^{a x} \),
\[
\int_{-\infty}^{\infty} e^{a x} (f(z) - g(z))^2 dz \ldots \text{min}.
\]

If \( i \)-th moment of \( Z \) is given,
\[
\mu_i = \int_{-\infty}^{\infty} z^i f(x) dz
\]
then,
\[
b_0 = \alpha \sqrt{\pi} \mu_0, \quad b_1 = \alpha \sqrt{\pi} \mu_1, \quad b_2 = \frac{\alpha}{4 \sqrt{\pi}} (2 \alpha^2 \mu_2 - \mu_0)
\]
\[
b_3 = \frac{\alpha}{12 \sqrt{\pi}} (2 \mu_3 - 3 \mu_1)
\]
\[
b_4 = \frac{\alpha}{96 \sqrt{\pi}} (4 \mu_4 - 12 \mu_2 + 3).
\]

In our case, we put
\[
\alpha^2 = \frac{1}{2 \sigma_f^2}, \quad \alpha = \frac{1}{\sqrt{2} \sigma_f}
\]
\[
\mu_0 = 1, \quad \mu_1 = 0, \quad \mu_2 = \sigma_f^2, \quad \mu_3 = \xi_f^3
\]
\[
Z = f(x, z) - Ef(x, z),
\]
then,
\[
b_0 = \frac{1}{\sqrt{2\pi} \sigma_f}, \quad b_1 = 0, \quad b_2 = \frac{1}{4 \sqrt{2\pi} \sigma_f} \left( \frac{2 \sigma_f^2}{\sigma_f^2} - 1 \right) = 0
\]
\[
b_3 = -\frac{\xi_f^3}{6 \sqrt{2\pi} \sigma_f}.
\]

We have from (A 3, 1),
\[
g(z) dz = e^{-\frac{z^2}{2 \sigma_f^2}} \left( b_0 + b_4 H_4 \left( \frac{z}{\sqrt{2 \sigma_f}} \right) \right)
\]

Inserting (A 3, 6),
\[
g(z) dz = \frac{1}{\sqrt{2\pi} \sigma_f} e^{-\frac{z^2}{2 \sigma_f^2}} \left[ 1 + \frac{\xi_f^3}{6} H_4 \left( \frac{z}{\sqrt{2 \sigma_f}} \right) \right] dz.
\]

Put
\[
\frac{z}{\sigma_f} = x, \quad f(x): \text{density distribution function}
\]
\[
f(x) dx = \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{\xi_f^3}{3 \sqrt{\pi}} e^{-\frac{3x^2}{2}} (x^3 - 3x) \right\} dx,
\]
then distribution function of \( x, \psi(x) \)
\[
\psi(x) = \int_{-\infty}^{x} f(x) dx
\]
\[
= \psi(x) + \frac{\xi_f^3}{3 \sqrt{\pi}} \psi_3(x)
\]

---

\[ \psi_2(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \]

\[ \psi_3(x) = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \left( x^3 - 3x \right) \, dx \]

\[ = e^{-\frac{x^2}{2}} (1 - x^3) \]

**References**


---

**A PRELIMINARY STUDY OF A STOCHASTIC PROGRAMMING MODEL**

In this paper we propose a model of stochastic programming, which takes into account distribution of the objective function as well as expectation value. Also probabilistic constraints are introduced instead of linear conditions. Applying it to a transportation problem we have a non-linear objective function constrained by linear inequalities. In order to get a global maximum (or minimum), several simple assumptions are made.

**I. Introduction**

When mathematical programming is applied to a system to obtain an optimal policy, one of the difficulties is that coefficients in the formulation are not constant but fluctuating or uncertain values. Some may result from fluctuation in outside of the system, and others may come from the system itself. For instance demand and price in an open market are considered as the former, and productivity and consumption rate of a plant or machine the latter. Also uncertainty frequently comes from smallness of sample size, especially in business statistics.

Recently this problem has been studied and developed to stochastic (or chance con-
strained) programming by many authors,\(^1\text{,}2\text{,}3\text{,}4\text{,}5\text{,}6\) who introduced probabilistic consideration into making constraints modulated by the uncertainty.

The purpose of this paper is the following:

1. To introduce a new concept of maximization process of the objective function, considering not only its expectation value but also distribution function of it.

2. To apply this model to a transportation type problem and formulate it as an example of our stochastic programming, taking the probabilistic constraints introduced by Charnes and Cooper.\(^2\text{,}4\) (1

3. To extend it to more general problems constrained by probability.

Let us consider the following problem. Suppose \(X(x_1, x_2, ..., x_n)\) be a vector. (we call it "policy" later)

\[
\max_x \sum_j p_j x_j \quad j=1, 2, ..., n 
\]

under the conditions

\[
\sum_j a_{ij} x_j \leq b_i \quad i=1, 2, ..., m 
\]

\(x_j \geq 0\). (1.3)

If the quantities \(p_j, a_{ij}, b_i\) are all definite constants, this is of course a typical linear programming problem. Let us take an example in a production process in which these quantities have the following meanings respectively: profit of \(j\)-th activity \(p_j\), and its \(i\)-th commodity production rate \(a_{ij}\), amount of demand of \(i\)-th commodity \(b_i\). As mentioned before these are not constant but some randomly fluctuating quantities. How should we make formulation in order to realize this situation? Difficulty occurs immediately in the explanation of (2), which is not always satisfied by a fixed policy \(X\), while we have to compute the value of the objective function for the policy.

In order to avoid this difficulty, two types of formulation were introduced.\(^1\text{,}3\)

1. Additional penalty cost is subtracted from the total profit for each violation of the constraints.

\[
\max_x E\{ \sum_j p_j x_j - \sum_i f_i(\sum_j a_{ij} x_j - b_i) \}, \quad \text{(1.4)}
\]

where \(f_i(x)\) is some function for the penalty cost, for instance it may have the following expression.

\[
f_i(b_i - \sum_j a_{ij} x_j) = \begin{cases} 
\beta_i(b_i - \sum_j a_{ij} x_j) & \text{for } a_{ij} x_j > b_i \\
0 & \text{for } a_{ij} x_j \leq b_i.
\end{cases}
\]

If \(f_i(x)\) is convex, \(E\{ \sum_j p_j x_j - \sum_i f_i(\sum_j a_{ij} x_j - b_i) \}\) is proved to be convex,\(^4\) subsequently, a global maximum coincides with a local one.

2. The second method introduces the following probabilistic inequalities instead of (2).

\[
\text{Prob} \{ \sum_j a_{ij} x_j \leq b_i \} \geq \alpha_i, \quad \text{(1.6)}
\]

where \(\alpha_i\) is a prescribed probability to the \(i\)-th constraint of (2). For some value of \(\alpha_i\), say 0.95, a set of \(X\) is defined such that the possibility of the \(i\)-th constraint holds is greater than 95\%. In other words we can say that less than only 5\% probability of overstock is allowed to the production manager in the factory.

In this paper the second method is preferred, because the concept of the penalty cost is not well-defined in general. For example stock shortage in inventory would not be able to
be evaluated by the simple linear penalty cost like (1.5).

Now let us proceed to discuss the main theme of this paper. It seems that few of discussions have been concerned with the objective function of stochastic programming so far. We have to point out that expectation value of profit may not be always a good \textit{optimality criterion} for a policy. Let us show an example. Suppose we have two policies $X_1$ and $X_2$ for which expectation values of profit have the following relationship,

$$EPX_1 > EPX_2 \quad (1.7)$$

where $P$ is a profit vector. Can we simply say that $X_1$ is better than $X_2$? Even though (1.7) is satisfied, the probability that actual profit of $X_1$ takes a value less than some prescribed limit may be greater than that of $X_2$. In other words $X_1$ may be more risky than $X_2$ in this case. When one activity which makes high average profit has also very large variance of it, we would not take this activity simply because of its high profit, if we are conservative managers.

For this reason we propose here the following objective function for our stochastic programming.

$$\max L_p \quad (1.8)$$

under

$$\text{Prob} \left( \sum_j p_j x_j \leq L_p \right) = \alpha \quad (1.9)$$

and

$$\text{Prob} \left( \sum_j a_{ij} x_j \leq b_i \right) \geq \alpha, \quad x_j \geq 0, \quad i=1, 2, \ldots, m \quad j=1, 2, \ldots, n \quad (1.10)$$

The additional constraint (9) is explained as follows: we try to maximize a lower limit $L_p$ such that probability that profit drops in the interval below the limit is restricted by $\alpha$. This condition will keep us from selecting unstable policy mentioned before. In the following section we formulate a transportation type problem in the stochastic programming.

II. \textit{Transportation Problem}

We consider a transportation problem in which supplies and demands at origins and destinations are random variables and transportation cost also, while structure matrix ($a_{ij}$ in (1.2)) is constant. In this case the objective function is an upper limit $H_c$ beyond which total cost is permitted to exceed in probability $\alpha$.

Problem:

$$\min H_c \quad (2.1)$$

under

$$\text{Prob} \left( \sum_j c_j x_j \leq H_c \right) = \alpha \quad (2.2)$$

and

$$\text{Prob} \left( \sum_j a_{ij} x_j \leq b_i \right) \geq \alpha, \quad x_j \geq 0, \quad i=1, 2, \ldots, m \quad j=1, 2, \ldots, n \quad (2.3) \quad x_j \geq 0, \quad (2.4)$$
where \(c_j\)'s and \(b_i\)'s are random variables, while \(a_0\)'s, \(\alpha_i\) are constants. Now in order to proceed to the step of analytical formulation, let us make some assumptions on these \(c_i\), \(b_i\), \(\alpha\) and \(\alpha_i\).

Assumption (1):

\(c_j\) and \(b_i\) have normal distributions respectively with mean value \(\bar{c}_j\), variance \(\sigma_j^2\) and \(\bar{b}_i\), \(\sigma_i^2\).

Assumption (2):

\(c_j\)'s are independent to each other.

Assumption (3):

\[ \alpha \leq 0.5 \]

This assumption is plausible because we usually desire the risk to be less than 50%.

Under these conditions we can derive an interesting convex non-linear programming problem which is equivalent to the original stochastic one.

Let us consider (2.2) first. Transforming (2.2),

\[ \text{Prob}(\sum_j c_j x_j \geq H_0) = \text{Prob} \left( \frac{\sum_j c_j x_j - \sum_j \bar{c}_j x_j}{\sqrt{\sum_j \sigma_j^2 x_j^2}} \geq \frac{H_0 - \sum_j \bar{c}_j x_j}{\sqrt{\sum_j \sigma_j^2 x_j^2}} \right), \]  

(2.5)

where \(\sqrt{\sum_j \sigma_j^2 x_j^2}\) is the variance of \(\sum_j c_j x_j\) obtained from the independency of \(c_j\)'s. The probability (2.5) is evaluated as a normal error function, because the left hand side of the inequality of the argument is a normalized random variable.

Then we have

\[ \text{Prob}(\sum_j c_j x_j \geq H_0) = G \left( \frac{H_0 - \sum_j \bar{c}_j x_j}{\sqrt{\sum_j \sigma_j^2 x_j^2}} \right), \]  

(2.6)

where

\[ G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy. \]  

(2.7)

Because \(G(x)\) is a monotone decreasing function of \(x\) and \(0 < G(x) < 1\) for \(-\infty < x < +\infty\), we have one and only one real root \(k\) such that

\[ G(k) = \alpha, \]  

(2.8)

or

\[ k = G^{-1}(\alpha) \geq 0 \]  

(2.9)

Again using the monotone decreasing property of \(G(x)\), we get that

\[ G \left( \frac{H_0 - \sum_j \bar{c}_j x_j}{\sqrt{\sum_j \sigma_j^2 x_j^2}} \right) = \alpha \]

is equivalent to the following:

\[ H_0 - \sum_j \bar{c}_j x_j = k \geq 0. \]  

(2.10)

Then the objective function is

\[ H_f = \sum_j \bar{c}_j x_j + k \sqrt{\sum_j \sigma_j^2 x_j^2}. \]  

(2.11)

Because \(\sqrt{\sum_j \sigma_j^2 x_j^2}\) is a convex function of \(X\) and \(k \geq 0\), the left hand side is a convex function. (see Appendix)

Similar discussion on the conditions (2.3) simplifies probability as follows. In this case
The left hand side of the argument is normalized, then
\[ \text{Prob}(\sum_j a_{ij}x_j \geq b_i) = G\left(\frac{b_i - \sum_j a_{ij}x_j}{\sigma_{b_i}}\right). \tag{2.13} \]

Let \( l_i \) be a solution of the equation,
\[ G(l_i) = \alpha_i. \tag{2.14} \]

Then the inequality (2.3)
\[ b_i - \sum_j a_{ij}x_j \geq \alpha_i \tag{2.15} \]
is equivalent to the following:
\[ \frac{b_i - \sum_j a_{ij}x_j}{\sigma_{b_i}} \leq l_i, \tag{2.16} \]
or
\[ a_{ij}x_j \geq b_i - l_i\sigma_{b_i}. \]

It is enough for us to put \( b_i - l_i\sigma_{b_i} \) in the place of random variable \( b_i \). Now we have an analytical expression of the stochastic programming for a transportation problem, that is,
\[ \min \sum_j c_jx_j + k\sqrt{\sum_j \sigma_j^2 x_j^2} \tag{2.17} \]
under
\[ a_{ij}x_j \geq b_i - l_i\sigma_{b_i}, \tag{2.18} \]
\[ x_j \geq 0. \tag{2.19} \]

As mentioned before (2.17) is a convex function of \( X \) because of the convexity of \( \sqrt{\sum_j \sigma_j^2 x_j^2} \) and non-negativity of \( k \) (or \( \alpha \leq 0.5 \)), then the piecewise approximation of the objective function enables us to attain a global maximum under constraints (2.18) and (2.19) by linear programming.

III. General Stochastic Programming

Now let us get back to the general stochastic programming (1.8), (1.9), (1.10) in which the structure matrix \((a_{ij})\) is not constant any more. For the same reason mentioned in the preceding section, we have to put a restriction on \( \alpha \)'s in order to obtain convexity of the feasible domain of (1.10).

Assumption (4):
\[ \alpha \leq 0.5 \]
\[ \alpha_i \leq 0.5 \]

By the similar method in Sec. II, we have,
\[ \text{Prob}(\sum_j p_jx_j \leq L_p) = \int \frac{L_p - \sum_j p_jx_j}{\sqrt{\sum_j \sigma_j^2 x_j^2}} = \alpha, \tag{3.1} \]
where

\[ I(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2} dy. \]

Then we will find the objective function, the lower limit of profit,

\[ L_p = \sum_j \tilde{p}_j x_j - k \sqrt{\sum_j \sigma_j^2 x_j^2} \]  \hspace{1cm} (3.2)

where \( k \) is a root of the equation

\[ I^{-1}(\alpha) = -k \] \hspace{1cm} (3.3)

Also we have,

\[ \text{Prob}\left( \sum a_{ij} x_j \leq b_i \right) \geq \int \left( \frac{\tilde{b}_i - \sum a_{ij} x_j}{\sqrt{\sigma_{ij}^2 + \sum_j \sigma_j^2 x_j^2}} \right) \] \hspace{1cm} (3.4)

then from (1.10), we have,

\[ I\left( \frac{\tilde{b}_i - \sum a_{ij} x_j}{\sqrt{\sigma_{ij}^2 + \sum_j \sigma_j^2 x_j^2}} \right) \geq \alpha_i \] \hspace{1cm} (3.5)

or an equivalent inequality,

\[ (\tilde{b}_i - \sum a_{ij} x_j) - l_i \sqrt{\sigma_{ij}^2 + \sum_j \sigma_j^2 x_j^2} \geq 0 \] \hspace{1cm} (3.6)

where \( l_i \) is a root of the following equation,

\[ I(l_i) = \alpha_i. \]

The same discussion can be applied to the concavity of the objective function \( L_p \) and the left hand side of the constraint (3.6), because \( k \) and \( l_i \) are non-negative.

**Assumption (5):**

Let us assume that a convex domain,

\[ \tilde{b}_i - \sum a_{ij} x_j \geq 0 \]

\[ x_j \geq 0 \] \hspace{1cm} (3.7)

is closed.

This assumption is also consistent with the nature of this problem. Because (3.7) is just a feasible domain for the linear programming in which all random variables are replaced by their average values, it is reasonable to require that this original linear programming problem does not have an unusual feasible domain. Under this assumption, it is clear that any \( X \) which satisfies (3.6) and \( x_j \geq 0 \) is included in the closed domain (3.7), because of the non-negativity of \( l_i \).

Thus we have an explicit formulation of the general stochastic programming,

\[ \max \left( \sum_j \tilde{p}_j x_j - k \sqrt{\sum_j \sigma_j^2 x_j^2} \right) \] \hspace{1cm} (3.8)

under

\[ (\tilde{b}_i - \sum a_{ij} x_j) - l_i \sqrt{\sigma_{ij}^2 + \sum_j \sigma_j^2 x_j^2} \geq 0 \]

\[ x_j \geq 0. \] \hspace{1cm} (3.9)

Suppose \( F(X, U) \) be a Lagrangian function,

\[ F(X, U) = \sum_j \tilde{p}_j x_j - k \sqrt{\sum_j \sigma_j^2 x_j^2} \]

\[ + \sum_j u_j \left( \tilde{b}_i - \sum a_{ij} x_j - c_i \sqrt{\sigma_{ij}^2 + \sum_j \sigma_j^2 x_j^2} \right) \] \hspace{1cm} (3.10)

where \( U(u_1, u_2, \ldots, u_m) \) is a non-negative vector. Then we can write down Kuhn-Tucker
condition for the optimality of a solution. Let \( X^o \) and \( U^o \) be a set of solutions which satisfies the following conditions.

\[
\begin{align*}
\left( \frac{\partial F}{\partial x_j} \right)_{x_j = x^o_j} & = 0 \quad \text{for} \quad x^o_j > 0 \\
\left( \frac{\partial F}{\partial x_j} \right)_{x_j = x^o_j} & \leq 0 \quad \text{for} \quad x^o_j = 0 \\
\left( \frac{\partial F}{\partial u_i} \right)_{u_i = u^o_i} & = 0 \quad \text{for} \quad u^o_i > 0 \\
\left( \frac{\partial F}{\partial u_i} \right)_{u_i = u^o_i} & \geq 0 \quad \text{for} \quad u^o_i = 0,
\end{align*}
\]

(3.11) (3.12)

where

\[
\begin{align*}
\frac{\partial F}{\partial x_j} & = \tilde{b}_j - \frac{b_j x_j}{\sqrt{\sum \sigma_i^2 x_i^2}} - \sum_i \alpha_i c_i \sigma_i^2 x_i - \sum_i \sigma_i^2 x_i^2 \\
\frac{\partial F}{\partial u_i} & = (\tilde{b}_i - \sum \tilde{a}_{ij} x_j) - c_i \sqrt{\sum \sigma_i^2 + \sum \sigma_i^2 x_i^2}.
\end{align*}
\]

(3.13) (3.14)

Then \( X \) is assured to be the global optimal solution of (38), (39). The most promising method to solve this tremendous non-linear programming is supposed to be the gradient method at this moment.\(^{16\text{th}}\)

Let us consider the following \( n+m \) simultaneous differential equations,

\[
\begin{align*}
\frac{\partial x_j}{\partial t} & = \begin{cases} 
0 & \text{if} \quad \frac{\partial F}{\partial x_j} < 0 \quad \text{and} \quad x_j = 0 \\
\frac{\partial F}{\partial x_j} & \text{otherwise},
\end{cases} \\
\frac{\partial u_i}{\partial t} & = \begin{cases} 
0 & \text{if} \quad \frac{\partial F}{\partial u_i} > 0 \quad \text{and} \quad u_i = 0 \\
-\frac{\partial F}{\partial u_i} & \text{otherwise}.
\end{cases}
\end{align*}
\]

(3.15) (3.16)

Fortunately Arrow, Hurwicz and Uzawa proved that these differential equations provide a saddle point \( X, U \) in the stationary state under the concavity conditions, which are satisfied by the objective function and left hand side of constraints (ref. 10. Theorem 3, p.137). Numerical computation for an example is now in progress by this author.

IV. Discussions

In the previous sections we made five assumptions for the convenience of computational procedures, especially for the convexity or concavity of the problem. Assumption (3), (4), and (5) are plausible and consistent with the nature of each optimizing problem: Total cost larger than some prescribed value is not preferable and total profit less than some desired level either. Assumption (1) is also deemed to be reasonable, because most of random variables and sums of them are classified in this category. However, the second assumption (independency) has to be examined more carefully. There exists, especially in price mechanism, a prominent correlation, negative or positive. Subsequently we can not use the Assumption (2) in general. In the following paper, this assumption will be removed and a general case will be treated, introducing a dispersion matrix.

Another problem to be solved is to find an effective computational procedure which avoids an oscillational solution and has rapid convergency to the saddle point.
Appendix

(a) First we prove the convexity of \( \sqrt{\sum_j \sigma_j^2 x_j^2} \)

Suppose \( X, Y \) and \( Z \) be vectors such that

\[
Z = \lambda X + (1-\lambda) Y
\]

or

\[
z_j = \lambda x_j + (1-\lambda)y_j
\]

\( 0 \leq \lambda \leq 1 \)

\[
W = \text{sign} \left\{ \sqrt{\sum_j \sigma_j^2 x_j^2} - (\lambda \sqrt{\sum_j \sigma_j^2 x_j^2} + (1-\lambda) \sqrt{\sum_j \sigma_j^2 y_j^2}) \right\}
\]

Multiplying a positive value, to the argument of the sign,

\[
\sqrt{\sum_j \sigma_j^2 x_j^2} + (\lambda \sqrt{\sum_j \sigma_j^2 x_j^2} + (1-\lambda) \sqrt{\sum_j \sigma_j^2 y_j^2})
\]

we have

\[
W = \text{sign} \left\{ 2\lambda(1-\lambda)\{\Sigma \sigma_j^2 x_jy_j - \sqrt{\sum_j \sigma_j^2 x_j^2 \sqrt{\sum_j \sigma_j^2 y_j^2}} \} \right\}
\]

by Schwarz's inequality,

\[
\sum_j \sigma_j^2 x_j y_j \leq \sqrt{\sum_j \sigma_j^2 x_j^2 \sqrt{\sum_j \sigma_j^2 y_j^2}}
\]

then we have,

\[
W \leq 0.
\]

(b) Next we prove convexity of the domain \( g(X) \geq 0 \), if \( g(X) \) is concave.

Let \( X \) and \( Y \) be \( g(X) \geq 0, \) and \( g(Y) \geq 0 \) and \( Z = \lambda X + (1-\lambda) Y, \) then

\[
g(\lambda X + (1-\lambda) Y) \geq \lambda g(X) + (1-\lambda) g(Y) \geq 0.
\]

Q. E. D.

References

A STOCHASTIC PROGRAMMING MODEL

In this paper we propose a stochastic programming model which considers the distribution of an objective function and probabilistic constraints. Applying it to a transportation type problem, we derive a non-linear programming problem constrained by linear inequalities, and show that it can be solved by iteration of linear programming.

I. Introduction

When mathematical programming is applied to a problem, one of the difficulties is that the coefficients in the formulation are not constants but fluctuating or uncertain values. Recently this problem has been studied and developed to stochastic programming by many authors, by introducing probabilistic constraints.

The purpose of this paper is the following:

(1) To introduce a new concept of maximization process by an objective function, considering not only its expectation value but also the distribution of it.

(2) To apply this model to a transportation type problem, defined later, and obtain a computational procedure for it.

Let us consider the following problem. Suppose \(x(x_1, x_2, \ldots, x_n)\) and \(p(p_1, p_2, \ldots, p_n)\) be \(n\)-vectors, \(b(b_1, b_2, \ldots, b_m)\) an \(m\)-vector and \(A(a_{ij})\) be an \(m \times n\)-matrix. Furthermore, let the components of these vectors and matrix be random variables. Then we define the basic formulation of our stochastic programming as follows:

**Problem 1:**

\[
\text{Max } f, \quad (1.1)
\]

subject to

\[
\text{Prob } (p'x \leq f) = \alpha, \quad (1.2)
\]

and

\[
\text{Prob } (\sum_j a_{ij}x_j \leq b_i) \geq \beta_i, \quad (1.3)
\]

\[
x_i \geq 0, \quad i=1, 2, \ldots, m, \quad j=1, 2, \ldots, n, \quad (1.4)
\]

where \(p'\) is the transposed vector of \(p\), and \(\alpha\) and \(\beta_i\) are prescribed probabilities and have...
the following meaning.

Under the uncertain situation, inequalities in linear programming,

\[ \sum_j a_{ij}x_j \leq b_i, \quad i = 1, 2, \ldots, m, \quad (1.5) \]

are not always satisfied by a fixed \( x \). Instead of using (1.5), we restrict the domain of \( x \) in such a way that probability of the \( i \)-th condition of (1.5) being held is greater than \( \beta_i \) (which we sometimes call a control level). This probabilistic condition was introduced by Charnes and Cooper. It seems, however, that few discussions have been concerned with the objective function. We have to point out that the expectation value of profit is not always considered as a good measure for optimality criterion. Even though a policy \( x \) dominates other policies in the expectation of profit, it may be more risky in such a way that the chance of getting a very low profit may be greater than the others because of the broadness of its distribution. For this reason, we maximize the lower allowable limit \( f \) defined by (1.2) for given probability \( \alpha \), instead of the expectation value of the profit.

This paper consists of the following sections besides this introduction.

II. Transportation Type problem
   (a) Assumptions and Formulation
   (b) Kuhn-Tucker Condition
   (c) Some Properties of Optimal Solution
   (d) Subsidiary Quadratic Programming
   (e) Numerical Example

III. Outline of Computational Procedures
   (a) Iteration
   (b) Interpolation

IV. Discussions

Appendix and References

II. Transportation Type Problem

In this section we consider a case in which \( b_i \)'s and \( p_j \)'s in Problem 1 are random variables, but \( a_{ij} \)'s are constants. Transportation and production horizon problems will be classified into this category if customer demand and price of commodity are random. We call these Transportation Type Problem.

(a) Assumptions and Formulation

Assumption 1: The random variable \( b_i \) has a normal distribution with mean value \( \bar{b}_i \) and variance \( \sigma_{b_i}^2 \).

Then the probability in (1.3) is transformed by simple subtraction and division as follows:

\[ \text{Prob} \left( \sum_j a_{ij}x_j \leq b_i \right) = \text{Prob} \left( \frac{\sum a_{ij}x_j - \bar{b}_i}{\sigma_{b_i}} \geq \frac{b_i - \bar{b}_i}{\sigma_{b_i}} \right) \quad (2.1) \]

From Assumption 1, the left hand side of the argument is found to be a normalized random variable with \( \mathcal{N}(0,1) \). Hence the probabilistic condition,
is replaced by
\[
G\left(\sum \frac{a_{ij}x_j - \bar{b}_i}{\sigma_{bi}}\right) \geq \beta_i,
\]
(2.2)
or
\[
\sum \frac{a_{ij}x_j - \bar{b}_i}{\sigma_{bi}} \leq G^{-1}(\beta_i),
\]
(2.3)
where
\[
G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy,
\]
(2.4)
Usually we take \( \beta_i \geq 0.5 \), then \( G^{-1}(\beta_i) \geq 0 \), and we put
\[
G^{-1}(\beta_i) = -\alpha_i,
\]
(2.5)
However, \( \beta_i \geq 0.5 \) is not a necessary condition in this case. It will be discussed in the section IV.

Let us define mean values \( \bar{p}_j \)'s and a dispersion matrix \( V \) such that
\[
\bar{p}_j = \mathbb{E}p_j, \quad V = \mathbb{V}(v_{ij}).
\]
(2.6)
(2.7)
Then the variance of \( \sum p_jx_j \) is computed as follows:
\[
\text{Var}(\sum p_jx_j) = \mathbb{E}(\sum (\bar{p}_j - \bar{p})x_j)^2
\]
\[= \sum \mathbb{E}(p_j - \bar{p})(p_j - \bar{p})x_jx_j = \sum v_{ij}x_ix_j = x'Vx.
\]
(2.8)

Assumption 2: The vector \( p \) has a multinormal distribution with a mean value vector \( \bar{p}(p_1, p_2, \ldots, p_n) \) and a positive definite dispersion matrix \( V \).

Then it is known that \( p'x \) has a normal distribution with mean value \( \bar{p}x \) and variance \( x'Vx \).

Hence,
\[
\text{Prob}(p'x \leq f) = \text{Prob}\left(\frac{p'x - \bar{p}'x}{\sqrt{x'Vx}} \leq \frac{f - \bar{p}'x}{\sqrt{x'Vx}}\right)
\]
\[= I\left(\frac{f - \bar{p}'x}{\sqrt{x'Vx}}\right), \quad \text{Prob}(1.2)
\]
where
\[
I(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.
\]
Then we have an equation for (1.2)
\[
I\left(\frac{f - \bar{p}'x}{\sqrt{x'Vx}}\right) = \alpha
\]
or
\[
f = \bar{p}'x + I^{-1}(\alpha)\sqrt{x'Vx}.
\]
(2.11)
Since \( \sqrt{x'Vx} \) is a convex function, as is proved in Appendix, if \( \alpha \leq 0.5 \) or \( I^{-1}(\alpha) \leq 0 \), (2.11) will be concave. So we assume the following:

Assumption 3:
\[
\alpha \leq 0.5
\]
Finally we have a maximization problem for the stochastic programming:

**Problem 2.1:**

\[
\text{Max } f_1 = \tilde{p}'x - q\sqrt{x'\tilde{x}},
\]

subject to

\[
Ax \leq b^*,
\]

where

\[
b^*_i = \tilde{b}_i - q_i\sigma_{b_i},
\]

and

\[
q = -I^{-1}(\alpha) \geq 0, q_i = -G^{-1}(\beta_i).
\]

Before proceeding, let us define a linear programming problem and make an assumption.

**Problem 2.0:**

\[
\text{Max } \tilde{p}'x,
\]

subject to

\[
Ax \leq b^*,
\]

\[
x \geq 0.
\]

**Assumption 4:** Problem 2.0 has a finite optimal solution \(\hat{x}_0\).

(b) Kuhn-Tucker Condition

As mentioned before, since the objective function of Problem 2.1 is concave and its constraints are linear, then there must exist only one maximum (or degenerate maxima) which satisfies the following Kuhn-Tucker Condition. \(^7\)

Let \(F(x, u)\) be a Lagrangian function and \(x(x_1, x_2, \ldots, x_n)\), and \(u(u_1, u_2, \ldots, u_n)\) be non-negative vectors.

\[
F(x, u) = \sum_j \tilde{p}_j x_j - q \sqrt{\sum_j v_{ij} x_i x_j} + \sum_i u_i (b_i^* - \sum_j a_{ij} x_j). \tag{2.14}
\]

If \((\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)\) and \((u_1, u_2, \ldots, u_n)\) satisfy the following conditions:

\[
\left. \begin{array}{l}
\sum_j v_{ij} \hat{x}_i = \tilde{p}_j - \frac{q \sqrt{\sum_j v_{ij} \hat{x}_i \hat{x}_j}}{\sum_i u_i a_{ij}} \\
\quad \leq 0 \quad \text{for } \hat{x}_j = 0, \\
\quad = 0 \quad \text{for } \hat{x}_j > 0,
\end{array} \right\} \tag{2.15}
\]

\[
\left. \begin{array}{l}
\frac{\partial F(x, u)}{\partial u_i} \hat{x} = b_i^* - \sum_j a_{ij} \hat{x}_j \\
\quad \geq 0 \quad \text{for } a_i = 0, \\
\quad = 0 \quad \text{for } a_i > 0,
\end{array} \right\} \tag{2.16}
\]

then we can say that \(\hat{x}\) is an optimal solution of Problem 2.1.

(c) Some Properties of Optimal Solution

**Theorem 1:** If Assumption 4 holds, the optimal solution of Problem 2.1 is finite.

**Proof:** Suppose there exists a feasible solution \(x\) of Problem 2.1 such that

\[
\tilde{p}'x - q\sqrt{x'\tilde{x}} > \tilde{p}'\hat{x}_0,
\]

then we have

\[
\tilde{p}'x > \tilde{p}'\hat{x}_0.
\]

Since \(\hat{x}_0\) is optimal in Problem 2.0, it contradicts the assumption. Hence for any feasible \(x\),

\[
\tilde{p}'x - q\sqrt{x'\tilde{x}} \leq \tilde{p}'\hat{x}_0. \tag{2.17}
\]

Q.E.D.
Theorem 2: If \( \hat{x} \) and \( \hat{u} \) are optimal solutions of the objective function \( f \), the optimal value of \( f \) is equal to \( u'Ax \).

Proof: Multiplying \( \hat{x} \) to each of Kuhn-Tucker Conditions in (2.15), and summing them up, we have

\[
\frac{\bar{p}'}{\sqrt{\hat{x}'V\hat{x}}} - q\hat{x}'V\hat{x} = u'Ax = 0.
\]

Again we find

\[
\frac{\bar{p}'}{\sqrt{\hat{x}'V\hat{x}}} - q\hat{x}'V\hat{x} = u'Ax.
\]  

Q.E.D.

This property is quite similar to that of linear programming, and it will give us a checking procedure in numerical computation.

(d) Subsidiary Quadratic Programming

In order to find a method of computation for the non-linear programming Problem 2.1, we consider the following quadratic programming:

Problem 2.11:

Max \( f_{11}=p'x - \frac{q}{2R} \cdot x'Vx \),  

subject to

\[
Ax \leq b^*, \quad x \geq 0,
\]

where \( R \) is a positive parameter.

When \( x'Vx \) is positive semi-definite, the objective function is concave, and consequently it has only one maximum (or degenerate maxima) for a given parameter \( R \). Now we have the following theorem:

Theorem 3: Suppose an optimal solution \( \hat{x}(R) \) of Problem 2.11 satisfies the following:

\[
R = \sqrt{\hat{\lambda}(R)'V\hat{\lambda}(R)},
\]

then \( \hat{x}(R) \) is also an optimal solution for Problem 2.1, and vice versa.

Proof: Let us write down Kuhn-Tucker conditions for Problem 2.11 to be satisfied by \( \hat{x}(R) \). Lagrangian function,

\[
F_{11}(x,u) = \sum_j \bar{p}_j \hat{x}_j - \frac{q}{2R} \sum_i v_i \hat{x}_i \hat{x}_j + \sum_i \hat{u}_i (b_i^* - \sum_j a_{ij} \hat{x}_j),
\]

\[
\left( \frac{\partial F_{11}}{\partial x_j} \right) \hat{x}_j \hat{u} = \bar{p}_j - \frac{q}{R} \sum_i v_i \hat{x}_i - \sum_i \hat{u}_i a_{ij}\ 
\begin{cases} 
\leq 0 & \text{for } \hat{x}_j = 0, \\
= 0 & \text{for } \hat{x}_j > 0,
\end{cases}
\]

\[
\left( \frac{\partial F_{11}}{\partial u_i} \right) \hat{x}_j = b_i^* - \sum_j a_{ij} \hat{x}_j
\begin{cases} 
\geq 0 & \text{for } \hat{u}_i = 0, \\
= 0 & \text{for } \hat{u}_i > 0.
\end{cases}
\]

Suppose \( \hat{x}(R) \) satisfies (2.21), the pairs of conditions (2.15), (2.16) and (2.23), (2.24), are completely equivalent. Therefore \( \hat{x}(R) \) is also the solution of Problem 2.1. Conversely, for an optimal solution \( \hat{x} \) of Problem 2.1, if we put \( R \) such that

\[
R = \sqrt{\hat{x}'V\hat{x}},
\]

\( \hat{x} \) also satisfies (2.23) and (2.24), because (2.15) and (2.16) apparently have the same structure as the latter.

Q.E.D.
According to Wolfe and Beale, since any concave quadratic programming can be transformed into linear programming, then our non-linear, originally stochastic programming, is to be computed by the same way.

Let us make a notice on some properties of the function

\[ r(R) = \sqrt{\tilde{x}'(R)\tilde{V}x(R)} \]

For infinite \( R \), Problem 2.11 is reduced to Problem 2.0 which has a finite optimal solution \( \tilde{x}_0 \), so

\[ \lim_{R \to \infty} r(R) = \sqrt{\tilde{x}_0'\tilde{V}x_0} = R_0. \]

Theorem 4: The function \( r(R) \) is monotone non-decreasing in \( R \) and has a finite limit \( R_0 \).

Proof: The latter part has been proved above. Supposing \( \tilde{x}, \tilde{y} \) be optimal solutions of Problem 2.11 for \( R_x \) and \( R_y \) respectively, we have the following inequalities:

\[ \tilde{p}'\tilde{x} - \frac{q}{2R_x} \tilde{x}'V\tilde{x} \geq \tilde{p}'\tilde{y} - \frac{q}{2R_y} \tilde{y}'V\tilde{y}, \]

\[ \tilde{p}'\tilde{x} - \frac{q}{2R_y} \tilde{y}'V\tilde{y} \geq \tilde{p}'\tilde{x} - \frac{q}{2R_y} \tilde{x}'V\tilde{x}. \]

Adding both sides and rearranging, we have

\[ (R_x - R_y)(\tilde{p}'\tilde{x} - \frac{q}{2R_y} \tilde{x}'V\tilde{x}) \geq 0. \]

(2.25)

Then if \( R_x > R_y \), it follows that \( \tilde{x}'V\tilde{x} \geq \tilde{y}'V\tilde{y} \) immediately. Since a square root function is monotone increasing, \( r(R) \) is found to be monotone non-decreasing.

Q.E.D.

Now, the conditions (2.23) and (2.24) imply that the components of the optimal solution \( x_j, s \) and \( u_i, s \) are to be proportional to \( R \). Since the inequality of each condition of (2.24) is held by \( x_j s \) for a sufficiently small \( R \), all \( u_i s \) become zeros. Hence, we have, for a small \( R \),

\[ \tilde{x}_j = \frac{R}{q} t_j, \quad j=1,2,\ldots,n, \quad (2.26) \]

where \( t_j \) is an independent variable of \( R \). Multiplying \( \tilde{x}_j \) to each condition of (2.23), and summing them up, we have

\[ \sum_j \tilde{p}_j\tilde{x}_j = \frac{R}{q} \sum_j v_{ij}\tilde{x}_j, \]

hence,

\[ r = \sqrt{\tilde{x}'V\tilde{x}} = \sqrt{\frac{R}{q} \tilde{p}'\tilde{x}} = \frac{R}{q} \sqrt{\tilde{p}'\tilde{t}}. \]

(2.27)

Theorem 5: If the following condition is satisfied by \( q, p \) and \( t \) for a sufficiently small \( R \):

\[ \frac{dr}{dR} = \frac{1}{q} \sqrt{\tilde{p}'\tilde{t}} \geq 1, \]

(2.28)

there exists one (degenerate) non-zero solution in Problem 2.1.

Proof: The quantity \( r(R) \) converges to a constant value \( R_0 \) when \( R \) goes to infinity. Therefore, if \( R < r(R) \) for a small \( R \), the curve \( r(R) \) will intersect with the line \( r = R \) at a point or on a line segment. Since the curve \( r(R) \) is monotone non-decreasing in \( R \), it does not have two or more separate intersections with the line \( r = R \). (see Fig. 3)

Q.E.D.

(e) Numerical Example

In order to figure out the nature of the solution of the non-linear programming Problem
Let us show a simple example:
\[
\begin{align*}
\bar{b}_1 &= 10, & \bar{b}_2 &= 12, & q &= 1.645 \text{ for } \alpha = 0.05 \\
v_{11} &= 10, & v_{12} &= 7, & q &= 2.323 \text{ for } \alpha = 0.01 \\
v_{21} &= 7, & v_{22} &= 20, \\
a_{11} &= 2, & a_{12} &= 1, & b_{1*} &= 3
\end{align*}
\]
Then we have the following non-linear programming:

\[
\begin{align*}
\text{Max } f_1(x_1, x_2) &= 10x_1 + 12x_2 - q\sqrt{10x_1^2 + 14x_1x_2 + 20x_2^2}, \\
\text{subject to } &\quad 2x_1 + x_2 \leq 3, \quad x_1 \geq 0, \quad x_2 \geq 0.
\end{align*}
\]

As the subsidiary quadratic programming, we have

\[
\begin{align*}
\text{Max } f_{11}(x_1, x_2) &= 10x_1 + 12x_2 - \frac{q}{2R}(10x_1^2 + 14x_1x_2 + 20x_2^2), \\
\text{subject to } &\quad 2x_1 + x_2 \leq 3, \quad x_1 \geq 0, \quad x_2 \geq 0.
\end{align*}
\]

First of all, let us compute the center of the ellipse \( f_{11} = \text{const.} \)

\[
\begin{align*}
\frac{\partial f_{11}}{\partial x_1} &= 0: \quad 10x_1 + 7x_2 = 10 \frac{R}{q}, \\
\frac{\partial f_{11}}{\partial x_2} &= 0: \quad 7x_1 + 20x_2 = 12 \frac{R}{q},
\end{align*}
\]

Solving these, we have
\[
\begin{align*}
x_1 &= \frac{116}{151} \frac{R}{q}, & x_2 &= \frac{50}{151} \frac{R}{q}
\end{align*}
\]

For a small \( R \), these \( x_1, x_2 \) satisfy Kuhn-Tucker condition, as is stated in the previous subsection 2. d. When \( R \) increases, the center moves from the origin to the boundary of the domain and hits the point \( P \) (Fig. 2), where we have
\[
\begin{align*}
x_1 &= 1.234, & x_2 &= 0.532, & \frac{R}{q} &= 1.6064, \\
r(R) &= 5.5038, & R &= \begin{cases}
2.6425 & \text{for } \alpha = 0.05 \\
3.7361 & \text{for } \alpha = 0.01.
\end{cases}
\end{align*}
\]

After reaching \( P \), the optimal point changes the direction of movement toward the point \( Q \) and arrives at it. Between \( P \) and \( Q \), we find the optimal solution (a contact point of the ellipse with the line segment \( PQ \)) as the intersection of the following lines:

\[
\frac{df_{11}}{dx_1} \frac{dx_1}{dx_2} + \frac{df_{11}}{dx_2} \frac{dx_2}{dx_1} = 0: \quad 4x_1 + 33x_2 = 14 \frac{R}{q},
\]

and
\[
2x_1 + x_2 = 3.
\]

Solving these, we have
\[
\begin{align*}
x_1 &= \frac{99}{62} \frac{14}{q}, & x_2 &= \frac{12}{62} + \frac{28}{62} \frac{R}{q},
\end{align*}
\]

At the point \( Q \),
\[
\begin{align*}
x_1 &= 0, & x_2 &= 3.0, & \frac{R}{q} &= \frac{99}{14},
\end{align*}
\]

\[
r(R) = \sqrt{20 \times 9} = 13.416, & \quad R &= \begin{cases}
11.633 & \text{for } \alpha = 0.05 \\
16.427 & \text{for } \alpha = 0.01.
\end{cases}
\]

For an \( R \) greater than these values the optimal solution remains at this point, which is
nothing but the solution of Problem 2.0. The curves of \( r(R) \) for \( \alpha = 0.05 \) and 0.01 are illustrated in Fig. 3. For \( \alpha = 0.05 \), the curve \( r(R) \) intersects with the line \( r = R \) after reaching \( Q \). Hence the linear programming solution \( \hat{x}_1 = 0, \hat{x}_2 = 3 \) is optimal for Problem 2.1. For \( \alpha = 0.01 \), however, there exists a solution between \( P \) and \( Q \). Solving the following quadratic equation for \( R \):

\[
R^2 = 10\hat{x}_1^2 + 14\hat{x}_1\hat{x}_2 + 20\hat{x}_2^2,
\]

(2.39)

Where \( \hat{x}_1, \hat{x}_2 \) are given by (2.37), we have \( R = 7.274 \).

After all, the final solutions of Problem 2.1, (2.30) and (2.31) for \( \alpha = 0.05 \) and 0.01 are

\[
\begin{align*}
\alpha &= 0.05 & \hat{x}_1 &= 0 & \hat{x}_2 &= 3 & a_1 &= 4.6434 & r &= R = 13.416 \\
q &= 1.645 & \beta'\beta &= 36 & f(\hat{x}_1, \hat{x}_2) &= 13.931 & a'A\hat{x} &= 4.6434 \times 3 &= 13.900, \\
\alpha &= 0.01 & \hat{x}_1 &= 0.8897 & \hat{x}_2 &= 1.2206 & a_1 &= 2.2150 & r &= R = 7.274 \\
q &= 2.323 & \beta'\beta &= 23.534 & f(\hat{x}_1, \hat{x}_2) &= 6.636 & a'A\hat{x} &= 2.215 \times 3 &= 6.645.
\end{align*}
\]

The same results are obtained by direct computation of the value of the objective function on the line \( PQ \) in Fig. 2.

Although this example is a special and unrealistic one, we can find several interesting facts. Since the profit of activity \( x_2, 12 \), is greater than that of \( x_1, 10 \), for a lower control level, \( \alpha = 0.05 \), we should only use the activity \( x_2 \). However, for a higher control level, \( \alpha = 0.01 \), we have to be cautious in choosing \( x_2 \) only, because its variance of profit \( v_{x_2} \) is comparatively large. The results of our computation say, "Do not use \( x_2 \) only, but take \( x_1 \) and \( x_2 \) together, because \( x_2 \) is more risky than \( x_1 \)."

### III. Outline of Computational Procedures

In the previous section we showed a graphical method for a simple example. A large scale problem, however, requires some other computational procedures which are suitable for a computer. Two candidates are considered at this moment. One is iteration and another
is interpolation. They have the following schemes:

(a) Iteration

Step 1: Start by solving the linear programming Problem 2.0, and obtain an initial value of $R, R_0$, and store it in $R$:

$$R_0 = \sqrt{\bar{x}_0'}\sqrt{\bar{x}_0} \to R.$$

Step 2: Using this $R$, solve Problem 2.II. If $r = \sqrt{\bar{x}'}\bar{x} = R, x(R)$ is an optimal solution of Problem 2.I. Jump to Step 3. If $r < R$, store the computed $r$ in $R$ and iterate this step.

Step 3: In order to check the degeneracy of the solution, we should test if for a small decrement of $R, \Delta R$, an intersection still appears on the line $r = R$. If it is degenerate, repeat this Step 3 from the beginning.

If there is no further degenerate solution, stop computation. In the case when we can not obtain an intersection for a sufficiently small $R$, we would have a trivial solution $\bar{x} = 0$.

(b) Interpolation

Step 1: Compute the linear programming Problem 2.0, and obtain $R_0$. Solving Problem 2.II for $R_0$, get $r(R_0)$. If $R_0 = r(R_0)$, that is all. If $r(R_0) < R_0$, replace $R_0$ by $R_0$ and $r_h$ by $r(R_0)$, where $R_0$ and $r_h$ are the coordinates of the point $H$ in Fig. 4. Suppose (2.28) in Theorem $\delta$ is satisfied, obtain a point for a small $R$. Let $R_0$ and $r_l$ be the coordinates of the point $L$.

Step 2: Compute the intersection $M$ of the line segment $LH$ with the line $r = R$. Let $R_m$ be the abscissa of $M$ and solve Problem 2.II for $R_m$ and obtain $r$. If $r < R_m$, replace $R_h$ by $R_m$ and $r_h$ by $r$. If $R_m < r$, store $R_m$ in $R_l$ and $r$ in $r_l$. Repeat this Step 2 from the beginning. If $r = R_m$, check the degeneracy by the same procedure as in (a). When there is no further degeneracy, stop computation.

In both cases, one is able to use Wolfe's quadratic programming method, which is essentially linear programming, as a subroutine.

IV. Discussions

We proposed a stochastic programming model and gave a fundamental procedure to solve it. As is shown numerically in the section 2.e, stochastic programming does not only choose an activity for its high expectation of profit, but also takes into account the risk of the activity.
Although it is difficult to state the effect of change of parameters on the optimal policy in detail, we find some properties of the solution from the experience by the hand computation and inspection. For a low control level (large $\alpha$) linear programming gives us the same result as stochastic programming, but as the level grows there appears a difference between them. Moreover, at too high a level of control (too small $\alpha$, too large $q$) a trivial solution $\delta = 0$ would be attained. However, this is reasonable, because too severe a restriction would make any activity impossible. The interaction effect between activities is another interesting thing to study, although it does not appear explicitly in the example.

Theoretically, a simpler necessary and sufficient condition for existence of a non-zero solution is desirable to avoid unnecessary effort in computation. A slightly different condition from (2.28) can be obtained as follows: differentiating $f_I$ with respect to $x_j$, we have

$$\frac{\partial f_I}{\partial x_j} = \bar{p}_i - q \frac{\sum v_{ij} x_i}{\sum q v_{ij} x_i x_j}$$

Letting $x_j$ be sufficiently small, we consider a point $(0, 0, \ldots, x_j, 0, \ldots, 0)$ where the following condition holds:

$$\frac{\partial f_I}{\partial x_j} > 0$$

Inserting the values of the coordinates of the point into (4.1), we find

$$p_j > q \sqrt{\bar{v}_{jj}} \quad \text{or} \quad p_j / \sqrt{\bar{v}_{jj}} > q$$

If (4.2) is satisfied by at least one $j$, we are sure that the origin is not optimal. Because the mean and standard deviation ratio would be more than 2.323 ($\alpha = 0.01$) in a practical case (otherwise the assumption of a normal distribution would fail as a matter of fact), (4.2) would be useful, although it seems to be slightly stronger than (2.28).

We have only discussed the maximization problem so far, but almost the same procedure can be applied to a minimization problem. Instead of (1.1), (1.2) and (1.3) we use the following expression:

$$\text{Min } f,$$

subject to

$$\text{Prob } (c'x \geq f) = a,$$

$$\text{Prob } (\sum a_{ij} x_j \geq b_i) = \bar{p}_i,$$

$x_j \geq 0$

These can be easily proved to be equivalent to the following:

$$\text{Min } f = c'x + q \sqrt{x'Vx},$$

subject to

$$\sum a_{ij} x_j \geq b_i + q \alpha_{ij},$$

$x_j \geq 0$

where $\bar{c}$ and $V$ are the mean and variance of $c$ defined similarly to (2.6) and (2.7).

It is to be noticed that the objective function is just the negative of that of the maximization problem, and that the right hand side of the constraint has a positive sign followed by $q \alpha_{ij}$ instead of negative.

Finally we give a glance to more general stochastic programming in which the components of the matrix $A$, $a_{ij}$s, are not constants but random variables. In this case the probabilistic condition,
is also transformed into an inequality, assuming a multinormal distribution of \( a_{ij} \) and \( b_i \), and \( \beta_i \geq 0.5 \).

\[
(b_i - \sum_j a_{ij} x_j) - q_i \sqrt{a_i \Sigma_j^2 + \sum_j v_{ij} x_j + \sum_k v_{ijk} x_j x_k} = 0,
\]

where

\[
q_i = I^{-1}(\beta_i), \quad a_{ij} = E(a_{ij})
\]

\[
v_{ij} = E(b_i - \bar{b}_i)(a_{ij} - \bar{a}_{ij}), \quad v_{ijk} = E(a_{ik} - \bar{a}_{ik})(a_{jk} - \bar{a}_{jk}).
\]

We find that the square root function in the above inequality is proved to be still convex: the quadratic function under the square root can be expressed in the same form as \( \sum_j v_{ij} x_i x_j \) by introducing a dummy variable for \( b_i \), which can be replaced by unity after the finishing proof without loss of its validity. Hence, a global maximum (or degenerate maxima) still coincides with a local maximum (or degenerate maxima). At this moment the most promising method for solving this kind of non-linear programming is supposed to be the gradient method.\(^{10}\)

### Appendix

**Convexity of \( \sqrt{x'Vx} \)**

Suppose \( x, y \) and \( z \) be vectors which have the following relation:

\[
z = \lambda x + (1 - \lambda)y, \quad 0 \leq \lambda \leq 1.
\]

Then we have

\[
\text{Sign} \left\{ \sqrt{x'Vz} - (\lambda \sqrt{x'Vx} + (1 - \lambda) \sqrt{y'Vy}) \right\}
\]

\[
= \text{Sign} \left\{ \sqrt{x'Vz} - (\lambda \sqrt{x'Vx} + (1 - \lambda) \sqrt{y'Vy})^2 \right\}
\]

\[
= \text{Sign} \left\{ \lambda x'Vx + 2\lambda - \lambda^2 \sqrt{x'Vx} \sqrt{y'Vy} + (1 - \lambda) y'Vy \right\}
\]

\[
- \sqrt{x'Vx} - 2(1 - \lambda) \sqrt{x'Vx} \sqrt{y'Vy} + (1 - \lambda)^2 y'Vy \right\}
\]

\[
= \text{Sign} \left\{ 2\lambda - (1 - \lambda) (x'Vy - \sqrt{x'Vx} \sqrt{y'Vy}) \right\}. \quad \text{(a)}
\]

Since the matrix \( V \) is positive semi-definite, for an arbitrary number \( t \) we have

\[
(tx + y)'V(tx + y) = t x'Vx + 2tx'Vy + y'Vy \geq 0
\]

Hence,

\[
(x'Vy)^2 \leq (x'Vx)(y'Vy)
\]

then

\[
x'Vy \leq \sqrt{x'Vx \sqrt{y'Vy}}
\]

because \( x'Vx \) and \( y'Vy \) are non-negative. Accordingly the sign function \( (a) \) is non-positive.

Q.E.D.

### References


