

Second Degree Chance Constraints with Lognormal Random Variables – An Application to Fisher's Discriminant Function for Separation of Populations

Vaskar Sarkar^{1,*}, Kripasindhu Chaudhuri², Rathindra Nath Mukherjee³

¹Department of Mathematics, Aryabhatta Institute of Engineering and Management Durgapur, Panagarh, Dist.- Burdwan, Pin - 713148, West Bengal, India

²Department of Mathematics, Jadavpur University, Jadavpur, Kolkata, 700032, West Bengal, India

³Department of Mathematics, Burdwan University, Burdwan, West Bengal, India

Abstract In this paper, we have discussed a transformation procedure of the second-degree chance constraints to the deterministic constraints for mathematical programming problems having general second degree chance constraints with lognormal random variables. We have used geometric inequality for this transformation. The transformed deterministic problem having non-linear constraints and linear or non-linear objective function can be solved using non-linear programming algorithm. Also we have applied this model to Fisher's discriminant function for separation of populations. A numerical simulation have been considered along with a graphical representation of the reduced solution region.

Keywords Stochastic Programming, Chance Constraint Programming, Lognormal Random Variable, Geometric Inequality, Deterministic Reduction, Non-linear Programming, Discriminant Function and Separation of Population

1. Introduction

Stochastic or probabilistic programming[13] deals with the situations where some or all the parameters of the mathematical programming problem are described by stochastic or random variables rather than by deterministic ones. Several models have been presented in the field of stochastic programming[21]. Two major approaches to stochastic programming[13,14] are recognized as:

1. Chance constrained programming
2. Two-stage programming

The chance constrained programming (CCP) can be used to solve problems involving chance constraints, i.e., constraints having finite probability of being violated. Its main feature is that the resulting decision ensures the probability of satisfying the constraints, i.e. the confidence level of being feasible. Thus, using CCP, the relationship between profitability and reliability can be quantified. The use of probabilistic constraints was initially introduced by Charnes et al.[6,7], while an early use of probabilistic programming in environmental economics is by Maler[17]. Since then, a number of environmental management case studies use probabilistic constraints[5,10,11,19,22,23]. Also the CCP, in recent years, has been generalized in several

directions and has various applications[21].

Many authors studied and developed this problem for the parameters having normal distribution, uniform distribution and exponential distribution, where the chance constraints are linear[3, 4, 21]. But lognormal distribution (with two parameters) has a significant role in human and ecological risk assessment for many reasons, of which the main three reasons are (i) many physical, chemical, biological, toxicological and statistical processes tend to create random variables that follow lognormal distributions[12], (ii) lognormal distributions are self-replicating under multiplication and division, i.e. products and quotients of lognormal random variables themselves follow lognormal distributions[1, 9] and (iii) when the conditions of the Central Limit Theorem hold[16], the mathematical process of multiplying a series of random variables will produce a new random variable (the product), which tends (in limit) to be lognormal in character, regardless of the distribution from which the input variables arise[2].

Also we may consider the fact that many environmental variables are non-negative means that they are generated by a skewed distribution. Several skewed probability models have been used to describe environmental data, including the Poisson, negative binomial, Weibull, gamma, exponential and the lognormal. Among these distributions, the lognormal has been the most widely applied[18]. In the year 1996, Cooper et.al.[8] noted the importance of using skewed distributions, to represent environmental variables within mathematical programming and lognormal distribution was

* Corresponding author:

sarkar_vaskar@rediffmail.com (Vaskar Sarkar)

Published online at <http://journal.sapub.org/ajcam>

Copyright © 2013 Scientific & Academic Publishing. All Rights Reserved

one among those skewed distributions.

So here in our model, we have considered that the parameters of the chance constraints follow lognormal distribution having known mean and standard deviation. In this paper, we have tried to reduce the general second degree (with respect to decision variables) chance constraints to non-linear deterministic constraints, but the objective function, in both of the considered and transformed models, may be linear or nonlinear with deterministic parameters and also we have applied this model to Fisher's discriminant function for separation of populations[20].

2. Chance Constrained Programming

The chance constrained programming (CCP) involves constraints expressed in terms of probabilities. Chance constrained programming represents risk in a qualitative way, i.e., in chance constrained programming, only the possibility of infeasibility is at stake regardless of the amount by which the constraints are violated. The simplest form of a chance constraint can be given as,

$\Pr\{\underline{x}^t A^k \underline{x} + B^k \underline{x} + c_k \leq t_k\} \geq \beta_k$ where, A^k is an $m \times n$ matrix usually termed to as a technology matrix, \underline{x} is an n -component vector of decision variables, B^k is an n -component vector and c_k & t_k are scalars and the scalar β_k stands for the probability with which constraint $\underline{x}^t A^k \underline{x} + B^k \underline{x} + c_k \leq t_k$ must be satisfied. The scalar β_k represents the reliability of decision \underline{x} and $1 - \beta_k$ gives the risk of infeasibility associated with decision \underline{x} . The choice of β_k is at the discretion of the decision-maker (DM), in other words, it is a policy maker's choice, which can be interpreted as an expression of the regulator's aversion to uncertainty[15]. In this paper we consider that a_{ij}^k (the ij^{th} element of A^k) and b_i^k (the i^{th} component of B^k) and c_k & t_k are independent lognormal random variables with known mean and standard deviation (s.d.).

3. Deterministic Reduction of the Mathematical Model

The generalized form of the second degree chanced constrained programming problem can be written as follows:

To find $\mathbf{X} = (x_1, x_2, x_3, x_4, \dots, x_n)$ so as to,

$$\text{Maximize } f(\mathbf{X}) \quad (1)$$

subject to the constraints,

$$\Pr\left[\sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j + \sum_{l \in S_3} b_l^k x_l + c_k \leq t_k\right] \geq 1 - p_k, k \in S_4 \quad (2)$$

$$x_i \geq L_i > 0, i = 1, 2, 3, \dots, n \quad (3)$$

where, $0 < p_k < 1, \forall k \in S_4$ and

S_1, S_2, S_3, S_4 are index sets and $S_1, S_2, S_3, S_4 \subseteq \{1, 2, 3, \dots, n\}$.

Here $a_{ij}^k, b_l^k, c_k, t_k$ are all independent lognormal random variables. Let us consider $\#S_1 = n_1, \#S_2 = n_2, \#S_3 = n_3, \#S_4 = n_4$ and also consider that $a_{ij}^k (> 0), b_l^k (> 0), c_k (> 0), t_k (> 0)$ are independent lognormal random variables $\forall i \in S_1, j \in S_2, l \in S_3, k \in S_4$.

Here our target is to transform the probabilistic second-degree constraints (2) in to deterministic constraints. We use Geometric Inequality (GI) i.e. $AM \geq GM$ to reduce the constraints in (2) in the following way:

We have

$$\frac{1}{n_2} \sum_{j \in S_2} a_{ij}^k x_j \geq \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}, \quad \forall i \in S_1, \forall k \in S_4$$

i.e.,

$$\frac{1}{n_2} x_i \sum_{j \in S_2} a_{ij}^k x_j \geq x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}, \quad \forall i \in S_1, \forall k \in S_4$$

i.e.,

$$\sum_{i \in S_1} \frac{1}{n_2} x_i \sum_{j \in S_2} a_{ij}^k x_j \geq \sum_{i \in S_1} \left\{ x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}} \right\}, \quad \forall k \in S_4$$

i.e.,

$$\frac{1}{n_2} \sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j \geq \sum_{i \in S_1} \left\{ x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}} \right\}, \quad \forall k \in S_4 \quad (4)$$

From the RHS of (4) using GI we can get,

$$\frac{1}{n_1} \sum_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \geq \left(\prod_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \right)^{\frac{1}{n_1}}, \forall k \in S_4 \quad (5)$$

Thus from (4) and (5) we get,

$$\begin{aligned} \frac{1}{n_2} \sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j &\geq \sum_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \\ &\geq n_1 \left(\prod_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \right)^{\frac{1}{n_1}}, \forall k \in S_4 \\ \Rightarrow \sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j &\geq n_1 n_2 \left(\prod_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \right)^{\frac{1}{n_1}}, \quad \forall k \in S_4 \end{aligned} \quad (6)$$

Also we can write

$$\frac{1}{n_3} \sum_{l \in S_3} (b_l^k x_l) \geq \left(\prod_{l \in S_3} (b_l^k x_l) \right)^{\frac{1}{n_3}} \quad (7)$$

Thus we get

$$\begin{aligned} \sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j + \sum_{l \in S_3} (b_l^k x_l) + c_k \\ \geq n_1 n_2 \left(\prod_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \right)^{\frac{1}{n_1}} + n_3 \left(\prod_{l \in S_3} (b_l^k x_l) \right)^{\frac{1}{n_3}} + c_k \end{aligned} \quad (8)$$

Now on the RHS of (8) again applying GI we can get

$$\begin{aligned} \frac{1}{3} \left[n_1 n_2 \left(\prod_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \right)^{\frac{1}{n_1}} + n_3 \left(\prod_{l \in S_3} (b_l^k x_l) \right)^{\frac{1}{n_3}} + c_k \right] \\ \geq \left[n_1 n_2 \left(\prod_{i \in S_1} \{x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}}\} \right)^{\frac{1}{n_1}} n_3 \left(\prod_{l \in S_3} (b_l^k x_l) \right)^{\frac{1}{n_3}} c_k \right]^{\frac{1}{3}} \end{aligned} \quad (9)$$

Therefore from (8) & (9) we get

$$\sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j + \sum_{l \in S_3} (b_l^k x_l) + c_k$$

$$\geq 3 \left[n_1 n_2 \left(\prod_{i \in S_1} \left\{ x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}} \right\} \right)^{\frac{1}{n_1}} n_3 \left(\prod_{l \in S_3} (b_l^k x_l) \right)^{\frac{1}{n_3}} c_k \right]^{\frac{1}{3}} \quad (10)$$

Now from (2) the chance constraints implies

$$\Pr \left[3 \left[n_1 n_2 \left(\prod_{i \in S_1} \left\{ x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}} \right\} \right)^{\frac{1}{n_1}} n_3 \left(\prod_{l \in S_3} (b_l^k x_l) \right)^{\frac{1}{n_3}} c_k \right]^{\frac{1}{3}} \leq t_k \right] \geq 1 - p_k, k \in S_4 \quad (11)$$

Taking natural logarithm over the elements under probability inside the bracket of (11) we have

$$\ln 3 + \frac{1}{3} [\ln(n_1 n_2 n_3) + \frac{1}{n_1} \ln \left(\prod_{i \in S_1} \left\{ x_i \left(\prod_{j \in S_2} a_{ij}^k x_j \right)^{\frac{1}{n_2}} \right\} \right) + \frac{1}{n_3} \ln \left(\prod_{l \in S_3} (b_l^k x_l) \right) + \ln c_k] \leq \ln t_k$$

i.e.,

$$3 \ln 3 + \ln(n_1 n_2) + \ln n_3 + \frac{1}{n_1} \sum_{i \in S_1} \{ \ln x_i + \frac{1}{n_2} \ln \left(\prod_{j \in S_2} a_{ij}^k x_j \right) \} + \frac{1}{n_3} \sum_{l \in S_3} (\ln b_l^k + \ln x_l) + \ln c_k \leq 3 \ln t_k$$

, $k \in S_4$

$$\Leftrightarrow \frac{1}{n_1 n_2} \sum_{i \in S_1} \sum_{j \in S_2} \ln a_{ij}^k + \frac{1}{n_3} \sum_{l \in S_3} \ln b_l^k + \ln c_k - 3 \ln t_k$$

$$\leq -(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \frac{1}{n_1} \sum_{i \in S_1} \ln x_i$$

$$- \frac{1}{n_2} \sum_{j \in S_2} \ln x_j - \frac{1}{n_3} \sum_{l \in S_3} \ln x_l, k \in S_4 \quad (12)$$

Now from (12) the chance constraints becomes

$$\Pr \left[\frac{1}{n_1 n_2} \sum_{i \in S_1} \sum_{j \in S_2} \ln a_{ij}^k + \frac{1}{n_3} \sum_{l \in S_3} \ln b_l^k + \ln c_k - 3 \ln t_k \right.$$

$$\left. \leq -(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \frac{1}{n_1} \sum_{i \in S_1} \ln x_i - \frac{1}{n_2} \sum_{j \in S_2} \ln x_j - \frac{1}{n_3} \sum_{l \in S_3} \ln x_l \right]$$

$$\geq 1 - p_k, k \in S_4 \quad (13)$$

Since each of $a_{ij}^k, b_l^k, c_k, t_k$ are independent lognormal random variables having known mean and standard deviation (s.d.), so $\frac{1}{n_1 n_2} \sum_{i \in S_1} \sum_{j \in S_2} \ln a_{ij}^k + \frac{1}{n_3} \sum_{l \in S_3} \ln b_l^k + \ln c_k - 3 \ln t_k$ is a random variable having normal distribution with known mean and s.d., say μ_k and σ_k respectively $k \in S_4$.

Thus we have the reduced form of the chance constraints as follows

$$\Pr \left[\frac{\frac{1}{n_1 n_2} \sum_{i \in S_1} \sum_{j \in S_2} \ln a_{ij}^k + \frac{1}{n_3} \sum_{l \in S_3} \ln b_l^k + \ln c_k - 3 \ln t_k - \mu_k}{\sigma_k} \right. \\ \left. \leq \frac{-(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \frac{1}{n_1} \sum_{i \in S_1} \ln x_i - \frac{1}{n_2} \sum_{j \in S_2} \ln x_j - \frac{1}{n_3} \sum_{l \in S_3} \ln x_l - \mu_k}{\sigma_k} \right] \geq 1 - p_k, k \in S_4. \quad (14)$$

Now, let $\Phi(z) = \int_{-\infty}^z \phi(t) dt$, where $\phi(t)$ be the probability density function of standard normal variate then we can write,

$$\Phi \left(\frac{-(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \frac{1}{n_1} \sum_{i \in S_1} \ln x_i - \frac{1}{n_2} \sum_{j \in S_2} \ln x_j - \frac{1}{n_3} \sum_{l \in S_3} \ln x_l - \mu_k}{\sigma_k} \right) \\ \geq \Phi(Z_k), k \in S_4$$

where, $\Phi(Z_k) = 1 - p_k, k \in S_4$

i.e.

$$\frac{-(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \frac{1}{n_1} \sum_{i \in S_1} \ln x_i - \frac{1}{n_2} \sum_{j \in S_2} \ln x_j - \frac{1}{n_3} \sum_{l \in S_3} \ln x_l - \mu_k}{\sigma_k} \\ \geq Z_k, k \in S_4$$

i.e.

$$\frac{1}{n_1} \sum_{i \in S_1} \ln x_i + \frac{1}{n_2} \sum_{j \in S_2} \ln x_j + \frac{1}{n_3} \sum_{l \in S_3} \ln x_l \\ \leq -(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \mu_k - \sigma_k Z_k, k \in S_4 \quad (15)$$

Hence using (1), (15) & (3) the deterministic implication with nonlinear constraints of the probabilistic problem is as follows:

To find $\mathbf{X}=(x_1, x_2, x_3, x_4, \dots, x_n)$, so as to,

$$\text{Maximize } f(\mathbf{X}) \quad (16)$$

subject to the constraints

$$\frac{1}{n_1} \sum_{i \in S_1} \ln x_i + \frac{1}{n_2} \sum_{j \in S_2} \ln x_j + \frac{1}{n_3} \sum_{l \in S_3} \ln x_l \\ \leq \underset{k \in S_4}{\text{Min}} \{ -(3 \ln 3 + \ln(n_1 n_2) + \ln n_3) - \mu_k - \sigma_k Z_k \} \quad \forall k \in S_4, \quad (17)$$

$$x_i \geq L_i > 0, i = 1, 2, 3, \dots, n. \quad (18)$$

If the solution of the problem stated in (16), (17) & (18) satisfies the condition stated in (2), then the same is the optimum solution of the given problem.

This model can be easily modified to the following models:

Model - A: Model having only second-degree terms in the chance constraints.

Model - B: Model having only linear chance constraints

4. Model – A: Model Having Only Second-degree terms in the Chance Constraints

In this case the form of the chance constrained model, obtained from (1), (2) & (3), is as follows:

To find $X=(x_1, x_2, x_3, x_4, \dots, x_n)$, so as to,

$$\text{Maximize } f(X) \quad (19)$$

subject to the constraints,

$$\Pr\left[\sum_{i \in S_1} \sum_{j \in S_2} a_{ij}^k x_i x_j \leq t_k\right] \geq 1 - p_k, k \in S_4 \quad (20)$$

$$x_i \geq L_i > 0, i = 1, 2, 3, \dots, n \quad (21)$$

where, $0 < p_k < 1, \forall k \in S_4$ and S_1, S_2, S_4 are index sets and all a_{ij}^k and t_k follows Log-normal distribution with known mean & s.d..

So there is no existence of the index set S_3 .

Thus the deterministic reduction, having non-linear constraint, of the above model can be obtained from (16), (17) & (18) by just omitting the terms involving n_3 and the term $3 \ln 3$, which is as follows:

To find $X=(x_1, x_2, x_3, x_4, \dots, x_n)$, so as to,

$$\text{Maximize } f(X) \quad (22)$$

subject to the constraints,

$$\frac{1}{n_1} \sum_{i \in S_1} \ln x_i + \frac{1}{n_2} \sum_{j \in S_2} \ln x_j \quad (23)$$

$$\leq \min_{k \in S_4} \{-\ln(n_1 n_2) - \mu_k - \sigma_k Z_k\}$$

$$x_i \geq L_i > 0, i = 1, 2, 3, \dots, n \quad (24)$$

where, μ_k & σ_k are the mean and s.d. of the normal

random variable $\frac{1}{n_1 n_2} \sum_{i \in S_1} \sum_{j \in S_2} \ln a_{ij}^k - \ln t_k$ and

$$\Phi(Z_k) = 1 - p_k, \forall k \in S_4.$$

5. Model – B: Model Having Only Linear Terms in the Chance Constraints

Here the form of the chance constrained model, obtained from (1), (2) & (3), is as follows:

To find $X=(x_1, x_2, x_3, x_4, \dots, x_n)$, so as to,

$$\text{Maximize } f(X) \quad (25)$$

subject to the constraints,

$$\Pr\left[\sum_{l \in S_3} b_l^k x_l \leq t_k\right] \geq 1 - p_k, k \in S_4 \quad (26)$$

$$x_l \geq L_l > 0, \forall l \in S_3 \quad (27)$$

where, $0 < p_k < 1, \forall k \in S_4$ and S_4 is index set.

So there is no existence of the index sets S_1, S_2 .

Thus the deterministic reduction, having linear constraint, of the above model can be obtained from (25), (26) & (27) by just omitting the terms involving n_1, n_2 and the term $3 \ln 3$, which is as follows:

To find $X=(x_1, x_2, x_3, x_4, \dots, x_n)$ so as to,

$$\text{Maximize } f(X) \quad (28)$$

subject to the constraints,

$$\frac{1}{n_3} \sum_{l \in S_3} \ln x_l \leq -\ln n_3 - \mu_k - \sigma_k Z_k, \quad (29)$$

$$x_l \geq L_l > 0, \forall l \in S_3 \quad (30)$$

where, μ_k & σ_k are the mean and s.d. of the normal

random variable $\frac{1}{n_3} \sum_{l \in S_3} \ln b_l^k - \ln t_k$ and

$$\Phi(Z_k) = 1 - p_k, \forall k \in S_4.$$

We can rewrite the above deterministic reduction as follows:

To find $X=(x_1, x_2, x_3, x_4, \dots, x_n)$, so as to,

$$\text{Maximize } f(X) \quad (31)$$

subject to the constraints,

$$\sum_{l \in S_3} \ln x_l \leq \min_{k \in S_4} \{-n_3 \ln n_3 - \mu'_k - \sigma'_k Z_k\} \quad (32)$$

$$x_l \geq L_l > 0, \forall l \in S_3 \quad (33)$$

where, μ'_k & σ'_k are the mean and s.d. of the normal

random variable $\sum_{l \in S_3} \ln b_l^k - n_3 \ln t_k$ and

$$\Phi(Z_k) = 1 - p_k, \forall k \in S_4.$$

6. Application of the Model in Fisher's Discriminant Analysis function – Separation of Population

Fisher derived the linear classification statistic (Johnson and Wichern (2001)) using the following idea:

His idea was to transform the multivariate observations X to univariate observations y such that the y 's derived from

population π_1 and π_2 were separated as much as possible.

Fisher suggested to take linear combination of X to create y 's, because they are simple enough functions of the component of X . Also Fisher's approach does not assume that the populations are Normal and it assumes that the population covariance matrices are equal.

So the objective is to select the linear combination of X to achieve the maximum separation of the population mean vectors $\underline{\mu}_1$ and $\underline{\mu}_2$ i.e.

$$\max_{\underline{L}} [\underline{L}'(\underline{\mu}_1 - \underline{\mu}_2)]^2. \text{ But this does not admit}$$

finite solution for \underline{L} . To make the problem meaningful,

Fisher considered the standardized ratio $\frac{[\underline{L}'(\underline{\mu}_1 - \underline{\mu}_2)]^2}{\underline{L}' \hat{\Sigma} \underline{L}}$,

that is, to find \underline{L} so as to $\max_{\underline{L}} \frac{[\underline{L}'(\underline{\mu}_1 - \underline{\mu}_2)]^2}{\underline{L}' \hat{\Sigma} \underline{L}}$.

This admits finite solution for \underline{L} . Alternatively, one may write the above problem, in case as follows Σ is unknown and can be estimated by $\hat{\Sigma} = (\hat{\sigma}_{ij})_{n \times n}$ and

$\underline{\mu}_1$ and $\underline{\mu}_2$ are known, as follows:

To find \underline{L} so as to

$$\max_{\underline{L}} [\underline{L}'(\underline{\mu}_1 - \underline{\mu}_2)]^2$$

subject to, $\Pr[\underline{L}' \hat{\Sigma} \underline{L} \leq t] \geq 1 - p$, $\underline{L} > 0$, where $p (> 0)$ be very small.

Here, let us consider that, $\hat{\Sigma}$ component-wise follows lognormal distribution with known mean and s.d. where $\hat{\Sigma} = (\hat{\sigma}_{ij})_{n \times n}$.

So the above problem reduces to the following form:

To find $(\underline{L}_1, \underline{L}_2, \underline{L}_3, \dots, \underline{L}_n)$, so as to

$$\max_{\underline{L}} f(\underline{L})$$

subject to,

$$\Pr[\sum_i \sum_j \hat{\sigma}_{ij} l_i l_j \leq t] \geq 1 - p,$$

$$i, j = 1, 2, 3, \dots, n,$$

$$\underline{L}_i > 0, \forall i$$

where, $f(\underline{L}) = [\underline{L}'(\underline{\mu}_1 - \underline{\mu}_2)]^2$.

Here $\hat{\sigma}_{ij}$ follows lognormal distribution for all i, j , with known mean and s.d. and $t (> 0)$ is some suitable constant (very small). Then using Model – A, the above problem can be solved.

7. Managerial Application

Waste water pollution can be controlled, at the river and associated land, either by emission restrictions of waste water or by land and water use restrictions. Typically the direct control of emissions is impossible due to the high cost of observing industry emissions. A further problem is that waste water emissions, which are determined by production events of industries, are stochastic and this leads to wide variations in the waste water pollution concentrations in surface and groundwater. Thus in practice the regulatory body is unable to apply deterministic water quality standards, but instead, must set reliability level which states that the minimum standard must be met for some proportion of the time. In setting this reliability level the regulatory body trades off reliability against associated costs. Also in the study of waste water pollution control, we focus our attention to probabilistic programming, because in most of the environmental management problems, quantitative data on the cost of exceeding emission standards are impossible to estimate. This is the case with waste water emissions where the costs of exceeding the standard would include health costs and a range of costs due to ecological damage. Our models can be used suitably to such problems based on relevant data collected through field work.

8. Numerical Example

Let us consider the following maximization problem:

To find x_1, x_2, x_3 so as to

$$\text{Maximize } f(X) = x_1 + x_2 + x_3 \quad (34)$$

subject to,

$$\Pr[\sum_{i=1}^3 \sum_{j=1}^3 a_{ij}^k x_i x_j \leq t_k] \geq 1 - p_k, \quad (35)$$

$$k = 1, 2,$$

$$x_i \geq 0.8, i = 1, 2, 3 \quad (36)$$

Where a_{ij}^k and t_k follows independent lognormal distribution with known mean & s.d. and $p_1 = 0.05, p_2 = 0.02$.

Table 1. For $k = 1$ mean of a_{ij}^k and t_k

$E(a_{11}^1) = 3$	$E(a_{12}^1) = 2$	$E(a_{13}^1) = 6$	$E(t_1) = 50$
$E(a_{21}^1) = 5$	$E(a_{22}^1) = 7$	$E(a_{23}^1) = 9$	
$E(a_{31}^1) = 10$	$E(a_{32}^1) = 12$	$E(a_{33}^1) = 5$	

Table 2. For $k = 1$ variance of a_{ij}^k and t_k

$V(a_{11}^1) = 8$	$V(a_{12}^1) = 5$	$V(a_{13}^1) = 25$	$V(t_1) = 10$
$V(a_{21}^1) = 20$	$V(a_{22}^1) = 30$	$V(a_{23}^1) = 40$	
$V(a_{31}^1) = 45$	$V(a_{32}^1) = 50$	$V(a_{33}^1) = 20$	

Table 3. For $k = 2$ mean of a_{ij}^k and t_k

$E(a_{11}^2) = 5$	$E(a_{12}^2) = 7$	$E(a_{13}^2) = 3$	$E(t_1) = 70$
$E(a_{21}^2) = 9$	$E(a_{22}^2) = 3$	$E(a_{23}^2) = 6$	
$E(a_{31}^2) = 7$	$E(a_{32}^2) = 16$	$E(a_{33}^2) = 30$	

Table 4. For $k = 2$ variance of a_{ij}^k and t_k

$V(a_{11}^2) = 4$	$V(a_{12}^2) = 6$	$V(a_{13}^2) = 22$	$V(t_1) = 20$
$V(a_{21}^2) = 12$	$V(a_{22}^2) = 30$	$V(a_{23}^2) = 40$	
$V(a_{31}^2) = 30$	$V(a_{32}^2) = 40$	$V(a_{33}^2) = 21$	

In this problem we can fit the Model – A (Section 4) discussed above and using (22) – (24) the problem becomes,

$$\text{Maximize } f(X) = x_1 + x_2 + x_3$$

subject to,

$$\frac{1}{3}(\ln(x_1) + \ln(x_2) + \ln(x_3)) + \frac{1}{3}(\ln(x_1) + \ln(x_2) + \ln(x_3)) \leq 1.304473,$$

$$x_1 \geq 0.8, x_2 \geq 0.8, x_3 \geq 0.8$$

The global solution obtained using optimization software ‘Mathematica 5.2’ is as follows:

$$z_{\max} = 12.6563 \quad \text{and} \quad x_1 = 0.8, \quad x_2 = 0.8, \quad x_3 = 11.0563.$$

Also using generic package ‘MATLAB 7’ we have verified that, this solution satisfies the constraints (35). Thus the optimal solution to the given problem is as obtained above.

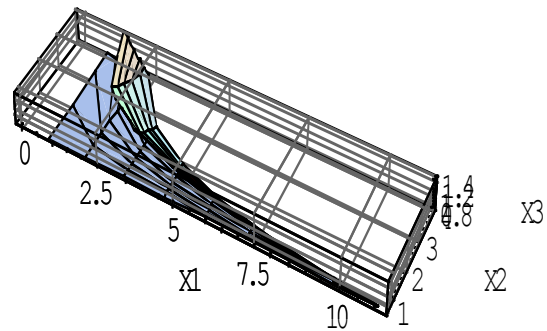
9. Discussion

In this paper, we have reduced the probabilistic constraints to deterministic constraints. The reduced deterministic constraints along with the feasibility conditions i.e.,

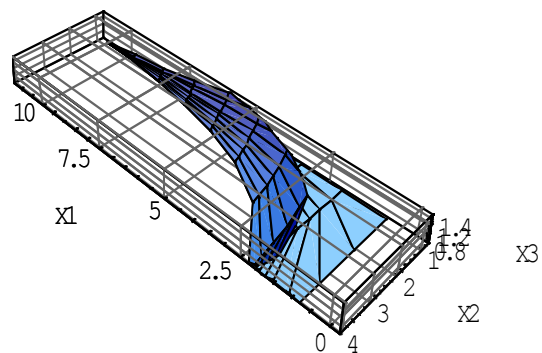
$$\frac{1}{3}(\ln(x_1) + \ln(x_2) + \ln(x_3)) + \frac{1}{3}(\ln(x_1) + \ln(x_2) + \ln(x_3)) \leq 1.304473,$$

$$x_1 \geq 0.8, x_2 \geq 0.8, x_3 \geq 0.8$$

generates the solution region as shown in the following figures.

**Figure 1.** Reduced solution region

The solution obtained in the above numerical example lies within the solution region shown in the following figure. Also we verified that this optimal solution of the reduced problem satisfies the probabilistic constraints of the original problem. This verification is very much required due to change in solution space (event space) at the time of deterministic reduction and it corroborates the optimal solution of the reduced problem as an optimal solution of the original problem.

**Figure 2.** Different angle view of reduced solution region

10. Conclusions

The basic objective of this work is to reduce the probabilistic constraints to deterministic constraints through

implicative relationship. While reducing the probabilistic constraints to deterministic constraints, the event space under consideration has been enlarged. As a result, implicative reduction calls for verification of whether the optimal solution under extended region satisfies the original chance constraints or not. Once this verification gives a positive response, the obtained optimal solution of the extended problem becomes the optimal solution of the original problem. In case of a tight extension of the original problem, the positive response during verification becomes a likely one. Indirectly, tight reduction through implicative relationship calls for use of sharp inequalities including separation of coefficient parameters, if needed under distributional assumption. For a lognormal setup, separation of coefficients from the lognormal variables is a must, because often wise evaluation of the resultant distribution becomes a tedious task. For a normal setup, this separation is redundant. Moreover for some distribution the joint probability distribution for large number of random variable can be evaluated using mathematical induction method[3]. But for some distribution, like lognormal, it is often become impossible to find joint probability distribution for large number of random variable. In that case our method is very much useful to reduce the probabilistic constraints to deterministic one. Further, more refinement of the solution may be possible applying the concept of genetic algorithm to the present solution.

REFERENCES

- [1] Aitchison, J., Brown, J.A.C. (1957) 'The Lognormal Distribution', Cambridge University Press, Cambridge, UK.
- [2] Benjamin, J.R. and Cornell, C.A. (1970) 'Probability, Statistics, and Decision for Civil Engineers', McGraw Hill, New York, NY.
- [3] Biswal, M.P., Biswal, N.P. and Li, Duan (1998) 'Probabilistic linear programming problems with exponential random variables: A technical note', European Journal of Operational Research, Vol. 111, pp.589–597.
- [4] Biswal, M.P., Biswal, N.P. and Li, Duan (2005) 'Probabilistic linearly constrained programming problems with log normal random variables', OPSEARCH, Vol. 42(1), pp.70–76.
- [5] Bouzaher, A., O utt, S. (1992) 'A stochastic linear programming model for corn residue production', Journal of the Operational Research Society, Vol. 43(9), pp.227–257.
- [6] Charnes, A., and Cooper, W.W. (1959) 'Chance constrained programming', Management Science, Vol. 6, pp.227–243.
- [7] Charnes, A., Cooper, W., Symonds, G. (1958) 'Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil', Management Science, Vol. 4, pp.235–263.
- [8] Cooper, W., Hemphill, H., Huang, Z., Li, S., Lelas, V., Sullivan, D. (1996) 'Survey of mathematical programming models in air pollution', European Journal of Operational Research, Vol. 96(1), pp.1–35.
- [9] Crow, E.L. and Shimizu, K. Eds. (1998) 'Lognormal Distributions: Theory and Application', Dekker, New York.
- [10] Ellis, J. (1987) 'Stochastic water quality optimization using imbedded chance constraints', Water Resources Research, Vol. 23(12), pp.2227–2238.
- [11] Hanley, N., Faichney, R., Munro, A., Shortle, J. (1998) 'Economic and environmental modelling for pollution control in an estuary', Journal of Environmental Management, Vol. 52(3), pp.211–225.
- [12] Hattis, D.B. and Burmaster, D.E. (1994) 'Assessment of variability and uncertainty distributions for practical risk assessments', Risk Analysis, Vol. 14(5), pp.713–730.
- [13] Kall, P. (1976) 'Stochastic programming', Springer, Berlin.
- [14] Kambo, N.S. (1984) 'Mathematical programming Techniques', Allied East-West Press Pvt. Ltd.
- [15] Lichtenberg, E., Zilberman, D. (1988) 'Efficient regulation of environmental risks', Quarterly Journal of Economics, Vol. 103, pp.167–178.
- [16] Mood, A.M., Graybill, F.A., and Boes, D.C. (1974) 'Introduction to the Theory of Statistics', Third Edition, McGraw Hill, New York, NY.
- [17] Maler, K. (1974) 'Environmental Economics: A Theoretical Inquiry', The Johns Hopkins University Press, Baltimore, MD.
- [18] Parkin, T., Robinson, J. (1992) 'Analysis of lognormal data. In: Stewart, B. (Ed.)', Advances in Soil Science, Springer, New York, pp. 193–235.
- [19] Pinter, J. (1991) 'Stochastic modelling and optimization for environmental management', Annals of Operations Research, Vol. 31, pp.527–544.
- [20] Johnson, R.A., Wichern, D.W. (2001) 'Applied multivariate statistical analysis', Third Edition, PHI Private Limited, India.
- [21] Stancu-Minasian, I.M., and Wets, M.J. (1976) 'A research bibliography in stochastic programming, 1955–1975', Operations Research, Vol. 24, pp.1078–1119.
- [22] Wagner, H.M. (1985) 'Principles of Operations Research with Applications in Managerial Decisions', Second Edition, Prentice Hall of India Pvt. Ltd., India.
- [23] Wagner, B., Gorelick, S. (1987) 'Optimal groundwater management under parameter uncertainty', Water Resources Research, Vol. 23(7), pp.1162–1174.