International Journal of
Operations Research

International Journal of Operations Research Vol. 11, No. 4, 100–111 (2014)

# A Fuzzy Chance Constraint Programming Approach for Optimal Allocation in Multivariate Stratified Surveys: A Compromise Solution

**Shamsher Khan*  and M.M. Khalid**

Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh 202002, India

*Received May 2014; Revised October 2014; Accepted October 2014*

***Abstract*** — Optimal allocation of sample size among various strata is an important step to get the precise estimates for population parameters and to reduce the cost of the survey. A reasonable criterion for optimal allocation is the minimization of the variances of the estimates for a specified cost or to minimize the cost of survey for desired precision of the estimates. The total cost of survey is a function of sample sizes allocated to various strata and the unitary cost of collecting information/measurement associated to particular stratum. The measurement cost $c_h$ which vary from stratum to stratum and affected by some factors such as nature of climate, weather conditions which occurs randomly is considered as fuzzy random variable (FRV). The survey is taken as multivariate in which we want to study more than one characteristic. Thus, in this paper the problem of optimum allocation in multivariate stratified sampling is formulated as a multiobjective fuzzy chance constrained programming (MOFCCP) problem with measurement cost $c_h$ as a normally distributed FRV. Compromise solution of its deterministic equivalent is obtained by goal programming technique. In addition, an illustrative example is also given to demonstrate the correctness of the proposed approach.

***Keywords*** — Multivariate stratified surveys, Optimum allocation, Compromise allocation, Stochastic programming, Multiobjective fuzzy chance constraint programming, Goal programming technique

## 1. INTRODUCTION

Variability or heterogeneity is inherent in the population units. The problem of obtaining a good estimator for population mean or total one should attempt either to increase the sample size or to devise certain method of sampling by which variability or heterogeneity can be reduced. One such method is stratified sampling. It consists in dividing the population into subpopulations called strata and each subpopulation as stratum. The problem of allocating the appropriate sample size to the respective stratum is known as the problem of optimum allocation.

Yates (1960) suggested two useful approaches for optimal allocation. One approach is to "Minimize the variances of the estimates subject to a cost function or to a given sample size" and another approach is to "minimize the total cost for the desired precision of the estimates".

In large scale sample surveys one is generally concerned with estimation of more than one population characteristic. If these characteristics are highly correlated then optimal allocation may differ little among themselves, Swain (2003). But if not so, then optimal allocation for one character may not be optimal for others. In such situations we obtain a compromise allocation, optimal for all characters in some sense.

The problem of optimum allocation of sample sizes to various strata is treated as mathematical programming problem by Kokan (1663) and a solution is proposed using nonlinear programming technique. Kokan and Khan (1967) have given an analytical solution for the optimum allocation in multivariate surveys. Many authors such as Chatterjee (1967, 1968), Ahsan and Khan (1977), Khan *et al.* (1997), Khan and Ahsan (2003) either suggested new approaches or explored exiting approaches further.

As we know that in real life situations we face uncertainty. Uncertainty arises due to lack of knowledge or due to inherent vagueness. To make decisions in such type of situations, we use probability theory and fuzzy set theory respectively. Many problems in statistics such as regression analysis, sample surveys, cluster analysis, estimation and so on can be viewed as a mathematical programming problem, Arthanari and Dodge (1981). The mathematical programming problems in which some or all of the parameters are described by stochastic (or random or probabilistic) variables are called stochastic programming or probabilistic programming problems, (see Dantzig (1955), Rao (1978), Prekopa (1995), Uryasev and

* Corresponding author's email: shamsherstats@gmail.com

Pardalos (2001)). There are two important techniques for solving stochastic programming namely two stage programming and chance constraint programming, Charnce and Cooper (1959). Chance constraint programming is a particular case of stochastic programming in which each constraint has a specified probability of being satisfied.

In stratified sampling, a lot of work has been done by applying stochastic programming. Díaz-García and Garay-Tapia (2007) worked out on optimum allocation in stratified sampling by treating the problem as a stochastic programming problem in which sample variances are considered as random variables with asymptotic normal distribution when population variances were unknown.

During the course of survey, per unit measurement cost may, however vary and affected by random causes so it can be considered as random variable. In this regard, Bakhshi *et al.* (2010) considered the measurement cost $c_h$ as normally distributed random variable and formulate the problem of optimum allocation as to minimize the variances of the estimates for a probabilistic cost constraint. Some authors also used stochastic programming for optimal allocation such as Javaid *et al.* (2011), Ali *et al.* (2011), and Díaz-García and Ramos-Quiroga (2012). Khan and Khalid (2013) used multi choice for precision in multivariate stratified sampling and obtained compromise allocation.

The measurement cost includes labour cost, travel expenses (including meals and lodging, if out of station) and will increase with the complexity of the survey and amount of data to be carried out. In multivariate survey, the cost of measurement of an individual may vary highly form one character to another (such as biological investigations), Ahsan and Khan (1982). There are some other factors which rarely affect the cost very much such as weather condition, nature of climate, flood etc., which occur randomly. In such situations the cost of approaching a unit in any stratum will be completely uncertain and estimated form the sampled data. In the presence of such factors and due to the complexity of the survey, the probability (cost is high) will be high. Then the random variable representing "high" is a fuzzy random variable. Thus the cost for each stratum which is estimated form the sampled data is a random variable as well as a fuzzy number i.e. the cost is a fuzzy random variable (FRV). The concept of fuzzy set theory was initially introduced by Zadeh (1965). After Zadeh, the concept of fuzzy random variable discussed by Kwakernaak (1978, 1979), Puri and Ralesku (1986), Liu and Liu (2003) according to different requirement of measurability.

Many authors have been used fuzzy random variables according to their requirements. Studies on linear programming problems with fuzzy random variable coefficients, called fuzzy random linear programming problems, were initiated by Wang and Qiao (1993) and Qiao *et al.* (1994). Optimization models of fuzzy stochastic programming were found in the paper by Luhandjula (2006). In this work, we are considering the cost as an independent normally distributed random variable with fuzzy mean and variance and the probability of satisfying the cost constraint is also fuzzy probability.

Garcia *et al.* (2001) considered the problem of estimating the expected value of fuzzy random variables in the stratified random sampling form finite populations but the use of fuzzy random variable to obtain the optimum allocation in stratified sampling is not being seen yet in literature.

Thus in this paper, we are working on the problem of optimal allocation of sample size to various strata in multivariate stratified surveys to minimize the variances of the estimates of population parameters for each characteristic to a given cost constraint. Our objective is to optimize all the objectives simultaneously, so this problem can be considered as a multiobjective programming problem. The cost constraint will be a fuzzy chance constraint as the cost is a fuzzy random variable (FRV) and the satisfying probability of constraint is also a fuzzy number. All the objectives are nonlinear (as variance function is nonlinear in sample size $n_h$) and we required integer solutions (sample size in fraction have no sense), so the problem of optimal allocation is formulated as an all integers nonlinear Multiobjective Fuzzy Chance Constraint Programming problem (MOFCCPP). A MOFCCPP can be defined as a mathematical programming problem in which we want to optimize more than one objective simultaneously with a fuzzy chance constraint. For solving this problem, first we transformed it into its deterministic equivalent and then obtained the compromise solutions by applying goal programming technique. A numerical illustration is also given to demonstrate the proposed approach. Remainder of this paper is organized in the following sections. Section 2 contains some preliminaries concepts on fuzzy random variable (FRV). Some notations related to multivariate stratified sampling are given in section 3. In section 4, mathematical formulation of the problem of optimal allocation in fuzzy environment is presented and its deterministic equivalent is obtained in section 5. Section 6 contains a solution procedure by goal programming technique and a numerical example is also presented in section 7 to justify the correctness of the given approach. In section 8, sensitivity analysis of the numerical example is conducted with discussion about the obtained results. In the last section, some concluding remarks with future research direction related to this work are presented.

## 2. PRELIMINARIES

From the various definitions of the concept of fuzzy number we choose following definitions:

**Definition 1. (Zadeh, 1965)** *Let* $X$ *be the classical crisp set of objects called the universe whose generic elements are denoted by* $x$ . $A$

*fuzzy set* $A$ *in* $X$ *is a function* $A : X \to [0,1]$. *The set* $[0,1]$ *is called a valuation set. The fuzzy set* $A$ *is characterized by the set of all pairs of points denoted by*

$$A = \left\{x, \mu_A(x)\right\}, \quad x \in X,$$

*where* $\mu_A(x)$ *is called the membership function of* $x$ *in* $A$. *The close the value of* $\mu_A(x)$ *is to* $1$, *the more* $x$ *belongs to* $A$.

**Definition 2. (Bector and Chandra, 2005)** *Let* $A$ *is a fuzzy set in* $X$. *Then the support of* $A$, *denoted by* $S(A)$, *is the crisp set given by*

$$S(A) = \left\{x \in X : \mu_A(x) > 0\right\}.$$

**Definition 3. (Normal Fuzzy Set)** *Let* $A$ *be a fuzzy set in* $X$, *the height* $h(A)$ *of* $A$ *is defined as*

$$h(A) = \sup_{x \in X} \mu_A(x).$$

*If* $h(A) = 1$, *then the fuzzy set* $A$ *is called a normal fuzzy set, otherwise it is called subnormal.*

**Definition 4. (Bector and Chandra, 2005)** *Let* $A$ *be a fuzzy set in* $X$ *and* $\alpha \in (0,1]$. *The* $\alpha$*-cut of the fuzzy set* $A$ *is the crisp set* $A_\alpha = \{x \in X : \mu_A(x) \geq \alpha\}$.

**Definition 5. (Bector and Chandra, 2005)** *A fuzzy set* $A$ *in* $R$ *is called a fuzzy number if it satisfies the following conditions*

*(i)* $A$ *is normal*

*(ii)* $A_\alpha$ *is closed interval for every* $\alpha \in (0,1]$,

*(iii) the support of* $A$ *is bounded.*

**Definition 6. (Equipossible fuzzy number)** *A fuzzy number* $A$ *is called an equipossible fuzzy number and denoted by* $A = (a_l, a_u)$ *if its membership function is given by*

$$\mu_A(x) = \begin{cases} 1 \ if \ a_l \leq x \leq a_u \\ 0 \ otherwise \end{cases}$$

**Definition 7. (Triangular fuzzy number)** *A fuzzy number* $A$ *is called a triangular fuzzy number and denoted by* $A = (a_l, a, a_u)$ *if its membership function is given by*

$$\mu_A(x) = \begin{cases} \dfrac{x - a_l}{a - a_l} \ if \ a_l \leq x \leq a \\ \dfrac{x - a_u}{a - a_u} \ if \ a \leq x \leq a_u \\ 0 \qquad otherwise \end{cases}$$

**Definition 8. (Fuzzy Probability)** *Let* $S = (x_1, x_2, x_3, ...., x_n)$ *be a discrete finite set and* $P$ *be the probability function defined on* $\Omega$ *which is the set of all subsets of* $S$ *with*

$$P\{(x_i)\} = p_i, \quad 0 \leq p_i \leq 1, \quad \forall\, i, \sum_{i=1}^{n} p_i = 1.$$

*Then* $\Omega$ *together with* $P$ *is said to be a fuzzy probability space if at least one of these* $p_i$ *is a fuzzy number.*

**Definition 9. (Fuzzy Random Variable)** *After Zadeh (1965), Kwakernaak (1978, 1979) suggested the approach of fuzzy random variables. Let* $(\Omega, A, P)$ *be the probability space, and* $X$ *be a random variable on* $(\Omega, A, P)$ *with probability density function* $f(x)$. *Fuzzy* *random variable* $X$ *is a mapping from* $\Re$ *to a family of fuzzy numbers. They can be viewed as an extension of any random variable where the set of its values is viewed not as real number but as fuzzy numbers.*

**Definition 10. (Fuzzy Normal Distribution)** (Buckley and Eslami, 2004). *The normal density* $N(\mu,\sigma^2)$ *has density function* $f(x:\mu,\sigma^2),\ x\in R$ *with mean* $\mu$ *and variance* $\sigma^2$. *So consider the fuzzy normal* $N(\tilde{\mu},\tilde{\sigma}^2)$ *for fuzzy numbers* $\tilde{\mu}$ *and* $\tilde{\sigma}^2>0$. *We wish to compute the fuzzy probability of obtaining a value in the interval* $[c,d]$. *We write this fuzzy probability as* $\tilde{P}[c,d]$. *For* $\alpha\in[0,1]$, $\mu\in\tilde{\mu}[\alpha]$ *and* $\sigma^2\in\tilde{\sigma}^2[\alpha]$.

Let $z_1=\dfrac{(c-u)}{\sigma}$ and $z_2=\dfrac{(d-u)}{\sigma}$, then

$$\tilde{P}[c,d][\alpha]=\left\{\int_{z_1}^{z_2} f(x;0,1)dx \mid \mu\in\tilde{\mu}[\alpha],\sigma^2\in\tilde{\sigma}^2[\alpha]\right\},\ \text{for}\ \ 0\le\alpha\le 1.$$

The above equation gets the $\alpha$-cuts of $\tilde{P}[c,d]$. Also, in the above equation $f(x;0,1)$ stands for the standard normal density with zero mean and unit variance. Let $\tilde{P}[c,d][\alpha]=[p_1[\alpha],p_2[\alpha]]$; then the minimum (maximum) of the expression on the right side of the above equation is $p_1[\alpha]_*, p_2[\alpha]^*$.

Let the fuzzy mean is $\tilde{M}$, then its $\alpha$-cuts are

$$\tilde{M}[\alpha]=\left\{\int_{-\infty}^{\infty} x\, f(x:\mu,\sigma^2)dx \mid \mu\in\tilde{\mu}[\alpha],\ \sigma^2\in\tilde{\sigma}^2[\alpha]\right\}=\left[m_*[\alpha],m^*[\alpha]\right].$$

Let the fuzzy variance is $\tilde{V}$, then its $\alpha$-cuts are

$$\tilde{V}[\alpha]=\left\{\int_{-\infty}^{\infty} (x-\mu)^2 f(x:\mu,\sigma^2)dx \mid \mu\in\tilde{\mu}[\alpha],\ \sigma^2\in\tilde{\sigma}^2[\alpha]\right\}=\left[V_*[\alpha],V^*[\alpha]\right].$$

We have two important results on fuzzy means and variances from Nanda *et al.* (2006). Let $\tilde{X}$ and $\tilde{Y}$ be two fuzzy random variables with means $m_x(\tilde{\theta}),m_y(\tilde{\theta})$ and variances $\sigma_x^2(\tilde{\theta}),\ \sigma_y^2(\tilde{\theta})$ respectively then $\tilde{Z}=a\tilde{X}+b\tilde{Y}$, a linear combination of the *FRVs* $\tilde{X}$ and $\tilde{Y}$ will also be a fuzzy random variable (*FRV)*, whose mean and variance are fuzzy numbers denoted by $m_z(\tilde{\theta})$ and $\sigma_z^2(\tilde{\theta})$, respectively. $m_z(\tilde{\theta})=\ a m_x(\tilde{\theta})+b m_y(\tilde{\theta})$ with $\alpha$-cuts,

$$m_z(\tilde{\theta})[\alpha]=\ [m_{z^*}(\theta)[\alpha],m_z^*(\theta)[\alpha]\ ],$$

and $\sigma_z^2(\tilde{\theta})=a^2\sigma_x^2(\tilde{\theta})+b^2\,\sigma_y^2(\tilde{\theta})$ with $\alpha$-cuts,

$$\sigma_z^2(\tilde{\theta})=[\sigma_{z*}^2(\theta)[\alpha],\sigma_z^{*2}(\theta)[\alpha]].$$

**Definition 11. (Partial Order Relation $\preceq$)** We are using the partial order relation by Nanda *et al.* (2006). Let $\tilde{A}$ and $\tilde{B}$ be two fuzzy numbers with $\alpha$-cuts $\tilde{A}[\alpha]=[a_*[\alpha],a^*[\alpha]]$ and $\tilde{B}[\alpha]=[b_*[\alpha],b^*[\alpha]]$ respectively, then $\tilde{A}\preceq\tilde{B}$ iff $a^*[\alpha]\le b_*[\alpha]$ for each $\alpha$.

## 3. NOTATIONS

In stratified sampling the population having $N$ units is divided into $L$ subpopulations having $N_1,N_2,N_3,...,N_L$ units respectively (symbols have their usual meaning from the standard book Cochran (1977), otherwise stated). Let the suffix $h$ denotes the stratum and $i$ the unit within the stratum. Also let,

| | |
|---|---|
| $N_h$ | total number of units in $h^{th}$ stratum |
| $n_h$ | number of units in sample from $h^{th}$ stratum |
| $y_{hi}$ | value obtained for the $i^{th}$ unit in the $h^{th}$ stratum |

$W_h = \dfrac{N_h}{N}$ stratum weight

$f_h = \dfrac{n_h}{N_h}$ sampling fraction in the $h^{th}$ stratum

$\overline{Y}_h = \dfrac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$ true mean

$\overline{y}_h = \dfrac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$ sample mean

$S_h^2 = \dfrac{\sum_{i=1}^{N_h} (y_{hi} - \overline{Y}_h)^2}{N_h - 1}$ true variance.

## 4. MATHEMATICAL FORMULATION OF THE PROBLEM OF OPTIMAL ALLOCATION IN UNCERTAIN ENVIRONMENT

In stratified sampling the values of the sample sizes $n_h$ (sample size from $h^{th}$ stratum) in the respective strata are chosen by the sampler. They may be selected to minimize $V(\overline{y}_{st})$ for a specified cost of taking the sample or to minimize the cost for a specified value of $V(\overline{y}_{st})$.

$\overline{y}_{st}$ is the estimate used in stratified sampling for population mean per unit and its variance is given as

$$V(\overline{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \frac{W_h^2 S_h^2}{N_h} = \sum_{h=1}^{L} \left( \frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2$$

Arthanari and Dodge (1981) have given the mathematical formulation of the above problem for minimizing the sum of variances subject to budget restriction (with overhead cost) as follows

$$\begin{aligned} \min \quad & V(\overline{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} \\ s.t. \quad & \sum_{h=1}^{L} c_h n_h + c_0 = C \\ & 1 \le n_h \le N_h,\ n_h\text{: integers for } h = 1,2,3,\ldots,L; \end{aligned} \tag{1}$$

where $S_h^2\,(h = 1,2,3,\ldots,L)$ are the true population variances, and term independent of $n_h$ has been ignored here.

In multivariate surveys, we have more than one character under study say $p$ characters $(p \ge 2)$, then for a fixed budget $C$, the problem of optimal allocation as a multiobjective programming (MOP) problem can be represented as follows

$$\begin{aligned} \min \quad & V(\overline{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} \text{ for each character } j = 1,2,3,\ldots,p \\ s.t. \quad & \sum_{h=1}^{L} c_h n_h + c_0 \le C \\ & 2 \le n_h \le N_h,\ n_h\text{: integers for } h = 1,2,3,\ldots,L; \end{aligned} \tag{2}$$

where $S_{jh}^2$ are the true population variances for $j^{th}$ characteristic and $c_h = \sum_{j=1}^{p} c_{jh}$ denotes the cost of measuring all the $p$ characters on a sampled unit from the $h^{th}$ stratum and $c_{jh}$ is the per unit cost of measuring the $j^{th}$ characteristic in $h^{th}$ stratum. The restrictions $2 \le n_h \le N_h$ for $h = 1,2,3,\ldots,L$ are introduced here to avoid the problem of oversampling.

Bakhshi *et al.* (2010) have formulated the problem (2) as a stochastic programming problem by considering the measurement cost $c_h$ as normally distributed random variable and used the chance constraint programming technique to transformed it into deterministic equivalence. Mathematically this problem can be stated as

$$\begin{aligned} \min \quad & V(\overline{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} \qquad (j = 1,2,3,\ldots,p) \\ s.t. \quad & P\left(\sum_{h=1}^{L} c_h n_h + c_0 \le C\right) \ge P_0 \\ & 2 \le n_h \le N_h,\ n_h\text{: integers for } h = 1,2,3,\ldots,L; \end{aligned} \tag{3}$$

where the cost $c_h$ is a normally distributed random variable and $P_0$ is constraint satisfying probability.

As we considered that, the cost $c_h$ is a fuzzy random variable (FRV) i.e. normally distributed random variable with approximated mean and variance (defined in section 2) and the probability $P_0$ is also a fuzzy number, so the above problem (3), with these assumptions will be

$$\begin{aligned} \min \quad & V(\overline{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} \qquad (j = 1,2,3,\ldots,p) \\ s.t. \quad & \tilde{P}\left(\sum_{h=1}^{L} \tilde{c}_h n_h + c_0 \le C\right) \succeq \tilde{P}_0, \\ & 2 \le n_h \le N_h,\ n_h\text{: integers for } h = 1,2,3,\ldots,L; \end{aligned} \tag{4}$$

where '~' indicates the fuzziness of the measurement cost and satisfying probability also.

Let $V_j\ \forall j = 1,2,3,\ldots p$ ; denotes the variance function then, problem (4) can be represented as a multiobjective fuzzy chance constrained programming problem (MOFCCPP) as follows

$$\begin{aligned} \min \quad & \left\{V_1, V_2, V_3, \ldots, V_p\right\} \\ s.t. \quad & \tilde{P}\left(\sum_{h=1}^{L} \tilde{c}_h n_h \le C_0\right) \succeq \tilde{P}_0, \\ & 2 \le n_h \le N_h,\ n_h\text{: integers for } h = 1,2,3,\ldots,L; \end{aligned} \tag{5}$$

where $C_0 = C - c_0$ is the budget not included overhead cost.

## 5. DETERMINISTIC EQUIVALENT

Since, uncertain optimization problems can't be optimized directly rather than their deterministic equivalent are obtained. Uncertainty may be in the form of fuzziness or randomness. We are considering both type of uncertainty under one roof in the form of FRV. A great deal of study has been done by obtaining the deterministic equivalent of the problems containing FRVs. Some authors have been worked out with FRVs in inventory control such as Dutta *et al.* (2005, 2007), Bag *et al.* (2009) and obtained the deterministic equivalent by graded mean integration representation of fuzzy numbers proposed by Chen and Hsieh (1999). Chakraborty (2002) has been defined the chance constraint programming in fuzzy environment where the parameters are random variables, but the probability of the constraint is imprecise. In our problem, we have fuzzy chance constraint in which parameters are FRVs and the probability is a fuzzy number. Nanda *et al.* (2006) used Buckley (2003) approach of fuzzy probability to convert the fuzzy chance constraint programming to its deterministic equivalent. The aim of this work is to solve the problem of optimal allocation in the presence of FRVs, not to develop any procedure/method to obtain deterministic equivalent of fuzzy chance constraint so we are adopting Nanda approach for the deterministic equivalent of the constraint of problem (5) in the following manner.

We have considered here that $\tilde{c}_h\,(h = 1,2,3,\ldots,L)$ to be normally distributed FRVs whose mean and variance are fuzzy numbers denoted by $\tilde{m}_h$, $\tilde{\sigma}_h^{\,2}$ (both are triangular fuzzy numbers) i.e. $\tilde{m}_h = (a_{lh}, a, a_{uh})$ and $\tilde{\sigma}_h^{\,2} = (b_{lh}, b, b_{uh})$, cost constraint satisfying probability is also a fuzzy number i.e. $\tilde{P}_0 = (p_l, p, p_u)$ .

Let $u_i = \sum_{h=1}^{L} \tilde{c}_h n_h$ is a linear combination of $L$ number of fuzzy random variables then $u_i$ will also be a fuzzy random variable with mean and variance are fuzzy numbers denoted by $\tilde{m}_{u_i}$ and $\tilde{\sigma}_{u_i}^{\,2}$ respectively.

Let

$$\tilde{m}_{u_i}[\alpha] = \sum_{h=1}^{L} n_h \tilde{m}_h = [m_{u_{i*}}[\alpha], m_{u_i}^{\,*}[\alpha]]$$

are the $\alpha$-cuts of mean,

$$\tilde{\sigma}_{u_i}^{\ 2}[\alpha] = \sum_{h=1}^{L} n_h^2\, \tilde{\sigma}_h^2 = [\sigma_{u_i*}^{\ 2}[\alpha], \sigma_{u_i}^{\ *2}[\alpha]]$$

are the $\alpha$-cuts of variance and

$$\tilde{p}_0[\alpha] = [p_{0*}[\alpha], p_0^{\ *}[\alpha]]$$

are the $\alpha$-cuts of probability, where

$$m_{u_{i*}}[\alpha] = [(n_1 a_1 + n_2 a_2 + \ldots + n_L a_L) - (n_1 a_{l1} + n_2 a_{l2} + \ldots + n_L a_{lL})]\alpha + (n_1 a_{l1} + n_2 a_{l2} + \ldots + n_L a_{lL}),$$

$$m_{u_i}^{\ *}[\alpha] = -[(n_1 a_{u1} + n_2 a_{u2} + \ldots + n_L a_{uL}) - (n_1 a_1 + n_2 a_2 + \ldots + n_L a_L)]\alpha + (n_1 a_{u1} + n_2 a_{u2} + \ldots + n_L a_{uL}) \tag{6}$$

$$\sigma_{u_i*}^{\ 2}[\alpha] = [(n_1^2 b_1 + n_2^2 b_2 + \ldots + n_L^2 b_L) - (n_1^2 b_{l1} + n_2^2 b_{l2} + \ldots + n_L^2 b_{lL})]\alpha + (n_1^2 b_{l1} + n_2^2 b_{l2} + \ldots + n_L^2 b_{lL}),$$

$$\sigma_{u_i}^{\ *2}[\alpha] = -[(n_1^2 b_{u1} + n_2^2 b_{u2} + \ldots + n_L^2 b_{uL}) - (n_1^2 b_1 + n_2^2 b_2 + \ldots + n_L^2 b_L)]\alpha + (n_1^2 b_{u1} + n_2^2 b_{u2} + \ldots + n_L^2 b_{uL}) \tag{7}$$

and

$$\begin{cases} p_{0*}[\alpha] = (p - p_l)\alpha + p_l \\ p_0^{\ *}[\alpha] = -(p_u - p)\alpha + p_u \end{cases}. \tag{8}$$

Now, the deterministic equivalent of the fuzzy chance constraint will be

$$F\left(\frac{C_0 - m_{u_i}^{\ *}[\alpha]}{\sigma_{u_i}^{\ *2}[\alpha]}\right) \ge p_0^{\ *}[\alpha]$$

for each $\alpha \in [0,1]$, where $F$ is the cumulative distribution function of $N(0,1)$ distribution.

Thus the problem of optimal allocation to be worked out will be

$$\begin{aligned} \min \quad & [V_1, V_2, V_3, \ldots, V_p] \\ s.t. \quad & F\left(\frac{C_0 - m_{u_i}^{\ *}[\alpha]}{\sigma_{u_i}^{\ *2}[\alpha]}\right) \ge p_0^{\ *}[\alpha] \ \text{ for each } \alpha \in [0,1] \\ & 2 \le n_h \le N_h,\ n_h\text{: integers for } h = 1, 2, 3, \ldots, L; \end{aligned} \tag{9}$$

where $m_{u_i}^{\ *}[\alpha]$, $\sigma_{u_i}^{\ *2}[\alpha]$, $p_0^{\ *}[\alpha]$ are given by (6), (7) and (8) respectively.

## 6. SOLUTION BY GOAL PROGRAMMING TECHNIQUE

The problem (9) is a multiobjective programming problem and we want to optimize all the objectives simultaneously. A single solution may not be optimal for all the objectives due to the conflict nature of objectives, so goal programming technique is taken into account to obtain compromise solution. There are various methods for solving multiobjective programming problem to obtain the compromise solutions such as global criterion method, weighted sum method, lexicographic ordering and goal programming etc., Miettinen (1998). The global criterion method is suitable for the situations where the decision maker does not have any specific expectations of the solution. In weighted sum method, decision maker specifies a weighting vector representing his/her preference information. In lexicographic, decision maker makes arrangement of objective functions according to their absolute importance. The problems arise when we have no idea about the preference and importance of objectives. In goal programming, our expectations about the objectives are taken into account and treated as goal. The purpose of sampling theory is to spring up the methods of sample selection and of estimation, that provide, at the lowest possible cost, estimates that are precise enough for our purpose. It is the guiding principle for optimal allocation. In order to apply this principle, we must be able to predict, for any sampling procedure that is under consideration, the precision and the cost to be expected (see Cochran (1977), page 9). In our case, we sustain a fixed total cost/budget, so the expected precision is set for each characteristic and treated as goal. This is the reason for adopting goal programming. Goal programming technique is one of the most popular techniques developed by Charnes *et al.* (1955), but the term goal programming was fixed in Charnes and Cooper (1961). In goal programming technique the decision maker fixed his (her) aspiration levels for each of the objectives and the deviations from these aspiration levels are minimized. Some authors have been used goal programming technique to solve the problem of optimum allocation such as Khan *et al.* (2010), Ansari *et al.* (2011). We are also proceeding in the similar manner.

Let $n_{jh}^*$ be the optimal sample size in the $h^{th}$ stratum for $j^{th}$ characteristic and $V_{jh}^*$ is the variance corresponding to $n_{jh}^*$. These variances are our expected/required precision for each characteristic. Also let $n_h$ be the size of the sample obtained by using the compromise allocation and $V^0_{jh}$ is the value of variance corresponding to $n_h$.

Obviously,

$$V^0_{jh} \geq V^*_{jh} \quad \text{or} \quad V^0_{jh} - V^*_{jh} \geq 0 .$$

A reasonable compromise criterion may be to find $n_h$ that minimize,

$$\sum_{j=1}^{p} \left(V^0_{jh} - V^*_{jh}\right) = \sum_{j=1}^{p} \left(V^0_{j} - V^*_{j}\right) \ ;$$

where $V^{\circ}_{jh} = V^{\circ}_{j}$ and $V^*_{jh} = V^*_{j}$ are used.

Our goal will be "Minimize the sum of the increases in the variances due to use of compromise allocation, i.e. for $j^{th}$ character the increase in the variance should not exceed $x_j$ ".

To achieve this goal $n_h$ must satisfy

$$V^0_j - V^*_j \leq x_j ;$$

or

$$V^0_j - \ x_j \leq V^*_j ;$$

or

$$\sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - x_j \leq V^*_j ; \ \ j = 1,2,3,\ldots p \ .$$

So we can be express the problem (9) as a goal programming problem as follows

$$\begin{aligned}
\min \quad & Z = \sum_{j=1}^{p} x_j \\
s.t. \quad & \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} - x_j \leq V^*_j, \ j = 1,2,3,\ldots,p \\
& F\left(\frac{C_0 - m_{u_i}{}^*[\alpha]}{\sigma_{u_i}{}^{*2}[\alpha]}\right) \geq p_0{}^*[\alpha] \ \text{ for each } \alpha \in [0,1] \\
& 2 \leq n_h \leq N_h, \ n_h\text{: integers for } h = 1,2,3,\ldots,L;
\end{aligned} \tag{10}$$

This problem can be solved by LINGO (13.0), which is a friendly user software package without any deeply study of mathematical programming. For more details about the software Lingo user's guide (2011) is available.

## 7. NUMERICAL ILLUSTRATION

For the purpose of numerical illustration we considered the data collected from a stratified random sample survey conducted in Varanasi district of Uttar Pradesh (UP), India, to study the distribution of manurial resources among different crops and cultural practices Sukhatme *et al.* (1984). Relevant data with respect to the two characteristics '*area under rice*' and '*total cultivated area*' are given in Table1. The total number of villages in the district was 4190.

To demonstrate our approach, we assume that total budget available for conducting the survey is $C = 1800$ units with overhead cost $c_0 = 300$ units i.e. $C - c_0 = 1500 = C_0$.

We also assume that $c_h$ as a normally distributed fuzzy random variable with mean and variances as fuzzy numbers (both are triangular) i.e. $\tilde{c}_h \sim N(\tilde{m}_h, \tilde{\sigma}^2_h)$.

Let $\tilde{m}_1 = (2,3,4)$ , $\tilde{m}_2 = (1,2,3)$ , $\tilde{m}_3 = (2,3,4)$ , $\tilde{m}_4 = (1,2,3)$ and $\tilde{\sigma}^2_1 = (0.25, 0.30, 0.35)$ , $\tilde{\sigma}^2_2 = (0.20, 0.25, 0.30)$ , $\tilde{\sigma}^2_3 = (0.25, 0.30, 0.35)$ , $\tilde{\sigma}^2_4 = (0.20, 0.25, 0.30)$ are the mean and variances of the cost associated with respective stratum, and the satisfying probability of cost constraint is also a fuzzy number i.e. $\tilde{P}_0 = (0.85, 0.90, 0.95)$.

**Table 1: Data of the Survey**

| Stratum $h$ | $N_i$ | $W_i$ | $S_{i1}^2$ | $S_{i2}^2$ |
|---|---|---|---|---|
| 1 | 1419 | 0.3387 | 4817.72 | 130121.15 |
| 2 | 619 | 0.1477 | 6251.26 | 7613.52 |
| 3 | 1253 | 0.2990 | 3066.16 | 1456.40 |
| 4 | 899 | 0.2146 | 56207.25 | 66977.72 |

The $\alpha$-cuts for mean, variance and probability will be

$$[(2n_1+n_2+2n_3+n_4)+\alpha(n_1+n_2+n_3+n_4),(4n_1+3n_2+4n_3+3n_4)-\alpha(n_1+n_2+n_3+n_4)],$$

$$[(0.25n_1^2+0.20n_2^2+0.25n_3^2+0.20n_4^2)+\alpha(0.5n_1^2+0.5n_2^2+0.5n_3^2+0.5n_4^2),$$

$$(0.35n_1^2+0.30n_2^2+0.35n_3^2+0.30n_4^2)-\alpha(0.5n_1^2+0.5n_2^2+0.5n_3^2+0.5n_4^2)],$$

and

$$[0.5\alpha+0.85, 0.95-0.5\alpha]$$

respectively. Let $V_1^*$ and $V_2^*$ are the expected variances for the estimates of character '*area under rice*' and '*total cultivated area*' respectively. We calculate these variances by taken into account only one objective at a time, for the given cost constraint. At $\alpha = 0.4$, we solve the following problems

$$\begin{aligned} \min \quad & \left(\frac{552.640}{n_1}+\frac{136.277}{n_2}+\frac{274.114}{n_3}+\frac{2588.343}{n_4}\right) \\ s.t. \quad & 3.6n_1+2.6n_2+3.6n_3+2.6n_4+1.48\sqrt{(0.33n_1^2+0.28n_2^2+0.33n_3^2+0.28n_4^2)} \le 1500 \\ & 2\le n_1\le 1419, 2\le n_2\le 619, 2\le n_3\le 1253, 2\le n_4\le 899 \end{aligned}$$

and

$$\begin{aligned} \min \quad & \left(\frac{14926.197}{n_1}+\frac{165.39747}{n_2}+\frac{130.202}{n_3}+\frac{3048.324}{n_4}\right) \\ s.t. \quad & 3.6n_1+2.6n_2+3.6n_3+2.6n_4+1.48\sqrt{(0.33n_1^2+0.28n_2^2+0.33n_3^2+0.28n_4^2)} \le 1500 \\ & 2\le n_1\le 1419, 2\le n_2\le 619, 2\le n_3\le 1253, 2\le n_4\le 899 \end{aligned}$$

for $V_1^*$ and $V_2^*$ respectively. After solving by LINGO (13.0), we get $V_1^* = 24.17$ and $V_2^* = 103.07$. These are our required goals for the variances of estimates.

The problem of optimal allocation to be solved as a goal programming problem will be

$$\begin{aligned} \min \quad & Z = x_1 + x_2 \\ s.t. \quad & \left(\frac{552.640}{n_1}+\frac{136.277}{n_2}+\frac{274.114}{n_3}+\frac{2588.343}{n_4}\right) - x_1 \le 24.17 \\ & \left(\frac{14926.197}{n_1}+\frac{165.39747}{n_2}+\frac{130.202}{n_3}+\frac{3048.324}{n_4}\right) - x_2 \le 103.07 \\ & 3.6n_1+2.6n_2+3.6n_3+2.6n_4+1.48\sqrt{(0.33n_1^2+0.28n_2^2+0.33n_3^2+0.28n_4^2)} \le 1500 \\ & 2\le n_1\le 1419, 2\le n_2\le 619, 2\le n_3\le 1253, 2\le n_4\le 899 \end{aligned}$$

By solving the above problem by LINGO (13.0) package, we obtained the following compromise allocation

$$n_1 = 197;\ n_2 = 35;\ n_3 = 34 \text{ and } n_4 = 143$$

with a total of $409$. The value of $Z$ is $11.5127$ with $x_1 = 8.6913$ and $x_2 = 2.8213$. Corresponding to this allocation the values of variances for two characters are $V_1 = 32.86$ and $V_2 = 105.89$. Also, the total expanse for conducting the survey is $1499.78$ units.

**Table 2. Results by increment in the value of $C_0$**

| $C_0$ | $Z$ | $x_1$ | $x_2$ | $V_1$ | $V_2$ | Total Sample Size |
|---|---|---|---|---|---|---|
| 1500 | 11.5127 | 8.6913 | 2.8213 | 32.86 | 105.89 | 409 |
| 1575 | 5.4575 | 5.3676 | 0.0899 | 29.54 | 103.16 | 433 |
| 1650 | 2.0409 | 2.0252 | 0.0157 | 26.19 | 103.09 | 459 |
| 1725 | 0.0000 | 0.0000 | 0.0000 | 24.15 | 102.85 | 482 |

**Table 3. Results by increment in the value of $\alpha$**

| $\alpha$ | $Z$ | $x_1$ | $x_2$ | $V_1$ | $V_2$ | Total Sample Size |
|---|---|---|---|---|---|---|
| 0.4 | 11.5127 | 8.6913 | 2.8213 | 32.86 | 105.89 | 409 |
| 0.5 | 7.105 | 7.03 | 0.0738 | 31.201 | 103.14 | 424 |
| 0.6 | 3.97 | 3.97 | 0.0000 | 28.14 | 103.07 | 443 |
| 0.7 | 1.66 | 1.66 | 0.0000 | 25.83 | 103.06 | 463 |
| 0.8 | 0.00 | 0.00 | 0.0000 | 24.156 | 102.81 | 484 |

## 8. SENSITIVITY ANALYSIS AND DISCUSSION

The impact of total cost $C_0$ and $\alpha$ on the value of objective function and variances of the estimates is discussed here. The values of $C_0$ and $\alpha$ are increased and obtained results are summarized in the Table 2 and 3. On the basis of results summarized in table 2 and 3, we observe that $C_0$ and $\alpha$ have a reasonable impact on the objective function. The impact of $C_0$ is evident from the fact that a $5\%$ increment in $C_0$ is resulting approximately $50\%$ decrement in the value of objective function and at $C_0 = 1725$, we attain our goals. Similarly, we can observe the impact of $\alpha$ on the value of objective function. A gradually increment in $\alpha$ minimizes the value of objective function and at $\alpha = 0.8$, we attain our required goals. On the other hand, we can state that an increment in $C_0$ and $\alpha$ is increasing the total size of sample allocated to various strata and reducing the variances. This statement can be couple with the principle of sampling viz. the sampling error usually decreases with the increment in sample size and we obtained more precise results.

## 9. CONCLUSION

The main feature of this paper is the consideration of randomness and fuzziness under one roof in the cost function. The problem of optimal allocation in multivariate stratified sampling is considered as multiobjective fuzzy chance constraint programming (MOFCCP) problem. By solving the deterministic equivalent by goal programming technique with a numerical illustration, it can be realized that the proposed approach gives an appropriate compromise solution. A sensitivity analysis is conducted to understand the relationship between cost of survey and variances of the estimates. In this work, we considered a linear cost function and worked out with multivariate stratified sampling. This work can be extended with nonlinear cost function for other sampling designs and may be considered as the future work direction of the present work.

## ACKNOWLEDGEMENTS

The authors are highly grateful to the area editor and anonymous referees for their comments and valuable suggestions that improved this manuscript in its present form.

## REFERENCES

1. Ahsan, M.J. and Khan, S.U. (1977). Optimum Allocation in Multivariate Stratified Random Sampling Using Prior Information. *Journal of Indian Statistical Association*, 15:57–67.
2. Ahsan, M.J. and Khan, S.U. (1982). Optimum Allocation in Multivariate Stratified Random Sampling with Overhead Cost. *Metrika*, 29:71–78.
3. Ali, I., Raghav, Y.S., Bari, A. (2011).Compromise allocation in multivariate stratified surveys with stochastic quadratic cost function. *Journal of Statistical Computation and Simulation*, 83(5):962-976.
4. Ansari, A. H., Varshney, R., Najmussehar, Ahsan, M.J. (2011). An optimum multivariate multiobjective stratified sampling design. *METRON*, 69(3): 227-250.
5. Arthanari, T.S. and Dodge, Y. (1981). *Mathematical Programming in Statistics*. New York: John Wiley.
6. Bakhshi, Z.H., Khan, M.F. and Ahmad, Q.S. (2010). Optimal sample numbers in multivariate stratified sampling with a probabilistic cost constraint. *International Journal of Mathematics and Applied Statistics*, 1(2):111–120.
7. Bag, S, Chakraborty, D., Roy, A. R. (2009). A production inventory model with fuzzy random demand and with flexibility and reliability consideration, *Computers & Industrial Engineering*, 56(1):411-416.
8. Bector, C. R. and Chandra, S. (2005). *Fuzzy Mathematical Programming and Fuzzy Matrix Games* © Springer-Verlag Berlin Heidelberg Printed in Germany.
9. Buckley, J. J. (2003). *Fuzzy Probabilities. New approach and applications.* Physica-Verlag Heidelberg.
10. Buckley, J. J. and Eslami, E. (2004). Uncertain probabilities-II. *Soft computing,* 8:193–199.
11. Charnes A. and W.W. Cooper (1959). Chance constrained programming. *Management Science*, 5:73–79.
12. Charnes A., Cooper W.W., Ferguson R. (1955). Optimal Estimation of Executive Compensation by Linear Programming. *Management Science*, 1:138-151.
13. Charnes, A. and Cooper W.W. (1961). *Management Models and Industrial Applications of Linear Programming, Vols. 1& 2,* John Wiley & Sons, New York.
14. Chatterjee, S. (1967). A Note on Optimum Allocation. *Skandinavisk Aktuarietidskrift*, 50:40–44.
15. Chatterjee, S. (1968). Multivariate Stratified Surveys. *Journal of American Statistical Association*, 63:530–534.
16. Chakraborty, D. (2002). Redefining chance-constrained programming in fuzzy environment, *Fuzzy Sets & Systems*, 25 (3): 327-333.
17. Chen, S.H. and Hsieh, C.H. (1999). Graded mean integration representation of generalized fuzzy number. *Journal of Chinese Fuzzy Systems Association*, 5: 1-7.
18. Cochran, W.G. (1977). *Sampling Techniques (Third Edition).* John Wiley and Sons, New York.
19. Dantzig, G.B. (1955). Linear programming under uncertainty, *Management Science,* 1 (3&4):197–206.
20. Díaz-García, J. A., Ramos-Quiroga, R. (2012). Stochastic optimal design in multivariate stratified sampling. *Optimization: A Journal of Mathematical Programming and Operations Research*, 61(1):1-24.
21. Díaz-García, J.A. and Garay-Tapia, Ma. M. (2007). Optimum allocation in stratified surveys: Stochastic programming. *Computational Statistics & Data Analysis,* 51: 3016-3026.
22. Dutta, P., Chakraborty, D., Roy, A. R. (2005). A single period inventory model with fuzzy random variable demand. *Mathematical and Computer Modelling*, 41: 915-922.
23. Dutta, P., Chakraborty, D., Roy, A. R. (2007). An inventory model for single-period products with reordering opportunities under fuzzy demand. *Computers and Mathematics with Applications*, 53: 1502-1517.
24. Garcia, D., Lubiano, M. A., Alonso, M. C. (2001). Estimating the expected value of fuzzy random variables in stratified random sampling form finite populations. *Information Sciences,* 138:165-184.
25. Javaid, S., Bakhshi Z. H., Khalid, M.M. (2011). Use of stochastic programming for sample allocation in two-stage stratified sampling. *International Journal of Agricultural and Statistical Sciences*, 7(2): 379-392.
26. Khan, M.G.M. and Ahsan, M.J. (2003). A Note on Optimum Allocation in Multivariate Stratified Sampling. *The South Pacific Journal of Natural Science,* 21: 91–95.
27. Khan, M.G.M., Ahsan, M.J., and Jahan, N. (1997). Compromise Allocation in Multivariate Stratified Sampling: An Integer Solution. *Naval Research Logistic,* 44:69–79.
28. Khan, M.G.M., Maiti T. and Ahsan M.J. (2010). An Optimal Multivariate Stratified Sampling Design Using Auxiliary Information: An Integer Solution Using Goal Programming Approach. *Journal of Official Statistics*, 26: 695–708.
29. Khan, S., Khalid, M.M. (2013). Multi Choice for Precision in Multivariate Stratified Surveys: A Compromise Solution. *International Journal of Operations Research*, 10(4):171-181.
30. Kokan, A. R. (1963). Optimum allocation in multivariate surveys. *Journal of Royal Statistical Society*, A, 126: 557-565.
31. Kokan, A.R. and Khan, S.U. (1967).Optimum Allocation in Multivariate Surveys: An Analytical Solution. *Journal of Royal Statistical Society*, *B (Statistical Methodology)*, 29:115–125.
32. Kwakernaak, H. (1978). Fuzzy random variable-I. *Information Sciences*, 15:1–29.
33. Kwakernaak, H. (1979). "Fuzzy random variable-II: Algorithm and Examples for the discrete case," *Information Sciences*, 17: 253–278.

34. LINGO (2011). User's Guide, Lindo Systems Inc., Chicago, IL, USA.
35. Liu, Y. K. and Liu, B. (2003). Fuzzy Random Variables: A scalar expected value operator. *Fuzzy Optimization and Decision Making*, 2(2): 143-160.
36. Luhandjula, M.K. (2006). Fuzzy stochastic linear programming: survey and future research directions. *European Journal of Operations Research*, 174:1353–1367.
37. Miettinen, K. M. (1998). *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers.
38. Nanda S., Panda G. and Dash J. K. (2006). A new solution method for fuzzy chance constrained programming problem. *Fuzzy Optimization and Decision Making* 5:355–370.DOI 10.1007/s10700-006-0018-8.
39. Prekopa, A., (1995). *Stochastic Programming*. Kluwer Academic Publishers (Series Mathematics and its Applications), Dordrecht.
40. Puri, M. L. And Ralescu, D. A. (1984). Fuzzy Random Variables. *Journal of Mathematical Analysis and Applications*, 114(2): 409-422.
41. Qiao, Z., Zhang Y., and Wang, G.-Y. (1994). On fuzzy random linear programming. *Fuzzy Sets and Systems*, 65:31–49.
42. Rao, S.S. (1978): *Optimization: Theory and Applications*, John Wiley & Sons, New York.
43. Sukhatme, P.V., Sukhatme, B.V., S. Sukhatme, and Ashok, C. (1984). *Sampling Theory of Surveys with Applications*, 3rd ed., Iowa State University Press and Indian Society of Agricultural Statistics, Ames, IW and New Delhi.
44. Swain, A.K.P.C. (2003). *Finite Population Sampling: Theory and Methods*. South Asian Publishers Pvt. Ltd., New Delhi.
45. Uryasev, S. and Pardalos, P. M. (2001): *Stochastic Optimization*. Kluwer Academic Publishers, Dordrecht.
46. Wang, G.-Y., and Qiao, Z. (1993). Linear programming with fuzzy random variable coefficients. *Fuzzy Sets and Systems*, 57:295–311.
47. Yates, F. (1960). *Sampling Methods for Census and Surveys (Second Edition)*. London: Charles Griffin and Co. Ltd.
48. Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8: 338-353.