

OPTIMUM ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING: MULTI-OBJECTIVE PROGRAMMING

José Antonio Díaz García & Liliana Ulloa Cortez

Comunicación Técnica No I-06-07/28-03-2006
(PE/CIMAT)



Optimum allocation in multivariate stratified sampling: Multi-objective Programming

José A. Díaz-García

*Universidad Autónoma Agraria Antonio Narro
Department of Statistics and Computation
25315 Buenavista, Saltillo
Coahuila, México*

Liliana Ulloa Cortez

*Universidad Autónoma Agraria Antonio Narro
Department of Statistics and Computation
25315 Buenavista, Saltillo
Coahuila, México*

Abstract

This paper considers optimum allocation in multivariate stratified sampling as a problem of the multi-objective optimisation of integers, under three different scenarios, those of complete, partial or zero information. The paper concludes with an example showing the implementation of each of the techniques proposed.

Key words: Multivariate stratified sampling, optimum allocation, multi-objective optimisation, value function method, lexicographic method, ε -constraint method, distance-based method.

PACS: 62E15.

1 Introducción

One of the areas of statistics that is most commonly used in all fields of scientific investigation is that of probabilistic sampling. Obtaining good results in medical, social or other research depends on a successful, well-implemented sampling process. The use of effective sampling techniques within a population

Email addresses: jadiaz@uaaan.mx (José A. Díaz-García), lullcor@uaaan.mx (Liliana Ulloa Cortez).

is thus seen as the obtention of data that is useful for gaining knowledge of important aspects of such a population. Stratified sampling is one of the most frequently utilised of these techniques. In order to determine the sample size within each stratum of a stratified sampling scheme, one possibility is that of optimum allocation. This is achieved by means of a non-linear optimisation problem, in which the objective function is the variance, which is subject to a cost restriction, or vice versa. Traditionally, this problem has been resolved by using the Cauchy-Schwarz (Stuart 1954) inequality cited in Cochran (1977) or the Multiplier method of Lagrange, see Sukhatme *et al.* (1984).

The usual way these sampling techniques are used, and in particular that of stratified sampling, is the univariate form, that is, when the size of the sample and its allocation into strata is proposed taking into account a single decision variable or characteristic, see Cochran (1977), Sukhatme *et al.* (1984) and Thompson (1997). In the context of stratified sampling, attempts have been made to establish the sample size and its allocation within strata, taking into account various characteristics, see Sukhatme *et al.* (1984) and Arthanari and Dodge (1981), among others.

When optimum allocation is performed, and a cost function that is subject to the restrictions of the variances in the different characteristics is proposed as the objective function, the problem is then reduced to one of classical mathematical programming, and has been treated as such by Arthanari and Dodge (1981), from a deterministic standpoint, and by Prékopa (1978) from a stochastic position. In this latter case, the problem can be approached using any of the techniques presented in Díaz-García and Garay (2006).

Alternatively, when the goal is to minimise the variances that are subject to a cost function, or to a given sample size, then the problem has been resolved in various ways, see Sukhatme *et al.* (1984). But as we shall see, all these approaches previously described in the literature are only particular cases of one of the techniques of multi-objective optimisation, which in the sampling context has a prime criterion that the population being studied must be totally identified, to the degree that it is possible to propose a scalar function establishing a relation between the variances of each characteristic. Such conditions, in practice, are rarely encountered. Furthermore, the above approaches do not safeguard against the problem of over-sampling, i.e. when the sample size in one or more strata is larger than the stratum size; furthermore, the sample sizes obtained are not integers, and must be rounded.

On the basis of the above considerations, in this paper we examine the problem of optimum allocation in multivariate stratified sampling, while simultaneously minimising the variances subject to a cost function or to a given sample size, as a problem of the multi-objective optimisation of integers. We study different techniques to resolve this problem, taking into account the prior knowledge

of the population, which is classified as complete, partial or zero information. Finally, all the techniques studied are applied to a standard problem, the solutions being obtained by means of the LINGO computer program.

2 Multivariate stratified sampling

When a sample is obtained, a typical problem that arises is that of estimating various characteristics of the population. This is usually complicated by the fact that the different characteristics may have different variances, which means that the sample sizes for each characteristic may vary. To formally set out the problem of the optimum allocation in stratified sampling, consider the following notation, see Cochran (1977), Sukhatme *et al.* (1984) and Thompson (1997).

Notation

The subindex $h = 1, 2, \dots, H$ denotes the stratum, and $i = 1, 2, \dots, N_h$ the unit within stratum h . Moreover:

N_h	Total number of units within stratum h
n_h	Number of units from the sample in stratum h
y_{hi}	Value obtained for the i -th unit in stratum h
$\mathbf{n} = (n_1, n_2, \dots, n_H)'$	Vector of the number of units in the sample
$W_h = \frac{N_h}{N}$	Relative size of stratum h
$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$	Population mean in stratum h
$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$	Sample mean in stratum h
$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}$	Populational variance in stratum h

c_h Cost per sampling unit in stratum h

$\bar{y}_{st} = \sum_{h=1}^H W_h \bar{y}_h$ Estimator of the populational mean
in the stratified sampling.

$V(\bar{y}_{st})$ Variance of \bar{y}_{st} .

where

$$V(\bar{y}_{st}) = \sum_{h=1}^H \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^H \frac{W_h S_h^2}{N}$$

Formally, the problem of optimum allocation in stratified sampling can be presented as the following programme of multi-objective, non-linear optimisation:

$$\begin{aligned} \min_{\mathbf{n}} \hat{V}(\bar{\mathbf{y}}_{st}) = \min_{\mathbf{n}} & \begin{pmatrix} \hat{V}(\bar{y}_{st}^1) \\ \vdots \\ \hat{V}(\bar{y}_{st}^G) \end{pmatrix} \\ & \text{subject to} \\ & \mathbf{c}'\mathbf{n} + c_0 = C \end{aligned} \quad (1)$$

where C is the total cost, c_0 is a fixed cost and $\mathbf{c}' = (c_1, \dots, c_H)$.

Although this problem has not been presented exactly so in the statistical literature, different approaches have been put forward to resolve the problem of optimum allocation in multivariate stratified sampling, including compromise allocation, compromise allocation minimising the total relative loss, and compromise allocation taking the mean of the values. Another approach is based on minimising the generalised variance, while yet another seeks to minimise the trace of the variance and covariance matrix, always subject to a predetermined cost function or sample size. These methods, and some additional ones, have been examined in detail by Sukhatme *et al.* (1984).

Note that the solutions proposed for programme (1) take real values, and thus the sample sizes n_h must be integers. We must also address the problem of over-sampling, that is, when $n_h \geq N_h$ for at least some h ; finally, there is the problem of estimating the variance on the basis of the sample size in each stratum. In order to overcome these three complications, we propose the

following, as an alternative to (1).

$$\begin{aligned}
\min_{\mathbf{n}} \widehat{V}(\bar{\mathbf{y}}_{st}) = & \min_{\mathbf{n}} \begin{pmatrix} \widehat{V}(\bar{y}_{st}^1) \\ \vdots \\ \widehat{V}(\bar{y}_{st}^G) \end{pmatrix} \\
& \text{subject to} \\
& \mathbf{c}'\mathbf{n} + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N},
\end{aligned} \tag{2}$$

where \mathbb{N} denotes the set of natural numbers.

The methods for resolving a multi-objective optimisation programme can be classified by considering the amount of information available to the programme, in the context of the sampling exercise; in practice, this concept is taken as the amount of information possessed concerning the study population, with three different scenarios, namely complete, partial or zero information, see Ríos, Ríos Insua and Ríos Insua (1989), Miettinen (1999) and Steuer (1986), among others. Let us now consider the problem (2) from the standpoint of each of the multi-objective optimisation methods, using the following classification:

$$\begin{array}{l}
\text{Multi-Objective} \\
\text{Optimisation Methods}
\end{array}
\left\{ \begin{array}{l}
\text{Complete information} \\
\text{Partial information} \\
\text{Zero information}
\end{array} \right.
\left\{ \begin{array}{l}
\text{Value function} \\
\text{Lexicographic} \\
\varepsilon - \text{constraint} \\
\text{Distances.}
\end{array} \right.$$

Optimum allocation via multi-objective optimisation

Note that $V(\bar{y}_{st}^j)$ is defined using the populational variances S_h^2 , $h = 1, 2, \dots, H$, which are usually unknown, and therefore these are substituted by the sample variances s_h^2 , $h = 1, 2, \dots, H$, defined as

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2.$$

And thus $V(\bar{y}_{st}^j)$ is substituted by the estimated variance $\widehat{V}(\bar{y}_{st}^j)$, which is given by

$$\widehat{V}(\bar{y}_{st}^j) = \sum_{h=1}^H \frac{W_h^2 s_{hj}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{hj}^2}{N}.$$

Value function

This method is applicable in experiments in which the information on the characteristics to be evaluated is complete, i.e. when the importance of each one is fully known. For example, when using the value function with the weighting method, its hierarchy must be so perfectly understood that the evaluator is able to assign an appropriate weight to each characteristic, see Ríos, Ríos Insua and Ríos Insua (1989), Miettinen (1999) and Steuer (1986), among others.

Under the value function technique, programme (2) is expressed as follows:

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{v}(\widehat{V}(\bar{\mathbf{y}}_{st})), \\ & \text{subject to} \\ & \sum_{h=1}^H c_h n_h + c_0 = C \quad (3) \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}. \end{aligned}$$

where $v(\cdot)$ is a scalar function that summarises the importance of each of the variances of the G characteristics.

Clearly, all the approaches described in the literature constitute particular cases of the above method.

Evidently, for every problem the value function $v(\cdot)$ may take an infinite number of forms, and this is what constitutes the difficulty for the evaluator in defining such a function. However, some simple functions have given excellent results in the applications, providing the evaluator with a relatively straightforward task. One of these particular forms is the weighting method. Under

this approach, problem (3) can be expressed as:

$$\begin{aligned} \min_{\mathbf{n}} \quad & \sum_{j=1}^G \lambda_j \widehat{V}(\bar{y}_{st}^j), \\ \text{subject to} \quad & \\ & \sum_{h=1}^H c_h n_h + c_0 = C \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}. \end{aligned}$$

such that $\sum_{j=1}^G \lambda_j = 1$, $\lambda_j \geq 0 \quad \forall \quad j = 1, 2, \dots, G$; where λ_j weights the importance of each characteristic.

In the context of multi-objective optimisation, this is without doubt the method that has been most thoroughly studied. Its popularity is due to the fact that the *value function* is not unique, and thus the investigator can select the function that is most suitable for the occasion. Furthermore, this method can be applied in experiments in which there are antecedents that help us to weight the characteristics to be evaluated; in other words, the value function method is utilised for recurrent studies in which, over time, the results obtained help us reach a better inference for future experiments, in which the appropriate weighting can be applied.

Lexicographic method

This method, like the previous one, requires complete information on the phenomenon in order to create an importance-ordered hierarchy of the characteristics evaluated, these being measured in this case by means of their variances. Unlike the value function method, it is not necessary to know what weight to allocate to each characteristic, but only the order of importance they represent in obtaining the sample, see Ríos, Ríos Insua and Ríos Insua (1989), Steuer (1986), among others. In practice, this is very useful, as on various occasions the evaluator will not know the value of the weight of each characteristic to be evaluated, but only the order in which each one affects the study.

In this case, to optimise programme (2), the evaluator must order the variances, beginning with the one presenting the most important characteristics, and then by descending order of importance, thus obtaining

$$\widehat{V}(\bar{y}_{st}^{i_1}), \widehat{V}(\bar{y}_{st}^{i_2}), \dots, \widehat{V}(\bar{y}_{st}^{i_G}),$$

where i_1, \dots, i_G is a permutation with the desired, descending order of the

set of superindices $1, 2, \dots, G$. Now, it is necessary to resolve the following programme:

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^{i_1}) \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned} \tag{4}$$

If the minimum of problem (4) is v_1 , in the next stage we must resolve the problem

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^{i_2}) \\
& \text{subject to} \\
& \sum_{h=1}^H \frac{W_h^2 s_{h1}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{h1}^2}{N} = v_1 \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned} \tag{5}$$

Now, let v_2 be the minimum of problem (5), and in the third stage we resolve the problem

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^{i_3}) \\
& \text{subject to} \\
& \sum_{h=1}^H \frac{W_h^2 s_{h1}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{h1}^2}{N} = v_1 \\
& \sum_{h=1}^H \frac{W_h^2 s_{h2}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{h2}^2}{N} = v_2 \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned}$$

Thus, we reach stage G , where the next problem to be solved is

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^{i_G}) \\
& \text{subject to} \\
& \sum_{h=1}^H \frac{W_h^2 s_{h1}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{h1}^2}{N} = v_1 \\
& \sum_{h=1}^H \frac{W_h^2 s_{h2}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{h2}^2}{N} = v_2 \\
& \sum_{h=1}^H \frac{W_h^2 s_{h3}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{h3}^2}{N} = v_3 \\
& \quad \vdots \\
& \sum_{h=1}^H \frac{W_h^2 s_{hG-1}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{hG-1}^2}{N} = v_{G-1} \\
& \quad \sum_{h=1}^H c_h n_h + c_0 = C \\
& \quad 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& \quad n_h \in \mathbb{N}.
\end{aligned}$$

Thus, the vector obtained in this stage is the optimum solution to the problem.

ε -constraint method

This is a method to resolve the problem when only partial information is available. In order to apply this method, the investigator need only identify the most important characteristic, see Ríos, Ríos Insua and Ríos Insua (1989) and Miettinen (1999). This method is extremely useful in studies in which the investigator has been able to identify the characteristic that has most influence on the obtention of the sample, and the limits to be attributed to the other characteristics.

Let us again start from problem (2), and assume that the most important characteristic in the study is the k -th one, $k \in \{1, 2, \dots, G\}$. Under this technique,

the problem can then be restated as follows:

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^k) \\
& \text{subject to} \\
& \widehat{V}(\bar{y}_{st}^r) \leq v_r, \quad r \neq k, \quad r = 1, 2, \dots, G \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N},
\end{aligned} \tag{6}$$

where, in this case, v_r is a pre-established bound for each of the $G-1$ remaining variances, which are given as constraints.

For all practical purposes, these v_r values can be taken as the upper limit of the confidence interval for each variance or, alternatively, they can be defined as the minimum individual values of the following problems:

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^r) \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}, \quad r = 1, 2, \dots, G, \quad r \neq k.
\end{aligned}$$

Note that the choice of the k characteristic and the lower limits v_r represent the evaluator's subjective preferences, and so if there were no solution to problem (6), this would mean that the v_r limits had been set too low and that at least one must be revised, see Ríos, Ríos Insua and Ríos Insua (1989).

Distance-based method

On many occasions, the investigator comes up against the problem that no antecedents are available with which to address it, or otherwise it might be difficult to decide which of the characteristics being evaluated is the most important. In such cases, the method presented in this section is the most suitable. No antecedent is required, as the only requirement for solving a problem using this method is a vector of *ideal* goals, which is determined with the null information expressed in the problem, see Ríos, Ríos Insua and Ríos Insua (1989) and Steuer (1986).

Then, from problem (2), with this method it is possible to obtain the optimum values, minimising the distance between the optimum and the vector of targets, simultaneously.

Let v_j be the ideal point or goal for the objective $\widehat{V}(\bar{y}_{st}^j)$, $j = 1, \dots, G$, i.e. the vector of targets \mathbb{V} is given as

$$\mathbb{V} = \begin{pmatrix} v_1 \\ \vdots \\ v_G \end{pmatrix}.$$

A great advantage of this method is that this vector of targets \mathbb{V} can be calculated without additional information. This is done by minimising, separately, each objective $\widehat{V}(\bar{y}_{st}^j)$, $j = 1, \dots, G$, such that the vector \mathbb{V} is defined as the vector of its individual minima, which is achieved on resolving the following G non-linear minimisation programmes for integers, see Rao (1978):

$$\begin{aligned} & \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^j), \\ & \text{subject to} \\ & \sum_{h=1}^H c_h n_h + c_0 = C, \\ & 2 \leq n_h \leq N_h, \\ & h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}. \end{aligned}$$

for $j = 1, \dots, G$.

When the vector \mathbb{V} has been established, we proceed to examine the problem to be optimised with the new objective function, namely

$$\begin{aligned} & \min_{\mathbf{n}} d(V(\bar{y}_{st}^j), \mathbb{V}) \\ & \text{subject to} \\ & \sum_{h=1}^H c_h n_h + c_0 = C \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}. \end{aligned}$$

where d corresponds to a weighted norm. More generally, we will consider the weighted norm L_q -, and so the problem to be optimised takes the following

form:

$$\begin{aligned} \min_{\mathbf{n}} \left[\sum_{j=1}^G \lambda_j |V(\bar{y}_{st}^j) - v_j|^q \right]^{\frac{1}{q}} \\ \text{subject to} \\ \sum_{h=1}^H c_h n_h + c_0 = C \\ 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ n_h \in \mathbb{N}, \end{aligned}$$

with $1 \leq q \leq \infty$ and $\lambda_j \geq 0$, which is the weight or priority given to each objective j . To illustrate this technique, let us take $\lambda_j = 1$, that is, we shall give the same priority to each characteristic, and we shall use the particular cases $q = 1$, $q = 2$ and $q = \infty$.

Thus, with $q = 1$, we have the following problem:

$$\begin{aligned} \min_{\mathbf{n}} \left[\sum_{j=1}^G |V(\bar{y}_{st}^j) - v_j| \right] \\ \text{subject to} \\ \sum_{h=1}^H c_h n_h + c_0 = C \\ 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ n_h \in \mathbb{N}. \end{aligned}$$

As v_j are constant for all $j = 1, \dots, G$, the problem is reduced to

$$\begin{aligned} \min_{\mathbf{n}} \sum_{j=1}^G V(\bar{y}_{st}^j) \\ \text{subject to} \\ \sum_{h=1}^H c_h n_h + c_0 = C \\ 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ n_h \in \mathbb{N}. \end{aligned}$$

With $q = \infty$, we need only take into account the maximum deviation, and so

the problem to be optimised will be

$$\begin{aligned}
& \min_{\mathbf{n}} \max_{j=1,2,\dots,G} [V(\bar{y}_{st}^j) - v_j] \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned}$$

and for $q = 2$ the problem is

$$\begin{aligned}
& \min_{\mathbf{n}} \left[\sum_{j=1}^G [V(\bar{y}_{st}^j) - v_j]^2 \right]^{1/2} \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned}$$

Alternatively, another distance has been proposed by Khuri and Cornell (1987):

$$\begin{aligned}
& \min_{\mathbf{n}} \sum_{j=1}^G \left[\frac{(V(\bar{y}_{st}^j) - v_j)^2}{v_j^2} \right] \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned}$$

Note that with all these optimisation methods, we have utilised the cost restriction $\sum_{h=1}^H c_h n_h + c_0 = C$. However, on some occasions, the restrictions do not apply to the costs but rather to the availability of man-hours for carrying out a survey, or simply to the total time available for performing the survey. These limitations can be described using the following expression, see Arthanari and

Dodge (1981):

$$\sum_{h=1}^H n_h = n.$$

3 Alternative approach

Note that in problem (2) it is assumed that the covariances between the different characteristics are zero, which under the assumption of normality is equivalent to assuming the characteristics to be stochastically independent, a fact that is not necessarily so in the applications used. This idea led us to propose the problem of optimum allocation under multivariate stratified sampling as follows:

$$\begin{aligned} & \min_{\mathbf{n}} \Theta \\ & \text{subject to} \\ & \sum_{h=1}^H c_h n_h + c_0 = C \quad (7) \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}. \end{aligned}$$

where $\Theta = \text{Cov}(\bar{\mathbf{y}}_{st})$ is the matrix of variances - covariances of the vector $\bar{\mathbf{y}}_{st} = (\bar{y}_{st}^1, \dots, \bar{y}_{st}^G)'$.

Obviamente la dificultad de plantear de esta forma el problema es como definir que significa el mínimo de una matriz. Sin embargo, si dicho problema puede ser tratado a través del método de la función de valor, fácilmente surgen interpretaciones del mínimo de una matriz. De esta forma, el programa (7) empleando la técnica de la función de valor está dado por

Obviously, the difficulty of expressing the problem in this way lies in defining the meaning of the minimum of a matrix. However, if this problem can be dealt with using the value function method, then it is easy to arrive at interpretations of the minimum of a matrix. Thus, programme (7), using the value

function technique, is given by

$$\begin{aligned}
& \min_{\mathbf{n}} \mathbf{v}(\Theta) \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned}$$

Note that, in particular $\mathbf{v}(\Theta) = \text{tr}(\Theta)$, and all the characteristics are given the same weighting, such that $\sum_{j=1}^G \lambda_j = 1, \quad \lambda_j \geq 0 \quad \forall \quad j = 1, 2, \dots, G$, then we obtain the particular solution described in the section on the value function. Similarly, as a particular case, too, we obtain the solution described by Dalenius, for which we simply define $\mathbf{v}(\Theta) = \det(\Theta) = |\Theta|$, to obtain the following programme under approach (2):

$$\begin{aligned}
& \min_{\mathbf{n}} |\Theta| \\
& \text{subject to} \\
& \sum_{h=1}^H c_h n_h + c_0 = C \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\
& n_h \in \mathbb{N}.
\end{aligned}$$

Note that the trajectory and the determinant are not the only value functions that can be used. Alternative definitions of the value function that have been used in other statistical contexts include:

- (1) The sum of all the elements of the matrix $\Theta = (\theta_{kl})$; $\mathbf{v}(\Theta) = \sum_{k,l=1}^G \theta_{kl}$.
- (2) $\mathbf{v}(\Theta) = \lambda_1(\Theta)$, where λ_1 is the maximum eigenvalue of the matrix of covariances Θ .
- (3) $\mathbf{v}(\Theta) = \lambda_G(\Theta)$, where λ_G is the minimum eigenvalue of the matrix of covariances Θ .

4 Example

Let us now use the data from the example described by Sukhatme *et al.* (1984, p. 164) to calculate sample sizes by means of the five methods examined above.

The results obtained with the LINGO software, for all the techniques, are shown in Table 1.

For this example, the multi-objective optimisation programme under approach (2) is

$$\begin{aligned}
& \min_{\mathbf{n}} \begin{pmatrix} \widehat{V}(\bar{y}_{st}^1) \\ \widehat{V}(\bar{y}_{st}^2) \end{pmatrix} \\
& \text{subject to} \\
& \sum_{h=1}^4 n_h = 382 \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\
& n_h \in \mathbb{N}.
\end{aligned} \tag{8}$$

Furthermore, consider the following two programmes for the non-linear minimising of integers:

$$\begin{aligned}
& \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^1) & \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^2) \\
& \text{subject to} & \text{subject to} \\
& \sum_{h=1}^4 n_h = 382 & \sum_{h=1}^4 n_h = 382 \\
& 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\
& n_h \in \mathbb{N}. & n_h \in \mathbb{N},
\end{aligned} \tag{9}$$

the individual minima of which are concentrated in the following vector

$$\mathbb{V} = \begin{pmatrix} 23.86 \\ 92.50 \end{pmatrix}. \tag{10}$$

Value Function

Under this method, it is assumed that both characteristics are equally important, and so the same weighting is given to the two characteristics being

evaluated, $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$. From the above, we have

$$\begin{aligned} \min_{\mathbf{n}} & \left(0.5 \left(\widehat{V}(\bar{y}_{st}^1) \right) + 0.5 \left(\widehat{V}(\bar{y}_{st}^2) \right) \right) \\ & \text{subject to} \\ & \sum_{h=1}^4 n_h = 382 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\ & n_h \in \mathbb{N}. \end{aligned}$$

Lexicographic Method

The first step in resolving the problem by means of the lexicographic method is to order the characteristics by order of importance, assuming that $\widehat{V}(\bar{y}_{st}^1)$ is the most important. In this first stage, we must resolve the first programme in (9). Then, in the second stage, and in the present case this is the final stage, we have the following optimisation programme:

$$\begin{aligned} \min_{\mathbf{n}} & \widehat{V}(\bar{y}_{st}^2) \\ & \text{subject to} \\ & \widehat{V}(\bar{y}_{st}^1) \leq 23.86 \\ & \sum_{h=1}^4 n_h = 382 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\ & n_h \in \mathbb{N}. \end{aligned}$$

ε -constraint methods.

Let us again assume characteristic 1 to be highest in the hierarchy, programme (8), and then the problem to be resolved using this technique is

$$\begin{aligned} & \min_{\mathbf{n}} \widehat{V}(\bar{y}_{st}^1) \\ & \text{subject to} \\ & \widehat{V}(\bar{y}_{st}^2) \leq 92.50 \\ & \sum_{h=1}^4 n_h = 382 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\ & n_h \in \mathbb{N}, \end{aligned}$$

where $v_2 = 92.50$ is the upper bound for the variance of characteristic 2, and this is calculated by minimising the second problem in (9).

Distance-based Method

Starting from problem (8), we apply the procedure for the distance $q = 1$ (norm of the absolute value), $q = 2$ and the distance proposed by Khuri and Cornell (1987). Taking (10) as the vector of targets, we then have:

for $q = 1$, the distance to be minimised is

$$\begin{aligned} & \min_{\mathbf{n}} \left(|\widehat{V}(\bar{y}_{st}^1) + \widehat{V}(\bar{y}_{st}^2)| \right) \\ & \text{subject to} \\ & \sum_{h=1}^4 n_h = 382 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\ & n_h \in \mathbb{N}. \end{aligned}$$

With $q = 2$, we have

$$\begin{aligned} \min_{\mathbf{n}} & \left[\left(\widehat{V}(\bar{y}_{st}^1) - 23.86 \right)^2 + \left(\widehat{V}(\bar{y}_{st}^2) - 92.50 \right)^2 \right] \\ & \text{subject to} \\ & \sum_{h=1}^4 n_h = 382 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\ & n_h \in \mathbb{N}, \end{aligned}$$

and with the distance proposed by Khuri and Cornell (1987), the problem to be optimised is

$$\begin{aligned} \min_{\mathbf{n}} & \left[\frac{\left(\widehat{V}(\bar{y}_{st}^1) - 23.86 \right)^2}{23.86^2} + \frac{\left(\widehat{V}(\bar{y}_{st}^2) - 92.50 \right)^2}{92.50^2} \right] \\ & \text{subject to} \\ & \sum_{h=1}^4 n_h = 382 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, 3, 4 \\ & n_h \in \mathbb{N}. \end{aligned}$$

TABLE 1: Sample sizes for the different allocations calculated.

Allocation	n_1	n_2	n_3	n_4
Value Function	201	28	32	121
Lexicographic	207	27	31	117
ε -constraints	120	39	55	168
Distances $q=1$	201	28	32	121
Distances $q=2$	190	29	36	127
Khuri and Cornell distance	191	29	35	127

Note that the strata with the highest numbers of units allocated are numbers 1 and 4, irrespective of the multi-objective optimisation technique implemented. This is due to the fact that the variances of the characteristics being evaluated are larger in these strata, see Sukhatme *et al.* (1984, p. 164).

It should also be noted that all the methods are illustrated using the same

example, but in real life this does not occur, because when considering the method to be utilized it is necessary to decide whether complete, partial or zero information is available, and then apply the most appropriate method to each situation, according to the investigator's criteria and experience.

5 Conclusions

The problem of optimum allocation in multivariate stratified sampling has been examined previously in statistical literature, but the solutions proposed have been particular cases of a multi-objective optimisation technique. Furthermore, all the solutions proposed have been merely particular cases of the technique known as the value function, within the context of multi-objective optimisation. In the solution to a programme for multi-objective optimisation, there are three possible scenarios, namely that the investigator possesses no information, partial information or complete information. Unfortunately, the value function technique has been proposed within a context of complete information. To extend this idea to a sampling context would require the investigator to be perfectly informed of the study population, such that it would be possible to propose a value function reflecting the importance of every single one of the variances of the characteristics being studied, and this possibility, today, is very rarely encountered. Taking these circumstances into account, the present study examines alternative techniques within the contexts of partial information (whereby it may be sufficient to know about the most important characteristic) and of zero information (in which it is not necessary to possess any information other than the estimators of the parameter being studied), in order to reach a more appropriate solution. Having addressed the problem of optimum allocation within stratified sampling as a programme of multi-objective optimisation, the techniques proposed for the resolution of the problem are illustrated by means of an example from the standard literature in the field, namely Sukhatme *et al.* (1984). In addressing the problem of optimum allocation in multivariate sampling, it is first necessary to determine in which of the three contexts (i.e., total information, partial information or zero information) this problem resides. Once this has been done, we then decide the technique to be applied, on the basis of the information available. It is important to note that the solution for an allocation problem should be achieved by the implementation of a single method. For this reason, the results obtained for the example are comparable only within the context in which the example was established.

Bearing these considerations in mind, note that the sample sizes allocated by the *value function* and *lexicographic methods* (both in the context of complete information) vary only slightly. Similarly, it can be shown that with the sample sizes allocated by the *distance method* (zero information) for the three different

distances examined ($q = 1$, $q = 2$ and the Khuri and Cornell distance), the sample sizes for the different strata are very similar, particularly so for the $q = 2$ distance and that proposed by Khuri and Cornell, in which the strata sizes differ by only one unit in stratum 1 and in stratum 3.

It is important to stress that the alternative approach based on the philosophy put forward by Dalenius is somewhat closer to a real-world situation, as it takes into account the possible correlation between the characteristics being studied.

Finally, we can explain the recent vertiginous development of the multiobjective optimisation by the adaptation of the metaheuristics techniques, thus, we have now methods for finding the solution of the problem, or better, for determining the set of efficient solutions (in the sense of Pareto's optimisation), or for obtaining a good approximation. This wide group of alternatives provides to the researcher more freedom for taking the best solution of the particular problem.

In fact, the development of the multiobjective optimisation in the above-cited direction could be used in the context of the optimal location in the stratified survey, see Jones et al. (2002).

Acknowledgment

This research work was partially supported by CONACYT-México, research grant no. 45974-F and IDI-Spain, grant MTM2005-09209.

References

- Arthanari, T.S., and Dodge Yadolah, 1981. *Mathematical Programming in Statistics*. A Wiley-Interscience Publication John Wiley and Sons, Inc.
- Díaz-García and Garay Tapia, M. M. 2006. Optimum allocation in stratified surveys: Stochastic programming. *C. Statist. and Data Analysis*. In press. See also <http://www.sciencedirect.com>.
- Cochran, W. G. 1977. *Técnicas de muestreo*. C.E.C.S.A. México.
- Jones, D. F., Mirrazavi, S. K., and Tamiz, M. 2002. Multi-objective metaheuristics: An overview of the current state-of-the art. *European J. Operational Res.* 137(1), pp. 1-9.
- Khuri, A.I. and Cornell, J.A. 1987. *Response Surface: Designs Analysis*. Marcel Dekker, Inc. New York.

- Miettinen, K. M. 1999. Non linear multiobjective optimization. Kluwer Academic Publishers, Boston.
- Prékopa A. 1978. "The use of stochastic programming for the solution of the some problems in statistics and probability". Technical Summary report 1983. University of Wisconsin-Madison.
- Rao, S. S. 1978-79. Optimization theory and applications. Wiley Eastern Limited.
- Ríos, S., Ríos Insua, S., y Ríos Insua, M.J. 1989. Procesos de decisión Multicriterio. EUDEMA, Madrid (in Spanish).
- Steuer, R. E. 1986. Multiple criteria optimization: Theory, computation and applications. John Wiley, New York.
- Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., and Asok, C. 1984. Sampling theory of surveys with application. State University Press.
- Thompson , M.E. 1997. Theory of sample surveys. Chapman & Hall