Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Multi-objective optimisation for optimum allocation in multivariate stratified sampling

by José A. Díaz-García and Liliana Ulloa Cortez

December 2008

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2008
Vol. 34, No. 2, pp. 215-222
Statistics Canada, Catalogue No. 12-001-X

Multi-objective optimisation for optimum allocation in multivariate stratified sampling

José A. Díaz-García and Liliana Ulloa Cortez [1]

Abstract

This paper considers the optimum allocation in multivariate stratified sampling as a nonlinear matrix optimisation of integers. As a particular case, a nonlinear problem of the multi-objective optimisation of integers is studied. A full detailed example including some of proposed techniques is provided at the end of the work.

Key Words: Multivariate stratified sampling; Optimum allocation; Multi-objective optimisation.

1. Introduction

One of the areas of statistics that is most commonly used in all fields of scientific investigation is that of probabilistic sampling. An effective sampling technique within a population represents an appropriate extraction of useful data which provides meaningful knowledge of the important aspects of the population. Stratified sampling is one of the classical methods for obtaining such information. This method considers the computation of the stratum sample size, which can be computed by various procedures, but optimum allocation has been found to be a useful approach. Optimum allocation is considered as a non-linear optimisation problem in which the objective function is the variance subject to a cost restriction, or vice versa. Traditionally, this problem has been solved by using the Cauchy-Schwarz (Stuart 1954) inequality, cited in Cochran (1977) or Lagrange's multiplier method, see Sukhatme, Sukhatme, Sukhatme and Asok (1984).

Classical sampling theory considers a single decision variable or parameter; for example, in our case, univariate stratified sampling studies one parameter, the sample size and its strata allocation, see Cochran (1977), Sukhatme *et al.* (1984) and Thompson (1997). Moreover, in the context of stratified sampling, some multivariate approaches have been proposed whereby the sample size and its allocation within strata take diverse characteristics into consideration, see Sukhatme *et al.* (1984) and Arthanari and Dodge (1981), among others.

When the optimum allocation is performed, and the cost function is the objective function, subject to certain variance restrictions in the different characteristics, then the problem can be reduced to a question of classical mathematical programming, and for this purpose there are two well-known approaches: Arthanari and Dodge (1981), from a deterministic point of view; and Prékopa (1978), from a stochastic position. In the latter case, the problem can be solved by using any of the techniques presented in Díaz-García and Garay (2007).

Alternatively, if we wish to minimise the variances subject to a cost function, or to a given sample size, then sereral approaches can be adopted to solve this, see Sukhatme *et al.* (1984). However, the above-mentioned approaches do not solve the over-sampling problem, *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 approximated. Moreover, as we shall see, all the previously published approaches in this area are particular cases of the multi-objective optimisation technique. If these problems could be overcome, then we would have a formal overview and a unified theory for resolving the problem of optimum allocation in multivariate stratified sampling, and would be able to consider all the literature (both theory and practice) on multi-objective optimisation and related questions.

In this paper we study optimum allocation in multivariate stratified sampling as a nonlinear problem of matrix optimisation of integers constrained by a cost function or by a given sample size. Making certain assumptions, we propose a way to solve the problem, through several particular techniques, see subsection 3.1. The second aim of the paper is related to the following fact: if we define a particular vectorial function of the objective function of the matrix optimisation problem, then in subsection 3.2 we show that the optimum allocation in multivariate stratified sampling also can be studied as a non-linear problem of the multi-objective optimisation of integers. In subsections 3.2.1 and 3.2.2 we propose different techniques for solving these problems. Finally, in section 4, some of the techniques described are applied to a numerical example from forestry.

1. José A. Díaz-García and Liliana Ulloa Cortez, Department of Statistics and Computation, Universidad Autónoma Agraria Antonio Narro, 25350 Buenavista, Saltillo, Coahuila, MÉXICO. E-mail: jadiaz@uaaan.mx and lullcor@uaan.mx.

2. Multivariate stratified sampling

Consider a population of size N, divided into H sub-populations (strata). We wish to find a representative sample of size n and an optimum allocation in the strata meeting the following requirements: i) to minimise the variance of the estimated mean subject, to a budgetary constraint; or ii) to minimise the cost subject to a constraint on the variances; this is the classical problem in optimum allocation in univariate stratified sampling, see Cochran (1977), Sukhatme *et al*. (1984) and Thompson (1997). However, if we consider more than one characteristic (variable) then the problem is known as optimum allocation in multivariate stratified sampling. For a formal expression of the problem of optimum allocation in stratified sampling, consider the following notation.

2.1 Notation

The subindex $h = 1, 2, \ldots, H$ denotes the stratum, $i = 1, 2, \ldots, N_h$ the unit within stratum h and $j = 1, 2, \ldots, G$ denotes the characteristic (variable). Moreover:

N_h	Total number of units within stratum h
n_h	Number of units from the sample in stratum h
y_{hi}^j	Value obtained for the i^{th} unit in stratum h of the j^{th} characteristic
$\mathbf{n} = (n_1, \ldots, n_H)'$	Vector of the number of units in the sample
$W_h = \dfrac{N_h}{N}$	Relative size of stratum h
$\bar{Y}_h^j = \dfrac{\sum_{i=1}^{N_h} y_{hi}^j}{N_h}$	Population mean in stratum h of the j^{th} characteristic
$\bar{y}_h^j = \dfrac{\sum_{i=1}^{n_h} y_{hi}^j}{n_h}$	Sample mean in stratum h of the j^{th} characteristic
$\bar{\mathbf{y}}_h = (\bar{y}_h^1, \ldots, \bar{y}_h^G)'$	Sample mean vector in stratum h
$\bar{y}_{\text{ST}}^j = \sum_{h=1}^{H} W_h \bar{y}_h^j$	Estimator of the population mean in multivariate stratified sampling for the j^{th} characteristic
$\bar{\mathbf{y}}_{\text{ST}} = (\bar{y}_{\text{ST}}^1, \ldots, \bar{y}_{\text{ST}}^G)'$	Estimator of the population mean vector in multivariate stratified sampling
$s_{h_{jk}}$	Sample covariance in stratum h of the j^{th} and k^{th} characteristics, where

$$s_{h_{jk}} = \frac{\sum_{i=1}^{n_h} (y_{hi}^j - \bar{y}_h^j)(y_{hi}^k - \bar{y}_h^k)}{n_h - 1}, \text{ and}$$

$$s_{h_{jj}} = s_{hj}^2 \frac{\sum_{i=1}^{n_h} (y_{hi}^j - \bar{y}_h^j)^2}{n_h - 1}$$

$\operatorname{Cov}(\bar{\mathbf{y}}_{\text{ST}})$	Variance-covariance matrix of $\bar{\mathbf{y}}_{\text{ST}}$
$\widehat{\operatorname{Cov}}(\bar{\mathbf{y}}_{\text{ST}})$	Estimator of the variance-covariance matrix of $\bar{\mathbf{y}}_{\text{ST}}$, where $\widehat{\operatorname{Cov}}(\bar{\mathbf{y}}_{\text{ST}}) \equiv \widehat{\operatorname{Cov}(\bar{\mathbf{y}}_{\text{ST}})}$ and

$$\widehat{\operatorname{Cov}}(\bar{\mathbf{y}}_{\text{ST}}) = \begin{pmatrix} \widehat{\operatorname{Var}}(\bar{y}_{\text{ST}}^1) & \widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^1, \bar{y}_{\text{ST}}^2) & \cdots & \widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^1, \bar{y}_{\text{ST}}^G) \\ \widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^2, \bar{y}_{\text{ST}}^1) & \widehat{\operatorname{Var}}(\bar{y}_{\text{ST}}^2) & \cdots & \widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^2, \bar{y}_{\text{ST}}^G) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^G, \bar{y}_{\text{ST}}^1) & \widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^G, \bar{y}_{\text{ST}}^2) & \cdots & \widehat{\operatorname{Var}}(\bar{y}_{\text{ST}}^G) \end{pmatrix}$$

$\widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^j, \bar{y}_{\text{ST}}^k)$	Estimated covariance of $\bar{y}_{\text{ST}}^j$ and $\bar{y}_{\text{ST}}^k$ where

$$\widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^i, \bar{y}_{\text{ST}}^j) \equiv \widehat{\operatorname{Cov}(\bar{y}_{\text{ST}}^j, \bar{y}_{\text{ST}}^k)}, \quad \text{with}$$

$$\widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^j, \bar{y}_{\text{ST}}^k) = \sum_{h=1}^{H} \frac{W_h^2 s_{h_{jk}}}{n_h} - \sum_{h=1}^{H} \frac{W_h s_{h_{jk}}}{N}, \quad \text{and}$$

$$\widehat{\operatorname{Cov}}(\bar{y}_{\text{ST}}^j, \bar{y}_{\text{ST}}^j) \equiv \widehat{\operatorname{Var}}(\bar{y}_{\text{ST}}^j) = \sum_{h=1}^{H} \frac{W_h^2 s_{hj}^2}{n_h} - \sum_{h=1}^{H} \frac{W_h s_{hj}^2}{N}$$

c_h	Cost per sampling unit in stratum h

Finally, let $\mathbf{V}_u(\bar{\mathbf{y}}_{\text{ST}}) \in \Re^G$, such that $\mathbf{V}_u(\bar{\mathbf{y}}_{\text{ST}}) = (\widehat{\operatorname{Var}}(\bar{y}_{\text{ST}}^1), \ldots, \widehat{\operatorname{Var}}(\bar{y}_{\text{ST}}^G))'$, where if $\mathbf{a} \in \Re^G$, $\mathbf{a}'$ denotes the transpose of $\mathbf{a}$.

3. A new approach for the problem of optimum allocation in multivariate stratified sampling

In this section we propose optimum allocation in multivariate stratified sampling as a matrix optimisation problem, for which a number of possible solutions are studied. We observe that the multi-objective optimisation problem is a particular case of a matrix optimisation. In the same sense, we note that optimum allocation in multivariate stratified sampling can be seen as a multi-objective optimisation problem. In each case, the respective solutions are straightforwardly derived.

3.1 Matrix optimisation

Formally, optimum allocation in stratified sampling can be studied by performing the following nonlinear matrix optimisation problem:

$$\begin{aligned} &\min_{\mathbf{n}} \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}) \\ &\text{subject to} \\ &\mathbf{c}'\mathbf{n} + c_0 = C, \end{aligned} \tag{1}$$

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

Note that the solutions proposed for problem (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, see Arvanitis and Afonja (1971). In order to overcome these two complications, we propose the following alternative approach to (1).

$$\begin{aligned} &\min_{\mathbf{n}} \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}) \\ &\text{subject to} \\ &\mathbf{c}'\mathbf{n} + c_0 = C \\ &2 \leq n_h \leq N_h,\ h = 1, 2, \ldots, H \\ &n_n \in \mathbb{N}, \end{aligned} \tag{2}$$

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

Obviously, the difficulty of expressing the problem in this way lies in defining the meaning of the minimum of a matrix function. The idea of minimising a matrix function, and in particular the matrix of variance-covariance, has been studied with respect to various areas of statistical theory. For example, when the regression estimators are determined for a multivariate general linear model, this is done by minimising the determinant or the trace of sums of squares and sums of products matrix of the erro, see Giri (1977). Similarly, the choice or comparison of some experimental design models is done by minimising a function of the variance-covariance matrix of treatment estimators, see Khuri and Cornell (1987) and Azaïs and Druilhet (1997).

Fortunately, it is possible to reduce the nonlinear matrix minimisation problem (2) to a univariate nonlinear minimisation problem by taking into account the following considerations (note that the prodecure described here is just one of various possible options, see Ríos, Ríos Insua and Ríos Insua (1989) and Miettinen (1999)). Observe that $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}})$ is an explicit function of $\mathbf{n}$, and so it must be denoted as as $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}) \equiv \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}))$. Also, assume that $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}))$ is a positive definite matrix for all $\mathbf{n}$, $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n})) > \mathbf{0}$. Now, let $\mathbf{n}_1$ and $\mathbf{n}_2$ be two possible values of the vector $\mathbf{n}$ and let $\mathbf{B} = \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) - \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2))$. We say that

$$\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) < \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2)) \Leftrightarrow \mathbf{B} < \mathbf{0}, \tag{3}$$

i.e., if the matrix $\mathbf{B}$ is a negative definite matrix. Moreover, note that $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1))$ and $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2))$, are diagonalizable. Then, let $D_{\mathbf{n}_1}$ and $D_{\mathbf{n}_2}$ be the diagonal matrixes associated with $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1))$ and $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2))$, respectively, with $D_{\mathbf{n}_1} = \operatorname{diag}(\alpha_1, \ldots, \alpha_G), \alpha_1 > \cdots > \alpha_G > 0$ and $D_{\mathbf{n}_2} = \operatorname{diag}(\tau_1, \ldots, \tau_G), \tau_1 > \cdots > \tau_G > 0,$ where α_j and τ_j denote the eigenvalues of $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1))$ and $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2))$, respectively. Thus, expression (3) can alternatively be presented as:

$$\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) < \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2)) \Leftrightarrow D_{\mathbf{n}_1} - D_{\mathbf{n}_2} < \mathbf{0},$$

i.e.,

$$\begin{aligned} &\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) < \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2)) \Leftrightarrow \underset{j=1,\ldots,G}{\alpha_j - \tau_j} < 0 \\ &\text{and} \\ &\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) \neq \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2)), \end{aligned} \tag{4}$$

which defines a weak Pareto order, see Steuer (1986), Ríos *et al.* (1989) and Miettinen (1999). Then from Steuer (1986), Ríos *et al.* (1989) and Miettinen (1999), there exist a function $\mathbf{f}: S \to \Re$, such that

$$\begin{aligned} &\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) < \widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2)) \\ &\qquad \Leftrightarrow \mathbf{f}\left(\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1))\right) < \mathbf{f}\left(\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_2))\right). \end{aligned} \tag{5}$$

where $\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n})) \in S \subset \Re^{G(G+1)/2}$ and S is the set of positive definite matrixes. From (5), Steuer (1986), Ríos *et al.* (1989) and Miettinen (1999) proof that the non-linear matrix minimisation problem (2) is reduced in the following univariate non-linear minimisation problem

$$\begin{aligned} &\min_{\mathbf{n}} \mathbf{f}\left(\widehat{\operatorname{Cov}}(\overline{\mathbf{y}}_{\text{ST}})\right) \\ &\text{subject to} \\ &\sum_{h=1}^{H} c_h n_h + c_0 = C \\ &2 \leq n_h \leq N_h,\ h = 1, 2, \ldots, H \\ &n_h \in \mathbb{N}; \end{aligned} \tag{6}$$

Unfortunately or fortunately the function $\mathbf{f}(\cdot)$ is not unique. For example, in other statistical contexts we see the following commonly used functions $\mathbf{f}(\cdot)$, see Giri (1977):

1. The trace of the matrix $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))$; $\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = \mathrm{tr}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n})))$.
2. The determinant of the matrix $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))$; $\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = |\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))|$.
3. The sum of all the elements of the matrix $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))$;
$$\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = \sum_{j,k=1}^{G} \widehat{\mathrm{Cov}}(\overline{y}_{\mathrm{ST}}^{j}, \overline{y}_{\mathrm{ST}}^{k}).$$
4. $\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = \lambda_{\max}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n})))$, where $\lambda_{\max}$ is the maximum eigenvalue of the matrix $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))$.
5. $\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = \lambda_{\min}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n})))$, where $\lambda_{\min}$ is the minimum eigenvalue of the matrix $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))$.
6. $\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = \lambda_{j}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n})))$, where λ_j is the j^{th} eigenvalue of the matrix $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))$, among others.

In particular Dalenius (1957), studied the problem (6) when $\mathbf{f}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))) = |\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))|$, in other words, the minimisation of the generalised variance $|\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}}(\mathbf{n}))|$, see also Arvanitis and Afonja (1971).

3.2 Multi-objectibve optimisation

Let us now, consider the vectorial function $\mathbf{F}: S \to \Re^{G}$, such that $\mathbf{F}(\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}})) = \mathbf{V}_u(\overline{\mathbf{y}}_{\mathrm{ST}})$. An alternative way of establishing problem (2) is

$$\begin{aligned} &\min_{\mathbf{n}} \mathbf{V}_u(\overline{\mathbf{y}}_{\mathrm{ST}}) = \min_{\mathbf{n}} \begin{pmatrix} \widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{1}) \\ \vdots \\ \widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{G}) \end{pmatrix} \\ &\text{subject to} \\ &\mathbf{c}'\mathbf{n} + c_0 = C \qquad (7) \\ &2 \le n_h \le N_h,\ h = 1, 2, \ldots, H \\ &n_h \in \mathbb{N}, \end{aligned}$$

which is a nonlinear problem of the multi-objective optimisation of integers, see Steuer (1986), Ríos *et al.* (1989) and Miettinen (1999).

In the sampling context, observe that in multi-objective optimisation problems, there rarely exists a point $\mathbf{n}^*$ which is considered as an optimum, *i.e.*, few cases satisfy the requirement that $\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j}(\mathbf{n}^*))$ is minimum for all $j = 1, \ldots, G$. This justifies the following notion of the Pareto point (which is more weakly defined than an optimum point):

> *We say that* $\mathbf{V}_u^*(\overline{\mathbf{y}}_{\mathrm{ST}})$ *is a Pareto point of* $\mathbf{V}_u(\overline{\mathbf{y}}_{\mathrm{ST}})$, *if there is no other point* $\mathbf{V}_u^1(\overline{\mathbf{y}}_{\mathrm{ST}})$ such that $\mathbf{V}_u^1(\overline{\mathbf{y}}_{\mathrm{ST}}) \le \mathbf{V}_u^*(\overline{\mathbf{y}}_{\mathrm{ST}})$, *i.e.*, *for all j*, $\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j_1}) \le \widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j_*})$ *and* $\mathbf{V}_u^1(\overline{\mathbf{y}}_{\mathrm{ST}}) \neq \mathbf{V}_u^*(\overline{\mathbf{y}}_{\mathrm{ST}})$.

Existence criteria for Pareto points of a multi-objective optimisation problem are established in Ríos *et al.* (1989) and Miettinen (1999). In particular we have:

> *Given* $\mathbf{V}_u(\overline{\mathbf{y}}_{\mathrm{ST}}): \Re^H \to \Re^G$ *and let us consider a non empty compact* $\mathcal{N} \subset \mathbb{N}^H$ *such that* $\mathcal{N}$ *is the set of all possible values of* $\mathbf{n}$ *determined by the restrictions in* (7). *If* $\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j})$ *is an upper semi-continuous for each* $j = 1, \ldots, G$, *then the problem* (7) *has a Pareto optimal solution.*

On the other hand, Steuer (1986), Ríos *et al.* (1989) and Miettinen (1999) studied the extension of scalar optimisation (*Kuhn-Tucker's conditions*) to the vectorial case. In particular, they proposed necessary conditions for Pareto solutions, which become sufficient conditions if: *N* is convex; the functions $\widehat{\mathrm{Var}}(\overline{\mathbf{y}}_{\mathrm{ST}}^{j})$, $j = 1, \ldots, G$ are convex; and the Lagrange generalised multipliers δ_j, associated with each function $\widehat{\mathrm{Var}}(\overline{\mathbf{y}}_{\mathrm{ST}}^{j})$, are positive, $\delta_j > 0$ for all *j*. Note that the above results for the existence of a Pareto solution and for Kuhn-Tucker's conditions are valid when the optimisation problem is continuous, *i.e.*, when the variables n_h are continuous ones, for all $h = 1, \ldots, H$. However, it should be recalled that in order to obtain the solution to an integer optimisation problem, it is normal to make the initial assumption that such a problem is one of continuous optimisation. First we derive the solution to the problem of continuous optimisation, and then, by means of heuristic or branch-and-bound methods, progress to solving that of integer optimisation. In this context, in the case of optimising integers, in a practical case, it is sufficient to see that the corresponding problem of continuous optimisation has an optimum Pareto solution, and to confirm that the set $\mathcal{N}$ of all the possible values of $\mathbf{n}$ contains at least one $\mathbf{n} \in \mathcal{N}$ for which all the coordinates are integers.

Methods for solving a multi-objective optimisation problem are based on the information possessed about a particular problem. There are three possible scenarios; when the investigator possesses: complete, partial or null information, see Ríos *et al.* (1989), Miettinen (1999) and Steuer (1986) In a stratified sampling context, complete information means that, the investigator knows the population in such a way that it is possible to propose a value function (Value function: This is a function $\phi: \Re^H \to \Re$

such that denoting $\mathbf{V}_u(\overline{\mathbf{y}}_{\text{ST}}) \equiv \mathbf{V}_u(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}))$ we have that $\min \mathbf{V}_u(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}^*)) < \min \mathbf{V}_u(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1)) \Leftrightarrow \phi(\mathbf{V}_u(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}^*))) < \phi(\mathbf{V}_u(\overline{\mathbf{y}}_{\text{ST}}(\mathbf{n}_1))), \mathbf{n}^* \neq \mathbf{n}_1$.) reflecting the importance of each variance of the studied characteristics, this possibility, today, is very rarely encountered. In partial information, the investigator knows the main characteristic of the study very well and this is sufficient support for the research. Finally, under null information, which is the most common situation, the researcher only possesses information about the estimators of the parameter of the experiment, and with this material an appropriate solution can be found.

For reasons of space it is impossible to give an exhaustive explanation of all the techniques proposed for solving multi-objective optimisation problems (7), see Ríos *et al*. (1989), Miettinen (1999) and Steuer (1986) for a detailed description. Moreover, there are heuristic methods, instead of the classical methods, by which the problem may be addressed in an alternative way, see Jones *et al*. (2002). As an illustration, we present below a survey of two commonly used techniques; the first one studies the complete information stage (the value function, also termed the utility function) and the second one, the null information scenario (a method based on distances).

3.2.1 Value function

As mentioned above, this method belongs to the complete information case, in which the investigator is able to summarise the importance of all the studied characteristics in a real function, see the next paragraph (see also Ríos *et al*. (1989), Miettinen (1999) and Steuer (1986), among others).

Under the value function technique, problem (7) is expressed as follows:

$$
\begin{aligned}
&\min_{\mathbf{n}} \phi(\mathbf{V}_u(\overline{y}_{\text{ST}})),\\
&\text{subject to}\\
&\sum_{h=1}^{H} c_h n_h + c_0 = C \qquad (8)\\
&2 \leq n_h \leq N_h,\ h = 1, 2, \ldots, H\\
&n_h \in \mathbb{N},
\end{aligned}
$$

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

Clearly, many of the approaches described in the literature on the question of optimum allocation in multivariate stratified sampling, such as compromise assignation, compromise assignation minimising total relative loss, and compromise assignation taking the mean of the optimum values, see Sakhatme *et al.* (1984), are particular cases of the above-mentioned method.

Note that the value function $\phi(\cdot)$ may take an infinite number of forms, which represents a fundamental obstacle to defining it. However, some simple functions have given excellent results in the applications and they can be considered as promising approaches. One of these particular forms is the weighting method. Under this approach, problem (8) can be expressed as:

$$
\begin{aligned}
&\min_{\mathbf{n}} \sum_{j=1}^{G} w_j \widehat{\text{Var}}(\overline{y}_{\text{ST}}^{j}),\\
&\text{subject to}\\
&\sum_{h=1}^{H} c_h n_h + c_0 = C\\
&2 \leq n_h \leq N_h,\ h = 1, 2, \ldots, H\\
&n_h \in \mathbb{N},
\end{aligned}
$$

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

Among the multi-objective techniques we find that the value function method is, in general, the most commonly applied, because its properties have been studied with most detail, see Ríos *et al*. (1989), Miettinen (1999), Steuer (1986), and the references therein.

3.2.2 Distance-based method

Sometimes, the researcher does not have sufficient previous information about the variables, or it is difficult to decide which are the most important characteristics of the experiment. In such cases, the method of this section is very useful, because it does not need many antecedents; moreover, it only requires a vector of *ideal* goals, which is determined with the null information expressed in the problem, see Ríos *et al*. (1989) and Steuer (1986).

Then, problem (7) is solved by obtaining the optimum values via the minimisation of the distance between the optimum and the vector of targets.

Let θ_j be the ideal point or goal for the objective $\widehat{\text{Var}}(\overline{y}_{\text{ST}}^{j}),\ j = 1, \ldots, G,$ *i.e*., the vector of targets $\boldsymbol{\Theta}$ is given by

$$
\boldsymbol{\Theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_G \end{pmatrix}.
$$

Note that the vector of targets $\boldsymbol{\Theta}$ can be calculated without additional information, which is a great advantage of this method. In fact, it is computed by minimising each objective $\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j}),\ j=1,\ldots,G$ separately, such that the vector $\boldsymbol{\Theta}$ is defined as the vector of its individual minima, and this is achieved by solving the following G non-linear minimisation problems of integers, see Rao (1979):

$$
\begin{array}{l}
\min\limits_{\mathbf{n}} \widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j}),\\
\text{subject to}\\
\sum\limits_{h=1}^{H} c_h n_h + c_0 = C\\
2 \le n_h \le N_n\\
h = 1, 2, \ldots, H\\
n_h \in \mathbb{N},
\end{array}
$$

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

Once the vector $\boldsymbol{\Theta}$ has been computed, we study the optimisation problem with the new objective function, namely

$$
\begin{array}{l}
\min\limits_{\mathbf{n}} d(\mathbf{V}_u(\overline{\mathbf{y}}_{\mathrm{ST}}), \boldsymbol{\Theta})\\
\text{subject to}\\
\sum\limits_{h=1}^{H} c_h n_h + c_0 = C\\
2 \le n_h \le N_h,\ h = 1, 2, \ldots, H\\
n_h \in \mathbb{N},
\end{array}
\tag{9}
$$

where $d(\cdot,\cdot)$ denotes a general distance function. In particular, when the program (9) is applied to the Euclidean distance, we have

$$
\begin{array}{l}
\min\limits_{\mathbf{n}} \sum\limits_{j=1}^{G} \left[\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j}) - \theta_j\right]^2\\
\text{subject to}\\
\sum\limits_{h=1}^{H} c_h n_h + c_0 = C\\
2 \le n_h \le N_h,\ h = 1, 2, \ldots, H\\
n_h \in \mathbb{N}.
\end{array}
$$

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

$$
\begin{array}{l}
\min\limits_{\mathbf{n}} \sum\limits_{j=1}^{G} \dfrac{\left(\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j}) - \theta_j\right)^2}{\theta_j^2}\\
\text{subject to}\\
\sum\limits_{h=1}^{H} c_h n_h + c_0 = C\\
2 \le n_h \le N_h,\ h = 1, 2, \ldots, H\\
n_h \in \mathbb{N}.
\end{array}
$$

Remarks:

1. Note that we have used the cost restriction $\sum_{h=1}^{H} c_h n_h + c_0 = C$ in every optimisation method. However, in some situations, we do not restrict the costs but we have restrictions for the availability of man-hours for carrying out a survey, or restrictions on the total available time for performing the survey, *etc*. These limitations can be described by using the following expression, see Arthanari and Dodge (1981):
$$\sum_{h=1}^{H} n_h = n.$$
2. Note that the multi-objective optimisation methods proposed here are general and they need to be adjusted in some particular problems; for instance, we do not consider the unit (magnitude) of each variance in the respective sums for the value function. We suggest a solution, namely to replace the variance of each characteristic by its corresponding coefficient of variation
$$\sqrt{\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^{j})} / \overline{y}_{\mathrm{ST}}^{j},\ j = 1\ldots, G;$$
then, the use of Khuri and Cornell's distance is more recommendable than is the use of the Euclidean distance.

3. It is desirable to consider estimators other than a mean estimator, for example the national mean estimator, or the comparison of regional means, *etc*. In particular, the associated editor recommended estimators of the following type:
$$\widehat{\mathbf{T}} = \sum_{h=1}^{H} w_h \overline{\mathbf{y}}_h \in \Re^G \tag{10}$$
where several weights w_h could even be used for the same variable. For instance, if one of the weights w_h

is 1, another is -1, and the others are 0, then we can compute the difference between two means of two different strata. In general, we can optimise problem (2) substituting the objective function $\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}})$, by any function of interest. For example, we could use the estimated variance-covariance matrix $\widehat{\mathrm{Cov}}(\hat{\mathbf{T}})$ of the estimator (10), among many other options.

4. A numerical example

The input information was taken from Arvanitis and Afonja (1971) which is a forest survey conducted in Humbolt County, California. The population was subdivided into nine strata on the basis of the timber volume per unit area, as determined from aerial photographs. The two variables included in this example are the basal area (BA) (In forestry terminology, 'Basal area' is the area of a plant perpendicular to the longitudinal axis of a tree at 4.5 feet above ground) in square feet, and the net volume in cubic feet (Vol.), both expressed on a per acre basis. The variances, covariances and the number of units within stratum h are listed in Table 1.

Table 1
Variances, covariances and the number of units within each stratum

		Variance		
Stratum	N_h	**BA**	**Vol.**	**Covariance**
1	11,131	1,557	554,830	28,980
2	65,857	3,575	1,430,600	61,591
3	106,936	3,163	1,997,100	72,369
4	72,872	6,095	5,587,900	166,120
5	78,260	10,470	10,603,000	293,960
6	51,401	8,406	15,828,000	357,300
7	24,050	20,115	26,643,000	663,300
8	46,113	9,718	13,603,000	346,810
9	102,985	2,478	1,061,800	39,872

For this example, the matrix optimisation problem under approach (2) is

$$\min_{\mathbf{n}} \begin{pmatrix} \widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^1) & \widehat{\mathrm{Cov}}(\overline{y}_{\mathrm{ST}}^1, \overline{y}_{\mathrm{ST}}^2) \\ \widehat{\mathrm{Cov}}(\overline{y}_{\mathrm{ST}}^2, \overline{y}_{\mathrm{ST}}^1) & \widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^2) \end{pmatrix}$$

subject to

$$\sum_{h=1}^{9} n_h = 1{,}000 \tag{11}$$

$$2 \le n_h \le N_h,\ h = 1, \ldots, 9$$

$$n_h \in \mathbb{N}.$$

Table 2 shows the optimisation solutions obtained by some of the methods described in Sections 2 and 3; specifically, we present the solutions via the trace, the determinant, the value function, the Euclidean distance and the Khuri and Cornell distance. We also include the optimum allocation for each characteristic, BA and Vol. (the first two rows in Table 2). The last two columns show the minimum values of the individual variances for the respective optimum allocations identified by each method. The results were computed using the commercial software Hyper LINGO/PC, release 6.0, see Winston (1995). The default optimisation methods used by LINGO to solve the nonlinear integer optimisation programs are Generalised Reduced Gradient (GRG) and branch-and-bound methods, see Bazaraa *et al.* (2006). Some technical details of the computations are the following: the maximum number of iterations of the methods presented in Table 2 was 1,193 (determinant problem) and the mean execution time for all the programs was 1 second. Finally, note that the greatest discrepancy found by the different methods among the sizes of the strata occurred when minimising the generalised variance $|\widehat{\mathrm{Cov}}(\overline{\mathbf{y}}_{\mathrm{ST}})|$. Beyond doubt, this is because it is the only method presented in Table 2 that takes into account the covariance between the two characteristics studied.

Table 2
Sample sizes and estimator of variances for the different allocations calculated

Allocation	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	$\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^1)$	$\widehat{\mathrm{Var}}(\overline{y}_{\mathrm{ST}}^2)$
BA	10	94	144	136	191	113	81	109	122	5.591	5,441.105
Vol.	7	62	119	136	200	161	98	134	83	5.953	5,139.531
$\mathrm{tr}\left(\widehat{\mathrm{Cov}}(\overline{y}_{\mathrm{ST}})\right)$	7	62	119	135	200	161	98	134	84	5.591	5,139.531
$\lvert \widehat{\mathrm{Cov}}(\overline{y}_{\mathrm{ST}}) \rvert$	9	93	128	129	193	123	86	106	133	5.616	5,403.876
Value Function [a]	7	62	119	135	200	161	98	134	84	5.591	5,139.531
d_E [b]	7	62	119	136	200	160	98	134	84	5.944	5,139.557
d_{KC} [c]	10	86	137	135	192	126	86	115	113	5.613	5,308.11

[a] $w_1 = w_2 = 0.50$
[b] Euclidean distances
[c] Khuri and Cornell distance

5. Conclusions

It is difficult to suggest rules for choosing a method in matrix optimisation (2) when there are important numerical differences between two of them. For example, Table 2 shows opposing results in the optimum allocations and the minimum variances for the trace and the determinant techniques. A similar situation occurs in the criterion selection for testing hypotheses in the MANOVA problem, see Giri (1977). In fact, the existence of general criteria based on power tests is not sufficient for an objective decision to be made and the final choice depends on the skill of the investigator.

However, when the problem of optimum allocation in multivariate stratified sampling is considered as a nonlinear problem of the multi-objective optimisation of integers, we can give some general suggestions to reduce the number of appropriate methods in accordance with each situation. First we need to recognise the research context of the problem (*i.e.*, total information, partial information or null information). Then, we can decide the technique according to the available information. 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 any example are comparable only within the context in which the example was established.

Acknowledgement

This research work was partially supported by CONACYT-Mexico, Research Grant No. 45974-F, 81512 and IDI-Spain, Grant No. MTM2005-09209. The authors also wish to thank the referees and Editor who handled this article for their very helpful suggestions.

References

Arthanari, T.S., and Dodge, Y. (1981). *Mathematical Programming in Statistics*. A Wiley-Interscience Publication John Wiley & Sons, Inc.

Arvanitis, L.G., and Afonja, B. (1971). Use of the generalized variance and the gradient projection method in multivariate stratified sampling. *Biometrics*, 27, 119-127.

Azaïs, J.-M., and Druilhet, P. (1997). Optimality of neighbour balanced designs when neighbour effects are neglected. *Journal of Statistical Planning and Inference*, 64, 353-367.

Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (2006). *Nonlinear Programming*: *Theory and Algorithms*, 3rd Edition. Wiley-Interscience, 2006.

Cochran, W.G. (1977). *Técnicas de muestreo*. C.E.C.S.A. México (in Spanish).

Dalenius, T. (1957). *Sampling in Sweden. Contribution to the methods and theories of sample survey practice*. Almqvist and Wicksell, Stockholm.

Díaz-García, J.A., and Garay Tapia, M.M. (2007). Optimum allocation in stratified surveys: Stochastic programming. *Computational Statistics and Data Analysis*, 51, 3016-3026.

Giri, N.C. (1977). *Multivariate statistical inference*. New York: Academic Press.

Jones, D.F., Mirrazavi, S.K. and Tamiz, M. (2002). Multi-objective metaheuristics: An overview of the current state-of-the art. *European Journal of Operational Research*, 137, 1-9.

Khuri, A.I., and Cornell, J.A. (1987) *Response Surface*: *Designs Analysis*. New York: Marcel Dekker, Inc.

Miettinen, K.M. (1999). *Non linear multiobjective optimization*. Kluwer Academic Publishers, Boston.

Prékopa, A. (1978). The use of stochastic programming for the solution of some problems in statistics and probability. Technical Summary report 1983. University of Wisconsin-Madison.

Rao, S.S. (1979). *Optimization theory and applications*. Wiley Eastern Limited.

Ríos, S., Ríos Insua, S., and Ríos Insua, M.J. (1989). *Procesos de decisión Multicriterio*. EUDEMA, Madrid (in Spanish).

Steuer, R.E. (1986). *Multiple criteria optimization*: *Theory*, *computation and applications*. New York: John Wiley & Sons, Inc.

Stuart, A. (1954). A simple presentation of optimum sampling results. *Journal of the Royal Statistical Society*, B, 16, 239-241.

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.

Winston, W.L. (1995). *Introduction to mathematical programming*: *Applications and algorithms*. Duxbury Press.