TWO-WAY OPTIMAL STRATIFICATION USING DYNAMIC PROGRAMMING

Pierre Lavallée
Statistics Canada, Ottawa (Ontario), K1A 0T6, Canada

ABSTRACT

In sample surveys, one of the main reasons for stratifying the population is to produce a gain in precision of the estimates. The stratification is often based on several stratification variables. Moreover several variables of interest are measured. The problem considered is how to determine optimum points of stratification which would divide the population domain of two or more stratification variables into distinct subsets such that the precision of one or more variables of interest is maximized. Following the approach used by Bühler & Deutler (1975), the methods proposed are based principally on the use of Dynamic Programming. This computational technique has been found to be free from some of the disadvantages of other previously published methods. Results are presented for the two-way stratification case but can easily be extended to higher dimensions.

Keywords: Dynamic programming; two-way stratification; optimum points; global minima.

1. INTRODUCTION

Stratification is a technique commonly used in sample surveys where the population of interest is divided into sub-populations. As stated by Cochran (1977), one reason for stratification is that it may produce a gain in precision in the estimates of the characteristics of the total population. This article will mainly be concerned with the maximization of the gain in precision due to stratification in the case of stratification based on two variables. The variables will be assumed to come from infinite populations or large finite ones.

The use of a stratified sample survey basically involves five different design operations:

1. the choice of the stratification variables;
2. the choice of the number of strata;
3. the determination of the way in which the population is to be stratified;
4. the allocation of the total sample size n to the strata;
5. the choice of a sampling design within strata.

Although in theory any sampling design can be chosen, following earlier work, only Simple Random Sampling (SRS) without replacement is considered within each strata. The stratification variables can be either categorical or non-categorical (i.e. continuous, discrete). The latter will be the type of stratification variable for which maximization of precision of estimates will be performed. Along this line, the most effective variables on which to stratify would be the variables of interest themselves. However, since in practice this is not always feasible, the stratification variables should be auxiliary variables which are highly correlated with the variables of interest.

The importance of efficient stratification should certainly not be under rated. For example, Dalenius and Gurney (1951) showed that in some cases increasing the number of strata can lead to a loss in precision if stratification is not well chosen.

The optimum stratification of a population consists in dividing the joint domain of the stratification variables in such a way that the precision of the estimates is maximized. In acheiving this goal, it is usually required that this division be done by cutting the domain of each stratification variable into distinct intervals. Such a stratification has been refered to as a lattice (or interval) optimum stratification by Isii and Taga (1969). A lattice stratification can in fact be seen as being formed by straight lines parallel to each of the axis of an Euclidian space.

Considering the determination of the optimum stratification as an important problem in survey sampling, many authors developed different methods to solve it with varying degrees of mathematical rigour. A large number of articles, since Hayashi and Maruyama (1948) and Dalenius (1950), who worked on the one-way optimum stratification problem, have been published for particular applications. One approach that has been considered by Bühler and Deutler (1975) in the one-way case is to use the technique of Dynamic Programming (DP). This technique, which is relatively simple to use, has been found to be free of some of the disadvantages of the previously published methods (see Section 3.2 and 4.2). The DP approach will in fact be the one that will be used in this article to solve the problem of determining the two-way optimum stratification.

2. SOME ASPECTS OF DYNAMIC PROGRAMMING

2.1 Definitions and Concepts

DP has been fully described in Bellman (1957), and Bellman and Dreyfus (1965). No formal definition seems to exist for DP. However, a general definition may be that it is a computational method using recurrence relations for solving sequential decision optimization problems.

A very broad range of DP problems can be described in a formal way using the following notation:

$$\text{Min} \sum_{j=1}^{K} \phi_j(u_{j-1}, u_j) \tag{2.1}$$

$$\text{subject to: } \begin{array}{l} u_j = \tau_j(u_{j-1}, v_j) \\ u_j \varepsilon U_j \\ v_j \varepsilon D_j(u_{j-1}) \\ u_0 = u', \; j=1,\ldots,K. \end{array}$$

The function to be minimized is called the objective function. In the general approach of DP, the concept of stage is used to make the decisions ordered. Here, the subscript j is referring to the jth of the K stages. The optimization problem is then solved sequentially one stage at a time. The state of the jth stage is given by the state variable u_j. The set U_j is called the state space which can be continuous or discrete depending on the type of problem considered. The variable v_j is called the decision variable. Associated with the decision variable is the decision space $D_j(u_{j-1})$. Finally, to be able to describe the states from stage to stage, the stage transformation function τ_j is used.

2.2 The Use of Dynamic Programming

Expressing a problem in the form given by (2.1) is a sufficient condition for using the DP method to solve the problem. The whole idea behind the method of DP is a simple principle called the Bellman's Principle of Optimality. Stated by Bellman (1957), this principle was given as follows:

> "An optimal policy has the property that, whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decisions".

With respect to the DP problem (2.1), let $\Phi_j^*(u_{j-1})$ be the optimal value of the objective function for the stages j to K given the state u_{j-1}. To determine the values of $\Phi_j^*(u_{j-1})$, the Bellman's Principle of Optimality suggests one start at $j=K$ and then go backward down to $j=1$.

Along this process, the problem (2.1) can then be solved using the following recurrence relation:

$$\Phi_j^*(u_{j-1}) = \underset{v_j \in D_j(u_{j-1})}{\text{Min}} \{\phi_j(u_{j-1}, u_j) + \Phi_{j+1}^*(u_j) \mid u_j = \tau_j(u_{j-1}, v_j)\}. \quad (2.2)$$

This last equation is often referred to as the Functional Equation of Bellman. The optimal value $\Phi_1^*(u_0)$ obtained at the end of the process gives in fact the global minimal value of the objective function. Since the optimal values $\Phi_j^*(u_{j-1})$ are now known for all stages j and all states u_{j-1}, the optimal vector $\underline{v}^*=(v_1^*,\ldots,v_K^*)$ can then be obtained by induction. Using $\Phi_1^*(u_0)$, v_1^* is first obtained and then v_2^* is obtained from $\Phi_2^*(u_1^*)$ where $u_1^*=\tau_1(u_0,v_1^*)$ and so on up to the Kth stage. By following this process, the optimization problem is found to be completely solved.

Even if the method involves the use of the recursive Functional Equation of Bellman, DP can be implemented on a computer using either a language that allows recursiveness or not. A non-recursive DP algorithm can be found in Bellman and Dreyfus (1965).

3. TWO-WAY OPTIMUM STRATIFICATION WITH ONE CATEGORICAL STRATIFICATION VARIABLE

In survey sampling, the population is often divided using more than one stratification variable. While the stratification based on the variables of interest is mainly to improve the precision of the estimates, the use of one or more other stratification variables may also be dictated by other reasons such as administrative convenience or to ensure a certain representativeness in some subdivisions of the population. These other stratification variables are often found to be categorical.

In this section, we consider a single variable of interest X for which we want to estimate the mean. The population is assumed to be already stratified into M categories (or classes). Each of these categories is then to be subdivided into L substrata according to the variable of interest X. Each substratum is to be formed by a distinct subinterval of the domain, $[a,b]$ say, of the subpopulation of X for the kth category. That is, for a given category k, each substratum h is to be given by $[x_{h-1}^+, x_h]$ where $[x_{h-1}^+, x_h]=]x_{h-1}, x_h]$ for $h=2,\ldots,L$; $[x_0^+, x_1]=[x_0, x_1]$ and $a=x_0<x_1<\ldots<x_L=b$. The vector $\underline{x}=(x_0,\ldots,x_L)$ is called the vector of points of stratification.

The problem considered here is to obtain, for each category k, Optimal Points of Stratification (OPS) $\underline{x}_k^*$ in such a way that these points of stratification will be the same for all M categories, i.e. $\underline{x}_k^*=\underline{x}^*$. Recall that such a stratification is called a lattice stratification. It should be noted that the following results will also hold for the cases of two or more categorical stratification variables, the resulting M categories being formed by the intersection of the classes of all the categorical variables. It may also be noted that the case of $M=1$ corresponds simply to the case of one-way optimum stratification.

We assume that the variable of interest X is sampled, for the kth of the M categories, from a subpopulation of size N_k with cumulative distribution function (CDF) $F_k(x)$ defined on $[a,b]$ with mean μ_k and finite variance σ_k^2. Note that even if the CDF may be different for each category k, the domain of X is assumed to be the same. Let θ_k represent the relative size of the subpopulation of the kth category such that $\Sigma_{k=1}^M \theta_k=1$ and $\theta_k \geq 0$ for $k=1,\ldots,M$. If the overall population size $N(N=\Sigma_{k=1}^M N_k)$ is finite, for example, θ_k is simply given by

$$\theta_k=\frac{N_k}{N}, \quad k=1,\ldots,M. \quad (3.1)$$

Based on these assumptions, the mean μ of the overall underlying population of the variable of interest X can be expressed as

$$\mu=\sum_{k=1}^M \theta_k\mu_k. \quad (3.2)$$

An unbiased estimate of μ is given by

$$\bar{x}_{comb}=\sum_{k=1}^M \theta_k\bar{x}_k=\sum_{k=1}^M \theta_k \sum_{h=1}^L W_{(k)h}\bar{x}_{kh} \quad (3.3)$$

where

$$\bar{x}_{kh}=\frac{1}{n_{kh}}\sum_{i=1}^{n_{kh}} x_{khi}, \quad (3.4)$$

$$W_{(k)h}=\int_{x_{h-1}^+}^{x_h} dF_k(x). \quad (3.5)$$

The quantity $\bar{x}_{kh}$ represents the sample mean in stratum (k,h) and the quantity $W_{(k)h}$ is the stratum subpopulation proportion within category k. The population proportion for stratum (k,h), denoted by W_{kh}, can be obtained from

$$W_{kh}=\theta_k W_{(k)h}, \quad k=1,\ldots,M; \; h=1,\ldots,L. \quad (3.6)$$

Using the last equation, the estimator $\bar{x}_{comb}$, given by (3.3), can be simply written in the form

$$\bar{x}_{comb}=\sum_{k=1}^{M}\sum_{h=1}^{L}W_{kh}\bar{x}_{kh}. \tag{3.7}$$

As noted in the introduction, it is assumed that the subpopulation sizes N_k are infinite or at least large enough compared to the corresponding sample sizes n_k and the number L of substrata so that the finite population correction (f.p.c.) can be ignored. Hence, ignoring the f.p.c., the variance of $\bar{x}_{comb}$ is given by

$$Var(\bar{x}_{comb})=\sum_{k=1}^{M}\sum_{h=1}^{L}W_{kh}^2\frac{\sigma^2_{(k)h}}{n_{kh}} \tag{3.8}$$

where $$\sigma^2_{(k)h}=\frac{N_k}{N_kW_{(k)h}-1}\int_{x^+_{h-1}}^{x_h}(x-\mu_{(k)h})^2dF_k(x) \tag{3.9}$$

$$\mu_{(k)h}=\frac{1}{W_{(k)h}}\int_{x^+_{h-1}}^{x_h}xdF_k(x), \tag{3.10}$$

$$h=1,\ldots,L;\ k=1,\ldots,M.$$

The determination of the OPS $\underline{x}^*$ to construct the lattice stratification for the population of the variable X can be done by minimizing equation (3.8) with respect to the points of stratification $\underline{x}$.

3.1 Sample Allocation

Dalenius (1950) showed that the OPS generally depend on the type of allocation used for the total sample size n. In general, the allocation can be expressed as

$$n_{kh}=p_{kh}n \tag{3.11}$$

where $\sum_{k=1}^{M}\sum_{h=1}^{L}p_{kh}=1$ and $p_{kh}>0$ for $k=1,\ldots,M$; $h=1,\ldots,L$. The variance of $\bar{x}_{comb}$ under the general allocation is given by

$$Var(\bar{x}_{comb})=\sum_{k=1}^{M}\sum_{h=1}^{L}\frac{W^2_{kh}\sigma^2_{(k)h}}{p_{kh}n}. \tag{3.12}$$

It should be noted that p_{kh} may or not depend on the points of stratification $\underline{x}$.

3.2 Obtaining OPS with DP

The determination of the OPS $\underline{x}^*$ for lattice stratification is done by minimizing the variance of $\bar{x}_{comb}$ under a given type of allocation. One way to solve this problem is to use the approach of Dalenius (1950) as in the case of one stratification variable, which involves taking the partial derivatives of the variance of $\bar{x}_{comb}$ with respect to the points of stratification $\underline{x}$ to obtain the minimal equations. However, the studies of Schneeberger (1979), Goller (1981) and Schneeberger (1985) have shown that the minimal equations, which are only necessary but not sufficient conditions for a global minimum, can also lead to a local minimum, a saddle-point or even a local maximum. An alternative approach, which leads to a global minimum for the variance of $\bar{x}_{comb}$, is to use DP.

As suggested by Bühler and Deutler (1975) for the one-way stratification case, the DP approach discussed in general in Section 2 can easily be applied to the present problem for most allocations represented by (3.11). For the problem of determining the OPS $\underline{x}^*$ for lattice stratification with one categorical stratification variable, the objective function $G(\underline{x})$ to be minimized can be expressed in the form

$$G(\underline{x})=\sum_{h=1}^{L}\left(\sum_{k=1}^{M}g_k(x_{h-1},x_h)\right). \tag{3.13}$$

The determination of $g_k(x_{h-1},x_h)$ is done by considering the formula for the variance of $\bar{x}_{comb}$, given by (3.12), obtained under a given allocation. Unfortunately some allocations cannot be handled using the DP approach since they do not permit expressing the variance of $\bar{x}_{comb}$ in the form (3.13) or even more generally in the form (2.1). However, for most usual allocations such as proportional or Neyman allocation, the form $G(\underline{x})$ is relatively easy to obtain.

With the stages corresponding to the different intervals h, the problem of determining of the OPS $\underline{x}^*$ for lattice stratification can be expressed in a form similar to (2.1) by

$$\text{Min}\sum_{h=1}^{L}\sum_{k=1}^{M}g_k(x_{h-1},x_h) \tag{3.14}$$

subject to:
$$x_h=x_{h-1}+d_h$$
$$x_h\in[a,b]$$
$$d_h\in B_h(x_{h-1})=[0,b-x_{h-1}]$$
$$x_0=a,\ h=1,\ldots,L.$$

The sequential decision aspect of DP can in fact be seen as distributing to one interval h at a time a portion d_h of the domain $[a,b]$.

The next step in the DP formulation of the problem involves the determination of the Functional Equation of Bellman given in general by (2.2). Let $G^*_h(x_{h-1})$ be the optimal value of the objective function for the strata (k,h) to (k,L) considering all M categories, given that lower bound for the strata (k,h) for $k=1,\ldots,M$ is x_{h-1}. The Functional Equation of Bellman for determination of the OPS $\underline{x}^*$ for lattice stratification is then expressed as

$$G^*_h(x_{h-1})=\underset{d_h\in B_h(x_{h-1})}{\text{Min}}\left\{\sum_{k=1}^{M}g_k(x_{h-1},x_h)+G^*_{h+1}(x_h)\mid x_h=x_{h-1}+d_h\right\}. \tag{3.15}$$

Following Result 1 of Bühler and Deutler (1975), the solutions obtained from the DP formulation (3.14) of the two-way optimum stratification problem with one categorical stratification variable would in fact be the true OPS $\underline{x}^*$. It should be noted that no convexity assumptions have been used with respect to the objective function (3.13). In the optimum stratification problem, the main difficulty comes in fact from being unable to assume convexity in general. Based on equation (3.15), the DP approach can be implemented easily. A difficulty, however, is brought by the infinite nature of both the state space

$[a,b]$ and the decision spaces $B_h(x_{h-1})$. This problem can fortunately be overcome by crossing these intervals by discrete steps which makes them look like finite sets. The number of these steps is chosen by taking into account precision, computer space and time.

In practice, it often occurs that the population of the variable X is a real population in the sense that it is a set of measured values. The DP approach is then simply used by replacing the different quantities entering into the objective functions $G(\underline{x})$ by their finite population counterparts.

4. TWO-WAY OPTIMUM STRATIFICATION

The method of optimum stratification considered in the previous section is applicable to a single variable of interest X. However, as pointed out by Kish and Anderson (1978), most surveys are in practice multisubject, i.e. in a single survey, several variables are measured. Multisubject surveys lead to considerably reduced costs compared to individual surveys for each variable of interest. However, the determination of optimum points of stratification is more difficult than in the single variable case.

In this section, it is supposed that two variables of interest X and Y are to be measured in order to produce estimates of their means μ_X and μ_Y, respectively. These variables (X,Y) are assumed to be sampled from a population of size N with joint CDF $F(x,y)$ defined on $[a_X,b_X]\times[a_Y,b_Y]$ with mean (μ_X,μ_Y) and covariance matrix $\Sigma=\begin{bmatrix}\sigma_X^2 & \sigma_{XY}\\ \sigma_{XY} & \sigma_Y^2\end{bmatrix}$. The matrix Σ is assumed to be finite, i.e. $|\Sigma|<\infty$ where $|A|$ denotes the determinant of the matrix A. Given that the population is to be divided into LxM strata, the problem considered here is to obtain OPS $\underline{x}^*=(x_0^*,x_1^*,\ldots,x_L^*)$ with respect to the variable X together with OPS $\underline{y}^*=(y_0^*,y_1^*,\ldots,y_M^*)$ with respect to the variable Y to form the optimum lattice stratification for the estimation of the means μ_X and μ_Y.

The usual unbiased estimators of the means μ_X and μ_Y are given by $\bar{x}_{st}$ and $\bar{y}_{st}$, respectively. For a given vector of points of stratification $(\underline{x},\underline{y})$, the estimator $\bar{x}_{st}$ can be obtained from

$$\bar{x}_{st}=\sum_{h=1}^{L}\sum_{k=1}^{M}W_{hk}\bar{x}_{hk} \tag{4.1}$$

$$\text{where } \bar{x}_{hk}=\frac{1}{n_{hk}}\sum_{i=1}^{n_{hk}}x_{hki}, \tag{4.2}$$

$$W_{hk}=\int_{x_{h-1}^+}^{x_h}\int_{y_{k-1}^+}^{y_k} dF(x,y). \tag{4.3}$$

The estimator $\bar{y}_{st}$ is similarly obtained from (4.1) by substituting y's for x's. It should be noted that the stratum population proportion W_{hk} given by (4.3) differs from the one denoted by W_{kh} and used in Section 3. As in the previous section, it is again assumed that the population size N is infinite or at least large enough compared to the total sample size n and the number LxM of strata so that the f.p.c. can be ignored. Hence, the sampling variance related to the estimator $\bar{x}_{st}$ is simply given by

$$\mathrm{Var}(\bar{x}_{st})=\sum_{h=1}^{L}\sum_{k=1}^{M}\frac{W_{hk}^2\sigma_{Xhk}^2}{n_{hk}} \tag{4.4}$$

$$\text{where } \sigma_{Xhk}^2=\frac{N}{NW_{hk}-1}\int_{x_{h-1}^+}^{x_h}\int_{y_{k-1}^+}^{y_k}(x-\mu_{Xhk}^2)\,dF(x,y) \tag{4.5}$$

$$\mu_{Xhk}=\frac{1}{W_{hk}}\int_{x_{h-1}^+}^{x_h}\int_{y_{k-1}^+}^{y_k} x\,dF(x,y). \tag{4.6}$$

An expression similar to (4.4) can be derived for the variance of $\bar{y}_{st}$.

One of the main problems in the determination of two-way optimum stratification is the choice of the objective function to be minimized to obtain the vector of OPS $(\underline{x}^*,\underline{y}^*)$. Unlike the case of a single variable of interest, the objective function for two variables of interest is not uniquely defined because of the various possibilities of considering the variances and covariance of $\bar{x}_{st}$ and $\bar{y}_{st}$. One possible objective function proposed in the literature (Ghosh (1963) and Sadasivan and Aggarwal (1978)) is the generalized variance of $\bar{x}_{st}$ and $\bar{y}_{st}$. The generalized variance is defined by the determinant of the covariance matrix of $\bar{x}_{st}$ and $\bar{y}_{st}$. As stated by Dahmström and Hagnell (1978), a somewhat more natural measure is the sum of the variances of $\bar{x}_{st}$ and $\bar{y}_{st}$ or, more generally, a weighted sum of these variances where the weights indicate the relative importance of the two variables. Letting θ_X and θ_Y be the weights associated with the variances of $\bar{x}_{st}$ and $\bar{y}_{st}$, respectively, such that $\theta_X+\theta_Y=1$, $\theta_X\geq 0$ and $\theta_Y\geq 0$, this objective function can be expressed as

$$V_{XY}=\theta_X\mathrm{Var}(\bar{x}_{st})+\theta_Y\mathrm{Var}(\bar{y}_{st})$$
$$=\sum_{h=1}^{L}\sum_{k=1}^{M}\frac{W_{hk}^2(\theta_X\sigma_{Xhk}^2+\theta_Y\sigma_{Yhk}^2)}{n_{hk}}. \tag{4.7}$$

The weighted average V_{XY} is often used as a basis to obtain an optimal sample size allocation (e.g., Dalenius (1957) and Cochran (1977)). In this article, the determination of OPS is considered only under the weighted average V_{XY}.

4.1 Sample Allocation

Since the values of the OPS generally depend on the type of allocation used, the minimization of the weighted average (4.7) should be done by considering a given allocation.

Again, the allocation can be expressed in general as $n_{hk} = n\,p_{hk}$. The weighted average V_{XY} under the general allocation is simply expressed as

$$V_{XY}=\sum_{h=1}^{L}\sum_{k=1}^{M}\frac{W_{hk}^{2}(\theta_X\sigma_{Xhk}^{2}+\theta_Y\sigma_{Yhk}^{2})}{p_{hk}n}. \qquad (4.8)$$

As mentioned in Section 3.1, p_{hk} may or not depend on the points of stratification $(\underline{x},\underline{y})$.

4.2 Obtaining OPS with DP

Ghosh (1963) first considered the problem of two-way optimum stratification. Using the same approach as Dalenius (1950), he obtained minimal equations by taking the partial derivatives of the generalized variance under the proportional allocation.

However, as noted earlier, the minimal equations are only necessary but not sufficient conditions for achieving a global minimum of the objective function. In fact, Schneeberger and Pollot (1985) showed that, for the bivariate normal distribution, Ghosh's result would lead to a saddle-point. To solve the problem of determining the OPS $(\underline{x}^*,\underline{y}^*)$ minimizing the weighted average V_{XY}, the method to be used should provide at least a local minimum (if not a global one). As in the preceding section, such a method can again be developed by the use of DP.

The objective function for the present problem can be expressed in the form

$$G(\underline{x},\underline{y})=\sum_{h=1}^{L}\sum_{k=1}^{M}g(x_{h-1},x_h,y_{k-1},y_k). \qquad (4.9)$$

The expression for the function $g(x_{h-1},x_h, y_{k-1},y_k)$ is simply determined from the formula of the weighted average V_{XY} given by (4.8). Again some allocations do not permit expressing V_{XY} given by (4.8) in the form (4.9). The objective function (4.9), however, is relatively easy to obtain for most usual allocations.

The general equation (4.9) for the objective function suggests that, for the determination of two-way optimum stratification, a two-dimensional DP approach should be used. Employing the general concepts of DP given in Section 2 with the state and decision variables being now two-dimensional, the stages may be specified by the pairs (h,k). The problem of two-way optimum stratification can then be expressed as

$$\mathrm{Min}\sum_{h=1}^{L}\sum_{k=1}^{M}g(x_{h-1},x_h,y_{k-1},y_k) \qquad (4.10)$$

subject to: $(x_h,y_k)=(x_{h-1}+d_h,y_{k-1}+t_k)$

$(x_h,y_k)\in[a_X,b_X]\times[a_Y,b_Y]$

$(d_h,t_k)\in B_h^{(X)}(x_{h-1})\times B_k^{(Y)}(y_{k-1})$

$=[0,b_X-x_{h-1}]\times[0,b_Y-y_{k-1}]$

$(x_0,y_0)=(a_X,a_Y)$,

$h=1,\ldots,L;\ k=1,\ldots,M.$

Although the formulation (4.10) seems to correspond to the one given by (2.1), a difficulty can in fact be seen with respect to the decision space $B_h^{(X)}(x_{h-1})\times B_k^{(Y)}(y_{k-1})$. In the formulation (2.1), the decision space depends only on the state of the previous stage. However, in the formulation (4.10), the decision space $B_h^{(X)}(x_{h-1})\times B_k^{(Y)}(y_{k-1})$ depends on the states of both past and future stages since the variable x_{h-1} is present in the decision spaces of the M stages $(h,1)$ to (h,M) and the variable y_{k-1} in the decision spaces of the L stages $(1,k)$ to (L,k).

Hence, the Bellman's Principle of Optimality stated in Section 2.2, which forms the basis for the use of DP, is not applicable. It should be noted that the problem is partially due to the way the stages have been defined. In fact, instead of L×M stages related to each stratum (h,k), there are only (L+M) stages which can be viewed through the decision variables d_h and t_k. However, due to the nature of the objective function given in general by (4.9), its transformation to reflect the (L+M) stages does not seem to be mathematically tractable for most allocations.

We propose a simple approach which permits a solution to the problem (4.10) using the unidimensional DP iteratively. Before the first iteration, some trial values, say $\underline{x}^{(0)}$ and $\underline{y}^{(0)}$ such that $a_X = x_0^{(0)} < x_1^{(0)} < \ldots < x_{h-1}^{(0)} < x_L^{(0)} = b_X$ and $a_Y = y_0^{(0)} < y_1^{(0)} < \ldots < y_{M-1}^{(0)} < y_M^{(0)} = b_Y$, are chosen for the initial points of stratification with respect to the variables X and Y, respectively. Then for the ith iteration $(i=1,2,\ldots)$, the points of stratification $\underline{y}^{(i-1)}$ are first considered as fixed. Note that the points of stratification $\underline{x}^{(i-1)}$ could also be chosen instead of $\underline{y}^{(i-1)}$. Fixing the values of $\underline{y}^{(i-1)}$ has in fact the effect of reducing the problem exactly to the one of two-way optimum stratification with one categorical stratification variable discussed in Section 3. This can be seen by comparing the formulation (4.10) to the one given by (3.14) with the values of the points of stratification $\underline{y}$ taken as constants in (4.10).

Let $G^*_{Xh}(x_{h-1},\underline{y}^{(i-1)})$ be the optimal value for the objective function (4.9) for the strata (h,k) to (L,k) for all $k=1,\ldots,M$ given that the lower bound for the strata (h,k) for $k=1,\ldots,M$ is x_{h-1}. The Functional Equation of Bellman with respect to the first part of the ith iteration is then given by

$$G^*_{Xh}(x_{h-1},\underline{y}^{(i-1)})=\underset{d_h\in B_h^{(X)}(x_{h-1})}{\mathrm{Min}}\Big\{\sum_{k=1}^{M}g(x_{h-1},x_h,y_{k-1}^{(i-1)},y_k^{(i-1)})+G^*_{Xh+1}(x_h,\underline{y}^{(i-1)})\ \Big|\ x_h=x_{h-1}+d_h\Big\} \qquad (4.11)$$

where $B_h^{(X)}(x_{h-1})$ is given in (4.10).

Using this last equation, new points of stratification $\underline{x}^{(i)}$ with respect to the variable X can be obtained to replace the preceding values $\underline{x}^{(i-1)}$. Hence, the OPS for the first part of the ith iteration are given by $(\underline{x}^{(i)},\underline{y}^{(i-1)})$. For the second part of the ith iteration, the points of stratification $\underline{x}^{(i)}$ are, in turn, considered as fixed. Using again the similarity that exists with the formulations (4.10) and (3.14) but now with the values of $\underline{x}$ taken as constants in (4.10), let $G^*_{Yk}(\underline{x}^{(i)},y_{k-1})$ be the optimal value for the objective function (4.9) for the strata (h,k) to (h,M) for all $h=1,\ldots,L$ given that the lower bound for the strata (h,k) for $h=1,\ldots,L$ is y_{k-1}.

The Functional Equation of Bellman for the second part of the ith iteration is then given by

$$G^*_{Yk}(\underline{x}^{(i)},y_{k-1}) = \underset{t_k \in B_k^{(Y)}(y_{k-1})}{\text{Min}} \{ \sum_{h=1}^{L} g(x_{h-1}^{(i)},x_h^{(i)}, y_{k-1},y_k) + G^*_{Yk+1}(\underline{x}^{(i)},y_k) \mid y_k = y_{k-1}+t_k \} \quad (4.12)$$

where $B_k^{(Y)}(y_{k-1})$ is given in (4.10).

Using expression (4.12), the values of the OPS $\underline{y}^{(i)}$ can be determined to replace the previous values $\underline{y}^{(i-1)}$ so that, for the ith iteration, the OPS are given by $(\underline{x}^{(i)},\underline{y}^{(i)})$.

This process iterates until some convergence criterion is satisfied. A possible convergence criterion as follows:

$$\max_{h=1}^{L}(|x_h^{(i)}-x_h^{(i-1)}|/x_h^{(i-1)})<\varepsilon_X \quad (4.13)$$

and

$$\max_{k=1}^{M}(|y_k^{(i)}-y_k^{(i-1)}|/y_k^{(i-1)})<\varepsilon_Y \quad (4.14)$$

where $\varepsilon_X>0$ and $\varepsilon_Y>0$.

A proof of convergence to the correct limit is given in Lavallée (1987). It is shown under simple assumptions that, if the unidimensional iterative DP approach converges, the convergence limit is a local minimum with probability 1 (or a saddle-point but with probability 0). It should be noted that, when the objective function is not convex, which is the case here, most of the iterative methods starting with arbitrary values cannot guarantee to lead to a global minimum.

The main advantage of the unidimensional iterative DP approach is that it can be easily implemented using the DP approach of Section 3 as a guideline. One of its disadvantages is that convergence cannot be insured. Note that this is true in most other cases where a univariate iterative approach is used instead of a multivariate one.

When the population of the variables X and Y is a real population, the OPS $\underline{x}^*$ and $\underline{y}^*$ are then obtained following the same approach as the one described in Section 3.2. Following this approach, the OPS are obtained using the proposed unidimensional iterative DP method, but the different quantities entering into the objective function (4.9) are replaced by their finite population counterparts.

REFERENCES

Bellman, R.E. (1957): Dynamic Programming. Princetown University Press, New Jersey.

Bellman, R.E., Dreyfus, S.E. (1965): La programmation dynamique et ses applications. Dunod, Paris, 385 pages.

Bühler, A., Deutler, M. (1975): Optimal Stratification and Grouping by Dynamic Programming. Metrika, Vol. 22, p. 161-175.

Cochran, W.G. (1977): Sampling Techniques, 3rd Edition. John Wiley & Sons, New York, 428 pages.

Dahmström, P., Hagnell, M. (1978): Multivariate Stratification of Primary Sampling Units in a General Purpose Survey. Scand. J. Statist., Vol. 5, p. 53-56.

Dalenius, T. (1950): The Problem of Optimum Stratification. Skand. Aktuarietidskr, Vol. 33, p. 201-213.

——— (1957): Sampling in Sweden. Almqvist and Wiksell, Stockholm.

Dalenius, T., Gurney, M. (1951): The Problem of Optimum Stratification. II. Skand. Aktuarietidskr., Vol. 34, p. 131-148.

Ghosh, S.P. (1963): Optimum Stratification with Two Characters. Ann. Math. Statist., Vol. 34, p. 866-872.

Goller, W. (1981): Eine Einfache Symmetrische Verteilung, Bei Der Das Symmetriezentrum Nicht Der Optimale Schichtungspunkt Sein Muß. Allgemeines Statistisches Archiv., Vol. 65, p. 368-375.

Hayashi, C., Maruyama, F. (1948): On Some Methods for Stratification. Research Memoir, Inst. Statist. Math., Vol. 4, p. 399-411.

Isii, K., Taga, Y. (1969): On Optimal Stratification for Multivariate Distributions. Skand. Aktuarietidskr., p. 24-38.

Kish, L., Anderson, D.W. (1978): Multivariate and Multipurpose Stratification. J.A.S.A., Vol. 73, No. 361, p. 24-34.

Lavallée, P. (1987): Some Contributions to Optimum Stratification. M. Sc. Thesis, Carleton University, Ottawa, 109 pages.

Nedzela, M. (1981): Introduction à la science de la gestion. Les presses de l'Université du Québec, Québec, 373 pages.

Sadasivan, G., Aggarwal, R. (1978): Optimum Points of Stratification in Bivariate Populations. Sankhyã, Vol. 40, Ser. C, part 1, p. 84-97.

Schneeberger, H. (1979): Saddle-Points of the Variance of the Sample Mean in Stratified Sampling. Sankhyã, Vol. 41, Ser. C, part 1, p. 92-96.

——— (1985): Maxima, Minima Und Sattlepunkte Bei Optimaler Schichtung und Optimaler Aufteilung. Allgemeines Statist. Archiv., Vol. 69, p. 286-297.

Schneeberger, H., Pollot, J.P. (1985): Optimum Stratification with Two Variates. Statist. Hefte, Vol. 26, No. 2, p. 97-113.