1

MINOR TARGET MARKS
MICRO-D
MADE IN THE U.S.A.

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30

SOME CONTRIBUTIONS
TO
OPTIMAL STRATIFICATION .

by

Pierre Lavallée

A thesis submitted to the Faculty of
Graduate Studies in partial fulfilment
of the requirements for the degree of
Master of Science
Department of Mathematics & Statistics

Carleton University
Ottawa, Ontario
November, 1987
© 1987

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-40630-5

The undersigned recommend to the Faculty of Graduate Studies and Research acceptance of this thesis

Some Contributions to Optimal Stratification

submitted by Pierre Lavalée, B.Sc., M.Sc., in partial fulfillment of the requirements for the degree of Master of Science

---

Dr. J.N.K. Rao, *Thesis Supervisor*

Dr. M. Alvo, *External Examiner*

Dr. C. Garner, *Chairman*
Department of Mathematics & Statistics

ABSTRACT

In sample surveys, one of the main reasons for stratifying the population is to produce a gain in precision of the estimates. The problem considered in this thesis is how to determine optimum points of stratification which would divide the (finite or infinite) population domain of the stratification variable into distinct intervals, such that the precision of an estimate of the variable of interest is maximized. Following the approach used by Bühler and Deutler (1975), the methods proposed are based principally on the use of Dynamic Programming.

After providing an overview of the Dynamic Programming approach in general, this thesis treats the following subjects: optimum stratification techniques in one-way stratification cases, optimum stratification with respect to variance estimation, two-way optimum stratification where one stratification variable is categorical and two-way optimum stratification where both stratification variables are non-categorical.

ACKNOWLEDGEMENTS

I would like first to gratefully thank Dr. J.N.K. Rao for his willingness to be my supervisor during my master studies and for spending some of his precious time in useful discussions and in giving advice. I also thank Edward Hughes of the Ontario Research Foundation for his help in the analysis of the unidimensional iterative Dynamic Programming approach which is presented in Chapter 6 of this thesis. A grateful thanks also goes to Dr. Shirley Mills for her encouragement and for providing facilities to do the necessary computing in the Mathematical and Statistical Computing Laboratory. I must also mention Diane Berezowski and Nancy Darcovich who put their time, respectively, in typing and proof reading this thesis. Finally, but certainly not least, I would like to thank my wife Marie-Claude for being so patient during the time spent on my master studies.

TABLE OF CONTENTS

1. Introduction	1
2. The Nature of Dynamic Programming	5
2.1 Definitions and Concepts	5
2.2 The Use of Dynamic Programming	6
2.3 Advantages and Disadvantages	7
3. One-Way Optimum Stratification	9
3.1 Infinite Population Case	11
3.1.1 Literature Review	11
3.1.2 Identification of Problems	13
3.1.3 The Dynamic Programming Approach	13
3.1.4 Dynamic Programming with Real Populations	15
3.1.5 Computer Program and Illustrations	16
3.2 Finite Population Case	21
3.2.1 Literature Review	21
3.2.2 Identification of Problems	22
3.2.3 The Dynamic Programming Approach	22
3.2.4 Dynamic Programming with Real Populations	25
3.2.5 Computer Program and Illustrations	25
3.3 Use of an Auxiliary Variable	29
3.3.1 Literature Review	31
3.3.2 The Dynamic Programming Approach	33
4. Optimum Stratification With Respect to Variance Estimation	36
4.1 Infinite Population Case	38
4.1.1 Possible Comparison Setups	38
4.1.2 The Use of Dynamic Programming	40
4.1.3 Computer Program and Empirical Studies	42
4.2 Finite Population Case	48
4.2.1 Possible Comparison Setups	48
4.2.2 The Use of Dynamic Programming	49

4.3 Use of an Auxiliary Variable.....	50
4.3.1 Possible Comparison Setups.....	52
4.3.2 The Use of Dynamic Programming	53
5. Two-Way Optimum Stratification With One Categorical Stratification Variable	55
5.1 Infinite Population Case.....	61
5.1.1 The Dynamic Programming Approach.....	61
5.1.2 Computer Program and Illustrations.....	64
5.2 Finite Population Case.....	70
5.2.1 The Dynamic Programming Approach	71
5.3 Use of an Auxiliary Variable.....	75
5.3.1 The Dynamic Programming Approach	77
6. Two-Way Optimum Stratification	78
6.1 Infinite Population Case.....	83
6.1.1 Literature Review	83
6.1.2 Identification of Problems	83
6.1.3 The Dynamic Programming Approach	84
6.1.4 Computer Program and Illustrations.....	88
6.2 Finite Population Case.....	95
6.2.1 The Dynamic Programming Approach	95
6.3 Use of an Auxiliary Variable	100
6.3.1 Literature Review	103
6.3.2 Identification of Problems.....	103
6.3.3 The Dynamic Programming Approach	103
References.....	105
Appendix A: The OPTIBND1 Program	
Appendix B: The OPTIBND2 Program	
Appendix C: The OPTIBND3 Program	
Appendix D: The OPTIBND4 Program	
Appendix E: The OPTIBND5 Program	
Appendix F: Convergence of the Unidimensional Iterative Dynamic Programming Approach	

SOME CONTRIBUTIONS TO OPTIMAL STRATIFICATION

by Pierre Lavallée

1. INTRODUCTION

Stratification is a technique commonly used in sample surveys where the population of interest is divided into sub-populations. Sampling is then done independently in the different strata. Cochran (1977) gives four main reasons for stratification. First, if estimates of known precision are desired for certain subdivisions, it may be advisable to treat each subdivision as a "population" to ensure a certain representativeness. Secondly, stratification may be dictated by administrative convenience if, for example, the survey agency has field offices. Thirdly, sampling problems may differ markedly in different parts of the population such that each part should be considered independently. Finally, stratification may produce a gain in precision in the estimates of the characteristics of the total population. This thesis will mainly be concerned with the maximization of the gain in precision due to stratification.

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

1. the choice of a stratification variable;
2. the choice of the number L 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, Simple Random Sampling (SRS) without replacement is only considered within each strata. The latter design is extensively used for sampling firms from lists in business surveys conducted by Statistics Canada. It should be noted, however, that some work by Singh and Sukhatme (1972) and Singh (1975) also treats sampling with varying probabilities. The stratification variable 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.

As pointed by Hansen, Hurwitz and Madow (1953), the choices involved in the five design operations are a matter in which effective use can be made of prior knowledge, personal intuition, judgement and statistical information that may be available. 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 L can lead to a loss in precision if stratification is not well chosen. Regarding the choice of the stratification variable, the most effective variable on which to stratify would be the variable of interest itself, say

X . However, since in practice this is not always feasible, the stratification variable should be an auxiliary variable which is highly correlated with the variable of interest X .

The maximization of the gain due to stratification can be directly related to the minimization of the variance of an estimate of the mean μ of the variable of interest X . Throughout this thesis, it will be assumed that X is sampled from a population of size N with cumulative distribution function (CDF) $F(x)$ defined on $[a, b]$ with mean μ and finite variance σ^2 . In Chapter 4, however, the assumption on $F(x)$ will be extended to finite fourth moment. Assuming that the population is to be divided into L strata, each stratum will be formed by a distinct subinterval of $[a, b]$. That is, each stratum h will be given by $[x_{h-1}^+, x_h]$ where $[x_{h-1}^+, x_h] =]x_{h-1}, x_h]$ for $h = 2, \dots, L$; $[x_0^+, x_1] = [x_0, x_1]$ and $a = x_0 < x_1 < \dots < x_L = b$. The vector $\underline{x} = (x_0, \dots, x_L)$ is called the vector of points of stratification. If an auxiliary variable Z defined on $[a_Z, b_Z]$ is used, the points of stratification will be the vector $\underline{z} = (z_0, \dots, z_L)$ where $a_Z = z_0 < z_1 < \dots < z_L = b_Z$. An unbiased estimate of the population mean μ is given by

$$(1.1) \quad \bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h$$

$$(1.2) \quad \text{where} \quad \bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi},$$

$$(1.3) \quad W_h = \int_{x_{h-1}^+}^{x_h} dF(x).$$

It should be noted that $\bar{x}_h$ is the sample mean in stratum h while W_h represents the population proportion falling in stratum h . The variance of $\bar{x}_{st}$ is given by

$$(1.4) \quad \text{Var}(\bar{x}_{st}) = \sum_{h=1}^L \frac{W_h^2 \sigma_h^2}{n_h} - \sum_{h=1}^L \frac{W_h \sigma_h^2}{N}.$$

$$(1.5) \quad \text{where } \sigma_h^2 = \frac{N}{NW_h - 1} \int_{x_{h-1}^+}^{x_h} (x - \mu_h)^2 dF(x),$$

$$(1.6) \quad \mu_h = \frac{1}{W_h} \int_{x_{h-1}^+}^{x_h} x dF(x).$$

The value μ_h represents the stratum mean and the value σ_h^2 , the stratum variance. In the case of an infinite population ($N = +\infty$), the second term of (1.4) becomes zero so that only the first term is considered. That second term is usually called the finite population correction (f.p.c.). Although in practice infinite populations

seldom occur, statisticians often deal with large finite populations so that the f.p.c. becomes negligible. The problem considered in this thesis will mainly be to find the Optimal Points of Stratification (OPS), denoted by x^* (or ξ^* if an auxiliary variable is used), such that the variance of $\bar{x}_{st}$ will be minimized. Such a stratification will be called an optimum stratification.

Before trying to develop some methods to obtain OPS, it may be worthwhile to see whether or not such points exist. Taga (1967) considered the general optimum stratification of the population of X . Such stratification is defined as the general division of the domains of X without restrictions to interval divisions. Under simple assumptions on $F(x)$, Taga (1967) proved that, under proportional allocation of the sample size n , the general optimal stratification exists and that, moreover, it coincides with the interval optimum stratification given by the OPS. For Neyman allocation, however, Isii (1966) showed that the general optimum stratification may not be the same as the interval optimum stratification. In the case where the stratification is based on one or several auxiliary variables, the existence of a general optimum stratification under proportional allocation was also proved by Taga (1967) where, in the univariate case, the optimum stratification was again coinciding with the interval optimum stratification. Isii and Taga (1969) showed that, under some assumptions, the general optimum stratification exists, under both proportional and Neyman allocations, for the multivariate case where there are several variables of interest and several stratification variables. Although the interval optimum stratification may not lead to the largest possible reduction in the variance of $\bar{x}_{st}$, this thesis will be restricted to the determination of OPS. This is mainly due to the complexity that may exist in obtaining the general stratification, and the operational difficulties that may occur in practice in implementing such a stratification.

The problem of determining OPS is not an isolated one in the sense that it is found to be linked to the problems of optimum grouping and optimum spacing. Following Eubank (1982), the problem of optimum grouping is solved by choosing group boundaries $a = x_0 < x_1 < \dots < x_{L-1} < x_L = b$ that minimize the asymptotic variance of the maximum likelihood estimator of a location or scale parameter obtained using only the group boundaries and the proportion of sample elements within each group. On the other hand, the problem of optimum spacing is solved by selecting percentile points $0 = u_0 < u_1 < \dots < u_{L-1} < u_L = 1$ (frequently called a spacing) corresponding to quantiles which maximize the asymptotic relative Fisher efficiency of the Asymptotic Best Linear Unbiased Estimate (ABLUE) of a location or scale parameter. The three problems have been shown by Adatia and Chan (1981) to be in fact equivalent under an absolutely continuous distribution for X . Hence, the solution of one of the three problems can directly lead to the solution of the two others. Anderson, Kish and Cornell (1980) argued that their joint treatment is useful since the numerical results and tables available for one problem can be used in all three problems. It is also efficient to share the

mathematical results among the three problems. The solutions proposed in this thesis for the optimal stratification might then be extended to the problems of optimum grouping and optimum spacing.

Considering the determination of the OPS as an important problem in survey sampling, many authors developed different methods to solve it with varying degree of mathematical rigour. A large number of articles, since Hayashi and Maruyama (1948) and Dalenius (1950) who worked independently, have been published on this problem for particular applications. One approach that has been considered by Bühler and Deutler (1975) 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. The DP approach will in fact be the one that will be described and extended in this thesis to solve the problem of determining the OPS.

Chapter 2 of this thesis will give an overview of the DP approach in general. Chapter 3 will treat the problem of the one-way optimum stratification. It should be noted that this part will be based mainly on the work of Bühler and Deutler (1975). Chapter 4 will deal with the problem of the optimum stratification with respect to variance estimation. In this case, the problem will be to find OPS in order to maximize the precision of an estimate of $\text{Var}(\bar{x}_{st})$ instead of the precision of $\bar{x}_{st}$ itself. Some empirical results will be given. Chapter 5 of this thesis will look at the two-way optimum stratification where one stratification variable is categorical. This problem arises when, for example, a national survey is already stratified by province or state. Finally, Chapter 6 of this thesis will consider two-way optimum stratification when the two stratification variables are non-categorical. To solve this problem an algorithm based on an unidimensional iterative DP approach will be given.

2. THE NATURE OF DYNAMIC PROGRAMMING

Dynamic Programming (DP) is a computational method involving the use of recurrence relations. Developed to a considerable extent by Richard Bellman in the early 1950's, this method has been named after the type of problems that were originally solved using this technique. DP evolved in fact as a result of studying non-linear optimization problems in which changes over time were important, i.e. problems with a dynamic nature. However, up to now this method has been applied to various problems in which time was in no way relevant. The kind of problems actually solved by DP are sequential decision optimization problems or, in other words, optimization problems where some "decisions" have to be made in a given order.

2.1 Definitions and Concepts

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. One typical example of the use of DP is the problem of allocation of resources discussed by Bellman and Dreyfus (1965). In this problem, a quantity Q of a particular resource has to be allocated to K different activities where a cost is associated to each activity depending on the quantity of resource allocated. The problem is then to determine the allocation of the resource in order to minimize the overall cost. It is assumed that the overall cost is the sum of the cost of each activity and that the cost of each activity is independent of the quantity of resource allocated to the other activities. This last assumption can however be relaxed in the general DP approach.

The preceding example and, in fact, a very broad range of DP problems can be described in a formal way using the following notation:

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

$$\text{subject to: } u_j = \tau_j(u_{j-1}, v_j)$$

$$u_j \in U_j$$

$$v_j \in D_j(u_{j-1})$$

$$u_0 = u', \quad j = 1, \dots, K$$

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 j th stage is referring to the j th of the K activities. The resource is then allocated sequentially one stage at a time. At each stage, the concept of state is used to describe the allocated stock (i.e. the quantity of resource already allocated)

at the given stage. The state of the j th stage is given by the state variable u_j . Note that, in the example, the initial state u_0 is set to zero since the stock level is full before beginning the allocation. The set U_j is called the state space which can be continuous or discrete depending on the type of problem considered. The act of allocating, at a given stage, a certain quantity of resource is referred to as a decision in the general approach of DP. The variable v_j , which gives the quantity of resource allocated is then called the decision variable. Associated with the decision variable is the decision space $D_j(u_{j-1})$ which is the set of possible quantities that can be allocated at stage j . It should be noted that the state space depends on the quantity u_{j-1} of resource already allocated to the previous $j - 1$ stages. The cost of allocating v_j amount of resource at state j is given by the stage return function ϕ_j , which may be different from stage to stage. The stage return function may also depend on the state variable u_{j-1} , i.e. the quantity allocated in the preceding stages. This last condition is the way in which the independence assumption of the allocation problem can be relaxed. Finally, to be able to describe the states from stage to stage, the stage transformation function τ_j is used. For the example, the stage transformation function can easily be seen to be $u_j = u_{j-1} + v_j$.

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".

This principle can be illustrated by the following example: Suppose that a traveller has to go from city A to city C in an optimal way. Assume now that the optimal way from another city B to city C is known. If the traveller arrives at city B then, whatever way the traveller has used to go from A to B, the optimal way from B to C is unchanged.

With respect to the DP problem of the form (2.1), let $\Phi_j^*(u_{j-1})$ be the optimal value of the objective function for the stages j to K given that the state of the stage $j - 1$ is u_{j-1} . The function $\Phi_j^*(u_{j-1})$ corresponds in the problem of allocation of resources to the minimal total cost of the activities j to K given that the quantity u_{j-1} has already been allocated to the first $j - 1$ activities. 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 $j = 1$. For $j = K$, the value of $\Phi_j^*(u_{j-1})$ for a given state u_{K-1} is simply obtained by

$$(2.2) \quad \Phi_K^*(u_{K-1}) = \min_{v_K \in D_K(u_{K-1})} \{ \phi_K(u_{K-1}, v_K) \mid u_K = \tau_K(u_{K-1}, v_K) \}.$$

The resource is then optimally allocated to the last activity given that u_{K-1} of the total quantity Q has already been allocated to the $K-1$ previous activities. It should be noted that $\phi_K(u_{K-1}, u_K)$ may not be monotonically increasing so that the minimal cost is not necessarily obtained by allocating zero quantity of resources. For $j = K-1$, since the value of $\Phi_K^*(u_{K-1})$ is already known for each state u_{K-1} , the value of $\Phi_{K-1}^*(u_{K-2})$ for a given state u_{K-2} is then obtained by

(2.3)

$$\Phi_{K-1}^*(u_{K-2}) = \min_{v_{K-1} \in D_{K-1}(u_{K-2})} \{ \phi_{K-1}(u_{K-2}, u_{K-1}) + \Phi_K^*(u_{K-1}) | u_{K-1} = \tau_{K-1}(u_{K-2}, v_{K-1}) \}$$

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

$$(2.4) \quad \Phi_j^*(u_{j-1}) = \min_{v_j \in D_j(u_{j-1})} \{ \phi_j(u_{j-1}, u_j) + \Phi_{j+1}^*(u_j) | u_j = \tau_j(u_{j-1}, v_j) \}$$

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 minimal cost of allocating the quantity Q of resources to the K activities. It then remains to obtain the optimal allocation to each activity j , say v_j^* , for which the optimal value $\Phi_1^*(u_0)$ has been found. 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 $\underline{v}^* = (v_1^*, \dots, v_K^*)$ 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 K th stage. By following this process, the problem of allocation of resources is found to be completely solved since the minimal cost is now known as well as the optimal quantity to be allocated to each activity.

The process just described forms the basis of the DP method. 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) and Nedzela (1981). One important remark made by Hadley (1964) about DP problems is that their nature is unchanged regardless of how many stages there are. This remark will be found to be important in the application of DP to the determination of optimum stratification.

2.3 Advantages and Disadvantages

One of the main advantages of the DP approach is its flexibility. It can in fact handle any optimization problem that can be put into the form given by (2.1). It should be noted that the summation in the objective function of (2.1) could

be changed to multiplication. In general, once an optimization problem has been expressed as a Functional Equation of Bellman (i.e. in a recursive fashion), the problem can usually be solved by DP. Another advantage of the DP approach is that, as noted by Bellman and Dreyfus (1965), the method always gives the absolute minimum (or maximum if the problem is one of maximization) and not a local minimum. Obtaining a local minimum using the DP approach would directly contradict Bellman's Principle of Optimality. A third advantage of DP is that it does not require any convexity assumption about the objective function given by (2.1) unlike most of the well-known non-linear optimization methods. Bellman and Dreyfus (1965), however, gives a way where the assumption of convexity can be used to increase the speed of the DP process. The last advantage of DP is that, even without making any assumption on the objective function, the DP approach turns out to be faster than direct enumeration of the different possibilities. Assuming that each of the K state spaces U_j contains N elements, the direct enumeration of the different possibilities would require computations of order N^K . However, because of the use of Bellman's Principle of Optimality, the DP approach would only require computations of the order KN . Hence, the number of computations increases linearly with the number K of stages for DP while it increases exponentially for direct enumeration.

One disadvantage related to the use of DP is that it cannot directly handle the cases where the state and decision spaces are infinite. This is mainly because, during the process, the value of $\Phi_j^*(u_{j-1})$ has to be computed for each u_{j-1} of the state space U_{j-1} . Hence, for example, if U_{j-1} is an interval of the real space, the computation of $\Phi_j^*(u_{j-1})$ for each u_{j-1} would be impossible. In the case where U_j is an interval, say $[a, b]$, Bellman and Dreyfus (1965) suggested that one cross the interval by discrete steps, i.e. by considering the values $u_{j-1} = a, a+\Delta, a+2\Delta, \dots, b-\Delta, b$ where $\Delta = \frac{b-a}{k}$ for some k . The value of Δ (or k) would then be chosen taking into account precision, computer space and time. It should be noted that perfectly continuous processes seldom occur in practice because of the finite precision of digital computers. Note also that this disadvantage with infinite spaces is related to the use of computers and not to the solution of the DP problem itself. A last disadvantage with the DP approach is that it requires the use of large quantities of space to store the values of the $\Phi_j^*(u_{j-1})$ for each stage and each state. However, with the increase in the computer power, this problem is becoming less and less important.

3. ONE-WAY OPTIMUM STRATIFICATION

The one-way optimum stratification certainly constitutes the basis for the optimum stratification with respect to two or more stratification variables. It may thus be worthwhile to give a detailed treatment of this case before considering other aspects of optimum stratification.

As in Chapter 1, a survey with a single variable of interest X is considered. It is assumed that X is sampled from a population of size N with CDF $F(x)$ defined on $[a, b]$ with mean μ and finite variance σ^2 . In theory, having the domain of X defined on a bounded interval may be considered as restrictive since many distributions such as the normal or the exponential have unbounded domains. However, in practice, those unbounded domains can usually be truncated to bounded ones so that the remaining part may be considered as negligible. Assuming that the population is to be divided into L strata, the problem considered in this chapter is to find OPS $\underline{x}^*$ (or $\underline{z}^*$ if the auxiliary variable Z is used for stratification instead of X) in order to minimize the variance of $\bar{x}_{st}$ given by equation (1.4).

Dalenius (1950) showed that the OPS generally depends on the type of allocation used for the total sample size n among the L strata. In the literature, the four basic types of allocation that are usually discussed are the proportional allocation, the Neyman allocation, the equal-sample-size allocation and the optimum allocation (with a cost constraint). Also called the representative allocation, the proportional allocation distributes the total sample size n according to the population proportion of each stratum. Formally,

$$(3.1) \quad n_h = W_h n, \quad h = 1, \dots, L$$

where W_h is given by (1.3).

Replacing the stratum sample size n_h in (1.4) by (3.1), the variance of $\bar{x}_{st}$ can then be expressed as

$$(3.2) \quad \text{Var}(\bar{x}_{st}) = \sum_{h=1}^L \frac{W_h \sigma_h^2}{n} - \sum_{h=1}^L \frac{W_h \sigma_h^2}{N}.$$

It should be noted that $\text{Var}(\bar{x}_{st})$ under proportional allocation becomes a function of the points of stratification $\underline{x}$ only. The Neyman allocation, named after Neyman (1934), distributes the total sample size in order to minimize the variance of $\bar{x}_{st}$. This allocation is given by

$$(3.3) \quad n_h = \frac{W_h \sigma_h}{\sum_{h=1}^L W_h \sigma_h} n, \quad h = 1, \dots, L$$

where W_h and σ_h^2 are given by (1.3) and (1.5), respectively.

Unfortunately, with this allocation it is possible to obtain $n_h > N_h$ which is undesirable. Cochran (1977), however, gives a method to deal with such a case. Assuming that $n_h \leq N_h$, the variance of $\bar{x}_{st}$ under Neyman allocation is then given by

$$(3.4) \quad \text{Var}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L W_h \sigma_h \right)^2 - \sum_{h=1}^L \frac{W_h \sigma_h^2}{N}$$

The third type of allocation is the equal-sample-size allocation which simply distributes the total sample size equally among the strata. Formally,

$$(3.5) \quad n_h = \frac{n}{L}, \quad h = 1, \dots, L.$$

This allocation gives equal work loads in each stratum. As with Neyman allocation, it is also possible to obtain $n_h > N_h$ with this allocation. Assuming that $n_h \leq N_h$, the variance of $\bar{x}_{st}$ under the equal-sample-size allocation is given by

$$(3.6) \quad \text{Var}(\bar{x}_{st}) = \frac{L}{n} \sum_{h=1}^L W_h^2 \sigma_h^2 - \sum_{h=1}^L \frac{W_h \sigma_h^2}{N}$$

Finally, the optimum allocation is the allocation obtained by minimizing the variance of $\bar{x}_{st}$ under a linear cost function of the form

$$(3.7) \quad C = c_0 + \sum_{h=1}^L c_h n_h.$$

It should be noted that c_0 is an overall cost that may depend on L . The allocation then obtained is given by

$$(3.8) \quad n_h = (C - c_0) \frac{W_h \sigma_h / \sqrt{c_h}}{\sum_{h=1}^L W_h \sigma_h \sqrt{c_h}}, \quad h = 1, \dots, L.$$

Assuming that $n_h \leq N_h$, the variance of $\bar{x}_{st}$ under optimum allocation is given by

$$(3.9) \quad \text{Var}(\bar{x}_{st}) = \frac{\left(\sum_{h=1}^L W_h \sigma_h \sqrt{c_h} \right)^2}{(C - c_0)} - \sum_{h=1}^L \frac{W_h \sigma_h^2}{N}$$

The OPS can be obtained under each of these four types of allocations. Although not discussed here, some authors have, however, considered other types

of allocation (see Hayashi, Maruyama and Ishida (1951) and Espejo (1985) for equiprecisional allocation and Lavallée (1985) for power allocation).

3.1 Infinite Population Case

In this section, the OPS are found for the case of infinite populations (i.e. $N = +\infty$) or for at least a large enough populations so that the f.p.c. can be ignored. It is also assumed that the stratification variable is the variable of interest, X , itself.

3.1.1 Literature Review: The problem of finding OPS for the infinite population case has first been considered by Hayashi and Maruyama (1948) and independently by Dalenius (1950). Assuming that the CDF $F(x)$ has a continuous density $f(x)$, Dalenius (1950) obtained expressions for the OPS for both proportional and Neyman allocations by taking the partial derivatives of (3.2) and (3.4) respectively, and equating them to zero. These expressions have been called minimal equations. For proportional allocation, the minimal equations have been found to be

$$(3.10) \quad x_h = \frac{\mu_h + \mu_{h+1}}{2}, \quad h = 1, \dots, L-1.$$

On the other hand, the minimal equations for the Neyman allocation are given by

$$(3.11) \quad \frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}, \quad h = 1, \dots, L-1.$$

Dalenius (1950) solved these minimal equations by an iterative trial-and-error method. Hayashi, Maruyama and Ishida (1951), following the work of Hayashi and Maruyama (1948), also obtained the minimal equations under proportional allocation. At the same time, they suggested the use of $\Delta_h = x_h - x_{h-1} = \text{constant}$ as a rule to compute approximate OPS for proportional as well as Neyman allocations. Dalenius and Gurney (1951), on the other hand, conjectured that the rule $W_h \sigma_h = \text{constant}$ would provide approximate OPS in the case of Neyman allocation. As a footnote, Mahalanobis (1952) proposed that "(...) an optimal or nearly optimum solution (to the minimal equations) would be obtained when the expected contribution of each stratum to the total aggregate value of x is made equal for each stratum". This rule, which can be expressed as $W_h \mu_h = \text{constant}$, has also been suggested by Hansen, Hurwitz and Madow (1953). In analyzing equation (3.11), Aoyama (1954) proposed as did Hayashi, Maruyama and Ishida (1951), the rule $\Delta_h = x_h - x_{h-1} = \text{constant}$. Kitagawa (1956) analyzed the principle of equipartition in order to justify the rule $W_h \mu_h = \text{constant}$. Dalenius and Hodges (1957) gave a rigorous proof that the rule $W_h \sigma_h = \text{constant}$ satisfies asymptotically the minimal equations given by (3.11). In their proof, they used the fact that making $\int_{x_{h-1}}^{x_h} \sqrt{f(x)} dx = \text{constant}$ would be asymptotically equivalent to $W_h \sigma_h = \text{constant}$. Commonly referred as the cum $\sqrt{f}$ rule, this rule

has been formally stated and also extended by Dalenius and Hodges (1959). As another attempt to solve the minimal equations (3.11), Ekman (1959) suggested the rule $W_h(x_h - x_{h-1}) = \text{constant}$ to obtain approximate OPS when the distribution of X is not highly skewed. Durbin (1959), in a review of Dalenius' doctoral thesis, proposed the use of equal intervals on the cumulative of $\frac{F(x)}{b-a} + f(x)$ or, in other words, the cum $\left(\frac{F}{b-a} + f\right)$ rule. In a well-known paper, Cochran (1961) compared empirically the cum $\sqrt{f}$ rule, the rule $W_h\mu_h = \text{constant}$, the rule $W_h(x_h - x_{h-1}) = \text{constant}$ and the cum $\left(\frac{F}{b-a} + f\right)$ rule, using actual skewed populations. It was found in this study that, under Neyman allocation, good results in terms of variance minimization can be obtained with both cum $\sqrt{f}$ rule or the rule $W_h(x_h - x_{h-1}) = \text{constant}$. As pointed out by Cochran (1961), this last rule of Ekman (1959) can, however, be a little troublesome computationally to apply. Sethi (1963) gave an iterative method, based on successive approximations, to solve the minimal equations for proportional, Neyman and equal-sample-size allocations. For the equal-sample-size allocation, Sethi (1963) obtained the following minimal equations:

$$(3.12) \quad W_h(\sigma_h^2 + (x_h - \mu_h)^2) = W_{h+1}(\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2) \quad , h = 1, \dots, L-1.$$

Serfling (1968), using the cum $\sqrt{f}$ rule, computed the optimal number of strata based on a given cost function. His results agreed with some conclusions that were made by Dalenius (1957). Moving away from the use of the minimal equations, Schneeberger (1971) expressed the problem of determining the OPS as a nonlinear programming problem. Along the same lines, Bühler and Deutler (1975) proposed the DP approach to obtain the OPS. The various possibilities and the simplicity of their approach were the principal motivation for the development of this thesis. Oslo (1976), dissatisfied with the cum $\sqrt{f}$ rule for proportional allocation, proposed the cum $\sqrt[3]{f}$ rule. Unnithan (1978) suggested the use of Shanno's iterative method to obtain the OPS. This method developed by Shanno (1970) is in fact a modification of Newton's method. The ASOS17-Algorithm, which is a 17-step computer algorithm to obtain approximate OPS under Neyman allocation, has been proposed by Wienhöld (1982). Wang and Aggarwal (1984), noting that a large number of skewed populations can be approximated by a simple Pareto distribution, provided a graphical method to obtain immediate OPS under Neyman allocation. In a further study on Pareto distributions, Aggarwal and Singh (1984) showed that the cum $\sqrt[3]{f}$ rule proposed by Oslo (1976) yields the exact OPS under a certain class of Pareto distributions. Under this same class of distributions, they also proved that the rule of Mahalanobis (1952) given above is true. Finally, more

recently, Espejo (1985) and Lavallée (1985) obtained the minimal equations under the equiprecisional allocation and the power allocation, respectively.

3.1.2 Identification of Problems: Before the beginning of the 1970's, it was seen that the methods proposed to obtain the OPS were mainly based on solving the minimal equations given by (3.10), (3.11) or (3.12). However, as noted by Kpedekpo (1973), these minimal equations are only necessary but not sufficient conditions to obtain a minimum for the variance of $\bar{x}_{st}$. In fact as early as 1958, Fisher (1958) gave an example of a distribution for X giving several relative minima for different points of stratification $\underline{x}$. Schneeberger (1979) showed with an example that the cum $\sqrt{f}$ rule of Dalenius and Hodges (1957) could lead to a saddle-point. He also detected the same problem with the iterative method of Unnithan (1978). On the other hand, Goller (1981) obtained an example of a simple symmetric distribution where the center of symmetry was not the OPS (in the case of two strata ($L = 2$)) which contradicted the rule of $W_h \sigma_h = \text{constant}$ conjectured by Dalenius and Gurney (1951). Schneeberger (1985) gave an example of a mixture of normal distributions where the minimal equations would be satisfied under a local minimum, a saddle-point or a local maximum. From these studies, it should be clear that solving the minimal equations is not sufficient to obtain the OPS. Hence the use of a method that will ensure a global minimum is required. As noted in Chapter 2, the method of DP proposed by Bühler and Deutler (1975) can achieve this goal.

3.1.3 The Dynamic Programming Approach: Following Bühler and Deutler (1975), the method of DP can be easily applied to the optimum stratification problem in the case of an infinite population. The first step in applying this method is to identify the objective function to minimize in order to express the problem as in (2.1) of Chapter 2. For the optimum stratification problem, the objective function can be expressed in the following form:

$$(3.13) \quad G(\underline{x}) = \sum_{h=1}^L g(x_{h-1}, x_h),$$

where the function $g(x_{h-1}, x_h)$ depends in fact on the choice of allocation that is to be used in the sampling design. For the proportional allocation, for example, the objective function is obtained from the variance of $\bar{x}_{st}$ given by (3.2) with the f.p.c. ignored. The function $g(x_{h-1}, x_h)$ for the proportional allocation is then given by

$$(3.14) \quad g(x_{h-1}, x_h) = W_h \sigma_h^2, \quad h = 1, \dots, L.$$

For the Neyman allocation, since the f.p.c. is ignored, it is clear that minimizing the term $\left(\sum_{h=1}^L W_h \sigma_h \right)^2$ of the variance of $\bar{x}_{st}$ given by (3.4) is the same as minimizing $\sum_{h=1}^L W_h \sigma_h$. Hence the function $g(x_{h-1}, x_h)$ for the Neyman allocation is

given by

$$(3.15) \quad g(x_{h-1}, x_h) = W_h \sigma_h, \quad h = 1, \dots, L.$$

The functions $g(x_{h-1}, x_h)$ for the equal-sample-size and optimum allocations can be obtained in a similar manner from (3.6) and (3.9), respectively.

With respect to the different concepts used in DP, the stages here clearly correspond to the different strata. The sequential decision aspect of DP can in fact be seen as distributing to one stratum at a time a portion of the domain of X , i.e. the interval $[a, b]$. With stratum h corresponding to $[x_{h-1}, x_h]$, the point of stratification x_h indicates the portions of $[a, b]$ distributed up to stratum h so that the variable x_h corresponds to the state variable u_h of Chapter 2. The initial state u_0 is then clearly given by $x_0 = a$. Since each point of stratification can in theory be any point of $[a, b]$, the state space of each stratum simply corresponds to $[a, b]$. The decision variable v_h in the present problem can be seen as the width, say d_h , given to the interval of stratum h . For stratum h , it is then clear that $x_h = x_{h-1} + d_h$ which corresponds to the stage transformation function. Since the domain of X is bounded, the variable d_h is restricted within the decision space $B_h(x_{h-1}) = [0, b - x_{h-1}]$. Finally, the optimum stratification problem for infinite populations can be expressed in the same form as (2.1) by

$$(3.16) \quad \begin{aligned} & \text{Min} \sum_{h=1}^L g(x_{h-1}, x_h) \\ & \text{subject to: } x_h = x_{h-1} + d_h \\ & \quad x_h \in [a, b] \\ & \quad d_h \in B_h(x_{h-1}) = [0, b - x_{h-1}] \\ & \quad x_0 = a, \quad h = 1, \dots, L. \end{aligned}$$

It was seen in Chapter 2 that the resolution of an optimization problem by DP requires the use of the Functional Equation of Bellman given by (2.4). Having expressed the present problem as a DP problem of the form of (2.1), as the second step let $G_h^*(x_{h-1})$ be the optimal value of the objective function for strata h to L given that the lower bound of stratum h is x_{h-1} . Following Chapter 2, the Functional Equation of Bellman for the optimum stratification problem for infinite populations is then given by

$$(3.17) \quad G_h^*(x_{h-1}) = \text{Min}_{d_h \in B_h(x_{h-1})} \{g(x_{h-1}, x_h) + G_{h+1}^*(x_h) | x_h = x_{h-1} + d_h\}$$

Bühler and Deutler (1975) stated, as their Result 1, that the solution obtained from the DP formulation (3.16) of the optimum stratification problem for infinite population would in fact give the OPS. 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 (3.17), the DP approach can be implemented easily. A difficulty, however, is brought by the fact that the state and decision spaces are infinite. This problem can fortunately be overcome by using the discrete step approach discussed in Section 2.3.

One of the main advantages of using the DP approach for optimum stratification comes from the possibility of determining the optimal number of strata. Since with DP the nature of the problem is unchanged regardless of how many stages there are, the optimum stratification can then be obtained for $L = 2, 3, 4, \dots$ strata until the difference between the optimal values $G_h^*(a)$ and $G_{h+1}^*(a)$ be less than a preassigned level, i.e.

$$(3.18) \quad |G_h^*(a) - G_{h+1}^*(a)|/G_{h+1}^*(a) < \epsilon \text{ for some } \epsilon > 0.$$

It should be noted that this process can be performed in a single run where L then refers to the maximum permitted number of strata. Note also that in $G_h^*(a)$, the value a is given to the point of stratification x_{h-1} so that all the previous points of stratification are equal to a , i.e. $a = x_0 = x_1 = \dots = x_{h-1}$.

3.1.4. Dynamic Programming with Real Populations: In practice, it often occurs that the population of the variable of interest X is a real population in the sense that it is a set $\{x\}$ of measured values. Methods like the one of Unnithan (1978) can hardly be applied since they require the knowledge of the density $f(x)$. Fortunately, the DP approach can, however, easily be used in such a case.

Let the variable of interest X be measured from the real population $\{x\}$ of N units. The use of the DP approach is then simply made by replacing the quantity W_h, μ_h and σ_h^2 of equations (1.3), (1.6) and (1.5), respectively, by their finite population counterparts:

$$(3.19) \quad W_h : \quad \frac{N_h}{N} \approx \frac{1}{N} \{\text{number of } x\text{'s in } I_h\}$$

$$(3.20) \quad \mu_h : \quad \bar{X}_h = \frac{1}{N_h} \sum_{x \in I_h} x$$

$$(3.21) \quad \sigma_h^2 : \quad S_h^2 = \frac{1}{N_h - 1} \sum_{x \in I_h} (x - \bar{X}_h)^2, \quad h = 1, \dots, L$$

where $I_h = [x_{h-1}^+, x_h]$ and $a = \min\{x\} = x_0 \leq x_1 \leq \dots \leq x_{L-1} \leq x_L = \max\{x\} = b$. Assuming that the population size N is large enough, compared to the total sample size n and the number L of strata, so that the f.p.c. can be ignored, the objective function keeps the same form as in (3.13) except that the quantity W_h, μ_h, σ_h^2 are replaced by $\frac{N_h}{N}, \bar{X}_h$ and S_h^2 , respectively. For example, in the case

of Neyman allocation, $g(x_{h-1}, x_h)$ is here given by

$$g(x_{h-1}, x_h) = \frac{N_h}{N} S_h, \quad h = 1, \dots, L.$$

The relationships with the concepts of the general DP approach are basically the same as the ones described in Section 3.1.3. The only difference may be that the state and decision spaces are, with the real population $\{x\}$, discrete sets. The state space which was $[a, b]$ in Section 3.1.3 now becomes the real population $\{x\}$ while the decision space $B_h(x_{h-1})$ is now given by $\{d | d \geq 0 \text{ and } d = x - x_{h-1} \text{ for } x \in \{x\}\}$. It should also be noted that the initial state becomes $x_0 = \min\{x\}$. Finally, the Functional Equation of Bellman remains of the same form as equation (3.17) except that the various quantities are replaced by their finite population counterparts.

One of the nice aspects of using the DP approach with real populations is that the state and decision spaces are discrete. The optimal value $G_h^*(x_{h-1})$ of the objective function for strata h to L given the lower bound x_{h-1} for stratum h can, in this case, be computed for each possible value of x_{h-1} (i.e. the complete population $\{x\}$) and for each stratum h . It should be noted that if the complete real population is too large to be handled rapidly by the computer, the population can be sorted and crossed by considering, for example, steps of two units instead of only one. This has in fact essentially the same effect as taking a systematic sample of the ordered population. Finally, as in Section 3.1.3, the optimal number of strata can still be easily determined.

3.1.5 Computer Program and Illustrations: In order to be able to actually compute the OPS in the case of an infinite population, the program OPTIBND1 has been written. A listing of this program can be found in Appendix A. The program, as all the other programs contained in this thesis, has been written in version 3.0 of Turbo Pascal which can be almost considered as standard Pascal. The Pascal language has been chosen mainly because it permits recursive statements so that the Functional Equation of Bellman can be programmed directly. Implemented on an IBM PC compatible micro computer, OPTIBND1 involves basically five steps. It first reads the data (the real population $\{x\}$, for example) from an external file. Secondly, it sorts the data by increasing order of the variable X using the procedure SORTSHLL. Thirdly, it selects a systematic sample from the data if the number of observations is too large to be handled by the user's computer. Next, it computes the OPS using the procedure OPTIBND. Finally, it displays the OPS with other descriptive statistics. The program OPTIBND1 can handle the case of a population with an hypothetical CDF $F(x)$ as well as the case of a real population. The latter case is in fact the default one where the finite population statistics $\frac{N_h}{N}$, $\bar{X}_h$ and S_h^2 given by (3.19), (3.20) and (3.21), respectively, are computed from the input data. The default allocation is the Neyman allocation but proportional, equal-sample-size and optimum allocations can also easily be considered. It should

be noted that the program computes only a single vector of OPS $\mathbf{x}^*$ even when several are possible. Finally, it may be added that OPTIBND1 can be considered as fast, taking into account that it has been implemented on a micro computer.

A first illustration of the use of OPTIBND1 employs the same distribution used by Bühler and Deutler (1975), for their first example. The distribution is symmetric and can be called, because of the shape of its density, an overlapping-triangle distribution. A plot of the density function of this distribution together with some descriptive statistics can be found in Figure 3.1.

Figure 3.1

The Overlapping-Triangle Distribution.

$$\text{density } f(x) = \begin{cases} \frac{49}{72}(x+1), & -1 \leq x \leq -1/7 \\ \frac{1}{3}(49x + \frac{35}{4}), & -1/7 < x \leq 0 \\ \frac{1}{3}(-49x + \frac{35}{4}), & 0 < x \leq 1/7 \\ \frac{49}{72}(-x+1), & 1/7 < x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

mean $\mu = 0.000$

variance $\sigma^2 = 0.114$

skewness = 0.000

kurtosis = 0.460

minimum value = -1.000

maximum value = 1.000

The results obtained using the program OPTIBND1 correspond exactly to the ones of Bühler and Deutler (1975). It should be noted that since the domain of X is continuous, the interval $[-1,1]$ has been crossed by discrete steps of 0.01. These results are presented in Table 3.1 for Neyman Allocation.

Table 3.1

Optimum Points of Stratification x^* for the Overlapping-Triangle Distribution, under Neyman Allocation. The Points $\hat{x}$ are Obtained from the Cum $\sqrt{f}$ Rule.

Stratum	h	1	2	3	4	$\text{Var}(\bar{x}_{st})/\text{Var}(\bar{x})$	marginal gain of precision (in %)
$L = 2$	x_h^*	0.15				0.463	53.7
	W_h	0.754					
	n_h/n	0.786					
	$\hat{x}_h$	0.00				0.488	
	W_h	0.500					
	n_h/n	0.500					
$L = 3$	x_h^*	-0.20	0.20			0.149	67.8
	W_h	0.218	0.564				
	n_h/n	0.314	0.371				
	$\hat{x}_h$	-0.17	0.17			0.152	
	W_h	0.234	0.531				
	n_h/n	0.348	0.304				
$L = 4$	x_h^*	-0.21	0.10	0.40		0.102	31.5
	W_h	0.212	0.498	0.167			
	n_h/n	0.366	0.334	0.139			
	$\hat{x}_h$	-0.32	0.00	0.32		0.106	
	W_h	0.157	0.343	0.343			
	n_h/n	0.229	0.271	0.271			
$L = 5$	x_h^*	-0.43	-0.12	0.12	0.43	0.064	37.3
	W_h	0.111	0.157	0.465	0.157		
	n_h/n	0.173	0.165	0.323	0.165		
	$\hat{x}_h$	-0.41	-0.08	0.08	0.41	0.068	
	W_h	0.118	0.201	0.362	0.201		
	n_h/n	0.187	0.226	0.175	0.226		

By looking at Table 3.1, it can be seen that for $L = 2$ strata, the OPS is not the point of symmetry of the distribution. This corroborates the result of Goller

(1981). This table also shows that the gain in increasing the number of strata is large up to $L \leq 3$ but decreases rapidly as L increases beyond 3. In fact, using criterion (3.18) with, for example, $\epsilon = 0.35$, the optimal number of strata would be $L = 4$. The choice of criterion (3.18) as a stopping rule for the optimal number of strata may, however, be questioned since the marginal gain of precision is greater for $L = 5$ than for $L = 4$. A criterion based on the decrease in the variance of $\bar{x}_{st}$ compared to the one with no stratification might be more appropriate. In comparing the use of DP with respect to the cum $\sqrt{f}$ rule of Dalenius and Hodges (1957), it is found that the former gives better results for small number of strata but, as L increases, the difference becomes small. However, it is expected that with skewed populations, the superiority of the DP approach will become more evident.

The second illustration of the use of OPTIBND1 employs population data on the sales of Incorporated Tobacconist Retailers in Canada for the year 1985. This highly skewed population of $N = 185$ units can be considered as large enough to ignore the f.p.c. as long as the total sample size n and the number of strata L are relatively small compared to N . For example, up to $L = 5$, n may be assumed to be, say, 20. An histogram of this population together with descriptive statistics can be found in Figure 3.2.

Figure 3.2

Distribution of the Sales of Incorporated Tobacconist Retailers in Canada for the Year 1985 (in Thousands of Dollars.)

population size $N = 185$

mean $\bar{X} = 521.0$

variance $S^2 = 213171.9$

skewness = 2.301

kurtosis = 7.484

minimum value = 1

maximum value = 2875

Using the program OPTIBND1, the OPS have been computed under Neyman allocation up to $L = 5$. These results can be found in Table 3.2.

Table 3.2

Optimum Points of Stratification $\underline{x}^*$ for the Sales of Incorporated Tobacconist Retailers in Canada for the Year 1985 ($N = 185$), under Neyman Allocation (f.p.c. ignored). The Points $\hat{\underline{x}}$ are Obtained from the Cum $\sqrt{f}$ Rule Applied to a Frequency Table of 15 Cells.

Stratum	h	1	2	3	4	$\text{Var}(\bar{x}_{st})/\text{Var}(\bar{x})$	marginal gain of precision (in %)
$L = 2$	x_h^*	595				0.363	63.7
	W_h	0.660					
	n_h/n	0.384					
	$\hat{x}_h$	575				0.364	
	W_h	0.649					
	n_h/n	0.369					
$L = 3$	x_h^*	466	1066			0.137	62.3
	W_h	0.578	0.357				
	n_h/n	0.472	0.319				
	$\hat{x}_h$	383	767			0.186	
	W_h	0.438	0.362				
	n_h/n	0.229	0.207				
$L = 4$	x_h^*	300	579	1066		0.079	42.3
	W_h	0.378	0.276	0.281			
	n_h/n	0.261	0.157	0.296			
	$\hat{x}_h$	192	575	958		0.095	
	W_h	0.227	0.422	0.254			
	n_h/n	0.091	0.324	0.175			
$L = 5$	x_h^*	300	567	875	1433	0.051	35.4
	W_h	0.378	0.265	0.243	0.076		
	n_h/n	0.319	0.174	0.208	0.141		
	$\hat{x}_h$	192	383	575	1150	0.060	
	W_h	0.227	0.211	0.211	0.287		
	n_h/n	0.117	0.109	0.110	0.341		

Again, as for the overlapping-triangle distribution, it is found that the marginal gain of precision decreases rapidly as the number L increases. Based on criterion (3.18) with, for example, $\epsilon = 0.36$, Table 3.2 shows that the optimal number of strata would be $L = 5$. The comparison of the DP approach with the use of the cum $\sqrt{f}$ rule again shows that the DP approach provides a more accurate determination of the OPS. Note that the cum $\sqrt{f}$ rule has been used by dividing the population into 15 cells based on Sales. As expected, the gain in efficacy of using the DP approach compared to the cum $\sqrt{f}$ rule with this skewed population is much more evident than with the overlapping-triangle distribution for $L \geq 3$.

3.2 Finite Population Case

In this section, the problem of finding OPS is considered when the population size N is small enough compared to the total sample size n and the number of strata L so that the f.p.c. cannot be ignored in the variance of $\bar{x}_{st}$. As in Section 3.1, it is assumed that the stratification variable is the same as the variable of interest X .

3.2.1 Literature Review: The finite population case has not received much attention in the literature on optimum stratification. However, in his first paper on optimum stratification, Dalenius (1950) stated that "the appearance of a finite population correction (in the variance of $\bar{x}_{st}$) will to some degree influence the choice of the optimum stratification point". On the other hand, Dalenius and Gurney (1951), after obtaining the minimal equations for the OPS under Neyman allocation, in the finite population case, and by considering three different distributions, concluded that the effect of the f.p.c on the OPS is small. They, however, added that a problem may occur in using the minimal equations if the overall sampling fraction $q = \frac{n}{N}$ is large enough so that a stratum sample size n_h becomes greater than the corresponding stratum population size N_h . Dalenius (1952) explicitly considered the case where the population is to be divided into $L = 2$ strata with one of the strata sampled with certainty, i.e. with $n_h = N_h$. As he pointed out, such sampling designs are of practical interest when sampling from a skewed population, such as a population of business establishments. It should be noted that the stratum with $n_h = N_h$ is commonly called a take-all stratum while the other stratum is called a take-some stratum. Setting the take-all stratum as the subinterval $[x_1^+, b]$ of $[a, b]$, Dalenius (1952) obtained the minimal equation related to this type of design:

$$(x_1 - \mu_1)^2 = \frac{W_1 \sigma_1^2}{\frac{n}{N} - 1 + W_1}.$$

Note that no special allocation is required for this type of design since the sample size for the take-some stratum is simply $n - N_2$, where N_2 is the size of the take-all stratum. Using a different approach, Glasser (1962) obtained essentially

the same result as Dalenius (1952) but using finite population arguments. Bühler and Deutler (1975), on the other hand, proposed the use of DP to solve the problem of optimum stratification under Neyman allocation for finite populations. In considering the problem of minimizing the total sample size n for a given variance of $\bar{x}_{st}$, Hidioglou (1986) used the same approach as Glasser (1962) to obtain the OPS x_1^* between the take-all and the take-some strata. Finally, Lavallée and Hidioglou (1986) derived the minimal equations for a sampling design with one take-all stratum and several take-some strata under the X -proportional power allocation.

3.2.2 Identification of Problems: As stated earlier, Dalenius and Gurney (1951) noticed that some problems may occur in using the minimal equations to obtain the OPS when the overall sampling fraction $q = \frac{n}{N}$ is large enough so that $n_h > N_h$ for some strata. Besides the fact that the minimal equations are necessary but not sufficient conditions, Schneeberger and Goller (1979) showed, using the log-normal distribution, that the minimal equations for $L = 2$ strata can lead to non-feasible solutions if the overall sampling fraction q is greater than a critical sampling fractions q_c . The value q_c depends on the CDF of the population of the variable X . It should be noted that this problem is relevant for Neyman, equal-sample-size or optimum allocation but not for proportional allocation where n_h is always less than N_h . Schneeberger (1981) gave, for the log-normal distribution, the critical sampling fractions q_c for up to $L = 10$ in the case of Neyman allocation. Based on these studies, it should be clear that the method of obtaining OPS in the case of finite populations must consider the possibility of obtaining $n_h > N_h$ during the computations. Fortunately, the DP approach of Bühler and Deutler (1975) can deal with this possibility.

3.2.3 The Dynamic Programming Approach: As in the method proposed by Bühler and Deutler (1975), a two-dimensional DP approach is needed in order to solve the problem of optimum stratification for Neyman allocation. The first dimension is the one related to the determination of the OPS while the second dimension refers to the optimal allocation of the total sample size n in order to minimize the variance of $\bar{x}_{st}$. (This should not be confused with the optimum allocation under a cost constraint given by (3.8).) Recalling the definition of Neyman allocation, the second dimension directly permits one to achieve this goal and, at the same time, allows the use of constraints to insure that $n_h \leq N_h$ for all stratum h . Moreover, by using a discrete state space for this dimension, it is then possible to obtain integer stratum sample size n_h , as is required in survey sampling. For the equal-sample-size allocation where for a given number of strata L the stratum sample size n_h is fixed at $n_h = \frac{n}{L}$, the use of a second dimension is not required. In fact, the approach given in Section 3.1 can be used except with a modification of the state space $B_h(x_{h-1})$ such that each stratum contains at least $\frac{n}{L}$ units. On the other hand, the optimum allocation with the cost constraint (3.7) can be easily handled by the use of the second dimension. The only difference from Ney-

man allocation is that instead of obtaining the stratum sample sizes n_h such that $\sum_{h=1}^L n_h = n$, with the optimum location, the stratum sample sizes are obtained so that $\sum_{h=1}^L c_h n_h = C - c_0$ according to (3.7) for given budget.

As done previously in Section 3.1, the first step in solving the problem by DP is to identify the objective function to be minimized in order to express the problem as in (2.1). For the optimum stratification problem in the case of finite populations, the objective function can be expressed in the form:

$$(3.22) \quad G(\underline{x}, \underline{n}) = \sum_{h=1}^L g(x_{h-1}, x_h, n_h)$$

where $\underline{n} = (\bar{n}_1, \dots, n_L).$

The vector of sample size $\underline{n}$ is subject to the constraint $\sum_{h=1}^L n_h = n$ for Neyman allocation, and also subject to the constraint that each n_h must be an integer less than or equal to the stratum population size N_h . Formally, this last constraint can be written as $n_h \leq [W_h N]$ where $[A]$ denotes the greatest integer less or equal to A . In determining the optimum stratification such that the variance of $\bar{x}_{st}$ is minimized, the function $g(x_{h-1}, x_h, n_h)$ of (3.22) is simply composed of each term of the variance of $\bar{x}_{st}$ itself, given in equation (1.4). That is, the function $g(x_{h-1}, x_h, n_h)$ is given by:

$$(3.23) \quad g(x_{h-1}, x_h, n_h) = W_h \sigma_h^2 \left(\frac{W_h}{n_h} - \frac{1}{N} \right)$$

Although the process should create a take-all stratum, if one is required, to minimize the variance of $\bar{x}_{st}$, a sample design which sets one stratum as a take-all and the others as take-somes can also be handled by the two dimensional DP approach. Assuming that the L th stratum is to be the take-all one, the objective function can be obtained by setting the L th term to 0 in the variance of $\bar{x}_{st}$ given by (1.4). The function $g(x_{h-1}, x_h, n_h)$ for a sampling scheme with a take-all stratum is then given by:

$$(3.24) \quad g(x_{h-1}, x_h, n_h) = \begin{cases} W_h \sigma_h^2 \left(\frac{W_h}{n_h} - \frac{1}{N} \right) & \text{for } h = 1, \dots, L-1 \\ 0 & \text{for } h = L. \end{cases}$$

The restriction that $n_L = W_L N$ has also to be included to the set of constraints applying to the vector $\underline{n}$.

In regard to the different concepts used in the DP theory, the stages again correspond to the different strata. However, contrarily to the DP approach used in Section 3.1, the state and decision spaces are both two-dimensional. The sequential decision aspect of DP can be seen as distributing to one stratum at a time a portion of the domain of $X, [a, b]$, and, for each portion distributed, as allocating

the stratum sample sizes subject to the given constraints. Since the point of stratification x_h indicates the portion of $[a, b]$ distributed up to stratum h , the first dimension of the state of stratum h can then be given by x_h . Also, letting m_h be the sample size allocated up to stratum h , the two-dimensional state of stratum h is obtained as (x_h, m_h) . Note that the initial state is simply given by $(a, 0)$. Since the stratum sample sizes n_h have to be integers, the state space for each stratum h corresponds to $[a, b] \times \{0, 1, 2, \dots, n\}$. As for the state variable, the decision variable is found to be two-dimensional and is given, in the present problem, as (d_h, n_h) where d_h is defined as the width of the interval $[x_{h-1}, x_h]$ of stratum h . The state transformation function τ_h of (2.1) is then given by $(x_h, m_h) = (x_{h-1} + d_h, m_{h-1} + n_h)$ for each stratum h . Finally, to satisfy the different constraints associated with the points of stratification $\underline{x}$ and the sample sizes $\underline{n}$, the decision variable (d_h, n_h) is, for the design without any take-all strata, restricted to the decision space $B_h^{(1)}(x_{h-1}) \times B_h^{(2)}(x_{h-1}, x_h, m_{h-1})$ where $B_h^{(1)}(x_{h-1}) = [0, b - x_{h-1}]$ and $B_h^{(2)}(x_{h-1}, x_h, m_{h-1}) = \{1, 2, \dots, n - m_{h-1}\} \cap \{1, 2, \dots, [W_h N]\}$. The problem of optimum stratification in the case of finite populations can then be put into the form of (2.1) (assuming that the variable u_h of (2.1) is a vector of two components) as

$$(3.25) \quad \text{Min} \sum_{h=1}^L g(x_{h-1}, x_h, n_h)$$

subject to: $(x_h, m_h) = (x_{h-1} + d_h, m_{h-1} + n_h)$
 $(x_h, m_h) \in [a, b] \times \{0, 1, 2, \dots, n\}$
 $(d_h, n_h) \in B_h^{(1)}(x_{h-1}) \times B_h^{(2)}(x_{h-1}, x_h, m_{h-1})$
 $(x_0, m_0) = (a, 0) \quad h = 1, \dots, L.$

Having expressed the problem in the form of (2.1), the next step is to obtain the Functional Equation of Bellman. Let $G_h^*(x_{h-1}, m_{h-1})$ be the optimal value of the objective function for strata h to L given that the lower bound of stratum h is x_{h-1} and $n - m_{h-1}$ units are allocated to strata h to L . Following the method of equation (2.4), the Functional Equation of Bellman for the optimum stratification problem for finite populations is given by:

$$(3.26) \quad G_h^*(x_{h-1}, m_{h-1}) = \text{Min}_{d_h \in B_h^{(1)}(x_{h-1})} \text{Min}_{n_h \in B_h^{(2)}(x_{h-1}, x_h, m_{h-1})} \{g(x_{h-1}, x_h, n_h) + G_{h+1}^*(x_h, m_h) | x_h = x_{h-1} + d_h, m_h = m_{h-1} + n_h\}$$

Bühler and Deütler (1975) showed, as their Result 2, that the solution obtained by the DP formulation (3.25) of the optimum stratification problem for finite populations is in fact the OPS $\underline{x}^*$ and the allocation $\underline{n}^*$ corresponding to Neyman allocation. Like the approach given in Section 3.1, this two-dimensional approach

of DP can also be easily implemented on a computer using equation (3.26). The technique of crossing the interval $[a, b]$ by discrete steps has to be employed because of the continuous aspect of the first dimension of the problem. On the other hand, since the second dimension related to the stratum sample sizes is discrete, the DP approach can in this dimension, be applied directly.

An advantage of using the DP approach is that the number of strata does not have to be specified. The optimal number of strata can then be obtained according to the criterion of stopping the increase in the number L of strata whenever the difference between the objective functions becomes negligible, i.e.

$$(3.27) \quad |G_h^*(a, 0) - G_{h+1}^*(a, 0)| / G_{h+1}^*(a, 0) < \epsilon \text{ for some } \epsilon > 0.$$

The main disadvantage of using the two-dimensional approach is that because of the introduction of this second dimension, the process of obtaining the OPS can be considerably slowed down, especially if the total sample size n is large. A possible solution may be to allocate two or more units at a time to the strata.

3.2.4 Dynamic Programming with Real Populations: Especially when the real population $\{x\}$ under consideration is small, the determination of the density $f(x)$ can become very difficult so that methods like the one proposed by Unnithan (1978) can hardly be applied. The possibility to rely on the finite population quantities $\frac{N_h}{N}$, $\bar{X}_h$, S_h^2 given by (3.19), (3.20) and (3.21), respectively, then becomes important.

As in Section 3.1.4, the DP approach is the same as the one described in Section 3.2.3 except that the quantities W_h , μ_h and σ_h^2 are replaced by their finite population counterparts. For example, the function $g(x_{h-1}, x_h, n_h)$ given by (3.23) then becomes

$$(3.28) \quad g(x_{h-1}, x_h, n_h) = \frac{N_h^2}{N^2} S_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \quad h = 1, \dots, L.$$

The relationships to the concepts of the general DP approach discussed in Chapter 2 are basically the same as in the previous section. The only differences are that the state space of each stratum h is now $\{x\} \times \{0, 1, \dots, n\}$ instead of $[a, b] \times \{0, 1, \dots, n\}$ and that the decision space $B_h^{(1)}(x_{h-1}) = [0, b - x_{h-1}]$ is replaced by $B_h^{(1)}(x_{h-1}) = \{d | d \geq 0 \text{ and } d = x - x_{h-1} \text{ for } x \in \{x\}\}$. It should be noted that the initial state is now given by $(\min\{x\}, 0)$ instead of $(a, 0)$.

The use of discrete state and decision spaces has the nice aspect that DP can then be applied directly without having to divide the domain of the variable X into discrete points. Also, the optimal number of strata can again be found, as in Section 3.2.3, using the criterion (3.27).

3.2.5 Computer Program and Illustrations: To compute the actual OPS x^* for the case of a finite population, the program OPTIBND2 has been written. This program which can be found in Appendix B has been implemented using, as for

OPTIBND1, the Pascal language. The program OPTIBND2 also involves basically the same five steps as OPTIBND1. First, it reads the data from an external file. This data can be a finite population $\{x\}$ or some discrete points of the domain of X , $[a, b]$. Secondly, it sorts the data by increasing order using the procedure SORTSHLL. Thirdly, if the volume of input data is too large to be handled by the user's computer, it selects a systematic sample. Next, it computes the OPS $\bar{x}^*$ as well as the optimal allocation n^* to minimize the variance of $\bar{x}_s$. Finally, it displays the results along with some descriptive statistics. The program OPTIBND2 can handle either the case of an hypothetical distribution $F(x)$ for the variable X or the case of a real population $\{x\}$. The latter case is the default one. As for OPTIBND1, the Neyman allocation is again the default allocation used by the program. Apart from the number L of strata and the total sample size n , the user can specify a minimal sample size to be allocated to each strata. For example, setting the minimal sample size to 2 gives the possibility of doing unbiased variance estimation. It should finally be noted that some results may take much more time to obtain compared to OPTIBND1 because of the use of the two-dimensional DP approach.

A first illustration of the use of the program OPTIBND2 is based on the assumption that the underlying distribution of X is the exponential distribution with parameter $\lambda = 2.0$. Since the domain of this distribution is $[0, +\infty[$, it was truncated to $[0, 10]$ to correspond to $[a, b]$. Note that the ignored portion $]10, +\infty[$ represents only 0.67% of the whole distribution. Figure 3.3 provides a plot of the density of X with some descriptive statistics.

Figure 3.3

The Exponential Distribution with parameter $\lambda = 2.0$.

Assuming that the population size N is 30 and that the total sample size n is 15 (so that $q = \frac{n}{N} = 50\%$), the OPS together with the optimal sample sizes n^* were computed using OPTIBND2. Since the domain of X is continuous, the interval $[0,10]$ is crossed by discrete steps of 0.1. A minimal sample size of 2 from each stratum has also been set. The results obtained under the above assumptions are reported in Table 3.3.

Table 3.3

Optimum Points of Stratification $\underline{x}^*$ and Optimal Sample Sizes $\underline{n}^*$ for the Exponential Distribution with $\lambda = 2.0$. (Here $N = 30$ and $n = 15$.) The Underlined Sample Size $\underline{n}_h$ Indicates that the Stratum is Take-All.

Stratum	h	1	2	3	4	5	$\text{Var}(\bar{x}_{st})/\text{Var}(\bar{x})$	marginal gain of precision (in %)
$L = 2$	x_h^*	0.4					0.173	82.7
	W_h	0.567						
	n_h	3	12					
$L = 3$	x_h^*	0.2	0.5				0.049	71.8
	W_h	0.333	0.300					
	n_h	2	2	<u>11</u>				
$L = 4$	x_h^*	0.2	0.4	0.7			0.024	50.8
	W_h	0.333	0.233	0.200				
	n_h	3	2	3	<u>7</u>			
$L = 5$	x_h^*	0.1	0.3	0.5	0.7		0.016	30.9
	W_h	0.167	0.267	0.167	0.133			
	n_h	2	2	2	2	<u>7</u>		

It is seen from Table 3.3 that, for this skewed population, only the case with $L = 2$ strata does not lead to the creation of a take-all stratum. Table 3.3 also shows that, for all the strata, the program OPTIBND2 has satisfied the constraints that the stratum sample size n_h should be less than or equal to the corresponding stratum population size N_h , and greater than or equal to the minimal sample size 2. By considering the optimal number of strata according to criterion (3.27), it can be seen that, with $\epsilon = 0.35$, the present example would lead to $L = 5$ strata.

As a second illustration, the distribution of the sales of the Incorporated Wholesalers of Grain in Canada for the year 1985 is considered as a population $\{x\}$ for the variable X . This population of $N = 80$ units, from which sales range from 3 to 307042 thousands of dollars, is found to be highly skewed. A histogram, together with some descriptive statistics, can be found in Figure 3.4.

Figure 3.4

Distribution of the Sales of the Incorporated Wholesalers of Grain in Canada
for the Year 1985 (in Thousands of Dollars).

population size $N = 80$

mean $\bar{X} = 16610.7$

variance $S^2 = 2.568 \times 10^9$

skewness = 4.528

kurtosis = 21.744

minimum value = 3

maximum value = 307042

In using the program OPTIBND2 to obtain the OPS $\bar{x}^*$ and the optimal sample sizes n^* , a total sample size n of 15 units has been assumed. Note that this corresponds to an overall sampling fraction q of only $\frac{15}{80} = 18.75\%$. As in the previous illustration, a minimal sample size of 2 has also been assumed to permit unbiased variance estimation. The results obtained for the present population are summarized in Table 3.4.

Table 3.4

Optimum Points of Stratification x^* and Optimal Sample Sizes n^* for the Sales of Incorporated Wholesalers of Grain in Canada for the Year 1985 ($N = 80$ and $n = 15$). The Underlined Sample Size n_h Indicates that the Stratum is Take-All.

Stratum	h	1	2	3	4	5	$\text{Var}(\bar{x}_{st})/\text{Var}(\bar{x})$	marginal gain of precision (in %)
$L = 2$	x_h^* N_h n_h	14809 70 5	<u>10</u>				0.0142	98.6
$L = 3$	x_h^* N_h n_h	3909 58 3	14809 12 2	<u>10</u>			0.0021	85.2
$L = 4$	x_h^* N_h n_h	3909 58 4	14809 12 3	37845 4 2	<u>6</u>		0.0015	28.6
$L = 5$	x_h^* N_h n_h	1008 40 2	4733 20 2	14809 10 3	37845 4 2	<u>6</u>	0.0006	60.0

Although the overall sampling fraction $q = 18.75\%$ is relatively small, Table 3.4 shows that a take-all stratum is formed with only $L = 2$ strata since the present population is very skewed. It is also seen from Table 3.4 that, by passing from one to two strata, the variance of $\bar{x}_{st}$ is reduced by 98.6%. Based on the criterion (3.27) with $\epsilon = 0.35$, the optimal number of strata is found to be $L = 4$. However, Table 3.4 also shows that, by passing from $L = 4$ to $L = 5$, the variance of $\bar{x}_{st}$ can be reduced by 60.0% which is greater than the stopping criterion $\epsilon = 0.35$. From this observation, the use of criterion (3.27) may therefore be questioned. Another criterion based on the decrease in the variance of $\bar{x}_{st}$ compared to the one with $L = 1$ might be more appropriate. Finally, it may be noted that the program OPTIBND2 satisfied for all the strata the constraints that the stratum sample size n_h must lie between the minimal sample size 2 and the stratum population size N_h . Thus OPTIBND2 solves the problems described by Schneeberger and Goller (1979) and Schneeberger (1981) for one-way optimum stratification in the finite population case.

3.3 Use of an Auxiliary Variable

It has been seen in Chapter 1 that the most effective stratification variable in order to minimize the variance of $\bar{x}_{st}$ would be X itself. However, as noted by

Dalenius (1950), the stratification variable cannot in practice be the same as the variable of interest. The stratification variable must then be an auxiliary variable chosen from highly correlated variables, for example, from a census.

In this section, it is assumed that the stratification is based on an auxiliary variable Z . The population of N units of this variable is assumed to be with CDF $F_Z(z)$ defined on $[a_Z, b_Z]$ with mean μ_Z and finite variance σ_Z^2 . It should be noted that the notation will be basically the same as before but with the use of the subscripts X and Z . The problem considered here is to determine OPS $\underline{z}^*$ for the variable Z such that the variance of $\bar{x}_{st}$ given by (1.4) is minimized. To achieve this goal, some assumptions must however be made about the nature of the relationship between the variables X and Z .

The model between X and Z that is usually assumed in the literature was first given by Dalenius and Gurney (1951) and extended by Singh and Sukhatme (1969). This model, which will be referred to as the general model, is given by:

$$(3.29) \quad X = \Psi(Z) + e.$$

The variable e is an error term with conditional mean $E(e|Z) = 0$ and with known conditional variance $\text{Var}(e|Z) = \nu(Z)$. This conditional variance is often assumed to be of the form $\nu(Z) = \delta, \delta Z$ or δZ^2 where $\delta > 0$ is a constant. Denoting the expectation of $\nu(Z)$ by μ_ν , as noted by Anderson, Kish and Cornell (1976), it is easy to see that $\text{Var}(e) = \mu_\nu$. On the other hand, the mean and the variance of $\Psi(Z)$ will be denoted by μ_Ψ and σ_Ψ^2 , respectively. Based on these quantities, it follows that

$$(3.30) \quad \mu_X = E(X) = E(\Psi(Z) + e) = \mu_\Psi,$$

$$(3.31) \quad \sigma_X^2 = \text{Var}(X) = \text{Var}(\Psi(Z) + e) = \sigma_\Psi^2 + \mu_\nu.$$

A simpler model, that is often seen and which is a special case of the general model (3.29), is the linear model

$$(3.32) \quad X = \alpha + \beta Z + e.$$

Dalenius and Gurney (1951) justified the use of this linear model by pointing out that, in practice, statisticians often try to use a stratification variable Z which is linearly related to the variable of interest X . Using the linear model (3.32), the mean and variance of X are then given by:

$$(3.33) \quad \mu_X = \alpha + \beta \mu_Z,$$

$$(3.34) \quad \sigma_X^2 = \beta^2 \sigma_Z^2 + \mu_\nu.$$

Considering the problem of determining the OPS $\underline{z}^*$ such that the variance of $\bar{x}_{st}$ is minimized, the approach to be used is to express the variance of $\bar{x}_{st}$, given by

(1.4), in terms of the points of stratification z . Based on the general model (3.29), the stratum proportion W_h , the stratum mean μ_{Xh} and the stratum variance σ_{Xh}^2 can then be obtained by the following:

$$(3.35) \quad W_h = \int_{z_{h-1}^+}^{z_h} dF_Z(z),$$

$$(3.36) \quad \mu_{Xh} = \frac{1}{W_h} \int_{z_{h-1}^+}^{z_h} \Psi(z) dF_Z(z) = \mu_{\Psi h},$$

$$(3.37) \quad \sigma_{Xh}^2 = \sigma_{\Psi h}^2 + \mu_{\nu h}$$

where $\sigma_{\Psi h}^2$ and $\mu_{\nu h}$ are obtained following (1.5) and (1.6).

The variance of $\bar{x}_{st}$, using these last quantities, can be expressed as

$$(3.38) \quad \text{Var}(\bar{x}_{st}) = \sum_{h=1}^L \frac{W_h^2(\sigma_{\Psi h}^2 + \mu_{\nu h})}{n_h} - \sum_{h=1}^L \frac{W_h(\sigma_{\Psi h}^2 + \mu_{\nu h})}{N}.$$

As seen previously, the OPS generally depends on the type of allocation used for the total sample size n among the L strata. The four types of allocation given in the beginning of this chapter (i.e. proportional, Neyman, equal-sample-size and optimum with respect to a cost function) can again be used based on the auxiliary variable Z . For example, the Neyman allocation given by (3.3), which has been obtained in order to minimize the variance of $\bar{x}_{st}$, is now given by:

$$(3.39) \quad n_h = \frac{W_h \sqrt{\sigma_{\Psi h}^2 + \mu_{\nu h}}}{\sum_{h=1}^L W_h \sqrt{\sigma_{\Psi h}^2 + \mu_{\nu h}}} n \quad h = 1, \dots, L$$

For this allocation, the variance of $\bar{x}_{st}$ can then be written as

$$(3.40) \quad \text{Var}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L W_h \sqrt{\sigma_{\Psi h}^2 + \mu_{\nu h}} \right)^2 - \sum_{h=1}^L \frac{W_h(\sigma_{\Psi h}^2 + \mu_{\nu h})}{N}.$$

This last expression can then be used to determine the OPS in terms of the auxiliary variable Z , under the Neyman allocation.

3.3.1 Literature Review: Most of the authors that have worked on the problem of optimum stratification based on an auxiliary variable have only considered the infinite population case, i.e. when the f.p.c. can be ignored. The problem was first looked at by Dalenius and Gurney (1951) where the minimal equations have been obtained under Neyman allocation for both the general and linear models given

by (3.29) and (3.32), respectively. For the general model, the minimal equations have been found to be:

$$(3.41) \quad \frac{[\Psi(z_h) - \mu_{\Psi h}]^2 + \sigma_{\Psi h}^2 + \nu(z_h) + \mu_{\nu h}}{\sqrt{\sigma_{\Psi h}^2 + \mu_{\nu h}}} = \frac{[\Psi(z_h) - \mu_{\Psi h+1}]^2 + \sigma_{\Psi h+1}^2 + \nu(z_h) + \mu_{\nu h+1}}{\sqrt{\sigma_{\Psi h+1}^2 + \mu_{\nu h+1}}},$$

$$h = 1, \dots, L-1$$

In continuing the work of Hayashi, Maruyama and Hishida (1951), Aoyama (1954) obtained the minimal equations under proportional allocation assuming the linear model (3.32). Under this allocation, the minimal equations have been expressed as

$$(3.42) \quad z_h = \frac{1}{2}(\mu_{Zh} + \mu_{Zh+1}), \quad h = 1, \dots, L-1.$$

Serfling (1968), using the linear model (3.32) and assuming that the intra-stratum correlation ρ_h between X and Z is large enough so that the cum $\sqrt{f_Z}$ rule can be applied on Z , obtained the optimal number of strata for a given cost function. Based on the work of Dalenius and Gurney (1951), Singh and Sukhatme (1969) derived the minimal equations under proportional allocation for the model (3.29). They also provided some approximate methods to solve the minimal equations under both Neyman and proportional allocations. Singh (1971) suggested the use of a cum $\sqrt{p_Z}$ rule to obtain approximate OPS where $p_Z(z)$ is based on a cost function involving the value of Z itself. In particular, if this cost function is of the form $C(Z) = \delta Z$ for some $\delta > 0$, this rule turns out to be a cum $\sqrt{f_Z}$ rule where $f_Z(z)$ is the density of Z . In considering the equal-sample-size allocation, Singh and Parkash (1975), after obtaining the minimal equations, proposed the cum $\sqrt{\nu}$ rule to compute approximate OPS under the general model where $\nu(Z)$ is the conditional variance of the error term. Oslo (1976) derived the cum $\sqrt{f_Z}$ rule of Singh (1971) without assuming any cost function but using the linear model (3.32) under proportional allocation. Following the work of Serfling (1968), Oslo (1976) also obtained the optimal number of strata under proportional allocation for a given cost function. Anderson, Kish and Cornell (1976) compared the cum $\sqrt{f_Z}$ rule with the cum $\sqrt{p_Z}$ rule in an empirical study where the variable X and Z followed a bivariate normal distribution. Under proportional allocation, it turned out that the cum $\sqrt{f_Z}$ rule was performing somewhat better than the cum $\sqrt{p_Z}$ rule. Considering the case where the underlying population of Z can be approximated by a Pareto distribution, Wang and Aggarwal (1984) suggested an iterative method developed under the linear model given by (3.32). Finally, Yadava and Singh (1984) considered the problem of determining the OPS z^* under the Z -proportional allocation (not discussed in this thesis).

One point that should be noted is that up to now the problem of optimum stratification has been restricted to the minimization of the variance of the estimate

$\bar{x}_{st}$ given by (1.1). However, as pointed out by Cochran (1977), other estimation methods can increase the precision of the estimate of μ_X by taking advantage of the correlation between X and Z . The most commonly-used of these methods in stratified sampling are the combined ratio estimate $\bar{x}_r$ and the combined regression estimate $\bar{x}_{lin}$. The combined ratio estimate $\bar{x}_r$ of μ_X is given by

$$(3.43) \quad \bar{x}_r = \frac{\sum_{h=1}^L W_h \bar{x}_h}{\sum_{h=1}^L W_h \bar{z}_h} \mu_Z$$

where $\bar{x}_h$ and $\bar{z}_h$ are the stratum sample means based on the same sample units, and W_h is given by (3.35).

On the other hand, the combined regression estimate $\bar{x}_{lin}$ can be written as

$$(3.44) \quad \bar{x}_{lin} = \sum_{h=1}^L W_h \bar{x}_h + \hat{\beta} \left(\mu_Z - \sum_{h=1}^L W_h \bar{z}_h \right)$$

where $\hat{\beta} = \left[\sum_{h=1}^L \frac{W_h^2 \sigma_{XZ_h}}{n_h} \right] \div \left[\sum_{h=1}^L \frac{W_h^2 \sigma_{X_h}^2}{n_h} \right]$

The quantity σ_{XZ_h} is the stratum covariance between X and Z which is given for the linear model (3.32) by $\beta \sigma_{Z_h}^2$. The use of the combined ratio estimate $\bar{x}_r$ in the optimum stratification problem was again first considered by Dalenius and Gurney (1951) where the minimal equations were obtained under Neyman allocation. Singh and Sukhatme (1973) shown, for Neyman allocation, that the problem of determining OPS for ratio and regression estimates is simply a particular case of minimizing the variance of $\bar{x}_{st}$ based on the auxiliary variable Z . This is important because it reduces the optimization problem with estimators $\bar{x}_r$ and $\bar{x}_{lin}$ to the initial problem with the estimator $\bar{x}_{st}$. Singh (1977) showed that this property also holds for the equal-sample-size allocation. He also noted that, under proportional allocation, the variances of $\bar{x}_r$ and $\bar{x}_{lin}$ are independent of the points of stratification z , assuming the linear model (3.32).

Most of the methods developed to obtain the OPS z^* that minimize the variance of $\bar{x}_{st}$ solve the minimal equations. It has been seen, however, in Sections 3.1 and 3.2 that the minimal equations were only necessary but not sufficient conditions to obtain a minimum for the variance of $\bar{x}_{st}$. Thus the points of stratification obtained could lead to a saddle-point or even a maximum. This remark also holds for the case where the stratification variable is the variable Z . Consequently, the use of a method like DP is required in order to ensure a global minimum for the variance of $\bar{x}_{st}$.

3.3.2 The Dynamic Programming Approach: The DP method for the case where the stratification is based on the auxiliary variable Z can be used in the same

manner as in Sections 3.1 and 3.2. One difference that arises is that all the required computations are now based on the CDF $F_Z(z)$ of the population of the variable Z instead of X . A second difference is that the different objective functions with respect to the various allocations are now related to the variance of $\bar{x}_h$ in terms of the points of stratification z instead of x .

For the infinite population case, the objective function is now of the form

$$(3.45) \quad G(z) = \sum_{h=1}^L g(z_{h-1}, z_h).$$

The function $g(z_{h-1}, z_h)$, which depends on the type of allocation used, can simply be obtained from the variance of $\bar{x}_h$ based on the quantities W_h , μ_{Xh} and σ_{Xh}^2 given by (3.35), (3.36) and (3.37), respectively. For Neyman allocation, for example, the objective function is obtained from equation (3.40) ignoring the f.p.c.. In this case, the function $g(z_{h-1}, z_h)$ is given by

$$(3.46) \quad g(z_{h-1}, z_h) = W_h \sqrt{\sigma_{\Psi h}^2 + \mu_{\nu h}}.$$

The DP formulation (2.1) of the optimum stratification problem based on the auxiliary variable Z is then given, for the infinite population case, by

$$(3.47) \quad \begin{aligned} & \text{Min } \sum_{h=1}^L g(z_{h-1}, z_h) \\ & \text{subject to: } z_h = z_{h-1} + d_h \\ & \quad z_h \in [a_Z, b_Z] \\ & \quad d_h \in B_h(z_{h-1}) = [0, b_Z - z_{h-1}] \\ & \quad z_0 = a_Z, \quad h = 1, \dots, L. \end{aligned}$$

By looking at this last formulation, it is clear that the Functional Equation of Bellman for this problem is simply given by equation (3.17) replacing the x 's by z 's. If the population of the variable Z is a real population $\{z\}$ of N units, the same approach described in Section 3.1.4 can be used. That is, assuming that the model (3.29) is true, the quantities W_h , $\mu_{\Psi h}$ and $\sigma_{\Psi h}^2$ defined in (3.35), (3.36) and (3.37) can be replaced by their finite population counterparts:

$$(3.48) \quad W_h : \quad \frac{N_h}{N} = \frac{1}{N} \{\text{number of } z\text{'s in } I_{Zh}\}$$

$$(3.49) \quad \mu_{\Psi h} : \quad \bar{\Psi}_h = \frac{1}{N_h} \sum_{z \in I_{Zh}} \Psi(z)$$

$$(3.50) \quad \sigma_{\Psi h}^2 : \quad S_{\Psi h}^2 = \frac{1}{N_h - 1} \sum_{z \in I_{Zh}} (\Psi(z) - \bar{\Psi}_h)^2.$$

where $I_{Zh} = [z_{h-1}^+, z_h]$.

For the finite population case, the objective function based on the points of stratification $\underline{z}$ can be expressed in the form

$$(3.51) \quad G(\underline{z}, \underline{n}) = \sum_{h=1}^L g(z_{h-1}, z_h, n_h).$$

As in Section 3.2, it is again possible to consider either the Neyman allocation or the optimum allocation based on a cost constraint. A sampling design involving a take-all stratum and several take-some strata can also be treated. The only work that remains is to replace the quantities W_h, μ_{Xh} and σ_{Xh}^2 given by (1.3), (1.6) and (1.5), respectively, by those given by (3.35), (3.36) and (3.37). For the Neyman allocation, the function $g(z_{h-1}, z_h, n_h)$ is then given, assuming a design with no take-all strata, by

$$(3.52) \quad g(z_{h-1}, z_h, n_h) = W_h(\sigma_{\psi h}^2 + \mu_{\nu h}) \left(\frac{W_h}{n_h} - \frac{1}{N} \right), h = 1, \dots, L.$$

Having determined the objective function, the problem can then be expressed in the form (2.1) in order to solve it using the two-dimensional DP method discussed in Section 3.2. The DP formulation (2.1) for the optimum stratification problem based on an auxiliary variable Z is then given, for the finite population case, by

$$(3.53) \quad \begin{aligned} & \text{Min } \sum_{h=1}^L g(z_{h-1}, z_h, n_h) \\ & \text{subject to: } (z_h, m_h) = (z_{h-1} + d_h, m_{h-1} + n_h) \\ & \quad (z_h, m_h) \in [a_Z, b_Z] \times \{0, 1, 2, \dots, n\} \\ & \quad (d_h, n_h) \in B_h^{(1)}(z_{h-1}) \times B_h^{(2)}(z_{h-1}, z_h, m_{h-1}) \\ & \quad (z_0, m_0) = (a_Z, 0), \quad h = 1, \dots, L. \end{aligned}$$

This formulation is in fact the same as (3.25) replacing the x 's by z 's. Because of this similarity, the Functional Equation of Bellman for this problem can then be simply obtained from equation (3.26). As in the previous paragraph, if the population of the auxiliary variable Z is a real population $\{z\}$, the problem of optimum stratification can be solved by DP by replacing the quantities $W_h, \mu_{\psi h}$ and $\sigma_{\psi h}^2$ by their finite population counterparts given by (3.48), (3.49) and (3.50), respectively.

All the advantages and disadvantages of using the DP approach given in Sections 3.1 and 3.2 are still applicable in the present case of stratification using the auxiliary variable Z . The major difficulty when using an auxiliary variable as the stratification variable is the determination of the form of the model given in general by (3.29). To determine the equation $\nu(Z)$ for the conditional variance of the error term of (3.29) may also be a difficult task.

4. OPTIMUM STRATIFICATION WITH RESPECT TO VARIANCE ESTIMATION

In survey sampling, effort is usually put in trying to estimate the mean (or the total) of the variable of interest in an optimum way. Letting X be the variable of interest, the allocation of the total sample size n and the points of stratification are then chosen in order to minimize the variance of an estimate of the mean of X . However, in general, this variance also has to be estimated from the survey. A measure of precision is then also associated with the estimated variance so that the allocation of the sample and the choice of the points of stratification may in fact be made in order to maximize the precision of the estimated variance. As noted by Ross (1961), placing more emphasis on the precision of variance estimation may be worthwhile because of the growing interest in using sample surveys for analytic studies. The point to consider is how much would the variance of the estimate of the mean be affected if the optimality criterion is applied to the minimization of the variance of the estimated variance. Conversely, the loss in precision of the estimated variance can also be measured if the allocation and points of stratification are chosen in order to minimize the variance of the estimated mean.

As in the previous chapters, the variable of interest X is assumed to be sampled from a population of N units with CDF $F(x)$ defined on $[a, b]$ with mean μ and finite variance σ^2 . It is also assumed, in this chapter, that the distribution of X has the finite central fourth moment γ . Considering the population as being divided into L strata by the points of stratification $\underline{x}$ (or $\underline{z}$ if an auxiliary variable Z is used as the stratification variable), the mean μ is estimated as before by $\bar{x}_{st}$ given by equation (1.1). The variance of $\bar{x}_{st}$ is given by equation (1.4). The variance of $\bar{x}_{st}$ can be estimated from the sample using the unbiased estimator

$$(4.1) \quad v_{st} = \text{var}(\bar{x}_{st}) = \sum_{h=1}^L \frac{W_h^2 s_h^2}{n_h} - \sum_{h=1}^L \frac{W_h s_h^2}{N}$$

$$\text{where} \quad s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2,$$

$\bar{x}_h$ is the sample mean given by (1.2) and W_h is given by (1.3).

It should be noted that, as for the variance of $\bar{x}_{st}$ given by (1.4), the second term of v_{st} corresponds to the f.p.c.. From Hansen, Hurwitz, and Madow (1953), the variance of v_{st} is given by the following

$$\begin{aligned}
\text{Var}(v_{st}) = \sum_{h=1}^L \frac{W_h^4}{n_h^2} \frac{(N_h - 1)^2}{N_h^2 (n_h - 1)^2} \left\{ \left[\frac{(n_h - 1)^2}{n_h} - \frac{(n_h - 1)}{n_h(N_h - 1)} ((n_h - 2)(n_h - 3) - (n_h - 1)) \right. \right. \\
(4.2) \quad \left. \left. - 4 \frac{(n_h - 1)(n_h - 2)(n_h - 3)}{n_h(N_h - 1)(N_h - 2)} - 6 \frac{(n_h - 1)(n_h - 2)(n_h - 3)}{n_h(N_h - 1)(N_h - 2)(N_h - 3)} \right] \gamma_h \right. \\
\left. + \left[\frac{(n_h - 1)N_h((n_h - 1)^2 + 2)}{n_h(N_h - 1)} + 2 \frac{(n_h - 1)(n_h - 2)(n_h - 3)N_h}{n_h(N_h - 1)(N_h - 2)} \right. \right. \\
\left. \left. + \frac{3(n_h - 1)(n_h - 2)(n_h - 3)N_h}{n_h(N_h - 1)(N_h - 2)(N_h - 3)} - \frac{N_h^2(n_h - 1)^2}{(N - 1)^2} \right] \sigma_h^4 \right\}
\end{aligned}$$

$$(4.3) \quad \text{where} \quad \gamma_h = \frac{N}{NW_h - 1} \int_{x_{h-1}^+}^{x_h} (x - \mu_h)^4 dF(x),$$

μ_h and σ_h^2 are given by (1.6) and (1.5), respectively, and $N_h = W_h N$.

This last equation is fairly complicated. However, if the population is infinite ($N = +\infty$) or at least large enough so that its finiteness can be ignored, the variance of v_{st} turns out to be simply given by

$$(4.4) \quad \text{Var}(v_{st}) = \sum_{h=1}^L \frac{W_h^4}{n_h^3} \left\{ \gamma_h - \frac{(n_h - 3)}{(n_h - 1)} \sigma_h^4 \right\}.$$

Moreover, assuming that the allocated stratum sample sizes n_h are also large, an approximation to the variance of v_{st} can be written as

$$(4.5) \quad \text{Var}(v_{st}) \approx \sum_{h=1}^L \frac{W_h^4}{n_h^3} \sigma_h^4 (\kappa_h - 1)$$

$$(4.6) \quad \text{where} \quad \kappa_h = \frac{\gamma_h}{\sigma_h^4}.$$

The quantity $(\kappa_h - 3)$, which is called kurtosis, is a measure of the heaviness of the tails of the population distribution of X in stratum h . The larger the kurtosis, the heavier the tails will be. Note that the normal distribution has a kurtosis of 0. It can also be shown that, for any distribution $F(x)$, the quantity κ_h is always greater or equal to 1. A good discussion about the properties of the kurtosis can be found in Scheffé (1959).

Having obtained the expressions for the variance of v_{st} for both finite and infinite population cases, the problem of determining the OPS, say $\underline{x}^{**}$ (or $\underline{z}^{**}$ if an auxiliary variable is used as the stratification variable) which will minimize the variance of v_{st} , can then be considered. This result will then allow the comparison with the OPS to be obtained by minimizing the variance of $\bar{x}_{st}$.

4.1 Infinite Population Case

In this section, it is assumed that the population size N is infinite (i.e., $N = +\infty$) or at least large enough, compared to the total sample size n and the number L of strata, so that its finiteness can be ignored in order to be able to approximate the variance of v_{st} by (4.4). The stratifying variable will also be assumed to be the variable of interest X itself, as in Section 3.1 for $\bar{x}_{st}$.

4.1.1 Possible Comparison Setups: As pointed out in Chapter 3, the OPS generally depend on the type of allocation used. The comparison of OPS obtained by minimizing the variance of v_{st} to others obtained by the minimization of the variance of $\bar{x}_{st}$ must then be made by considering different types of allocation. The four types of allocation discussed in Chapter 3, i.e. the proportional, equal-sample-size, Neyman and optimum (based on a cost constraint) allocations, can clearly be used. However, two other types of allocation based on the minimization of v_{st} will also be considered.

The proportional allocation is given by equation (3.1). Since it distributes the total sample size n based only on the population proportion of each stratum, it can be called "neutral" in the sense that it minimizes neither the variance of $\bar{x}_{st}$ nor the variance of v_{st} . Under proportional allocation, the variance of $\bar{x}_{st}$ is given by (3.2), ignoring the f.p.c.. On the other hand, the variance of v_{st} is obtained as

$$(4.7) \quad \text{Var}(v_{st}) = \sum_{h=1}^L \frac{W_h}{n^3} \left\{ \gamma_h - \frac{(nW_h - 3)}{(nW_h - 1)} \sigma_h^4 \right\}.$$

The OPS $\underline{x}^{**}$ can then be determined by minimizing equation (4.7) and then substituting into equation (3.2) in order to measure the loss of precision of $\bar{x}_{st}$ in not using the OPS $\underline{x}^*$ obtained by minimizing (3.2), and vice versa. It should be noted that if the loss of precision is found to be considerable, it may then be worthwhile to consider a weighted sum of the two minimizing functions, i.e.

$$(4.8) \quad V = \theta_1 \text{Var}(\bar{x}_{st}) + \theta_2 \text{Var}(v_{st})$$

where $\theta_1 + \theta_2 = 1$ and $\theta_1 \geq 0, \theta_2 \geq 0$. The weight θ_1 represents the importance of $\text{Var}(\bar{x}_{st})$ relative to $\text{Var}(v_{st})$.

The equal-sample-size allocation is given by (3.5). As with the proportional allocation, it can also be considered as "neutral". Under this allocation, the variance

of $\bar{x}_{st}$ is obtained by equation (3.6), ignoring the f.p.c.. The variance of v_{st} under the equal-sample-size allocation is given by

$$(4.9) \quad \text{Var}(v_{st}) = \frac{L^3}{n^3} \sum_{h=1}^L W_h^4 \left\{ \gamma_h - \frac{(n-3L)}{(n-L)} \sigma_h^4 \right\}.$$

Since the Neyman allocation given by (3.3) has been obtained by minimizing the variance of $\bar{x}_{st}$, it cannot be considered as "neutral". The variance of $\bar{x}_{st}$ under this allocation is obtained by equation (3.4), ignoring the f.p.c.. The result obtained by using equation (4.4) to express the variance of v_{st} under Neyman allocation is unfortunately messy. However, assuming that the stratum sample sizes n_h are large, the following nicer expression under Neyman allocation can be obtained from (4.5):

$$(4.10) \quad \text{Var}(v_{st}) \doteq \frac{1}{n^3} \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1) \right) \left(\sum_{h=1}^L W_h \sigma_h \right)^3.$$

The optimum allocation with a cost constraint, given by equation (3.8), is not "neutral", as the Neyman allocation. For this allocation, the variance of $\bar{x}_{st}$ is expressed by (3.9) for fixed budget. Again, assuming that the stratum sample sizes n_h are large, the variance of v_{st} is given, from equation (4.5), by

$$(4.11) \quad \text{Var}(v_{st}) \doteq \frac{1}{(C - c_0)^3} \left(\sum_{h=1}^L W_h \sigma_h c_h^{3/2} (\kappa_h - 1) \right) \left(\sum_{h=1}^L W_h \sigma_h \sqrt{c_h} \right)^3.$$

In studying optimum sampling designs with respect to variance estimation, Ross (1961) defined an allocation which minimizes the variance of v_{st} . By differentiating equation (4.5) under the constraint $\sum_{h=1}^L n_h = n$, Ross (1961) obtained the allocation:

$$(4.12) \quad n_h = \frac{W_h \sigma_h (\kappa_h - 1)^{1/4}}{\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4}} n, \quad h = 1, \dots, L.$$

This allocation, which is not "neutral", will be referred to as the Ross allocation. It should be noted that it is assumed that the allocated n_h are large enough so that equation (4.5) can be used for the variance of v_{st} . Under the Ross allocation, the variance of $\bar{x}_{st}$ is given by

$$(4.13) \quad \text{Var}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L W_h \sigma_h / (\kappa_h - 1)^{1/4} \right) \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} \right).$$

Furthermore, the variance of v_{st} can be expressed, under this allocation, as

$$(4.14) \quad \text{Var}(v_{st}) = \frac{1}{n^3} \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} \right)^2.$$

One important point about the Ross allocation (4.12) is that it depends on the kurtosis of each stratum h in such a way that if κ_h is equal for all h , the resulting allocation is then simply the Neyman allocation. Ross (1961) obtained an expression for the relative efficiency for the allocation (4.12) with respect to any other allocation.

Ross (1961) also defined an allocation that minimizes the variance of v_{st} under a cost constraint of the form (3.7). The resulting allocation is given by

$$(4.15) \quad n_h = (C - c_0) \frac{W_h \sigma_h (\kappa_h - 1)^{1/4} / c_h^{1/4}}{\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} c_h^{3/4}}, \quad h = 1, \dots, L.$$

Referred to as the Ross optimum allocation, this allocation gives the following expression for the variance of $\bar{x}_{st}$:

$$(4.16) \quad \text{Var}(\bar{x}_{st}) = \frac{1}{(C - c_0)} \left(\sum_{h=1}^L W_h \sigma_h c_h^{1/4} / (\kappa_h - 1)^{1/4} \right) \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} c_h^{3/4} \right).$$

The variance of v_{st} under the Ross optimum allocation can be given, using equation (4.5), by

$$(4.17) \quad \text{Var}(v_{st}) = \frac{1}{(C - c_0)^3} \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} c_h^{3/4} \right)^4.$$

A comparison between the OPS $\underline{x}^*$ that minimize the variance of $\bar{x}_{st}$ and the OPS $\underline{x}^{**}$ that minimize the variance of v_{st} can now be made by assuming one of the six allocations just described. Since this comparison requires the computation of the OPS for a given CDF $F(x)$ of the variable X , we first obtain these OPS using the DP approach.

4.1.2 The Use of Dynamic Programming: The DP approach that has to be used in order to obtain the OPS $\underline{x}^*$ and $\underline{x}^{**}$, that minimize the variances of $\bar{x}_{st}$ and v_{st} , respectively, is exactly the same as the one described in Section 3.1. The only work needed is to express the objective function to minimize in the form given by (3.13). The relationships with the different concepts used in the general DP approach are still the same. Hence, it is not surprising to find that the Functional

Equation of Bellman for this problem is simply given by equation (3.17). It should also be noted that all the advantages and disadvantages described in Section 3.1 still apply to the present problem.

As in Section 3.1, the objective function $G(\underline{x})$ is simply obtained from the equation of variance expressed under a given allocation. The choice of the objective function should in fact be made such that they can answer the question raised in this chapter. That is, how much the variance of $\bar{x}_{st}$ can be affected if the optimality criterion is placed on the minimization of the variance of v_{st} , and vice versa? This is an important point since it turns out that the determination of the OPS in order to minimize the variance of v_{st} cannot be solved by DP if the allocation is based on the minimization of the variance of $\bar{x}_{st}$. This also holds for OPS that minimize the variance of $\bar{x}_{st}$ with an allocation based on minimizing the variance of v_{st} . The reason for this problem is simply that, in such cases, the objective function can not be expressed in the form (2.1) or, more specifically, in the form (3.13). It should be noted that it may not be appropriate to consider, for example, the minimization of v_{st} under Neyman allocation to answer the question of this chapter. However, considering the determination of the OPS $\underline{x}^{**}$ such that the variance of v_{st} is minimized under the Ross allocation, for example, the function $g(x_{h-1}, x_h)$ of equation (3.13) is simply obtained from equation (4.14) as

$$(4.18) \quad g(x_{h-1}, x_h) = W_h \sigma_h (\kappa_h - 1)^{1/4}, \quad h = 1, \dots, L.$$

In practice, it often occurs that the population under consideration is a real population $\{x\}$. Assuming that the population size N is large enough so that this real population $\{x\}$ can be considered infinite, the approach discussed in Section 3.1.4 of replacing the different quantities of the objective function by their finite population counterparts can again be used. The stratum proportion W_h is then replaced by $\frac{N_h}{N}$ given by (3.19), μ_h by $\bar{X}_h$ given by (3.20) and σ_h^2 by S_h^2 given by (3.21). Moreover, the finite population counterpart of γ_h can be obtained as

$$(4.19) \quad J_h = \frac{1}{(N_h - 1)} \sum_{x \in [x_{h-1}^+, x_h]} (x - \bar{X}_h)^4, \quad h = 1, \dots, L.$$

Using J_h , the value κ_h can then be replaced by K_h given by

$$(4.20) \quad K_h = J_h / S_h^4, \quad h = 1, \dots, L.$$

Similarly, the function $g(x_{h-1}, x_h)$ given by equation (4.18) becomes

$$(4.21) \quad g(x_{h-1}, x_h) = \frac{N_h}{N} S_h (K_h - 1)^{1/4}, \quad h = 1, \dots, L.$$

It should finally be noted that all the advantages and disadvantages of using this approach remain the same as the ones discussed in Section 3.1.4.

4.1.3 Computer Program and Empirical Studies: In order to compute the OPS that minimize the variance of $\bar{x}_{st}$ or v_{st} under one of the allocations presented in Section 4.1.1, the computer program OPTIBND3 has been written. This program, whose listing can be found in Appendix C, is essentially the same as the program OPTIBND1 described in Section 3.1.5. It involves in fact the same five steps. The only feature that OPTIBND3 has different from OPTIBND1 is that it displays in its fifth step the overall kurtosis and the within-strata kurtoses. It also shows both variances of $\bar{x}_{st}$ and v_{st} . As for OPTIBND1, this program can be used for the case of an hypothetical CDF for the variable of interest X as well as for the case of a real population $\{x\}$. This last case, which is the default, is based on the finite population counterparts of W_h, μ_h, σ_h^2 and κ_h given by (3.19), (3.20), (3.21) and (4.20), respectively. The program OPTIBND3 computes, as its default comparison setup, the OPS $\underline{x}^{**}$ minimizing the variance of v_{st} under the Ross allocation. A change in the comparison setup (induced by a change in the choice of allocation, for example) can however simply be made by modifying the indicated code within the program.

The increase in the variance $\bar{x}_{st}$ using the OPS for minimizing the variance of v_{st} is assessed, first using the overlapping-triangle distribution (Bühler and Deutler (1975)). A description of this distribution can be found in Figure 3.1.

Using the program OPTIBND3, the OPS $\underline{x}^{**}$ and $\underline{x}^*$ were computed under Ross and Neyman allocations, which are non-neutral allocations. As in Section 3.1.5, since the domain of X is continuous, the interval $[-1, 1]$ has to be crossed by discrete steps. The results are displayed in Table 4.1.

From the third last column of Table 4.1, it can be seen that some gain in precision in the estimation of variance can be achieved by computing the OPS with respect to variance estimation rather than mean estimation, especially for $L = 2$ strata where this gain of precision is about 12.6%. On the other hand, the last column of Table 4.1 shows that the use of $\underline{x}^{**}$ leads to some loss of precision in the estimation of the mean. However, the loss of precision in the estimation of the mean is found to be relatively lower than the gain in precision in the estimation of the variance. At least for the overlapping-triangle distribution, it is therefore preferable to compute the OPS $\underline{x}^{**}$ for minimizing the variance of v_{st} rather than $\underline{x}^*$ for minimizing the variance of $\bar{x}_{st}$, under the Ross and Neyman allocations. By comparing the OPS $\underline{x}^{**}$ with the OPS $\underline{x}^*$, one interesting point can be made in the case where $L = 2$. The OPS x_1^{**} in this case has exactly the value 0.00 which is the point of symmetry of the overlapping-triangle distribution. Recalling the studies of Schneeberger (1979) and Goller (1981) where the OPS x_1^* was not necessarily the center of a symmetric distribution, it may then be guessed that this point of symmetry would in fact be the one where the variance of v_{st} , instead of $\bar{x}_{st}$, would be minimized. Further investigations with other symmetric distributions would however be needed to confirm this point.

Table 4.1

Optimum Points of Stratification $\underline{x}^{**}$ with respect to Variance Estimation and Optimum Points of Stratification $\underline{x}^*$ with respect to Mean Estimation for the Overlapping-Triangle Distribution. Ross Allocation is Used for $\underline{x}^{**}$, Neyman Allocation is Used for $\underline{x}^*$.

Stratum	h	1	2	3	4	$n^3 \text{Var}(v_{st})$ ($\times 10^{-3}$)	% gain of pre- cision	$n \text{Var}(x_{st})$ ($\times 10^{-3}$)	% loss of pre- cision	
$L = 2$	x_h^{**}	0.00				6.240	12.6	55.9	5.5	
	W_h	0.500								
	n_h/n	0.500								
	x_h^*	0.15				7.140		53.0		
W_h	0.754									
n_h/n	0.786									
$L = 3$	x_h^{**}	-0.18	0.18			0.427	2.3	17.2	0.6	
	W_h	0.229	0.542							
	n_h/n	0.335	0.330							
	x_h^*	-0.20	0.20			0.437		17.1		
	W_h	0.218	0.564							
	n_h/n	0.314	0.371							
$L = 4$	x_h^{**}	-0.42	-0.10	0.18			0.181		5.2	11.8
	W_h	0.115	0.175	0.481						
	n_h/n	0.147	0.137	0.300						
	x_h^*	-0.21	0.10	0.40		0.191	11.7			
	W_h	0.212	0.498	0.167						
	n_h/n	0.366	0.334	0.139						
$L = 5$	n_h^{**}	-0.42	-0.10	0.10	0.42			0.060	3.2	7.5
	W_h	0.115	0.175	0.420	0.175					
	n_h/n	0.194	0.181	0.251	0.181					
	x_h^*	-0.43	-0.12	0.12	0.43	0.062	7.4			
	W_h	0.111	0.157	0.465	0.157					
	n_h/n	0.173	0.165	0.323	0.165					

Again using the overlapping-triangle distribution, the OPS $\underline{x}^{**}$ and $\underline{x}^*$ were computed under proportional allocation. Recall that proportional allocation is a neutral allocation in the sense that it is not based on the minimization of either the variance of v_{st} or that of $\bar{x}_{st}$. The results reported in Table 4.2 were computed using OPTIDND3.

Table 4.2

Optimum Points of Stratification $\underline{x}^{**}$ with respect to Variance Estimation and Optimum Points of Stratification $\underline{x}^*$ with respect to Mean Estimation for the Overlapping-Triangle Distribution. Proportional Allocation is Used for Both $\underline{x}^{**}$ and $\underline{x}^*$.

Stratum	h	1	2	3	4	$n^3 \text{Var}(v_{st})$ ($\times 10^{-3}$)	% gain of pre- cision	$n \text{Var}(\bar{x}_{st})$ ($\times 10^{-3}$)	% loss of pre- cision
$L = 2$	x_h^{**}	0.00				6.240	18.0	55.9	4.9
	W_h	0.500							
	x_h^*	0.14				7.610		53.3	
	W_h	0.748							
$L = 3$	x_h^{**}	-0.26	0.28			0.657	2.5	19.0	1.1
	W_h	0.186	0.637						
	x_h^*	-0.25	0.25			0.674		18.8	
	W_h	0.191	0.617						
$L = 4$	x_h^{**}	-0.42	-0.04	0.34		0.314	9.8	13.7	4.6
	W_h	0.114	0.282	0.455					
	x_h^*	-0.27	0.12	0.45		0.348		13.1	
	W_h	0.181	0.551	0.165					
$L = 5$	x_h^{**}	-0.52	-0.16	0.16	0.52	0.087	12.2	8.1	2.5
	W_h	0.078	0.162	0.520	0.162				
	x_h^*	-0.47	-0.15	0.15	0.47	0.099		7.9	
	W_h	0.096	0.150	0.508	0.150				

By comparing the results of Table 4.2 with those of Table 4.1, it can be seen that, as expected, the minimum variance of v_{st} , for given L , is always greater

under proportional allocation than under Ross allocation. Similarly, the minimum variance of $\bar{x}_{st}$, for given L , is always less under Neyman allocation than under proportional allocation. Table 4.2 shows that, as in Table 4.1, the gain in precision in the estimation of the variance is, in all cases, greater than the loss of precision in the estimation of the mean when the OPS $\underline{x}^{**}$, for minimizing the variance of v_{st} , are used. Hence, under proportional allocation also, it might be preferable to determine the OPS $\underline{x}^{**}$ instead of the OPS $\underline{x}^*$ minimizing the variance of $\bar{x}_{st}$. Note that this result is based only on the overlapping-triangle distribution, and may not hold for other distributions. Finally, the OPS $\underline{x}_1^{**}$ for the case of $L = 2$ strata is again found to be the center 0.00 of the overlapping-triangle distribution.

A second study was undertaken using the distribution of the sales of the Incorporated Tobacconist Retailers in Canada for the year 1985. A description of this skewed population of $N = 185$ units can be found in Figure 3.2. As in Section 3.1.3, this population can be treated as infinite, provided that the total sample size n and the number of strata L are relatively small compared to the population size N .

The OPS $\underline{x}^{**}$ and $\underline{x}^*$ were computed under the Ross and Neyman allocations respectively, using the computer program OPTIBND3. The results are summarized in Table 4.3.

Again, Table 4.3 shows some gain in precision for the variance estimate by computing the OPS with respect to variance estimation rather than mean estimation. However, unlike the case of the overlapping-triangle distribution, this gain in precision is not always greater than the loss of precision in the mean estimate. Hence, one should not suggest the use of $\underline{x}^{**}$ in preference to $\underline{x}^*$ for all distributions or populations. By comparing the OPS $\underline{x}^*$ and the OPS $\underline{x}^{**}$, for each L , it can finally be seen that they tend to differ considerably (except for $L = 3$).

The OPS $\underline{x}^{**}$ and $\underline{x}^*$ were also computed under proportional allocation for the present population, using the program OPTIBND3. The results are reported in Table 4.4.

Table 4.3

Optimum Points of Stratification $\underline{x}^{**}$ with respect to Variance Estimation and Optimum Points of Stratification $\underline{x}^*$ with respect to Mean Estimation for the Sales of Incorporated Tobacconist Retailers in Canada for the Year 1985 ($N = 186$). Ross Allocation is Used for $\underline{x}^{**}$, Neyman Allocation is Used for $\underline{x}^*$ (f.p.c. ignored).

Stratum	h	1	2	3	4	$n^3 \text{Var}(v_{st})$ ($\times 10^{-3}$)	% gain of pre- cision	$n \text{Var}(\bar{x}_{st})$ ($\times 10^{-3}$)	% loss of pre- cision
$L = 2$	x_h^{**}	875				7.140	25.0	80.8	4.5
	W_h	0.886							
	n_h/n	0.728							
	x_h^*	595				25.300		77.3	
	W_h	0.660							
	n_h/n	0.384							
$L = 3$	x_h^{**}	465	1066			0.880	0.8	28.7	1.1
	W_h	0.573	0.362						
	n_h/n	0.433	0.356						
	x_h^*	466	1066			0.887		28.4	
	W_h	0.578	0.357						
	n_h/n	0.472	0.319						
$L = 4$	x_h^{**}	326	712	1066		0.262	9.0	16.9	9.7
	W_h	0.395	0.384	0.157					
	n_h/n	0.283	0.300	0.132					
	x_h^*	300	579	1066		0.288		15.4	
	W_h	0.378	0.276	0.281					
	n_h/n	0.261	0.157	0.296					
$L = 5$	x_h^{**}	292	573	875	1373	0.086	4.4	10.7	3.9
	W_h	0.373	0.276	0.238	0.065				
	n_h/n	0.302	0.193	0.174	0.099				
	x_h^*	300	567	875	1433	0.090		10.3	
	W_h	0.378	0.265	0.243	0.076				
	n_h/n	0.319	0.174	0.208	0.141				

Table 4.4

Optimum Points of Stratification $\underline{x}^{**}$ with respect to Variance Estimation and Optimum Points of Stratification $\underline{x}^*$ with respect to Mean Estimation for the Sales of Incorporated Tobacconist Retailers in Canada for the Year 1985 ($N = 185$). Proportional Allocation is Used for Both $\underline{x}^{**}$ and $\underline{x}^*$ (f.p.c. ignored).

Stratum	h	1	2	3	4	$n^3 \text{Var}(v_{st})$ ($\times 10^{-3}$)	% gain of pre- cision	$n \text{Var}(\bar{x}_{st})$ ($\times 10^{-3}$)	% loss of pre- cision
$L = 2$	x_h^{**} W_h	1066 0.935				12.70	0.0	88.5	0.0
	x_h^* W_h	1066 0.935				12.70		88.5	
$L = 3$	x_h^{**} W_h	686 0.757	1373 0.195			4.72	25.4	47.1	23.0
	x_h^* W_h	465 0.573	1287 0.368			6.33		38.3	
$L = 4$	x_h^{**} W_h	465 0.573	875 0.314	1373 0.065		3.43	6.8	28.8	12.5
	x_h^* W_h	326 0.395	712 0.384	1373 0.173		3.68		25.6	
$L = 5$	x_h^{**} W_h	292 0.373	595 0.286	948 0.238	1373 0.054	3.25	0.3	20.2	1.0
	x_h^* W_h	292 0.373	573 0.276	875 0.238	1373 0.065	3.26		20.0	

As in the case of Table 4.3, the gain in precision in the estimation of variance is in general not greater than the loss of precision in the estimation of the mean. An optimality criterion based on the weighted average (4.8) may, therefore, be preferable as a compromise between the minimization of the variance of v_{st} and the minimization of the variance of $\bar{x}_{st}$.

4.2 Finite Population Case

The case considered in this section is the one where the population size N is small enough that the finiteness of the population cannot be ignored. In this case, the formula that has to be used for the variance of v_{st} is given by equation (4.2). It is also assumed throughout this section that the stratification variable is the variable of interest, X .

4.2.1 Possible Comparison Setups: In computing the OPS to assess how much the variance of $\bar{x}_{st}$ can be affected if the optimality criterion is the minimization of the variance of v_{st} (and vice versa), some types of allocation must be assumed. In Section 4.1, the six types of allocation considered were the proportional, equal-sample-size, Neyman, optimum (with a cost constraint), Ross and Ross optimum allocations. Essentially the same allocations can be used in the case where the population is finite.

The proportional and equal-sample-size allocations given by (3.1) and (3.5), respectively, can be used in a simple way for the finite population case. The variance of $\bar{x}_{st}$ is expressed by (3.2) for the proportional allocation and by (3.6) for the equal-sample-size allocation. To obtain the variance of v_{st} under proportional allocation, it suffices to replace, in equation (4.2), the stratum sample size n_h by (3.1). The n_h are replaced by (3.5), under equal-sample-size allocation. It should be noted that both allocations are "neutral" since they are not defined by the minimization of the variance of $\bar{x}_{st}$ or v_{st} . Note also that with the equal-sample-size allocation, it is possible to obtain $n_h > N_h$, where N_h is the stratum population size.

The Neyman and optimum (with a cost constraint) allocations defined by equations (3.3) and (3.8), respectively, can also be considered for the finite population case. Under Neyman allocation, the variance of $\bar{x}_{st}$ is given by (3.4) while under optimum allocation with the cost constraint (3.7), it is given by (3.9). The variance of v_{st} for each of these allocations can be obtained by replacing, in equation (4.2), the stratum sample size n_h with the selected allocation. Note that for both Neyman and optimum allocations, the resulting expressions for the variance of v_{st} cannot be expressed in the DP form (3.22), as seen below. However, since these allocations are "non-neutral" because of minimizing the variance of $\bar{x}_{st}$, it may not be appropriate to consider them for the minimization of the variance of v_{st} , as discussed in Section 4.1. It should also be noted that with the Neyman and optimum (with a cost constraint) allocations, it may happen that $n_h > N_h$ for some of the strata.

The Ross allocation given by equation (4.12) as well as the Ross optimum allocation of equation (4.15) have been obtained by Ross (1961) by differentiating the formula (4.5) for the variance of v_{st} with respect to n_h . Recall that this formula holds for the infinite population case. For the finite population case, the differentiation of equation (4.2) with respect to n_h is clearly not mathematically tractable so that no expressions equivalent to (4.12) and (4.15) can be obtained

if the population cannot be considered as infinite. One possibility, however, is to still use equations (4.12) and (4.15) for the Ross and Ross optimum allocations, respectively, even if the population is finite. Based on these equations, the variances of $\bar{x}_{st}$, for the Ross and Ross optimum allocations, are respectively given by the following expressions:

(4.22)

$$\text{Var}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L W_h \sigma_h / (\kappa_h - 1)^{1/4} \right) \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} \right) - \sum_{h=1}^L \frac{W_h}{N} \sigma_h^2,$$

and

$$\begin{aligned} \text{Var}(\bar{x}_{st}) = & \frac{1}{(C - c_0)} \left(\sum_{h=1}^L W_h \sigma_h c_h^{1/4} / (\kappa_h - 1)^{1/4} \right) \left(\sum_{h=1}^L W_h \sigma_h (\kappa_h - 1)^{1/4} c_h^{3/4} \right) \\ & - \sum_{h=1}^L \frac{W_h}{N} \sigma_h^2. \end{aligned}$$

On the other hand, the variance of v_{st} for each of these allocations can simply be obtained by replacing within equation (4.2) the expression (4.12) for the Ross allocation and the expression (4.15) for the Ross optimum allocation. It should be noted that for both of these "non-neutral" allocations, it is possible to obtain $n_h > N_h$.

Based on these allocations, it is then possible to compare the OPS $\underline{x}^{**}$ obtained by minimizing the variance of v_{st} with the OPS $\underline{x}^*$ obtained by minimizing the variance of $\bar{x}_{st}$. Since the computation of these OPS includes the constraints that $n_h \leq N_h$ for all of the L strata, the use of the two-dimensional DP approach discussed in Section 3.2 is considered.

4.2.2 The Use of Dynamic Programming: For the proportional allocation, since the stratum sample size n_h is always less than the corresponding stratum population size N_h , the same approach as for the infinite population case can be used. The objective function to be minimized should however be based on the variance formula derived for the finite population case in Section 4.2.1. For the equal-sample-size allocation, the same DP approach as for the infinite population case can also be used because, for a given number L of strata, the stratum sample size is unchanged whatever the value of the points of stratification $\underline{x}$. As pointed out in Section 3.2.3, it suffices in this case to define the state space $B_h(x_{h-1})$ of form (3.16) to insure that the stratum population contains at least $\frac{n}{L}$ units. For the four other types of allocation, however, because of the constraint that the stratum sample size n_h must be less than or equal to the stratum population size N_h for all h , the two-dimensional DP approach discussed in Section 3.2 then has to be used.

Recalling that the first dimension is related to the determination of the OPS while the second one is used to obtain the optimal sample sizes $\underline{n}^*$ that minimize the objective function, the two-dimensional DP approach of Section 3.2 can be directly applied to the present problem without any changes. The only work required is in fact to set up the objective function to be minimized and the constraints to be applied on the sample sizes $\underline{n}$. Note that all the relationships to the general concepts of DP are still the same. The Functional Equation of Bellman is also given by equation (3.26).

As discussed in Section 3.2.3, since the second dimension directly handles the allocation of the total sample size n , the objective function $G(\underline{x}, \underline{n})$ of equation (3.22) is simply obtained from the variance formula that is to be minimized, without assuming any allocation. Hence, for example, to determine the OPS $\underline{x}^{**}$ that minimize the variance of v_{st} , the function $g(x_{h-1}, x_h, n_h)$ of the objective function (3.22) is obtained by directly taking, for each stratum h , the corresponding term of the equation of the variance of v_{st} given by (4.2). The use of the second dimension permits to obtain, with the constraint $\sum_{h=1}^L n_h = n$, the Neyman allocation if the variance of $\bar{x}_{st}$ is minimized and the Ross allocation if the variance of v_{st} is minimized. With the cost constraint given by (3.7), on the other hand, the optimum allocation (3.8) is obtained if the variance of $\bar{x}_{st}$ is minimized and the Ross optimum allocation is obtained if the variance of v_{st} is minimized. It should be noted that the Ross allocation and the Ross optimum allocation obtained through the two-dimensional DP approach are in fact the true allocations that would have been derived from the differentiation of equation (4.2) in the finite population case and not the approximations using equations (4.12) and (4.15) for the infinite population case. Also, with the two-dimensional DP approach, it is not possible to minimize the variance of $\bar{x}_{st}$ under an allocation that minimizes the variance of v_{st} , and vice versa.

It often occurs that the finiteness of a population cannot be ignored when dealing with a real population $\{x\}$. In this case, the approach discussed in Section 3.1.4 (of replacing the different quantities in the objective function by their finite population counterparts) can again be used. The quantities W_h, μ_h, σ_h^2 and κ_h can then be replaced respectively by $\frac{N_h}{N}$ given by (3.19), $\bar{X}_h$ given by (3.20), S_h^2 given by (3.21) and K_h given by (4.20). The advantages and disadvantages of using this approach are still the same as those discussed in Section 3.2.4 for the mean estimation in the finite population case.

It should finally be noted that no computer program has been written to compute the OPS $\underline{x}^{**}$ that minimize the variance of v_{st} contrarily the infinite population case. However, since the DP approach is exactly the same as the one used in Section 3.2, some minor changes could be made to the program OPTIBND2 in order to compute these OPS.

4.3 Use of an Auxiliary Variable

In practice, it is not feasible to use the variable of interest X as the stratifi-

cation variable. The stratification is based on an auxiliary variable Z chosen from highly correlated data already available.

As in Chapter 3, it is assumed that the population of N units of the variable Z is with CDF $F_Z(z)$ defined on $[a_Z, b_Z]$ with mean μ_Z and finite variance σ_Z^2 . It is also assumed throughout this section that the CDF $F_Z(z)$ has a finite central fourth moment γ_Z . The purpose of this section is to provide an approach to assess how much the variance of $\bar{x}_{st}$ can be affected if the optimality criterion is to minimize the variance of v_{st} when the stratification is based on the auxiliary variable Z . For this purpose, the OPS $\underline{z}^{**}$ that minimize the variance of v_{st} and the OPS $\underline{z}^*$ that minimize the variance of $\bar{x}_{st}$ can be computed based on a given allocation.

The nature of the relationship between the variables X and Z can be represented, as in Section 3.3, by the general model (3.29) or by the linear model (3.32) which is a special case of the general model. Apart from the assumptions given by Singh and Sukhatme (1969) that the error term e in (3.29) has the conditional mean $E(e|Z) = 0$ and the conditional variance $\text{Var}(e|Z) = \nu(Z)$, some additional assumptions on the error term are needed. First, it is assumed that the distribution of the error e , for a given Z , is symmetric about the mean 0 so that its conditional third moment (which is also central because of $E(e|Z) = 0$) is also 0, i.e. $E(e^3|Z) = 0$. Second, it is assumed that the error term has conditional fourth moment equal to $\eta(Z)$, i.e. $E(e^4|Z) = \eta(Z)$. On the other hand, for the general model (3.29), since the mean and the variance of $\Psi(z)$ are denoted by μ_Ψ and σ_Ψ^2 respectively, the central fourth moment of $\Psi(z)$ will be denoted by γ_Ψ . Based on these assumptions, the mean μ_X and the variance σ_X^2 of the variable X has already been given, under the general model, by equations (3.30) and (3.31), respectively. The central fourth moment γ_X can also be obtained, under the general model and these assumptions, from the following equations:

$$\begin{aligned}
 \gamma_X &= E(X - \mu_X)^4 = E(\Psi(Z) + e - \mu_\Psi)^4 \\
 &= E(\Psi(Z) - \mu_\Psi)^4 + 4E[(\Psi(Z) - \mu_\Psi)^3 e] + 6E[(\Psi(Z) - \mu_\Psi)^2 e^2] \\
 &\quad + 4E[(\Psi(Z) - \mu_\Psi) e^3] + E(e^4) \\
 &= \gamma_\Psi + 0 + 6E[(\Psi(Z) - \mu_\Psi)^2 \nu(Z)] + 0 + \mu_\eta \\
 (4.23) \quad &= \gamma_\Psi + 6E[(\Psi(Z) - \mu_\Psi)^2 \nu(Z)] + \mu_\eta
 \end{aligned}$$

where $\mu_\eta = E(\eta(Z)) = E(E(e^4|Z)) = E(e^4)$.

Note that equation (4.23) is simply obtained using conditional arguments on Z . It should also be noted that in the case where the conditional variance $\nu(Z)$ is a constant, say δ , the expression (4.23) simplifies to

$$(4.24) \quad \gamma_X = \gamma_\Psi + 6\delta\sigma_\Psi^2 + \mu_\eta.$$

The population of the auxiliary variable Z is assumed to be divided, as usual, into L strata. Using the general model (3.29), the expressions for the variance of $\bar{x}_{st}$

and the variance of v_{st} based on the points of stratification z can be derived using the same approach as in Section 3.3. Hence, it suffices to express, in equations (1.4) and (4.2), the quantities W_h by (3.35), μ_{Xh} by (3.36) and σ_{Xh}^2 by (3.37). On the other hand, the quantity γ_{Xh} is found to be simply expressed by

$$(4.25) \quad \gamma_{Xh} = \gamma_{\Psi h} + 6E_h [(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)] + \mu_{\eta h}, \quad h = 1, \dots, L.$$

where $\gamma_{\Psi h}$ and $\mu_{\eta h}$ are obtained following (4.3) and (1.6) and

$$(4.26) \quad E_h [(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)] = \frac{1}{W_h} \int_{z_{h-1}^+}^{z_h} (\Psi(Z) - \mu_{\Psi})^2 \nu(Z) dF_Z(z).$$

For example, if the population of Z can be considered as infinite, the variance of v_{st} can then be obtained, using equation (4.4), by

$$(4.27) \quad \text{Var}(v_{st}) = \sum_{h=1}^L \frac{W_h^4}{n_h^3} \left\{ \gamma_{\Psi h} + 6E_h [(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)] + \mu_{\eta h} - \frac{(n_h - 3)}{(n_h - 2)} (\sigma_{\Psi h}^2 + \mu_{\nu h})^2 \right\}.$$

By setting a particular allocation to determine the stratum sample sizes n_h , this last equation can be used to determine the OPS $\underline{z}^{**}$ based on the auxiliary variable Z .

4.3.1 Possible Comparison Setups: It has been stated in Section 4.1.1 and 4.2.1 that the assessment of how much the variance of $\bar{x}_{st}$ can be affected if the variance of v_{st} is minimized (and vice versa) has to be done by assuming some types of allocation for the total sample size n . The different possible comparisons of the OPS $\underline{z}^{**}$ that minimize the variance of v_{st} with the OPS $\underline{z}^*$ minimizing the variance of $\bar{x}_{st}$ are then related to the chosen allocations. The types of allocation discussed in Sections 4.1.1. and 4.2.1, i.e. proportional, equal-sample-size, Neyman, optimum (with a cost constraint), Ross and Ross optimum allocations, can again be considered when the stratification is based on the auxiliary variable Z . In fact, for both finite and infinite population cases, the chosen allocations can simply be expressed by replacing the quantities $W_h, \mu_{Xh}, \sigma_{Xh}^2$ and γ_{Xh} by the equations (3.35), (3.36), (3.37) and (4.25), respectively, in the definition of the allocation. For example, the Ross allocation defined by (4.12) is then given by

$$(4.28) \quad n_h = \frac{W_h (\sigma_{\Psi h}^2 + \mu_{\nu h})^{1/2} (\kappa_{Xh} - 1)^{1/4}}{\sum_{h=1}^L W_h (\sigma_{\Psi h}^2 + \mu_{\nu h})^{1/2} (\kappa_{Xh} - 1)^{1/4}} n.$$

$$(4.29) \quad \text{with} \quad \kappa_{Xh} = \frac{\gamma_{Xh}}{\sigma_{Xh}^4} = \frac{\gamma_{\Psi h} + 6E_h [(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)] + \mu_{\eta h}}{(\sigma_{\Psi h}^2 + \mu_{\nu h})^2}.$$

The equations for the variances of $\bar{x}_{st}$ and v_{st} under one of the six types of allocation directly follow using the same replacement approach.

4.3.2 The Use of Dynamic Programming: Since the comparison between the OPS $\underline{z}^{**}$ and $\underline{z}^*$ minimizing the variance of v_{st} and $\bar{x}_{st}$, respectively, requires the actual computation of these OPS, it is worthwhile to consider the use of DP. The DP approaches to be used for the infinite and finite population cases for the auxiliary variable Z are in fact exactly the same as the ones described in Section 3.3.2. Again the only work required is to express the variance to be minimized under a given allocation in the form (3.45) for the infinite population case. For the finite population case, the variance to be minimized has to be expressed in the form (3.51). The form (3.45) or (3.51) can simply be obtained from the variances of v_{st} and $\bar{x}_{st}$ by replacing the quantities $W_h, \mu_{Xh}, \sigma_{Xh}^2$ and γ_{Xh} by the equations (3.35), (3.36), (3.37) and (4.25), respectively. As an example, under the Ross allocation in the infinite population case, the function $g(z_{h-1}, z_h)$ of (3.45) is given, for the variance of v_{st} (4.14), by

$$(4.30) \quad g(z_{h-1}, z_h) = W_h(\sigma_{\Psi h}^2 + \mu_{\nu h})^{1/2}(\kappa_{Xh} - 1)^{1/4}$$

where κ_{Xh} is given by (4.29).

It should be noted that all the relationships to the general concepts of DP are still the same as in Section 3.3.2. The Functional Equation of Bellman can also be obtained from equation (3.17) for the infinite population case and by equation (3.26) for the finite population case, replacing the x 's by z 's.

When the population of the auxiliary variable Z is a real population $\{z\}$ of N units, the same approach as the one described in Section 3.3.2 can be used. With this approach, the different quantities of the objective function are replaced by their finite population counterparts. Hence, the population proportion W_h is replaced by $\frac{N_h}{N}$ given by (3.48), the mean $\mu_{\Psi h}$ by $\bar{\Psi}_h$ given by (3.49) and the variance $\sigma_{\Psi h}^2$ by $S_{\Psi h}^2$ given by (3.50), assuming that the general model (3.29) is true. On the other hand, the quantities $\gamma_{\Psi h}$ and $E_h [(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)]$ can be replaced by

$$(4.31)$$

$$\gamma_{\Psi h} : J_{\Psi h} = \frac{1}{N_h - 1} \sum_{z \in I_{Zh}} (\Psi(z) - \bar{\Psi}_h)^4,$$

$$(4.32)$$

$$E_h [(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)] : \overline{[(\Psi(Z) - \mu_{\Psi})^2 \nu(Z)]}_h = \frac{1}{N_h - 1} \sum_{z \in I_{Zh}} (\Psi(z) - \bar{\Psi}_h)^2 \nu(z)$$

where $I_{Zh} = [z_{h-1}^+, z_h]$.

As noted in Section 3.3, the main difficulty when using an auxiliary variable as the stratification variable is to determine the form of the model given in general by (3.29). The form of the conditional variance $\nu(Z)$ of the error term is usually

difficult to determine, and the form of the conditional central fourth moment $\eta(Z)$ is often an even more difficult task. These drawbacks hinder the applicability of the model approach when the OPS z^{**} that minimize the variance of v_{st} have to be computed.

5. 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 variable of interest X is mainly to improve the precision of the estimate, the use of one or more other stratification variables may 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. For example, Canadian socio-economic and agricultural surveys are stratified by province. On the other hand, business surveys use, as a stratification variable, a Standard Industrial Classification. It should be noted that stratification which involves two variables is referred to as two-way stratification.

In this chapter, the estimation of the mean of the variable of interest X based on two-way stratification with one categorical stratification variable will be studied. The population is assumed to be already stratified into M categories (or classes). Each of these categories is then to be subdivided in L substrata according to the variable of interest X (or Z if an auxiliary variable is used). The problem considered here is to obtain, for each category k , OPS $\underline{x}_k^*$ (or $\underline{z}_k^*$) in such a way that these points of stratification will be the same for all M categories, i.e. $\underline{x}_k^* = \underline{x}^*$. From Isii and Taga (1969), such a stratification is called a lattice stratification which can be seen as being formed by straight lines parallel to x - or y -axis in the two-dimensional Euclidian space. Because these strata are in some sense symmetric, as noted by Kish and Anderson (1978), some increase in precision of the estimate could be obtained by introducing asymmetries. These asymmetries could be introduced by determining a different vector of OPS for each of the M categories. The use of lattice stratification has however the advantage that, in practice, it is found to be usually more convenient for displaying results since the produced tables can easily be square or rectangular. Also, from an implementation point of view, lattice stratification is generally simpler to manage and to program than general two-way stratification. One question may then be to assess how much precision in the estimate is lost by determining a single vector of OPS $\underline{x}^*$ for all the M categories. 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.

We assume that the variable of interest X is sampled, for the k th of the M categories, from a subpopulation of size N_k with 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 from each category k , the domain of X is assumed to be the same. Let θ_k represent the relative size of the population of the k th category such that $\sum_{k=1}^M \theta_k = 1$ and $\theta_k > 0$ for $k = 1, \dots, M$. If the population size N ($N = \sum_{k=1}^M N_k$) is finite, for

example, θ_k is simply given by

$$(5.1) \quad \theta_k = \frac{N_k}{N}, \quad k = 1, \dots, M.$$

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

$$(5.2) \quad \mu = \sum_{k=1}^M \theta_k \mu_k.$$

An unbiased estimate of μ is then given by

$$(5.3) \quad \bar{x}_{comb} = \sum_{k=1}^M \theta_k \bar{x}_k$$

$$(5.4) \quad \text{where} \quad \bar{x}_k = \sum_{h=1}^L W_{(k)h} \bar{x}_{kh},$$

$$(5.5) \quad \bar{x}_{kh} = \frac{1}{n_{kh}} \sum_{i=1}^{n_{kh}} x_{khi},$$

$$(5.6) \quad W_{(k)h} = \int_{x_{k-1}^+}^{x_h} dF_k(x).$$

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

$$(5.7) \quad W_{kh} = \theta_k W_{(k)h}, \quad h = 1, \dots, M; k = 1, \dots, L.$$

Using the last equation, the estimator $\bar{x}_{comb}$, given by (5.3), can be simply written in the form

$$(5.8) \quad \bar{x}_{comb} = \sum_{k=1}^M \sum_{h=1}^L W_{kh} \bar{x}_{kh}.$$

The form (5.8) has the same form as $\bar{x}_{st}$ given by (1.1) in the one-way stratification case. The variance of $\bar{x}_{comb}$ is obtained as

$$(5.9) \quad \text{Var}(\bar{x}_{comb}) = \sum_{k=1}^M \sum_{h=1}^L W_{kh}^2 \frac{\sigma_{(k)h}^2}{n_{kh}} - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} \sigma_{(k)h}^2}{N}$$

$$(5.10) \text{ where } \sigma_{(k)h}^2 = \frac{N_k}{N_k W_{(k)h} - 1} \int_{x_{k-1}^+}^{x_h} (x - \mu_{(k)h})^2 dF_k(x)$$

$$(5.11) \quad \mu_{(k)h} = \frac{1}{W_{(k)h}} \int_{x_{k-1}^+}^{x_h} x dF_k(x), \quad h = 1, \dots, L; k = 1, \dots, M.$$

Note that the second term in equation (5.9) is the usual f.p.c.. The determination of the OPS $\underline{x}^*$ (or $\underline{z}^*$ if an auxiliary variable is used) to construct the lattice stratification for the population of the variable X can be done by minimizing equation (5.9) with respect to the points of stratification $\underline{x}$.

In practice, two other approaches, different from minimizing the variance of $\bar{x}_{comb}$ given by (5.9), are often employed to approximate the single vector $\underline{x}^*$ of OPS for all M categories. In the first of these approaches, the OPS $\underline{x}_k^*$ for each category k are obtained separately and then these OPS are reconciled to finally obtain a single vector of points of stratification $\hat{\underline{x}}$. This reconciliation can either be done in an intuitive way or following a predetermined rule such as taking a weighted average of the points of stratification across the M categories. For example, this weighted average may be

$$(5.12) \quad \hat{x}_h = \sum_{k=1}^M \theta_k x_{kh}^*, \quad k = 1, \dots, L-1.$$

Note that $x_{k0}^* = a$ and $x_{kL}^* = b$ for all k because of the previous assumption on the domain of X . It should also be noted that this approach will seldom produce the true OPS $\underline{x}^*$. The second approach basically consists in ignoring the stratification into M categories and then computing the OPS for the overall unstratified population of the variable X . Clearly, this method will provide some points of stratification, say $\hat{\underline{x}}$, that may be used for each of the M categories. However, unless the CDF $F_k(x)$ of the subpopulation for each category k are equal (i.e. $F_k(x) = F(x)$), this approach will not generally provide the true OPS $\underline{x}^*$ for the lattice stratification. These two approaches are often used in practice because they in fact correspond to the one-way stratification problem discussed in Chapter 1 and hence are easy to implement. The techniques given in this chapter, however, can also be simply applied. It would be of interest to assess the increase in the variance of $\bar{x}_{comb}$, given in (5.9), by using the points of stratification $\hat{\underline{x}}$ or $\hat{\underline{z}}$ relative to using the OPS $\underline{x}^*$.

Dalenius (1950) showed that the OPS generally depends on the type of allocation used for the total sample size n . In two-way stratification, the allocation can be done in what may be called either a one-step basis or a two-step basis. For the one-step type of allocation, the total sample size n is distributed directly to each stratum (k, h) . On the other hand, for the two-step type of allocation, the total sample size n is first distributed among the M categories, and the resulting allocated sample size n_k of each category k is then redistributed among the L strata of the category.

The four different one-step allocations considered in this chapter correspond to the ones discussed in Chapter 3 for one-way stratification, i.e., the proportional, equal-sample-size, Neyman and optimum (with a cost constraint) allocations given by (3.1), (3.5), (3.3) and (3.8), respectively. The proportional allocation here is

obtained from

$$(5.13) \quad n_{kh} = W_{kh}n = \theta_k W_{(k)h}n, \quad k = 1, \dots, M; h = 1, \dots, L.$$

Under this allocation, the variance of $\bar{x}_{comb}$ is given by

$$(5.14) \quad \text{Var}(\bar{x}_{comb}) = \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh}\sigma_{(k)h}^2}{n} - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh}\sigma_{(k)h}^2}{N},$$

where W_{kh} and $\sigma_{(k)h}^2$ are given by (5.7) and (5.10), respectively.

On the other hand, the equal-sample-size allocation is simply expressed as

$$(5.15) \quad n_{kh} = \frac{n}{ML}, \quad k = 1, \dots, M; h = 1, \dots, L.$$

The variance of $\bar{x}_{comb}$ under this allocation is given by

$$(5.16) \quad \text{Var}(\bar{x}_{comb}) = ML \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh}^2 \sigma_{(k)h}^2}{n} - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} \sigma_{(k)h}^2}{N}.$$

Considering the minimization of the variance of $\bar{x}_{comb}$ with respect to the stratum sample size n_{kh} , the Neyman allocation is determined by

$$(5.17) \quad n_{kh} = \frac{W_{kh}\sigma_{(k)h}}{\sum_{k=1}^M \sum_{h=1}^L W_{kh}\sigma_{(k)h}} n, \quad k = 1, \dots, M; h = 1, \dots, L.$$

Under this allocation, the variance of $\bar{x}_{comb}$ can be expressed as

$$(5.18) \quad \text{Var}(\bar{x}_{comb}) = \frac{1}{n} \left(\sum_{k=1}^M \sum_{h=1}^L W_{kh}\sigma_{(k)h} \right)^2 - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh}\sigma_{(k)h}^2}{N}.$$

Finally, the optimum allocation is obtained from the minimization of the variance of $\bar{x}_{comb}$ under a cost constraint of the form

$$(5.19) \quad C = c_0 + \sum_{k=1}^M \sum_{h=1}^L c_{kh} n_{kh}.$$

The allocation obtained can be expressed as

$$(5.20) \quad n_{kh} = (C - c_0) \frac{W_{kh}\sigma_{(k)h}/\sqrt{c_{kh}}}{\sum_{k=1}^M \sum_{h=1}^L W_{kh}\sigma_{(k)h}\sqrt{c_{kh}}}, \quad k = 1, \dots, M; h = 1, \dots, L.$$

Under this allocation, the following formula is obtained for the variance of $\bar{x}_{comb}$:

$$(5.21) \quad \text{Var}(\bar{x}_{comb}) = \frac{1}{(C - c_0)} \left(\sum_{k=1}^M \sum_{h=1}^L W_{kh} \sigma_{(k)h} \sqrt{c_{kh}} \right)^2 - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} \sigma_{(k)h}^2}{N}.$$

Since these one-step allocations turn out to be basically the same as the ones discussed in Chapter 3, all the related comments presented for the one-way stratification still hold for the two-way case.

For the two-step type of allocation, each of the two steps corresponds, in fact, to an allocation in a one-way stratification. Hence, the two-step allocations can be obtained from any two of the four allocations given in Chapter 3, i.e., the proportional, equal-sample-size, Neyman and optimum (with a cost constraint) allocations given by (3.1), (3.5), (3.3) and (3.8), respectively. Since the first step is used to allocate the total sample size n among the M categories, the expressions of the allocations for this step can simply be obtained by replacing the quantities used in the expressions given in Chapter 3 by the quantities related to the subpopulation underlying each category. The stratum proportion W_h is replaced by the relative size θ_k of category k , the stratum variance σ_h^2 by the variance σ_k^2 of the subpopulation of category k and the cost c_h of the cost function (3.7) by a cost c_k related to each category k . For example, using Neyman allocation in the first step would be expressed as

$$(5.22) \quad n_k = \frac{\theta_k \sigma_k}{\sum_{k=1}^M \theta_k \sigma_k} n, \quad k = 1, \dots, M.$$

Assuming that a sample size n_k has been allocated to each category k , the expressions of the allocations for the second step can also be obtained from the ones given in Chapter 3. In this case, the quantity W_h is replaced by the stratum subpopulation proportion $W_{(k)h}$ given by (5.6), the quantity σ_h^2 by the stratum variance $\sigma_{(k)h}^2$ given by (5.10) and the cost c_h by c_{kh} taken from (5.19). Considering again the Neyman allocation as an example, the expression for the second step would be given by

$$(5.23) \quad n_{kh} = \frac{W_{(k)h} \sigma_{(k)h}}{\sum_{h=1}^L W_{(k)h} \sigma_{(k)h}} n_k, \quad k = 1, \dots, M; h = 1, \dots, L.$$

The two-step allocations are obtained by combining the definition of the allocation of the first step with the one of the second step. For convenience, the two-step allocations may be labeled by first putting the type of allocation of the first step

and then, the type of allocation of the second step, separated by a dash. As an example, the Neyman-Neyman allocation is defined, using (5.22) and (5.23), by

$$n_{kh} = \frac{W_{(k)h} \sigma_{(k)h} \theta_k \sigma_k}{\left(\sum_{h=1}^L W_{(k)h} \sigma_{(k)h} \right) \left(\sum_{k=1}^M \theta_k \sigma_k \right)} n, \quad k = 1, \dots, M; h = 1, \dots, L.$$

It should be noted that the Neyman-Neyman allocation differs from the Neyman allocation given by (5.17). Under the Neyman-Neyman allocation, the variance of $\bar{x}_{comb}$ can be obtained from

(5.24)

$$\text{Var}(\bar{x}_{comb}) = \frac{1}{n} \left[\sum_{k=1}^M \frac{\theta_k}{\sigma_k} \left(\sum_{h=1}^L W_{(k)h} \sigma_{(k)h} \right)^2 \right] \left[\sum_{k=1}^M \theta_k \sigma_k \right] - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} \sigma_{(k)h}^2}{N}.$$

Another example of a two-step allocation is the proportional-Neyman allocation given by

(5.25)

$$n_{kh} = \frac{W_{(k)h} \sigma_{(k)h}}{\sum_{h=1}^L W_{(k)h} \sigma_{(k)h}} \theta_k n = \frac{W_{kh} \sigma_{(k)h}}{\sum_{h=1}^L W_{(k)h} \sigma_{(k)h}} n, \quad k = 1, \dots, M; h = 1, \dots, L.$$

Under this allocation, the variance of $\bar{x}_{comb}$ can be expressed as

$$(5.26) \quad \text{Var}(\bar{x}_{comb}) = \frac{1}{n} \sum_{k=1}^M \theta_k \left(\sum_{h=1}^L W_{(k)h} \sigma_{(k)h} \right)^2 - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} \sigma_{(k)h}^2}{N}.$$

As a third example, the Neyman-proportional allocation is defined as

$$n_{kh} = \frac{W_{(k)h} \theta_k \sigma_k}{\sum_{k=1}^M \theta_k \sigma_k} n = \frac{W_{kh} \sigma_k}{\sum_{k=1}^M \theta_k \sigma_k} n, \quad k = 1, \dots, M; h = 1, \dots, L.$$

The variance of $\bar{x}_{comb}$ under this allocation is obtained as

(5.27)

$$\text{Var}(\bar{x}_{comb}) = \frac{1}{n} \left(\sum_{k=1}^M \frac{\theta_k}{\sigma_k} \sum_{h=1}^L W_{(k)h} \sigma_{(k)h}^2 \right) \left(\sum_{k=1}^M \theta_k \sigma_k \right) - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} \sigma_{(k)h}^2}{N}.$$

Note that the proportional-proportional and the equal-sample-size-equal-sample-size allocations turn out to be exactly the same as the one-step allocations given by (5.13) and (5.15), respectively. In the case of a finite population, some care must be taken if one of the two steps uses the equal-sample-size, Neyman or optimum

(under a cost constraint) allocation because it may be possible to obtain $n_k > N_k$ or $n_{kh} > N_{kh}$. Such situations are undesirable because, in practice, they are not feasible and also because they invalidate the expressions of the variance of $\bar{x}_{comb}$ under the given allocation.

5.1 Infinite Population Case

In this section, it is assumed that the overall population size N ($N = \sum_{k=1}^M N_k$) is infinite or at least large enough, compared to the total sample size n and the number ML of strata, so that the f.p.c. in the variance of $\bar{x}_{comb}$ can be ignored. As before, the substratification variable within each category is considered as being the variable of interest X itself.

The problem of determining the OPS $\underline{x}^*$ for lattice stratification where one of the stratification variable is categorical does not seem to have been considered in the literature. As pointed out earlier, 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 based on the use of DP.

5.1.1 The Dynamic Programming Approach: The DP approach discussed in general in Chapter 2 can easily be applied to the present problem for most of the one-step and two-step allocations presented earlier. As stated in Section 3.1.3, the first step is to identify the objective function to minimize in a form similar to the one given by (2.1). For the problem of determining the OPS $\underline{x}^*$ for lattice stratification with one categorical stratification variable, the objective function $G(\underline{x})$ can be expressed in the form

$$(5.28) \quad G(\underline{x}) = \sum_{h=1}^L \left(\sum_{k=1}^M g_k(x_{h-1}, x_h) \right).$$

The determination of $g_k(x_{h-1}, x_h)$ is done by considering the formula for the variance of $\bar{x}_{comb}$, ignoring the f.p.c., obtained under a given allocation. For the one-step allocations, it turns out that all of the proportional, equal-sample-size, Neyman and optimum allocations given by (5.13), (5.15), (5.17) and (5.20), respectively, can be handled by the DP approach. For the proportional allocation,

for example, the function $g_k(x_{h-1}, x_h)$ is expressed, using (5.14), as

$$(5.29) \quad g_k(x_{h-1}, x_h) = W_{kh} \sigma_{(k)h}^2, \quad k = 1, \dots, M; h = 1, \dots, L.$$

In a similar manner, using equation (5.18), the function $g_k(x_{h-1}, x_h)$ is obtained for the Neyman allocation as

$$(5.30) \quad g_k(x_{h-1}, x_h) = W_{kh} \sigma_{(k)h}, \quad k = 1, \dots, M; h = 1, \dots, L.$$

Unfortunately for the two-step allocations, some of them cannot be handled using the DP approach. In fact, all the allocations involving the use of Neyman or optimum (with a cost constraint) allocation as the second step allocation make the formula for the variance of $\bar{x}_{comb}$ such that it can not be put into the form (5.28) or even more generally into the form (2.1). This becomes clear by examining, for example, the variances of $\bar{x}_{comb}$ under Neyman-Neyman allocation and under proportional-Neyman allocation given by (5.24) and (5.26), respectively. However, for the remaining two-step allocations, the objective function $G(\underline{x})$ is relatively easy to obtain. As an example, the function $g_k(x_{h-1}, x_h)$ of (5.28) can be determined for the Neyman-proportional allocation by rearranging equation (5.27), ignoring the f.p.c.. The function $g_k(x_{h-1}, x_h)$ is given by

$$(5.31) \quad g_k(x_{h-1}, x_h) = \frac{\theta_k W_{(k)h} \sigma_{(k)h}^2}{\sigma_k} = \frac{W_{kh} \sigma_{(k)h}^2}{\sigma_k}, \quad k = 1, \dots, M; h = 1, \dots, L.$$

The relationships to the general concepts of DP presented in Chapter 2 are basically the same in the present case as the ones described in Section 3.1.3 for one-way stratification. With the stages corresponding to the different intervals h , the problem of determining of the OPS $\underline{x}^*$ for lattice stratification can in fact be expressed in a form similar to (2.1) by

$$(5.32) \quad \begin{aligned} & \text{Min } \sum_{h=1}^L \sum_{k=1}^M g_k(x_{h-1}, x_h) \\ & \text{subject to : } x_h = x_{h-1} + d_h \\ & \quad x_h \in [a, b] \\ & \quad d_h \in B_h(x_{h-1}) = [0, b - x_{h-1}] \\ & \quad x_0 = a, \quad h = 1, \dots, L. \end{aligned}$$

The second step in the DP formulation of the problem involves the determination of the Functional Equation of Bellman given in general by (2.4). 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 the lower bound for the strata (k, h) for $k = 1, \dots, M$

is x_{h-1} . The Functional Equation of Bellman for determining of the OPS $\underline{x}^*$ for lattice stratification is then expressed as

(5.33)

$$G_h^*(x_{h-1}) = \min_{d_h \in B_h(x_{h-1})} \left\{ \sum_{k=1}^M g_k(x_{h-1}, x_h) + G_{h+1}^*(x_h) \mid x_h = x_{h-1} + d_h \right\} *$$

Because of the similarity in the DP formulation of this problem with the one for one-way optimum stratification, all the advantages and disadvantages of the use of the DP approach are exactly the same as the ones described in Section 3.1.3. The optimal number L of intervals in each category can again be determined based on a criterion of the form (3.18).

When the subpopulation of each category k is a real subpopulation $\{x\}_k$ of N_k units, the same approach as the one given in Section 3.1.4 can be used. This approach consists basically of replacing the different quantities entering into the objective functions $G(\underline{x})$ by their finite population counterparts. The value of θ_k is then given by (5.1) while the quantities σ_k^2 , $W_{(k)h}$, $\sigma_{(k)h}^2$ and W_{kh} can be replaced as follows:

$$(5.34) \quad \sigma_k^2 : S_k^2 = \frac{1}{N_k - 1} \sum_{x \in \{x\}_k} (x - \bar{X}_k)^2$$

$$(5.35) \quad \text{where} \quad \bar{X}_k = \frac{1}{N_k} \sum_{x \in \{x\}_k} x,$$

$$(5.36) \quad W_{(k)h} : \frac{N_{kh}}{N_k} = \frac{1}{N_k} \{\text{number of } x\text{'s in } I_{(k)h}\},$$

$$(5.37) \quad \sigma_{(k)h}^2 : S_{(k)h}^2 = \frac{1}{N_{kh} - 1} \sum_{x \in I_{(k)h}} (x - \bar{X}_{kh})^2$$

$$(5.38) \quad \text{where} \quad \bar{X}_{kh} = \frac{1}{N_{kh}} \sum_{x \in I_{(k)h}} x,$$

$$\text{and} \quad I_{(k)h} = \{x\}_k \cap [x_{h-1}^+, x_h],$$

$$(5.39) \quad W_{kh} : \frac{N_{kh}}{N}, \quad k = 1, \dots, M; h = 1, \dots, L.$$

As an example, the function $g_k(x_{h-1}, x_h)$ for the Neyman allocation can be expressed as

$$g_k(x_{h-1}, x_h) = \frac{N_{kh}}{N} S_{(k)h}^2, \quad k = 1, \dots, M; h = 1, \dots, L.$$

As noted in Section 3.1.4, with this approach, the DP formulation for the problem given by (5.32) is basically unchanged except that the state space $[a, b]$ is now

replaced by the population $\{x\}$ (where $\{x\} = \cup_{k=1}^M \{x\}_k$) and the decision space $B_k(x_{k-1})$ by $\{d | d \geq 0 \text{ and } d = x - x_{k-1}, \text{ for } x \in \{x\}\}$. It should be noted that, although each category k contains only N_k units, the use of DP for lattice stratification requires one to cross the complete population $\{x\}$ of N units.

5.1.2 Computer Programs and Illustrations: To be able to actually compute the OPS $\underline{x}^*$ for lattice stratification where one stratification variable is categorical, the program OPTIBND4 has been written for the case of an infinite population. A listing of this program can be found in Appendix D. Because of the similarity of the DP approach related to the present problem with the one of the one-way optimum stratification problem, OPTIBND4 has been programmed by simply modifying the program OPTIBND1 given in Appendix A. Hence, the program OPTIBND4 involves basically the same five steps of OPTIBND1 but with some changes to the definition of the objective function and to the display of the results and descriptive statistics. The first step is to read the data from an external file. Note that two variables are then read: the category k and the value of the variable X . Secondly, OPTIBND4 sorts the input data in increasing order of the variable X , irrespective of the categories. Thirdly, it selects a systematic sample from the data if the number of observations is too large to be handled by the user's computer. Next, it computes the OPS $\underline{x}^*$ using the procedure OPTIBND. Finally, it displays the results with some descriptive statistics. It should be noted that OPTIBND4 can handle both the cases of hypothetical CDF's $F_k(x)$ for the subpopulations of the variable X and the case of real subpopulations $\{x\}_k$. The latter case is the default. The program has been written to compute the OPS under the Neyman allocation given by (5.17) but all other possible one-step or two-step allocations can be considered by simply modifying the statements indicated within the program.

As a first illustration of the use of the program OPTIBND4, $M = 3$ categories were considered. The underlying subpopulations of the categories 1, 2 and 3 were assumed to be exponential distributions with parameters $\lambda_1 = 1.0$, $\lambda_2 = 2.0$ and $\lambda_3 = 3.0$, respectively. A description of this distribution can be found in Figure 3.3. As in Section 3.2.5, the domain $[0, +\infty[$ of this distribution is truncated to $[0, 10]$ for each of the three categories, such that the maximum of the neglected parts over the categories is negligible, 0.0045%. The OPS were computed under proportional allocation which is both a one-step and a two-step allocation. On the other hand, since the population size is infinite, the relative size θ_k of each category k has been set arbitrarily to 1/3. Since the domain of X is continuous, the interval $[0, 10]$ was crossed by discrete steps of 0.1, as in Section 3.2.5. The results obtained for the present distributions for the three categories are displayed in Table 5.1.

Table 5.1

Optimum Points of Stratification $\underline{x}^*$ for 3 Categories under Exponential Distributions with $\lambda_1 = 1.0, \lambda_2 = 2.0$ and $\lambda_3 = 3.0$, respectively, under Proportional Allocation. The Relative Sizes are Set to $\theta_1 = \theta_2 = \theta_3 = 1/3$. The Points $\hat{\underline{x}}$ are Obtained by Taking the Weighted Averages of the OPS $\underline{x}_k^*$ for Each Category k . The Points $\hat{\hat{\underline{x}}}$ are Obtained by Ignoring the Categories.

Stratum	h	1	2	3	4	5	$\frac{\text{Var}(\hat{\underline{x}}_{comb})}{\text{Var}(\hat{\underline{x}}_{st})}$	marginal gain of precision (in %)
$L = 2$	x_h^*	1.2						
	W_{1h}	0.2329	0.1004				0.408	59.2
	W_{2h}	0.3021	0.0302					
	W_{3h}	0.3242	0.0091					
	x_{1h}^*	1.6					0.350	
	x_{2h}^*	0.8						
	x_{3h}^*	0.5						
	$\hat{x}_h$	1.0					0.417	
	$\hat{\hat{x}}_h$	1.2					0.408	
$L = 3$	x_h^*	0.7	2.1					
	W_{1h}	0.1678	0.1247	0.0408			0.216	47.2
	W_{2h}	0.2511	0.0772	0.0050				
	W_{3h}	0.2925	0.0402	0.0006				
	x_{1h}^*	1.0	2.6				0.178	
	x_{2h}^*	0.5	1.3					
	x_{3h}^*	0.3	0.8					
	$\hat{x}_h$	0.6	1.6				0.235	
	$\hat{\hat{x}}_h$	0.7	2.1				0.216	

Table 5.1 (continued)

Stratum	h	1	2	3	4	5	$\frac{\text{Var}(\bar{x}_{comb})}{\text{Var}(\bar{x}_{st})}$	marginal gain of precision (in %)
$L = 4$	x_{1h}^*	0.5	1.3	2.8			0.133	38.2
	W_{1h}	0.1312	0.1113	0.0706	0.0203			
	W_{2h}	0.2107	0.0979	0.0235	0.0012			
	W_{3h}	0.2590	0.0676	0.0067	0.0001			
	x_{1h}^*	0.8	1.8	3.4			0.108	
	$\hat{x}_{2h}^*$	0.4	0.9	1.7				
	x_{3h}^*	0.3	0.6	1.1				
	$\hat{x}_h$	0.5	1.1	2.1			0.150	
$L = 5$	$\hat{\hat{x}}_h$	0.5	1.3	2.8			0.133	
	x_{1h}^*	0.4	1.0	1.9	3.4		0.090	32.8
	W_{1h}	0.1099	0.1008	0.0728	0.0387	0.0111		
	W_{2h}	0.1836	0.1047	0.0377	0.0071	0.0004		
	W_{3h}	0.2329	0.0838	0.0155	0.0011	< .00005		
	x_{1h}^*	0.6	1.3	2.3	3.8		0.073	
	x_{2h}^*	0.3	0.7	1.2	2.0			
	x_{3h}^*	0.2	0.5	0.8	1.3			
	$\hat{x}_h$	0.4	0.8	1.4	2.4		0.109	
	$\hat{\hat{x}}_h$	0.4	1.0	1.9	3.4		0.0090	

Table 5.1 shows that a non-negligible loss of precision in the estimate $\bar{x}_{comb}$ is found by imposing the constraint that the stratification must be a lattice one instead of a general one. Taking $L = 2$ for example, the ratio of the minimum variance for lattice stratification over the variance obtained under SRS for each category (i.e. $\text{Var } \bar{x}_{st}$) is found to be 0.408 while the ratio of the minimum variance for general stratification over $\text{Var } (\bar{x}_{st})$ is only equal to 0.350. In some applications, this loss of precision might suggest the choice of general stratification over lattice stratification. It should be noted that the variance obtained under general stratification is in fact the smallest possible variance for a given number L of strata in each category. Considering the reconciliation of the OPS $\bar{x}_1^*$, $\bar{x}_2^*$ and $\bar{x}_3^*$ of the three categories to obtain a single vector of points of stratification $\hat{\bar{x}}$, Table 5.1 shows that this task may be difficult to implement intuitively because of the differences in the points of stratification of the three vectors of OPS. However, using formula (5.12),

a single vector of points of stratification $\hat{x}$ was computed which, as expected, leads to a larger variance of $\bar{x}_{comb}$ than the one based on the OPS $\underline{x}^*$. For the present distributions, computing a single vector of OPS $\hat{x}$ by ignoring the categories is in fact equivalent to the one-way optimum stratification problem where the underlying population is a mixture of the three exponential distributions. It is seen from Table 5.1 that, for a given L , the points of stratification $\hat{x}$ are exactly the same as the OPS $\underline{x}^*$. This equality is however only due to the degree of precision used in computing the points of stratification. Some further investigations have shown that, by crossing the interval $[0,10]$ by steps of 0.001, these points of stratification were, in fact, different. However, the points of stratification $\hat{x}$ clearly constitutes a fairly good approximation to the OPS $\underline{x}^*$ for the present distribution.

As a second illustration of the use of the program OPTIBND4, we took the distribution of the sales of Incorporated Retailers of General Merchandise Stores in the Province of Québec as category 1 and the one for the Province of Ontario as category 2. A description of these two skewed populations of sizes $N_1 = 195$ and $N_2 = 280$, respectively, can be found in Figure 5.1 (1) and Figure 5.1 (2). If the total sample size n and the number L of strata are relatively small, these populations can be considered as infinite. As in the previous illustration, the proportional allocation is used. On the other hand, the relative sizes θ_1 and θ_2 are equal to $\frac{195}{475} = 0.411$ and $\frac{280}{475} = 0.589$ based on the definition (5.1). The results obtained are summarized in Table 5.2.

Figure 5.1 (1)

Distribution of the Sales of Incorporated Retailers of General Merchandise Stores in the Province of Québec for the Year 1985 (in Thousands of Dollars).

population size $N_1 = 195$

mean $\bar{X}_1 = 629.8$

variance $S_1^2 = 226132.4$

skewness = 0.861

kurtosis = 0.193

minimum value = 2

maximum value = 1976

Figure 5.1 (2)

Distribution of the Sales of Incorporated Retailers of General Merchandise Stores in the Province of Ontario for the Year 1985 (in Thousands of Dollars).

population size $N_2 = 280$

mean $\bar{X}_2 = 507.4$

variance $S_2^2 = 174308.4$

skewness = 1.114

kurtosis = 1.006

minimum value = 2

maximum value = 1999

Table 5.2 shows again that some loss of precision in the estimate $\bar{x}_{comb}$ is found by considering a lattice stratification rather than a general stratification. However, this loss of precision is found to be much smaller for the present real subpopulations than for the exponential distributions of the previous example. This is mainly due to the fact that the distributions of the sales for the Provinces of Québec and Ontario are found to be not so different. Using formula (5.12) to reconcile the OPS $\underline{x}_1^*$ and $\underline{x}_2^*$ obtained for categories 1 and 2 respectively, the single vector of points of stratification $\hat{\underline{x}}$ was computed for each $L = 2$ to 5. Table 5.2 shows that these points of stratification $\hat{\underline{x}}$, as expected, lead to a larger variance of $\bar{x}_{comb}$ than the one under the OPS $\underline{x}^*$. As before, the OPS $\hat{\underline{x}}$ obtained by ignoring the categories are quite close to the OPS $\underline{x}^*$. Table 5.2 shows that $\hat{\underline{x}}$ provides, at least for the present distributions and the ones of the previous illustration, a better approximation to the true $\underline{x}^*$ than the compromise points of stratification $\hat{\underline{x}}$.

Table 5.2

Optimum Points of Stratification $\underline{x}^*$ for the Sales of the Retailers of General Merchandise Stores in Canada for the Year 1985 (Category 1 is Québec with $N_1 = 195$, Category 2 is Ontario with $N_2 = 280$), under Proportional Allocation (f.p.c. ignored). The Relative Sizes are $\theta_1 = 195/475$ and $\theta_2 = 280/475$. The Points $\hat{\underline{x}}$ are Obtained by Taking the Weighted Averages of the OPS $\underline{x}_k^*$ for Each Category k . The Points $\hat{\underline{x}}$ are Obtained by Ignoring the Categories.

Stratum	h	1	2	3	4	5	$\frac{\text{Var}(\hat{\underline{x}}_{\text{comp}})}{\text{Var}(\hat{\underline{x}}_t)}$	marginal gain of precision (in%)
$L = 2$	x_h^*	689					0.3172	68.3
	W_{1h}	0.263	0.148					
	W_{2h}	0.419	0.170					
	x_{1h}^* x_{2h}^*	689 670					0.3169	
	$\hat{x}_h$	678					0.3176	
	$\hat{\hat{x}}_h$	689					0.3172	
$L = 3$	x_h^*	459	1048				0.1512	52.3
	W_{1h}	0.206	0.144	0.61				
	W_{2h}	0.294	0.225	0.070				
	x_{1h}^* x_{2h}^*	516 397	1048 962				0.1492	
	$\hat{x}_h$	446	997				0.1551	
	$\hat{\hat{x}}_h$	463	1072				0.1523	

Table 5.2 (continued)

Stratum	h	1	2	3	4	5	$\frac{\text{Var}(\bar{x}_{comb})}{\text{Var}(\bar{x}_i)}$	marginal gain of precision (in%)
$L = 4$	x_{1h}^*	340	701	1235			0.0922	39.0
	W_{1h}	0.138	0.125	0.113	0.035			
	W_{2h}	0.269	0.156	0.113	0.051			
	x_{1h}^*	248	618	1235			0.0888	
	x_{2h}^*	397	889	1442				
	$\hat{x}_h$	336	778	1357			0.0966	
	$\hat{\hat{x}}_h$	325	701	1204			0.0936	
$L = 5$	x_{1h}^*	277	618	977	1442		0.0572	38.0
	W_{1h}	0.094	0.163	0.075	0.050	0.0029		
	W_{2h}	0.225	0.156	0.125	0.069	0.014		
	x_{1h}^*	248	618	977	1319		0.0569	
	x_{2h}^*	295	594	962	1442			
	$\hat{x}_h$	276	604	968	1391		0.0621	
	$\hat{\hat{x}}_h$	267	594	977	1442		0.0575	

5.2 Finite Population Case

In this section the finite population case is considered, where the f.p.c. of the variance of $\bar{x}_{comb}$ cannot be ignored. As in Section 5.1, the variable used to substratify each category is assumed to be the variable of interest X .

In considering the problem of determining, for the finite population case, the OPS x^* for lattice stratification with one categorical stratification variable, the same approach as Dalenius and Gurney (1951) can be used. This approach basically consists of taking the partial derivatives of the formula for the variance of $\bar{x}_{comb}$ (with the f.p.c.) under a given type of allocation to obtain the minimal equations. However, apart from the fact that the minimal equations are only necessary conditions for a global minimum, some problems may occur when using a one-step or two-step allocation involving the equal-sample-size, Neyman or optimum (with a cost constraint) allocations. The studies of Schneeberger and Goller (1979) and Schneeberger (1981) reveal that the use of the minimal equations obtained under such allocations can lead to non-feasible solutions for the OPS. That is, the computed OPS may produce allocations of sample sizes n_{kh} to some strata

greater than the corresponding stratum population sizes N_{kh} . The determination of the OPS $\underline{x}^*$ under the finite population case must therefore explicitly consider the constraints $n_{kh} \leq N_{kh}$. Fortunately, the DP approach can be used to achieve this goal, as in Chapter 3. -

5.2.1 The Dynamic Programming Approach: The use of DP for determining the OPS $\underline{x}^*$ in the finite population case consists of two parts. The first part uses the two-dimensional DP approach described in Section 3.2.3. The first dimension is related to the determination of the OPS $\underline{x}^*$ while the second dimension refers to the allocation of sample sizes n_h to each interval $[x_{h-1}^+, x_h]$ across the M categories, where

$$(5.40) \quad n_h = \sum_{k=1}^M n_{kh}.$$

The second part of this DP approach is used to distribute each sample size n_h among the M categories, which provide the stratum sample sizes n_{kh} . The second dimension of the first part and the second part itself in fact permit the control of the stratum sample sizes n_{kh} in order to satisfy the constraints $n_{kh} \leq N_{kh}$. It should be noted that, unlike the two-step allocations discussed previously, this DP approach first allocates the sample sizes n_h instead of the sample sizes n_k related to each category k . The allocation of the sample sizes n_k to the categories first is, however, a more natural way to proceed since, as noted in the beginning of the chapter, the population is usually assumed to be first stratified with respect to the categories. Note that even if the sample sizes n_h are allocated first, the integration of the two parts of this DP approach makes the stratum sample sizes n_{kh} to be allocated as if a one-step allocation would be used, as it will be seen below.

Since the proportional allocation given by (5.13) produces stratum sample sizes n_{kh} always less than or equal to the corresponding stratum population sizes N_{kh} , there is no need to use this two-part DP approach under this allocation. In fact, the approach of Section 5.1 can simply be used, except adding the f.p.c. in the minimization of the variance of $\bar{x}_{comb}$ given by (5.14). This is also the case for the equal-sample-size allocation given by (5.15) because the allocated stratum sample sizes n_{kh} are fixed with respect to the points of stratification $\underline{x}$. By considering the definition of the Neyman allocation given by (5.17), it is clear that this allocation can be realized by the proposed two-part DP approach. Note that by using integer spaces for the second dimension of the first part and for the second part itself, it is possible to obtain integer stratum sample size n_{kh} , which is usually required in sample surveys. The optimum allocation (5.20) can also be considered using the two-part DP approach. The only difference with the Neyman allocation is that instead of allocating the sample sizes n_h and n_{kh} such that $n = \sum_{h=1}^L n_h$ and $n_h = \sum_{k=1}^M n_{kh}$, respectively, these are allocated based on cost constraints of the form $C = c_0 + \sum_{h=1}^L c_h n_h$ and (5.19), respectively, for given budget. Since this

DP approach allocates the sample sizes n_h instead of the sample sizes n_k of each category k , unfortunately, none of the two-step allocations discussed earlier can be handled using the proposed two-part DP method. The exception to this are, of course, the proportional-proportional and the equal-sample-size-equal-sample-size allocations which are the same as the one-step allocations given by (5.13) and (5.15), respectively. The two-part DP approach could in fact be used if the two steps of the allocation would have been changed to first allocate the sample sizes n_h .

As seen before, the DP method requires the identification of the objective function to be minimized. For the problem of the determination of the OPS for lattice stratification where one stratification variable is categorical, the objective function is of the form

$$(5.41) \quad G(\underline{x}, \underline{n}) = \sum_{h=1}^L g(x_{h-1}, x_h, \underline{n}_h)$$

$$(5.42) \quad \text{where} \quad g(x_{h-1}, x_h, \underline{n}_h) = \sum_{k=1}^M t_k(x_{h-1}, x_h, n_{kh}),$$

$$\underline{n} = (n_{11}, n_{12}, \dots, n_{1L}, n_{21}, \dots, n_{ML})$$

$$\underline{n}_h = (n_{1h}, n_{2h}, \dots, n_{Mh}), \quad h = 1, \dots, L.$$

The ML -vector $\underline{n}$ is the vector of the stratum sample sizes for all ML strata while the M -vectors $\underline{n}_h$ are the vectors of the stratum sample sizes for each interval $[x_{h-1}^+, x_h]$. Even if the DP approach is to be based on two parts, those are to be integrated in the same minimization process so that only one objective function is required. Because the allocation of the total sample size n is handled by the DP process itself, the function $t_k(x_{h-1}, x_h, n_{kh})$ can simply be obtained from equation (5.9). Hence, the function $t_k(x_{h-1}, x_h, n_{kh})$ is given by

$$(5.43) \quad t_k(x_{h-1}, x_h, n_{kh}) = W_{kh} \sigma_{(k)h}^2 \left(\frac{W_{kh}}{n_{kh}} - \frac{1}{N} \right)$$

where W_{kh} and $\sigma_{(k)h}^2$ are given by (5.7) and (5.10), respectively.

The first part of the proposed DP method is based on a two-dimensional approach similar to the one described in Section 3.2.3 where the first dimension is related to the points of stratification $\underline{x}$ and the second one to sample sizes n_h . With regard to the different concepts of DP discussed in Chapter 2, the stages of this first part are directly related to each of the L intervals given by $[x_{h-1}^+, x_h]$. As in Section 3.2.3, both state and decision spaces are two-dimensional. Letting m_h be the sample size allocated up to the interval h across the M categories

(i.e. the strata (k, j) , $k = 1, \dots, M$; $j = 1, \dots, h$), the state of the stage h is given by the two-dimensional state variable (x_h, m_h) . Note that the initial state is simply defined as $(a, 0)$. Since the sample sizes n_h have to be integers, the state space for the h th stage is then given by $[a, b] \times \{0, 1, \dots, n\}$. As for the state variable, the decision variable is two-dimensional. It is defined by (d_h, n_h) where d_h is, as in Section 3.2.3, the width of the interval $[x_{h-1}^+, x_h]$. For this decision variable, the stage transformation function is simply defined as $(x_h, m_h) = (x_{h-1} + d_h, m_{h-1} + n_h)$. In order to satisfy the constraint that the sample size n_h must be less than the corresponding population size N_h ($N_h = \sum_{k=1}^M N_{kh}$) and the other constraints related to the construction of the intervals, the decision variable (d_h, n_h) is restricted to the decision space $B_h^{(1)}(x_{h-1}) \times B_h^{(2)}(x_{h-1}, x_h, m_{h-1})$ where $B_h^{(1)}(x_{h-1}) = [0, b - x_{h-1}]$ and $B_h^{(2)}(x_{h-1}, x_h, m_{h-1}) = \{1, 2, \dots, n - m_{h-1}\} \cap \{1, 2, \dots, [W_h N]\}$. It should be noted that W_h is defined as

$$(5.44) \quad W_h = \sum_{k=1}^M \theta_k W_{(k)h} = \sum_{k=1}^M W_{kh}.$$

The second part of the proposed DP method being related to the allocation of the sample size n_h among the M categories, the DP approach to be used in this part can simply be a one-dimensional one. In terms of the general concepts of DP, the stages clearly correspond to each of the M categories for a given interval h . The sequential decision aspect can then be seen as distributing a portion of the sample size n_h to each category k . The state of each category k is then given by the state variable m_{kh} which is the sample size allocated up to category k , within the interval $[x_{h-1}^+, x_h]$. For this state variable, the initial state m_{0h} is simply zero while the state space corresponds to the set $\{0, 1, \dots, n_h\}$, for each of the L intervals. On the other hand, the decision variable of the k th category of the h th interval is the stratum sample size n_{kh} . Related to this decision variable, the stage transformation function is clearly given, for each interval h , by $m_{kh} = m_{k-1h} + n_{kh}$. In order to satisfy the constraint $n_{kh} \leq N_{kh}$, the decision variable n_{kh} is restricted to the decision space $B_{kh}^{(3)}(x_{h-1}, x_h, n_h, m_{k-1h})$ given by $B_{kh}^{(3)}(x_{h-1}, x_h, n_h, m_{k-1h}) = \{1, 2, \dots, n_h - m_{k-1h}\} \cap \{1, 2, \dots, [W_{kh} N]\}$. Using the above formulations of the two parts of the proposed DP approach, the problem of determining the OPS $\underline{x}^*$ and the optimal sample sizes $\underline{n}^*$ for the finite population case can then be expressed in the following form:

$$\begin{aligned}
(5.45) \quad & \text{Min} \sum_{h=1}^L \sum_{k=1}^M t_k(x_{h-1}, x_h, n_{kh}) \\
& \text{subject to: } (x_h, m_h) = (x_{h-1} + d_h, m_{h-1} + n_h) \dots \\
& (x_h, m_h) \in [a, b] \times \{0, 1, \dots, n\} \\
& (d_h, m_h) \in B_h^{(1)}(x_{h-1}) \times B_h^{(2)}(x_{h-1}, x_h, m_{h-1}) \\
& (x_0, m_0) = (a, 0) \\
& \text{and } m_{kh} = m_{k-1h} + n_{kh} \\
& m_{kh} \in \{0, 1, \dots, n_h\} \\
& n_{kh} \in B_{kh}^{(3)}(x_{h-1}, x_h, n_h, m_{k-1h}) \\
& m_{0h} = 0, \quad k = 1, \dots, M; h = 1, \dots, L.
\end{aligned}$$

The above form of the problem can be used to work out the second step in the DP formulation of the problem, i.e. the determination of the Functional Equation of Bellman. With respect to the first part of the proposed DP approach, let $G_h^*(x_{h-1}, m_{h-1})$ be the optimal value for the objective function (5.41) for intervals h to L given that the lower bound for the h th interval is x_{h-1} and $n - m_{h-1}$ units are allocated to the strata (k, h) to (K, L) across the M categories. On the other hand, with respect to the second part, let $T_k^*(x_{h-1}, x_h, m_{k-1h})$ be the optimal value for the minimization of the function (5.42) for category k to M , within interval h , given that $n_h - m_{k-1h}$ units are allocated to these $(M - k + 1)$ categories. The Functional Equation of Bellman for the solution of the present problem by the two-part DP approach is given by

$$\begin{aligned}
(5.46) \quad & G_h^*(x_{h-1}, m_{h-1}) = \text{Min}_{d_h \in B_h^{(1)}(x_{h-1})} \text{Min}_{n_h \in B_h^{(2)}(x_{h-1}, x_h, m_{h-1})} \\
& \{T_1^*(x_{h-1}, x_h, 0) + G_{h+1}^*(x_h, m_h) | x_h = x_{h-1} + d_h, m_h = m_{h-1} + n_h\} \\
(5.47) \quad & \text{where } T_k^*(x_{h-1}, x_h, m_{k-1h}) = \text{Min}_{n_{kh} \in B_{kh}^{(3)}(x_{h-1}, x_h, n_h, m_{k-1h})} \\
& \{t_k(x_{h-1}, x_h, n_{kh}) + T_{k+1}^*(x_{h-1}, x_h, m_{kh}) | m_{kh} = m_{k-1h} + n_{kh}\}.
\end{aligned}$$

By looking at this Functional Equation of Bellman, the existence of the two part of the proposed DP approach then becomes clear.

The proposed method for solving the problem of two-way optimum stratification with one categorical stratification variable has the advantage, as with the approach given in Section 3.2.3, that the optimal number of intervals L can simply be obtained based on a criterion of the form (3.27). Note, however, that the

number of categories M is assumed to be fixed. On the other hand, because of the use of the second part, the proposed algorithm may become much slower than the two-dimensional DP approach when the number of categories M gets large.

When the subpopulation of each of the categories is a real population $\{x\}_k$, the same approach as in Section 3.1.4 can be used. Based on this approach, the quantities of the objective function are replaced by their finite population counterparts. Hence, the quantities W_{kh} and $\sigma_{(k)h}^2$ are replaced respectively by $\frac{N_{kh}}{N}$ given by (5.39) and $S_{(k)h}^2$ given by (5.37). The function $t_k(x_{h-1}, x_h, n_{kh})$ of equation (5.43) then becomes

$$(5.48) \quad t_k(x_{h-1}, x_h, n_{kh}) = \frac{N_{kh}^2}{N^2} S_{(k)h}^2 \left(\frac{1}{n_{kh-}} - \frac{1}{N_{kh}} \right).$$

Note finally that this two-step DP method can be programmed by merging together the two programs OPTIBND1 and OPTIBND2 given in Appendices A and B, respectively.

5.3 Use of an Auxiliary Variable

Since, in practice, the stratification variable cannot be the same as the variable of interest, the points of stratification must be based on an auxiliary variable Z which is highly correlated with the variable of interest X . On the other hand, the categorical variable used for the stratification can be assumed to be the unchanged.

In this section, it is assumed that the auxiliary variable Z is sampled, for the k th of the M categories, from a subpopulation of size N_k with CDF $F_{Zk}(z)$ defined on $[a_Z, b_Z]$ with mean μ_{Zk} and finite variance σ_{Zk}^2 . It should be noted that since the subpopulation sizes N_k are the same as the ones for the variable X given in the beginning of the chapter, the relative sizes θ_k of the subpopulations remain unchanged. The purpose of this section is to describe the approach to be used to determine the OPS z^* which minimize the variance of $\bar{x}_{comb}$ for lattice stratification with one categorical stratification variable.

In defining the nature of the relationship between the variables X and Z , two different kinds of models can be assumed. The first kind may be to suppose that the relationship between the variables X and Z is the same for all the M categories. In this case, a model similar to the general model given by (3.29) can be used. On the other hand, the second kind of models assumes that the relationship between X and Z is different from category to category. The model for each category k is then given by

$$(5.49) \quad X = \Psi_k(Z) + e, \quad k = 1, \dots, M.$$

The error term e in the model (5.49) is assumed to have, for the k th category, the conditional mean $E_k(e|Z) = 0$ and the conditional variance $\text{Var}_k(e|Z) = \nu_k(Z)$. Note that this model is in fact the same one as the general model (3.29) but with

the subscript k added. Since the case where the relationship between X and Z is the same across the M categories is simply a special case of the model (5.49), the emphasis will be put on the latter model. Denoting the mean and the variance of $\Psi_k(Z)$ by $\mu_{\Psi k}$ and $\sigma_{\Psi k}^2$, respectively, the mean and the variance of the variable of interest X can then be expressed, for each category k , by

$$(5.50) \quad \mu_{Xk} = E_k(X) = E_k(\Psi_k(Z) + e) = \mu_{\Psi k}$$

$$(5.51) \quad \text{and} \quad \sigma_{Xk}^2 = \text{Var}_k(X) = \text{Var}_k(\Psi_k(Z) + e) = \sigma_{\Psi k}^2 + \mu_{\nu k}$$

$$(5.52) \quad \text{where} \quad \mu_{\nu k} = E_k(\nu_k(Z)) = \text{Var}_k(e).$$

Considering the problem of determining the OPS $\underline{z}^*$ to minimize the variance of $\bar{x}_{comb}$, the approach to be used is to express the different quantities entering in the variance of $\bar{x}_{comb}$ in terms of the points of stratification $\underline{z}$. Note that this corresponds exactly to what has been done in Section 3.3 for the minimization of the variance of $\bar{x}_s$ given by (1.4). Based on the model (5.49), the stratum subpopulation proportion $W_{(k)h}$, the stratum mean $\mu_{X(k)h}$, the stratum variance $\sigma_{X(k)h}^2$ and the stratum population proportion W_{kh} can be obtained by the following:

$$(5.53) \quad W_{(k)h} = \int_{z_{h-1}^+}^{z_h} dF_{Zk}(z),$$

$$(5.54) \quad \mu_{X(k)h} = \frac{1}{W_{(k)h}} \int_{z_{h-1}^+}^{z_h} \Psi_k(z) dF_{Zk}(z) = \mu_{\Psi(k)h},$$

$$(5.55) \quad \sigma_{X(k)h}^2 = \sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h}$$

where $\sigma_{\Psi(k)h}^2$ and $\mu_{\nu(k)h}$ are obtained following (5.10) and (5.11), and

$$(5.56) \quad W_{kh} = \theta_k W_{(k)h}$$

with $W_{(k)h}$ given by (5.53).

Using these last quantities, the variance of $\bar{x}_{comb}$ can then be expressed as a function of the points of stratification $\underline{z}$ as

$$(5.57) \quad \text{Var}(\bar{x}_{comb}) = \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh}^2 (\sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h})}{n_{kh}} - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} (\sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h})}{N}.$$

It should be noted that all the one-step or two-step allocations described in the beginning of this chapter can also be expressed as a function of the points of stratification $\underline{z}$ using exactly the same replacement process. For example, the Neyman allocation given by (5.17) can be expressed as

$$(5.58) \quad n_{kh} = \frac{W_{kh} (\sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h})^{1/2}}{\sum_{k=1}^M \sum_{h=1}^L W_{kh} (\sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h})^{1/2}} n, \quad k = 1, \dots, M; h = 1, \dots, L.$$

Under this allocation, the variance of $\bar{x}_{comb}$ becomes
(5.59)

$$\text{Var}(\bar{x}_{comb}) = \frac{1}{n} \left(\sum_{k=1}^M \sum_{h=1}^L W_{kh} (\sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h})^{1/2} \right)^2 - \sum_{k=1}^M \sum_{h=1}^L \frac{W_{kh} (\sigma_{\Psi(k)h}^2 + \mu_{\nu(k)h})}{N}$$

5.3.1 The Dynamic Programming Approach: As noted in Section 3.3.2, the DP method to be used when the stratification variable is the auxiliary variable Z is exactly the same as the one where the stratification variable is the variable of interest X . The only differences are that all the required computations are now based, for each category k , on the CDF's $F_{Zk}(z)$ of the subpopulations of the variable Z instead of X , and that the objective function under a given type of allocation is expressed in terms of the points of stratification z . Hence for both infinite and finite population cases, the OPS z^* can be obtained using exactly the same DP approaches as the ones given in Section 5.1 and 5.2, respectively. Note that for the infinite population case, the objective function is then of the form

$$(5.60) \quad G(z) = \sum_{h=1}^L \left(\sum_{k=1}^M g_k(z_{h-1}, z_h) \right)$$

On the other hand, for the finite population case, the objective function used in the two-part DP approach can be expressed as

$$(5.61) \quad G(z, n) = \sum_{h=1}^L \sum_{k=1}^M t_k(z_{h-1}, z_h, n_{kh}).$$

In the case where the subpopulations of the auxiliary variable Z are real populations $\{z\}_k$, the different quantities entering into the objective function to be minimized can again be replaced by their finite population counterparts. Assuming that the model (5.49) holds, the quantities $W_{(k)h}$, $\mu_{\Psi(k)h}$, $\sigma_{\Psi(k)h}^2$ can then be replaced by the following:

$$(5.62) \quad W_{(k)h} : \quad \frac{N_{kh}}{N_k} = \frac{1}{N_k} \{\text{number of } z\text{'s in } I_{Z(k)h}\},$$

$$(5.63) \quad \mu_{\Psi(k)h} : \quad \bar{\Psi}_{(k)h} = \frac{1}{N_k} \sum_{z \in I_{Z(k)h}} \Psi_k(z),$$

$$(5.64) \quad \sigma_{\Psi(k)h}^2 : \quad S_{\Psi(k)h}^2 = \frac{1}{N_k - 1} \sum_{z \in I_{Z(k)h}} (\Psi_k(z) - \bar{\Psi}_{(k)h})^2$$

where $I_{Z(k)h} = \{z\}_k \cap [z_{h-1}^+, z_h]$.

It should finally be noted that this analysis using the auxiliary variable Z is based on the assumption that the model (5.49) is completely known for each of the M categories. The determination of the model (5.49) as well as the form of the conditional variance $\nu_k(Z)$ of the error term may, in practice, be a difficult task.

6. TWO-WAY OPTIMUM STRATIFICATION

The methods of optimum stratification considered in the previous chapters are 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 chapter, 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 finite covariance matrix $\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$, i.e. $|\Sigma| < \infty$ where $|A|$ denotes the determinant of the matrix A . The marginal distributions of X and Y will be denoted by $F_X(x)$ and $F_Y(y)$, respectively. Given that the population is to be divided into $L \times M$ strata, the problem considered here is to obtain OPS $\underline{x}^* = (x_0, x_1, \dots, x_L)$ with respect to the variable X together with OPS $\underline{y}^* = (y_0, y_1, \dots, y_M)$ with respect to the variable Y to form the optimum lattice stratification for the estimation of the means μ_X and μ_Y . For the case where the stratification is based on auxiliary variables Z and U , these OPS will be $\underline{z}^*$ and $\underline{u}^*$, respectively. Although there may be a loss of precision compared to the use of a general optimum stratification, the lattice stratification offers simplicity and practical convenience.

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

$$(6.1) \quad \bar{x}_{st} = \sum_{h=1}^L \sum_{k=1}^M W_{hk} \bar{x}_{hk}$$

$$(6.2) \quad \text{where} \quad \bar{x}_{hk} = \frac{1}{n_{hk}} \sum_{i=1}^{n_{hk}} x_{hki},$$

$$(6.3) \quad W_{hk} = \int_{x_{k-1}^+}^{x_k} \int_{y_{k-1}^+}^{y_k} dF(x, y).$$

The estimator $\bar{y}_{st}$ is similarly obtained from (6.1) by substituting y 's for x 's. It should be noted that the stratum population proportion W_{hk} given by (6.3) differs from the one denoted by W_{kh} used in Chapter 5. Related to each of these estimators is its sampling variance which is given, for the estimator $\bar{x}_{st}$, by

$$(6.4) \quad \text{Var}(\bar{x}_{st}) = \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}^2 \sigma_{Xhk}^2}{n_{hk}} - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} \sigma_{Xhk}^2}{N}$$

$$(6.5) \quad \text{where} \quad \sigma_{Xhk}^2 = \frac{N}{NW_{hk} - 1} \int_{x_{k-1}^+}^{x_k} \int_{y_{k-1}^+}^{y_k} (x - \mu_{Xhk})^2 dF(x, y),$$

$$(6.6) \quad \mu_{Xhk} = \frac{1}{N_{hk}} \int_{x_{k-1}^+}^{x_k} \int_{y_{k-1}^+}^{y_k} x dF(x, y).$$

An expression similar to (6.4) can be derived for the variance of $\bar{y}_{st}$. Note that the stratum mean μ_{Yhk} and the stratum variance σ_{Yhk}^2 for the variable Y can also be obtained using equations (6.6) and (6.5), respectively. The sampling covariance between $\bar{x}_{st}$ and $\bar{y}_{st}$ is defined as

$$(6.7) \quad \text{Cov}(\bar{x}_{st}, \bar{y}_{st}) = \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}^2 \sigma_{XYhk}}{n_{hk}} - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} \sigma_{XYhk}}{N}$$

$$(6.8) \quad \text{where} \quad \sigma_{XYhk} = \frac{N}{NW_{hk} - 1} \int_{x_{hk-1}^+}^{x_h} \int_{y_{k-1}^+}^{y_k} (x - \mu_{Xhk})(y - \mu_{Yhk}) dF(x, y),$$

μ_{Xhk} and μ_{Yhk} are obtained from (6.6). As usual, the last terms of the variance of $\bar{x}_{st}$, the variance of $\bar{y}_{st}$ and their covariance are the f.p.c.

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 in the case of one-way stratification, the objective function for two-way stratification 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, say D_{XY} , is defined by the determinant of the covariance matrix of $\bar{x}_{st}$ and $\bar{y}_{st}$. Formally,

$$(6.9) \quad D_{XY} = \begin{vmatrix} \text{Var}(\bar{x}_{st}) & \text{Cov}(\bar{x}_{st}, \bar{y}_{st}) \\ \text{Cov}(\bar{x}_{st}, \bar{y}_{st}) & \text{Var}(\bar{y}_{st}) \end{vmatrix} = \text{Var}(\bar{x}_{st}) \text{Var}(\bar{y}_{st}) - (\text{Cov}(\bar{x}_{st}, \bar{y}_{st}))^2.$$

The generalized variance is in fact a measure of the size of the covariance matrix. 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

$$(6.10) \quad \begin{aligned} V_{XY} &= \theta_X \text{Var}(\bar{x}_{st}) + \theta_Y \text{Var}(\bar{y}_{st}) \\ &= \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}^2 (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{n_{hk}} - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N} \end{aligned}$$

It should be noted that the last objective function (6.10) does not take into account the covariance between $\bar{x}_{st}$ and $\bar{y}_{st}$ unlike the generalized variance (6.9). However, due to its simple form and the fact that it is often used as a basis to obtain an optimal sample size allocation (e.g., Dalenius (1957) and Cochran (1977)), the weighted average V_{XY} , given by (6.10), is to be preferred to the generalized variance D_{XY} given by (6.9). In this chapter, the determination of OPS is considered only under the weighted average V_{XY} .

In practice, another approach different from minimizing an objective function like the weighted average V_{XY} or the generalized variance D_{XY} is often used. This approach basically determines the points of stratification with respect to the variable X independently

from those with respect to the variable Y . By considering on one hand the marginal CDF $F_X(x)$, some OPS $\hat{x}$ are obtained by minimizing the variance of $\bar{x}_h$ given by (1.4). On the other hand, some other OPS $\hat{y}$ are similarly obtained based on the marginal CDF $F_Y(y)$. This approach is generally justified by assuming that the correlation (or the covariance) between the two variables of interest X and Y is negligible. Note that this approach permits, in fact, the use of the methods for the one-way optimum stratification problem discussed in Chapter 3. Because this approach is usually not optimal in the sense that, in general it will not minimize the weighted average V_{XY} given by (6.10), it may be interesting to assess the loss in precision by using such an approach.

Since the values of the OPS generally depend on the type of allocation used, the minimization of the weighted average (6.10) may be done by considering each of a given number of allocations. As in Chapter 5, the allocation of the total sample size n for two-way stratification based on the points of stratification $(\underline{x}, \underline{y})$ (or $(\underline{z}, \underline{u})$ if the auxiliary variables Z and U are used) can be done either on a one-step or a two-step basis. The one-step type of allocation distributes the total sample size n directly to each of the LM strata. On the other hand, the two-step type of allocation is carried out by first distributing the total sample size n among the L intervals formed by the points of stratification $\underline{x}$, say, and then the resulting sample sizes n_h are redistributed among the M strata of each interval. Note that this can also be performed by first considering the M intervals formed by the points of stratification $\underline{y}$, which produces in general a different allocation.

The four different one-step allocations considered in this chapter are in fact the same ones as those discussed in Chapter 5 but applied to the two-way stratification formed by the vector of points of stratification $(\underline{x}, \underline{y})$ (or $(\underline{z}, \underline{u})$ if Z and U are used as stratification variables). Recall that those one-step allocations were the proportional, equal-sample-size, Neyman and optimum (based on a cost constraint) allocations. For the present problem, the proportional allocation is given by

$$(6.11) \quad n_{hk} = W_{hk}n, \quad h = 1, \dots, L; k = 1, \dots, M$$

where W_{hk} is given by (6.3). Under this allocation, the weighted average V_{XY} , given by (6.10), becomes

$$(6.12) \quad V_{XY} = \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{n} = \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N}$$

The equal-sample-size allocation, on the other hand, is again given by (5.15). The weighted average V_{XY} under this allocation is simply expressed as

$$(6.13) \quad V_{XY} = \frac{LM}{n} \sum_{h=1}^L \sum_{k=1}^M W_{hk}^2 (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2) = \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N}$$

Isii and Taga (1969) defined the Neyman allocation in the multi-way stratification problem to be the one which will minimize the chosen objective function. By taking the partial

derivatives of the weighted average (6.10) subject to the constraint $\sum_{h=1}^L \sum_{k=1}^M n_{hk} = n$, the Neyman allocation for the present problem is defined as

$$(6.14) \quad n_{hk} = \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2}}{\sum_{h=1}^L \sum_{k=1}^M W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2}} n, \quad h = 1, \dots, L; k = 1, \dots, M.$$

Replacing the stratum sample size n_{hk} in (6.10) by (6.14), the weighted average V_{XY} under this allocation is given by

$$(6.15) \quad V_{XY} = \frac{1}{n} \left(\sum_{h=1}^L \sum_{k=1}^M W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2} \right)^2 - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N}.$$

The optimum allocation can be obtained by minimizing the weighted average V_{XY} under the cost constraint (5.19). The resulting allocation is defined as

$$(6.16) \quad n_{hk} = \frac{(C - c_0) W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2} / \sqrt{c_{hk}}}{\sum_{h=1}^L \sum_{k=1}^M W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2} \sqrt{c_{hk}}} n, \quad h = 1, \dots, L; k = 1, \dots, M.$$

Under this allocation, the weighted average V_{XY} can be expressed as

$$(6.17) \quad V_{XY} = \frac{1}{(C - c_0)} \left(\sum_{h=1}^L \sum_{k=1}^M W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2} \sqrt{c_{hk}} \right)^2 - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N}.$$

Since these one-step allocations are basically the same as the ones discussed in Chapter 3, all the related comments for one-way stratification still hold for the two-way case.

Turning to the two-step type of allocation, each of the two steps corresponds, in fact, to an allocation in a one-way stratification. A two-step allocation can simply be formed by any two of the four allocations given in Chapter 3, i.e., the proportional, equal-sample-size, Neyman and optimum allocations given by (3.1), (3.5), (3.3) and (3.8), respectively. Note that a two-step allocation is dependent on the order in which the variables are considered, i.e., distributing first the total sample size n among the L intervals formed by the points of stratification x will usually yield a different allocation than distributing first among the M intervals formed by y . Because of this, a convenient way to label the two-step allocations is to first put the name of the allocation of the first step followed by the name of one of the second step separated by a dash, as in Chapter 5, but with the subscripts X and Y added to indicate the variable involved. Since each step is an allocation in a one-way stratification, the allocation of each step is defined in terms of the marginal CDF $F_X(x)$ or $F_Y(y)$. The proportional, equal-sample-size, Neyman or optimum (with a cost constraint) allocations to be used in the first or second step can simply be obtained directly from equations (3.1), (3.5), (3.3) and (3.8), respectively, since the marginal CDF's $F_X(x)$ and $F_Y(y)$ are in fact equivalent to the CDF $F(x)$ assumed for X in Chapter 3. Considering

the Neyman_X-Neyman_Y allocation, for example, the distribution of the total sample size n among the L intervals formed by the points of stratification $\underline{x}$ can be obtained from

$$(6.18) \quad \hat{n}_h = \frac{W_{Xh}\sigma_{Xh}}{\sum_{h=1}^L W_{Xh}\sigma_{Xh}} n, \quad h = 1, \dots, L,$$

where W_{Xh} and σ_{Xh}^2 are obtained following (1.3) and (1.5), respectively.

For the second step, the allocation of the sample size n_h among the M strata of each interval h is expressed, for this example, by

$$(6.19) \quad n_{hk} = \frac{W_{Yk}\sigma_{Yk}}{\sum_{k=1}^M W_{Yk}\sigma_{Yk}} n_h, \quad h = 1, \dots, L; k = 1, \dots, M,$$

where W_{Yk} and σ_{Yk}^2 are obtained following (1.3) and (1.5), respectively.

Combining (6.18) and (6.19), the resulting Neyman_X-Neyman_Y allocation is given by

$$(6.20) \quad n_{hk} = \frac{W_{Yk}\sigma_{Yk}W_{Xh}\sigma_{Xh}}{\left(\sum_{k=1}^M W_{Yk}\sigma_{Yk}\right)\left(\sum_{h=1}^L W_{Xh}\sigma_{Xh}\right)} n, \quad h = 1, \dots, L; k = 1, \dots, M.$$

Under this allocation, the weighted average (6.10) can be expressed as

$$(6.21) \quad V_{XY} = \frac{1}{n} \left[\sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}(\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{W_{Yk}W_{Xh}\sigma_{Yk}\sigma_{Xh}} \right] \left[\sum_{k=1}^M W_{Yk}\sigma_{Yk} \right] \left[\sum_{h=1}^L W_{Xh}\sigma_{Xh} \right] - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}(\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N}.$$

Another example of a two-step allocation is the equal-sample-size_Y-proportional_X allocation which reduces to

$$(6.22) \quad n_{hk} = W_{Xh} \frac{n}{M}, \quad h = 1, \dots, L; k = 1, \dots, M.$$

Note that this allocation differs from the equal-sample-size_X-proportional_Y allocation. Under the allocation (6.22), the weighted average V_{XY} , given by (6.10), becomes

$$(6.23) \quad V_{XY} = \frac{M}{n} \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}(\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{W_{Xh}} - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}(\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{N}.$$

In the finite population case, some attention must be given if one of the two steps uses the equal-sample-size, Neyman or optimum (under a cost constraint) allocation because it may happen that the stratum sample size n_{hk} is greater than the corresponding stratum

population size N_{hk} for one or more strata. When $n_{hk} > N_{hk}$, the previous expressions for V_{XY} are unfortunately not valid. It should finally be noted that, instead of using the marginal CDF $F_Y(y)$ for example, the allocation in the second step could also be done using the conditional distribution of Y within each interval formed by the points of stratification $\underline{x}$ (i.e. conditioning on $X \in [x_{h-1}^+, x_h]$ for each interval h). The resulting two-step type of allocation would be, in general, different from the one using the marginal CDF for the allocation at the second step. These allocations are not considered in this thesis.

6.1 Infinite Population Case

This section considers the case where the population size is infinite or at least large enough, compared to the total sample size n and the number L of strata, so that the f.p.c. can be ignored. It is also assumed that the stratification variables are the variables of interest X and Y .

6.1.1 Literature Review: 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 D_{XY} given by (6.9) ignoring the f.p.c., and under the proportional allocation. Schneeberger (1970) expressed the problem of minimizing the generalized variance D_{XY} as a non-linear programming problem and proposed algorithms to solve the cases where there are two or more stratification variables. Considering the problem of determining a general optimum stratification with several stratification variables, Golder and Yeomans (1973) suggested the use of cluster algorithms, which minimize the total Within-Sum-Of-Squares of the stratification variables. These algorithms are generally based on the re-allocation of units to different strata until no further reduction of the objective function is found. Along the same lines, Dahmström and Hagnell (1978) suggested an iterative algorithm to minimize a linear combination of the variances of the mean estimate of each variable under PPS sampling with replacement. Sadasivan and Aggarwal (1978), in trying to extend the work of Ghosh (1963), obtained the minimal equations for the generalized variance D_{XY} under an approximation to Neyman allocation (defined on the minimization of D_{XY}). An approximate rule, similar to Ekman's (1959) rule for the univariate case, was obtained. Finally, Jarque (1981) suggested different objective functions that can be used for the determination of a general optimum stratification using cluster algorithms.

6.1.2. Identification of Problems: The approach of Ghosh (1963) has often been quoted in the literature especially the result, given as Corollary 2, that the OPS for a two by two ($L = M = 2$) lattice stratification of a symmetric bivariate distribution, under proportional allocation, is the center of gravity. Ghosh (1963) proved this result by simply showing that the minimal equations are satisfied by taking the points of stratification as the center of gravity. However, as noted earlier, the minimal equations are only necessary but not sufficient conditions for achieving a global minimum of the objective function. In considering the bivariate normal distribution with mean $(\mu_X, \mu_Y) = (0, 0)$ and covariance matrix $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, Schneeberger and Pollöt (1985) showed in fact that, for reasonably large correlation ρ , Corollary 2 of Ghosh (1963) would lead to a saddle-point. Note that some examples of symmetric distributions could also be constructed for which the center

of gravity would give a local maximum for the objective function. 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 chapters, such a method can again be developed by the use of DP.

6.1.3 The Dynamic Programming Approach: As in the previous chapters, the first step in formulating the present problem in one similar to (2.1) is to identify the objective function to be minimized. The problem being the determination of the vector of OPS $(\underline{x}^*, \underline{y}^*)$ for two-way optimum stratification, the objective function can be expressed in the form

$$(6.24) \quad G(\underline{x}, \underline{y}) = \sum_{h=1}^L \sum_{k=1}^M g(x_{h-1}, x_h, y_{k-1}, y_k).$$

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} obtained under a given allocation. The one-step allocations, namely the proportional, equal-sample-size, Neyman and optimum allocations given by (6.11), (5.15), (6.14) and (6.16), respectively, all yield equations for the weighted average V_{XY} that can be put in the form (6.24). As an example, for the proportional allocation, the function $g(x_{h-1}, x_h, y_{k-1}, y_k)$ is obtained from (6.12), as

$$(6.25) \quad g(x_{h-1}, x_h, y_{k-1}, y_k) = W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2) \quad , h = 1, \dots, L; k = 1, \dots, M.$$

Similarly, using equation (6.15), the function $g(x_{h-1}, x_h, y_{k-1}, y_k)$ for the Neyman allocation is given by

$$(6.26) \quad g(x_{h-1}, x_h, y_{k-1}, y_k) = W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)^{1/2} \quad , h = 1, \dots, L; k = 1, \dots, M.$$

Unfortunately, for the two-step allocations, the only ones that permit the resulting expression of V_{XY} in the form (6.24) are those which do not involve Neyman nor optimum allocations in any of the two steps. For example, the formula (6.21) for V_{XY} under Neyman_X-Neyman_Y allocation cannot be put in the form (6.24), even if the f.p.c. is ignored. However, under the equal-sample-size_Y-proportional_X allocation, the functions $g(x_{h-1}, x_h, y_{k-1}, y_k)$, using equation (6.23), is simply given by

$$(6.27) \quad g(x_{h-1}, x_h, y_{k-1}, y_k) = \frac{W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2)}{W_{Xh}} \quad , h = 1, \dots, L; k = 1, \dots, M.$$

The general equation (6.24) for the objective function suggests that, for the determination of two-way optimum stratification, a two-dimensional DP approach, along the same lines as the one given in Section 3.2.3, should be used. Employing the general concepts of DP given in Chapter 2, the stages may be specified by the pairs (h, k) . The state of the stage (h, k) may be described by the two-dimensional state variable (x_h, y_k) . Note that the initial state (x_0, y_0) is then clearly given by (a_X, a_Y) . Since x_h can take in theory any value in the interval $[a_X, b_X]$ and y_h any value in $[a_Y, b_Y]$, the state space at each stage (h, k) can be given by $[a_X, b_X] \times [a_Y, b_Y]$. On the other hand, in the two-dimensional decision

variable (d_h, c_k) , the variable d_h may correspond to the width of the interval $[x_{h-1}^+, x_h]$ while the variable c_k to the width of the interval $[y_{k-1}^+, y_k]$. (Do not confuse with the costs entering into the optimum allocation (6.16).) The state transformation function is simply given by $(x_h, y_k) = (x_{h-1} + d_h, y_{k-1} + c_k)$. Finally, since the domain of (X, Y) , i.e. $[a_X, b_X] \times [a_Y, b_Y]$, is bounded, the decision variable (d_h, c_k) should be restricted to the decision space $B_h^{(X)}(x_{h-1}) \times B_k^{(Y)}(y_{k-1})$, where $B_h^{(X)}(x_{h-1}) = [0, b_X - x_{h-1}]$ and $B_k^{(Y)}(y_{k-1}) = [0, b_Y - y_{k-1}]$. Based on these definitions, the problem of the two-way optimum stratification in the infinite population case can be expressed as

$$(6.28) \quad \text{Min} \sum_{h=1}^L \sum_{k=1}^M g(x_{h-1}, x_h, y_{k-1}, y_k)$$

subject to:

$$(x_h, y_k) = (x_{h-1} + d_h, y_{k-1} + c_k)$$

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

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

$$(x_0, y_0) = (a_X, a_Y) \quad , h = 1, \dots, L; k = 1, \dots, M.$$

Although the formulation (6.28) seems to correspond to the one given by (2.1), a problem 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 (6.28), the decision space $B_h^{(X)}(x_{h-1}) \times B_k^{(Y)}(y_{k-1})$ in fact 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 contradicted. It should be noted that the problem is partially due to the way the stages have been defined. In fact, instead of $L \times M$ stages related to each stratum (h, k) , there are in reality $(L + M)$ stages which can be viewed through the decision variables d_h and c_k . However, due to the nature of the objective function given in general by (6.24), its transformation to reflect the $(L + M)$ stages does not seem to be mathematically tractable except for a particular case. This is the case where, under proportional allocation, the weighted average V_{XY} given by (6.12) is to be minimized when the variables X and Y are independent. In this case, $\mu_{Xhk} = \mu_{Xh}$ for $k = 1, \dots, M$ and $\mu_{Yhk} = \mu_{Yk}$ for $h = 1, \dots, L$ so that, ignoring the f.p.c., the weighted average V_{XY} simply becomes

$$(6.29) \quad V_{XY} = \frac{1}{n} \sum_{h=1}^L \theta_X \int_{x_{h-1}^+}^{x_h} (x - \mu_{Xh})^2 dF_X(x) + \frac{1}{n} \sum_{k=1}^M \theta_Y \int_{y_{k-1}^+}^{y_k} (y - \mu_{Yk})^2 dF_Y(y)$$

$$= \sum_{h=1}^L g_X(x_{h-1}, x_h) + \sum_{k=1}^M g_Y(y_{k-1}, y_k).$$

Even though this special form permits direct use of the method of DP, its special nature is too restrictive with respect to the initial problem considered in this chapter. Another DP approach is therefore needed.

We propose a simple approach which permits a solution to the problem (6.28) using the unidimensional DP iteratively. This approach consists in first fixing the points of stratification with respect to one of X or Y and then computing the OPS for the non-fixed variable. In the second step, the resulting OPS are fixed and new OPS are then determined for the other variable which was first considered as fixed. This iterative process is continued until convergence is attained under on a predetermined criterion. It should be noted that a similar approach has been used by Goodman (1979) to obtain solutions to maximum likelihood equations for log-linear models where the method used was the well-known unidimensional Newton's method. In a formal way, the unidimensional iterative DP approach can be described as follows: Before the first iteration, some arbitrary values, say $\underline{x}^{(0)}$ and $\underline{y}^{(0)}$ such that $a_X = x_0^{(0)} < x_1^{(0)} < \dots < x_{h-1}^{(0)} < x_L^{(0)} = b_X$ and $a_Y = y_0^{(0)} < y_1^{(0)} < \dots < y_{M-1}^{(0)} < y_M^{(0)} = b_Y$, are chosen for the initial points of stratification with respect to the variable X and Y , respectively. Then for the i th iteration (where $i = 1, 2, \dots$), 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 Chapter 5. This can be seen by comparing the formulation (6.28) to the one given by (5.32) with the values of the points of stratification $\underline{y}$ taken as constants in (6.28). Let $G_{Xh}^*(x_{h-1}, \underline{y}^{(i-1)})$ be the optimal value for the objective function (6.24) for the strata (h, k) to (L, k) for all $k = 1, \dots, M$ given that the lower bound for the strata (h, k) for $k = 1, \dots, M$ is x_{h-1} . The Functional Equation of Bellman with respect to the first part of the i th iteration is then given by

(6.30)

$$G_{Xh}^*(x_{h-1}, \underline{y}^{(i-1)}) = \min_{d_h \in B_h^{(X)}(x_{h-1})} \left\{ \sum_{k=1}^M g(x_{h-1}, x_h, y_{k-1}^{(i-1)}, y_k^{(i-1)}) + G_{Xh+1}^*(x_h, \underline{y}^{(i-1)}) | x_h = x_{h-1} + d_h \right\}$$

where $B_h^{(X)}(x_{h-1})$ is given in (6.28).

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 i th iteration are given by $(\underline{x}^{(i)}, \underline{y}^{(i-1)})$. For the second part of i th iteration, the points of stratification $\underline{x}^{(i)}$ are, in turn, considered as fixed. Using again the similarity that exists with the formulations (6.28) and (5.32) but now with the values of $\underline{x}$ taken as constants in (6.28), let $G_{Yk}^*(\underline{x}^{(i)}, y_{k-1})$ be the optimal value for the objective function (6.24) for the strata (h, k) to (h, M) for all $h = 1, \dots, L$ given that the lower bound for the strata (h, k) for $h = 1, \dots, L$ is y_{k-1} . The Functional Equation of Bellman for the second part of the i th iteration is then given by

(6.31)

$$G_{Yk}^*(\underline{x}^{(i)}, y_{k-1}) = \min_{c_k \in B_k^{(Y)}(y_{k-1})} \left\{ \sum_{h=1}^L g(x_{h-1}^{(i)}, x_h^{(i)}, y_{k-1}, y_k) + G_{Yk+1}^*(\underline{x}^{(i)}, y_k) | y_k = y_{k-1} + c_k \right\}$$

where $B_k^{(Y)}(y_{k-1})$ is given in (6.28).

Using expression (6.31), the values of the OPS $\underline{y}^{(i)}$ can be determined to replace the previous values $\underline{y}^{(i-1)}$ so that, for the i th iteration, the OPS are given by $(\underline{x}^{(i)}, \underline{y}^{(i)})$. As stated earlier, this process iterates until some convergence criterion is satisfied. A possible convergence criterion is to stop iterating when the maximum relative differences between the points of stratification of the $(i-1)$ th and the i th iterations for the variables X and Y are less than preassigned levels ϵ_X and ϵ_Y respectively, i.e.

$$(6.32) \quad \max_{h=1}^L \left(|x_h^{(i)} - x_h^{(i-1)}| / x_h^{(i-1)} \right) < \epsilon_X$$

$$(6.33) \quad \text{and} \quad \max_{k=1}^M \left(|y_k^{(i)} - y_k^{(i-1)}| / y_k^{(i-1)} \right) < \epsilon_Y$$

where $\epsilon_X > 0$ and $\epsilon_Y > 0$.

In Appendix F a proof of convergence to the correct limit is given. It is shown under simple assumptions that, if the unidimensional iterative DP approach converges, the convergence limit is a local minimum with probability 1. 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. McCormick (1983) gives some ways by which general optimization methods can insure the determination of a global minimum.

One of the main disadvantages in using the unidimensional iterative DP approach is that convergence is slow. Note that this is true in most other cases where a univariate iterative approach is used instead of a multivariate one. Another disadvantage in using the proposed approach is that it does not easily permit the determination of the optimum number of strata, unlike in the previous chapters. This is unfortunate since this possibility was in fact one of the main advantages of using DP. Hence, when using the proposed approach, the values of L and M have to be specified in the statement of the problem.

In practice, it often occurs that the population of the variables X and Y is a real population $\{(x, y)\}$ of N units. The OPS $\underline{x}^*$ and $\underline{y}^*$ are then obtained following the same approach as the one described in Section 3.1.4. Following this approach, the OPS are obtained using the proposed unidimensional iterative DP method, but the different quantities entering into the objective function (6.24) are replaced by their finite population counterparts. The quantities W_{hk} , σ_{Xhk}^2 and W_{Xh} are then replaced by

$$(6.34) \quad W_{hk} : \quad \frac{N_{hk}}{N} = \frac{1}{N} \{ \text{number of } (x, y) \text{ such that } (x, y) \in I_{hk} \}$$

$$(6.35) \quad \sigma_{Xhk}^2 : S_{Xhk}^2 = \frac{1}{N_{hk} - 1} \sum_{(x,y) \in I_{hk}} (x - \bar{X}_{hk})^2$$

$$(6.36) \quad \text{where } \bar{X}_{hk} = \frac{1}{N_{hk}} \sum_{(x,y) \in I_{hk}} x$$

and $I_{hk} = [x_{h-1}^+, x_h] \times [y_{k-1}^+, y_k]$,

$$(6.37) \quad W_{Xh} : \frac{N_h}{N} = \frac{1}{N} \{\text{number of } x\text{'s in } [x_{h-1}^+, x_h]\}.$$

The quantities σ_{Yhk}^2 and W_{Yk} are replaced similarly by S_{Yhk}^2 and $\frac{N_k}{N}$, respectively. The lower bounds $(x_0, y_0) = (a_X, a_Y)$ are now given by $(x_0, y_0) = (\min \{x\}, \min \{y\})$ while the upper bounds $(x_L, y_M) = (b_X, b_Y)$ are given by $(x_L, y_M) = (\max \{x\}, \max \{y\})$ where $\{x\}$ and $\{y\}$ are the marginal populations with respect to X and Y , respectively. Finally, it may be noted that, by setting ϵ_X and ϵ_Y of (6.32) and (6.33) to be respectively less than the minimum difference between two consecutive x 's and two consecutive y 's in the population $\{(x, y)\}$, the process will stop iterating in a given number of steps.

6.1.4. Computer Program and Illustrations: In order to actually compute the OPS (x^*, y^*) for the two-way optimum stratification problem in the infinite population case, the program OPTIBND5 has been written. This program, for which a listing can be found in Appendix E, is essentially a modification of the program OPTIBND4 used for the two-way stratification problem with a categorical stratification variable. However, OPTIBND5 involves nine steps instead of the five in the preceding programs. First, it reads the data (the population $\{(x, y)\}$ for example) from an external file. Secondly, it sorts the input data by ascending order of the first variable (i.e. X) using the procedure SORTSHLL. Thirdly, if the number of observations is too large to be handled by the user's computer, it selects a systematic sample from the input data. Fourthly, it sorts the input data in ascending order of the second variable (i.e. Y) using again the procedure SORTSHLL. Next, if the number of observations has been found to be too large to be handled by the user's computer, it selects a second systematic sample, which is different from the first one since the data is not sorted according to the same variable. At the sixth step, it computes the OPS with respect to the first variable using CUMFSTAR1. Next, it displays the OPS with other descriptive statistics for the first part of the iteration. At the eighth step, it computes the OPS with respect to the second variable using CUMFSTAR2. Finally, it displays, for the ninth step, the OPS obtained for the second part of the iteration along with other descriptive statistics. The program OPTIBND5 performs the steps 6 to 9 until convergence is attained according to the criteria (6.32) and (6.33). As the other programs, OPTIBND5 can be used in the case of a population with an hypothetical CDF $F(x, y)$ or in the case of a real population $\{(x, y)\}$. The later is the default option. The default allocation used by this program is the one-step Neyman allocation given by (6.14). Note that all the other one-step or two-step allocations that can be considered for the DP approach can also be handled by OPTIBND5 by simply changing the indicated code within the program.

Two illustrations of the use of the program OPTIBND5 are given. The first is based on a bivariate exponential distribution with parameters $\lambda_X = 1.0$ and $\lambda_Y = 2.0$ for the

variables X and Y , respectively. This distribution, which basically involves the product of two univariate exponential distributions, is described in Figure 6.1.

Figure 6.1

The Bivariate Exponential Distribution with Parameters $\lambda_X = 1.0$ and $\lambda_Y = 2.0$.

$$\text{density } f(x) = \begin{cases} \lambda_X \lambda_Y e^{-\lambda_X x - \lambda_Y y} & , x \geq 0, y \geq 0, \lambda_X > 0, \lambda_Y > 0 \\ 0 & , \text{elsewhere.} \end{cases}$$

means : $1/\lambda_X = 1.000$ $1/\lambda_Y = 0.500$
 variance : $1/\lambda_X^2 = 1.000$ $1/\lambda_Y^2 = 0.250$
 covariance : 0.000

symbol	F	symbol	F	symbol	F
.....	0.0 - 0.2	+++++	0.8 - 1.0	.....	1.6 - 1.8
.....	0.2 - 0.4	ooooo	1.0 - 1.2	.....	1.8 - 2.0
.....	0.4 - 0.6	xxxxx	1.2 - 1.4		
.....	0.6 - 0.8	wwwww	1.4 - 1.6		

Since the domain of (X, Y) is the area $[0, +\infty] \times [0, +\infty]$, it is truncated as $[0, 10] \times [0, 10]$ such that the neglected part represents only 0.0045% of the whole distribution. On the other hand, since the domain area is continuous, each of the intervals $[0, 10]$ is crossed by discrete steps of 0.25. It should be noted that the value 0.25 is used only for illustration of OPTIBND5 and is not designed to provide precise results for the OPS (x^*, y^*) . Regarding the weighted average V_{XY} to be minimized, the relative importances θ_X and θ_Y were set equal to 0.5, giving equal importances to X and Y . The total sample size n was allocated by the Neyman allocation which is a one-step allocation. The results obtained from the use of the program OPTIBND5 are summarized in Table 6.1.

2 of/de 2

Table 6.1

Optimum Points of Stratification ($\underline{x}^*, \underline{y}^*$) for the Bivariate Exponential Distribution with $\lambda_X = 1.0$ and $\lambda_Y = 2.0$, under Neyman allocation. The relative importances are set to $\theta_X = \theta_Y = 1/2$. The Points $\hat{x}$ and $\hat{y}$ are Obtained by Considering X and Y Separately.

M = 2				
L = 2	y ₁ [*] = 0.75			V _{XY} /V _{SRS}
	x ₁ [*] = 1.25	W ₁₁ = 0.554	W ₁₂ = 0.159	0.306
		n ₁₁ /n = 0.362	n ₁₂ /n = 0.157	
	W ₂₁ = 0.223	W ₂₂ = 0.064		
		n ₂₁ /n = 0.336	n ₂₂ /n = 0.115	
	x̂ ₁ = 1.25	ŷ ₁ = 0.75		0.306
M = 2				
L = 3	y ₁ [*] = 0.75			V _{XY} /V _{SRS}
	x ₁ [*] = 0.75	W ₁₁ = 0.410	W ₁₂ = 0.118	0.182
		n ₁₁ /n = 0.255	n ₁₂ /n = 0.134	
x ₂ [*] = 2.00	W ₂₁ = 0.262	W ₂₂ = 0.075		
	n ₂₁ /n = 0.222	n ₂₂ /n = 0.096		
	W ₃₁ = 0.105	W ₃₂ = 0.030		
	n ₃₁ /n = 0.223	n ₃₂ /n = 0.070		
	x̂ ₁ = 0.75	ŷ ₁ = 0.75		0.182
	x̂ ₂ = 2.00			
M = 3				
L = 2	y ₁ [*] = 0.50 y ₂ [*] = 1.25			V _{XY} /V _{SRS}
	x ₁ [*] = 1.25	W ₁₁ = 0.451	W ₁₂ = 0.204	0.268
		n ₁₁ /n = 0.292	n ₁₂ /n = 0.142	
W ₂₁ = 0.181		W ₂₂ = 0.082		
n ₂₁ /n = 0.315		n ₂₂ /n = 0.144		
		W ₁₃ = 0.059	0.062	
		W ₂₃ = 0.024		
		n ₂₃ /n = 0.045		
	x̂ ₁ = 1.25	ŷ ₁ = 0.50	ŷ ₂ = 1.25	0.268
M = 3				
L = 3	y ₁ [*] = 0.50 y ₂ [*] = 1.25			V _{XY} /V _{SRS}
	x ₁ [*] = 0.75	W ₁₁ = 0.334	W ₁₂ = 0.151	0.146
		n ₁₁ /n = 0.200	n ₁₂ /n = 0.105	
x ₂ [*] = 2.00	W ₂₁ = 0.213	W ₂₂ = 0.096		
	n ₂₁ /n = 0.188	n ₂₂ /n = 0.091		
	W ₃₁ = 0.086	W ₃₂ = 0.039		
	n ₃₁ /n = 0.201	n ₃₂ /n = 0.092		
		W ₁₃ = 0.043	0.055	
		W ₂₃ = 0.28		
		n ₂₃ = 0.040		
		W ₃₃ = 0.011		
		n ₃₃ /n = 0.029		
	x̂ ₁ = 0.75	ŷ ₁ = 0.50	ŷ ₂ = 1.25	0.146
	x̂ ₂ = 2.00			

Table 6.1 shows that considerable reduction in the weighted average V_{XY} can be achieved by stratifying on the variables (X, Y) . However, by comparing the ratio of the weighted average V_{XY} over the weighted average under SRS (i.e. V_{SRS}), it can be seen that the

case $L = 3$ and $M = 2$ provides a larger gain in precision than the case $L = 2$ and $M = 3$. Hence, in general, increasing the number of intervals for one of the two variables X and Y should not be done arbitrarily. All the runs of the program OPTIBND5 to produce Table 6.1 used the same starting values for the initial vector of points of stratification $(\underline{x}^{(0)}, \underline{y}^{(0)})$ which, for the first interval (i.e. $x_1^{(0)}$ or $y_1^{(0)}$) was 0.25, and 0.50 for the second interval when $L = 3$ or $M = 3$ (i.e. $x_2^{(0)}$ or $y_2^{(0)}$).

To illustrate the mechanism underlying the program OPTIBND5, a contour plot of the value of the objective function V_{XY} is given in Figure 6.2 for the case $L = 2$ and $M = 2$. It should be noted that in this case, the only variables involved are x_1 and y_1 .

Figure 6.2

In Figure 6.2, the circled value is the global minimum of the objective function. By following the different arrows, the path followed by the program OPTIBND5 can easily be seen. Note that in this case, and in fact for all the combinations of L and M considered, the algorithm converged in a single iteration to a global minimum with the stopping criteria ϵ_X and ϵ_Y both set equal to 0.001. Unfortunately, this is not in general the case since the convergence cannot be insured.

It can be seen from Table 6.1 that the points of stratification $\hat{x}$ and $\hat{y}$ obtained by taking each variable separately turned out to be the same as the OPS x^* and y^* , respectively, due to lack of precision associated with the determination of these points of stratification. In general, $(\hat{x}, \hat{y})$ is different from (x^*, y^*) since $(\hat{x}, \hat{y})$ is determined by the one-way optimum stratification approach which involves a different objective function than the two-way stratification problem. In fact, by crossing the interval $[0, 10]$ by discrete steps of 0.01, the points of stratification $(\hat{x}, \hat{y})$ are different from (x^*, y^*) .

The second illustration of the use of the computer program OPTIBND5 is based on the bivariate distribution of the sales and assets of Incorporated Tobacconist Retailers in Canada for the year 1985. This population of $N = 185$ units is described in Figure 6.3.

Figure 6.3

Bivariate Distribution of the Sales and Assets of Incorporated Tobacconists Retailers in Canada for the Year 1985 (in Thousands of Dollars).

population size $N = 185$

	Sales	Assets
means :	521.0	124.0
variances :	213171.9	16549.1
covariance :	41701.5	
minimum :	1	1
maximum :	2875	993

symbol	FREQU
.....	0.000000000 - 10.200391901
.....	10.200391901 - 20.400783802
.....	20.400783802 - 30.601175702
.....	30.601175702 - 40.801567603
.....	40.801567603 - 51.001959504
.....	51.001959504 - 61.202351405
.....	61.202351405 - 71.402743305
.....	71.402743305 - 81.603135206
.....	81.603135206 - 91.803527107
.....	91.803527107 - 102.003919009

Provided that the total sample size n and the number LM of strata are small relative to N , the present population can be considered as infinite. It should, however, be noted that with $L = 3$ and $M = 3$, the number of strata is already equal to 9 which can invalidate the above assumption of infinite population size. Moreover, the fact that the units of the population $\{(x, y)\}$ are not uniformly spread within the domain area can cause some strata to be empty. The constraint that each stratum contains at least two population units is used to permit the computation of $S_{X_{hk}}^2$ and $S_{Y_{hk}}^2$. As in the previous illustration, the relative importances θ_X and θ_Y were both set equal to 0.5 and the Neyman allocation was used. Table 6.2 gives the results of the use of OPTIBND5 for the present distribution.

Table 6.2

Optimum Points of Stratification ($\bar{x}^*$, $\bar{y}^*$) for the Sales and Assets, respectively, of Tobacco Retailers in Canada for the Year 1985 ($N = 185$), under Neyman allocation (f.p.c. ignored). The Relative Importances are Set to $\theta_X = \theta_Y = 1/2$. The Points $\hat{\bar{x}}$ and $\hat{\bar{y}}$ are Obtained by Considering X and Y Separately. ((1) indicates non-feasible solutions.)

$M = 2$				
$L = 2$	$y_1^* = 152$			V_{XY}/V_{SRS}
	$x_1^* = 642$	$W_{11} = 0.613$	$W_{12} = 0.102$	0.234
		$n_{11}/n = 0.431$	$n_{12}/n = 0.060$	
		$W_{21} = 0.151$	$W_{22} = 0.134$	
		$n_{21}/n = 0.091$	$n_{22}/n = 0.418$	
	$\hat{x}_1 = 595$	$\hat{y}_1 = 237$		0.265
$M = 2$				
$L = 3$	$y_1^* = 141$			V_{XY}/V_{SRS}
	$x_1^* = 292$	$W_{11} = 0.366$	$W_{12} = 0.032$	0.148
		$n_{11}/n = 0.182$	$n_{12}/n = 0.024$	
	$x_2^* = 668$	$W_{21} = 0.258$	$W_{22} = 0.102$	
		$n_{21}/n = 0.156$	$n_{22}/n = 0.061$	
		$W_{31} = 0.102$	$W_{32} = 0.140$	
		$n_{31}/n = 0.048$	$n_{32}/n = 0.529$	
	$\hat{x}_1 = 466$	$\hat{y}_1 = 237$		$+\infty(1)$
	$\hat{x}_2 = 1066$			
$M = 3$				
$L = 2$	$y_1^* = 67 \quad y_2^* = 177$			V_{XY}/V_{SRS}
	$x_1^* = 642$	$W_{11} = 0.350$	$W_{12} = 0.306$	0.209
		$n_{11}/n = 0.218$	$n_{12}/n = 0.249$	
		$W_{21} = 0.038$	$W_{22} = 0.140$	
		$n_{21}/n = 0.020$	$n_{22}/n = 0.096$	
		$W_{13} = 0.059$	$W_{23} = 0.108$	
		$n_{13}/n = 0.037$	$n_{23}/n = 0.380$	
	$\hat{x}_1 = 595$	$\hat{y}_1 = 107$	$\hat{y}_2 = 277$	$+\infty(1)$
$M = 3$				
$L = 3$	$y_1^* = 50 \quad y_2^* = 141$			V_{XY}/V_{SRS}
	$x_1^* = 292$	$W_{11} = 0.161$	$W_{12} = 0.204$	0.131
		$n_{11}/n = 0.059$	$n_{12}/n = 0.097$	
	$x_2^* = 642$	$W_{21} = 0.070$	$W_{22} = 0.151$	
		$n_{21}/n = 0.018$	$n_{22}/n = 0.084$	
		$W_{31} = 0.016$	$W_{32} = 0.124$	
		$n_{31}/n = 0.001$	$n_{32}/n = 0.067$	
		$W_{13} = 0.032$	$W_{23} = 0.097$	
		$n_{13}/n = 0.025$	$n_{23}/n = 0.061$	
		$W_{33} = 0.145$	$n_{33}/n = 0.587$	
	$\hat{x}_1 = 466$	$\hat{y}_1 = 107$	$\hat{y}_2 = 277$	$+\infty(1)$
	$\hat{x}_2 = 1066$			

Table 6.2 shows that considerable gain in precision can be obtained by stratifying the present bivariate population. However, it can be seen that the case $L = 3$ and $M = 2$ and the case $L = 3$ and $M = 3$ give similar gains in precision so that the optimum number of

strata may be taken as 6 with $L = 3$ and $M = 2$. Also, as in the previous illustration, the case $L = 3$ and $M = 2$ provides a larger gain in precision than the case $L = 2$ and $M = 3$ so that, for a given total number LM of strata, the number of intervals for the variables X and Y should not be chosen arbitrarily. For the different runs of the program OPTIBND5 that produced Table 6.2, the initial points of stratification ($\bar{x}^{(0)}, \bar{y}^{(0)}$) were taken as the medians of the marginal populations for the case of two intervals (i.e. $L = 2$ or $M = 2$), and by taking the 33th and 67th percentiles of the marginal populations for the case of three intervals (i.e. $L = 3$ or $M = 3$).

As in the previous illustration, the mechanism underlying the program OPTIBND5 is illustrated by drawing a contour plot of the value of the objective function V_{XY} for the case $L = 2$ and $M = 2$. This contour plot is given in Figure 6.4.

Figure 6.4

As before, the circled value represents the global minimum of the objective function. Following the arrows which describe the path used by OPTIBND5, it can be seen that the

algorithm converges to a local minimum in a single iteration. Unfortunately, this local minimum is not the same as the global minimum. However, the value of the objective function at this local minimum is quite close to the value at the global minimum. Using the stopping criteria $\epsilon_X = 0.001$ and $\epsilon_Y = 0.001$, the algorithm converged in almost two iterations for all cases studied here.

The points of stratification $\hat{x}$ and $\hat{y}$, obtained by considering the sale and asset variables separately, differ quite substantially from the OPS x^* and y^* . The infinite sign in Table 6.2 indicates that the points of stratification $\hat{x}$ and $\hat{y}$ obtained do not satisfy the constraint of at least two population units in each stratum so that the value of the weighted average V_{XY} cannot be computed.

6.2 Finite Population Case

In this section, the finite population case is considered, where the population size N is small enough so that the f.p.c. of the weighted average V_{XY} , given by (6.10), cannot be ignored. As the previous section, the two-way lattice stratification is based on the variables of interest X and Y themselves.

In the literature, the only studies that can be found related to the finite population case are those based on the use of cluster algorithms, discussed in Section 6.1.1. Recall that these studies were performed by Golder and Yeomans (1973), Dahmström and Hagnell (1978) and Jarque (1981). Since the cluster algorithms are generally based on re-allocation of units, these algorithms are, by their nature, more convenient to be used with small populations than with large ones. However, cluster algorithms are usually not designed to handle the problem where the stratum sample size n_{hk} is found to be greater than the corresponding stratum population size N_{hk} . Moreover, the stratifications obtained using these algorithms are general ones in the sense that they are not determined under the constraint of lattice stratification. The approach of Ghosh (1963) could be used under lattice stratification. Minimal equations would then be obtained by taking the partial derivatives of the expression of the weighted average V_{XY} under a given allocation, including the f.p.c.. As noted by Schneeberger and Pollöt (1985), these minimal equations, however, are only necessary but not sufficient conditions for obtaining a global, or even a local minimum, i.e., saddle-points or local maxima can also be obtained as solutions. Using the bivariate normal distribution with means $(\mu_X, \mu_Y) = (0, 0)$ and covariance matrix $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, Schneeberger and Pollöt (1985) also showed that, for a given value of ρ , the minimal equations obtained under the Neyman allocation (6.14) can also lead to non-feasible solutions if the overall sampling fraction $q = \frac{n}{N}$ is greater than a critical value q_c . This critical value q_c depends on the underlying population for the variable (X, Y) . It is clear from the previous comments that a method is required to solve the problem of determining the OPS (x^*, y^*) minimizing the weighted average V_{XY} for the finite population case. Fortunately, as before, DP can be used which will at least provide a local minimum to the present problem.

6.2.1 The Dynamic Programming Approach: For the finite population case, a three-dimensional approach is needed if the constraint $n_{hk} \leq N_{hk}$ is not satisfied by the allocation. The first two dimensions are in fact related to the determination of the OPS for

the variables X and Y , respectively, while the third dimension refers to the allocation of the total sample size n among the different strata. As noted in Section 3.2.3, the use of another dimension for the allocation of the total sample size n permits the control of allocation to insure that the stratum sample size n_{hk} is less than or equal to the corresponding stratum population size N_{hk} . Also, by defining discrete state and decision spaces for this dimension, it is then possible to obtain integer sample sizes n_{hk} .

For the one-step allocations discussed in the beginning of this chapter, the proportional allocation given by (6.11) clearly does not require the use of the third dimension since the constraint $n_{hk} \leq N_{hk}$ is always satisfied. Therefore, the two-way optimum stratification problem can be simply handled, in this case, using the DP approach discussed in Section 6.1.3. For the equal-sample-size allocation given by (5.15), the DP approach of Section 6.1.3 can also be used because the allocated sample sizes are fixed with respect to the points of stratification $(\underline{x}, \underline{y})$. The only change required is to modify the decision spaces $B_h^{(X)}(x_{h-1})$ and $B_k^{(Y)}(y_{k-1})$ defined in (6.28) to insure stratum population sizes N_{hk} of at least $\frac{n}{ML}$ units. For the Neyman allocation (6.14), it is clear that this allocation can be realized by the use of the third dimension. This is also true for the optimum allocation (6.16) defined under the cost constraint (5.19): The only difference is that, under Neyman allocation, the stratum sample sizes n_{hk} are allocated such that $\sum_{h=1}^L \sum_{k=1}^M n_{hk} = n$ while, under optimum allocation, they are allocated in order to satisfy $\sum_{h=1}^L \sum_{k=1}^M c_{hk} n_{hk} = C - c_0$.

In considering the two-step allocations, all the ones involving only the proportional or equal-sample-size allocation for each of the two steps do not require the use of the third dimension. These two-step allocations can simply be treated using the DP approach of Section 6.1.3. On the other hand, none of the remaining two-step allocations, which use Neyman or optimum (with a cost constraint) allocation in at least one of the two steps, can be handled by the three-dimensional approach described below.

It has been seen earlier that the DP formulation of a problem in the form (2.1) first involves the identification of the objective function to be minimized. For the two-way optimum stratification problem in the finite population case, this objective function is of the form

$$(6.38) \quad G(\underline{x}, \underline{y}, \underline{n}) = \sum_{h=1}^L \sum_{k=1}^M t(x_{h-1}, x_h, y_{k-1}, y_k, n_{hk})$$

where $\underline{n} = (n_{11}, n_{12}, \dots, n_{1M}, n_{21}, \dots, n_{LM})$.

Since the problem is to determine the OPS $(\underline{x}^*, \underline{y}^*)$ such that the weighted average V_{XY} is minimized, the objective function $G(\underline{x}, \underline{y}, \underline{n})$ can simply be obtained from equation (6.10). Note that the allocation of the total sample size n is directly performed by the use of the third dimension. Using (6.10), the function $t(x_{h-1}, x_h, y_{k-1}, y_k, n_{hk})$ is given by

$$(6.39) \quad t(x_{h-1}, x_h, y_{k-1}, y_k, n_{hk}) = W_{hk} (\theta_X \sigma_{Xhk}^2 + \theta_Y \sigma_{Yhk}^2) \left(\frac{W_{hk}}{n_{hk}} - \frac{1}{N} \right).$$

The general form of the objective function given by (6.38) is similar to the form given by (3.22) for the one-way optimum stratification problem in the finite population case.

The only difference is that the latter uses only one dimension for the determination of the OPS while the former uses two dimensions. Using the general concepts of DP discussed in Chapter 2, the stages may be given by each pair (h, k) . The state of each stage (h, k) can be expressed by the three-dimensional state variable (x_h, y_k, m_{hk}) where m_{hk} is defined as the total sample size allocated up to stage (h, k) . Based on this state variable, the initial state (x_0, y_0, m_{00}) can be clearly be given by $(a_X, a_Y, 0)$. Since the stratum sample sizes n_{hk} have to be integers, the state space of each stage is found to be $[a_X, b_X] \times [a_Y, b_Y] \times \{0, 1, \dots, n\}$. The decision variable is also three-dimensional and is given by the variable (d_h, c_k, n_{hk}) where, as in Section 6.1.3, d_h represents the width of the interval $[x_{h-1}^+, x_h]$ and c_k the width of the interval $[y_{k-1}^+, y_k]$. Using these definitions, the stage transformation function can simply be given by $(x_h, y_k, m_{hk}) = (x_{h-1} + d_h, y_{k-1} + c_k, m_{h-1k-1} + n_{hk})$. Finally, to satisfy the different constraints associated with the points of stratification $(\underline{x}, \underline{y})$ and the sample sizes $\underline{n}$, the decision variable (d_h, c_k, n_{hk}) is restricted to the decision space $B_h^{(X)}(x_{h-1}) \times B_k^{(Y)}(y_{k-1}) \times B_{hk}(x_{h-1}, x_h, y_{k-1}, y_k, m_{h-1k-1})$ where $B_h^{(X)}(x_{h-1})$ and $B_k^{(Y)}(y_{k-1})$ are as in formulation (6.28) and where $B_{hk}(x_{h-1}, x_h, y_{k-1}, y_k, m_{h-1k-1}) = \{1, 2, \dots, n - m_{h-1k-1}\} \cap \{1, 2, \dots, [W_{hk}N]\}$. The problem of two-way optimum stratification in the case of a finite population can then be expressed as

(6.40)

$$\text{Min } \sum_{h=1}^L \sum_{k=1}^M t(x_{h-1}, x_h, y_{k-1}, y_k, n_{hk})$$

$$\text{subject to: } (x_h, y_k, m_{hk}) = (x_{h-1} + d_h, y_{k-1} + c_k, m_{h-1k-1} + n_{hk})$$

$$(x_h, y_k, m_{hk}) \in [a_X, b_X] \times [a_Y, b_Y] \times \{0, 1, 2, \dots, n\}$$

$$(d_h, c_k, n_{hk}) \in B_h^{(X)}(x_{h-1}) \times B_k^{(Y)}(y_{k-1}) \times B_{hk}(x_{h-1}, x_h, y_{k-1}, y_k, m_{h-1k-1})$$

$$(x_0, y_0, m_{00}) = (a_X, a_Y, 0), \quad h = 1, \dots, L; k = 1, \dots, M.$$

Unfortunately, formulation (6.40) has exactly the same problem as formulation (6.28) since the decision space $B_h^{(X)}(x_{h-1}) \times B_k^{(Y)}(y_{k-1}) \times B_{hk}(x_{h-1}, x_h, y_{k-1}, y_k, m_{h-1k-1})$ again depends on the states of both past and future stages. Recall that it directly contradicts the Bellman's Principle of Optimality stated in Section 2.2. As noted in Section 6.1.3, the problem arises from the fact that, instead of $L \times M$ stages, there are in reality only $(L + M)$ stages which can be viewed through the decision variables d_h and c_k .

To be able to express the problem given by (6.40) into one which can be handled by DP, a similar iterative approach as in Section 6.1.3 can be used. The points of stratification $\underline{y}$ are first fixed and then the OPS $\underline{x}^*$ and the optimal sample sizes $\underline{n}^*$ are computed. In the second part, the resulting OPS $\underline{x}^*$ are themselves considered as fixed and new vectors of OPS $\underline{y}^*$ and sample sizes $\underline{n}^*$ are obtained. This process is iterated until convergence is attained. Since fixing one dimension leaves to two others unfixed, this method can be named as a two-dimensional iterative DP approach. Formally, before the first iteration, the points of stratification $\underline{x}$ and $\underline{y}$ are initialized to arbitrary values $\underline{x}^{(0)}$ and $\underline{y}^{(0)}$, respectively, such that $a_X = x_0^{(0)} < x_1^{(0)} < \dots < x_{L-1}^{(0)} < x_L^{(0)} = b_X$ and $a_Y = y_0^{(0)} < y_1^{(0)} < \dots < y_{M-1}^{(0)} < y_M^{(0)} = b_Y$. For the i th iteration, the values of the points

of stratification $\underline{y}^{(i-1)}$ are considered as fixed, which in fact transforms the problem to one similar to the two-way optimum stratification problem with one categorical stratification variable. The fixed intervals $[y_{k-1}^{+(i-1)}, y_k^{(i-1)}]$ can be viewed as the categories discussed in Chapter 5. Because of this similarity, the two-part DP approach proposed in Section 5.2.1 for the finite population case can directly be used almost without any changes. Based on the same relationships with the general concepts of DP, the problem of determining the OPS $\underline{x}^*$ for the first part of the i th iteration can then be expressed as

(6.41)

$$\text{Min} \sum_{h=1}^L \sum_{k=1}^M t(x_{h-1}, x_h, y_{k-1}^{(i-1)}, y_k^{(i-1)}, n_{hk})$$

$$\text{subject to: } (x_h, m_h) = (x_{h-1} + d_h, m_{h-1} + n_h)$$

$$(x_h, m_h) \in [a_X, b_X] \times \{0, 1, \dots, n\}$$

$$(d_h, n_h) \in B_h^{(X)}(x_{h-1}) \times B_h^{(X)}(x_{h-1}, x_h, m_{h-1})$$

$$\text{where } B_h^{(X)}(x_{h-1}, x_h, m_{h-1}) = \{1, 2, \dots, n - m_{h-1}\} \cap \{1, 2, \dots, [W_{Xh}N]\}$$

$$(x_0, m_0) = (a_X, 0)$$

$$\text{and } m_{hk} = m_{hk-1} + n_{hk}$$

$$m_{hk} \in \{0, 1, \dots, n_h\}$$

$$n_{hk} \in B_{hk}(x_{h-1}, x_h, y_{k-1}^{(i-1)}, y_k^{(i-1)}, n_h, m_{hk-1}) =$$

$$\{1, 2, \dots, n_h - m_{hk-1}\} \cap \{1, 2, \dots, [W_{hk}N]\}$$

$$m_{h0} = 0, \quad h = 1, \dots, L; k = 1, \dots, M.$$

This formulation can be used to obtain the Functional Equation of Bellman required for the use of DP. Let $G_{Xh}^*(x_{h-1}, \underline{y}^{(i-1)}, m_{h-1})$ be the optimal value of the objective function of (6.41) for the strata (h, k) to (L, k) for all $k = 1, \dots, M$ given that the lower bound for the strata (h, k) for $k = 1, \dots, M$ is x_{h-1} , and given that a total of $n - m_{h-1}$ units have been allocated to the strata (h, k) to (L, k) for $k = 1, \dots, M$. As in Section 5.2.1, also let $T_k^*(x_{h-1}, x_h, \underline{y}^{(i-1)}, m_{hk-1})$ be the optimal value for the strata (h, k) to (h, M) in the interval $[x_{h-1}^+, x_h]$, given that $n_h - m_{hk-1}$ units have been allocated to these $(M - k + 1)$ strata. The Functional Equation of Bellman with respect to the first part of the i th iteration is then given by

(6.42)

$$G_{Xh}^*(x_{h-1}, \underline{y}^{(i-1)}, m_{h-1}) = \underset{d_h \in B_h^{(X)}(x_{h-1})}{\text{Min}} \underset{n_h \in B_h^{(X)}(x_{h-1}, x_h, m_{h-1})}{\text{Min}} \{T_1^*(x_{h-1}, x_h, \underline{y}^{(i-1)}, 0) + G_{Xh+1}^*(x_h, \underline{y}^{(i-1)}, m_h)\}$$

where

$$T_k^*(x_{h-1}, x_h, \underline{y}^{(i-1)}, m_{hk-1}) = \underset{n_{hk} \in B_{hk}(x_{h-1}, x_h, y_{k-1}^{(i-1)}, y_k^{(i-1)}, n_h, m_{hk-1})}{\text{Min}} \{t(x_{h-1}, x_h, y_{k-1}^{(i-1)}, y_k^{(i-1)}, n_{hk}) + T_{k+1}^*(x_{h-1}, x_h, \underline{y}^{(i-1)}, m_{hk})\}$$

Based on (6.42), new points of stratification $\underline{x}^{(i)}$ can be computed to replace the preceding values $\underline{x}^{(i-1)}$ so that the OPS for the variables (X, Y) are $(\underline{x}^{(i)}, \underline{y}^{(i-1)})$ for this first part of the i th iteration.

For the second part of the i th iteration, the values of the points of stratification $\underline{x}^{(i)}$ are considered as fixed. Since the fixed intervals $[x_{h-1}^{(i)}, x_h^{(i)}]$, in turn, can be viewed as the categories discussed in Chapter 5, the two-part DP approach discussed in Section 5.2.1 can again be applied to determine OPS with respect to the variable Y . The problem of obtaining the OPS $\underline{y}^{(i)}$ for the second part of the i th iteration can be formulated as follows:

(6.43)

$$\text{Min } \sum_{h=1}^L \sum_{k=1}^M t(x_{h-1}^{(i)}, x_h^{(i)}, y_{k-1}, y_k, n_{hk})$$

subject to: $(y_k, m_k) \triangleq (y_{k-1} + c_k, m_{k-1} + n_k)$

$$(y_k, m_k) \in [a_Y, b_Y] \times \{0, 1, \dots, n\}$$

$$(c_k, n_k) \in B_k^{(Y)}(y_{k-1}) \times B_k^{(Y)}(y_{k-1}, y_k, m_{k-1})$$

$$\text{where } B_k^{(Y)}(y_{k-1}, y_k, m_{k-1}) = \{1, 2, \dots, n - m_{k-1}\} \cap \{1, 2, \dots, [W_{Yk}N]\}$$

$$(y_0, m_0) = (a_Y, 0)$$

and $m_{hk} = m_{h-1k} + n_{hk}$

$$m_{hk} \in \{0, 1, \dots, n_k\}$$

$$n_{hk} \in B_{hk}^{(X)}(x_{h-1}^{(i)}, x_h^{(i)}, y_{k-1}, y_k, n_k, m_{h-1k}) =$$

$$\{1, 2, \dots, n_k - m_{h-1k}\} \cap \{1, 2, \dots, [W_{hk}N]\}$$

$$m_{0k} = 0, \quad k = 1, \dots, L; \quad k = 1, \dots, M$$

Based on this formulation of the problem, the Functional Equation of Bellman can be obtained in a similar manner as in the first part of the i th iteration. Let $G_{Yk}^*(\underline{x}^{(i)}, y_{k-1}, m_{k-1})$ be the optimal value of the objective function of (6.43) for the strata (h, k) to (h, M) for all $h = 1, \dots, L$ given that the lower bound for the strata (h, k) for $h = 1, \dots, L$ is y_{k-1} and given that a total of $n - m_{k-1}$ units have been allocated to the strata (h, k) to (h, M) for $h = 1, \dots, L$. As in Section 5.2.1, also let $T_h^*(\underline{x}^{(i)}, y_{k-1}, y_k, m_{h-1k})$ be the optimal value for the strata (h, k) to (L, k) in the interval $[y_{k-1}^+, y_k]$, given that $n_k - m_{h-1k}$ units have been allocated to these $(L - h + 1)$ strata. The Functional Equation of Bellman with respect to the second part of the i th iteration is then given by

(6.44)

$$G_{Yk}^*(\underline{x}^{(i)}, y_{k-1}, m_{k-1}) = \min_{c_h \in B_k^{(Y)}(y_{k-1})} \min_{n_k \in B_k^{(Y)}(y_{k-1}, y_k, m_{k-1})} \{T_1^*(\underline{x}^{(i)}, y_{k-1}, y_k, 0) + G_{Yk+1}^*(\underline{x}^{(i)}, y_k, m_k)\}$$

where

$$T_h^*(\underline{x}^{(i)}, y_{k-1}, y_k, m_{h-1k}) = \min_{n_{hk} \in B_{hk}(x_{h-1}^{(i)}, x_h^{(i)}, y_{k-1}, y_k, m_{h-1k})} \{t(x_{h-1}^{(i)}, x_h^{(i)}, y_{k-1}, y_k, n_{hk}) + T_{h+1}^*(\underline{x}^{(i)}, y_{k-1}, y_k, m_{hk})\}$$

Using (6.44), the values of the OPS $\underline{y}^{(i)}$ can be computed to replace the preceding values $\underline{y}^{(i-1)}$. Hence, the OPS for i th iteration are given by $(\underline{x}^{(i)}, \underline{y}^{(i)})$. As described earlier, this process iterates until some convergence criterion is satisfied. For this two-dimensional iterative DP approach the convergence criterion can be defined as (6.32) and (6.33), which is the one used for the unidimensional iterative DP approach of Section 6.1.3.

The iterative process underlying the two-dimensional iterative DP approach is basically the same as the one underlying the one-dimensional iterative DP approach. That is, both approaches fix the points of stratification with respect to one variable while they determine the OPS for the other variable. Because of this, the proof of convergence to the correct limit, given in Appendix F, still applies for the two-dimensional iterative DP approach. The convergence of this method, however, is slower than in the unidimensional iterative DP approach since, at each iteration, the two-part DP approach developed in Section 5.2.1 is used. Hence, this iterative process should in most cases be used only for small populations and for small numbers $L \times M$ of strata. Another disadvantage is that, unlike the DP approaches given in the previous chapters, it does not permit easy determination of the optimum number of strata. Hence, the values of L and M have to be specified in the statement of the problem.

In practice, it often occurs that the population under consideration is a real population $\{(x, y)\}$ of size N . The OPS $(\underline{x}^*, \underline{y}^*)$ can then still be obtained using the proposed two-dimensional iterative DP method but with the different quantities entering into the objective function (6.38) replaced by their finite population counterparts, as before. It should be noted that the process will stop iterating in a given number of steps if ϵ_X and ϵ_Y of (6.32) and (6.33) are defined to be less than the minimum difference between two consecutive x 's and two consecutive y 's, respectively. This is in fact true no matter how small ϵ_X and ϵ_Y are chosen.

6.3 Use of Auxiliary Variables

The stratification variables cannot in practice be the same as the variables of interest. The stratification variables, say Z and U , must therefore be chosen from highly correlated data already available. These data can come from the most recent census, for example.

It is assumed that the population of size N of the auxiliary variables Z and U has a joint CDF $F_{ZU}(z, u)$ defined on $[a_Z, b_Z] \times [a_U, b_U]$ with mean vector (μ_Z, μ_U) and finite

covariance matrix $\Sigma_{ZU} = \begin{bmatrix} \sigma_Z^2 & \sigma_{ZU} \\ \sigma_{ZU} & \sigma_U^2 \end{bmatrix}$. As usual, the notation used will be basically the same as in Sections 6.1 and 6.2 but with the use of the subscripts X, Y, Z and U . The problem considered here is to determine the OPS $(\underline{z}^*, \underline{u}^*)$ in order to form the lattice stratification which will minimize the weighted average V_{XY} given by (6.10). This, however, requires the knowledge of the nature of the relationship between the auxiliary variables (Z, U) and the variables of interest (X, Y) .

One possible model to describe the relationship between (X, Y) and (Z, U) was considered by Ghosh (1963). This model, which is quite general, is defined as follows:

$$(6.45) \quad \begin{aligned} X &= \Psi(Z, U) + e \\ Y &= \zeta(Z, U) + \eta \end{aligned}$$

Where the error terms e and η are such that the conditional means $E(e|Z, U)$ and $E(\eta|Z, U)$ are both zero. The conditional variances of e and η are also supposed to be of the forms $\text{Var}(e|Z, U) = \nu_e(Z, U)$ and $\text{Var}(\eta|Z, U) = \nu_\eta(Z, U)$. As it is usually done in multivariate models, the conditional covariance between e and η , i.e. $\text{Cov}(e, \eta|Z, U)$, is assumed to be 0. On the other hand, the covariance between $\Psi(Z, U)$ and $\zeta(Z, U)$, $\text{Cov}(\Psi(Z, U), \zeta(Z, U))$, in general, is different from 0. It should be noted that the above assumptions differ from those given by Ghosh (1963) who simply stated that $E(e) = E(\eta) = 0$ and $\text{Cov}(e, \eta) = \text{Cov}(e, \Psi(Z, U)) = \text{Cov}(e, \zeta(Z, U)) = \text{Cov}(\eta, \Psi(Z, U)) = \text{Cov}(\eta, \zeta(Z, U)) = 0$ but $\text{Cov}(\Psi(Z, U), \zeta(Z, U)) \neq 0$. It can, however, be verified that the two sets of assumptions are equivalent, so that, for consistency with the previous chapters, the former set of assumptions will be used. Since $\text{Var}(e) = E(\nu_e(Z, U))$ and $\text{Var}(\eta) = E(\nu_\eta(Z, U))$, the unconditional variances of e and η will be denoted by μ_{ν_e} and μ_{ν_η} , respectively. On the other hand, as usual, the mean and the variance of $\Psi(Z, U)$ will be denoted by μ_Ψ and σ_Ψ^2 , respectively, while those of $\zeta(Z, U)$ by μ_ζ and σ_ζ^2 . Based on these quantities the means and the variances of the variables of interest X and Y can then be expressed as

$$(6.46) \quad \mu_X = E(X) = E(\Psi(Z, U) + e) = \mu_\Psi,$$

$$(6.47) \quad \sigma_X^2 = \text{Var}(X) = \text{Var}(\Psi(Z, U) + e) = \sigma_\Psi^2 + \mu_{\nu_e}$$

$$(6.48) \quad \text{and } \mu_Y = E(Y) = E(\zeta(Z, U) + \eta) = \mu_\zeta,$$

$$(6.49) \quad \sigma_Y^2 = \text{Var}(Y) = \text{Var}(\zeta(Z, U) + \eta) = \sigma_\zeta^2 + \mu_{\nu_\eta}.$$

To determine the OPS $(\underline{z}^*, \underline{u}^*)$ such that the weighted average V_{XY} is minimized, the approach to use is to simply express (6.10) in terms of the points of stratification $\underline{z}$ and $\underline{u}$. Proceeding as in Section 3.3, the stratum proportion W_{hk} , the stratum mean μ_{Xhk} and the stratum variance σ_{Xhk}^2 of the variable X , under the model (6.45), are given by

$$(6.50) \quad W_{hk} = \int_{z_{k-1}^+}^{z_k^+} \int_{u_{k-1}^+}^{u_k^+} dF_{ZU}(z, u),$$

$$(6.51) \quad \mu_{Xhk} = \frac{1}{W_{hk}} \int_{z_{k-1}^+}^{z_k^+} \int_{u_{k-1}^+}^{u_k^+} \Psi(z, u) dF_{ZU}(z, u) = \mu_{\Psi hk},$$

$$(6.52) \quad \sigma_{Xhk}^2 = \sigma_{\Psi hk}^2 + \mu_{\nu_e hk}$$

where $\sigma_{\Psi hk}^2$ and $\mu_{\nu_e hk}$ are obtained from (6.5) and (6.6).

The stratum mean μ_{Yhk} and the stratum variance σ_{Yhk}^2 of the variable Y can similarly be expressed, under the model (6.45), as

$$(6.53) \quad \mu_{Yhk} = \mu_{\zeta hk}$$

$$(6.54) \quad \text{and} \quad \sigma_{Yhk}^2 = \sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk}.$$

Using these expressions, the weighted average V_{XY} to be minimized is given by

$$(6.55) \quad V_{XY} = \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}^2}{n_{hk}} [\theta_X(\sigma_{\Psi hk}^2 + \mu_{\nu_e hk}) + \theta_Y(\sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk})] \\ - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}}{N} [\theta_X(\sigma_{\Psi hk}^2 + \mu_{\nu_e hk}) + \theta_Y(\sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk})]$$

This formula can be used to determine the OPS ($\underline{z}^*, \underline{u}^*$) that will form an optimum lattice stratification such that the weighted average V_{XY} is minimized. According to Ghosh (1963), if in the model (6.45) the function Ψ involves the variable Z only, and the function ζ the variable U only (or vice-versa), and if both functions Ψ and ζ are assumed to be one-to-one, then an important invariance property is preserved. That is, the lattice stratification obtained from the points of stratification ($\underline{z}, \underline{u}$) is still a lattice stratification with respect to the variable (X, Y) . If one of θ_X or θ_Y is set to zero then the case of a single variable of interest and two stratification variables can directly be treated. This case often occurs in practice.

As mentioned earlier, the determination of the OPS usually depends on the allocation used. All the different one-step or two-step allocations together with their associated expression for the weighted average V_{XY} can be expressed in terms of the points of stratification ($\underline{z}, \underline{u}$) using exactly the same replacement process as for equation (6.55). For example, the Neyman allocation, given by (6.14), can be directly expressed, under the model (6.45), as

$$(6.56) \quad n_{hk} = \frac{W_{hk} [\theta_X(\sigma_{\Psi hk}^2 + \mu_{\nu_e hk}) + \theta_Y(\sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk})]^{1/2}}{\sum_{h=1}^L \sum_{k=1}^M W_{hk} [\theta_X(\sigma_{\Psi hk}^2 + \mu_{\nu_e hk}) + \theta_Y(\sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk})]^{1/2}} n, \quad h = 1, \dots, L; k = 1, \dots, M.$$

Also, under this allocation, the weighted average V_{XY} becomes

$$(6.57) \quad V_{XY} = \frac{1}{n} \left(\sum_{h=1}^L \sum_{k=1}^M W_{hk} [\theta_X(\sigma_{\Psi hk}^2 + \mu_{\nu_e hk}) + \theta_Y(\sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk})]^{1/2} \right)^2 \\ - \sum_{h=1}^L \sum_{k=1}^M \frac{W_{hk}}{N} [\theta_X(\sigma_{\Psi hk}^2 + \mu_{\nu_e hk}) + \theta_Y(\sigma_{\zeta hk}^2 + \mu_{\nu_\eta hk})].$$

6.3.1 Literature Review: Aoyama (1954) considered a single variable of interest but a lattice stratification based on several auxiliary variables. Using a linear multivariate model, he obtained minimal equations by taking the partial derivatives of the variance of $\bar{x}_{st}$, under proportional allocation. Assuming the model (6.45), Ghosh (1963) evaluated the derivatives of the generalized variance D_{XY} under proportional allocation to obtain the related minimal equations. Thomsen (1977) considered the use of the cum $\sqrt{f}$ rule of Dalenius and Hodges (1959) to determine approximate points of stratification for each stratification variable. Assuming that a single variable of interest is linearly related to two stratification variables, he also obtained the optimum number of strata based on a given cost function. Finally, Kish and Anderson (1978) studied various approaches of increasing the precision of the estimates of multiple characteristics by using a lattice stratification based on several auxiliary variables. In one approach, the use of principal component analysis for the reduction of the number stratification variables was investigated.

6.3.2 Identification of Problems: Although the minimization of the weighted average V_{XY} given by (6.55) under the model (6.45) has not been considered in the literature, some difficulties might be expected if the approaches followed by Aoyama (1954) or Ghosh (1963) are used since the minimal equations are only necessary but not sufficient conditions for a global minimum. Moreover, in the finite population case, the minimal equations could also lead to a non-feasible solution if the allocation used is one for which it is possible to obtain a stratum sample size n_{hk} greater than the corresponding stratum population size N_{hk} . The use of a method which would handle these problems is therefore needed. Fortunately, the DP approaches developed in Sections 6.1 and 6.2 can again be applied when the lattice stratification is based on the auxiliary variables Z and U .

6.3.3 The Dynamic Programming Approach: The DP methods developed in Sections 6.1 and 6.2 are used basically the same way when the stratification is based on the auxiliary variables Z and U , except for two differences. First, all the computations are now based on the quantities obtained under the CDF $F_{ZU}(z, u)$ of the population of the auxiliary variables (Z, U) instead of the population of (X, Y) . Secondly, the different objective functions obtained by considering one of the one-step or two-step allocations are now expressed in terms of the points of stratification (z, u) . Hence, in the infinite population case, the objective function takes the form

$$(6.58) \quad G(z, u) = \sum_{h=1}^L \sum_{k=1}^M g(z_{h-1}, z_h, u_{k-1}, u_k).$$

On the other hand, the objective function for the finite population case is expressed as

$$(6.59) \quad G(z, u, n) = \sum_{h=1}^L \sum_{k=1}^M t(z_{h-1}, z_h, u_{k-1}, u_k, n_{hk}).$$

Note that the DP formulations of the two-way optimum stratification problem when the auxiliary variables Z and U are used can simply be obtained from those of Sections 6.1 and 6.2 by replacing the x 's and y 's by z 's and u 's, respectively.

In the case where the population of (Z, U) is a real population $\{(z, u)\}$ of size N , the different quantities entering into the objective function are again replaced by their finite population counterparts. Using the approach of Section 3.1.4, the quantities W_{hk} , $\mu_{\Psi hk}$ and $\sigma_{\Psi hk}^2$ are replaced by

$$(6.60) \quad W_{hk} : \quad \frac{N_{hk}}{N} = \frac{1}{N} \{ \text{number of } (z, u) \text{ in } I_{ZU hk} \},$$

$$(6.61) \quad \mu_{\Psi hk} : \quad \bar{\Psi}_{hk} = \frac{1}{N_{hk}} \sum_{(z, u) \in I_{ZU hk}} \Psi(z, u),$$

$$(6.62) \quad \sigma_{\Psi hk}^2 : \quad S_{\Psi hk}^2 = \frac{1}{N_{hk}-1} \sum_{(z, u) \in I_{ZU hk}} (\Psi(z, u) - \bar{\Psi}_{hk})^2$$

where $I_{ZU hk} = [z_{h-1}^+, z_h] \times [u_{k-1}^+, u_k]$. Similarly, the quantities $\mu_{\zeta hk}$ and $\sigma_{\zeta hk}^2$ are replaced by $\bar{\zeta}_{hk}$ and $S_{\zeta hk}^2$, respectively, which are defined as above replacing Ψ by ζ . Once the replacement is done, the unidimensional iterative DP approach or the two-dimensional iterative DP approach can directly be used to obtain the OPS (z^*, u^*) minimizing the weighted average V_{XY} .

Finally, it should be noted that this development is based upon the assumption that the nature of the relationship between (X, Y) and (Z, U) (expressed by the model (6.45)) is completely known. However, the forms of the functions $\Psi(Z, U)$ and $\zeta(Z, U)$ and the conditional variances $\nu_e(Z, U)$ and $\nu_\eta(Z, U)$ may, unfortunately, be difficult to determine in practice.

REFERENCES

- Adatia, A., Chan, L.K. (1981), "Relations between stratified, grouped and selected order statistics samples", *Scand. Actuar. J.*, p. 193-202.
- Aggarwal, V., Singh, R. (1984) "On optimum stratification with proportional allocation for a class of Pareto distributions", *Commun. Statist. - Theor. Meth.*, Vol. 13, No. 24, p. 3107-3116.
- Anderson, D.W., Kish, L., Cornell, R.G. (1976), "Quantifying gains from stratification for optimum and approximately optimum strata using a bivariate normal model", *J.A.S.A.*, Vol. 71, No. 356, p. 887-892.
- (1980), "On stratification, grouping and matching", *Scand. J. Statist.*, Vol. 7, p. 61-66.
- Aoyama, H. (1954), "A study of the stratified random sampling", *Ann. Inst. Statist. Math.*, Vol. 6, p. 1-36.
- Bellman, R.E. (1957), "Dynamic programming", *Princeton 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. (1961), "Comparison of methods for determining stratum boundaries", *Bull. Intern. Statist. Inst.*, Vol. 38, part 2, p. 345 - 358.
- (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.
- (1952), "The problem of optimum stratification in a special type of design", *Skand. Aktuarietidskr.*, Vol. 35, p. 61-70.
- (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.
- Dalenius, T., Hodges, J.L. (1957), "The choice of stratification points", *Skand. Aktuarietidskr.*, Vol. 40, p. 198-203.

- (1959), "Minimum variance stratification", J.A.S.A., Vol. 54, p. 88-101.
- Durbin, J. (1959), "Review of "Sampling in Sweden"", Journal Roy. Statist. Soc. A, Vol. 122, p. 246-248.
- Ekman, G. (1959), "An approximation useful in univariate stratification", Ann. Math. Statist., Vol. 30, p. 219-229.
- Espejo, M.R. (1985), "Equiprecisional allocation and optimum stratification", Statistics, Vol. 16, No. 4, p. 559-562.
- Eubank, R.L. (1982), "A quantile domain perspective on the relationships between optimal grouping, spacing and stratification problems", Statist. & Prob. Letters, Vol. 1, p. 69-73.
- Fisher, W.D. (1958), "On grouping for maximum homogeneity", J.A.S.A., Vol. 53, p. 789-798.
- Ghosh, S.P. (1963), "Optimum stratification with two characters", Ann. Math. Statist., Vol. 34, p. 866-872.
- Glasser, G.J. (1962), "On the complete coverage of large units in a statistical study", Rev. Intern. Statist. Inst., Vol. 30, No. 1, p. 28-32.
- Golder, P.A., Yeomans, K.A. (1973), "The use of cluster analysis for stratification", Appl. Statist., Vol. 22, p. 213-219.
- 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.
- Goodman, L.A. (1979), "Simple models for the analysis of association in cross-classification having ordered categories", J.A.S.A., Vol. 74, p. 537-552.
- Hadley, G. (1964), "Nonlinear and dynamic programming", Addison-Wesley Publishing Company Inc., Massachusets, 484 pages.
- Hansen, M.H., Hurwitz, W.N., Madow, W.G. (1953), "Sample survey, methods and theory", Vol. I and II, John Wiley & Sons, New York, 332 pages.
- Hayashi, C., Maruyama, F. (1948), "On some methods for stratification", Research Memoir, Inst. Statist. Math., Vol. 4, p. 399-411.
- Hayashi, C., Maruyama, F., Ishida, M. (1951), "On some criteria for stratification", Ann. Inst. Statist. Math., Vol. 2, p. 77-86.
- Hidiroglou, M.A. (1986), "The construction of a self-representing stratum of large units in survey design", Amer. Statist., Vol. 40, No. 1, p. 27-31.
- Isii, K. (1966), "Mathematical programming and statistical inference", Sûgaku (Mathematics), Vol. 18, p. 85-95.

Isii, K., Taga, Y. (1969), "On optimal stratification for multivariate distributions", *Skand. Aktuarietidskr.*, p. 24-38.

Jarque, C.M. (1981), "A solution to the problem of optimum stratification in multivariate sampling", *Appl. Statist.*, Vol. 30, No. 2, p. 163-169.

Kish, L., Anderson, D.W. (1978), "Multivariate and multipurpose stratification", *J.A.S.A.*, Vol. 73, No. 361, p. 24-34.

Kitagawa, T. (1956), "Some contributions to the design of sample surveys", *Sankhyā*, Vol. 17, part 1, p. 1-36.

Kpedekpo, G.M.K. (1973), "Recent advances on some aspects of stratified sample designs. A review of the literature", *Metrika*, Vol. 20, No. 1, p. 54-64.

Lavallée, P. (1985), "Stratification based on power allocation", unpublished paper.

Lavallée, P., Hidirolou, M.A. (1986), "On the stratification of skewed populations", paper presented at the 1987 Amer. Stat. Assoc. Annual Meeting.

Mahalanobis, P.C. (1952), "Some aspects of the design of sample surveys", *Sankhyā*, Vol. 12, parts 1 and 2, p. 1-7.

Marsden, J.E., (1974), "Elementary classical analysis", W.H. Freeman and Company, San Francisco, 549 pages.

Marsden, J.E., Tromba, A.J. (1976), "Vector calculus, 2nd edition", W.H. Freeman and Company, San Francisco, 591 pages.

McCormick, G.P. (1983), "Nonlinear programming", John Wiley & Sons, New York, 444 pages.

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

Neyman, J. (1934), "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection", *Jour. Roy. Statist. Soc.*, Vol. 97, p. 558-606.

Oslo, I.T. (1976), "A comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation", *Metrika*, Vol. 23, p. 15-25.

Ross, A. (1961), "Variance estimates in "optimum" sample designs", *J.A.S.A.*, Vol. 63, p. 135-142.

Sadasivan, G., Aggarwal, R. (1978), "Optimum points of stratification in bivariate populations", *Sankhyā*, Vol. 40, Ser. C, part 1, p. 84-97.

Scheffé, H. (1959), "The analysis of variance", John Wiley & Sons, New York, 477 pages.

Schneeberger, H. (1970), "Optimierung in der stichprobentheorie durch schichtung", Statist. Hefte, Vol. 11, p. 242-253.

—— (1971), "Optimierung in der stichprobentheorie durch schichtung und aufteilung", Unternehmensforschung, Vol. 15, No. 5, p. 240-253.

—— (1979), "Saddle-points of the variance of the sample mean in stratified sampling", Sankhyā, Vol. 41, Ser. C, part 1, p. 92-96.

—— (1981), "The problem of optimum stratification and allocation with $q = n/N > 0$ ", Metrika, Vol. 28, p. 179-189.

—— (1985), "Maxima, minima und saddlepunkte bei optimaler schichtung und optimaler aufteilung", Allgemeines Statist. Archiv., Vol. 69, p. 286-297.

Schneeberger, H., Goller, W. (1979), "On the problem of the feasibility of optimal stratification points according to Dalenius", Statist. Hefte, Vol. 20, No. 4, p. 250-256.

Schneeberger, H., Pollöt, J.P. (1985), "Optimum stratification with two variates", Statist. Hefte, Vol. 26, No. 2, p. 97-113.

Serfling, R.J. (1968), "Approximately optimal stratification", J.A.S.A., Vol. 63, p. 1298-1309.

Sethi, V.K. (1963), "A note on optimum stratification of populations for estimating the population means", Austral. Jour. Statist., Vol. 5., p. 20-33.

Shanno, D.F. (1970), "Parameter selection of modified Newton methods for function minimization", Siam Jour. Numer. Anal., Vol. 7.

Singh, R. (1971), "Approximately optimum stratification on the auxiliary variable", J.A.S.A., Vol. 66, No. 336, p. 829-833.

—— (1975), "A note on optimum stratification in sampling with varying probabilities", Austral. Jour. Statist., Vol. 17, No. 1, p. 12-21.

—— (1977), "A note on optimum stratification for equal allocation with ratio and regression methods of estimation", Austral. Jour. Statist., Vol. 19, No. 2, p. 96-104.

Singh, R., Parkash, D. (1975), "Optimum stratification for equal allocation", Ann. Inst. Statist. Math., Vol. 27, p. 273-280.

Singh, R., Sukhatme, B.V. (1969), "Optimum stratification", Ann. Inst. Statist. Math., Vol. 21, p. 515-528.

—— (1972), "Optimum stratification in sampling with varying probabilities", *Ann. Inst. Statist. Math.*, Vol. 24., p. 485-494.

—— (1973), "Optimum stratification with ratio and regression methods of estimation", *Ann. Inst. Statist. Math.*, Vol. 25, p. 627-633.

Taga, Y. (1967), "On optimum stratification for the objective variable based on concomitant variables using prior information", *Ann. Inst. Statist. Math.*, Vol. 19, p. 101-130.

Thomsen, I. (1977), "On the effect of stratification when two stratifying variables are used", *J.A.S.A.*, Vol. 72, No. 357, p. 149-153.

Unnithan, G. (1978), "The minimum variance boundary points of stratification", *Sankhyā*, Vol. 40, Ser. C, part 1, p. 60-72.

Wang, M.C., Aggarwal, V. (1984), "Stratification under a particular Pareto distribution", *Commun. Statist.-Theor. Meth.*, Vol. 13, No. 6, p. 711-735.

White, D.J., (1969), "Dynamic programming", Holden-Day, San Francisco, 181 pages.

Wienhold, P. (1982), "Der ASOS17-algorithmus ein verfahren zur approximativen simultanen optimierung der schichtenbildung in lagerkollektiven", *allgemeines Statist. Archiv.*, Vol. 66, p. 274-288.

Yadava, S., Singh, R. (1984), "Optimum stratification for allocation proportional to strata totals for simple random sampling scheme", *Commun. Statist.-Theor. Meth.*, Vol. 13, No. 22, p. 2793-2806.

APPENDIX A

The OPTIBND1 Program


```
program optibndl(CON,dfile,output);
```

```
( This program computes optimum bounds of stratification for infinite )
( populations in the case of an hypothetical distribution           )
( or in the case of a large real population.                      )
( The default allocation is Neyman allocation but proportional    )
( or equal allocations can also be used by changing the code at the )
( indicated places. )
```

[illegible]

```
var dfile: filetype;
    fildata: glnarray;
    popul: poparray;
    bound, stmean, stvar, sts2, stw: bndtype;
    stnsample: intarray;
    stratind: array[0..1] of integer;
    ndata, npopul, nsamp, h, i, j, totnsamp: integer;
    sampfrac, next, minvar, variance, weight, totw, totmean, totvar,
    filler1, filler2, filler3, filler4: real;
```

```
procedure sortshll(var x:glnarray;n:integer);
```

```

        temp:=-x[i];
        x[i]:=-x[j];
        x[j]:=temp;
        alldone:=false;
    end
end
until alldone
end
end;

```

```

(+++++)

```

```

función min(a,b:real):real;
begin

```

```

    if a<b then min:=-a
    else min:=-b;

```

```

end;

```

```

function max(a,b:real):real;

```

```

begin

```

```

    if a>b then max:=-a
    else max:=-b;

```

```

end;

```

```

function p3(a:real):real;

```

```

begin

```

```

    p3:=-a*a*a;

```

```

end;

```

```

function p4(a:real):real;

```

```

begin

```

```

    p4:=-a*a*a*a;

```

```

end;

```

```

(*****

```

```

function w(x:poparray;n,low,high:integer):real;

```

```

begin

```

```

    w:=(high-low+1)/n;

```

```

end;

```

```

(+++++)

```

```

function m0(x:poparray;n,low,high:integer):real;

```

```

( for hypot. distr. )

```

```

var tot,lval,hval:real;

```

```

begin

```

```

    tot:=0.0;

```

```

    lval:=max(-1.0,x[low]);

```

```

    hval:=min(-1/7,x[high]);

```

```

    if lval<hval then tot:=-tot+49/72*((sqr(hval)-sqr(lval))/2+hval-lval);

```

```

    lval:=max(-1/7,x[low]);

```

```

    hval:=min(0.0,x[high]);

```

```

    if lval<hval then

```

```

        tot:=-tot+1/3*(49/2*(sqr(hval)-sqr(lval))+35/4*(hval-lval));

```

```

    lval:=max(0.0,x[low]);

```

```

    hval:=min(1/7,x[high]);

```

```

    if lval<hval then

```

```

        tot:=-tot+1/3*(49/2*(sqr(lval)-sqr(hval))+35/4*(hval-lval));

```

```

    lval:=max(1/7,x[low]);

```

```

    hval:=min(1.0,x[high]);

```

```

    if lval<hval then tot:=-tot+49/72*((sqr(lval)-sqr(hval))/2+hval-lval);

```

```

    m0:=-tot;
end;

```

```

(*****)

```

```

function mu(x:poparray;n,low,high:integer):real;

```

```

var i:integer;
    totx:real;

```

```

begin

```

```

    totx:=0.0;

```

```

    for i:=low to high do

```

```

        totx:=totx+x[i];

```

```

    mu:=-totx/(high-low+1);

```

```

end;

```

```

(+++++)

```

```

function m1(x:poparray;n,low,high:integer):real;

```

```

{ for hypot. distr. }

```

```

var tot,lval,hval:real;

```

```

begin

```

```

    tot:=0.0;

```

```

    lval:=max(-1.0,x[low]);

```

```

    hval:=min(-1/7,x[high]);

```

```

    if lval<hval then tot:=-tot+49/72*((p3(hval)-p3(lval))/3
                                         +(sqr(hval)-sqr(lval))/2);

```

```

    lval:=max(-1/7,x[low]);

```

```

    hval:=min(0.0,x[high]);

```

```

    if lval<hval then tot:=-tot+1/3*(49/3*(p3(hval)-p3(lval))
                                         +35/4*(sqr(hval)-sqr(lval))/2);

```

```

    lval:=max(0.0,x[low]);

```

```

    hval:=min(1/7,x[high]);

```

```

    if lval<hval then tot:=-tot+1/3*(49/3*(p3(lval)-p3(hval))
                                         +35/4*(sqr(hval)-sqr(lval))/2);

```

```

    lval:=max(1/7,x[low]);

```

```

    hval:=min(1.0,x[high]);

```

```

    if lval<hval then tot:=-tot+49/72*((p3(lval)-p3(hval))/3
                                         +(sqr(hval)-sqr(lval))/2);

```

```

    m1:=-tot;

```

```

end;

```

```

(*****)

```

```

function sigma2(x:poparray;n,low,high:integer):real;

```

```

var i:integer;

```

```

    totx,totx2:real;

```

```

begin

```

```

    totx:=0.0;

```

```

    totx2:=0.0;

```

```

    if low=high then sigma2:=0.0

```

```

    else begin

```

```

        for i:=low to high do begin

```

```

            totx:=totx+x[i];

```

```

            totx2:=totx2+sqr(x[i])

```

```

        end;

```

```

        sigma2:=-(totx2-sqr(totx)/(high-low+1))/(high-low)

```

```

end;
end;

```

```

(+++++ )

```

```

function m2(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var tot,lval,hval:real;
begin
  tot:=0.0;
  lval:=max(-1.0,x[low]);
  hval:=min(-1/7,x[high]);
  if lval<hval then tot:=tot+49/72*((p4(hval)-p4(lval))/4
                                     +(p3(hval)-p3(lval))/3);
  lval:=max(-1/7,x[low]);
  hval:=min(0.0,x[high]);
  if lval<hval then tot:=tot+1/3*(49/4*(p4(hval)-p4(lval))
                                   +35/4*(p3(hval)-p3(lval))/3);
  lval:=max(0.0,x[low]);
  hval:=min(1/7,x[high]);
  if lval<hval then tot:=tot+1/3*(49/4*(p4(lval)-p4(hval))
                                   +35/4*(p3(hval)-p3(lval))/3);
  lval:=max(1/7,x[low]);
  hval:=min(1.0,x[high]);
  if lval<hval then tot:=tot+49/72*((p4(lval)-p4(hval))/4
                                     +(p3(hval)-p3(lval))/3);
  m2:=tot;
end;

```

```

(+++++ )

```

```

function wtrig(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
begin
  wtrig:=m0(x,n,low,high);
end;
function mutrig(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
begin
  if low=high then mutrig:=0.0
  else mutrig:=m1(x,n,low,high)/m0(x,n,low,high);
end;
function s2trig(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var wh:real;
begin
  if low=high then s2trig:=0.0
  else begin
    wh:=m0(x,n,low,high);
    s2trig:=m2(x,n,low,high)/wh-sqr(m1(x,n,low,high)/wh);
  end;
end;

```

```

(***** )

```

```

procedure printall(x:poparray;n,lstrat:integer;b:bndtype);
begin
  writeln(CON);writeln(CON);
  writeln(CON,' :2, '*** DYNAMIC PROGRAMING SIMULATION ***');

```

```

writeln(CON);
writeln(CON, ' ':2, 'data from file ', dosfile);
writeln(CON);
writeln(CON, ' ':2, '*** univariate statistics ***');
writeln(CON);
writeln(CON, ' ':2, 'number of observations: ', n:1); -
writeln(CON, ' ':2, 'population mean: ', mu(x,n,1,n):8:3);
variance:=sigma2(x,n,1,n);
writeln(CON, ' ':2, 'population variance (S2): ', variance:8:3);
writeln(CON, ' ':2, 'sample size: ', nsample:1);
variance:=variance/nsample;
writeln(CON, ' ':2, 'variance for unstratified estimate (Vsrs): ', variance:8);
writeln(CON);
writeln(CON, ' ':2, '*** optimal stratification ',
              '(with Neyman allocation) ***');
writeln(CON);
writeln(CON, ' ':2, 'number of strata: ', lstrat:1);
writeln(CON, ' ':2, 'optimal bounds of stratification: ');
write(CON, ' ':2);
for h:=0 to lstrat do
  write(CON, ' b(' , h:1, ')');
writeln(CON);
write(CON, ' ':2);
for h:=0 to lstrat do
  write(CON, ' ', b[h]:8:2);
writeln(CON);writeln(CON);

stratind[0]:=-0; { -1; for hypot. distr. }
h:=-1;
for i:=1 to n do begin
  if x[i]>b[h] then begin
    stratind[h]:=-i-1;
    h:=h+1;
  end;
end;
stratind[lstrat]:=-n;

weight:=0.0;
for h:=1 to lstrat do begin
  i:=stratind[h-1]+1; { +0; for hypot. distr. }
  j:=stratind[h];
  stw[h]:=-w(x,n,i,j);
  stmean[h]:=-mu(x,n,i,j);
  sts2[h]:=-sigma2(x,n,i,j);
  {weight is given for Neyman allocation}
  weight:=weight+stw[h]*sqrt(sts2[h]);
end;
totw:=0.0;
totmean:=0.0;
totnsamp:=0;
totvar:=0.0;
for h:=1 to lstrat do begin
  totw:=totw+stw[h];
  totmean:=totmean+stw[h]*stmean[h];
  { Neyman allocation is assumed }
  stnsample[h]:=-round(nsample*stw[h]*sqrt(sts2[h])/weight);
  totnsamp:=totnsamp+stnsample[h];
  stvar[h]:=-sts2[h]/stnsample[h];
  totvar:=totvar+sqr(stw[h])*stvar[h];
end;

```

```

write(CON,' ':2, 'stratum number: ');
for h:=1 to lstrat do
  write(CON,' ':8,h:2);
writeln(CON,' ':7,'all');
write(CON,' ':2, '% stratum proportion: ');
for h:=1 to lstrat do
  write(CON,' ':2,100*stw[h]:8:2);
writeln(CON,' ':2,100*totw:8:2);
write(CON,' ':2, 'stratum mean: ');
for h:=1 to lstrat do
  write(CON,' ':2,stmean[h]:8:3);
writeln(CON,' ':2,totmean:8:3);
write(CON,' ':2, 'stratum variance (S2h): ');
for h:=1 to lstrat do
  write(CON,' ':2,sts2[h]:8:3);
writeln(CON,' * ');
write(CON,' ':2, 'stratum sample size: ');
for h:=1 to lstrat do
  write(CON,' ':2,stnsample[h]:8);
writeln(CON,' ':2,totnsamp:8);
write(CON,' ':2, 'stratum variance of estimate: ');
for h:=1 to lstrat do
  write(CON,' ':2,stvar[h]:8);
writeln(CON,' ':2,totvar:8);
writeln(CON);
variance:=100*totvar/variance;
writeln(CON,' ':2,'% variance ratio (Vstr/Vsrs): ',variance:8:2,'%');
writeln(CON);writeln(CON);
writeln(CON,'*****',
  '*****');

```

end;

(*****)

procedure optibnd(z:poparray;n,lstrat:integer;var b:bnctype;var fstar0:real);

```

var fstar:array[1..1,1..dsize] of real;
nextbnd:array[0..1,1..dsize] of integer;
compflag:array[1..1,1..dsize] of boolean;
preslbnd,substrat:integer;

```

(.....)

function objf(low,high:integer):real;

begin

```

( if low<1 then low:=-low-1; ,for hypot. distr. )
if low>high then objf:=maxreal
else begin
  ( objf for Neyman allocation )
  objf:=-w(z,n,low,high)*sqrt(sigma2(z,n,low,high));
end;

```

end;

(.....)

function cumfstar(stage,lbound:integer):real;

var f,minval:real;

```

    nextbnd,hbound,lowb:integer;

begin
    if compflag[stage,lbound] then minval:=-fstar[stage,lbound]
    else begin
        if stage=lstrat then begin
            minval:=-objf(lbound,n);
            nextbnd:=-n+1;
        end
        else begin
            minval:=-maxreal;
            if lbound=1 then lowb:=0
            else lowb:=-lbound;
            for hbound:=-lowb to n-lstrat+stage do begin
                f:=-objf(lbound,hbound)+cumfstar(stage+1,hbound+1);
                if f<minval then begin
                    minval:=f;
                    nextbnd:=-hbound+1;
                end;
            end;
            end;
            fstar[stage,lbound]:=-minval;
            nextbnd[stage,lbound]:=-nextbnd;
            compflag[stage,lbound]:=-true;
            WRITELN('STAGE: ',STAGE:1,' LBND: ',LBOUND:1, ' CUMF*: ',MINVAL:8:4);
            if lbound=1 then begin
                substrat:=-lstrat-stage+1;
                b[0]:=-z[1];
                preslbnd:=1;
                j:=-1;
                for h:=stage to lstrat do begin
                    preslbnd:=-nextbnd[h,preslbnd];
                    b[j]:=-z[preslbnd-1];
                    j:=-j+1;
                end;
                printall(z,n,substrat,b);
            end;
            end;
            cumfstar:=-minval;
        end;

    { ..... }

begin (procedure optibnd)
    for h:=-1 to lstrat do
        for j:=-1 to n do
            compflag[h,j]:=-false;

    fstar0:=-cumfstar(1,1);
end;

{*****}

begin (main program)
    assign(dfile,dosfile);
    reset(dfile);
    i:=-1;
    readln(dfile,filler1,filler2,filler3,filedata[i],filler4);
    while not eof(dfile) do begin
        i:=-i+1;

```

```

--      readln(dfile,filler1,filler2,filler3,fildata[i],filler4);
end;
ndata:=-1;
close(dfile);

sortshll(fildata,ndata);
if ndata > dsize then begin
    npopul:=-dsize;
    sampfrac:=-dsize/ndata;
end
else begin
    npopul:=-ndata;
    sampfrac:=-1.0;
end;
next:=-1.0;
for j:-1 to npopul-1 do begin
    popul[j]:=-fildata[round(next)];
    next:=-next+1/sampfrac;
end;
popul[npopul]:=-fildata[ndata];
FOR I:-1 TO NPOPUL DO WRITELN(I,' ',POPUL[I],6:3);

    optibnd(popul,npopul,1,bound,minvar);

end.

```


APPENDIX B

The OPTIBND2 Program


```

        for i:=1 to n-jump do begin
            j:=i+jump;
            if x[i]>x[j] then begin
                temp:=x[i];
                x[i]:=x[j];
                x[j]:=temp;
                alldone:=false;
            end
        end
    until alldone
end
end;

```

(+++++)

```

function min(a,b:integer):integer;
begin
    if a<b then min:=-a
    else min:=-b;
end;
function rmin(a,b:real):real;
begin
    if a<b then rmin:=-a
    else rmin:=-b;
end;
function max(a,b:real):real;
begin
    if a>b then max:=-a
    else max:=-b;
end;
function p3(a:real):real;
begin
    p3:=-a*a*a;
end;
function p4(a:real):real;
begin
    p4:=-a*a*a*a;
end;

```

(*****)

```

function w(x:poparray;n,low,high:integer):real;
begin
    w:=(high-low+1)/n;
end;

```

(+++++)

```

function wexpon(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var lambda:real;
begin
    lambda:=2.0;
    wexpon:=-exp(-lambda*x[low])-exp(-lambda*x[high]);
end;

```

(*****)

```

function mu(x:poparray;n,low,high:integer):real;

```

```

var i:integer;
    totx:real;

begin
    totx:=0.0;
    for i:=low to high do
        totx:=totx+x[i];
    mu:=-totx/(high-low+1);
end;

(+++++)

function muexpon(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var lambda:real;
begin
    lambda:=2.0;
    muexpon:=(exp(-lambda*x[low]))*(x[low]+1/lambda)-
        exp(-lambda*x[high]))*(x[high]+1/lambda))/wexpon(x,n,low,high);
end;

(*****)

function sigma2(x:poparray;n,low,high:integer):real;

var i:integer;
    totx,totx2:real;

begin
    totx:=0.0;
    totx2:=0.0;
    if low=high then sigma2:=0.0
    else begin
        for i:=low to high do begin
            totx:=totx+x[i];
            totx2:=totx2+sqr(x[i])
        end;
        sigma2:=(totx2-sqr(totx)/(high-low+1))/(high-low)
    end;
end;

(+++++)

function s2expon(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var lambda:real;
begin
    lambda:=2.0;
    s2expon:=(exp(-lambda*x[low]))*(sqr(x[low])+2*x[low]/lambda+2/sqr(lambda))-
        exp(-lambda*x[high]))*(sqr(x[high])+2*x[high]/lambda+2/sqr(lambda)))/
        wexpon(x,n,low,high) - sqr(muexpon(x,n,low,high));
end;

(*****)

procedure printall(x:poparray;n,lstrat:integer;b:bndtype;stnsample:intarray);

begin
    writeln(CON);writeln(CON);

```

```

writeln(CON, ' ':2, '*** DYNAMIC PROGRAMMING SIMULATION ***');
writeln(CON);
writeln(CON, ' ':2, 'data from file ', dosfile);
writeln(CON);
writeln(CON, ' ':2, '*** univariate statistics ***');
writeln(CON);
writeln(CON, ' ':2, 'population size: ', ndata:1);
writeln(CON, ' ':2, 'population mean: ', mu(x,n,1,n):8:3);
variance:=-sigma2(x,n,1,n);
writeln(CON, ' ':2, 'population variance (S2): ', variance:8:3);
writeln(CON, ' ':2, 'sample size: ', nsample:1);
variance:=(ndata-nsample)*variance/(ndata*nsample);
writeln(CON, ' ':2, 'variance for unstratified estimate (Vsrs): ', variance, 8);
writeln(CON);
writeln(CON, ' ':2, '*** optimal stratification ',
      '(with Neyman allocation) ***');
writeln(CON);
writeln(CON, ' ':2, 'number of strata: ', lstrat:1);
writeln(CON, ' ':2, 'optimal bounds of stratification: ');
write(CON, ' ':2);
for h:=0 to lstrat do
  write(CON, ' b(' , h:1, ')');
writeln(CON);
write(CON, ' ':2);
for h:=0 to lstrat do
  write(CON, ' ', b[h]:8:2);
writeln(CON);writeln(CON);

stratind[0]:=-0; ( -1 for hypot. distr. )
h:=-1;
for i:-1 to n do begin
  if x[i]>b[h] then begin
    stratind[h]:=-i-1;
    h:=-h+1;
  end;
end;
stratind[lstrat]:=-n;

weight:=0.0;
for h:-1 to lstrat do begin
  i:=-stratind[h-1]+1; ( +0 for hypot. distr. )
  j:=-stratind[h];
  stnpopul[h]:=-round(ndata*w(x,n,i,j));
  stmean[h]:=-mu(x,n,i,j);
  sts2[h]:=-sigma2(x,n,i,j);
end;
totnpopul:=0;
totmean:=0.0;
totnsamp:=0;
totvar:=0.0;
for h:-1 to lstrat do begin
  totnpopul:=totnpopul+stnpopul[h];
  totmean:=totmean+stnpopul[h]*stmean[h]/ndata;
  totnsamp:=totnsamp+stnsample[h];
  if stnpopul[h]=0 then stvar[h]:=-0.0
  else stvar[h]:=(stnpopul[h]*stnsample[h])*sts2[h]/
    (stnpopul[h]*stnsample[h]);
  totvar:=totvar+sqr(stnpopul[h]/ndata)*stvar[h];
end;

```

```

write(CON, ' :2, 'stratum number: ');
for h:=1 to lstrat do
  write(CON, ' :8,h:2);
writeln(CON, ' :7,'all');
write(CON, ' :2, 'stratum population size: ');
for h:=1 to lstrat do
  write(CON, ' :2,stnpopul[h]:8);
writeln(CON, ' :2,totnpopul:8);
write(CON, ' :2, 'stratum mean: ');
for h:=1 to lstrat do
  write(CON, ' :2,stmean[h]:8:3);
writeln(CON, ' :2,totmean:8:3);
write(CON, ' :2, 'stratum variance (S2h): ');
for h:=1 to lstrat do
  write(CON, ' :2,sts2[h]:8:3);
writeln(CON, ' * ');
write(CON, ' :2, 'stratum sample size: ');
for h:=1 to lstrat do
  write(CON, ' :2,stnsample[h]:8);
writeln(CON, ' :2,totnsamp:8);
write(CON, ' :2, 'stratum variance of estimate: ');
for h:=1 to lstrat do
  write(CON, ' :2,stvar[h]:8);
writeln(CON, ' :2,totvar:8);
writeln(CON);
variance:=-100*totvar/variance;
writeln(CON, ' :2,'% variance ratio (Vstr/Vsrs): ',variance:8:2,'%');
writeln(CON);writeln(CON);
writeln(CON, '*****');
  '*****');
end;

{*****}

procedure optibnd(z:poparray;n,popsize,nsamp,lstrat:integer;
  var b:bndtype;var ns:intarray;var fstar0:real);

var fstar:array[1..1,1..dsize,minsamp..nsample] of real;
  nextbnd,nextsamp:array[0..1,1..dsize,minsamp..nsample] of integer;
  compflag:array[1..1,1..dsize,minsamp..nsample] of boolean;
  preslbnd,pressamp,substrat:integer;

  (.....)

function objf(low,high,sampsize:integer):real;

var wh:real;

begin
  ( if low<1 then low:=-low-1; for hypot. distr. )
  wh:=w(z,n,low,high);
  if (low>high)or(sampsize<minsamp)or(sampsize>round(popsize*wh)) then
    objf:=maxreal
  else objf:=-wh*sigma2(z,n,low,high)*(round(popsize*wh)-sampsize)/
    (sampsize*popsize);
end;

  (.....)

function cumfstar(stage,lbound,sampsize:integer):real;

```

```

var f,minval:real;
    availsmp,nxtbnd,hbound,lowb,nxtsamp,stpop,stsamp:integer;

begin
  if compflag[stage,lbound,sampsize] then minval:=-fstar[stage,lbound,sampsize]
  else begin
    stpop:=round(popsizew(z,n,lbound-0,n)); ( -1, for hypot. distr. )
    if stage=lstrat then begin
      stsamp:=min(sampsize,stpop);
      minval:=objf(lbound,n,stsamp);
      nxtbnd:=n+1;
      nextsamp:=sampsize-stsamp;
    end
    else begin
      minval:=maxreal;
      if lbound=1 then lowb:=0
      else lowb:=lbound;
      availsmp:=sampsize-minsamp*(lstrat-stage);
      for hbound:=lowb to (n-minsamp*(lstrat-stage)) do
        for stsamp:=0 to min(availsmp,stpop) do begin
          f:=objf(lbound,hbound,stsamp)
            +cumfstar(stage+1,hbound+1,(sampsize-stsamp));
          if f<minval then begin
            minval:=f;
            nxtbnd:=hbound+1;
            nextsamp:=sampsize-stsamp;
          end;
        end;
      end;
    end;
    fstar[stage,lbound,sampsize]:=-minval;
    nextbnd[stage,lbound,sampsize]:=-nxtbnd;
    nextsmp[stage,lbound,sampsize]:=-nextsmp;
    compflag[stage,lbound,sampsize]:=true;
    WRITELN('STAGE: ',STAGE:1,' LBND: ',LBOUND:1,
      ' SAMPSIZE: ',SAMPSIZE:1,' CUMF*: ',MINVAL:8:4,
      ' NXTBND: ',NXTBND:1,' NXTSAMP: ',NXTSAMP:1);
    if (lbound=1)and(sampsize=nsamp) then begin
      substrat:=lstrat-stage+1;
      b[0]:=z[1];
      preslbnd:=1;
      pressamp:=nsamp;
      k:=1;
      for h:=stage to lstrat do begin
        i:=nextbnd[h,preslbnd,pressamp];
        j:=nextsmp[h,preslbnd,pressamp];
        b[k]:=-z[i-1];
        ns[k]:=-pressamp-j;
        preslbnd:=i;
        pressamp:=j;
        k:=k+1;
      end;
      printall(z,n,substrat,b,ns);
    end;
  end;
  cumfstar:=-minval;
end;
(.....)

```

```

begin (procedure optibnd)
  for h:-1 to lstrat do
    for i:-1 to n do
      for j:-minsamp to nsamp do
        compflag[h,i,j]:=false;

      fstar0:=cumfstar(1,1,nsamp);
    end;

  (*****)

begin (main program)
  assign(dfile,dosfile);
  reset(dfile);
  i:-1;
  readln(dfile,filler1,filler2,filler3,fildata[1],filler4);
  while not eof(dfile) do begin
    i:=i+1;
    readln(dfile,filler1,filler2,filler3,fildata[i],filler4);
  end;
  ndata:=i;
  close(dfile);

  sortshll(fildata,ndata);

  if ndata > dsize then begin
    npopul:=dsize;
    sampfrac:=dsize/ndata;
  end
  else begin
    npopul:=ndata;
    sampfrac:=1.0;
  end;
  next:=-1.0;
  for j:-1 to npopul-1 do begin
    popul[j]:=fildata[round(next)];
    next:=-next+1/sampfrac;
  end;
  popul[npopul]:=fildata[ndata];
  FOR I:-1 TO NPOPUL DO WRITELN(I,' ',POPUL[I]:8);

  optibnd(popul,npopul,ndata,nsample,1,bound,stnsample,minvar);

end.

```


APPENDIX C

The OPTIBND3 Program

```
begin
  jump:=-n;
  while jump>1 do begin
    jump:=-jump div 2;
    repeat
      alldone:=-true;
```

```

        for i:=1 to n-jump do begin
            j:=i+jump;
            if x[i]>x[j] then begin
                temp:=x[i];
                x[i]:=x[j];
                x[j]:=temp;
                alldone:=false;
            end
        end
    until alldone
end
end;

(*****)

function w(x:poparray;n,low,high:integer):real;
begin
    w:=(high-low+1)/n;
end;

(*****)

function mu(x:poparray;n,low,high:integer):real;
var i:integer;
    totx:real;
begin
    totx:=0.0;
    for i:=low to high do
        totx:=totx+x[i];
    mu:=-totx/(high-low+1);
end;

(*****)

function sigma2(x:poparray;n,low,high:integer):real;
var i:integer;
    totx,totx2:real;
begin
    totx:=0.0;
    totx2:=0.0;
    if low=high then sigma2:=0.0
    else begin
        for i:=low to high do begin
            totx:=totx+x[i];
            totx2:=totx2+sqr(x[i])
        end;
        sigma2:=(totx2-sqr(totx)/(high-low+1))/(high-low)
    end;
end;

(*****)

function gamma(x:poparray;n,low,high:integer):real;
var i:integer;

```

```

totx4,muh:real;

begin
  muh:=-mu(x,n,low,high);
  totx4:=0.0;
  if low=high then gamma:=0.0
  else begin
    for i:=low to high do begin
      totx4:=-totx4+sqr(sqr(x[i]-muh));
    end;
    gamma:=-totx4/(high-low);
  end;
end;

(+++++)

function min(a,b:real):real;
begin
  if a<b then min:=-a
  else min:=-b;
end;
function max(a,b:real):real;
begin
  if a>b then max:=-a
  else max:=-b;
end;
function p3(a:real):real;
begin
  p3:=-a*a*a;
end;
function p4(a:real):real;
begin
  p4:=-a*a*a*a;
end;
function p5(a:real):real;
begin
  p5:=-a*a*a*a*a;
end;
function p6(a:real):real;
begin
  p6:=-a*a*a*a*a*a;
end;

(+++++)

function m0(x:poparray;n,low,high:integer):real;
{ for hypot. distr. }
var tot,lval,hval:real;
begin
  tot:=0.0;
  lval:=-max(-1.0,x[low]);
  hval:=-min(-1/7,x[high]);
  if lval<hval then tot:=-tot+49/72*((sqr(hval)-sqr(lval))/2+hval-lval);
  lval:=-max(-1/7,x[low]);
  hval:=-min(0.0,x[high]);
  if lval<hval then
    tot:=-tot+1/3*(49/2*(sqr(hval)-sqr(lval))+35/4*(hval-lval));
  lval:=-max(0.0,x[low]);
  hval:=-min(1/7,x[high]);
  if lval<hval then

```

```

    tot:=-tot+1/3*(49/2*(sqr(lval)-sqr(hval))+35/4*(hval-lval));
    lval:=-max(1/7,x[low]);
    hval:=-min(1.0,x[high]);
    if lval<hval then tot:=-tot+49/72*((sqr(lval)-sqr(hval))/2+hval-lval);
    m0:=-tot;
end;

```

(+++++)

```

function m1(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var tot,lval,hval:real;
begin
    tot:=-0.0;
    lval:=-max(-1.0,x[low]);
    hval:=-min(-1/7,x[high]);
    if lval<hval then tot:=-tot+49/72*((p3(hval)-p3(lval))/3
                                         +(sqr(hval)-sqr(lval))/2);

    lval:=-max(-1/7,x[low]);
    hval:=-min(0.0,x[high]);
    if lval<hval then tot:=-tot+1/3*(49/3*(p3(hval)-p3(lval))
                                         +35/4*(sqr(hval)-sqr(lval))/2);

    lval:=-max(0.0,x[low]);
    hval:=-min(1/7,x[high]);
    if lval<hval then tot:=-tot+1/3*(49/3*(p3(lval)-p3(hval))
                                         +35/4*(sqr(hval)-sqr(lval))/2);

    lval:=-max(1/7,x[low]);
    hval:=-min(1.0,x[high]);
    if lval<hval then tot:=-tot+49/72*((p3(lval)-p3(hval))/3
                                         +(sqr(hval)-sqr(lval))/2);

    m1:=-tot;
end;

```

(+++++)

```

function m2(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var tot,lval,hval:real;
begin
    tot:=-0.0;
    lval:=-max(-1.0,x[low]);
    hval:=-min(-1/7,x[high]);
    if lval<hval then tot:=-tot+49/72*((p4(hval)-p4(lval))/4
                                         +(p3(hval)-p3(lval))/3);

    lval:=-max(-1/7,x[low]);
    hval:=-min(0.0,x[high]);
    if lval<hval then tot:=-tot+1/3*(49/4*(p4(hval)-p4(lval))
                                         +35/4*(p3(hval)-p3(lval))/3);

    lval:=-max(0.0,x[low]);
    hval:=-min(1/7,x[high]);
    if lval<hval then tot:=-tot+1/3*(49/4*(p4(lval)-p4(hval))
                                         +35/4*(p3(hval)-p3(lval))/3);

    lval:=-max(1/7,x[low]);
    hval:=-min(1.0,x[high]);
    if lval<hval then tot:=-tot+49/72*((p4(lval)-p4(hval))/4
                                         +(p3(hval)-p3(lval))/3);

    m2:=-tot;
end;

```

(+++++)

```
function m3(x:poparray;n,low,high:integer):real;
```

```
( for hypot. distr. )
```

```
var tot,lval,hval:real;
```

```
begin
```

```
  tot:=0.0;
```

```
  lval:=max(-1.0,x[low]);
```

```
  hval:=min(-1/7,x[high]);
```

```
  if lval<hval then tot:=-tot+49/72*((p5(hval)-p5(lval))/5  
                                         +(p4(hval)-p4(lval))/4);
```

```
  lval:=max(-1/7,x[low]);
```

```
  hval:=min(0.0,x[high]);
```

```
  if lval<hval then tot:=-tot+1/3*(49/5*(p5(hval)-p5(lval))  
                                         +35/4*(p4(hval)-p4(lval))/4);
```

```
  lval:=max(0.0,x[low]);
```

```
  hval:=min(1/7,x[high]);
```

```
  if lval<hval then tot:=-tot+1/3*(49/5*(p5(lval)-p5(hval))  
                                         +35/4*(p4(hval)-p4(lval))/4);
```

```
  lval:=max(1/7,x[low]);
```

```
  hval:=min(1.0,x[high]);
```

```
  if lval<hval then tot:=-tot+49/72*((p5(lval)-p5(hval))/5  
                                         +(p4(hval)-p4(lval))/4);
```

```
  m3:=-tot;
```

```
end;
```

```
(+++++)
```

```
function m4(x:poparray;n,low,high:integer):real;
```

```
( for hypot. distr. )
```

```
var tot,lval,hval:real;
```

```
begin
```

```
  tot:=0.0;
```

```
  lval:=max(-1.0,x[low]);
```

```
  hval:=min(-1/7,x[high]);
```

```
  if lval<hval then tot:=-tot+49/72*((p6(hval)-p6(lval))/6  
                                         +(p5(hval)-p5(lval))/5);
```

```
  lval:=max(-1/7,x[low]);
```

```
  hval:=min(0.0,x[high]);
```

```
  if lval<hval then tot:=-tot+1/3*(49/6*(p6(hval)-p6(lval))  
                                         +35/4*(p5(hval)-p5(lval))/5);
```

```
  lval:=max(0.0,x[low]);
```

```
  hval:=min(1/7,x[high]);
```

```
  if lval<hval then tot:=-tot+1/3*(49/6*(p6(lval)-p6(hval))  
                                         +35/4*(p5(hval)-p5(lval))/5);
```

```
  lval:=max(1/7,x[low]);
```

```
  hval:=min(1.0,x[high]);
```

```
  if lval<hval then tot:=-tot+49/72*((p6(lval)-p6(hval))/6  
                                         +(p5(hval)-p5(lval))/5);
```

```
  m4:=-tot;
```

```
end;
```

```
(+++++)
```

```
function wtrig(x:poparray;n,low,high:integer):real;
```

```
( for hypot. distr. )
```

```
begin
```

```
  wtrig:=m0(x,n,low,high);
```

```
end;
```

```
function mutrig(x:poparray;n,low,high:integer):real;
```

```
( for hypot. distr. )
```

```

begin
  if low-high then mutrig:-0.0
  else mutrig:-m1(x,n,low,high)/m0(x,n,low,high);
end;
function s2trig(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var wh:real;
begin
  if low-high then s2trig:-0.0
  else begin
    wh:-m0(x,n,low,high);
    s2trig:-m2(x,n,low,high)/wh-sqr(m1(x,n,low,high)/wh);
  end;
end;
function gamtrig(x:poparray;n,low,high:integer):real;
( for hypot. distr. )
var m0h,mlh,muh:real;
begin
  if low-high then gamtrig:-0.0
  else begin
    m0h:-m0(x,n,low,high);
    mlh:-m1(x,n,low,high);
    muh:-mlh/m0h;
    gamtrig:-((m4(x,n,low,high)-4.0*muh*m3(x,n,low,high)+
      6.0*sqr(muh)*m2(x,n,low,high))/m0h-3.0*p4(muh));
  end;
end;

(*****)

procedure printall(x:poparray;n,lstrat:integer;b:bndtype);

begin
  writeln(LST);writeln(LST);
  writeln(LST,' ':2,'*** DYNAMIC PROGRAMMING SIMULATION ***');
  writeln(LST);
  writeln(LST,' ':2,'data from file ',dosfile);
  writeln(LST);
  writeln(LST,' ':2,'*** univariate statistics ***');
  writeln(LST);
  writeln(LST,' ':2,'number of observations: ',n:1);
  writeln(LST,' ':2,'population mean: ',mu(x,n,1,n):8:3);
  variance:-sigma2(x,n,1,n);
  writeln(LST,' ':2,'population variance (S2): ',variance:8:3);
  writeln(LST,' ':2,'sample size: ',nsample:1);
  variance:-gamma(x,n,1,n)/sqr(variance);
  writeln(LST,' ':2,'population kurtosis: ',(variance-3.0):8:4);
  writeln(LST);
  writeln(LST,' ':2,'*** optimal stratification (with Ross allocation) ***');
  writeln(LST,' ':2,'--- optimality w.r.t. variance estimation ---');
  writeln(LST);
  writeln(LST,' ':2,'number of strata: ',lstrat:1);
  writeln(LST,' ':2,'optimal bounds of stratification: ');
  write(LST,' ':2);
  for h:-0 to lstrat do
    write(LST,' b(' ,h:1,')');
  writeln(LST);
  write(LST,' ':2);
  for h:-0 to lstrat do
    write(LST,' ',b[h]:8:2);

```

```

writeln(LST);writeln(LST);

stratind[0]:=-0; ( -1; for hypot. distr. )
h:-1;
for i:-1 to n do begin
  if x[i]>b[h] then begin
    stratind[h]:=-i-1;
    h:-h+1;
  end;
end;
stratind[lstrat]:=-n;

weight:=0.0;
for h:-1 to lstrat do begin
  i:-stratind[h-1]+1; ( +0; for hypot. distr. )
  j:-stratind[h];
  stw[h]:=-w(x,n,i,j);
  stmean[h]:=-mu(x,n,i,j);
  sts2[h]:=-sigma2(x,n,i,j);
  kappa[h]:=-gamma(x,n,i,j)/sqr(sts2[h]);
  (weight is given for Ross allocation )
  weight:=-weight+stw[h]*sqrt(sts2[h]*sqrt(kappa[h]-1.0));
end;
totw:=0.0;
totmean:=0.0;
totnsamp:=0;
totvare:=0.0;
totvarv:=0.0;
for h:-1 to lstrat do begin
  totw:=totw+stw[h];
  totmean:=totmean+stw[h]*stmean[h];
  ( Ross allocation is assumed )
  stnsample[h]:=-round(nsample*stw[h]*
    sqrt(sts2[h]*sqrt(kappa[h]-1.0))/weight);
  totnsamp:=totnsamp+stnsample[h];
  if stw[h]≠0.0 then begin
    stvare[h]:=-0.0;
    stvarv[h]:=-0.0;
    stcv[h]:=-0.0;
  end
  else begin
    stvare[h]:=-sts2[h]/(stnsample[h]);
    stvarv[h]:=-sqr(sts2[h])/p3(stnsample[h])*
      (kappa[h]-(stnsample[h]-3.0)/(stnsample[h]-1.0));
    stcv[h]:=-100*sqr(stvarv[h])/stvare[h];
  end;
  totvare:=totvare+sqr(stw[h])*stvare[h];
  totvarv:=totvarv+p4(stw[h])*stvarv[h];
end;

write(LST,' ':2, 'stratum number: ');
for h:-1 to lstrat do
  write(LST,' ':8,h:2);
writeln(LST,' ':7,'all');
write(LST,' ':2, '% stratum proportion: ');
for h:-1 to lstrat do
  write(LST,' ':2,100*stw[h]:8:2);
writeln(LST,' ':2,100*totw:8:2);
write(LST,' ':2, 'stratum mean: ');
for h:-1 to lstrat do

```



```

        write(LST,' ':2,stmean[h]:8:3);
        writeln(LST,' ':2,totmean:8:3);
        write(LST,' ':2,'stratum variance (S2h): ');
        for h=-1 to lstrat do
            write(LST,' ':2,sts2[h]:8:3);
            writeln(LST,' * ');
            write(LST,' ':2,'stratum kurtosis (Kh-3): ');
            for h=-1 to lstrat do
                write(LST,' ':2,(kappa[h]-3.0):8:3);
            writeln(LST,' * ');
            write(LST,' ':2,'stratum sample size: ');
            for h=-1 to lstrat do
                write(LST,' ':2,stnsample[h]:8);
            writeln(LST,' ':2,totnsamp:8);
            write(LST,' ':2,'var of estimate of mean (Vm): ');
            for h=-1 to lstrat do
                write(LST,' ':2,stvare[h]:8);
            writeln(LST,' ':2,totvare:8);
            write(LST,' ':2,'var of estimate of var (Vv): ');
            for h=-1 to lstrat do
                write(LST,' ':2,stvarv[h]:8);
            writeln(LST,' ':2,totvarv:8);
            write(LST,' ':2,'% cv of estimate of variance ');
            for h=-1 to lstrat do
                write(LST,' ':2,stcv[h]:8:2);
            if totvare<0.0 then variance:-100*sqrt(totvarv)/totvare
            else variance:-0.0;
            writeln(LST,' ':2,variance:8:2);
            writeln(LST);writeln(LST);
            writeln(LST,'*****',
                '*****');
end;

(*****

procedure optibnd(z:poparray;n,lstrat:integer;var b:bndtype;var fstar0:real),
var fstar:array[1..1,1..dsize] of real;
    nextbnd:array[0..1,1..dsize] of integer;
    compflag:array[1..1,1..dsize] of boolean;
    preslbnd,substrat:integer;

(.....)

function objf(low,high:integer):real;
var wh,sigma2h,gammah:real;

begin
    ( if low<1 then low:-low-1; for hypot. distr. )
    if low>high then objf:-maxreal
    else if low=high then objf:-0.0
    else begin
        wh:-w(z,n,low,high);
        sigma2h:-sigma2(z,n,low,high);
        gammah:-gamma(z,n,low,high);
        if (gammah<sqrt(sigma2h))or(sigma2h=0.0) then objf:-maxreal
        (objf for Ross allocation)
        else objf:-wh*sqrt(sigma2h*sqrt(gammah/sqrt(sigma2h)-1.0));
    end;
end;

```

```

end;

(.....)

function cumfstar(stage,lbound:integer):real;

var f,minval:real;
    nextbnd,hbound,lowb:integer;

begin
  if compflag[stage,lbound] then minval:=-fstar[stage,lbound]
  else begin
    if stage=lstrat then begin
      minval:=objf(lbound,n);
      nextbnd:=n+1;
    end
    else begin
      minval:=maxreal;
      if lbound=1 then lowb:=0
      else lowb:=-lbound;
      for hbound:=lowb to n-lstrat+stage do begin
        f:=objf(lbound,hbound)+cumfstar(stage+1,hbound+1);
        if f<minval then begin
          minval:=f;
          nextbnd:=hbound+1;
        end;
      end;
      end;
      fstar[stage,lbound]:=-minval;
      nextbnd[stage,lbound]:=-nextbnd;
      compflag[stage,lbound]:=true;
      WRITELN('STAGE: ',STAGE:1, ' LBND: ',LBOUND:1, ' CUMF*: ',MINVAL:8:4);
      if lbound=1 then begin
        substrat:=lstrat-stage+1;
        b[0]:=z[1];
        preslbnd:=1;
        j:=1;
        for h:=stage to lstrat do begin
          preslbnd:=nextbnd[h,preslbnd];
          b[j]:=z[preslbnd-1];
          j:=j+1;
        end;
        printall(z,n,substrat,b);
      end;
      end;
      cumfstar:=-minval;
    end;
  end;

(.....)

begin (procedure optibnd)
  for h:=1 to lstrat do
    for j:=1 to n do
      compflag[h,j]:=false;

  fstar0:=cumfstar(1,1);
end;

(******)

```

```

begin (main program)
  assign(dfile,dosfile);
  reset(dfile);
  i:-1;
  readln(dfile,filler1,filler2,filler3,fildata[i],filler4);
  while not eof(dfile) do begin
    i:-i+1;
    readln(dfile,filler1,filler2,filler3,fildata[i],filler4);
  end;
  ndata:-i;
  close(dfile);

  sortshll(fildata,ndata);
  if ndata > dsize then begin
    npopul:=-dsize;
    sampfrac:=-dsize/ndata;
  end
  else begin
    npopul:=-ndata;
    sampfrac:=-1.0;
  end;
  next:-1.0;
  for j:-1 to npopul-1 do begin
    popul[j]:=-fildata[round(next)];
    next:=-next+1/sampfrac;
  end;
  popul[npopul]:=-fildata[ndata];
  FOR I:-1 TO NPOPUL DO WRITELN(I,' ',POPUL[I]:6:3);

  optibnd(popul,npopul,1,bound,minvar);

end.

```

APPENDIX D

The OPTIBND4 Program

{*****}

```
procedure sortshll(var x:glnarray;n:integer);
```

```
var jump:integer;  
    alldone:boolean;  
    temp:poprec;
```

```
begin  
    jump:=n;  
    while jump>1 do begin  
        jump:=jump div 2;  
        repeat  
            alldone:=true;  
            for i:=1 to n-jump do begin  
                j:=i+jump;  
                if x[i].pop>x[j].pop then begin  
                    temp:=x[i];  
                    x[i]:=x[j];  
                    x[j]:=temp;  
                    alldone:=false;  
                end  
            end  
        until alldone  
    end  
end;
```

```
(+++++)
```

```
function min(a,b:real):real;  
begin
```

```
    if a<b then min:=-a  
    else min:=-b;
```

```
end;
```

```
function max(a,b:real):real;
```

```
begin
```

```
    if a>b then max:=-a  
    else max:=-b;
```

```
end;
```

```
function p3(a:real):real;
```

```
begin
```

```
    p3:=-a*a*a;
```

```
end;
```

```
function p4(a:real):real;
```

```
begin
```

```
    p4:=-a*a*a*a;
```

```
end;
```

```
(*****)
```

```
procedure computw(x:poparray;n,mcateg,low,high:integer;var res:rcatarray);
```

```
( This procedure computes the stratum subpopulation proportion W(k)h )
```

```
var k,i:integer;
```

```
begin
```

```
    for k:=1 to mcateg do  
        res[k]:=0.0;
```

```
    for i:=low to high do
```

```
        res[x[i].categ]:=res[x[i].categ]+1.0;
```

```
    for k:=1 to mcateg do
```

```

      res[k]:=-res[k]/(n*teta[k]);
end;

```

```

(+++++)

```

```

procedure wexpon(x:poparray;n,mcateg,low,high:integer;var res:rcatarray);
( for hypot. distr. )
( this procedure computes the stratum subpopulation proportion W(k)h )
var lambda,lval,hval:real;
begin
  lval:=x[low].pop;
  hval:=x[high].pop;
  for k:=1 to mcateg do begin
    lambda:=k;
    res[k]:=(exp(-lambda*lval)-exp(-lambda*hval));
  end;
end;

```

```

(*****)

```

```

procedure computmu(x:poparray;n,mcateg,low,high:integer;var res:rcatarray);
var k,i,j:integer;
    count:icarray;
begin
  for k:=1 to mcateg do begin
    res[k]:=0.0;
    count[k]:=0;
  end;
  for i:=low to high do begin
    j:=x[i].categ;
    res[j]:=res[j]+x[i].pop;
    count[j]:=count[j]+1;
  end;
  for k:=1 to mcateg do
    if count[k]=0 then res[k]:=0.0
    else res[k]:=res[k]/count[k];
end;

```

```

(+++++)

```

```

procedure muexpon(x:poparray;n,mcateg,low,high:integer;var res:rcatarray);
( for hypot. distr. )
var lambda,lval,hval:real;
    w:rcatarray;
begin
  lval:=x[low].pop;
  hval:=x[high].pop;
  if lval=hval then
    for k:=1 to mcateg do
      res[k]:=0.0
  else begin
    wexpon(x,n,mcateg,low,high,w);
    for k:=1 to mcateg do begin
      lambda:=k;
      res[k]:=(exp(-lambda*lval)*(lval+1/lambda)-
        exp(-lambda*hval)*(hval+1/lambda))/w[k];
    end;
  end;
end;

```

end;

(*****)

procedure computs2(x:poparray;n,mcateg,low,high:integer;var res:rcatarray);

var k,i,j:integer;

res2:rcatarray;

count:icatarray;

begin

for k:-1 to mcateg do begin

res[k]:=-0.0;

res2[k]:=-0.0;

count[k]:=-0;

end;

for i:-low to high do begin

j:=x[i].categ;

res[j]:=-res[j]+x[i].pop;

res2[j]:=-res2[j]+sqr(x[i].pop);

count[j]:=-count[j]+1;

end;

for k:-1 to mcateg do

if count[k]<=-1 then res[k]:=-0.0

else res[k]:=(res2[k]-sqr(res[k])/count[k])/(count[k]-1);

end;

(*****)

procedure s2expon(x:poparray;n,mcateg,low,high:integer;var res:rcatarray);

(for hypot. distr.)

var lambda,lval,hval:real;

w,mu:rcatarray;

begin

lval:=x[low].pop;

hval:=x[high].pop;

if lval=hval then

for k:-1 to mcateg do

res[k]:=-0.0

else begin

wexpon(x,n,mcateg,low,high,w);

muexpon(x,n,mcateg,low,high,mu);

for k:-1 to mcateg do begin

lambda:=-k;

res[k]:=(exp(-lambda*lval)*(sqr(lval)+2*lval/lambda+2/sqr(lambda))-

exp(-lambda*hval)*(sqr(hval)+2*hval/lambda+2/sqr(lambda)))/w[k]

-sqr(mu[k]));

end;

end;

end;

(*****)

procedure printall(x:poparray;n,mcateg,lstrat:integer;b:bndtype);

begin

stratind[0]:=-0; (-1; for hypot. distr.)

h:-1;

for i:-1 to n do begin

if x[i].pop>b[h] then begin


```

        stratind[h]:=-i-1;
        h:=-h+1;
    end;
end;
stratind[lstrat]:=-n;

weight:=0.0;
for h:-1 to lstrat do begin
    i:=-stratind[h-1]+1; ( +0; for hypot. distr. )
    j:=-stratind[h];
    computw(x,n,mcateg,i,j,totw);
    computmu(x,n,mcateg,i,j,totmean);
    computs2(x,n,mcateg,i,j,totvar);
    for k:-1 to mcateg do begin
        ( stw[k,h] represents W(k)h )
        stw[k,h]:=-totw[k];
        stmean[k,h]:=-totmean[k];
        sts2[k,h]:=-totvar[k];
        ( weight is given for Neyman allocation )
        weight:=-weight+teta[k]*stw[k,h]*sqrt(sts2[k,h]);
    end;
end;
allw:=0.0;
allmean:=0.0;
allnsamp:=0;
allvar:=0.0;
allvarsrs:=0.0;
computs2(x,n,mcateg,1,n,tots2);
for k:-1 to mcateg do begin
    totw[k]:=0.0;
    totmean[k]:=0.0;
    totnsamp[k]:=0;
    totvar[k]:=0.0;
    for h:-1 to lstrat do begin
        totw[k]:=-totw[k]+stw[k,h];
        totmean[k]:=-totmean[k]+stw[k,h]*stmean[k,h];
        ( Neyman allocation is assumed )
        stnsample[k,h]:=-round(nsamp*teta[k]*stw[k,h]*
                                sqrt(sts2[k,h])/weight);
        totnsamp[k]:=-totnsamp[k]+stnsample[k,h];
        if stnsample[k,h]=0 then stvar[k,h]:=0.0
        else stvar[k,h]:=-sts2[k,h]/stnsample[k,h];
        totvar[k]:=-totvar[k]+sqr(stw[k,h])*stvar[k,h];
    end;
    allw:=-allw+teta[k]*totw[k];
    allmean:=-allmean+teta[k]*totmean[k];
    allnsamp:=-allnsamp+totnsamp[k];
    allvar:=-allvar+sqr(teta[k])*totvar[k];
    totvarsrs[k]:=-tots2[k]/totnsamp[k];
    allvarsrs:=-allvarsrs+sqr(teta[k])*totvarsrs[k];
end;

writeln(CON);writeln(CON);
writeln(CON,' ':2,'*** DYNAMIC PROGRAMMING SIMULATION ***');
writeln(CON);
writeln(CON,' ':2,'data from file ',dosfile);
writeln(CON);
for k:-1 to mcateg do begin
    writeln(CON,' ':2,'*** category ',k:1,' univariate statistics ***');
    writeln(CON);
end;

```

```

        writeln(CON, ' ':2, 'number of observations: ', round(teta[k]*n):1);
        writeln(CON, ' ':2, 'subpopulation mean: ', totmean[k]:8:3);
        writeln(CON, ' ':2, 'subpopulation variance (S2k): ', tots2[k]:8:3);
        writeln(CON, ' ':2, 'sample size: ', totnsamp[k]:1);
        writeln(CON, ' ':2, 'variance for unstratified estimate (Vsrsk): ',
            totvarsrs[k]:8);
        writeln(CON);
    end;
    writeln(CON, ' ':2, '*** optimal stratification ',
        '(with Neyman allocation) ***');
    writeln(CON);
    writeln(CON, ' ':2, 'number of intervals: ', lstrat:1);
    writeln(CON, ' ':2, 'optimal bounds of stratification: ');
    write(CON, ' ':2);
    for h:=0 to lstrat do
        write(CON, ' b(', h:1, ', '));
    writeln(CON);
    write(CON, ' ':2);
    for h:=0 to lstrat do
        write(CON, ' ', bound[h]:8:2);
    writeln(CON); writeln(CON);

    write(CON, ' ':2, 'interval number: ');
    for h:=1 to lstrat do
        write(CON, ' ':8, h:2);
    writeln(CON, ' ':7, 'all');
    writeln(CON, ' ':2, '% stratum proportion: ');
    for k:=1 to mcateg do begin
        write(CON, ' ':2, ' categ ', k:2, ' ');
        for h:=1 to lstrat do
            write(CON, ' ':2, 100*teta[k]*stw[k,h]:8:2);
            writeln(CON, ' ':2, 100*teta[k]*totw[k]:8:2);
        end;
        writeln(CON, ' ':2, ' ... overall proportion: ', 100*allw:8:2);
        writeln(CON, ' ':2, 'stratum mean: ');
        for k:=1 to mcateg do begin
            write(CON, ' ':2, ' categ ', k:2, ' ');
            for h:=1 to lstrat do
                write(CON, ' ':2, stmean[k,h]:8:3);
                writeln(CON, ' ':2, totmean[k]:8:3);
            end;
            writeln(CON, ' ':2, ' ... overall mean: ', allmean:8:3);
            writeln(CON, ' ':2, 'stratum variance (S2kh): ');
            for k:=1 to mcateg do begin
                write(CON, ' ':2, ' categ ', k:2, ' ');
                for h:=1 to lstrat do
                    write(CON, ' ':2, sts2[k,h]:8:3);
                    writeln(CON, ' * ');
                end;
                writeln(CON, ' ':2, ' ... * ');
            end;
            writeln(CON, ' ':2, 'stratum sample size: ');
            for k:=1 to mcateg do begin
                write(CON, ' ':2, ' categ ', k:2, ' ');
                for h:=1 to lstrat do
                    write(CON, ' ':2, stnsample[k,h]:8);
                    writeln(CON, ' ':2, totnsamp[k]:8);
                end;
                writeln(CON, ' ':2, ' ... overall sample size: ', allnsamp:1);
                writeln(CON, ' ':2, 'stratum variance of estimate: ');
                for k:=1 to mcateg do begin

```

```

        write(CON, ' ':2, ' categ ', k:2, ' ');
        for h:-1 to lstrat do
            write(CON, ' ':2, stvar[k,h]:8);
            writeln(CON, ' ':2, totvar[k]:8);
        end;
        writeln(CON, ' ':2, ' ... overall variance of estimate (Vstr):', allvar:8);
        writeln(CON);
        allvar:=-100*allvar/allvarsrs;
        writeln(CON, ' ':2, '% variance ratio (Vstr/Vsrs): ', allvar:8:2, '%');
        writeln(CON); writeln(CON);
        writeln(CON, '*****',
            '*****');
end;

(*****)

procedure optibnd(z:poparray;n,mcateg,lstrat:integer;var b:bnctype;
    var fstar0:real);

var fstar:array[1..1,1..dsize] of real;
    nextbnd:array[0..1,1..dsize] of integer;
    compflag:array[1..1,1..dsize] of boolean;
    preslbnd,substrat:integer;

(.....)

function objf(low,high:integer):real;

var w,sigma2:rcatarray;
    tot:real;

begin
    ( if low<1 then low:=-low-1; for hypot. distr. )
    if low>high then objf:=-maxreal
    else begin
        ( objf for Neyman allocation )
        computw(z,n,mcateg,low,high,w);
        computs2(z,n,mcateg,low,high,sigma2);
        tot:=0.0;
        for k:-1 to mcateg do
            tot:=-tot+teta[k]*w[k]*sqrt(sigma2[k]);
        objf:=-tot;
    end;
end;

(.....)

function cumfstar(stage,lbound:integer):real;

var f,minval:real;
    nextbnd,hbound,lowb:integer;

begin
    if compflag[stage,lbound] then minval:=-fstar[stage,lbound]
    else begin
        if stage=lstrat then begin
            minval:=-objf(lbound,n);
            nextbnd:=-n+1;
        end
        else begin

```

```

minval:=-maxreal;
if lbound=-1 then lowb:=-0
else lowb:=-lbound;
for hbound:=-lowb to n-lstrat+stage do begin
  f:=-objf(lbound,hbound)+cumfstar(stage+1,hbound+1);
  if f<minval then begin
    minval:=-f;
    nextbnd:=-hbound+1;
  end;
end;
end;
fstar[stage,lbound]:=-minval;
nextbnd[stage,lbound]:=-nextbnd;
compflag[stage,lbound]:=-true;
WRITELN('STAGE: ',STAGE:1,' LBND: ',LBOUND:1,' CUMF*: ',MINVAL:8:4);
if lbound=-1 then begin
  substrat:=-lstrat-stage+1;
  b[0]:=-z[1].pop;
  preslbnd:=-1;
  j:=-1;
  for h:=-stage to lstrat do begin
    preslbnd:=-nextbnd[h,preslbnd];
    b[j]:=-z[preslbnd-1].pop;
    j:=-j+1;
  end;
  printall(z,n,mcateg,substrat,b);
end;
end;
cumfstar:=-minval;
end;

(.....)

begin (procedure optibnd)
  for h:-1 to lstrat do
    for j:-1 to n do
      compflag[h,j]:=-false;
  fstar0:=-cumfstar(1,1);
end;

(*****)

begin (main program)
  assign(dfile,dosfile);
  reset(dfile);
  i:-1;
  readln(dfile,filler1,fildata[i].categ,filler2,fildata[i].pop,filler3);
  while not eof(dfile) do begin
    i:-i+1;
    readln(dfile,filler1,fildata[i].categ,filler2,fildata[i].pop,filler3);
  end;
  ndata:-i;
  close(dfile);

  sortshll(fildata,ndata);
  if ndata > dsize then begin
    npopul:=-dsize;
    sampfrac:=-dsize/ndata;
  end
end

```

```

else begin
    npopul:=-ndata;
    sampfrac:=1.0;
end;
next:=-1.0;
for j:-1 to npopul-1 do begin
    popul[j]:=-fildata[round(next)];
    next:=-next+1/sampfrac;
end;
popul[npopul]:=-fildata[ndata];
FOR I:-1 TO NPOPUL DO
    WRITELN(I:8,' ',POPUL[I].CATEG:4,' ',POPUL[I].POP:8:3);
( determination of teta[1],...,teta[m] )
for k:-1 to m do
    totpop[k]:=0;
for i:-1 to npopul do
    totpop[popul[i].categ]:=-totpop[popul[i].categ]+1;
for k:-1 to m do
    teta[k]:=-totpop[k]/npopul;

    optibnd(popul,npopul,m,1,bound,minvar);

end.

```

APPENDIX E

The OPTIBND5 Program

```
var dfile: filetype;
    fildata: glnarray;
    popsort1, popsort2: poparray;
    bound1, bound2: bndtype;
```

```

    stw,stmean1,stmean2,stvar,sts21,sts22:realarray;
    stnsample:intarray;
    iter,ndata,npopul,h,i,j,k,totnsamp:integer;
    sampfrac,next,minvar,variance,weight,totw,totmean1,totmean2,totvar,
    filler1,filler2,filler3:real;

(*****)

procedure sortshll(indic:integer;var x:glnarray;n:integer);

var jump:integer;
    alldone:boolean;
    temp:real;

begin
    jump:=n;
    while jump>1 do begin
        jump:=jump div 2;
        repeat
            alldone:=true;
            for i:=1 to n-jump do begin
                j:=i+jump;
                if ((indic=1)and(x[i].pop1>x[j].pop1))or
                    ((indic=2)and(x[i].pop2>x[j].pop2)) then begin
                    temp:=x[i].pop1;
                    x[i].pop1:=x[j].pop1;
                    x[j].pop1:=temp;
                    temp:=x[i].pop2;
                    x[i].pop2:=x[j].pop2;
                    x[j].pop2:=temp;
                    alldone:=false;
                end
            end
        until alldone
    end
end;

(+++++)

function w(x:poparray;n:integer;lowv1,highv1,lowv2,highv2:real):real;

var i,totn:integer;

begin
    totn:=0;
    for i:=1 to n do
        if (lowv1<x[i].pop1)and(x[i].pop1<=highv1)and
            (lowv2<x[i].pop2)and(x[i].pop2<=highv2) then totn:=totn+1;
    w:=-totn/n;
end;

(*****)

function mu(indic:integer;x:poparray;n:integer;
    lowv1,highv1,lowv2,highv2:real):real;

var totx:real;
    i,totn:integer;

begin

```



```

totx:=0.0;
totn:=0;
for i:=1 to n do
  if (lowv1<x[i].pop1)and(x[i].pop1<=highv1)and
    (lowv2<x[i].pop2)and(x[i].pop2<=highv2) then begin
    if indic=1 then totx:=totx+x[i].pop1
    else totx:=totx+x[i].pop2;
    totn:=totn+1;
  end;
  if totn>0 then mu:=-0.0
  else mu:=-totx/totn;
end;

```

{*****}

```

function sigma2(indic:integer;x:poparray;n:integer;
  lowv1,highv1,lowv2,highv2:real):real;

```

```

var i,totn:integer;
    totx,totx2:real;

```

```

begin
  totn:=0;
  totx:=0.0;
  totx2:=0.0;
  for i:=1 to n do
    if (lowv1<x[i].pop1)and(x[i].pop1<=highv1)and
      (lowv2<x[i].pop2)and(x[i].pop2<=highv2) then begin
      if indic=1 then begin
        totx:=totx+x[i].pop1;
        totx2:=totx2+sqr(x[i].pop1);
      end
      else begin
        totx:=totx+x[i].pop2;
        totx2:=totx2+sqr(x[i].pop2);
      end;
      totn:=totn+1;
    end;
    if totn<=1 then sigma2:=-0.0
    else sigma2:=(totx2-sqr(totx)/totn)/(totn-1);
  end;

```

{*****}

```

function min(a,b:real):real;

```

```

begin

```

```

  if a<b then min:=-a
  else min:=-b;

```

```

end;

```

```

function max(a,b:real):real;

```

```

begin

```

```

  if a>b then max:=-a
  else max:=-b;

```

```

end;

```

{*****}

```

function wexpon(x:poparray;n:integer;lowv1,highv1,lowv2,highv2:real):real;

```

```

{ for hypot. distr. }

```

```

var lambda1,lambda2,tot1,tot2:real;

```

```

begin
  lambdal:=-1.0;
  tot1:=exp(-lambdal*lowv1)-exp(-lambdal*highv1);
  lambda2:=-2.0;
  tot2:=exp(-lambda2*lowv2)-exp(-lambda2*highv2);
  wexpon:=-tot1*tot2;
end;

(+++++)

function muexpon(indic:integer;x:poparray;n:integer;
                 lowv1,highv1,lowv2,highv2:real):real;
( for hypot. distr. )
var lambdal,lambda2,expon1,exponh,tot,denom:real;
begin
  if indic=1 then begin
    lambdal:=-1.0;
    expon1:=exp(-lambdal*lowv1);
    exponh:=exp(-lambdal*highv1);
    tot:=-expon1*(lowv1+1/lambdal)-exponh*(highv1+1/lambdal);
    denom:=-expon1-exponh;
  end
  else begin
    lambda2:=-2.0;
    expon1:=exp(-lambda2*lowv2);
    exponh:=exp(-lambda2*highv2);
    tot:=-expon1*(lowv2+1/lambda2)-exponh*(highv2+1/lambda2);
    denom:=-expon1-exponh;
  end;
  if denom=0.0 then muexpon:=0.0
  else muexpon:=-tot/denom;
end;

```

```

(+++++)

function s2expon(indic:integer;x:poparray;n:integer;
                 lowv1,highv1,lowv2,highv2:real):real;
( for hypot. distr. )
var lambdal,lambda2,expon1,exponh,tot1,tot2,denom:real;
begin
  if indic=1 then begin
    lambdal:=-1.0;
    expon1:=exp(-lambdal*lowv1);
    exponh:=exp(-lambdal*highv1);
    tot1:=-expon1*(sqr(lowv1)+2*lowv1/lambdal+2/sqr(lambdal))-
           exponh*(sqr(highv1)+2*highv1/lambdal+2/sqr(lambdal));
    tot2:=-expon1*(lowv1+1/lambdal)-exponh*(highv1+1/lambdal);
    denom:=-expon1-exponh;
  end
  else begin
    lambda2:=-2.0;
    expon1:=exp(-lambda2*lowv2);
    exponh:=exp(-lambda2*highv2);
    tot1:=-expon1*(sqr(lowv2)+2*lowv2/lambda2+2/sqr(lambda2))-
           exponh*(sqr(highv2)+2*highv2/lambda2+2/sqr(lambda2));
    tot2:=-expon1*(lowv2+1/lambda2)-exponh*(highv2+1/lambda2);
    denom:=-expon1-exponh;
  end;
  if denom=0.0 then s2expon:=0.0
  else s2expon:=-tot1/denom-sqr(tot2/denom);
end;

```

end;

(*****)

procedure printall(iter:integer;xsort1,xsort2:poparray;
n,lstrat1,lstrat2:integer;b1,b2:bnctype);

begin

stvar[1,1]:=-tetal*sigma2(1,xsort1,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2)+
teta2*sigma2(1,xsort2,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2);
(both -0.0; for hypot. distr.)
stvar[2,2]:=-tetal*sigma2(2,xsort1,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2)+
teta2*sigma2(2,xsort2,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2);
(both -0.0; for hypot. distr.)

variance:=(tetal*stvar[1,1]+teta2*stvar[2,2])/nsample;

if iter=0 then begin

writeln(CON);writeln(CON);

writeln(CON,' ':2,'*** DYNAMIC PROGRAMMING SIMULATION ***');

writeln(CON);

writeln(CON,' ':2,'data from file ',dosfile);

writeln(CON);

writeln(CON,' ':2,'*** univariate statistics ***');

writeln(CON);

writeln(CON,' ':2,'*** population 1:');

writeln(CON);

writeln(CON,' ':2,'number of observations: ',n:1);

writeln(CON,' ':2,'population mean: ',

(tetal*mu(1,xsort1,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2)+
teta2*mu(1,xsort2,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2)):8:3);
(both -0.0; for hypot. distr.)

writeln(CON,' ':2,'population variance (S2): ',stvar[1,1]:8:3);

writeln(CON,' ':2,'sample size: ',nsample:1);

writeln(CON,' ':2,'variance for unstratified estimate (Vlsrs): ',
(stvar[1,1]/nsample):8);

writeln(CON);

writeln(CON,' ':2,'*** population 2:');

writeln(CON);

writeln(CON,' ':2,'number of observations: ',npopul:1);

writeln(CON,' ':2,'population mean: ',

(tetal*mu(2,xsort1,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2)+
teta2*mu(2,xsort2,n,(xsort1[1].pop1-1.0),xsort1[n].pop1,
(xsort2[1].pop2-1.0),xsort2[n].pop2)):8:3);
(both -0.0; for hypot. distr.)

writeln(CON,' ':2,'population variance (S2): ',stvar[2,2]:8:3);

writeln(CON,' ':2,'sample size: ',nsample:1);

writeln(CON,' ':2,'variance for unstratified estimate (V2srs): ',
(stvar[2,2]/nsample):8);

writeln(CON);

writeln(CON,' ':2,'weighted average for unstratified estimate (Vsrs): ',
variance:8);

writeln(CON);

end;

writeln(CON,' ':2,'***** ITERATION ',iter:1,' *****');

```

      '*****');
writeln(CON);
writeln(CON, ' :2, '*** optimal stratification (with Neyman",
      ' allocation) ***');
writeln(CON);
writeln(CON, ' :2, '*** population 1:');
writeln(CON);
writeln(CON, ' :2, 'number of intervals: ', lstrat1:1);
writeln(CON, ' :2, 'optimal bounds of stratification: ');
write(CON, ' :2);
for h:=0 to lstrat1 do
  write(CON, ' b(', h:1, ')');
writeln(CON);
write(CON, ' :2);
for h:=0 to lstrat1 do
  write(CON, ' ', b1[h]:8:2);
writeln(CON);writeln(CON);
writeln(CON, ' :2, '*** population 2:');
writeln(CON);
writeln(CON, ' :2, 'number of intervals: ', lstrat2:1);
writeln(CON, ' :2, 'optimal bounds of stratification: ');
write(CON, ' :2);
for h:=0 to lstrat2 do
  write(CON, ' b(', h:1, ')');
writeln(CON);
write(CON, ' :2);
for h:=0 to lstrat2 do
  write(CON, ' ', b2[h]:8:2);
writeln(CON);writeln(CON);

b1[0]:=-b1[0]-1.0; ( -0.0; for hypot. distr. )
b2[0]:=-b2[0]-1.0; ( -0.0; for hypot. distr. )
weight:=0.0;
for h:=1 to lstrat1 do
  for k:=1 to lstrat2 do begin
    stw[h,k]:=-tetal*w(xsort1,n,b1[h-1],b1[h],b2[k-1],b2[k])+
      teta2*w(xsort2,n,b1[h-1],b1[h],b2[k-1],b2[k]);
    stmean1[h,k]:=-tetal*mu(1,xsort1,n,b1[h-1],b1[h],b2[k-1],b2[k])+
      teta2*mu(1,xsort2,n,b1[h-1],b1[h],b2[k-1],b2[k]);
    stmean2[h,k]:=-tetal*mu(2,xsort1,n,b1[h-1],b1[h],b2[k-1],b2[k])+
      teta2*mu(2,xsort2,n,b1[h-1],b1[h],b2[k-1],b2[k]);
    sts21[h,k]:=-tetal*sigma2(1,xsort1,n,b1[h-1],b1[h],b2[k-1],b2[k])+
      teta2*sigma2(1,xsort2,n,b1[h-1],b1[h],b2[k-1],b2[k]);
    sts22[h,k]:=-tetal*sigma2(2,xsort1,n,b1[h-1],b1[h],b2[k-1],b2[k])+
      teta2*sigma2(2,xsort2,n,b1[h-1],b1[h],b2[k-1],b2[k]);
    ( Neyman allocation is assumed )
    weight:=-weight+stw[h,k]*sqrt(tetal*sts21[h,k]+teta2*sts22[h,k]);
  end;
totw:=-0.0;
totmean1:=-0.0;
totmean2:=-0.0;
totnsamp:=0;
totvar:=-0.0;
for h:=1 to lstrat1 do
  for k:=1 to lstrat2 do begin
    totw:=-totw+stw[h,k];
    totmean1:=-totmean1+stw[h,k]*stmean1[h,k];
    totmean2:=-totmean2+stw[h,k]*stmean2[h,k];
    ( weight is given for Neyman allocation )
    stnsample[h,k]:=-round(nsample*stw[h,k]*

```

```

      sqrt(tetal*sts21[h,k]+teta2*sts22[h,k])/weight);
totnsamp:=-totnsamp+stnsample[h,k];
stvar[h,k]:=-(tetal*sts21[h,k]+teta2*sts22[h,k])/stnsample[h,k];
totvar:=-totvar+sqr(stw[h,k])*stvar[h,k];
end;

writeln(CON,' ':2,'*** stratum statistics ***');
writeln(CON);
if iter=0 then begin
  writeln(CON,' ':2,'displayed as: variable 2 intervals');
  writeln(CON,' ':2,'
                                1          2');
  writeln(CON,' ':2,'
                                +-----+-----+');
  writeln(CON,' ':2,'
                                1 | var 1 | var 1 |');
  writeln(CON,' ':2,'
                                | var 2 | var 2 |');
  writeln(CON,' ':2,' variable 1
                                +-----+-----+');
  writeln(CON,' ':2,' intervals 2 | var 1 | var 1 |');
  writeln(CON,' ':2,'
                                | var 2 | var 2 |');
  writeln(CON,' ':2,'
                                +-----+-----+');
  writeln(CON);
end;
writeln(CON,' ':2,' % stratum proportion:');
for h:=1 to lstrat1 do begin
  write(CON,' ':2,' ');
  for k:=1 to lstrat2 do
    write(CON,' ':2,100*stw[h,k]:8:2);
  writeln(CON);writeln(CON);
end;
writeln(CON,' ':2,' ... overall proportion: ',100*totw:8:2);
writeln(CON);
writeln(CON,' ':2,' stratum means:');
for h:=1 to lstrat1 do begin
  write(CON,' ':2,' ');
  for k:=1 to lstrat2 do
    write(CON,' ':2,stmean1[h,k]:8:3);
  writeln(CON);
  write(CON,' ':2,' ');
  for k:=1 to lstrat2 do
    write(CON,' ':2,stmean2[h,k]:8:3);
  writeln(CON);writeln(CON);
end;
writeln(CON,' ':2,' ... overall means: ',totmean1:8:3);
writeln(CON,' ':2,' ',totmean2:8:3);
writeln(CON);
writeln(CON,' ':2,' stratum variance (S2hk):');
for h:=1 to lstrat1 do begin
  write(CON,' ':2,' ');
  for k:=1 to lstrat2 do
    write(CON,' ':2,sts21[h,k]:8:3);
  writeln(CON);
  write(CON,' ':2,' ');
  for k:=1 to lstrat2 do
    write(CON,' ':2,sts22[h,k]:8:3);
  writeln(CON);writeln(CON);
end;
writeln(CON,' ':2,' stratum sample size:');
for h:=1 to lstrat1 do begin
  write(CON,' ':2,' ');
  for k:=1 to lstrat2 do
    write(CON,' ':2,stnsample[h,k]:8);
  writeln(CON);writeln(CON);
end;

```

```

end;
writeln(CON,' ':2,' ... overall sample size: ',totnsamp:8);
writeln(CON);
writeln(CON,' ':2,' stratum weighted average (Vxyhk): ');
for h:-1 to lstrat1 do begin
  write(CON,' ':2,' ');
  for k:-1 to lstrat2 do
    write(CON,' ':2,stvar[h,k]:8);
  writeln(CON);writeln(CON);
end;
writeln(CON,' ':2,' ... overall weighted average (Vxy): ',totvar:8);
writeln(CON);
variance:=100*totvar/variance;
writeln(CON,' ':2,'% variance ratio (Vxy/Vsrs): ',variance:8:2,'%');
writeln(CON);writeln(CON);
end;

```

(*****)

```

procedure optibnd(zsort1,zsort2:poparray;n,lstrat1,lstrat2:integer;
  var blval,b2val:bndtype;var fstar0:real);

```

```

var fstar:array[1..1,1..dsize] of real;
nextbnd:array[0..1,1..dsize] of integer;
compflag:array[1..1,1..dsize] of boolean;
z:poparray;
preslbnd:integer;
bold,b2old:bndtype;

```

(.....)

```

function maxdiff(bnew,bold:bndtype;lstrat:integer):real;

```

```

var diff,maxim:real;

```

```

begin
  maxim:=-maxreal;
  for h:-1 to lstrat do begin
    if bold[h]=0.0 then diff:=-maxreal
    else diff:=abs((bnew[h]-bold[h])/bold[h]);
    if diff>maxim then maxim:=diff;
  end;
  maxdiff:=-maxim;
end;

```

(.....)

```

function objf(lowv1,highv1,lowv2,highv2:real):real;

```

```

var whk:real;

```

```

begin
  whk:=-w(z,n,lowv1,highv1,lowv2,highv2);
  if (lowv1>highv1) or (lowv2>highv2) or
    (round(n*whk)<=-1) then objf:=-maxreal
  else objf:=-whk*sqrt(teta1*sigma2(1,z,n,lowv1,highv1,lowv2,highv2)+
    teta2*sigma2(2,z,n,lowv1,highv1,lowv2,highv2));
end;

```

(.....)

```

function cumfstarl(stage,lbound:integer):real;

( This function computes the OPS with respect to variable 1 )

var f,minval,lbndv:real;
    nextbnd,hbound:integer;

begin
  if compflag[stage,lbound] then minval:=-fstar[stage,lbound]
  else begin
    if lbound=1 then lbndv:=-z[lbound].pop1-1.0 ( -0.0 for hypot. distr. )
    else lbndv:=-z[lbound-1].pop1;
    if stage=lstrat1 then begin
      minval:=-0.0;
      for k:=1 to lstrat2 do
        minval:=-minval+objf(lbndv,z[n].pop1,b2val[k-1],b2val[k]);
      nextbnd:=-n+1;
    end
    else begin
      minval:=maxreal;
      for hbound:=-lbound to n do begin
        f:=0.0;
        for k:=1 to lstrat2 do
          f:=-f+objf(lbndv,z[hbound].pop1,b2val[k-1],b2val[k]);
        f:=-f+cumfstarl(stage+1,hbound+1);
        if f<minval then begin
          minval:=-f;
          nextbnd:=hbound+1;
        end;
      end;
    end;
    fstar[stage,lbound]:=-minval;
    nextbnd[stage,lbound]:=-nextbnd;
    compflag[stage,lbound]:=true;
    WRITELN('POPULATION 1 ITERATIONS, STAGE: ',STAGE:1,
            ' LBND: ',LBND:1, ' CUMF*: ',MINVAL:8:4);
  end;
  cumfstarl:=-minval;
end;

```

(.....)

```

function cumfstar2(stage,lbound:integer):real;

( This function computes the OPS with respect to variable 2 )

var f,minval,lbndv:real;
    nextbnd,hbound:integer;

begin
  if compflag[stage,lbound] then minval:=-fstar[stage,lbound]
  else begin
    if lbound=1 then lbndv:=-z[lbound].pop2-1.0 ( -0.0 for hypot. distr. )
    else lbndv:=-z[lbound-1].pop2;
    if stage=lstrat2 then begin
      minval:=-0.0;
      for h:=1 to lstrat1 do
        minval:=-minval+objf(blval[h-1],blval[h],lbndv,z[n].pop2);
      nextbnd:=-n+1;
    end
  end
end;

```

```

end
else begin
  minval:=-maxreal;
  for hbound:=-lbound to n do begin
    f:=0.0;
    for h:-1 to lstrat1 do
      f:=f+objf(blval[h-1],blval[h],lbndv,z[hbound].pop2);
    f:=f+cumfstar2(stage+1,hbound+1);
    if f<minval then begin
      minval:=-f;
      nextbnd:=-hbound+1;
    end;
  end;
  end;
  fstar[stage,lbound]:=-minval;
  nextbnd[stage,lbound]:=-nextbnd;
  compflag[stage,lbound]:=true;
  WRITELN('POPULATION 2 ITERATIONS, STAGE: ',STAGE:1,
    ' LBND: ',LBOUND:1, ' CUMF*: ',MINVAL:8:4);
end;
cumfstar2:=-minval;
end;

```

(.....)

```

begin {procedure optibnd}

```

```

  blval[0]:=-zsort1[1].pop1;
  sampfrac:=lstrat1/n;
  next:=1/sampfrac;
  for h:-1 to lstrat1 do begin
    blval[h]:=-zsort1[round(next)].pop1;
    next:=next+1/sampfrac;
  end;
  b2val[0]:=-zsort2[1].pop2;
  b2old[0]:=-b2val[0];
  sampfrac:=lstrat2/n;
  next:=1/sampfrac;
  for k:-1 to lstrat2 do begin
    b2val[k]:=-zsort2[round(next)].pop2;
    b2old[k]:=-b2val[k];
    next:=next+1/sampfrac;
  end;

  iter:=0;
  printall(iter,zsort1,zsort2,n,lstrat1,lstrat2,blval,b2val);
  repeat
    iter:=iter+1;

    z:=zsort1;
    blold[0]:=-blval[0];
    for h:-1 to lstrat1 do begin
      blold[h]:=-blval[h];
      for j:-1 to n do
        compflag[h,j]:=false;
      end;

      b2val[0]:=-b2val[0]+1.0; ( -0.0; for hypot. distr. )
      fstar0:=cumfstar1(1,1);
      b2val[0]:=-b2val[0]+1.0; ( +0.0; for hypot. distr. )
    end;
  until iter=lstrat2;
end;

```



```

    blval[0]:=-z[1].pop1;
    preslbnd:=-1;
    for h:-1 to lstrat1 do begin
        preslbnd:=nextbnd[h,preslbnd];
        blval[h]:=-z[preslbnd-1].pop1;
    end;
    printall(iter,zsort1,zsort2,n,lstrat1,lstrat2,blval,b2val);

    if (maxdiff(blval,bldold,lstrat1)>epsilon1)or
       (maxdiff(b2val,b2old,lstrat2)>epsilon2) then begin

        z:=zsort2;
        b2old[0]:=-b2val[0];
        for k:-1 to lstrat2 do begin
            b2old[k]:=-b2val[k];
            for j:-1 to n do
                compflag[k,j]:=false;
        end;

        blval[0]:=-blval[0]-1.0; ( -0.0; for hypot. distr. )
        fstar0:=cumfstar2(1,1);
        blval[0]:=-blval[0]+1.0; ( +0.0; for hypot. distr. )

        b2val[0]:=-z[1].pop2;
        preslbnd:=-1;
        for k:-1 to lstrat2 do begin
            preslbnd:=nextbnd[k,preslbnd];
            b2val[k]:=-z[preslbnd-1].pop2;
        end;
        printall(iter,zsort1,zsort2,n,lstrat1,lstrat2,blval,b2val);

    end;
until ((maxdiff(blval,bldold,lstrat1)<epsilon1)
      and(maxdiff(b2val,b2old,lstrat2)<epsilon2))or(iter=maxiter);
writeln(CON,'*****');
      '*****');
end;

(*****)

begin (main program)
    assign(dfile,dosfile);
    reset(dfile);
    i:-1;
    readln(dfile,filler1,filler2,filler3,filedata[i].pop1,filedata[i].pop2);
    while not eof(dfile) do begin
        i:=i+1;
        readln(dfile,filler1,filler2,filler3,filedata[i].pop1,filedata[i].pop2);
    end;
    ndata:=i;
    close(dfile);

    tshll(1,filedata,ndata);
    if ndata > dsize then begin
        npopul:=dsize;
        sampfrac:=dsize/ndata;
    end
    else begin
        npopul:=ndata;

```

```

    sampfrac:=-1.0;
end;
next:=-1.0;
for j:-1 to npopul-1 do begin
    popsort1[j]:=-fildata[round(next)];
    next:=-next+1/sampfrac;
end;
popsort1[ npopul ]:=-fildata[ndata];
FOR I:-1 TO NPOPUL DO WRITELN(I, ' ',POPSORT1[I].POP1:6:3,
                                ' ',POPSORT1[I].POP2:6:3);

WRITELN;

sortsh11(2,fildata,ndata);
if ndata > dsize then begin
    npopul:=-dsize;
    sampfrac:=-dsize/ndata;
end
else begin
    npopul:=-ndata;
    sampfrac:=-1.0;
end;
next:=-1.0;
for j:-1 to npopul-1 do begin
    popsort2[j]:=-fildata[round(next)];
    next:=-next+1/sampfrac;
end;
popsort2[ npopul ]:=-fildata[ndata];
FOR I:-1 TO NPOPUL DO WRITELN(I, ' ',POPSORT2[I].POP1:6:3,
                                ' ',POPSORT2[I].POP2:6:3);

optibnd(popsort1,popsort2, npopul,11,12,bound1,bound2,minvar);

end.

```

APPENDIX F

Convergence of the Unidimensional Iterative Dynamic Programming Approach

APPENDIX F: CONVERGENCE OF THE UNIDIMENSIONAL ITERATIVE DYNAMIC PROGRAMMING APPROACH

In Section 6.1.3, the unidimensional iterative DP approach was proposed for solving the two-way optimum stratification problem. This approach consists basically in first fixing the points of stratification with respect to one of the stratification variables X or Y and then computing the OPS for the remaining variable. In the second step, the resulting OPS are considered themselves as fixed and new OPS are determined for the other stratification variable which was first considered as fixed. The process iterates until convergence is attained based on a predetermined criterion.

Unfortunately, the unidimensional iterative DP approach, which in principle is appealing because of its simplicity, cannot be proved to converge in general. In fact, examples can easily be constructed where the process will iterate indefinitely without ever converging to some point. However, it may be argued that most of the distributions encountered in practice are such that the unidimensional iterative DP approach converges to some point which is, with probability 1, a local minimum.

Although the proposed approach does not converge for all possible distributions, it may be proved that, assuming that the process converges to some points, this point is in fact a local minimum. A proof along these lines is first given for the case where $L = 2$ and $M = 2$.

F.1 Convergence to the Correct Limit for the Case $L = 2$ and $M = 2$.

Based on the general form of the objective function given by (6.24) for the two-way optimum stratification problem in the infinite population case, the objective function for the case $L = 2$ and $M = 2$ can be rewritten as

$$(F.1) \quad G(\underline{x}, \underline{y}) = g(x_0, x_1, y_0, y_1) + g(x_0, x_1, y_1, y_2) + g(x_1, x_2, y_0, y_1) + g(x_1, x_2, y_1, y_2).$$

In formula (F.1), the values of x_0, x_2, y_0 and y_2 are in fact constants since $x_0 = a_X, x_2 = b_X, y_0 = a_Y$ and $y_2 = b_Y$. Because of this, the objective function simply turns out to be a two variable function which can be expressed as

$$(F.2) \quad G(\underline{x}, \underline{y}) = p(x_1, y_1) = p(x, y).$$

The optimization problem can therefore be written as

$$(F.3) \quad \begin{aligned} &\text{Min } p(x, y) \\ &\text{subject to : } (x, y) \in C = [a_X, b_X] \times [a_Y, b_Y]. \end{aligned}$$

The unidimensional iterative DP approach can be viewed as finding the minimum value of $p(x, y)$ with respect to x for a given value of y , and vice-versa. For given values $(x^{(i)}, y^{(i)})$ at the i th iteration, the algorithm underlying the proposed DP approach may be expressed as

$$(F.4) \quad A_{pX}(C, y^{(i)}) = \{(x^{(i+1)}, y^{(i)}) | p(x^{(i+1)}, y^{(i)}) = \inf_{x \in [a_X, b_X]} p(x, y^{(i)})\}$$

and

$$(F.5) \quad A_{pY} \left(C, x^{(i+1)} \right) = \{ (x^{(i+1)}, y^{(i+1)}) | p(x^{(i+1)}, y^{(i+1)}) = \inf_{y \in [a_Y, b_Y]} p(x^{(i+1)}, y) \}.$$

We denote this algorithm by A_p for the function $p(x, y)$. It should be noted that the formulations (F.4) and (F.5) respectively come from the Functional Equations of Bellman given by (6.30) and (6.31). Note also that, on a closed set, taking the infimum is equivalent as taking the minimum.

Before stating a theorem (preceded by a lemma) on the local minimum from the unidimensional iterative DP approach, two other well-known theorems (see Marsden and Tromba (1976)) are given for completeness. The first theorem is related to sufficient conditions for local minima, local maxima and saddle-points, while the second one is the Second-Order Taylor's theorem.

Theorem 1: Let $p(x, y)$ be third-order differentiable on an open set D . A point (x^*, y^*) is a local minimum of p provided the following three conditions hold:

$$(F.6) \text{ i)} \quad \frac{\partial p}{\partial x}(x^*, y^*) = \frac{\partial p}{\partial y}(x^*, y^*) = 0$$

$$(F.7) \text{ ii)} \quad \frac{\partial^2 p}{\partial x^2}(x^*, y^*) > 0$$

$$(F.8) \text{ iii)} \quad \Delta = \left(\frac{\partial^2 p}{\partial x^2} \right) \left(\frac{\partial^2 p}{\partial y^2} \right) - \left(\frac{\partial^2 p}{\partial x \partial y} \right)^2 > 0 \text{ at } (x^*, y^*).$$

If in (ii), we have < 0 instead of > 0 and condition (iii) is unchanged, then we have a strict local maximum. On the other hand, if $\Delta < 0$ then we have a saddle-point.

The critical point (x^*, y^*) is called a non-degenerate critical point when $\Delta \neq 0$. Otherwise, the critical point (x^*, y^*) is called degenerate and then the sufficient conditions of Theorem 1 no longer hold so that local minima, local maxima and saddle-points must be identified using other techniques.

Theorem 2 (Second-Order Taylor's Theorem): Let $p(x, y)$ be third-order differentiable at a point $(x^*, y^*) \in D$, where D is an open set. Then, for any point $(x, y) \in D$, we may write

$$(F.9) \quad \begin{aligned} p(x, y) = & p(x^*, y^*) + (x - x^*) \frac{\partial p}{\partial x}(x^*, y^*) + (y - y^*) \frac{\partial p}{\partial y}(x^*, y^*) \\ & + \frac{(x - x^*)^2}{2} \frac{\partial^2 p}{\partial x^2}(x^*, y^*) + \frac{(y - y^*)^2}{2} \frac{\partial^2 p}{\partial y^2}(x^*, y^*) + (x - x^*)(y - y^*) \frac{\partial^2 p}{\partial x \partial y}(x^*, y^*) \\ & + R(x, y) \end{aligned}$$

where

$$(F.10) \quad \lim_{(x, y) \rightarrow (x^*, y^*)} \frac{R(x, y)}{(x - x^*)^2 + (y - y^*)^2} = 0.$$

These two theorems will be used in the proof of a theorem on the local minimum from the unidimensional iterative DP approach. This proof, however, also requires the use of the following lemma.

Lemma 1: Let $q(X, Y) = aX^2 + bXY + cY^2 + d$ be defined on an open disk $D_\epsilon(0, 0)$ of radius ϵ centered on $(0, 0)$. Then, if the algorithm A_q applied to the function q converges to $(0, 0)$, we have $a > 0$ and $c > 0$.

Proof: Assume that the algorithm A_q converges to $(0, 0)$. We shall first prove that $a \neq 0$ and $c \neq 0$. Suppose that $a = 0$ or $c = 0$, or both. Without loss of generality, we may take $a = 0$ so that

$$q(X, Y) = bXY + cY^2 + d.$$

Then, for fixed $Y = Y_0$ where $|Y_0| < \epsilon$, $q(X, Y_0)$ is a linear function of X with slope bY_0 . Therefore, if $Y_0 \neq 0$ and $b \neq 0$, then $A_{qX}(D_\epsilon(0, 0), Y_0) \subset bd(D_\epsilon(0, 0))$ where $bd(D_\epsilon(0, 0))$ is the set of boundary points of $D_\epsilon(0, 0)$ so that $(0, 0) \notin bd(D_\epsilon(0, 0))$. On the other hand, if $Y_0 = 0$ or $b = 0$, then $A_{qX}(D_\epsilon(0, 0), Y_0)$ can take any value of X for which $(X, Y_0) \in D_\epsilon(0, 0)$ since $q(X, Y_0)$ does not involve the variable X . Therefore, the algorithm A_q cannot converge to $(0, 0)$ since, for any $Y = Y_0$ where $|Y_0| < \epsilon$, $A_{qX}(D_\epsilon(0, 0), Y_0)$ cannot converge to $(0, 0)$. This leads to a contradiction so that $a \neq 0$ and $c \neq 0$.

We now prove that $a > 0$ and $c > 0$. Suppose that $a < 0$ or $c < 0$, or both. Then, since $a \neq 0$ and $c \neq 0$, by completing the squares, $q(X, Y)$ can be written as

$$(F.11) \quad q(X, Y) = a\left(X + \frac{b}{2a}Y\right)^2 + d + Y^2\left(\frac{4ac - b^2}{4a}\right)$$

$$(F.12) \quad \text{and} \quad q(X, Y) = c\left(Y + \frac{b}{2c}X\right)^2 + d + X^2\left(\frac{4ac - b^2}{4c}\right).$$

Without loss of generality, we may take $a < 0$. Then, for fixed $Y = Y_0$ where $|Y_0| < \epsilon$, $A_{qX}(D_\epsilon(0, 0), Y_0) \subset bd(D_\epsilon(0, 0))$ since $q(X, Y_0)$ is a parabola with its maximum at $(-\frac{b}{2a}Y_0, Y_0)$. Therefore the algorithm A_q cannot converge to $(0, 0)$ since for any $Y = Y_0$ where $|Y_0| < \epsilon$, $A_{qX}(D_\epsilon(0, 0), Y_0)$ cannot converge to $(0, 0)$. This again leads to a contradiction so that $a > 0$ and $b > 0$. $\square$

Assuming that the proposed unidimensional iterative DP approach converges to some point (x^*, y^*) , Theorem 3 below proves, for the case $L = 2$ and $M = 2$, that (x^*, y^*) is, with probability 1, a local minimum or, with probability 0, a saddle-point.

Theorem 3: Let the algorithm A_p applied to the function p be the unidimensional iterative DP approach for the case where $L = 2$ and $M = 2$. Also let $p(x, y)$ be third-order differentiable at a point $(x^*, y^*) \in D$ where D is the interior of the domain $[a_X, b_X] \times [a_Y, b_Y]$. Assume that all the critical points (i.e. points $(\tilde{x}, \tilde{y})$ where $\frac{\partial p}{\partial x}(\tilde{x}, \tilde{y}) = \frac{\partial p}{\partial y}(\tilde{x}, \tilde{y}) = 0$) contained in D are non-degenerate. Then, if for some starting values $(x^{(0)}, y^{(0)}) \in D$, the algorithm A_p converges to the point (x^*, y^*) , we have that (x^*, y^*) is a local minimum with probability 1 or a saddle-point with probability 0.

Proof: Assume that, for some starting values $(x^{(0)}, y^{(0)}) \in D$, the algorithm A_p converges to the point $(x^*, y^*) \in D$. We shall first show that (x^*, y^*) is in fact a critical point. Suppose that (x^*, y^*) is not a critical point. Then, $\frac{\partial p}{\partial x}(x^*, y^*) \neq 0$ or $\frac{\partial p}{\partial y}(x^*, y^*) \neq 0$, or both. Without loss of generality, we may assume that $\frac{\partial p}{\partial x}(x^*, y^*) \neq 0$. Since $\frac{\partial p}{\partial x}(x^*, y^*) \neq 0$, there exists a point $(x^{**}, y^*) \in D$ such that $p(x^{**}, y^*) < p(x^*, y^*)$, i.e. $\{(x^*, y^*)\} \not\subset A_{pX}(D, y^*)$ for all $y^* \in]a_Y, b_Y[$. Therefore, the algorithm A_p cannot converge to (x^*, y^*) , leading to a contradiction. Hence (x^*, y^*) is a critical point.

Now since the algorithm A_p converges to (x^*, y^*) , by definition, for every open disks $D_\epsilon(x^*, y^*)$ centered on (x^*, y^*) , there exists an N such that for all $i \geq N$, we have $(x^{(i)}, y^{(i)}) \in D_\epsilon(x^*, y^*)$. On the other hand, by theorem 2, for any $(x, y) \in D_\epsilon(x^*, y^*)$, $p(x, y)$ may be written as

$$(F.13) \quad p(x, y) = p(x^*, y^*) + \frac{(x - x^*)^2}{2} \frac{\partial^2 p}{\partial x^2}(x^*, y^*) + (x - x^*)(y - y^*) \frac{\partial^2 p}{\partial x \partial y}(x^*, y^*) \\ + \frac{(y - y^*)^2}{2} \frac{\partial^2 p}{\partial y^2}(x^*, y^*) + R(x, y),$$

since $\frac{\partial p}{\partial x}(x^*, y^*) = \frac{\partial p}{\partial y}(x^*, y^*) = 0$. Letting $a = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}(x^*, y^*)$, $b = \frac{\partial^2 p}{\partial x \partial y}(x^*, y^*)$, $c = \frac{1}{2} \frac{\partial^2 p}{\partial y^2}(x^*, y^*)$ and $d = p(x^*, y^*)$, (F.13) can be express as

$$p(x, y) = a(x - x^*)^2 + b(x - x^*)(y - y^*) + c(y - y^*)^2 + d + R(x, y).$$

Also, letting $X = x - x^*$ and $Y = y - y^*$, we have

$$(F.14) \quad p(x, y) = q(X, Y) = aX^2 + bXY + cY^2 + d + R^1(X, Y).$$

where $R^1(X, Y) = R(X + x^*, Y + y^*)$.

Now, as $\epsilon \rightarrow 0$, $(x, y) \rightarrow (x^*, y^*)$ and consequently $(X, Y) \rightarrow (0, 0)$. Therefore, for $\epsilon \rightarrow 0$, by Theorem 2, $q(X, Y) \rightarrow aX^2 + bXY + cY^2 + d$. Further, since $(X, Y) \in D_\epsilon(0, 0)$, by Lemma 1, for $\epsilon \rightarrow 0$, we have that $a > 0$ and $c > 0$. By completing the squares, for $\epsilon \rightarrow 0$, $p(x, y)$ can then be written as

$$(F.15) \quad p(x, y) = a\left(x - x^* + \frac{b}{2a}(y - y^*)\right)^2 + d + (y - y^*)^2 \left(\frac{4ac - b^2}{4a}\right)$$

$$(F.16) \quad \text{and} \quad p(x, y) = c\left(y - y^* + \frac{b}{2c}(x - x^*)\right)^2 + d + (x - x^*)^2 \left(\frac{4ac - b^2}{4c}\right).$$

We will now show, using these last two equations, that (x^*, y^*) is a local minimum, with probability 1, or a saddle-point, with probability 0, of the function p . First, since all critical points $(\tilde{x}, \tilde{y}) \in D$ are non-degenerate,

$$0 \neq \Delta = (2a)(2c) - (b)^2 = 4ac - b^2.$$

Thus, Δ can only be either positive or negative. Assume that $\Delta > 0$. Then, since $a > 0$ and $c > 0$, by Theorem 1, (x^*, y^*) is a local minimum of the function p . On the other hand, assume that $\Delta < 0$. Then, by Theorem 1, (x^*, y^*) is a saddle point of the function p . Now, since $a > 0$, considering $p(x, y)$ as a function of x only, $p(x, y)$ is a parabola with its minimum at $(-\frac{b}{2a}(y - y^*) + x^*, y)$. Let $y = y_0$ where $|y_0 - y^*| < \epsilon$. Therefore, as $\epsilon \rightarrow 0$, $A_{pX}(D, y_0) = \{(x_1, y_0)\}$ where $(x_1, y_0) \in D_\epsilon(x^*, y^*)$ and $x_1 = -\frac{b}{2a}(y_0 - y^*) + x^*$ since the algorithm A_p is assumed to converge to (x^*, y^*) . Now, since $c > 0$, considering $p(x, y)$ as a function of y only, $p(x, y)$ is a parabola with its minimum at $(x, -\frac{b}{2c}(x - x^*) + y^*)$. Therefore, since the algorithm A_p is assumed to converge to (x^*, y^*) , as $\epsilon \rightarrow 0$, $A_{pY}(D, x_1) = \{(x_1, y_1)\}$ where $(x_1, y_1) \in D_\epsilon(x^*, y^*)$ and $y_1 = \frac{b^2}{4ac}(y_0 - y^*) + y^*$. But $\Delta < 0$ or $b^2 > 4ac$ so that $|\frac{b^2}{4ac}(y_0 - y^*)| \geq |y_0 - y^*|$, i.e. $|y_1 - y^*| \geq |y_0 - y^*|$. Therefore, unless $y_0 = y^*$ or $b = 0$, the algorithm A_p cannot converge to (x^*, y^*) . But $b > 0$ since $b^2 > 4ac$ and $a > 0$ and $c > 0$. Therefore, y_0 must be equal to y^* , if y is fixed at y_0 . Similarly, x_0 must be equal to x^* , by first fixing $x = x_0$. Therefore, for $(x_0, y_0) \in D_\epsilon(x^*, y^*)$, we must have either $x_0 = x^*$ or $y_0 = y^*$, or both. Now, for a given ϵ , the set of points $B_\epsilon = \{(x, y) | (x, y) \in D_\epsilon(x^*, y^*) \text{ and } (x = x^* \text{ or } y = y^*)\}$ is a set of Lebesgue measure 0, i.e.

$$Pr [(x, y) \in B_\epsilon | (x, y) \in D_\epsilon(x^*, y^*)] = 0.$$

Therefore, if $\Delta < 0$ the algorithm A_p converges to (x^*, y^*) only with probability 0. Thus, the point of convergence (x^*, y^*) is, with probability 1, a local minimum or, with probability 0, a saddle-point. $\square$

Assuming that the unidimensional iterative DP approach converges, Theorem 3 establishes in fact that this approach converges, with probability 1, to the correct limit, i.e. to a local minimum. Note that the assumption of convergence is the most restrictive one since the proposed approach cannot be proved to converge in general, as noted before. It should also be noted that it might be feasible to prove Theorem 3 without the assumption that all the critical points are non-degenerate. However, this appears to be a much more difficult task.

F.2 Convergence to the Correct Limit for the Case $L \geq 2$ or $M \geq 2$.

A method similar to the one of Section F.1 may be used to prove the convergence to the correct limit of the unidimensional iterative DP approach for the case $L \geq 2$ or $M \geq 2$. That is, assuming that the process converges to some point (x^*, y^*) , this point, with probability 1, is a local minimum or, with probability 0, a saddle-point. However, due to the complexity of the problem, such a proof is outside the scope of this thesis.

END

08.08.88

FIN