ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

A study of optimal sampling schemes

By

Georgios Panagioti Sarris

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
June, 2009

https://doi.org/10.26219/heal.aueb.5554

https://doi.org/10.26219/heal.aueb.5554

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Μελέτη βέλτιστων μεθόδων δειγματοληψίας

Γεώργιος Παναγιώτη Σαρρής

ΔΙΑΤΡΙΒΗ
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Ιούνιος, 2009

https://doi.org/10.26219/heal.aueb.5554

https://doi.org/10.26219/heal.aueb.5554

DEDICATION

To my family

For existing, for educated me,
For encouraging,
For supporting me,
all my life….

https://doi.org/10.26219/heal.aueb.5554

ACKNOWLEDGEMENTS

I would like to thank a number of people who have encouraged me to write this thesis. This work would not have been possible without the support and encouragement of all them.

Firstly, I would like to express my deeper gratitude to my advisor Ioulia Papageorgiou under whose supervision I chose this topic and began the thesis. I thank my professor Ioulia for her invaluable support, encouragement and useful suggestions throughout this research work. I am very grateful for the cooperation and the interest of my professor. Her continuous guidance enabled me to complete my work successfully.

I am as ever, especially indebted to my parents, Panagiotis and Eleni for their love and support throughout my life. My mother, Eleni in the first place is the person who put the fundament my learning character, showing me the joy of intellectual pursuit ever since I was a child. My father, Panagiotis is the one who sincerely raised me with his caring and gently love. I also wish to thank my brother Antonis for his useful and invaluable advises for my life. Words fail me to express my appreciation to Vedrana whose dedication, love and persistent confidence in me, has taken the load off my shoulder. Their unflinching courage and conviction will always inspire me, and I hope to continue, in my own small way, the noble mission to which they gave their lives.

Georgios Sarris

https://doi.org/10.26219/heal.aueb.5554

VITA

I was born in 1983 in Athens Greece. I entered the department of Mathematics in the University of Athens in 2001 and I received my degree in 2006. The same year, I was accepted in the Master's Program in Statistics in the department of Statistics, Athens University of Economics and Business.

ABSTRACT

Georgios Sarris

A study of optimal sampling schemes

June 2009

This thesis comprises a research of optimal sampling designs. Firstly, I insert the theory of the standard sampling methods under finite population and I find the estimators of quantities of the population of these methods. Then, I insert the idea of the superpopulation approach. Optimal sampling methods under autocorrelated finite populations are presented in chapter 4. I analyse there, four optimal sampling designs. I present finally a short comparison of these optimal sampling schemes.

ΠΕΡΙΛΗΨΗ

Γεώργιος Σαρρής

Μελέτη βέλτιστων μεθόδων δειγματοληψίας

Ιούνιος 2009

Η διατριβή αποτελεί μία έρευνα βέλτιστων δειγματοληπτικών σχέδιων. Αρχικά εισάγω τη θεωρία των τυποποιημένων δειγματοληπτικών μεθόδων σε πεπερασμένο πληθυσμό και βρίσκω εκτιμητές ποσοτήτων του πληθυσμού από αυτές τις μεθόδους. Εν συνεχεία, εισάγω την ιδέα της προσέγγισης του υπερπληθυσμού. Βέλτιστοι δειγματοληπτικοί μέθοδοι σε αυτοσυσχετισμένους πεπερασμένους πληθυσμούς παρουσιάζονται στο κεφάλαιο 4. Εγώ αναλύω εκεί 4 βέλτιστα δειγματοληπτικά σχέδια. Στο τέλος παρουσιάζω μία σύγκριση αυτών των βέλτιστων δειγματοληπτικών σχεδίων.

TABLE OF CONTENTS

https://doi.org/10.26219/heal.aueb.5554

LIST OF TABLES

https://doi.org/10.26219/heal.aueb.5554

LIST OF FIGURES

1. CHAPTER

INTRODUCTION

In everyday life, we “meet” in media researches which are referred on a great spectrum of sciences like sociology, psychology, medicine, economics etc. However, the method of the completely census to implement a research (collection of all the information of the study) seems impracticable (cost, time).For this reason, the researchers apply sampling methods (part of population).

In chapter 2, we insert the theory of the standard sampling methods under finite population. We find the estimators of each sampling method of quantities of the population like the population total, the population mean, the ratio and the proportion. We also produce the estimators’ variance, a criterion for judging the produced estimator and the method of sampling. Finally, we refer to advantages and disadvantages of each sampling method.

In chapter 3, we insert the idea of the superpopulation approach, where the population vector ($y_1, y_2, \ldots, y_N$) is assumed to be a realization of a random unknown vector $(Y_1, Y_2, \ldots, Y_N)$ with a common distribution ξ.This approach has many advantages in relation with the approach of fixed population because we can make assumptions about this distribution. So, we can make use of a structure (it is rule and not an exception) like linear trend or autocorrelation among the units. We also refer the superpopulation inference for a model without clusters.

Sampling methods under autocorrelated finite populations are presented in chapter 4. We give optimal sampling schemes which provide us with the best estimators. Firstly, we refer to Blight’s model (1973). Here, the optimal sampling scheme is the centrally located systematic design. A generalization of the previous optimal sampling design is the model by I.Papageorgiou and K.X.Karakostas (1998). They consider with an autocorrelated finite population with an integer convex autocorrelation function $\rho(\cdot)$. Blight’s results hold and for this more general case. R.Mukerjee and S.Segupta (1989) prove that the optimal design is equivalent with a minimization problem for which they give a useful algorith. Finally, Chang-Tai Chao proposes two optimal designs which are based on the eigensystem of the population covariance matrix.

We finish our work in chapter 5 with a comparison of optimal designs which are presented in previous chapter. We suppose an autocorrelated finite population with an integer convex autocorrelation function. We compare the design of I.Papageorgiou and K.X.Karakostas (section 4.2) with the Design 1 of Chang-Tai-Chao (section 4.4). The comparison study is based on the efficiency of these sampling designs. The analytic numerical results are given from computational work with programming.

2. CHAPTER

FINITE POPULATION SAMPLING

Introduction

In this chapter, we develop the theory of the standard sampling methods. We will explore and present basic concepts for sampling from finite populations. We will refer the causes for which the sampling surveys are very important in practical every day problems and the scope of researcher. We will mention the estimators of quantities of the population like the population total, the population mean, the ratio and the proportion for every sampling method. We introduce the estimator's variance, a quantity associated with the estimator but also with the method of sampling and thus it can be used as a criterion for judging the produced estimator and the method of sampling. Finally, we make a reference about the advantages and the disadvantages of the sampling methods, something which comprises a comparison of these.

2.1. The idea of finite population sampling

2.1.1. Researches

In everyday life, we "meet" in newspapers, television and other media, results which are based on researches. These are referred on a great spectrum of applications like sociology, psychology, medicine, economics, political science, education, demography, husbandry etc.

Initially, one way to implement a research is to collect all the related with the subject of study information from a population. This procedure is the completely census. We take into account all the units of the population and then we proceed to analysis of this data (ex. to find the mean height of students in a class, we need the data with the height of all the students).

Nevertheless, if we consider that almost all the cases of researches concern thousands or millions units, the completely census seems impracticable. The major reasons are the cost and the time to collect and analyze the data. Moreover there is a need to repeat the same study occasionally and some times very often. Hence, it is necessarily to select a small part of the units with an appropriate way, where the collected units express the initial population. The researcher analyzes this part of units and generalizes the results for all the population. This part or subset of population is named sample, through which we implement the sampling research. Usually, we note the size of the sample as "n", while the size of all the units of the population is "N" (obviously $N \geq n$).

The principals and the methods for the collection and the analysis of the data from finite population is known as "Sample Survey Methods" or "Sampling". The analysis of sampling surveys is the most important tool for researches on a great spectrum of applications.

2.1.2. The concept of finite population and sample

The concept of population describes a group of people, animals, items or observations for which we are interested.

Element is defined every unit of a set, on which occurs a procedure of measurement or observation of a property. For example if we want to measure the height of footballers of a team, each footballer is an element.

Population is defined as an essential set of units which will be studied for one or more characteristics (set of elements).

Sampling units are collections of simple units like households, classes etc… The sampling unit may not coincide with an element .For example a household may have 3 or 4 members (elements). However, if every sampling unit contains one element of a population (household with a member), then the sampling unit and the element are the same.

Sampling frame is a list of elements which consists the population of research. There is list-frame like catalogue with number of telephones or a simple catalogue with names. There is area-frame like blocks of a town. Finally, there is notional-frame. For example, all the earthquakes which are occurred for a month in Greece.

Sample is a collection of sampling units from a frame.

2.1.3. Sampling researches

The sample is the approach which is selected for the sampling researches for the following basic reasons:

A. Reduced cost: If data are secured from only a small fraction of the aggregate, expenditures are smaller than if a complete census is attempted.

B. Greater speed: The data can be collected and summarized more quickly with a sample than with a complete count.

C. Greater accuracy: Because personnel of higher quality can be employed and given intensive training and because more careful supervision of the field work and processing of results becomes feasible when the volume of work is reduced, a sample may actually produce more accurate results than the kind of complete enumeration that can be taken.

D. Practical reasons: Sometimes the realization of a survey with complete enumeration may be impossible.

2.1.4. The Scope of Sampling

The scope of the researcher who realizes a sampling is:

A. The optimum choice of sample. A sample is optimum, if represents properly the population and provides estimator with great accuracy.

B. The statistical inference. Estimators for parameters of population, for which we are interested (population total for a characteristic, population mean for a characteristic, ratio of two characteristics of a population, population proportion for a characteristic), their variances, choice of appropriate sampling size etc…

2.1.5. Choice of the sample

The choice of the sample size and the selection of the units of the population which will be included in the sample are very important procedures in order to have representative sample of this population. By increasing the sample size, surely we have better accuracy to the estimators of the characteristics of the population. However, large sample size involves increase of the cost of the sampling and more intensive work. Small sampling size may lead to biased estimators. Though, we must notate that a careful choice of a small sample can give better results than a large one whose the units are not selected appropriate. The choice of the sample is affected from the below factors:

1) population size
2) variation of the studied population
3) cost of the survey
4) method of the choice of the sample
5) statistical error
6) confidence coefficient

For the choice of the sampling size, we take into account the margin error (d) of the variation of the estimator from the real estimated value and the confidence (confidence coefficient) $1-a$.

If θ is a parameter of the population and $\hat{\theta}$ is an estimator of it, then the margin error is defined as:

$$P[|\hat{\theta}-\theta|\leq d] \geq 1-a$$

Usually $a = 0.05$ or 0.10.

2.1.6. Errors of the Sampling

The major disadvantage of a sampling survey is that the estimators of the characteristics of the population may contain "errors of sampling". However we can reduce these errors with an appropriate method of the choice of the sample. There are sampling errors and non-sampling errors. Sampling errors can be occurred from the non-correct choice of the appropriate sampling method and its sample. Aside from the sampling error associated with the process of selecting a sample, a survey is subject to a wide variety of errors. These errors are commonly referred to as non-sampling errors. Non-sampling errors can be defined as errors arising during the course of all survey activities other than sampling. Unlike sampling errors, they can be present in both sample surveys and censuses.

2.1.7. Methods of the collection of the data

In a sampling survey, the information is collected with the help of questionnaires. The major ways which the researcher applies their survey to take the answers of the questionary is:

- Face to Face
- Telephone interview
- Postal interview
- Self completion

2.1.8. Estimation of parameters of population

The objective scope of sampling survey is the estimation of parameters of population from the information which is contained in the sample. The major parameters of population with great interest are the population total for a characteristic, the population mean for a characteristic, the ratio of two characteristics of a population, population proportion for a characteristic.

Let Y the characteristic which we study and N the number of population. $Y_1, Y_2, ..., Y_N$ are the values of Y for the units of the population.

Population total

$$\text{Y}=\sum_{i=1}^{N} Y_i$$

Example: The total of the incomings of citizens in Athens.

Population mean

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

Example: The mean of the height of students in a school.

Population proportion (for a characteristic)

$$P = \frac{A}{N}$$

A: The number of members of population with the characteristic.

Example: The proportion of students in a school who are smokers.

Ratio (for two characteristics of a population)

$$R = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i}$$

Example: The ratio of number of boys who like to play football towards these ones who like to play basketball.

Let θ is a characteristic of the population. We estimate θ from the sample, s, which we select, by estimator $\hat{\theta}(s)$. The major properties which lead to "good" estimators are the unbiasedness and the accuracy.

Unbiased is an estimator $\hat{\theta}$ when $E(\hat{\theta}) = \theta$. A measure of accuracy of an estimator is the variance, $Var(\hat{\theta}(s)) = E(\hat{\theta} - \theta)^2$. The smaller the variance is, more accurate the estimator is. If the estimator is not biased the quantity which determines the accuracy of the estimator is the MSE (mean square error).

$$MSE(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + bias^2(\hat{\theta})$$

2.2. SIMPLE RANDOM SAMPLING

2.2.1. Description

The most simple and essential sampling method is the simple random sampling (srs). When the size of the sample is n and the size of all the units of the population is N, the possible samples are $\binom{N}{n}$, assuming that the selection of units is without replacement. Simple random sampling is a method of selecting n units out of N such that every one of the possible samples has an equal chance to be chosen. The chance or in other words the possibility for every sample to be chosen is $\frac{1}{\binom{N}{n}}$.

2.2.2. Definition and notation

The values obtained for any specific item in the N units that comprise the population are denoted by $Y_1, Y_2, ..., Y_N$. The corresponding values for the units in the sample are denoted by $y_1, y_2, ..., y_n$. Note that the sample will not consist of the first n units in the population, except in the instance, usually rare, in which these units happen to be drawn.

Capital letters refer to characteristics of the population and lower case letters to those of the sample. In this chapter we will make use of the notation given in Table 2.1.

Table 2.1. Notation of population and sample quantities

	population	sample
Mean	$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$	$\overline{y} = \frac{1}{N}\sum_{i=1}^{n} y_i$
Total	$Y = \sum_{i=1}^{N} Y_i$	$y = \sum_{i=1}^{n} y_i$
Variance	$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$
Covariance	$S_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$	$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$

The random variable of y_i is denoted as "$y_i^{'}$". In simple random sampling we have that:

$$P[y_i^{'} = Y_k] = \frac{1}{N}, \quad i = 1,2,...,n \quad , \quad k = 1,2,...,N \tag{2.1}$$

In additional, we denote $f = \frac{n}{N}$, the sample fraction of the sampling.

Corollary 2.1. The correlation between the random variables $y_i^{'}$ and $y_j^{'}$, for $i \neq j$, is given from the following relation:

$$\rho(y_i^{'}, y_j^{'}) = -\frac{1}{N-1} \qquad (2.2)$$

2.2.3. Estimation of population mean

The population mean is a very important characteristic of the population. We will discuss about the unbiased estimator of the population mean, the variance of the estimation and the standard error of it.

Theorem 2.1.

i) The sample mean $\overline{y}$ is an unbiased estimate, with simple random sampling, of $\overline{Y}$.

$$E(\overline{y}) = \overline{Y} \quad . \qquad (2.3)$$

ii) The variance of the mean $\overline{y}$ from a simple random sampling is:

$$V(\overline{y}) = \frac{S^2}{n}(1-f), \quad f = \frac{n}{N} \text{ (sample fraction).} \qquad (2.4)$$

Proof:

i) The random variables $y_i^{'}$ take the values $Y_1, Y_2, ..., Y_N$ with the below probability:

$$P\,[y_i^{'} = Y_k] = \frac{1}{N}, \quad i = 1,2,...,n \quad ,k = 1,2,...,N$$

so

$$E(\overline{y}) = \frac{1}{n}\sum_{i=1}^{n} E(y_i^{'}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N}\sum_{k=1}^{N} Y_k = \frac{1}{n}n\overline{Y} = \overline{Y}$$

because we have

$$E(y_i^{'}) = \sum_{k=1}^{N} Y_k P[y_i^{'} = Y_k] = \frac{1}{N}\sum_{k=1}^{N} Y_k = \overline{Y} .$$

ii) We know from Corollary 2.1 that $\rho(y_i^{'}, y_j^{'}) = -\frac{1}{N-1}$.

Defining also

$$\sigma^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}(Y_i - \overline{Y})^2$$

, we calculate the variance

$$V(\overline{y}) = V(\frac{1}{n}\sum_{i=1}^{n} y_i^{'}) = \frac{1}{n^2}\left\{\sum_{i=1}^{n} V(y_i^{'}) + \sum_{i \neq j} Cov(y_i^{'}, y_j^{'})\right\} =$$

$$= \frac{1}{n^2}\left\{n\sigma^2 + n(n-1)\rho\sigma^2)\right\} = \frac{\sigma^2}{n}(1 - \frac{n-1}{N-1})$$

and we have

$$V(\overline{y}) = \frac{\sigma^2}{n}(\frac{N-n}{N-1}) .$$

We know that

$$\sigma^2 = \frac{N-1}{N} S^2 .$$

So if we replace the previous relation in the relation of $V(\overline{y})$, we finally have

$$V(\overline{y}) = \frac{S^2}{n}(1-f) .$$

The "f" is the **sample fraction** and the "$1-f$" is called as **finite population correction.**

If the size of the population N is large, the sample fraction is very small and it can be omitted.

Though the previous relation becomes

$$V(\overline{y}) = \frac{S^2}{n} \qquad (2.5)$$

https://doi.org/10.26219/heal.aueb.5554

Corollary 2.2. The standard error of $\bar{y}$ is

$$\sigma(\bar{y}) \;=\; \frac{S}{\sqrt{n}}\sqrt{1-f}\,. \tag{2.6}$$

Theorem 2.2. For a simple random sample s^2 is an unbiased estimate of S^2.

$$E(s^2) = \; S^2. \tag{2.7}$$

Proof:

We know the relations below

i) $E(y_i - \bar{Y})^2 = \sigma^2$,

ii) $\mathrm{E}(\bar{y} - \bar{Y})^2 = V(\bar{y}) = \;\dfrac{S^2}{n}(1-f)$,

iii) $E(y_i - \bar{Y})(\bar{y} - \bar{Y}) = Cov(y_i, \bar{y}) = Cov[y_i, (y_1 + y_2 + ... + y_n)\dfrac{1}{n}] =$

$\dfrac{1}{n}\{V(y_i) + \sum\limits_{i\neq j}\sum\limits_{i\neq j} Cov(y_i, y_j)\} = \dfrac{1}{n}\{\sigma^2 + (n-1)\rho\sigma^2\} =$

$\dfrac{\sigma^2}{n} - \dfrac{n-1}{n(N-1)}\sigma^2$,

iv) $\sigma^2 = \dfrac{N-1}{N}S^2$.

So we have,

$$E(s^2) = E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right\} =$$

$$\frac{1}{n-1}E\left\{\sum_{i=1}^{n}[y_i - \bar{Y} + \bar{Y} - \bar{y}]^2\right\} =$$

$$\frac{1}{n-1}E\left\{\sum_{i=1}^{n}(y_i - \bar{Y})^2 + n(\bar{Y} - \bar{y})^2 - 2\sum_{i=1}^{n}(y_i - \bar{Y})(\bar{y} - \bar{Y})\right\} =$$

$$\frac{1}{n-1}\left\{n\sigma^2 + (1-f)S^2 - 2\sigma^2 + \frac{2(n-1)}{N-1}\sigma^2\right\} =$$

$$\frac{1}{n-1}\left\{\frac{n(N-1)}{N}S^2+\frac{N-n}{N}S^2-\frac{2(N-1)}{N}S^2+\frac{2(n-1)}{N-1}\frac{N-1}{N}S^2\right\}=$$

$$\frac{S^2}{n-1}(\frac{nN-N}{N})=S^2$$

Corollary 2.3. An unbiased estimate of the variances of $\bar{y}$ is

$$\hat{V}(\bar{y}) \equiv s^2(\bar{y}) = \frac{s^2}{n}(1-f) \qquad (2.8)$$

where s^2 is an unbiased estimator of S^2.

Corollary 2.4. The standard error of $\bar{y}$ is

$$s(\bar{y}) = \frac{s}{\sqrt{n}}\sqrt{1-f} \qquad (2.9)$$

2.2.4. Estimation of population total

Another very important characteristic of the population is the population total Y. This is estimated by $\hat{Y} = N\bar{y}$, where $\bar{y}$ is the unbiased estimator of the population mean $\bar{Y}$.

Theorem 2.3. An unbiased estimate of the population total Y is $\hat{Y} = N\bar{y}$.

$$E(\hat{Y}) = Y \qquad (2.10)$$

with the following variance

$$V(\hat{Y}) = \frac{N^2S^2}{n}(1-f) \qquad (2.11)$$

Proof:

Straightforward consequence from the definition of $Y = N\bar{Y}$ and theorems 2.1, 2.2.

Corollary 2.5. The standard error of $\hat{Y}$ is

$$\sigma(\hat{Y}) = \frac{NS}{\sqrt{n}}\sqrt{1-f} \qquad (2.12)$$

When the value of the variance S^2 is unknown and is estimated from the sample variance s^2, we have the following corollary.

Corollary 2.6. An unbiased estimator of the $V(\hat{Y})$ is

$$V(\hat{\hat{Y}}) \equiv s^2(\hat{Y}) = \frac{N^2 S^2}{n}(1-f) \qquad (2.13)$$

The corresponding standard error is

$$s(\hat{Y}) = \frac{Ns}{\sqrt{n}}\sqrt{1-f} \qquad (2.14)$$

2.2.5. Estimation of ratio

Sometimes the ratio of two characteristics of the population is very important information. For example, the number (x) of the members of a household is the one characteristic when the other is the expenditures (y) of the household every week. So the ratio shows us the expenditure of each member of the household every week.

$$R = \frac{Y}{X} = \frac{\bar{y}}{\bar{x}}$$

The R is estimated by $\hat{R}$ where

$$\hat{R} = \frac{y}{x} = \frac{\bar{y}}{\bar{x}}$$

Theorem 2.4. In small samples the distribution of $\hat{R}$ is skew and $\hat{R}$ is usually a slightly biased estimate of R. If variates y_i, x_i are measured on each unit of a simple random sample of size n, assumed large, the bias of the estimation becomes negligible, so

$$E(\hat{R}) \approx R \tag{2.15}$$

with variance

$$V(\hat{R}) \approx \frac{1-f}{n\overline{X}^2} \sum_{i=1}^{N} \frac{(y_i - Rx_i)^2}{N-1} \tag{2.16}$$

(Proof is omitted, see Sampling Techniques (2nd edition) of William G. Cochran p.30)

Corollary 2.7. The variance $V(\hat{R})$ is estimated from

$$V\hat{(}\hat{R}) \equiv s^2(\hat{R}) = \frac{1-f}{n\overline{X}^2} \frac{\sum_{i=1}^{n}(y_i - \hat{R}x_i)^2}{n-1} \tag{2.17}$$

when $\overline{X}$ is unknown, it is estimated by $\overline{x}$.

Corollary 2.8. The standard error of $\hat{R}$ is estimated by $s(\hat{R})$, where for computational reasons it is written

$$s(\hat{R}) = \frac{\sqrt{1-f}}{\overline{x}\sqrt{n(n-1)}} \sqrt{\sum_{i-1}^{n} {y_i}^2 - 2\hat{R}\sum_{i=1}^{n} x_i y_i + \hat{R}^2 \sum_{i=1}^{n} {x_i}^2} \tag{2.18}$$

2.2.6. Estimation of proportion

In a sampling survey sometimes, it is very interesting to estimate the proportion of a population for a specific property.

Let

$$y_i = \begin{cases} 1, & \textit{if } i - \textit{unit has the specify property} \\ 0, & \textit{otherwise} \end{cases}$$

$A \equiv Y = \sum_{i=1}^{N} y_i$ = number of units of the population with the specify property

$a \equiv y = \sum_{i=1}^{n} y_i$ = number of units of the sample with the specify property

$\overline{Y} = \frac{A}{N} = P =$ proportion of the population with the specify property

$\overline{y} = \frac{a}{n} = p =$ proportion of the sample with the specify property

Theorem 2.5.

i) The sample proportion p is an unbiased estimator of P.

$$E(p) = P \qquad (2.19)$$

with variance

$$V(p) = \frac{S^2}{n}(1-f) \; = \; \frac{PQ}{n}(\frac{N-n}{N-1}) \qquad (2.20)$$

(Proof is omitted, see Damianou C., (1999) Sampling Methodology: Techniques and Applications, 3rd edition , Aithra: Athens, page 67)

Corollary 2.9. An unbiased estimator of $V(p)$ is

$$s^2(p) = \frac{pq}{n-1}(1-f) \qquad (2.21)$$

2.3. STRATIFIED RANDOM SAMPLING

2.3.1. Description

In stratified sampling the population of N units is first divided into L subpopulations of $N_1, N_2, ..., N_L$ units, respectively. These subpopulations comprise the whole population, so that

$$N_1 + N_2 + ... + N_L \; = N$$

The subpopulations are called strata. The sample sizes within the strata are denoted by $n_1, n_2, \ldots, n_L$ .If a simple random sample is taken in each stratum, the whole procedure is described as a stratified random sampling.

2.3.2. Notation

The suffix h denotes the stratum and i the unit within the stratum. The following symbols all refer to stratum h:

N_h, n_h: total number of units and number of units in sample correspondingly.

Y_{hi}: value obtained for the $i-th$ unit.

$W_h = \dfrac{N_h}{N}, w_h = \dfrac{n_h}{n}$: stratum weight in the population, stratum weight in the sample correspondingly.

$f_h = \dfrac{n_h}{N_h}$: sampling fraction in the stratum.

$\overline{Y}_h = \dfrac{\sum_{i=1}^{N_h} Y_{hi}}{N_h}, \overline{y}_h = \dfrac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$: true mean, sample mean correspondingly.

$S_h^2 = \dfrac{\sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2}{N_h - 1}, s_h^2 = \dfrac{\sum_{i=1}^{n_h} (y_{hi} - \overline{y}_h)^2}{n_h - 1}$: true variance, sample variance correspondingly.

2.3.3. Estimation of population mean

Theorem 2.6. The $\hat{\overline{Y}} \equiv \overline{y}_{st} = \sum_{h=1}^{L} W_h \overline{y}_h$ is an unbiased estimator of $\overline{Y}$ (the stratified random sampling is comprised of L strata with stratum weight W_h)

$$E(\overline{y}_{st}) = \overline{Y} \qquad (2.22)$$

with variance

$$V(\bar{y}_{st}) \;=\; \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h}(1-f_h) \qquad (2.23)$$

(Proof is omitted, see Sampling Techniques (2nd edition) of William G. Cochran p.89, 91)

Corollary 2.10. If $W_h = w_h$, this case is referred as proportional allocation. We substitute

$$n_h = \frac{nN_h}{N}$$

in (2.22).

The variance reduces to

$$V(\bar{y}_{prop}) \;=\; \frac{1-f}{n}\sum_{h=1}^{L} W_h {S_h}^2 \,. \qquad (2.24)$$

Finally, if the variances in all the strata have the same value, S_w^2, we obtain the simple result

$$V(\bar{y}_{prop}) \;=\; \frac{S_w^2}{n}(1-f)\,. \qquad (2.25)$$

Corollary 2.11. An unbiased estimator of $\bar{y}_{st}$ is

$$\hat{V}(\bar{y}_{st}) \;\equiv\; s^2(\bar{y}_{st}) \;=\; \sum_{h=1}^{L} W_h^2 \frac{s_h^2}{n_h}(1-f_h)\,. \qquad (2.26)$$

In the case of proportional allocation, an unbiased estimator of $V(\,\bar{y}_{prop})$ is

$$\hat{V}(\bar{y}_{prop}) \;=\; \frac{1-f}{n}\sum_{h=1}^{L} W_h s_h^2 \qquad (2.27)$$

2.3.4. Estimation of population total

We know that $Y = N\bar{Y}$, so

$$\hat{Y}_{st} = N\hat{\bar{Y}} = N\bar{y}_{st}$$

Theorem 2.6. An unbiased estimator of Y is

$$E(\hat{Y}_{st}) = Y \tag{2.28}$$

with variance

$$V(\hat{Y}_{st}) = \sum_{h=1}^{L} N_h^2 \frac{S_h^2}{n_h}(1-f_h) \tag{2.29}$$

Proof:

Straightforward consequence from the definition of $\hat{Y}_{st} = N\hat{\bar{Y}} = N\bar{y}_{st}$ and theorem 2.6**.**

Corollary 2.12. An unbiased estimator of $V(\hat{Y}_{st})$ is

$$\hat{V}(\hat{Y}_{st}) \equiv s^2(\hat{Y}_{st}) = \sum_{h=1}^{L} N_h^2 \frac{s_h^2}{n_h}(1-f_h) \tag{2.30}$$

2.3.5. Estimation of proportion

We define the below quantities:

A_h: number of units of the population in strata h with the specify property.

a_h: number of units of the sample in strata h with the specify property.

P_h: true proportion of the population in strata h with the specify property.

p_h: sample proportion of the population in strata h with the specify property.

We know that

$$\hat{P} \equiv p_h = \frac{a_h}{n_h}$$

and that

$$A = \sum_{h=1}^{L} A_h \ = \ \sum_{h=1}^{L} N_h P_h$$

which is estimated by

$$\hat{A}_{st} = \sum_{h=1}^{L} N_h \hat{P}_h \ = \ \sum_{h=1}^{L} N_h p_h \ .$$

So, P is estimated by

$$\hat{P} \ = p_{st} = \frac{\hat{A}}{N} \ = \ \sum_{h=1}^{L} W_h p_h$$

Theorem 2.7. An unbiased estimator of P is p_{st}

$$E(\ p_{st}\ ) \ = P \qquad (2.31)$$

with variance

$$V(\ p_{st}\ ) \ = \ \sum_{h=1}^{L} W_h^2 \frac{P_h Q_h}{n_h} \frac{(N_h - n_h)}{N_h - 1} \qquad (2.32)$$

Proof:

If we substitute the above relation for p_{st} it is straightforward that

$$E(\ p_{st}\ ) \ = P \ .$$

Now for the variance, from the definition of the proportion

$$S_h^2 \ = \ \frac{N_h}{N_h - 1} P_h Q_h$$

and the variance of p_h

$$V(\ p_h\ ) \ = \ \frac{N_h - n_h}{N_h - 1} \frac{P_h Q_h}{n_h} .$$

So from the relation of p_{st} the (2.31) is straightforward.

Corollary 2.13. In the case of proportional allocation, where $n_h = \mathrm{n} W_h$,

$$V(\ p_{st}\ ) \ = \ \frac{1-f}{n}\sum_{h=1}^{L} W_h P_h Q_h \qquad (2.33)$$

Corollary 2.14. An unbiased estimator of $V(p_{st})$, when P_h is unknown, is

$$V\hat{(}p_{st}) \equiv \ s^2(p_{st}) \ = \ \sum_{h=1}^{L} W_h \frac{p_h q_h}{n_h - 1}(1-f_h) \qquad (2.34)$$

In the case of proportional allocation, an estimator of $V(p_{prop})$, is

$$\hat{V}(p_{prop}) \ \equiv \ s^2\,(\,p_{prop}) \ \approx \ \frac{1}{n}\sum_{h=1}^{L} W_h p_h q_h \qquad (2.35)$$

2.3.6. Optimum Allocation

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

The simplest cost function is of the form

$$\text{Cost } = C = \ c_0 + \sum_{h=1}^{L} c_h n_h$$

Theorem 2.8. In stratified random sampling with a cost of the previous form, the variance of the estimated mean $\bar{y}_{st}$ is minimum when

$$n_h \ = \ n\frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h / \sqrt{c_h}} \text{ (optimum allocation)}$$

and if $c_h \ = c, \quad h = 1,\ldots,L$ then

$$n_h \ = \ n\frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h} \qquad \text{(Neyman)} \qquad (2.36)$$

(Proof is omitted, see Sampling Techniques (2nd edition) of William G. Cochran p.95-96 using the calculus method of Langrange multipliers).

A formula for the minimum variance is obtained by substituting the value of n_h (Neyman) into the general formula for $V(\bar{y}_{st})$.

$$V_{\min}(\bar{y}_{st}) = \frac{1}{n}(\sum_{h=1}^{L} W_h S_h)^2 - \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2$$

Theorem 2.9. In stratified random sampling with a specified value of the variance

$$V(\bar{y}_{st}) = V_0 = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h}(1 - f_h),$$

the cost is minimum when

$$n_h = n\frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h / \sqrt{c_h}}.$$

If $c_h = c, \quad h = 1,\ldots,L$ then

$$n_h = n\frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h}.$$

(Proof is omitted, see Damianou C., (1999) Sampling Methodology: Techniques and Applications, 3^{rd} edition, Aithra: Athens, page 143-144)

2.3.7. Comparison of simple random, proportional and optimum allocation of n

Theorem 2.10. If terms in $f_h = \dfrac{n_h}{N_h}$ are ignored,

$$V_{opt} \le V_{prop} \le V_{ran}$$

Where the optimum allocation is for fixed n, that is, with $n_h \propto N_h S_h$.

Proof:

We know that,

$$V_{ran} = \frac{S^2}{n}, \quad V_{prop} = \frac{1}{nN}\sum_{h=1}^{L} N_h S_h^2, \quad V_{opt} = \frac{1}{nN^2}(\sum_{h=1}^{L} N_h S_h)^2.$$

From the definition of S^2,

$$V_{ran} = \frac{S^2}{n} = \frac{1}{nN}\sum_{h=1}^{L} N_h S_h^2 + \frac{1}{nN}\sum_{h=1}^{L} N_h (\overline{Y}_h - \overline{Y})^2$$

$$V_{ran} = V_{prop} + \frac{1}{nN}\sum_{h=1}^{L} N_h (\overline{Y}_h - \overline{Y})^2$$

we also have that,

$$V_{prop} - V_{opt} = \frac{1}{nN}\sum_{h=1}^{L} N_h (S_h - \overline{S})^2$$

so,

$$V_{ran} = V_{opt} + \frac{1}{nN}\sum_{h=1}^{L} N_h (S_h - \overline{S})^2 + \frac{1}{nN}\sum_{h=1}^{L} N_h (\overline{Y}_h - \overline{Y})^2$$

$$V_{opt} \leq V_{prop} \leq V_{ran}$$

In stratified sampling, the variance of the estimator arises from optimum allocation of n, reduces in terms of the variance of simple random allocation as much as:

i) The variation between the means of the strata increases

ii) The inhomogeneity in the strata increases

2.4. Systematic Sampling

2.4.1. Description

This method of sampling is at first sight quite different from simple random sampling. The basic features of systematic sampling are:

a) the sampling interval k

We distinguish now two occasions:

i) If N is a multiple of n, let $k = \dfrac{N}{n}$

ii) If N is not a multiple of n, let $k = \left[\dfrac{N}{n}\right] + 1$

b) a random start r from the first k units

We must stress that if the value of sampling interval takes the value $\left[\dfrac{N}{n}\right] + 1$, there will be samples with size smaller than n. To avoid this problem we complete these samples with units, by starting from the beginning of the population.

Suppose that the N units in the population are numbered 1 to N in some order. To select a sample of units, we take a unit (r) at random from the first k units and every $k-th$ unit thereafter. So the sample is comprised by the subsequent units in the population with numbers

$$r, r+k, \ldots, r+(n-1)k$$

For instant, if k is 15 and if the first unit drawn is number 13, the subsequent units are numbers 28, 43, 58 and so on. The selection of the first unit determines the whole sample. This type is called an every $k-th$ systematic sample.

2.4.2. Notation

If r is the random start and k is the sampling interval, the sample is

$y_r, y_{r+k}, \ldots, y_{r+(n-1)k}$. There are k different possible systematic samples as many as the possible values of r.

Let $\bar{y}_{r.}$, $S_{r.}^2$ are the mean and the variance of the r-sample (sample with random start r)

$$\bar{y}_{r.} = \frac{1}{n}\sum_{j=1}^{n} y_{r+(j-1)k}$$

$$S_{r.}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_{r+(j-1)k} - \bar{y}_{r.})^2$$

Let S_{wsy}^2 is the variance among units that lie within the same systematic sample

$$S_{wsy}^2 = \frac{1}{k(n-1)}\sum_{r=1}^{k}\sum_{j=1}^{n}(y_{rj} - \bar{y}_{r.})^2 = \frac{1}{k}\sum_{r=1}^{k}S_{r.}^2$$

2.4.3. Estimation of population mean

Let $\bar{y}_{sy} = \frac{1}{n}\sum_{j=1}^{n} y_{r+(j-1)k}$ be the mean of a random systematic sample. The $\bar{y}_{sy}$ is an estimator of population total. Here, we must distinguish two occasions to examine the property of unbiasedness.

- If N is a multiple of n, the $\bar{y}_{sy}$ is an unbiased estimator of $\bar{Y}$.

- If N is not a multiple of n, we follow the next two ways to provide unbiasedness to the estimator.

a) We adjust the probabilities of selection for each of the systematic samples, according to the number of units that contain.

b) We complete the samples with smaller size, by starting from the beginning of the population.

For example, if $N = 10$ and $n = 3$, then $k = \left[\frac{N}{n}\right] + 1$ because N is not a multiple of n, so $k = 4$. Thus, with $N = 10$ and $k = 4$ the numbers of the units in the four systematic samples are shown in the following Table 2.2

Table 2.2. Systematic samples (1st way)

I	**II**	**III**	**IV**
1	2	3	4
5	6	7	8
9	10		

With the first way, if a probability of selection $\frac{3}{10}$ is given to each of the first two samples and a probability $\frac{2}{10}$ to each of the last two, the sample mean is unbiased.

With the second way, we complete the last two samples, with units, by starting from the beginning of the population. So, the numbers of the units in the four systematic samples are shown in the Table 2.3.

Table 2.3. Systematic samples (2nd way)

I	II	III	IV
1	2	3	4
5	6	7	8
9	10	1	2

Theorem 2.11. The $\hat{\bar{Y}} \equiv \bar{y}_{sy} = \frac{1}{n}\sum_{j=1}^{n} y_{r+(j-1)k}$ is an unbiased estimator of the population mean

$$E(\bar{y}_{sy}) = \bar{Y} \tag{2.37}$$

with variance

$$V(\bar{y}_{sy}) = \frac{N-1}{N}S^2 - \frac{k(n-1)}{N}S^2_{wsy} \tag{2.38}$$

(Proof is omitted, see Damianou C., (1999) Sampling Methodology: Techniques and Applications, 3^{rd} edition , Aithra: Athens, page.192).

To compare the sampling schemes discussed in this chapter, we distinguish three types of population structure and compare the sampling schemes with respect to the variances of the estimators they provide.

a) For random population, where the units in the population are in "random" order, without specific trend or structure, we have

$$V(\bar{y}_{sy}) = V(\ \bar{y}_{ran})$$

b) For ordered population, where the units are ordered with respect to a characteristic, we have

$$V(\bar{y}_{sy})\ \leq\ V(\bar{y}_{ran})$$

c) For periodic population, where the units have a circle rotation, we have

$$V(\bar{y}_{sy})\ \geq\ V(\bar{y}_{ran})$$

Corollary 2.15. If the systematic sampling is equivalent to the simple random sampling (random list without periodicity or order) an estimation of $V(\bar{y}_{sy}))$ is

$$\hat{V(\bar{y}_{sy})}\ \equiv\ s^2(\bar{y}_{sy}) = \frac{s^2}{n}(1-f) \qquad (2.39)$$

2.4.4. Estimation of population total Y, ratio R, proportion P

Supposing we have random population and we apply Theorem 2.3, 2.4, 2.5 correspondingly for the estimator $\bar{y}_{sy}$.

2.5. Cluster Sampling

2.5.1. Description

Several references have been made in books (Damianou C., (1999) Sampling Methodology: Techniques and Applications, 3^{rd} edition , Aithra: Athens, Sampling Methods of William G. Cochran etc.) to surveys in which the sampling unit consists of a group or cluster of smaller units that we have called elements. There are two main reasons for the widespread application of cluster sampling. Although the first intension may be to use the elements as sampling units, it is found in many surveys that no reliable list of elements in the population is available and that it would be prohibitively expensive to construct such a list. For example a region from maps can be divided into area units such as blocks in the cities. These clusters are often chosen to solve the problem of constructing a list of sampling units.

The population in cluster sampling is divided into groups-clusters and this seems to be similar with the method of stratified sampling. However, there are significant differences. In stratified sampling the elements within strata should be as homogeneous as possible and the strata in a whole should be as heterogeneous as possible. In cluster sampling the elements within clusters should be as heterogeneous as possible and the clusters in a whole should be as homogeneous as possible. Moreover, the mechanism of selecting units differs in these two cases. Details are given in the sequel of this paragraph.

Simple one-stage cluster sampling is a sampling plan in which clusters are chosen by simple random sampling and, within each sample cluster, all listing units are selected. Simple two-stage cluster sampling is a sampling plan in which clusters are selected at the first stage by simple random sampling and, within each sample cluster, listing units are selected at the second stage by simple random sampling. Finally, simple multi-stage cluster sampling involves more than two stages for the selection of the sample. In this chapter we will present the sampling design of simple one-stage cluster sampling.

2.5.2. Notation

N : number of clusters in the population

n : number of clusters in the sample

M_i : number of units in cluster $i, \quad i = 1, \ldots, N$

Y_{ji} : the $j-unit$ of the $i-cluster$

$\overline{m}$: the sample mean size of the clusters,

$$\overline{m} = \frac{1}{n} \sum_{i=1}^{n} m_i$$

M : the number of the units in the population,

$$M = \sum_{i=1}^{N} M_i$$

$\overline{M}$: the population mean size of the clusters,

$$\overline{M} = \frac{M}{N}$$

Y_i : the total of the characteristic Y in i-cluster,

$$Y_i = \sum_{j=1}^{m_i} Y_{ji}$$

$\overline{Y}$: the population total,

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{M}$$

Y : the population total,

$$Y = M\overline{Y} = \sum_{i=1}^{N} Y_i$$

$\overline{y}$: the sample mean,

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$$

A : the number of units in the population with the specify property

P : the proportion of the population with the specify property,

$$P = \frac{A}{M} = \frac{\sum_{i=1}^{N} A_i}{\sum_{i=1}^{M} M_i}$$

a_i : the number of units with the specify property in cluster i

p : the sample proportion with the specify property,

$$p = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

2.5.3. Estimation of population mean $\overline{Y}$ and population total Y

The population mean $\overline{Y}$ is estimated by $\overline{y}_C$, so

$$\hat{\overline{Y}}_C \equiv \overline{y}_C = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} \tag{2.40}$$

The population total Y is estimated by $M\overline{y}_C$ if M is known, so

$$\hat{Y}_C \equiv M\overline{y}_C = M\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} \tag{2.41}$$

The estimators $\hat{\overline{Y}}_C$ and $\hat{Y}_C$ of $\overline{Y}$ and Y correspondingly are not unbiased, like the ratio estimators.

Corollary 2.16. The variance of $\overline{y}_C$ is approximately

$$V(\overline{y}_C) \approx \frac{1-f}{n\overline{M}^2} \frac{\sum_{i=1}^{N} (Y_i - \overline{Y}m_i)^2}{N-1} \tag{2.42}$$

and is estimated by

$$\hat{V}(\bar{y}_C) \;\equiv\; s^2(\bar{y}_C) \;\approx\; \frac{1-f}{n\bar{M}^2}\,\frac{\sum_{i=1}^{N}(Y_i - \bar{y}_C M_i)^2}{n-1} \qquad (2.43)$$

(If $\bar{M}$ is unknown, we estimate it by $\bar{m}$)

Corollary 2.17. The variance $V(\hat{Y}_C) \;=\; V(M\bar{y}_C)$ is estimated by

$$\hat{V}(\hat{Y}_C) \;\equiv\; s^2(\hat{Y}_C)\; M^2 s^2(\bar{y}_C) \;=\; \frac{1-f}{n}\,\frac{\sum_{i=1}^{n}(y_i - \bar{y}_C m_i)^2}{n-1} \qquad (2.44)$$

Generally, these estimators are biased and provide good estimation for n$\geq$20.

2.5.4. Estimation of proportion

For the estimation of the proportion P of units in a population with a specify property, we have the following estimator

$$\hat{P}_C \;=\; p_C \;=\; \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} \qquad (2.45)$$

a_i is the number of units in i-cluster with the specify property

(The estimator $\hat{P}_C$ is not unbiased)

Corollary 2.18. The variance of $\hat{P}_C$ is approximately

$$V(\hat{P}_C) \;\approx\; \frac{1-f}{n\bar{M}^2}\,\frac{\sum_{i=1}^{N}(A_i - PM_i)^2}{N-1} \qquad (2.46)$$

and is estimated by

$$\hat{V}(\hat{P}_C) \;\equiv\; s^2(p_C) \;=\; \frac{1-f}{n(n-1)\bar{M}^2}\sum_{i=1}^{n}(a_i - p_C m_i)^2 \qquad (2.47)$$

2.6. Comparison of sampling methods

In the previous paragraphs, we gave a short description about the sampling methods. Now we want to compare these methods, so we will mention the advantages and the disadvantages of each one. This reference is towards the direction to aid in the critical point of the choice of the appropriate method in a sampling survey. Usually a combination of methods is chosen.

2.6.1. Simple Random Sampling

Advantages

- Provide good estimators for characteristics of the population, with low cost.
- The estimators of characteristics (like mean, total, proportion) are unbiased estimators of the parameters with unbiased estimators of their variances.
- This method is implemented in a later stage of other methods. It is fundamental method.

Disadvantages

- Provides estimators with large variance in relation with the estimator's variances of the other methods. The disadvantages of the simple random sampling are referred during the description of the advantages of the other methods.

2.6.2. Stratified Random Sampling

Advantages

- Better estimation in relation of the simple random sampling because the variance in strata is smaller than in the whole population.

- Representation of certain groups in the population.
- Because this method succeeds small variance, it is required smaller sample size, so lower cost.
- Provides separate estimation of the characteristics for each strata.
- Provides unbiased estimators.
- Every strata contains smaller size of units, so the selection of the sample becomes an easy procedure. We can also work with separate groups of researchers, every one for every strata.

Disadvantages

- Difficulties to the right separation of the population into strata and the strata's definition.
- If the strata do not appear inhomogeneity for the characteristics, the variance will be the same with this in simple random sampling.

2.6.3. Systematic Sampling

Advantages

- The selection of the sample is very easy for the interviewer.
- Very useful and applicable to cases with large population size.
- Provides unbiased estimators if we follow right steps.

Disadvantages

- In the case of periodic population, the variance of the estimators is increased. It could be more than the corresponding in simple random sampling.
- If we do not follow right steps, which we referred in the paragraph of the systematic sampling, this method may give unbiased estimators.
- When the size of the population is unknown we can not calculate with accuracy the sample interval k. So the samples may be smaller if k is overestimated.

2.6.4. Cluster Sampling

Advantages

- Reduced lower cost when the frame of the units of the population does not appear or it is constructed with great cost.
- In the case where the cost of the collection of the information increases as much as the distance between the individuals in the population increases, the cluster sampling ensures lower cost in relation with the other methods.

Disadvantages

- In contrast with the stratified sampling, in cluster sampling the elements within clusters should be as heterogeneous as possible and the clusters in a whole should be as homogeneous as possible.
- Generally provides biased estimators.

3. CHAPTER

THE IDEA OF THE SUPERPOPULATION APPROACH-INFERENCE FOR SUPERPOPULATION PARAMETERS

Introduction

In the following chapter, we adopt the superpopulation approach in sampling from the finite population. We develop the basic concepts of the superpopulation. Then, we make inference for superpopulation parameters like superpopulation mean. In this section, the superpopulation inference is referred, at first instance, for population models without clusters. Later we make inference for a model that accommodates multiple levels (stages) of clustering in the population. The model we will develop comprises only two levels but similarly the methodology works for more than two levels. In the inference, we mention properties of estimators by incorporating both the randomness due to the sampling of the population as well as the randomness due to the generation of the population by a model. We find estimators for the superpopulation mean as well as its variance.

3.1. The idea of the superpopulation approach

3.1.1. Definition and notation

In the previous chapter, we described sampling from finite population, an approach which is called in the literature as sampling under fixed population. In contrast, in the following chapter we will develop the idea of sampling from superpopulation (superpopulation approach).

According to the superpopulation approach the population vector $(y_1, y_2, \ldots, y_N)$ is assumed to be a realization of a random unknown vector $(Y_1, Y_2, \ldots, Y_N)$. In the following, we use the symbol ξ, for the common distribution of $(Y_1, Y_2, \ldots, Y_N)$.

This approach has many advantages in relation with the approach of fixed population. The main advantage of the superpopulation approach, in contrast to the classical one, is that it allows one to make assumptions about the distribution ξ. So, we can make use of a structure like linear trend or autocorrelation among the units that may appear in the population. The existence of structure in the population is quite common in practice. Especially in sciences like economy, agriculture, demography, industrial (quality control) etc. That makes the superpopulation approach appropriate for substantial.

In the simple case of the superpopulation approach, we assume that there is no correlation among the variables $Y_1, Y_2, ..., Y_N$. In that way, the population $(Y_1, Y_2, \ldots, Y_N)$ can be considered as a sample of size N from a hypothetical infinite population. We adopt the idea of superpopulation approach, to have the ability to make assumptions for the common distribution ξ of $Y_1, Y_2, \ldots, Y_N$ and especially for their correlation.

In practice, we assume the population vector $(y_1, y_2, \ldots, y_N)$ as a realization of a random unknown vector $(Y_1, Y_2, \ldots, Y_N)$ with common distribution ξ, for which we suppose that general characteristics (like first and second-order moments) are known. Frequently, instead to the term of superpopulation, we call the above form superpopulation model. This happens, when we refer to the conditions which corresponds to Y_i $(i = 1, 2, ..., N)$, which define a specify class of distributions. With the term conditions, we mean constrains or specific structure on the moments, means, variances etc.

In more mathematical terms, the superpopulation model in its general form assumes that

$$y_i = \mu_i + \varepsilon_i, \quad (i = 1, 2, ..., N) \tag{3.1}$$

where μ_i is the deterministic part and ε_i the random. For random vector $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_{\mathrm{N}})$, we assume that it has zero mean $(E_\xi(\varepsilon_i) = 0)$ and a positive-definite covariance matrix.

For example we refer to a very important superpopulation model, Cochran's model (1946), where

$$E_\xi(y_i) = \mu$$

$$E_{\xi}(y_i-\mu)(y_j-\mu)=E_{\xi}(\varepsilon_i\,\varepsilon_j)=\sigma^2\rho(|i-j|),\quad i\neq j=1,\ldots,N\ and\ \rho(0)=1),$$

where μ,σ^2 are unknown superpopulation parameters of the model and $\rho(\cdot)$ is the autocorrelation function. This is model (3.1) for $\mu_i=\mu\quad(i=1,2,\ldots,N)$ and $V_{ij}=\sigma^2\rho(|i-j|)$.

3.1.2. Necessity of the idea of superpopulation

Sample survey inference is historically concerned with finite-population parameters, that is, functions like means and totals of the observations for the individuals in the population. However, in many scientific applications, interest usually focuses on the superpopulation parameters associated with a stochastic mechanism hypothesized to generate the observations in the population rather than the finite-population parameters.

In classical sampling theory, the target of inference is finite-population parameters like the mean $\bar{Y}$ of the N unit values in the population. A stochastic model for the finite-population values is sometimes used to evaluate and suggest sample designs and estimators; see, for example, Cochran(1939, 1946). However, in scientific applications, the parameters associated with the stochastic model are of more interest than the finite-population parameters. Deming(1953) refers to inference for superpopulation parameters as an “analytic” use of survey data.

3.2. Model without cluster

3.2.1. Description and notation

In this section a superpopulation model without clusters is considered. Here we will consider the properties of estimators that incorporate both the randomness due to the sampling of the population and the randomness due to the generation of the population by a model. Letting the subscript RS refer to the (repeated) sampling randomness, the subscript F refer to the model randomness, and no subscript refer to both sources of randomness. So, the usual decompositions for the expectation and the variance of an estimator $\hat{\theta}$ that will be used throughout are:

$$E(\hat{\theta}) = E_F[E_{RS}(\hat{\theta})], \tag{3.2}$$

$$Var(\hat{\theta}) = E_F[Var_{RS}(\hat{\theta})] + Var_F[E_{RS}(\hat{\theta})] \tag{3.3}$$

The population values are $(Y_1, \eta_1),\ldots,(Y_K, \eta_K)$, where η is a stratum indicator with range $\{1,2,...,L\}$. We assume that the (Y_i, η_i) are independent and identically distributed, each with the same distribution as the random vector (Y, η), which has the bivariate distribution function F. We restrict our attention to stratified simple random sampling.

3.2.2. Estimating a superpopulation mean

The target parameter is $\mu = E_F(Y)$, which is to be estimated using stratified simple random sampling without replacement. Let K_h be the known number of observations in the $h-th$ stratum in the finite population, $h = 1,2,...,L$. Let k_h be the number of sampled observations in the $h-th$ stratum. The total number of observations in the sample, $k = k_1 + k_2 + ... + k_L$.

The stratified mean is

$$\bar{y} = \sum_{h=1}^{L} \frac{K_h}{K} \bar{y}_h, \tag{3.4}$$

where $\bar{y}_h$ is the mean of the sampled observations in stratum h. The stratified mean is an unbiased estimator of the population mean $\bar{Y}$, under repeated sampling of the same finite population. Letting the subscript wo, refer to the repeated sampling variance estimator treating the sample as if it had been a without-replacement sample. The repeated-sampling variance estimator of $\bar{y}$ is

$$\hat{\mathrm{var}}_{wo}(\bar{y}) = \sum_{h=1}^{L} \frac{K_h^2}{K^2} \frac{K_h - k_h}{K_h} \frac{s_h^2}{k_h}$$

where s_h^2 is the sample variance of the observations in the $h-th$ stratum. Under repeated sampling of the same finite population, $\hat{\mathrm{var}}_{wo}(\bar{y})$ is an unbiased estimator of the variance of $\bar{y}$.

However, it will be useful to consider the previous estimator treating the sample with replacement. The corresponding subscript to this estimator is wr. So, we have

$$\hat{\mathrm{var}}_{wr}(\bar{y}) = \sum_{h=1}^{L} \frac{K_h^2}{K^2} \frac{s_h^2}{k_h}$$

If we incorporate the randomness due to the sampling and the distribution function F, we have that $\bar{y}$ is unbiased for μ.

$$E(\ \bar{y}\ ) = E_F[E_{RS}(\bar{y})] = \mu \qquad (3.5)$$

where $\bar{y} = \sum_{h=1}^{L} \frac{K_h}{K} \bar{y}_h$.

3.2.3. The variance of the estimator

$$Var(\ \bar{y}\ ) = E_F[Var_{RS}(\bar{y})] + Var[E_{RS}(\bar{y})] =$$

$$= E_F\left[\sum_{h=1}^{L} \frac{K_h^2}{K^2} \frac{K_h - k_h}{K_h} \frac{S_h^2}{k_h}\right] + Var_F\left[\bar{Y}\right] =$$

$$= \frac{1}{K^2}\sum_{h=1}^{L} \sigma_h^2 \mathrm{E}_F\left[K_h \frac{K_h - k_h}{k_h}\right] + \frac{1}{K^2}\sum_{h=1}^{L} \sigma_h^2 \mathrm{E}_F(K_h) + \frac{1}{K^2} Var_F(\sum_{h=1}^{L} K_h \mu_h) =$$

$$= \frac{1}{K^2}\sum_{h=1}^{L} \sigma_h^2 \mathrm{E}_F\left[\frac{K_h^2}{k_h}\right] + \frac{1}{K}\Delta_{betw}$$

where $\Delta_{betw} = \sum_{h=1}^{L} \pi_h\,(\mu_h - \mu)^2$ is the between-strata variability of the μ_h, $\pi_h \equiv E(K_h / K) = P(\eta = h)$, S_h^2 is the variance of the population values in the $h-th$ stratum, μ_h and s_h^2 are the mean and the variance of Y in the $h-th$ stratum with respect to the F distribution. We note also that it is straightforward that:

$$\mu = \sum_{h=1}^{L} \pi_h \mu_h$$

We obtain an unbiased variance estimator of $\bar{y}$, if we add an unbiased estimator of $Var_F(\bar{Y})$ to $\hat{\text{var}}_{wo}(\bar{y})$ or equivalently, add an unbiased estimator of $\frac{1}{K}\Delta_{betw}$ to $\hat{\text{var}}_{wr}(\bar{y})$.

We have the following equations:

$$V\hat{a}r(\bar{Y}) = \frac{1}{K-1}\sum_{h=1}^{L}\frac{K_h}{K}(\bar{y}_h - \bar{y})^2 + \frac{1}{K}\sum_{h=1}^{L}\frac{K_h}{K}\left[1 - \frac{K-K_h}{k_h(K-1)}\right]s_h^2$$

$$\hat{\Delta}_{betw} = \frac{K}{K-1}\sum_{h=1}^{L}\frac{K_h}{K}(\bar{y}_h - \bar{y})^2 - \sum_{h=1}^{L}\frac{K_h(K-K_h)}{K(K-1)}\frac{s_h^2}{k_h}$$

$$E(V\hat{a}r(\bar{Y}) = E_F(\frac{1}{K-1}\sum_{h=1}^{L}\frac{K_h}{K}(\bar{Y}_h - \bar{Y})^2) + E_F(\frac{1}{K}\sum_{h=1}^{L}\frac{K_h - 1}{K-1}S_h^2) = Var_F(\bar{Y})$$

Therefore,

$$V\hat{a}r_{SP}(\bar{y}) = \hat{\text{var}}_{wo}(\bar{y}) + V\hat{a}r(\bar{Y}) \qquad (3.6)$$

or equivalently,

$$V\hat{a}r_{SP}(\bar{y}) = \hat{\text{var}}_{wr}(\bar{y}) + \frac{1}{K}\hat{\Delta}_{betw} \qquad (3.7)$$

is an unbiased estimator for $Var(\bar{y})$.

3.3. Two-stage Model with clusters

3.3.1. Description and notation

In the following section, we consider a model that accommodates multiple levels of clustering in the population. The model we will develop, comprises only two levels. The population consists of K primary clusters, the $i-th$ of which consists of N_i secondary clusters. In the $j-th$ secondary cluster of the $i-th$ primary cluster, there are M_{ij} population values with total $T_{ij} = Y_{ij1} + ... + Y_{ijM_{ij}}$. The $i-th$ primary cluster has associated with it a stratum variable $\eta_i \in \{1, 2, ..., L\}$ and a size variable Z_i which will be used for probability-proportional-to-size sampling. We assume that

$\{(M_{ij}, T_{ij})\}$, (note that $j = 1, 2, ..., N_i$), are independent and identically distributed with mean (a_i, r_i) and variances-covariances $(\sigma_{11i}, \sigma_{22i}, \sigma_{12i})$ and that the $(a_i, r_i, \sigma_{11i}, \sigma_{22i}, \sigma_{12i}, N_i, Z_i, \eta_i)$ are independent and identically distributed from an eight-dimensional random variable with distribution function G.

In this section we describe the multistage (two-stage) sampling considered and the notation for the estimation of the superpopulation mean using a weighted mean. At the first stage of sampling, k_h primary clusters are sampled from the K_h primary clusters in stratum h as a probability-proportional-to-size sample without replacement. That is the primary clusters are the primary sampling units(PSU's) and the inclusion probability of a given PSU in stratum h is proportional to its Z value. At the second stage of sampling, we assume that n_{hi} secondary clusters are sampled with replacement from $i-th$ sampled PSU from stratum h. We do assume that the sample weights (inverses of the inclusion probabilities) are known for all sampled observations.

3.3.2. Estimating a superpopulation mean

Let ,

$$t_{hij} = \sum_{l=1}^{m_{hij}} w_{hijl} y_{hijl} , \qquad d_{hij} = \sum_{l=1}^{m_{hij}} w_{hijl}$$

where y_{hijl} and w_{hijl} are the population value and the sample weight of the $i-th$ sampled observation in the $j-th$ sampled secondary cluster in the $i-th$ sampled PSU of stratum h, and m_{hij} is the number of sampled observations in that secondary cluster. Letting

$$t_{hi} = \sum_{j=1}^{n_{hi}} t_{hij} , \qquad d_{hi} = \sum_{j=1}^{n_{hi}} d_{hij}$$

So, we have under repeated sampling of the population

$$t = \sum_{h=1}^{L} \sum_{i=1}^{k_h} t_{hi} , \qquad d = \sum_{h=1}^{L} \sum_{i=1}^{k_h} d_{hi}$$

which are approximately unbiased estimators of the total of Y and the population size, respectively .

The weighted estimator of $\overline{Y}$ and μ is $\overline{y} = \frac{t}{d}$

$$E(\overline{y}) = E(\frac{t}{d}) = \mu \tag{3.8}$$

3.3.3. The variance of the estimator

In this section we will make inference for the superpopulation mean $(\overline{y})$. Under repeated sampling of the same finite population, a variance estimator of $\overline{y}$ is given by Korn and Graubard(1998)

$$V\hat{a}r_{wo}(\overline{y}) = \frac{1}{d^2}\left\{\sum_{h=1}^{L}\sum_{i=1}^{k_h}\sum_{j\prec i}^{k_h}\left[\frac{\lambda_{hi}\lambda_{hj}}{\lambda_{hij}} - 1\right]\cdot\left[(t_{hi} - \overline{y}d_{hi}) - (t_{hj} - \overline{y}d_{hj})\right]^2 + Ks_w^2\right\}$$

where

$$s_w^2 = \frac{1}{K}\sum_{h=1}^{L}\sum_{i=1}^{k_h}\lambda_{hi}n_{hi}s_{hi}^2 .$$

Furthermore, λ_{hi} is the inclusion probability of the $i-th$ PSU in stratum h, λ_{hij} is the joint inclusion probability of the $i-th$ and $j-th$ PSU's in stratum h and

$$s_{hi}^2 = \frac{1}{n_{hi}-1}\sum_{j=1}^{n_{hi}}\left[(t_{hij} - \overline{y}d_{hij}) - \frac{(t_{hi} - \overline{y}d_{hi})}{n_{hi}}\right]^2$$

So, we obtain an asymptotically unbiased estimator of the variance of $\overline{y}$ that incorporates the repeated-sampling variability as well as the model variability G

$$V\hat{a}r_{SP}(\overline{y}) = V\hat{a}r_{wo}(\overline{y}) + V\hat{a}r(\overline{Y}) \tag{3.9}$$

where

$$V\hat{a}r(\overline{Y}) = \frac{1}{d^2}\left[\frac{K}{K-1}\sum_{h=1}^{L}\sum_{i=1}^{k_h}\lambda_{hi}(t_{hi} - \overline{y}d_{hi})^2 - Ks_w^2\right]$$

To avoid the necessity of specifying the joint inclusion probabilities for variance estimation due to the difficulty of the computation of this quantity, the sampling design is frequently approximated as a with-replacement stratified probability-proportional-to-size sample of PSU's (Durbin, 1953). The repeated-sampling variance estimator is given by

$$V\hat{a}r_{wr}(\overline{y}) = \frac{1}{d^2}\sum_{h=1}^{L}\frac{k_h}{k_{h-1}}\sum_{i=1}^{k_h}\left[(t_{hi}-\overline{y}d_{hi})-\frac{1}{k_h}\sum_{j=1}^{k_h}(t_{hj}-\overline{y}d_{hj})\right]^2 .$$

A second approximately unbiased variance estimator that does not require specifying the joint inclusion probabilities is given by

$$V\hat{a}r_{SP-a}(\overline{y}) = V\hat{a}r_{wr}(\overline{y}) + V\hat{a}r_b - V\hat{a}r_w \qquad (3.10)$$

where

$$V\hat{a}r_b = \frac{1}{d^2}\sum_{h=1}^{L}\frac{1}{K_h}\left[\sum_{i=1}^{k_h}(t_{hi}-\overline{y}d_{hi})\right]^2$$

and

$$V\hat{a}r_w = \frac{1}{d^2}\sum_{h=1}^{L}\frac{k_h}{K_h(k_h-1)}\sum_{i=1}^{k_h}\left[(t_{hi}-\overline{y}d_{hi})-\frac{1}{k_h}\sum_{j=1}^{k_h}(t_{hj}-\overline{y}d_{hj})\right]^2 .$$

4. CHAPTER

OPTIMAL SAMPLING METHODS UNDER A GENERAL CORRELATED POPULATION

Introduction

This chapter is comprised by models of autocorrelated finite populations. For each model, we develop the optimal sampling scheme which provides us with the best estimators. In the first section we refer to Blight's model. We describe it and we mention that the optimal sampling scheme is the centrally located systematic design. This is proved by mathematical computations in which we use the Markov property and the Bayes' theorem. We finish the part by finding the best estimators that correspond and finally with a very important comparison between the optimal sampling schemes with the random located systematic sampling and the simple random sampling. The second section assumes a generalization of Blight's sampling designs. I.Papageorgiou and K.X.Karakostas (1998) consider with an autocorrelated finite population with an integer convex autocorrelation function $\rho(\cdot)$. The results, with respect to the optimal design and estimators are presented. It is proved that Blight's results hold for this more general case. The optimal design is a systematic one with an almost symmetrical structure. This design maintains and for the asymptotic case, respectively. The third section presents a research work of R.Mukerjee and S.Segupta (1989) who obtain the optimal estimation of finite population total under a fully general correlated model. They prove that the optimal design is equivalent with a minimization problem. The analytic solution of this nonlinear programming problem is not easy, however an algorithm may be very useful to this direction. However, the computational part of this work is very intensive. For this reason, Chang-Tai Chao proposes two designs which are based on the eigensystem of the population covariance matrix. We provide a detailed presentation of these designs and we describe how these designs select the sampling units. We show the sampling locations selected by these designs under the Gaussian model with Figures. Finally, we refer to the relative efficiencies of these designs to SRS under the Gaussian model from Chang-Tai Chao´s plots.

4.1. Optimal sampling scheme for an autocorrelated finite population (B.J.N.Blight's sampling designs) under Cochran's model

4.1.1. Description of the model with autocorrelated finite population

In our experience, in many problems of sampling, the existence of autocorrelation among the members of the population is the rule rather than the exception. Cochran (1946) refers to several practical studies in which the variance of a population has been observed to increase with its size, and argues that this may be explained by considering the finite population to have been generated in sequence from a superpopulation with a monotonically decreasing autocorrelation function. This concept is obvious due to the fact that the relationship between near observations is stronger than that between distant observations. In this section we consider the problem of estimating the mean for a finite population that is generated by a simple linear Markov process.

Let the deviations of the population values from the superpopulation mean μ, be θ_1 , θ_2 ,..., θ_N. We interest to estimate $\overline{\theta} = (\theta_1 + ... + \theta_N)$ /N from a sample of size n taken without replacement from the population. Assume $x = (x_1,..., x_n)$ is the sample and let the identification vector p be such that if k is the $i-th$ element of p, then $x_i = \theta_k$. If we have 'random sampling' the elements of p are sampled randomly without replacement from the first N integers. Another sampling scheme is the systematic, where p has the following form,

$$p = \{i, i+k, ..., i+(n-1)k\}.$$

We shall call a systematic sampling scheme 'centrally located' if $2i = N + 1 - (n-1)k$. In contrast we call it 'randomly located' if i is chosen at random from the integers $(1,...,k-1)$.

We want to find the appropriate function of the data which minimizes the mean square error for $\overline{\theta}$ as the efficient estimator. According to the normality assumptions described below it is easily shown that this estimator is $E(\bar{\theta}\,|x)$ and its mean square error is $\mathrm{var}(\bar{\theta}\,|x)$. Next, we determine the choice of p to minimize this variance.

Each member of the population is assumed to have been generated by the model

$$\theta_t = \lambda\, \theta_{t-1} + \varepsilon_t \quad (t = 2,\ldots,N,\ |\lambda| < 1) \qquad (4.1)$$

where the $\{\varepsilon_t\}$ form an uncorrelated series each being distributed about zero with a constant variance, σ^2. The initial value, θ_1, is assumed to have a normal distribution with zero mean and variance $\sigma^2\,(1-\lambda^2)^{-1}$. We must note here that it is necessary to have reliable estimators of the model parameters μ, λ and σ^2.

4.1.2. Basic results

In the following basic results we will make extensive use of Bayes' theorem and the Markov property of the model.

Let

$$S = \theta_2 + \ldots + \theta_h .$$

Initially, we want to determine $f\left(S \,|\, \theta_1, \theta_{h+1}\right)$, $f(S \,|\, \theta_1)$ and $f(S \,|\, \theta_{h+1})$. The integer h represents the distance between the two consecutive observed points, θ_1 and θ_{h+1}, in the population, and for the end intervals, h represents the distance of the first observed point from the end. The Markov property of the model, as a function of θ_i, gives

$$f(\theta_i \,|\, \theta_1, \theta_{h+1}) \propto f(\theta_{h+1} \,|\, \theta_i)\; f(\theta_i \,|\, \theta_1) \quad (i = 2,\ldots,h).$$

The functions on the right are easily derived from (4.1) and equating the means of both sides we have, for $2 \le i \le h$

$$E(\,\theta_i \,|\, \theta_1, \theta_{h+1}) = (1-\lambda^{2h})^{-1} \left\{ \left(\lambda^{i-1} - \lambda^{2h-i+1}\right)\theta_1 + (\lambda^{h-i+1} - \lambda^{h+i-1})\theta_{h+1} \right\}$$

Thus, it is straightforward that

$$E(S \,|\, \theta_1, \theta_{h+1}) = \sum_{i=2}^{h} E(\theta_i \,|\, \theta_1, \theta_{h+1}) = \frac{\lambda - \lambda^h}{(1-\lambda)(1+\lambda^h)} (\theta_1 + \theta_{h+1}) \qquad (4.2)$$

It may be shown similarly that

$$E(S \mid \theta_1) = \frac{\lambda - \lambda^h}{1-\lambda}\theta_1\,,\; E(S \mid \theta_{h+1}) = \frac{\lambda - \lambda^h}{1-\lambda}\theta_{h+1} \qquad (4.3)$$

Since (4.2) is the minimum mean square error estimator of S from θ_1 and θ_{h+1}, then

$$Var(S \mid \theta_1, \theta_{h+1}) = Var(S) - \left\{\frac{\lambda - \lambda^h}{(1-\lambda)(1+\lambda^h)}\right\}^2 Var(\theta_1 + \theta_{h+1}),$$

so we have

$$Var(S \mid \theta_1, \theta_{h+1}) = \sigma^2 \frac{h(1-\varphi)}{(1-\lambda^2)}, \qquad (4.4)$$

where

$$\varphi = \frac{1}{h}\frac{1+\lambda}{1-\lambda}\frac{1-\lambda^h}{1+\lambda^h} \qquad (4.5)$$

Furthermore,

$$Var(S \mid \theta_1) = \frac{\sigma^2}{(1-\lambda^2)(1-\lambda)^2}\left\{(h-1)(1-\lambda^2) - 2\lambda(1-\lambda^{h-1}) - (\lambda - \lambda^h)^2\right\} \qquad (4.6)$$

It follows that $Var(S \mid \theta_{h+1}) = Var(S \mid \theta_1)$.

Now we have the ability to determine $E(\overline{\theta} \mid x)$ and $Var(\overline{\theta} \mid x)$ by using any given identification vector p in the previous expressions. We may assume, without loss of generality, that the elements of p are in order of magnitude, so that if we define S_i to be obviously the sum of θ's between x_i and x_{i+h}, S_0 to be the sum of θ's preceding x_1 and S_n the sum of θ's following x_n, then, when p is known,

$$NE(\overline{\theta} \mid x) = n\overline{x} + E(S_0 \mid x_1) + \sum_{i=1}^{n-1} E(S_i \mid x_i, x_{i+1}) + E(S_n \mid x_n), \qquad (4.7)$$

$$N^2 Var(\overline{\theta} \mid x) = Var(S_0 \mid x_1) + \sum_{i=1}^{n-1} Var(S_i \mid x_i, x_{i+1}) + Var(S_n \mid x_n), \qquad (4.8)$$

where $\overline{x}$ is the sample mean.

4.1.3. Optimal sampling scheme for the model

The optimal sample allocation is this which minimizes the variance of the estimator, so we first work on the related problem of choosing an appropriate integer j to minimize the conditional variance,

$$Var(\sum_{i=1}^{m}\theta_i \,\Big|\, \theta_1, \theta_j, \theta_m) \quad (1 < j < m).$$

From equations (4.4), (4.5), (4.8), it follows that the above conditional variance is equal to

$$\sigma^2(1-\lambda)^{-2}\left[\left\{j-1-\frac{(1+\lambda)}{(1-\lambda)}\frac{(1-\lambda^{j-1})}{(1+\lambda^{j-1})}\right\}+\left\{m-j-\frac{(1+\lambda)}{(1-\lambda)}\frac{(1-\lambda^{m-j})}{(1+\lambda^{m-j})}\right\}\right]$$

Here, we can see that this function is minimized when we choose an integer j to maximize

$$\frac{1-\lambda^{j-1}}{1+\lambda^{j-1}}+\frac{1-\lambda^{m-j}}{1+\lambda^{m-j}}$$

Similarly, after a little manipulation, it may be shown that this is equivalent to choosing j to minimize

$$\psi(j) = \lambda^{j-1} + \lambda^{m-j} \quad (1 < j < m)$$

Now it is convenient to distinguish two cases for λ. The first one we assume that $\lambda > 0$ and the second one $\lambda < 0$. When $\lambda = 0$ the conditional variance of $\overline{\theta}$ does not depend on p and any design will suffice.

Case 1. $\lambda > 0$

Let

$$r = \begin{cases} \frac{1}{2}m, & \textit{if } m \textit{ is even} \\ \frac{1}{2}(m-1), & \textit{if } m \textit{ is odd} \end{cases}$$

Then

$$\psi(j) - \psi(r) = \lambda^{j-1} - \lambda^{m-r})(1-\lambda^{r-j}) \quad (1 < j < m)$$

which can be seen to be nonnegative. So, the optimum value of j is r.

Therefore, when $\lambda > 0$, the variance is minimized by choosing the central of the three sample points, so that the intervals on either side of it are equal, or differ by one unit. We can consider that the variance for a design based on n unequally spaced sample points can be reduced by moving a sample point so that it is equidistant or as nearly equidistant as possible, from its neighboring sample points. Finally, we conclude at a design in which the internal intervals are either equal or differ by one unit. We can see that the optimal allocation is systematic in the case in which the two extreme sample points are chosen so that the distance between them is a multiple of $(n-1)$.

Now to determine the systematic design we must find the optimal choice of the interval spacing, h, as well as the end spacings, i, j. From the definition of h, i, j, it applies that

$$(n-1)h+i+j=N+1$$

The variance of this allocation, from the equations (4.4), (4.6), is

$$\frac{(n-1)(1-\phi)h\sigma^2}{(1-\lambda)^2}+\frac{\sigma^2}{(1-\lambda^2)(1-\lambda)^2}\,[\,\left\{(i-1)(1-\lambda^2)-2\lambda(1-\lambda^{i-1})-(\lambda-\lambda^i)^2\right\}+ \left\{(j-1)(1-\lambda^2)-2\lambda(1-\lambda^{j-1})-2(\lambda-\lambda^j)^2\right\}] \qquad (4.9)$$

where φ is known from (4.5). We consider that h, $(i+j)$ are constant values. The variance is minimized with respect to i, if we choose i to maximize

$(\lambda-\lambda^i)^2-2\lambda^i+(\lambda-\lambda^j)^2-2\lambda^j$. This is equal to $(\lambda^i+\lambda^j-1-\lambda)^2-(1+\lambda)^2-2\lambda^{i+j}$, from which it follows that we wish to choose i to minimize $(\lambda^i+\lambda^j)$. So we finally have the solution

$$i=j \quad or \quad i=j+1$$

This result states the fact that the optimal systematic allocation is the centrally located. Now we want to determine the value of spacing, so we substitute $i=j$ into (4.9) and minimize with respect to h. After a little manipulation it follows that the optimum value for h is that which maximizes

$$(1-\frac{\lambda^i}{1+\lambda})^2+(\frac{n-1}{1+\lambda^h}) \qquad (4.10)$$

where we know that $2i=N+1-(n-1)h$. There is not a simple solution for h, but for small n, N the previous function can be maximized numerically. In contrast for large values of n, N we should choose h as large as possible so that it is greatest integer less than $\frac{N}{n}$.

Case 2. $\lambda < 0$

We observe that

$$\psi(j) - \psi(2) = (\lambda^{m-j} - \lambda)(1 - \lambda^{j-2}) \quad (1 < j < m),$$

is nonnegative for $\lambda < 0$. Hence, the optimal value of j is that which gives the minimum value to the previous function. There are two values, firstly $j = 2$, or by symmetry $j = (m-1)$.

We can conclude now that in this case the optimal deterministic allocation is a two cluster design. A design with more than two clusters can reduce its variance by moving sample points to the end two clusters. Similarly with the first case, the end spacings should be equal, with common value i. The optimum value of h is that which maximizes

$$(1 - \frac{\lambda^i}{1+\lambda})^2 + (1 + \lambda^{j-2})$$

where $2i = N - n - h + 3$. If $N - n$ is large then at least one of h and i must be large. When i is large the best h is 1 and in converse, so the overall optimum value for i is 1. Hence, the two clusters should be sampled at the extreme ends of the θ sequence.

4.1.4. The best estimator

We consider now the best estimator and its mean square error for the case 1, where $\lambda > 0$. In this case, the optimal sample allocation is a centrally systematic design as this has been shown. The elements of p has the form $\{i, i+h, ..., i+(n-1)h\}$, where $2i = N + 1 - (n-1)h$ and h is chosen to maximize (4.10). From (4.2), (4.3) and (4.7) we have that

$$E(\overline{\theta} \mid x) = \frac{nh}{N}\varphi\overline{x} + \{\frac{(1+\lambda)\lambda^h}{1+\lambda^h} - \lambda^i\}\frac{x_1 + x_n}{N(1-\lambda)} \qquad (4.11)$$

The previous estimator is comprised by two components. The first one is the principal component, which is $\frac{nh}{N}\phi\bar{x}$. The second one is negligible in most of the cases. For example, if population and sample sizes are large. Even for small values of n, N this component will be small if h, i are chosen to maximize (4.10).

Correspondingly, from (4.4), (4.6) and (4.8) the mean square error for this estimator is

$$Var(\bar{\theta} \mid x) = \frac{\sigma^2}{N^2(1-\lambda)^2(1-\lambda^2)}[(1-\lambda^2)(N-n) - 2\lambda\{n+1-2\lambda^{i-1}-(n-1)\lambda^{h-1}\}$$

$$-\left\{2(\lambda-\lambda^i)^2 + \frac{2(n-1)(\lambda-\lambda^h)^2}{1+\lambda^h}\right\}] \qquad (4.12)$$

4.1.5. Randomly located systematic sampling-Comparison with the optimal sampling scheme

In the following chapter we describe another commonly used sampling scheme which is the randomly located systematic sampling. Then, we find estimator for $\bar{\theta}$, its mean square error and we compare the efficiency of this scheme with the optimum centrally located systematic design.

A randomly located systematic sample has identification vector

$$p = \{j, j+h, ..., j+(n-1)h\}$$

where j is chosen at random from the integers, $1, ..., (2i-1)$, where $2i = N-(n-1)+1$.

The best estimator of $\bar{\theta}$ of the randomly located systematic design is

$$E(\bar{\theta} \mid x) = \frac{nh\varphi}{\mathrm{N}}\bar{x} + \{\frac{(1+\lambda)\lambda^h}{1+\lambda^h} - \frac{\lambda(1-\lambda^{2i-1})}{(2i-1)(1-\lambda)}\}\frac{x_1+x_n}{N(1-\lambda)}.$$

We notice that the first component of the estimator is the same with the case of the optimum design (centrally located systematic design). Hence, we confirm that the second term of the equation is affected by the randomization.

We would like now to compare the mean square errors of the two different systematic designs (randomly and centrally located). So, let Δ represent the difference of the mean square error of the randomly located systematic design from the mean square error of the centrally located systematic design. It may be shown that

$$\Delta = \frac{2\sigma^2}{N^2(1-\lambda^2)(1-\lambda)^2}\left\{\frac{\lambda(1-\lambda^{2i-1})}{(2i-1)(1-\lambda)} - \lambda^i\right\}\left\{2(1+\lambda) - \lambda^i - \frac{\lambda(1-\lambda^{2i-1})}{(2i-1)(1-\lambda)}\right\}$$

A very interesting comparison is made by B.J.N.Blight (1973) in paper "Sampling from an autocorrelated finite population" where he develop mathematically the reduction in the variance contribution of the end spacing due to the use of a centrally located systematic design rather than a randomly located design in a Table(4.1). These are expressed as percentage reductions of the end spacing variance of the randomly located design. The parameters are the autocorrelation parameter λ and the end spacing i.

Table 4.1. Reduction in the variance contribution of the end spacing

	i			
λ	2	5	8	10
0,1	6,76	0,71	0,24	0,15
0,25	17,15	2,71	0,91	0,55
0,5	35,33	10,99	4,38	2,69
0,9	77,46	55,94	43,84	37,95

We can notice in Table 4.1 that greater reduction occurs for high correlation λ and for small spacing i. However, usually the reductions are fairly small, apart from the previous extreme cases. Hence, we may conclude that the position of the first sample point is not critical. Furthermore for large sample sizes, these reductions are also insignificant. So, there is little to choose between the two types of systematic design.

4.1.6. Simple random sampling-Comparison with the optimal sampling scheme

It would be significant a comparison between the simple random sampling with the optimum design. The significance of this comparison is derived due to the commonly use of the simple random sampling.

The method of deriving the optimum estimator of $\overline{\theta}$ in the case of random sampling is not obvious due to the fact that the data vector x(sample) contains information on the unknown identification vector p. So, we concentrate to the best linear estimator. By minimizing $E\{(a\overline{x}-\overline{\theta})\}$ with respect to a, we have that

$$a=\frac{\operatorname{var}(\overline{\theta})}{\operatorname{var}(\overline{x})}$$

where

$$Var(\overline{\theta})=\frac{\sigma^2}{N^2(1-\lambda^2)(1-\lambda)^2}\left\{\mathrm{N}(1-\lambda^2)-2\lambda(1-\lambda^{\mathrm{N}})\right\},$$

$$Var(\overline{x})=\frac{N(n-1)}{n(N-1)}Var(\overline{\theta})+\frac{N-n}{n(N-1)}\frac{\sigma^2}{(1-\lambda^2)}$$

Then, the minimized mean square error for the sample mean is $\dfrac{(1-a)}{a}\operatorname{var}(\overline{\theta})$.

We will refer two basic different cases. Let us begin with the case where $N,n\to\infty$ and $h=\dfrac{N}{n}$ remains finite and nonzero. Then,

$$a=\frac{(1+\lambda)}{2\lambda+(1-\lambda)h},$$

and the mean square error of the best linear estimator is

$$\frac{(h-1)\sigma^2}{\mathrm{N}(1-\lambda)\{2\lambda+(1-\lambda)h\}}.$$

The second case is when $\lambda=1$. Then $\alpha=1$ and the mean square error is

$$\frac{(N-n)(N+1)\sigma^2}{6nN}.$$

An interesting comparison can be made between systematic and random sampling assuming that N,n are both large. B.J.N.Blight (1973) in paper “Sampling from an autocorrelated finite population" gives us percentage relative asymptotic efficiency of random to systematic sampling (Table 4.2).

Table 4.2. Percentage relative asymptotic efficiency of random to systematic sampling

	h^{-1}				
λ	0,2	0,1	0,04	0,02	0,01
0,1	72,64	73,42	73,6	73,63	73,64
0,25	42,54	44,44	44,92	44,98	44,99
0,5	12,73	15,57	16,5	16,63	16,66
0,9	0,07	0,13	0,32	0,46	0,51

The Table 4.2 shows us, specifically for small values of autocorrelation parameter, that there is a substantial reduction in mean square error when we choose systematic sampling rather than random sampling. This reduction decreases for values of autocorrelation parameter λ near 1. The conclusion is that it is better to prefer to choose systematic sampling to have better estimators (with lower mean square errors).

4.2. Optimum sampling design for autocorrelated finite population with integer convex autocorrelation function (I.Papageorgiou and K.X.Karakostas)

4.2.1. Description of the model

In this chapter, we develop the theory of the optimum sampling under Cochran's model when the autocorrelation function is convex. We will conclude that this is the centrally located systematic design, the same with Blight's optimum design. This is significant, morever, from the point of view that, this case comprises a generalisation of Blight's sampling designs.

We denote Y_i, $(i = 1,...,N)$ the value of the study variable of the $i-th$ unit of the finite population S with size N. Let $s = \{i_1, i_2, ..., i_n\}$ is the sample with size n from population S. We would like to estimate the population mean $\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$. We can assume that $i_1 < i_2 < ... < i_n$ and that Y_i's are a realisation of a superpopulation model and have a common distribution g such that according to Cochran's model

$$E_g(Y_i) = \theta, \quad \mathrm{E}_g\{(Y_i - \theta)(\mathrm{Y}_{i'} - \theta)\} = \sigma^2 \rho(|i - i'|) \quad (i, i' = 1, 2, ..., N) \qquad (4.13)$$

where θ, σ^2 are the unknown superpopulation parameters and $\rho(\cdot)$ is the autocorrelation function. We interest here for the case in which the autocorrelation function is integer convex function. So, it follows for $\rho(\cdot)$ that

$$\rho(i) - 2\rho(i+1) + \rho(i+2) \geq 0 \quad (i = 0, 1, 2, ..., N)$$

4.2.2. Optimal sampling scheme

We would like to find the optimum sampling scheme which minimizes the mean square error of the estimator. As we know, an estimator of $\overline{Y}$ is the least square estimator denoted by $\overline{y}_L = \frac{\sum_{i \in s} Y_i}{N}$. The mean square error of this estimator under the superpopulation approach is given by Karakostas (1990)

$$MSE(\overline{y}_L \mid s) = Var(\overline{Y}) + \frac{1}{nN}(\frac{N}{n} j_n' \Gamma_s j_n - 2 j_N' \Gamma_{S,s} j_n) \qquad (4.14)$$

where j_k $(k = n, N)$ is an appropriate vector of ones,

$\Gamma_s = Var(Y_s)$, $\Gamma_{S,s} = \mathrm{cov}(Y_S, Y_s)$, with $Y_s = (Y_{i_1}, Y_{i_2}, ..., Y_{i_n})'$ and $Y_S = (Y_1, Y_2, ..., Y_N)'$.

It is straightforward that the minimization of the right part of the equation(4.14) is equivalent with the minimization of the following form

$$\frac{N}{n} \sum_{i \in s} \sum_{j \in S} \rho(|i - j|) - 2 \sum_{i \in s} \sum_{j \in S} \rho(|i - j|). \qquad (4.15)$$

Theorem 4.2.1.(I.Papageorgiou and K.X.Karakostas). If $\rho(\cdot)$ is an integer autocorrelation function with $\rho(0)=1$ and $\lim_{h\to\infty}\rho(h)=0$, the sampling design which minimizes (4.2.3) is a systematic one with an almost symmetrical structure.

Before we start to prove this theorem we will explain what is systematic sampling with an almost symmetrical structure and we will mention a very useful Lemma.

Let $s=\{i_1,i_2,...,i_n\}$ is a sample and we denote $h_j=i_{j+1}-i_j$ $(j=1,...,n-1)$ to be the distance between two consecutive sampling points. We mention that $h_0=i_1-1$ and $h_n=N-i_n$. The symmetrical structure in a sampling design occurs if $h_0=h_n$, $h_1=h_{n-1}$ and so on. A sampling design has an almost symmetrical structure if $|h_i-h_{n-i}|\leq 1$, with $i=1,...,\frac{1}{2}n-1$ or $\frac{1}{2}(n-1)$ when n is even or odd, respectively.

Lemma 4.2.1(Karakostas and Wynn, 1989). Every integer convex function ρ with $\lim_{h\to\infty}\rho(h)=0$ can be written as

$$\rho(h)=\sum_{r=1}^{\infty}\alpha_r\varphi_r(h), \qquad (4.16)$$

where a_r are nonnegative constants with $\sum_{r=1}^{\infty}\alpha_r=1$ and $\phi_r(h)=\frac{(r-h)^+}{r}$ with

$$(r-h)^+=\begin{cases} r-h, & (r>h) \\ 0, & (r\leq h)\end{cases}$$

Proof of the Theorem 4.2.1.Lemma 4.2.1 shows us that the minimisation of (4.15) is equivalent to

$$\min[\sum_{r=1}^{\infty}\alpha_r\left\{\frac{N}{n}\frac{1}{r}\sum_{i\in s}\sum_{j\in s}(r-|i-j|)^+-2\frac{1}{r}\sum_{i\in s}\sum_{j\in s}(r-|i-j|)^+\right\}] \qquad (4.17)$$

If we now minimize every term in (4.17) then $\frac{N}{n}j_n^{'}\Gamma_s j_n-2j_N^{'}\Gamma_{S,s}j_n$ will also be minimized. For $r\geq N-1$ the minimization of (4.17) involves

$$\min\left\{-nN+\frac{1}{r}\Big(\frac{-N}{n}\sum_{i\in s}\sum_{j\in s}|i-j|+2\sum_{i\in s}\sum_{j\in s}|i-j|\Big)\right\}=\min\Big[-N+nN(N-1)-$$

$$-2\frac{N}{n}\sum_{i=1}^{n-1}i(n-i)h_i-2\left\{h_0(h_1+\ldots+h_n)+\ldots+(h_0+\ldots+h_{n-1})h_n\right\}\Big] \qquad (4.18)$$

By the minimization of (4.18) we take the optimum values of $h_k, (k=0,1,\ldots,n)$. We can find these values by using Lagrange multipliers. So let

$$Q(h_0,h_1,\ldots,h_n)=-2\frac{n}{N}\sum_{i=1}^{n-1}i(n-i)h_i-2\left\{h_0(h_1+\ldots+h_n)+\ldots+(h_0+h_1+\ldots+h_{n-1})h_n\right\}$$

$-\lambda(h_0+\ldots+h_n-N+1)$,

where λ is Lagrange multiplier. By equating to zero the first derivatives of Q with respect to $h_k, (k=0,1,\ldots,n)$ and λ and solving the resulting equations with respect to $h_0,\ldots,h_n$, we find first that

$$h_0=h_n,\;\; h_1=h_{n-1},\;\ldots,\;\; h_{\frac{1}{2}n-1}=h_{\frac{1}{2}n+1} \qquad (4.19)$$

when n is even. Similar relations hold and for n when is odd. Finally we get that

$$h_1=h_2=\ldots=h_{n-1}=\frac{N}{n},\;\; h_0=h_n=\frac{N-n}{2n}. \qquad (4.20)$$

The $h_i's$ must be integers so we have

$$h_i=\left[\frac{N}{n}\right] \;\; or \;\; h_i=\left[\frac{N}{n}\right]+1,\quad (i=1,2,\ldots,n-1), \qquad (4.21)$$

while h_0 and h_n are equal to or to $\left[(N-n)/(2n)\right]$ *or to* $\left[(N-n)/(2n)\right]+1$, where [] stands for the integer part. All the previous theory and the determination of h_i (4.20) concerns to the case where $r\geq N-1$. For the case where $r<N-1$ not all of the $h_i's$ will appear in the system of equations from the use of Lagrange multipliers. However the right-hand side of this system will continue to have the same symmetrical structure as for $r\geq N-1$, so $h_i's$ has the same form with (4.20) in the case $r<N-1$.

4.2.3. Asymptotic case

We now consider the case in which n, N tend to infinity in such a way that $f=\frac{n}{N}$ is a constant. Here in this case(Karakostas and Wynn, 1989), to find the optimum sampling design we have to minimize

$$\lim_{n\to\infty} j_n^{'}\Gamma_s j_n \qquad (4.21)$$

Firstly, we will find the sampling design which minimizes $j_n^{'}\Gamma_s j_n$ for any value of N, n with Theorem 4.2.2.

Theorem 4.2.2.For given values of N and n the sampling design which minimizes the quantity $j_n^{'}\Gamma_s j_n$ is a systematic design with an almost symmetrical structure.

The proof is based on the definition of a convex function. Here the convex function is the autocorrelation function $\rho(\cdot)$.

Consider any sampling design $s=\{h_0,h_1,...,h_{n-1},h_n\}$ of size n and assume another sampling design $s^*=\{h_0^*,h_1^*,...,h_{n-1}^*,h_n^*\}$ with the property that

$$h_0^*=h_n, \quad h_1^*=h_{n-1},\ldots,h_{n-1}^*=h_0$$

We write the quantity $j_n^{'}\Gamma_s j_n$ as

$$j_n^{'}\Gamma_s j_n = n+\Bigg[\sum_{i=1}^{n-1}\{\rho(h_i)+\rho(h_i^*)\}+\sum_{i=1}^{n-2}\{\rho(h_i+h_{i+1})+\rho(h_i^*+h_{i+1}^*)\} +...+$$

$$+\{\rho(h_1+h_2+...+h_n)+\rho(h_1^*+h_2^*+...+\rho(h_{n-1}^*)\} \quad \Bigg].$$

Since we examine the case which the autocorrelation function $\rho(\cdot)$ is an integer convex. By the definition of convexity it follows that

$$j_n^{'}\Gamma_s j_n \geq n+\Bigg\{\sum_{i=1}^{n-1}\{\rho(\frac{h_i+h_i^*}{2})+\sum_{i=1}^{n-2}\rho(\frac{h_i+h_i^*+h_{i+1}+h_{i+1}^*}{2})+...+$$

$$+\rho(\frac{h_1+h_1^*+h_2+h_2^*+\ldots+h_{n-1}+h_{n-1}^*}{2})\ \Big\} = j_n^{'}\Gamma_{\bar{s}}\, j_n,$$

where $\bar{s}$ is a sampling design of size n with $\bar{h}_i = \frac{1}{2}(h_i + h_i^*)$ for $i = 1,2,\ldots,n-1$. From the definition of s, s^* we have for this sampling design that

$$\bar{h}_i = \bar{h}_{n-i} \quad (i = 1,2,\ldots,\frac{1}{2}n-1 \quad if\ n\ is\ even\ or\ \frac{1}{2}(n-1) \quad if\ n\ is\ odd) \quad (4.22)$$

We will examine now what does it happen for the optimum sampling design when these values tend to infinity with $f = \frac{n}{N}$ constant. We have the following Theorem.

Theorem 4.2.3.In the asymptotic case, with $N \to \infty$, $n \to \infty$, $f = \frac{n}{N}$ is constant, and for an integer convex autocorrelation function $\rho(\cdot)$, the sampling sequence which minimizes (4.3.1) is a systematic one with the property that the distances between successive sampled units are equal or differ by one.

We will describe the proof of the Theorem (4.2.3) from I.Papageorgiou and K.X.Karakostas. They use the following notation given in Karakostas and Wynn (1989) and Karakostas (1990). Let X_i is the indicator variable,

$$X_i = \begin{cases} 0, & if\ i \in s \\ 1, & if\ i \notin s \end{cases}$$

then define

$$N_r^{11}(s) = \#\{X_{i+r} \mid X_i = 1 \quad and \quad X_{i+r=1}\},$$

$$N_r^{01}(s) = \#\{X_{i+r} \mid X_i = 0 \quad and \quad X_{i+r} = 1 \quad or \quad X_i = 1 \quad and \quad X_{i+r} = 09\},$$

$$N_r^{00}(s) = \#\{X_{i+r} \mid X_i = 0 \quad and \quad X_{i+r} = 0\}$$

So the expression (4.3.1) can be written as

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} N_i^{11}(s)\rho(i) \qquad (4.23)$$

We get from Theorem 4.2.2 that for any sampling design s,

$$\frac{1}{N}\sum_{i=1}^{N} N_i^{11}(s)\rho(i) \geq \frac{1}{N}\sum_{i=1}^{N} N_i^{11}(\bar{s})\rho(i).$$

If we take the limits of the previous inequality we have that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} N_i^{11}(s)\rho(i) \geq \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} N_i^{11}(\bar{s})\rho(i)$$

So we finished the proof of Theorem(4.2.3) by obtaining that the optimum sampling sequence is $\bar{s}$. The form of this sampling sequence is given by (4.22).

We can conclude that I.Papageorgiou and K.X.Karakostas show something very significant. We notice that Blight's optimum sampling design is still optimum for any integer convex autocorrelation function. So in this section we have a generalization of Blight's sampling designs.

4.3. Optimal estimation of finite population total under a general correlated model(R.Mukerjee and S.Segupta)

4.3.1. Description of the superpopulation model and notation

We denote with $Y_i \quad (i = 1,2,...,N)$ the value of the study variable of the $i-th$ unit of the finite population U with size N. We, then, consider a superpopulation model consisting of prior distributions a such that

$$E_\alpha(Y_i) = \mu_i, \;\; E_\alpha\{(Y_i - \mu_i)(\mathrm{Y}_j - \mu_j)\} = \upsilon_{ij}, \qquad (4.24)$$

where E_α denote expectation with respect to a. This model imposes no constrains to $V_{ij} \quad (i, j = 1,...,N)$. Consequently, the variance-covariance matrix Σ of Y_i $(i = 1,...,N)$ can be of any type. This in the most general case for a superpopulation model and it has been studied by R.Mukerjee and S.Segupta (1989). The objective here is again to estimate the population total $Y = \sum_{i=1}^{N} Y_i$, an equivalent with respect to inference quantity with the mean $\bar{Y}$, on the basis of a sample. The sample s is a subset of U population, drawn according to a sampling design p with positive inclusion probability π_i for every unit i. E_p denotes expectation with respect to p. Furthermore, let P_n denotes the class of designs p with fixed sample size n and L_u denotes the class of linear unbiased estimators

$$e = \alpha_s + \sum\nolimits_{i \in s} b_{si} Y_i \tag{4.25}$$

where the α_s and b_{si}'s are real constants satisfying the following equations

$$E_p(\alpha_s) = \sum\nolimits_s \alpha_s p(s) = 0\,, \quad \sum\nolimits_{s \supset i} b_{si} p(s) = 1 \quad (i = 1, \ldots, N) \tag{4.26}$$

Finally H_n is the class of strategies (p, e) with $p \in P_n$ and $e \in L_u$. We know that we derive the optimal strategy in the class of H_n under the model (4.24), when we will find the minimum of the expected variance $E_\alpha E_p\{(e - Y)^2\}$. So we will work in this direction.

4.3.2. Optimal estimator

We consider a strategy $(p, e) \in H_n$. Let b_s be a $n \times 1$ vector with elements b_{si} $(i \in s)$. Furthermore, we denote V_s the submatrix of V obtained by considering the units $i \in s$ and let 1 be a $N \times 1$ vector with all elements unity. Using the equations of (4.26) we have that

$$E_\alpha E_p\{(e - Y)^2\} = \sum\nolimits_s (\alpha_s - \sum_{i=1}^{N} \mu_i + \sum_{i \in s} b_{si} \mu_i)^2 \, p(s) + \sum_s b_s^{'} V_s b_s p(s) - 1^{'} V 1$$

$$\geq \sum\nolimits_s b_s^{'} V_s b_s p(s) - 1' V 1 \tag{4.27}$$

to hold, with the equality if and only if,

$$a_s = \sum_{i=1}^{N} \mu_i - \sum_{i \in s} b_{si} \mu_i \tag{4.28}$$

for every s with $p(s)$.

Let $V_s^{-1} = ((\upsilon_s^{ij}))$. Then we define for $i, j = 1, \ldots, N$

$$\phi_{ij} = \sum\nolimits_{s \supset ij} \upsilon_s^{ij} p(s) \qquad (4.29)$$

and it follows the $N \times N$ matrix Φ with its elements ϕ_{ij}.

Theorem 4.3.1.For a given $p \in P_n$, under the superpopulation model (4.24)

$$E_\alpha E_p\{(e-Y)^2\} \ \geq \ 1^{'}\Phi^{-1}1 - 1^{'}V1$$

for every $e \in L_n$, with equality if and only if $e = e^*$, where e^* is specified by (4.28). Further, a strategy (p,e) is optimal in H_n provided $(p,e) = (p^*, e^*)$, where p^* is a sampling design that minimizes $1^{'}\Phi^{-1}1$ with respect to $p \in P_n$.

We will prove the Theorem(4.3.1). Let

$$\lambda = (\lambda_1, \lambda_2, ..., \lambda_{\mathrm{N}})' = \Phi^{-1}1, \qquad (4.30)$$

where we denote λ_s as a $n \times 1$ subvector of λ given by the elements $i \in s$ and

$$b_s^* = V_s^{-1}\lambda_s \qquad (4.31)$$

with its elements $b_{si}^* \quad (i \in s)$.

Equations (4.26), (4.29)-(4.31) give us that

$$\sum\nolimits_s b_s^*{}'V_s b_s^* p(s) = \sum\nolimits_s \lambda_s{}'V_s^{-1}\lambda_s p(s) = \lambda'\Phi\lambda = 1'\Phi^{-1}1,$$

$$\sum\nolimits_s b_s{}'V_s b_s^* p(s) = \sum\nolimits_s b_s{}'\lambda_s p(s) = 1'\lambda = 1'\Phi^{-1}1$$

Straightforward, by the previous equations we obtain

$$E_\alpha E_p\{(e-Y)^2\} \geq \sum\nolimits_s (b_s - b_s^*)'V_s(b_s - b_s^*)p(s) + 1'\Phi^{-1}1 - 1'V1$$

$$\geq 1^{'}\Phi^{-1}1 - 1^{'}V1 \qquad (4.32)$$

with equality if and only if (4.28) holds and further

$$b_s = b_s^* \qquad (4.33)$$

for every s with $p(s) > 0$.

We finally conclude that the optimal estimator for a given p under the model (4.24), is given by (4.28), (4.33), with b_s^* given by (4.31). So, we finished the proof of the Theorem (4.3.1).

It is interesting to consider the special case of a model (4.24) which has for $1 \le i \ne j \le N$, $\upsilon_{ij} = \rho(\upsilon_{ii}\upsilon_{jj})^{\frac{1}{2}}$ with constant ρ, $\frac{-1}{N-1} < \rho < 1$.

Theorem 4.3.2.Under the superpopulation model (4.24) with $\upsilon_{ij} = \rho(\upsilon_{ii}\upsilon_{jj})^{\frac{1}{2}}$ $(1 \le i \ne j \le N)$ a strategy (p,e) is optimal in H_n if and only if $\pi_i = \pi_{i0}$ for every i $(1 \le i \le N)$ and e is given by

$$e = \sum_{i \in s}(Y_i - \mu_i)/\pi_{i0} + \sum_{i=1}^{N}\mu_i$$

for every s with $p(s) > 0$.

Now we give the proof of this Theorem. The relation (4.29) shows here that

$$\phi_{ii} = g_1 \upsilon_{ii}^{-1}\pi_i \quad (1 \le i \le N), \quad \phi_{ij} = g_2(\upsilon_{ii}\upsilon_{jj})^{-\frac{1}{2}}\pi_{ij} \quad (1 \le i \ne j \le N)$$

where

$$g_1 = \frac{1+(n-2)\rho}{(1-\rho)\{1+(n-1)\rho\}}, \quad g_2 = \frac{-\rho}{(1-\rho)\{1+(n-1)\rho\}}$$

and π_{ij} is the joint distribution probability of units i and j.

By defining $S = (\upsilon_{11}^{\frac{1}{2}},..,\upsilon_{\mathrm{NN}}^{\frac{1}{2}})$, it is shown that $S'\Phi S = g_1 n + g_2 n(n-1)$ and that by the Cauchy-Schwarz inequality, $S'\Phi S \ge (1'S^2)/S'\Phi S$. Hence

$$1'\Phi^{-1}1 - 1'V1 \ge (1-\rho)\left\{ n^{-1}(\sum_{i=1}^{N}\upsilon_{ii}^{\frac{1}{2}})^2 - \sum_{i=1}^{N}\upsilon_{ii} \right\}$$

with equality if and only if ΦS is proportional to 1 or equivalently

$$\pi_i = \frac{n\upsilon_{ii}^{\frac{1}{2}}}{(\sum_{i=1}^{N}\upsilon_{ii}^{\frac{1}{2}})} = \pi_{i0}$$

for every $i \quad (i = 1,...,N)$. Now it follows that for any p with $\pi_i = \pi_{i0} \quad (i = 1,...,N)$, it is easy to verify that $b_{si} = \pi_{i0}{}^{-1}$ so the theorem is obvious.

4.3.3. Optimal sampling design

In the previous paragraph we mentioned that the optimal design requires the minimization of $1'\Phi^{-1}1$ with respect to $p \in P_n$. The analytic solution of this nonlinear programming problem is not easy, however there are some algorithms may be available that help to this direction.

Let

$$S = \left\{\left(i_1, i_2, ..., i_n\right) : 1 \le i_1 < ... < i_n \le N\right\}$$

is the set of all the possible s. A design p in P_n may be represented by nonnegative quantities $\{p(s), s \in S\}$. It is obvious that $\sum_{s\in S} p(s) = 1$. By the definition of ϕ_{ij} we denote

$$\Phi = \sum_{s\in S} p(s)T(s),$$

where $s = (1,...,n)$ and the $N \times N$ matrix $T(1,...,n)$ is defined as

$$T(1,...,n) = \begin{pmatrix} V_{1...n}^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

$V_{1...n}$ is the $n \times n$ submatrix of V given by the n rows and columns.

Theorem 4.3.3.A design $\{p^*(s), s \in S\}$ is optimal in the sense of minimizing $1'\Phi^{-1}1$, that is maximizing -$1'\Phi^{-1}1$, in P_n if and only if

$$F(\Phi^*, s) = \lim_{c\to 0^+} c^{-1}[1'(\Phi^*)^{-1}1 - 1'\left\{\left(1-c\right)\Phi^* + cT(s)\right\}^{-1}1] \le 0 \qquad (4.34)$$

for every $s \in S$, where $\Phi^* = \sum_{s \in S} p^*(s)T(s)$.

The matrix $T(s)$ is nonnegative-definite for each s, so from (4.34) it follows that a design $\{p^*(s), s \in S\}$ is optimal in P_n if and only if

$$F(\Phi^*, s) = 1'(\Phi^*)^{-1}T(s)(\Phi^*)^{-1}1 - 1'(\Phi^*)^{-1}1 \le 0 \tag{4.35}$$

for every $s \in S$.

A numerical determination of the optimal design is very difficult. A version of W-algorithm (Silvey, 1980, pp 29-30) can help us to this determination.We mention a brief description of this version of W-algorithm.

Let δ be a pro-assigned positive quantity and $0 < c_k < 1$ be a real sequence which has $\lim c_k = 0$ and $\sum c_k$ divergent. Firstly we can assume that the design

$$p_1(s) = \binom{N}{n}^{-1}$$

for each $s \in S$. Let $\{p_\kappa(s), s \in S\}$ for $k = 1, 2...$ be the design in $k-th$ iteration and $\Phi_k = \sum_{s \in S} p_\kappa(s)T(s)$. The iteration of the algorithm stops at $k-th$ stage if $\max_{s \in S} F(\Phi_k, s) < \delta$. Otherwise, we continue to the $(k+1)$ stage of iteration with the design which follows

$$p_{k+1}(s) = \begin{cases} (1-c_{\kappa+1})p_\kappa(s), & s \neq s_{(k+1)}, \\ (1-c_{\kappa+1})p_\kappa(s_{(\kappa+1)}) + c_{\kappa+1}, & s = s_{(k+1)}, \end{cases}$$

where $s_{(\kappa+1)}$ maximizes $F(\Phi_\kappa, s)$ over the set S. It follows that

$$\Phi_{k+1} = (1-c_{k+1})\Phi_k + c_{k+1}\mathrm{T}(s_{(k+1)}).$$

We can see that when the algorithm terminates at the $k-th$ stage so we have that $1'(\Phi_k)^{-1}1 < 1'(\Phi^*)^{-1}1 + \delta$ where Φ^* corresponds to the optimal design. Thus the algorithm leads to the minimum possible value of $1'\Phi^{-1}1$.

Now we mention and give the mathematical solution of an example, which is relative with the theory of optimal design under a general correlated model

(Biometrika, paper ´Optimal estimation of finite population total under a general correlated model´ by Rahul Mukerjee and S.Sengupta, page 791).

Example 4.3.1.Let the population size $\mathrm{N}=4$ and the sample size $n=2$. Assume that ${\upsilon_{ii}}^2=\sigma^2$ for $i=1,...,4$, $\upsilon_{ij}=0.5\sigma^2$ for $1\le i\ne j\le 3$ and otherwise $\upsilon_{ij}=0$. So we have the following V matrix.

$$V=\sigma^2\begin{pmatrix}1 & 0.5 & 0.5 & 0\\ 0.5 & 1 & 0.5 & 0\\ 0.5 & 0.5 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

Let

$$p(1,2)=p(1,3)=p(2,3)=q_1,$$
$$p(1,4)=p(2,4)=p(3,4)=q_2.$$

It follows that

$$3q_1+3q_2=1\Leftrightarrow q_1+q_2=\frac{1}{3}\Leftrightarrow q_2=\frac{1}{3}-q_1 \qquad (4.36)$$

By the Theorem 4.3.1, a strategy (p,e) is optimal in H_n provided $(p,e)=(p^*,e^*)$, where p^* is a sampling design that minimizes $1^{'}\Phi^{-1}1$ with respect to $p\in P_n$. Hence, we would like to minimize $1^{'}\Phi^{-1}1$ to find the optimal design p^*. So we will find firstly the matrix Φ.

$$\Phi_{ij}=\sum_{s\supset i,j}\upsilon_s^{ij}p(s)$$

where $((\upsilon_s^{ij}))=V_s^{-1}$ is the $(i-j)-$element of the inverse of the submatrix of V which arises if we keep the rows and the columns which correspond to the elements of the sample. We compute now the V_s^{-1} for all $s\in S$.

$$V_{s\supset\{1,2\}}^{-1}=\left[\sigma^2\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}\right]^{-1}=\frac{1}{\sigma^2}\begin{pmatrix}\frac{4}{3} & \frac{-2}{3}\\ \frac{-2}{3} & \frac{4}{3}\end{pmatrix}=\frac{2}{3}\sigma^2\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}$$

Similarly

$$V^{-1}_{s\supset\{1,3\}} = \frac{2}{3}\sigma^2\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix},\ V^{-1}_{s\supset\{2,3\}} = \frac{2}{3}\sigma^2\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$$V^{-1}_{s\supset\{1,4\}} = \left[\sigma^2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Similarly

$$V^{-1}_{s\supset\{2,4\}} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ V^{-1}_{s\supset\{3,4\}} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Below we compute separately every element of the matrix Φ.

$$\Phi_{11} = \upsilon^{11}_{s_1} p(s_1) + \upsilon^{11}_{s_2} p(s_2) + \upsilon^{11}_{s_3} p(s_3)$$

where s_1, s_2, s_3 are the tree possible samples which contain the element 1. $\{1,2\}, \{1,3\}, \{1,4\}$ correspondingly. Hence, we have

$$\Phi_{11} = \frac{2}{3\sigma^2}2q_1 + \frac{2}{3\sigma^2}2q_1 + \frac{1}{\sigma^2}q_2 \Leftrightarrow \Phi_{11} = \frac{8}{3\sigma^2}q_1 + \frac{1}{\sigma^2}q_2$$

Similarly

$$\Phi_{12} = \Phi_{21} = \upsilon^{12}_{s_1} p(s_1) = \frac{2}{3\sigma^2}(-1)q_1 = -\frac{2}{3\sigma^2}q_1 \ ,$$

$$\Phi_{13} = \Phi_{31} = -\frac{2}{3}q_1,$$

$$\Phi_{23} = \Phi_{32} = -\frac{2}{3\sigma^2}q_1,$$

$$\Phi_{22} = \Phi_{33} = \frac{8}{3\sigma^2}q_1 + \frac{1}{\sigma^2}q_2,$$

$$\Phi_{14} = \Phi_{41} = \Phi_{24} = \Phi_{42} = \Phi_{34} = \Phi_{43} = 0,$$

$$\Phi_{44} = \frac{3}{\sigma^2}q_2.$$

$$\Phi=\begin{pmatrix} \frac{8}{3\sigma^2}q_1+\frac{1}{\sigma^2}q_2 & -\frac{2}{3\sigma^2}q_1 & -\frac{2}{3\sigma^2}q_1 & 0 \\ -\frac{2}{3\sigma^2}q_1 & \frac{8}{3\sigma^2}q_1+\frac{1}{\sigma^2}q_2 & -\frac{2}{3\sigma^2}q_1 & 0 \\ -\frac{2}{3\sigma^2}q_1 & -\frac{2}{3\sigma^2}q_1 & \frac{8}{3\sigma^2}q_1+\frac{1}{\sigma^2}q_2 & 0 \\ 0 & 0 & 0 & \frac{3}{\sigma^2}q_2 \end{pmatrix}$$

$$\Leftrightarrow \Phi=\frac{1}{\sigma^2}\begin{pmatrix} \frac{8}{3}q_1+q_2 & -\frac{2}{3}q_1 & -\frac{2}{3}q_1 & 0 \\ -\frac{2}{3}q_1 & \frac{8}{3}q_1+q_2 & -\frac{2}{3}q_1 & 0 \\ -\frac{2}{3}q_1 & -\frac{2}{3}q_1 & \frac{8}{3}q_1+q_2 & 0 \\ 0 & 0 & 0 & 3q_2 \end{pmatrix}$$

We substitute (4.36) in Φ we calculate $1'\Phi^{-1}1$ and we find that

$$1'\Phi^{-1}1=2\frac{13q_1-5}{(3q_1-1)(q_1+1)}$$

Finally, we want to minimize $1'\Phi^{-1}1$, so we find the first derivative of this.

$$(1'\Phi^{-1}1)'=\frac{-6(13q_1^{\,2}-10q_1+1)}{(3q_1-1)^2(q_1+1)}$$

We find q_1 which $(1'\Phi^{-1}1)'=0$. We have the following two solutions:

$a)\quad q_1=0.1182$

$b)\quad q_1=0.6511$

The solution b is rejected because from (4.36) it follows that q_2 is negative and this not possible because q_2 is probability. So we have that

$$q_1=0.1182\,,\ q_2=\frac{1}{3}-q_1=0.2152$$

It is shown that the optimal design is given by p^*, where

$$p^*(1,2)=p^*(1,3)=p^*(1,4)=0.1181$$
$$p^*(1,4)=p^*(2,4)=p^*(3,4)=0.2152$$

By (4.28) and (4.33), the optimal strategy in H_n is (p^*,e^*), where

$$e^{*}(s)=\begin{cases}1.7889\sum_{i\in s}(Y_i-\mu_i)+\sum_{i=1}^{4}\mu_i, & 1\leq i \text{ p } \; j\leq 3,\\ 2.6834(Y_i-\mu_i)+1.5489(Y_4-\mu_4)+\sum_{i=1}^{4}\mu_i, & 1\leq i\leq 3\end{cases}$$

It can be checked finally that this optimal design satisfies (4.35).

4.4. Optimal design under a correlated population based on the eigensystem of the population covariance matrix (Chang-Tai Chao).

4.4.1. Description of the model

In this chapter we assume a model with a correlated population and we examine the selection of n sampling units out of N units to predict the population quantity of interest. More specifically, we have a correlated spatial population and we show that we can obtain lower prediction mean-square error with careful sampling arrangement of the sampling sites. An example is that the systematic design can be used to select samples for better prediction results. However, it is only effective under certain population covariance structures. So, we mention in this chapter two sampling methods of Chang-Tai Chao, which are based on the eigensystem of the population covariance matrix. Other authors give computationally intensive algorithms to find the optimal sample by minimizing the mean-square error. The advantages of the two methods of Chang-Tai Chao, which we will develop below, are that it is not required computationally intensive algorithm. Furthermore, these methods require fewer population assumptions.

Assume that the population consists of N units labelled $1,2,...,N$. Let $y=(y_1,y_2,...,y_N)'$ be the vector of the values of the variable of interest which is considered as a realization of a random vector $Y=(Y_1,Y_2,...,Y_N)'$. Let s be the sample of n units selected from the population and y_s, the vector of y values associated with s, be the vector of observed values. We assume that $T(Y)$ is the population quantity of interest. We would like to find an optimal sampling strategy which contains an unbiased estimator $\hat{T}(Y)$ and a appropriate design that can select s to minimize the conditional mean-square error of $\hat{T}$ given s,

$$E[(T-\hat{T})^2|s]$$

Let Y be the population random vector with mean vector $\mu=(\mu_1,\mu_2,...,\mu_{\mathrm{N}})'$, where $\mathrm{E}(\mathrm{Y}_i)=\mu_i$ and Covariance matrix

$$Var(Y)=\Sigma=\{\sigma_{ij}\}_{i,j=1,...,N}$$

where $\sigma_{ij}=\begin{cases} Var(Y_i), & if\ i=j \\ Cov(Y_i,Y_j), & if\ i\neq j \end{cases}$

We consider here the prediction of the population total $T(Y)=\sum_{i=1}^{N}Y_i$. We select n sampling units out of N population units to predict the previous population quantity $T(Y)$.

4.4.2. Sampling designs based on the eigensystem of the population covariance matrix

The main object in this section is to find the best prediction result. So, we would like to select the sample, which give us the conditional mean-square error as small as possible. One could suggest searching one by one all the possible samples to examine its mean-square error. Since the population size is finite, the number of different possible samples is finite as well. However, the number of samples is $\binom{N}{n}$. This number is extremely large in the most of the cases. Hence, this idea of a complete enumeration is not convenient.

An idea is that the sampling units that can give lower mean-square prediction error are the units that have better prediction ability from the other unselected. Consequently, one would like to select the units that account for as much total population variability as possible.

Let

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\mathrm{N}}$$

be the ordered eigenvalues of Σ and

$$e_1, e_2, \ldots, e_N$$

be the associated normaliyed eigenvectors. The N-dimensional coordinate system can be rotated into a new N-dimensional orthogonal coordinate system, in which the N axes are the linear combinations of the original variables, such that the coefficients of the $i-th$ linear combination, denoted as X_i, $i = 1, 2, \ldots, N$, are the components of the ith eigenvectors e_i. It follows,

$$X_i = e_i'Y = e_{i1}Y_1 + e_{i2}Y_2 + \ldots + e_{iN}Y_N$$

where e_{ij} is the $j-th$ component of the $i-th$ eigenvector and we have the below matrix e,

$$e = \begin{pmatrix} e_{11} & e_{12} & \mathrm{K} & e_{1N} \\ e_{21} & \mathrm{O} & \mathrm{K} & \mathrm{M} \\ \mathrm{M} & \mathrm{K} & \mathrm{O} & \mathrm{M} \\ e_{N1} & \mathrm{K} & \mathrm{K} & e_{NN} \end{pmatrix}$$

The e_{ij} is also known as the $i-th$ principal component in the principal component analysis (PCA). The variability in Y is extracted into the variances of uncorrelated random variables, X_i. Furthermore,

$$\sum_{i=1}^{N} Var(Y_i) = \sum_{i=1}^{N} Var(X_i)$$

and the variance of X_i is

$$Var(X_i) = \lambda_i \quad \forall i = 1, 2, \ldots, N$$

If we are looking for the units which account more in the total variability, a reasonable candidate(s) would be those of the original variables that are strongly associated with the leading PCAs. The absolute value of the loadings of Y_i in the PCAs is an indicator of this association.

We described previously theoritically the general idea of the eigensystem of the population covariance matrix and we gave the reason why a design based on this intuition can offer us better prediction results. We will develop now two sampling designs of Chang-Tai Chao which are based on the eigensystem of the covariance matrix. In addition, we will see that these designs can select s to minimize the conditional mean-square error. Chang-Tai Chao propose these sampling designs to select

$$s = \{i_1, i_2, ..., i_n\}, \;\; i_j \in \{1,2,..., N\}$$

with a fixed sample size n.

Design 1

The n sampling units in this design, are selected based on the component that have the largest absolute value in each of the first n eigenvectors. We describe below the steps.

$$\textit{Step }1: \quad i_1 = j, \left|e_{1j}\right| = \max_i \left|e_{1i}\right|$$

M

$$\textit{Step } k: \quad i_k = j, \; \left|e_{kj}\right| = \max_{i, i \notin s} \left|e_{ki}\right|$$

repeat step k till $k = n$.

Design 2

In the design 1 the sampling units are selected based on the magnitude of their corresponding components in the first n eigenvectors. In this Design now, the units of the sample are selected depending not only on the magnitude but also on the sign of their corresponding components in the leading eigenvectors. We describe the following steps of this Design.

$$n = 1: \; s = \{j\}, \left|e_{1j}\right| = \max_i \left|e_{1i}\right|$$

$$n > 1: \textit{Step } 1: \textit{Let } s' = \{j_1, j_2, ..., j_m\}, m < N$$

where

$$\left|e_{1j_1}\right| \geq \left|e_{1j_2}\right| \geq ... \geq \left|e_{1j_m}\right| \geq ... \geq \left|e_{1j_N}\right|.$$

The number m is an integer which indicates the number of units in s' and it can be appropriately specified before the survey according to the population size N.

Step k: Let $s_{tmp} = \{l_1, l_2\}$ where $l_1,\ l_2$ satisfy:

1. $l_1,\ l_2$ not having been selected into s.

2. $$\left|e_{kl_1}\right| = \max_{i} \left|e_{ki}\right|$$

3. $$\left|e_{kl_2}\right| = \max_{\substack{i \\ e_{kj} \cdot e_{kl_1} \mathrm{p}\, 0}} \left|e_{kj}\right|$$

Units l_1, l_2 will be added into s by

$$\begin{cases} i_{2(k-1)} = l_1, i_{2k-1} = l_2, & if\ n \geq 2k-1 \\ i_{2(k-1)} = l_1, & if\ n = 2(k-1) \end{cases}$$

repeat step k till $n = 2k-1 \quad or \quad n = 2(k-1)$.

Final adjustement: Let $s_{-i_1} = \{i_2, ..., i_n\}$ and $i_1 = j_p, j_p \in s'$ such that j_p satisfies

$$mcor(j_p, s_{-i_1}) = \min_{\substack{j_k \in s' \\ j_k \notin s_{-i_1}}} mcor(j_k, s_{-i_1})$$

where $mcor$ is the **multiple correlation coefficient*** between unit j_k and the set s_{-i_1}.

4.4.3. Sampling locations in spatial Gaussian model

In the previous section, we analyzed the two sampling designs of Chang-Tai Chao. Now we give an example of Chang-Tai Chao which shows us the sampling locations selected by these sampling designs for spatial Gaussian model. We assume that the population size $N = 25$ and the sample size $n = 5$.

$*$The multiple correlation coefficient between $X_1, X_2, ..., X_k$ is denoted by $mcor(1,(2,...,k)) = \rho^2_{1,(2,...,k)} = 1 - \dfrac{|R|}{R_{11}}$, where $R = \left\|\rho_{ij}\right\|$, $|R|$ is the determinant of R, R_{11} is the determinant of the new matrix which arises if from the matrix R we eliminate the column and the row of the $11-$element of the matrix.

In the spatial Gaussian model, the population random vector Y follows a multivariate normal distribution

$$Y \sim N(\mu, \Sigma)$$

where

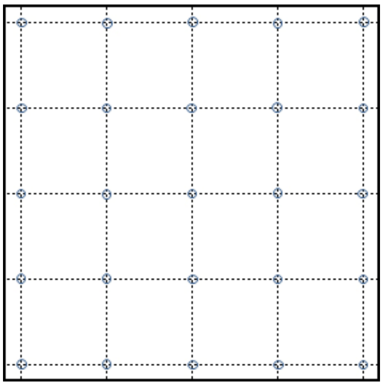
$$\mu = (\mu_1, \mu_2, ..., \mu_N)', \Sigma = \{\sigma_{ij}\}, \quad i, j = 1, ..., N$$

Here, a Gaussian-shaped spatial covariance function(Cressie, 1993) is used to generate Σ

$$\sigma_{ij} = \sigma^2 \exp(-\|h\|^2 / c^2)$$

where h is the Euclidean distance between sites i and j. The parameter c determines the strength of covariance in the study region. The larger the c is, the stronger the covariance between population units. In the simulation of Chang-Tai Chao, parameters values, $c = 3.5, \quad \mu_i = 0 \quad \forall i$ and $\sigma^2 = 1$ are used.

Figure 4.1



We show first that the sampling sites(population units) are the crosspoints of a 5×5 rectangular grid(Figure 4.1).

Figure 4.2

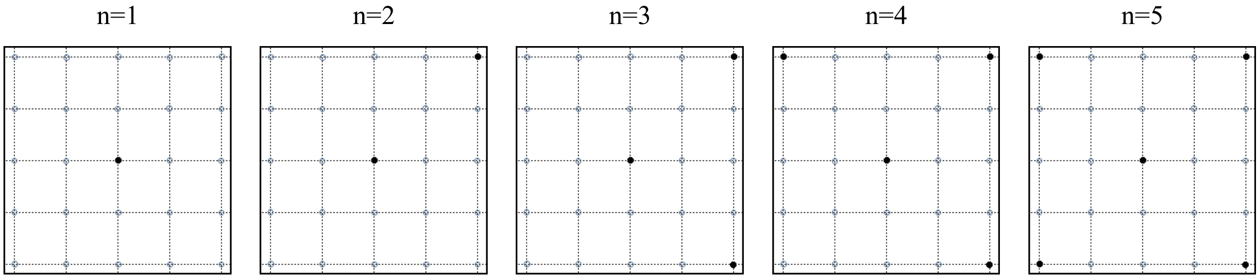


Figure 4.3

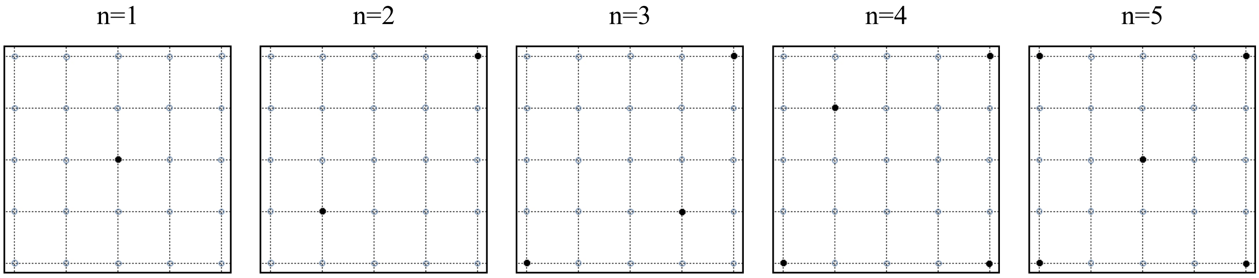


Figure 4.2 illustrates the sampling sites selected by the design 1 and Figure 4.3 illustrates the sampling sites selected by the design 2 respectively. It is assumed that m = 9 in design 2.

We can see in Figure 4.2 that the ideal sampling sites by design 1 should be spread symmetrically and evenly in the region study. However, this design does not give such an arrangement for n = 1,2,3,4. In Figure 4.3 the design 2 gives the same result as the design 1 for n=5. In design 2 we know that the selection of the sampling sites arises by taking into account the sign of e_{ij} and secting i_1 from s´according to the final adjustement. Figures 4.2 and 4.3 improve different arrangement of the sampling sites for n=1,2,3,4.

A quite common case is that where the sampling locations cannot be distributed as regular as in Figure 4.1. For example, an air pollution study where the monitoring sites might distribute irregularly at the study region. So, it would be very interesting to consider the case that the possible sampling locations are distributed randomly. We will describe an application of Chang-Tai Chao. The coordinates of the locations of the 25 population sites are generated by a bivariate uniform distribution((A_i, B_i) be the coordinates, A_i~Unif(1,5), B_i~(Unif(1,5)). Chang-Tai Chao in the above mentioned work conclude for this example that it is more difficult to arrange the sampling sites symmetrically. We see also that design 2 do it better and the sampling sites has more symmetrical figure.

4.4.4. Relative efficiency to simple random sampling

A comparison between the proposed designs and the simple random sampling can be made to examine the performance of these designs. This comparison can be made based on the relative efficiencies of design 1 and 2 to simple random sampling. The relative efficiency of a design to simple random sampling is defined as the ratio of the mean-square prediction error obtained with SRS to that obtained with the design. A value greater than 1 indicates that the proposed design is more efficient. The population size used in this section is $N = 81$ and the population quantity of interest is the population total.

$$\mathrm{T(Y)} = 1_N^{'}\,\mathrm{Y} = T(Y) = 1_N{}'Y = \sum_{i=1}^{N} Y_i$$

where 1_N is a vector of length N in which all elements are 1. The best linear unbiased predictor (BLUP) for the population total,

$$T_1 = 1_n^{'} w_s + 1_{N-n}^{'}[v_{\bar{s}} + \Lambda_{\bar{s}s}\Lambda_{ss}^{-1}(w_s - v_s)]$$

(Bolfarine and Zacks, 1992, p.25), where $\bar{s}$ is an index set containing the labels of the unselected units, w_s is the vector of observed values, v_s and $v_{\bar{s}}$ consist of the mean values associated with s and $\bar{s}$. $\Lambda_{\bar{s}s}$ is the covariance matrix between $W_{\bar{s}}$ and W_s and Λ_{ss} is the covariance matrix of W_s. The best unbiased predictor (BUP) is equivalent with the BLUP under the Gaussian model.

In order to construct the ratio for the efficiency we simulated the mean-square prediction error. The simulation was proceed by producing K realizations of the model and corresponding design and calculating the

$$E(T-\hat{T})^2 = \frac{1}{K}\sum_{j=1}^{K}(T_j-\hat{T}_j)^2\ )$$

where T_j and $\hat{T}_j$ are the true and the predicted population total of the $j-th$ realization.

Chang-Tai Chao considers for the Gaussian model two sampling situations: the regularly and the randomly distributed possible sampling locations. Suppose that $K=15000$, $N=81$, $c=3.5$. He calculates the relative efficiency of the proposed designs to SRS for the previous cases and for sample sizes from 1 to 40 and shows as the values with plots.

The performance of the design 1 is often better than that of SRS in both of cases(regularly and randomly distributed population sites). The performance of the design 1 could also be worse than SRS because sometimes the design 1 gives 'bad' arrangement of sampling sites for some sample sizes. In fact, it is well known that the sampling units should be arranged symmetrically and evenly in the study region under such a correlated population with equal variance.

It is shown through various plots given in this work that the performance of the design 2 is in general better than SRS in both cases. Although, design 2 does not perform as well in the case where the population sites are randomly distributed.

We conclude that both designs have no certain behaviour and are comparable with the SRS design, a design that is not proposed for its efficiency. A possible explanation can be drawn from the fact that the way both designs were obtained did not require any optimal properties of the resulting design. The underlying method was rather intuitive intending to produce an easy to implement solution.

If we make a comparison between design 1 and 2 we see that almost always design 2 is better and for regularly and for randomly distributed population sites.

CHAPTER

COMPARISON BETWEEN OPTIMAL DESIGNS

Introduction

Closing up this review study on superpopulation approach in sampling we proceed with a short comparison study between the optimal sampling designs that have been presented in the previous chapters. The comparison study is based on the efficiency of these sampling designs.

In practice, not always the best design or the best estimator is used. It is important to know, at least, how far is the result we are going to use from the optimal. The aim of this comparison study is to assess the efficiency of the available methodologies and also provide with a relative measure of “distance” of each method with the optimal one.

Initially, we suppose that we have an autocorrelated finite population with an integer convex autocorrelation function (in this study, $\rho(u)=e^{-0.1u}$). The observations of the population are generated by the multivariate normal distribution. We compare the centrally located systematic design of I.Papageorgiou and K.X.Karakostas (section 4.2) with the Design 1 of Chang-Tai-Chao (section 4.4). In this comparison, the optimal sampling design by R. Mukerjee and S.Sengupta (section 4.3) is not included because there are many practical problems. In fact this methodology, although quite general, can only work for small (far from realistic) sizes of n and N. So, we focus our comparison study to the previous sampling designs. We select the optimal samples of these designs from the data and we find the prediction of the mean. Then, we find the mean square error of the prediction of the centrally located systematic design of I.Papageorgiou and K.X.Karakostas and we compare it with the mean square error of the simple random sampling. The study procedure is followed for Chang-Tai-Chao design. We study these designs for different values of sampling size n $(n=5, n=10, n=20)$ to have a more completed idea. The population size N is 100.

5.1. Generation of the observations

5.1.1. Data

The population size of this study is $N = 100$. The values of the observations of the population X_i are generated by the multivariate normal distribution $N(M, \Sigma, N)$, where M is a matrix $1 \times N$ of means, Σ is the covariance matrix $N \times N$ and the N is the number of observations we generate. Without loss of generality, we suppose that $\sigma^2_{X_i} = 1$, $\forall\, i \in N$, where $\mathbf{X} = (X_1, ..., X_N)$ is the vector $1 \times N$ of the observations of the population. We know that

$$\rho(\mathrm{X}_i, X_j) = \frac{Cov(X_i, X_j)}{\sigma_{X_i} \sigma_{\mathrm{X}_j}}, \;\; i, j \in (1, ..., N) \qquad (5.1)$$

so,

$$\rho(X_i, X_j) = Cov(X_i, X_j)\,. \qquad (5.2)$$

We assume that we have an autocorrelated finite population with an integer convex autocorrelation function. In this study we suppose that $\rho(u) = e^{-0.1u}$ where $u = |i - j|$. This function is an integer convex. Now from the relation (5.2) we can calculate the covariance matrix.

We give the covariance matrix and a vector of zeros M (vector of means $1 \times N$). We simulate from the multivariate normal distribution and obtain each time vectors of dimension $1 \times N$, with the observations of the population for our study. So the generation of the data is straightforward.

Based on the optimal sampling designs, we select

$$s = \{i_1, ..., i_n\}, \;\; i_j \in \{1, ..., N\}, \quad i_j \neq i_{j'} \quad \forall j \neq j' \qquad (5.3)$$

with a fixed sample size n, where i_j is the selected position in the vector of the observations of all the population.

5.2. Application of the optimal sampling designs

5.2.1. Description of the sampling designs

In section 4.2 we referred to the optimal sampling design proposed by I.Papageorgiou and K.X.Karakostas. They develop the theory of the optimum sampling under Cochran's model when the autocorrelation function is convex. They conclude that this is the centrally located systematic design. In section 4.4, we developed the idea of Chang-Tai-Chao about the optimal sampling design, where the selection of the sampling units under a correlated population based on the eigensystem of the population covariance matrix. He presents two different designs. In the comparison below, we select sampling units by Design 1, as we describe it in section 4.4. In the comparison study, the optimal sampling design proposed by R.Mukerjee and S.Sengupta (section 4.3) is not included. This happens due to the difficulty of this method for large values of population and sampling size (when the value $\binom{N}{n}$ is large). So, the application of this method becomes impossible. It requires extremely intensive computation and complicated procedure for this case.

5.2.2. Centrally located systematic design (by I.Papageorgiou and K.X.Karakostas)

We have seen that the optimal sample by the design proposed by I.Papageorgiou and K.X.Karakostas, is the systematic design with an almost symmetrical structure. However, we'll apply here the more accurate centrally located systematic design. Firstly, we calculate the internal spacing

$$h = \left[\frac{N}{n}\right], \tag{5.4}$$

where $[\cdot]$ stands for the integer part.

Let

$$a = N + 1 - (n-1)h$$

then we find the ending spacing $i-1$ because

$$i = \left[\frac{a}{2}\right]. \tag{5.5}$$

We have i , h so we have the optimal sample of this systematic design

$$p_1 = \{i, i+h, ..., i+(n-1)h\}. \tag{5.6}$$

In the application we suppose that the population size is $N = 100$. We generate observations from the multivariate normal distribution. We take samples with sizes $n = 5, n = 10, n = 20$ from a vector $1 \times N$ of the observations. We find the mean of the sample vector. The mean of the design's sample is the prediction of the mean. The variance of this prediction is given by the function below (Model-complete strategies for sampling from convex autocorrelated finite populations by I.Papageorgiou and K.X.Karakostas, 2001)

$$V(\hat{\theta}) = \frac{1}{N^2}\left[V(i)+V(j)+(n-1)\left\{(h+1)+2\sum_{k=1}^{h}(h+1-k)\rho(k)-\frac{2(\sum_{k=0}^{h}\rho(k))^2}{(1+\rho(h))}\right\}\right] \tag{5.7}$$

where

$$V(m) = \frac{1}{N^2}\left[m+2\sum_{k=1}^{m-1}(m-k)\rho(k)-(\sum_{k=0}^{m-1}\rho(k))^2\right].$$

5.2.3. **Optimal sampling design based on the eigensystem** (by Chang-Tai-Chao**)**

The idea of Chang-Tai-Chao about the optimal sampling design is based on the eigensystem of the population covariance matrix. We find at first instance the eigenvectors of the covariance matrix. Then we follow the steps of the Design 1:

Step 1 : $i_1 = j$, $\left|e_{1j}\right| = \max\limits_{i}\left|e_{1i}\right|$

M

Step k : $i_k = j$, $\left|e_{kj}\right| = \max\limits_{i, i \notin s}\left|e_{ki}\right|$

repeat step k till $k = n$.

So we find now the optimal sample of this sampling design

$$q_1 = \{i_1, i_2, \ldots, i_n\} . \tag{5.8}$$

We take the sample from the same vector $1 \times N$ with the observations. We find the mean of the sample vector. The mean of this design's sample is the prediction of the mean of the design. Then, we do not find the variance of the prediction because we have not a function for it.

5. 3. Results of the study

5. 3.1. Numeral results

In this section, we mention the results of the study of the application of these two sampling designs when the sample sizes are $n = 5, n = 10, n = 15$ correspondingly. In the Table 5.1 we can see the prediction of the population mean ($\hat{\theta}$) for each sampling design when the sampling sizes are $n = 5, n = 10, n = 15$ correspondingly.

Table 5.1.Estimator $\hat{\theta}$ of the sampling designs ($n = 5, n = 10, n = 20$)

sampling design\sampling size	$n = 5$	$n = 10$	$n = 20$
Systematic centrally located	0.0682	0.1902	0.1972
Eigensystem's design	0.5741	0.4474	0.2646

In the Table 5.2 below, we give the mean square error of the prediction for all the previous cases of the sampling size. We also give the mean square error of the prediction for the simple random sampling ($S^2(1-f)/n$). We can see that the mean square error of the centrally located systematic design is smaller than this of the simple random sampling for all the sampling sizes. This is expected as the above mentioned result has been obtained as we have seen in chapter 4, based on the optimality of MSE criterion.

Table 5.2.Mean square error of the prediction for the systematic design and the eigensystem's design

sampling design\sampling size	$n = 5$	$n = 10$	$n = 20$
Systematic centrally located	0.0381	0.0135	0.0037
Eigensystem's design	0.19	0.09	0.04

5.3.2. Conclusion of the comparison

We mentioned previously the numerical results of the comparison study between the centrally located systematic design (by I.Papageorgiou and K.X.Karakostas) and the sampling design based on the eigensystem (by Chang-Tai-Chao).

We can compare these designs with the relative efficiency of the one design to another one. Here we'll define the relative efficiency as the ratio of the mean-square prediction error obtained with the sampling design based on the eigensystem to that obtained of the centrally located systematic design. A value greater than 1 indicates that this kind of systematic design is more efficient.

We can see that for all the cases of the sampling sizes, the relative efficiency is greater than 1. So, the design proposed by I.Papageorgiou and K.X.Karakostas is more efficient than the sampling design by Chang-Tai-Chao.

APPENDIX

A1. This program produces an integer convex autocorrelation function. Specifically this function is $\rho(u) = e^{-0.1u}$, where u is $u = |i - j|$. We suppose that $\mathbf{X} = (X_1, ..., X_N)$ and $i, j \in (1, ..., N)$.

```
function [sigma] = convexcovariance(k,N)

% k is a coefficient controlling the exponential function.
% N is the population size.

C = [];
for m = 1:1:N
   C = [C;m*ones(1,N)];
end;
P = C - C';
P = abs(P);
P = k * P;
sigma = exp(P);
```

A2. Here values of the observations of the population X_i are generated by the multivariate normal distribution $N(M, \Sigma, N)$, where M is a matrix $1 \times N$ of means, Σ is the covariance matrix $N \times N$ and the N is the number of observations we generate.

```
function [Y] = mvnormrnd(mu,sigma,n)

% MVNORMRND - Multivariate Normal - Random Number Generation %
Copyright (c) 1998, Harvard University. Full copyright in the file Copyright % % Y =
mvnormrnd(mu, sigma, n) %
```

```
%   mu = p by 1 mean column vector or n by p matrix of means
%   sigma = covariance matrix
%   n = number of observations to generate
%
%   Y = an n by p matrix of row vectors with mean mu and covariance sigma
%
% Note: works slightly different from Matlab builtin MVNRND.
%
%   if mu is a column vector, n rows will be returned, all with mean mu
%
%   if mu is a matrix, a matrix of the same size will be returned with
%   row Y(i,:) having mean mu(i,:) .
%
% See also: MVNORMPDF, MVNORMLPR

[d1,d2] = size(mu);
S = chol(sigma)';
if d2==1,
% then mu is a column vector
X = normrnd(0,1,n,d1);
Y = X*S' + ones(n,1)*mu' ;
else
X = normrnd(0,1,d1,d2);
Y = X*S' + mu ;
end
```

A3. This function firstly gives the eigenvectors of the covariance matrix. Then this program produces the optimal sample of the sampling design 1 of Chang-Tai-Chao.

```
function [p] = eigensystemsample(sigma,n)

[E,D] = eig(sigma);
E = abs(E);
[M,I] = max(E');
p = I([1:1:n]);
```

A4. This function produces the optimal sample of the systematic centrally located design by I.Papageorgiou and K.X.Karakostas**.**

```
function [p] = systematicsample(n,N)

h = floor(N/n);
a = (N+1) - ((n-1)*h);
i = floor(a/2);
p = i:h:N;
```

A5. This function gives the variance of the estimator of the sampling design 1 by Chang-Tai-Chao**.**

.

```
function [Vi] = V(i,N)

Nsquare = N*N;
A = [(i-1):-1:1];
K1 = -0.1*[1:1:(i-1)];
B = exp(K1);
K2 = -0.1*[0:1:(i-1)];
C = exp(K2);

Vi = i + (2 * (A*B')) - (sum(C)^2);
Vi = Vi / Nsquare;
```

A6. The program below calculates the variance of the estimator of the systematic centrally located design (I.Papageorgiou and K.X. Karakostas (2001).Model-complete strategies from convex autocorrelated finite populations, Journal of Statistics Planning and Inference (p.83)).

```
function [Fi] = F(i,n,N,h)


Nsquare = N*N;

A = [h:-1:1];

K1 = -0.1*[1:1:h];

B = exp(K1);

K2 = -0.1*[0:1:h];

C = exp(K2);


S = sum(C)^2;

S = (2 * S) / (1 + exp(-0.1*h));


Fi = (2 * (A*B')) - S;

Fi = (h + 1) + Fi;

Fi = (n - 1) * Fi;

Fi = (2 * V(i,N)) + Fi;

Fi = Fi / Nsquare;
```

A7. Here I calculate the mean and the variance of the optimal samples for both of designs by Chang-Tai-Chao and by I.Papageorgiou and K.X.Karakostas. ($N = 100$, $n_1 = 5$, $n_2 = 10$, $n_3 = 20$)

```
N = 100;

M = 1;

n1 = 5;

n2 = 10;
```

```
n3 = 20;

h1 = floor(N/n1);
h2 = floor(N/n2);
h3 = floor(N/n3);

k = -0.1;

sigma = convexcovariance(k,N);
mu = zeros(M,N);
X = mvnormrnd(mu,sigma,N);
p1 = eigensystemsample(sigma,n1);
p2 = eigensystemsample(sigma,n2);
p3 = eigensystemsample(sigma,n3);

q1 = systematicsample(n1,N);
q2 = systematicsample(n2,N);
q3 = systematicsample(n3,N);

i1 = q1(1);
i2 = q2(1);
i3 = q3(1);

Y1 = X(:,p1);
Y2 = X(:,p2);
Y3 = X(:,p3);

Z1 = X(:,q1);
Z2 = X(:,q2);
Z3 = X(:,q3);

meanY1 = mean(Y1');
meanY2 = mean(Y2');
meanY3 = mean(Y3');
```

```
meanZ1 = mean(Z1');
meanZ2 = mean(Z2');
meanZ3 = mean(Z3');

Vi1 = V(i1,N);
Vi2 = V(i2,N);
Vi3 = V(i3,N);

Fi1 = F(i1,n1,N,h1);
Fi2 = F(i2,n2,N,h2);
Fi3 = F(i3,n3,N,h3);
```

REFERENCES

Damianou C., (1999). Sampling Methodology: Techniques and Applications, 3rd Edition, Aithra, Athens.

William G.Cochran, (1953). Sampling Techniques, 2nd Edition, John and Sons, New York.

Paul S.Levy, Stanley Lemeshow, (1999). Sampling of populations: Methods and Applications, 3rd Edition, John Wiley and Sons, New York.

Barry I.Graubard and Edward L.Korn, (2002). Inference for superpopulation parameters-Using sample surveys, Statistical Science, Vol 17, No 1, 73-96.

David J.Bartholomew, Fiona Steele, Irini Moustaki and Jane Galbraith. The analysis and Interpretation of multivariate data for social scientists, Chapman and Hall.

I.Papageorgiou, (1998). Optimal sampling for autocorrelated finite population, Doctoral Thesis.

B.J.N.Blight, (1973). Sampling from an autocorrelated finite population, Biometrika, 60, 375-385.

I.Papageorgiou and K.X.Karakostas, (1998). On optimal sampling designs for autocorrelated finite populations, Biometrika, 85, 482-486.

R.Mukerjee and S.Segupta (1989). Optimal estimation of finite population total under a general correlated model, Biometrika, 76, 789-794.

I.Papageorgiou and K.X.Karakostas (2001). Model-complete strategies for sampling from convex autocorrelated finite populations, Journal of Statistical Planning and Inference, 99, 71-89.

Chang-Tai-Chao (2001). Selection of sampling units under a correlated population based on the eigensystem of the population covariance matrix, vol 15, no 8, 757-775.

https://doi.org/10.26219/heal.aueb.5554