SOME APPLICATIONS OF NON-LINEAR PROGRAMMING TECHNIQUES IN SAMPLE SURVEYS

ABSTRACT
OF
THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
LUCKNOW

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
LUCKNOW
प्रज्ञा शील करुणा
ESTABLISHED 1996

FOR THE AWARD OF DEGREE OF

Doctor of Philosophy
IN
STATISTICS

Submitted by

Arun Pal

Enrollment No: 727/18

Co-Supervisor
Dr. Subhash Kumar Yadav
Associate Professor

Supervisor
Dr. Rahul Varshney
Assistant Professor

DEPARTMENT OF STATISTICS
SCHOOL OF PHYSICAL & DECISION SCIENCES
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
Vidya Vihar, Raebareli Road, Lucknow-226025 (U.P.), India

2022

Abstract

This thesis entitled “Some Applications of Non-Linear Programming Techniques in Sample Surveys" is submitted to the Babasaheb Bhimrao Ambedkar University, Lucknow, India, to supplicate the degree of Doctor of Philosophy in Statistics. It embodies the research work carried out by me in the Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow.

Optimization is a technique to find the best solution among all available solutions for the formulated mathematical programming problems. In the modern world, sample survey play an indispensable role in almost every sphere of human activity. In sample surveys the precision suffers when too small a sample is selected and the cost increases when the sample is too large. A sample survey is the most efficient, scientific, economical and reliable data collection method for collecting information and suggesting the best estimator to predict unknown population parameters. The objective of the sample survey is to provide these information, about the population under study, with the use of minimum budget and other available resources in feasible time. The problem arises in conducting the sample surveys is to opt a sampling design that either assures the maximum precision of the sample estimates for the given budget of survey or suggest the minimum budget for a given level of precision on estimates as the case may be. In sampling theory various sampling designs are suggested for different situations and the nature of the population. The emphasis is given, in this thesis, on the formulation of non-linear programming problems of sample surveys for multivariate case. Also, the solutions of non-linear programming problems are suggested by using some suitable optimization techniques.

In stratified sample surveys, the problem of determining the sample sizes from the strata is known as the problem of allocations and several types of allocations i.e. equal, proportional, and optimum allocations are available in literature. Among these, an optimum allocation is considered as the best allocation. In multivariate stratified sampling, where more than one characteristic is defined on each population unit, the optimum allocation for one characteristic will not, in general, be optimum for other characteristics. Thus, the problem of allocation becomes more complex and the problem of sample allocations in multivariate stratified sampling has been discussed by various authors for a long time, starting apparently with Neyman (1934). In such circumstances, a compromise criterion is required, resulting in an optimum allocation for all the characteristics in some sense. Since this allocation is based on a compromise criterion, it is known as a compromise allocation in multivariate stratified sampling. Many authors such as Ali *et al.* (2015), Varshney *et al.* (2015), Haseen *et al.* (2016), Raghav *et al.* (2017), Varshney *et al.* (2017), Ansari *et al.* (2018) and others, worked out for compromise allocation by defining new estimators or used existing estimators in different situations for multivariate case. In this thesis, the problem of obtaining a compromise allocation in various situations is studied in detail. When population means are to be evaluated, Yates (1960) recommended the criterion.

In many decision-making problems, sample surveys, environmental, social, economic, and technical areas have more than one multiple objective functions. It is significant to realize that multiple objectives are often non commensurable and in conflict with each other in optimization problems. The problem of deriving the statistical information on population characteristics may be formulated as an optimization problem of minimizing the cost of survey subject to the restriction that the loss of precision is within the prescribed limit or alternately minimizing the loss in precision subject to the restriction that the cost of the survey remains within the given budget. Some authors formulated these problems as optimization problems and suggested using various non-linear programming techniques to solve them, such as dynamic programming, lexicographic goal programming, fuzzy programming, genetic algorithm, etc.

The aim of this thesis is to explore the formulations of non-linear programming problems in sample surveys and their solutions by using some suitable optimization techniques. It is also aimed to inquire whether optimization techniques have any significant effects on sampling theory. The discussions in this thesis are included on non-response problems under two stage sampling.

The thesis comprises six chapters. Among them, **Chapter 1** provides a primary introductory material to equip the reader for understanding the problems discussed in the following chapters.

In **Chapter 2**, we deals with the problem of obtaining compromise allocation for multivariate cluster sampling design. The authors worked out compromise allocation for gamma cost function. It is discussed for equal sized cluster and formulated the problem as All Integer Non-Linear Programming Problems (AINLPPs). The authors, also, suggested the use of appropriate technique to find the best solution. The solutions are obtained by using genetic algorithm, lexicographic goal programming and lagrange multiplier technique. A comparative study is carried out to find the best suited technique.

In **Chapter 3**, we deals with the problem of optimum allocation in two stage sampling design under non-response. This chapter addresses the issue of determining the compromise allocation for a multivariate two-stage sample design in the setting of non-response.

It is investigated for first stage units of equivalent size, and the problem is referred to as AINLPP. The authors also suggested the use of different strategy in order to get the best solution.

In **Chapter 4**, we proposed the improved ratio cum product type of exponential estimator for population mean in stratified random sampling under linear cost function. We formulated our problem as an AINLPP and we solve this by genetic programming technique using MATLAB. This chapter has been presented in different sections which involve introduction, review of literature, proposed estimator, linear cost function, theoretical efficiency comparison, numerical illustration based on real data, result and discussion, conclusion.

In **Chapter 5**, we worked on an improved memory type ratio estimator in stratified random sampling under linear and non-linear cost functions, drawing inspiration from Aslam *et al.* (2020). The intricacy is written in the AINLPP form. The recommended estimate's best allocation with mean square error is determined in this scenario, and the proposed estimator is compared to other comparable estimators.

Under **Chapter 6**, drawing inspiration from Aslam *et al.* (2020), we worked on an improved memory type product estimator in stratified random sampling under linear cost function. The problem is framed in the form AINLPP. The recommended estimate's best allocation with mean square error is determined in this scenario, and the estimator is compared to other comparable estimators in literature.

The research work presented in Chapters 2 to 6 of this thesis is based on my following published/accepted/communicated research papers.

1. VARSHNEY, R., PAL, A., MRADULA, AND ALI, I. (2022). Optimum allocation in multivariate cluster sampling design in the presence of gamma cost function. *Journal of Statistical Computation and Simulation.*

DOI: 10.1080/00949655.2022.2104845.

2. PAL, A., MRADULA, AND VARSHNEY, R. (2022). Compromise allocation in multivariate two-stage sampling design in presence of non-response. *International Journal of Agricultural and Statistical Sciences.* 18(1), 273–279.

3. VARSHNEY, R., PAL, A., YADAV, S.K., AND SINGH, L. (2022). Optimal strategy for improved estimation of population mean under stratified sampling design using genetic programming. *International Journal of Agricultural and Statistical Sciences,* 18(2), 799–804.

4. PAL, A., YADAV, S.K., VARSHNEY, R., AND DUBE, M. (2022). Improved memory type ratio estimator for population mean in stratified random sampling with linear and non-linear cost functions. *Journal of Statistics and Management Studies.* (Under Revision).

5. YADAV, S.K., VISHWAKARMA, G.K., VARSHNEY, R., AND PAL, A. (2022). Improved memory type product estimator for population mean in stratified random sampling with linear and non-linear cost functions. *Journal of Statistics and Management Studies.* (Communicated).

Bibliography

Ali, I., Shafiullah., and Haseen, S. (2015). Geometric programming approach in multivariate stratified sample surveys in case of non-response. *Electronic Journal of Applied Statistical Analysis*, 8 (1), 28–43.

Ansari, A. H., Varshney, R., and Ahsan, M. (2018). Compromise mixed allocation in multivariate stratified sampling using dynamic programming technique. *Journal of Advanced Statistics*, 3 (4), 71–79.

Aslam, I., Noor-Ul-Amin, M., Yasmeen, U., and Hanif, M. (2020). Memory type ratio and product estimators in stratified sampling. *Journal of Reliability and Statistical Studies*, 13 (1), 1–20.

Haseen, S., Ali, I., and Bari, A. (2016). Multiobjective stochastic multivariate stratified sampling in presence of nonresponse. *Communications in Statistics - Simulation and Computation*, 45 (8), 2810–2826.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97 (4), 558–625.

Raghav, Y. S., Khan, M. F., and Khalil, S. (2017). Sample allocation problem in multiobjective multivariate stratified sample survey under two stage randomized response model. *Journal of Mathematical and Computational Science*, 7 (6), 1074–1089.

Varshney, R., Gupta, S., and Ali, I. (2017). An optimum multivariate multiobjective stratified sampling design: fuzzy programming approach. *Pakistan Journal of Statistics and Operation Research*, 13 (4), 829–855.

Varshney, R., Khan, M. G. M., Fatima, U., and Ahsan, M. J. (2015). Integer compromise allocation in multivariate stratified surveys. *Annals of Operations Research*, 226 (1), 659–668.

Yates, F. (1960). *Sampling methods for censuses and surveys*. Charles Griffin and Co. Ltd., London.