SOME CONTRIBUTIONS TO TWO ESTIMATION PROBLEMS: ESTIMATION OF COMMON MEAN AND SIZE OF A FINITE POPULATION

By

MOLOY DE

Tata Consultancy Services, Kolkata

Thesis submitted to the University of Calcutta for the award of the degree of Doctor of Philosophy (in Science)

Kolkata

2005

PREFACE

This thesis is being submitted to the university of Calcutta in fulfillment of the primary requirement for the award of the degree of Doctor of Philosophy in Science.

No part of the thesis was submitted for a degree of any University or prize. Some of the results described in this dissertation are, however, contained in four papers published in Calcutta Statistical Association Bulletin (1995, 1998), Sankhya Series B (1997) and Sequential Analysis (2003) and one paper accepted for publication in Journal of Applied Statistical Science (to appear in 2005). One chapter of this dissertation was presented in a National Level Conference at Calcutta University (1999).

The thesis was prepared under the active guidance of Prof. Samindranath Sengupta, Department of Statistics, University of Calcutta, to whom I owe a deep debt of gratitude for his general supervision and numerous instructions which had directed the course of the research. The final version of the thesis owes much to his thorough editing of the entire manuscript.

I am grateful to Prof. Bikas Kumar Sinha of Indian Statistical Institute, Kolkata and Prof. Jyotirmoy Sarkar of Indiana University Purdue University, Indianapolis for their invaluable help in my research work. Both of them co-authored some of my published papers. I also thank Prof. C. G. Bhattacharya and Prof. Kamal Roy for their helpful comments on some areas of my research work.

I thank Mr. Sujoy Mukhoti for helping me out with the simulation work.

Let me mention of my parents Prof. Yudhisthir De and Mrs. Bina De, my wife Mrs. Sudeshna De and my sister Mrs. Manjima Mitra Mustafi for their constant inspiration and support without which I could not have completed my work.

I thank Mr. Ajoyendra Mukherjee and Mr. Sudhir Mallick of Tata Consultancy Services, Kolkata for granting me the necessary permission to carry out research work in Statistics.

I also thank all my friends and colleagues who have shown interest in my research work at various points of time during my research activity.

Moloy De

CONTENTS

INTRODUCTION

This dissertation deals with the application of Estimation Theory in two fields viz. 'Estimation of Common Mean' and 'Estimation of the Size of a Finite Population'. Both the fields already possess vast literature and still attract researchers for further challenging issues. We have tried to address some of these in this thesis which is divided into two parts. In the first four chapters [Part I] we consider the problem of estimation of common mean and in the next two chapters [Part II] we address the problem of estimation of the size of a finite population. All the references cited in the thesis are listed at the end.

Part I: Estimation of Common Mean

The problem of estimation of common mean (μ) of two or more univariate normal populations with possibly unknown unequal variances based on independent samples of fixed sizes has some unresolved problems. The crux of the problem lies in the fact that the family of distributions of the sufficient statistic is not complete as the expectation of the difference of the two sample means is zero for all μ. In fact the uniformly minimum variance unbiased estimator (UMVUE) of μ in this problem does not exist. When the population variances are known the BLUE (in fact the UMVUE) of μ depends on the ratio of the variances of the two populations. Since most often the ratio is unknown, a plug-in estimator was suggested by Graybill and Deal (1959) and it is known in the literature as Graybill-Deal Estimator. Attempts have been made to prove some decision theoretic properties like admissibility or restricted admissibility of the Graybill-Deal Estimator or to improve upon the same by considering different other estimators as well as sequential sampling procedures.

Bhattacharya (1978, 1979, 1980, 1981, 1988) considered the problem in details. His work evolved mainly from the design problems. The common mean estimators of Yates type or Cohen-Sakrowitz type were developed. Zacks (1970) developed the Bayes and Fiducial Equivariant estimators. A good review work on this problem of common mean estimation is available in Pal and Sinha (1996). Several researchers have tried to establish the admissibility of the Graybill-Deal estimator. Some results on restricted admissibility of the Graybill-Deal estimator are proved in Sinha-Mouqadem (1982). Sinha (1985) also worked on the unbiased estimation of the variance of the Graybill-Deal estimator of the common mean of several normal populations. Sequential estimation of common mean is discussed in Govindarajulu (1975), which covers the two-stage estimation as well. Many of the results are discussed in Ritcher (1960). Construction of Graybill-Deal estimator for common mean involves estimation of variance components to be plugged into the estimator with known variances. As far as the estimation of the variance components is concerned, the papers of Grubbs (1984), Khatri and Srivastava (1979) and of Das, Sinha and Sinha (1993) are worth mentioning. Sun Y (1999) in his PhD. dissertation considered the cost aspect while constructing the two-stage estimator of the common mean.

Considering two normal populations with unknown common mean μ and possibly unequal variances σ_1^2, σ_2^2, Sinha-Mouqadem (1982) considered four classes of unbiased estimators defined as follows:

$$C = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(s_1^2, s_2^2, D^2) \leq 1\}$$
$$C_0 = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(\tfrac{s_1^2}{s_2^2}) \leq 1\}$$
$$C_1 = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(s_1^2, s_2^2) \leq 1\}$$
$$C_2 = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(\tfrac{s_1^2}{D^2}, \tfrac{s_2^2}{D^2}) \leq 1\}$$

In the above, $\bar{X}$ refers to mean of X-sample, $D = \bar{Y} - \bar{X}$ where $\bar{Y}$ refers to mean of Y-sample, s_1^2 and s_2^2 are respectively unbiased variance estimators of the underlying X and Y populations.

Sinha-Mouqadem proved the admissibility of the Graybill-Deal Estimator in C_0 and the extended admissibility in C. They provided a class of admissible estimators in C_1, but produced a single admissible estimator in C_2. All the results in Sinha-Mouqadem (1982) are based on two normal samples of equal sample size n. In Chapter 1 we have extended all these results for two normal samples for the case of unequal sample sizes n_1 and n_2. We have developed a class of admissible estimators in C_2 as well which contains the estimator in Sinha-Mouqadem as a special case.

In Chapter 2 we have studied the performance (mainly by simulation results) of the Graybill-Deal Estimator for estimating the common mean of two univariate normal populations with unequal variances under a two-stage sampling scheme. In a two-stage sampling scheme a total of n observations are taken as follows: In the first stage, m observations are taken from each population, where $n > 2m$. The first-stage sample variances are denoted by s^2_{1m} and s^2_{2m} where $s^2_{1m} = \sum(X_i - \bar{X})^2/(m-1), s^2_{2m} = \sum(Y_i - \bar{Y})^2/(m-1)$. If $s^2_{1m} \leq s^2_{2m}$, then in the second stage, $n - 2m$ additional observations are taken from population 1; otherwise $n - 2m$ additional observations are taken from population 2. At the conclusion of the second stage sampling, the sample sizes are

$$N_1 = (n-m)\, I(R \leq 1) + m\, I(R > 1) \quad \text{and} \quad N_2 = n - N_1$$

where $R = s^2_{1m}/s^2_{2m}$ and I is the indicator function.

The two-stage estimators proposed by Richter (1960) are

$$\hat{\mu}_1 = I(R \leq 1)\; \bar{X}_{N_1} + I(R > 1)\; \bar{Y}_{N_2}$$

and

$$\hat{\mu}_2 = \frac{N_1}{N_1 + RN_2}\bar{X}_{N_1} + \frac{RN_2}{N_1 + RN_2}\bar{Y}_{N_2}.$$

It is interesting to note that $\hat{\mu}_2$ is not a proper Graybill-Deal estimator when $n > 2m$, because it includes sample variances based on the first stage data only. So $\hat{\mu}_2$ can possibly be improved by the two-stage Graybill-Deal estimator

$$\hat{\mu}_2^* = \frac{N_1}{N_1 + R^* N_2}\bar{X}_{N_1} + \frac{R^* N_2}{N_1 + R^* N_2}\bar{Y}_{N_2}$$

where $R^* = s^2_{1N_1}/s^2_{2N_2}$ is the ratio of the sample variances at the conclusion of second stage sampling; that is, it incorporates excess data in the second stage from the population with smaller first-stage sample variance. Notice that the Graybill-Deal estimator $\hat{\mu}_2^*$ does not take into account the nature of the sampling scheme; it simply imitates the estimator $\hat{\mu}_{GD}$ for the fixed sample size case. We have studied properties of $\hat{\mu}_2^*$ and compared its performance with the other two competitors listed above.

Further we compare various two-stage estimators with the Graybill-Deal estimator in terms of their variances. We provide a justification for the use of two-stage sampling over the one stage method, even when the second stage sampling cost per unit is (moderately) larger than that of the first stage. We also consider the problem of estimation of the variance of the Graybill-Deal estimator based on two-stage sampling. Finally, we discuss two modifications to the second stage sampling scheme :(1) use a fully sequential method and (2) include an indifference zone where equal observations are drawn from the two populations. It is observed that the performance of the Graybill-Deal estimator improves under the fully sequential procedure, but not under the inclusion of an indifference zone.

The estimation of the common mean of several normal populations with possibly different unknown variances was motivated by the estimation of treatment effect in a balanced incomplete block design (BIBD) with uncorrelated random block effects, by suitably combining the intra-block estimate and the interblock estimate. These estimates are derived under the usual assumptions of independence and normality with equal mean and possibly unequal variances (see Montgomery, 1991, pp. 184-186).

In all these studies, the populations are assumed to be normally distributed with identical means, while they are assumed heterogeneous with respect to variability. Typically, means are assumed identical because the populations came into existence from the same source, but the variances are assumed possibly unequal because the populations are exposed to different environmental or controlled conditions. For example, a soil engineer has theorized that samples from two different soils have the same mean because the soils were formed from the same geological phenomenon but have different variances because they were exposed to different meteorological conditions.

However, consider another example in which a chemical is split into several pools and sent to different laboratories for analysis. The laboratories may differ from one another not only in terms of their internal consistency (measured by standard deviation) due to instrument, method and/or human differences; but also in terms of their biases, departures from the true mean, due to location, time-lag, environmental conditions and/or technology. Thus the same environmental/controlled conditions which cause the population variances to differ from one another, are also likely to change the population means. In Chapter 3, we assume that the mean responses differ from a common underlying value (the true mean) by an extra term involving a linear function of one or more concomitant variable(s). In the presence of such covariate information, how shall one estimate the true normal mean based on data from multiple heterogeneous centers which differ in terms of covariate influences as well as unknown variances? This precisely is the problem studied in this chapter.

We illustrate the setup below. The linear models incorporate concomitant variable in the form of a linear regressor taking values over a continuum.

Suppose data on the response variable Y as well as on a covariate X are available from I different sources. We assume the following regression model with unequal error variances:

$$Y_{ij} = \mu + \gamma\, X_{ij} + \epsilon_{ij}; \quad j = 1, \ldots, n_i; \;\; i = 1, \ldots, I$$

where μ (the true mean response) and γ (the gradient on the regressor X) are unknown constants and ϵ_{ij}'s are independent random error variables distributed as $N(0, \sigma_i^2)$. Thus our model is an extension of the regression model with unknown variance components. The common mean may be estimated either by (i) a Graybill-Deal, or by (ii) a weighted least square estimator (WLS), provided we have first estimated the variance components.

In the process we have derived Grubbs' (1948) estimators, and also constructed Rao's (1970) Minimum Norm Quadratic Unbiased Estimators (MINQUE) with Invariance for the variance components for our model. We have presented estimation of the common population mean using these variance components estimates. Finally we have made a simulation study of the performance of these estimators of μ.

In the context of estimating the common mean of a bivariate normal distribution when the two component variables have different sampling costs per unit, the problem of allocating a fixed sampling budget among the first variable, the second variable and the bivariate data is also worth investigating. An interesting example involving differential costs of sampling from the two populations is found in Sun (1999). The Environment Protection Agency monitors gasoline quality based on Reid Vapor Pressure which is measured using two methods: a cheap and quick but crude measure obtained on site, and an expensive laboratory analysis yielding a measure of presumably higher precision. Sun (1999) studies the inference about μ by appropriately combining information from the bivariate data and additional data exclusively from X_1, but leaves open the question of optimum allocation of sampling budget to the two types of data.

For the known dispersion matrix case the optimal strategy can be developed. We show in Chapter 4 that the entire budget should be allocated to exactly one type of data depending on the relative values of the parameters relating to ratio of costs, ratio of standard deviations and the correlation coefficient. For the case of unknown dispersion matrix a two-stage strategy involving bivariate data in the first stage and the optimal type of data in the second stage was considered. At the conclusion of the second-stage sampling, the common mean was estimated either using the second-stage data only, or augmenting it with the appropriate part of the first-stage data. Standard errors of the estimators were obtained through a simulation study. Furthermore, the problem of choice of the optimal first-stage sample size has been addressed.

Part II: Estimation of the Size of a Finite Population

The problem of estimation of the population size (N) of a finite closed population is known to have great practical importance. Well known problems of this kind are the estimation of the total number of fish in a lake, the estimation of total number of wild animals in a forest etc. Several authors have already considered the problem in the past and suggested different methods of sampling with associated estimation procedures (see Boswell et al. (1988), Seber (1982) and the references therein). The basic procedure is to initially catch, mark and release k population units into the target population and then to recatch units randomly from the population in one or more samples.

For unbiased estimation of N, a simple procedure (to be called procedure I) is to recatch and release units one by one until m ($\leq k$) of k initially marked units are recaptured. If S_m denotes the number of trials required, then S_m follows a Negative Binomial distribution with success probability $\frac{k}{N}$ and the uniformly minimum variance unbiased estimator (UMVUE) of N is obtained from the well known results on Negative Binomial distribution (see Johnson and Kotz (1969), page 126) as $\hat{N}_I = \frac{kS_m}{m}$ with variance $V(\hat{N}_I) = \frac{N(N-k)}{m}$. Also the expected number of trials of the procedure is $ASN(I) = E(S_m) = \frac{mN}{k}$.

If units are sampled one by one without being replaced into the population until m ($\leq k$) of the k initially marked units are recaptured (to be called Procedure II), then S_m, the number of trials required, follows a Negative Hypergeometric distribution and in this case the UMVUE of N is given by $\hat{N}_{II} = \frac{(k+1)S_m}{m} - 1$ with $V(\hat{N}_{II}) = \frac{(N+1)(N-k)(k+1-m)}{m(k+2)}$ and $ASN(II) = E(S_m) = \frac{m(N+1)}{k+1}$ (see Johnson and Kotz (1969), page 157).

A simple modification of the Procedure I (to be called Procedure III) is also suggested in the literature as follows: initially k population units are marked and released into the target population and then units are sampled at random, marked and released one by one until m marked units are recaptured. The procedure is a special case of a more general procedure suggested in

Goodman (1953) and is termed as capture-mark-release-recapture (CMRR) sampling scheme. Using more general methods, Goodman (1953) obtained the UMVUE of N for this procedure as the quotient of two determinants and gave some simplified expression for $k = 1$. Darroch (1958) had shown that Goodman's estimator, for $k = 1$, can also be expressed as the ratio of two Sterling numbers or differences of zero. Hossain (1995) had considered the special case of $k = m = 1$ in which case the UMVUE of N is $\binom{S_1+1}{2}$, where S_1 is the number of trials required.

The purpose of Chapter 5 is to supplement these studies with various other results and to compare Procedure III with Procedures I and II in terms of the ASN and the variance of the UMVUE of N. It is demonstrated that the Procedure III is always better than the Procedure I and also appears to be better than Procedure II when N is considerably large.

The CMRR sampling scheme is basically a sequential estimation procedure. The sequential estimation of the population size can be analysed in a parallel framework to the sequential binomial estimation of p^{-1}. This itself has a vast literature including Gupta (1967), Sinha-Sinha (1975,) and Sinha-Bose (1985).

We recall the set-up of CMRR sampling for unbiased estimation of N, the (unknown) size of a finite closed population. In urn sampling terminology, we have initially an urn containing k black and N-k white balls where k is known and N is unknown. Balls are drawn at random one at a time, each time examining the colour of the ball drawn. At any stage, whenever a white ball is drawn, its colour is changed to black and is replaced into the urn. However, if a black ball is drawn it is replaced into the urn without any change of colour. Thus after each draw, whatever be the colour of the ball drawn, one black ball is replaced into the urn unless the sampling is stopped. The experiment provides dependent Bernoulli trials, the nature of dependence being manifested in the conditional probabilities of drawing black balls at various draws. All these probabilities are of course functions of N.

Borrowing ideas from the Bernoulli Sequential Estimation of p^{-1} under independent trials and using the notions of 'Closed' and 'Pushed-up' sampling plans, we have provided in Chapter 6 an unified approach to the problem of unbiased estimation of N, and in particular, have given a necessary and sufficient condition for unbiased estimability of N under an arbitrary stopping rule. Several examples have been provided in this regard.

PART I :
ESTIMATION OF COMMON MEAN OF NORMAL POPULATIONS

CHAPTER 1

ADMISSIBLE ESTIMATORS OF COMMON MEAN OF TWO UNIVARIATE NORMAL POPULATIONS

In this chapter we extend some admissibility results of Sinha and Mouqadem (*Commn. Stat. Theo. Meth.* **II**, 1982, 1603-1614) on the estimation of common mean of two independent univariate normal populations with unequal variances. We also present a new class of admissible estimators of the common mean within a subclass of unbiased estimators.

1. INTRODUCTION

Let

$$\Pi_i = N(\mu, \sigma_i^2), i = 1, 2 \tag{1.1}$$

be two univariate independent normal populations. The problem is to estimate the common mean μ. Let n_1 and n_2 be the sizes of two random samples from Π_1 and Π_2 respectively. The following notations are standard:

$$\begin{aligned} \bar{X} &= \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \\ \bar{Y} &= \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \\ s_1^2 &= \frac{1}{(n_1-1)}\sum_{i=1}^{n_1}(X_i - \bar{X}^2) \\ s_2^2 &= \frac{1}{(n_2-1)}\sum_{i=1}^{n_2}(Y_i - \bar{Y}^2) \\ D &= \bar{Y} - \bar{X} \end{aligned} \tag{1.2}$$

Here X_i's denote independent observations from Π_1 and Y_i's denote independent observations from Π_2.

When σ_1^2 and σ_2^2 are known $\frac{n_1\bar{X}}{\sigma_1^2} + \frac{n_2\bar{Y}}{\sigma_2^2}$ is sufficient for μ and the distribution of the sufficient statistic is complete. The UMVUE of μ is given by

$$\hat{\mu}_{UMVUE} = \frac{n_1\bar{X}/\sigma_1^2 + n_2\bar{Y}/\sigma_2^2}{n_1/\sigma_1^2 + n_2/\sigma_2^2} \tag{1.3}$$

When σ_1^2 and σ_2^2 are unknown but $\eta = \frac{\sigma_2^2}{\sigma_1^2}$ is known, $n_1\eta\bar{X} + n_2\bar{Y}$ is still sufficient for μ and the distribution of the sufficient statistic is complete. The UMVUE of μ is given by

$$\hat{\mu}_{UMVUE} = \frac{n_1\eta\bar{X} + n_2\bar{Y}}{n_1\eta + n_2} \tag{1.4}$$

This can be directly argued as follows. The class of error functions consists of (i) $\bar{X} - \bar{Y}$, (ii) contrasts involving X_i's and (iii) contrasts involving Y_i's. It is trivial to note that $n_1\eta\bar{X} + n_2\bar{Y}$ is uncorrelated with all the error functions listed above. Hence $\hat{\mu}$ in (1.4) is the UMVUE of μ.

When η is unknown $(\bar{X}, \bar{Y}, s_1^2, s_2^2)$ are jointly sufficient for μ but their joint distribution is not complete because E(D) = 0. The following Theorem [see Lehmann (1983) Page 132] shows that UMVUE of μ does not exist when η is unknown, or in other words, when σ_1^2 and σ_2^2 are unknown.

Theorem 1.1: When η is unknown the UMVUE of μ does not exist.

Proof: If possible let T be the UMVUE of μ in this case.
For a given $\eta_0 \in R$, $\hat{\mu}_{\eta_0} = \frac{n_1\eta_0\bar{X} + n_2\bar{Y}}{n_1\eta_0 + n_2}$ is the UMVUE of μ when $\eta = \eta_0$.
Since UMVUE is unique, $T = \hat{\mu}_{\eta_0}$ for any value of η_0 whatsoever. This is clearly a contradiction. So the UMVUE of the common mean does not exist. QED.

When η is unknown, one can consider the estimator

$$\hat{\mu}_0 = \frac{n_1\bar{X}/s_1^2 + n_2\bar{Y}/s_2^2}{n_1/s_1^2 + n_2/s_2^2} \tag{1.5}$$

which is basically a plug-in estimator. In literature $\hat{\mu}_0$ is referred to as the Graybill-Deal Estimator. $\hat{\mu}_0$ is unbiased since the conditional mean of $\hat{\mu}_0$ given s_1^2, s_2^2 and hence the unconditional mean is μ as $(\bar{X}, \bar{Y})$ is independent of (s_1^2, s_2^2) under normality. There are some results available in the literature on the properties of $\hat{\mu}_0$ or its variants in relation to other estimators [See Pal and Sinha (1996)]. As for example Graybill and Deal (1959) proved that the estimator $\hat{\mu}_0$ is uniformly better than the individual sample means $\bar{X}$ and $\bar{Y}$ if and only if $n_i \geq 11, \; i = 1, 2$.

Sinha and Mouqadem (1982) [abbreviated subsequently as SM (1982)] confined their attention to certain subclasses of unbiased estimators of the common mean and succeeded in producing admissible estimators within these subclasses. Our purpose is to extend their results and provide a class of admissible unbiased estimators.

In particular, SM(1982) considered the following classes of estimators of μ.

$$C = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(s_1^2, s_2^2, D^2) \leq 1\}$$

$$C_0 = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(\frac{s_2^2}{s_1^2}) \leq 1\}$$

$$C_1 = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(s_1^2, s_2^2) \leq 1\}$$

$$C_2 = \{\hat{\mu} : \hat{\mu} = \bar{X} + D\phi, 0 \leq \phi(\frac{s_1^2}{D^2}, \frac{s_2^2}{D^2}) \leq 1\} \tag{1.6}$$

It is not difficult to assert that the estimators proposed in (1.6) are all unbiased for μ. The proof is immediate for estimators belonging to C_0 and C_1 since D is independent of s_1^2 and s_2^2. It is less transparent for estimators belonging to C and C_2. Since $C_2 \subset C$, we indicate the proof for $\hat{\mu} \in C$. It is readily verified that $E(\hat{\mu}) = \mu \Leftrightarrow E(D\phi) = 0$. Next, we note that $E(D\phi) = E[\phi E(D|s_1^2, s_2^2, D^2)] = E[\phi E(D|D^2)]$. However, $E(D|D^2) = 0$ since $D \sim N(0, \sigma_D^2)$. Hence the assertion is justified.

The restriction to $0 \leq \phi \leq 1$ in (1.6) above is obvious as otherwise the estimators are trivially inadmissible. This can be seen as follows. First note that

$$Var(\hat{\mu}) = \frac{\sigma_1^2\sigma_2^2}{n_2\sigma_1^2 + n_1\sigma_2^2} + E\{D^2(\phi - \frac{\sigma_1^2}{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2})^2\}.$$

[For a proof of this assertion, see Lemma 2.1 below]. Therefore for general ϕ, defining ϕ^* as $\phi^* = \begin{cases} \phi, & 0 \leq \phi \leq 1 \\ 0, & \phi < 0 \\ 1, & \phi > 1 \end{cases}$,

we obtain $(\phi^* - \frac{\sigma_1^2}{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2})^2 \leq (\phi - \frac{\sigma_1^2}{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2})^2$ which implies $Var(\hat{\mu}_{\phi^*}) \leq Var(\hat{\mu}_\phi)$, thereby making $\hat{\mu}_\phi$ inadmissible.

It may be noted that the class C contains all the subclasses C_0, C_1 and C_2 and the subclass C_0 is contained in C_1 and C_2. Also the estimator $\hat{\mu}_0$

defined in (1.5) above belongs to *C₀*, as $\hat{\mu}_0$ can be expressed as $\bar{X} + D\phi_0$ where

$$\phi_0 = \frac{1}{1 + \frac{n_1 s_2^2}{n_2 s_1^2}} = \frac{1}{1 + \frac{n_1}{n_2} R}, \tag{1.7}$$

R being the sample variance ratio

$$R = \frac{s_2^2}{s_1^2} \tag{1.8}$$

Assuming the sample sizes are equal ($n_1 = n_2 = n > 2$ say) SM(1982) proved the admissibility of the Graybill-Deal Estimator $\hat{\mu}_0$ in C_0 and its extended admissibility in C. They also provided a class of admissible estimators in C_1 and a single admissible estimator in C_2. In section 2 and section 3 we extend these results for two normal samples of unequal sizes n_1 and n_2. In section 4 we develop a class of admissible estimators in C_2 which contains the estimator in SM(1982) as a particular case.

2. ADMISSIBILITY OF $\hat{\mu}_0$ IN C_0 AND ITS EXTENDED ADMISSIBILITY IN C FOR UNEQUAL SAMPLE SIZES

We begin with the following generalisations of the results of SM(1982) for unequal sample sizes.

Lemma 2.1. Let $\hat{\mu} = \bar{X} + D\phi \in C$. Then

$$Var(\hat{\mu}) = \frac{\sigma_1^2 \sigma_2^2}{n_2 \sigma_1^2 + n_1 \sigma_2^2} + E\{D^2(\phi - \frac{\sigma_1^2}{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2})^2\}. \tag{2.1}$$

Proof.

$$\begin{aligned} Var(\hat{\mu}) &= E(\hat{\mu} - \mu)^2 \\ &= E[(\bar{X} - \mu) + D\phi]^2 \\ &= \frac{\sigma_1^2}{n_1} + 2E[(\bar{X} - \mu)D\phi] + E[D^2\phi^2] \\ &= \frac{\sigma_1^2}{n_1} + 2E[\phi E\{DE(\bar{X} - \mu|D)|D^2\}] + E[D^2\phi^2] \end{aligned}$$

Note that $\bar{X} - \mu$ and D are jointly normally distributed and this yields $E[\bar{X} - \mu | D] = -\frac{\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2} D$. Then the above expression simplifies to

$$Var(\hat{\mu}) = \frac{\sigma_1^2}{n_1} - 2\frac{\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2} E[D^2\phi] + E[D^2\phi^2] \tag{2.2}$$

Next,

$$\begin{aligned}
\text{RHS of(2.1)} &= \frac{\sigma_1^2\sigma_2^2}{n_2\sigma_1^2 + n_1\sigma_2^2} + E\{D^2(\phi - \frac{\sigma_1^2}{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2})^2\} \\
&= \frac{\sigma_1^2\sigma_2^2}{n_2\sigma_1^2 + n_1\sigma_2^2} + E(D^2)\frac{\sigma_1^4}{(\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2)^2} \\
&\quad - \frac{2\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2} E(D^2\phi) + E(D^2\phi^2) \\
&= \text{Expression in (2.2)}
\end{aligned}$$

upon simplifications and using the fact that
$E(D^2) = Var(D) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. QED

We now proceed to prove the **admissibility of $\hat{\mu_0}$ in C_0** when the sample sizes are unequal. Note that when an estimator belongs to C_0, ϕ and D are independent. Using (2.1), for $\hat{\mu} \in C_0$, we obtain,

$$Var(\hat{\mu}) = \frac{\eta\sigma^2}{n_2(1 + \frac{n_1}{n_2}\eta)} + \frac{\sigma^2}{n_1}(1 + \frac{n_1}{n_2}\eta)E\{\phi(R) - \frac{1}{1 + \frac{n_1}{n_2}\eta}\}^2, \tag{2.3}$$

where, $\sigma^2 = \sigma_1^2, \eta = \frac{\sigma_2^2}{\sigma_1^2}$ and $R = \frac{s_2^2}{s_1^2}$, as defined in (1.8) and $\phi(R)$ indicates a function of R alone.

We will closely follow the arguments as in SM(1982). Our main result is given in Theorem 2.1. Prior to that we state two Lemmas which are versions of Lemma 2.2 and Lemma 2.3 in SM(1982) for unequal sample sizes. We omit the proofs which essentially follow along the same lines as in SM(1982).

Lemma 2.2. Suppose there exists H_0 on $(0, \infty)$ which is absolutely continuous with respect to the Lebesgue Measure and satisfies $H_0\{(a, b)\} > 0$ if $0 < a < b < \infty$ and

$$\int_0^\infty (1 + \frac{n_1}{n_2}\eta)E\{\phi(R) - \frac{1}{1 + \frac{n_1}{n_2}\eta}\}^2 H_0(d\eta) \tag{2.4}$$

is finite and minimum at ϕ_0. Then $\hat{\mu_0} = \bar{X} + D\phi_0$ is admissible in C_0.

Define,

$$I_i(r) = \int_0^\infty \eta^i f_\eta(r) H_0(d\eta),\ i = 0, 1, \tag{2.5}$$

where $f_\eta(r) = (\frac{\nu_2}{\eta\nu_1})^{\frac{1}{2}\nu_2} \frac{\Gamma(\frac{1}{2}(\nu_1+\nu_2))}{\Gamma(\frac{1}{2}\nu_1)\Gamma(\frac{1}{2}\nu_2)} \frac{r^{\frac{1}{2}\nu_2 - 1}}{(1+\frac{\nu_2 r}{\nu_1 \eta})^{\frac{1}{2}(\nu_1+\nu_2)}}$, $r > 0$. Note $f_\eta(r)$ is the density of $F = \frac{R}{\eta}$ with degrees of freedom ν_2, ν_1.

Lemma 2.3. Let H_0 satisfy $I_i(r) < \infty,\ i = 0, 1$ for $r > 0$ and

$$\int_0^\infty \frac{I_0(r) I_1(r)}{I_0(r) + \frac{n_1}{n_2} I_1(r)} dr < \infty, \int_0^\infty \frac{\eta}{1 + \frac{n_1}{n_2}\eta} H_0(d\eta) < \infty.$$

Then $\phi_0(r) = [1 + \frac{n_1}{n_2}\frac{I_1(r)}{I_0(r)}]^{-1}$ minimizes (2.4).

Theorem 2.1. For $n_1 + n_2 > 4, \hat{\mu_0}$ is admissible in C_0.

Proof. By (1.4) $\hat{\mu_0}$ corresponds to $\phi_0(r) = \frac{1}{1 + \frac{n_1}{n_2}\frac{(n_2-1)r}{n_1-1}}$. So for $\hat{\mu_0}$ to be admissible, by Lemma 2.3. we need to have $\frac{I_1(r)}{I_0(r)} = \frac{n_2-1}{n_1-1} r$. Putting $H_0(d\eta) = \eta^{p-1},\ \eta > 0,\ p \in R$, we get,

$$I_i(r) = \frac{\Gamma(\frac{1}{2}\nu_1 + i + p)\Gamma(\frac{1}{2}\nu_2 - i - p)}{\Gamma(\frac{1}{2}\nu_1)\Gamma(\frac{1}{2}\nu_2)} (\frac{\nu_2}{\nu_1})^{i+p} r^{i+p-1} \tag{2.6}$$

where $\nu_i = n_i - 1,\ i = 1, 2$. Now $\frac{I_1(r)}{I_0(r)} = \frac{n_1+2p-1}{n_2-2p-3}\frac{n_2-1}{n_1-1} r$. Hence $p = \frac{1}{4}(n_2 - n_1 - 2)$. For $I_0(r), I_1(r)$ to be defined we need $-\frac{1}{2}(n_1 - 1) < p < \frac{1}{2}(n_2 - 3)$, which implies $n_1 + n_2 > 4$. QED.

Next, we generalise Theorem (2.2) of SM(1982) for unequal sample sizes.

Theorem 2.2. For $n_1 > 4, \hat{\mu_0}$ is extended admissible in C.

Proof. For $\hat{\mu} \in C$, from (2.1), $Var(\hat{\mu}) = \frac{\sigma_1^2\sigma_2^2}{n_2\sigma_1^2 + n_1\sigma_2^2} + E[D^2(\phi - \frac{1}{1+\frac{n_1}{n_2}\eta})^2]$.

Simple calculations show,

$$E[D^2(\phi_0 - \frac{1}{1+\frac{n_1}{n_2}\eta})^2] \le (\frac{n_1-1}{n_2-1})^2 \frac{\sigma_1^2}{n_1(1+\frac{n_1}{n_2}\eta)} E(\frac{1-\frac{n_2-1}{n_1-1}F^*}{F^*})^2 \qquad (2.7)$$

where $F^* = \frac{1}{\eta}F \sim \chi^2_{n_2-1,n_1-1}$ and the expectation exists. Right hand side of (2.7) can be made as small as possible making σ_1 small and keeping η constant. Hence for any competitor $\hat{\mu}$ it is not possible to have $Var(\hat{\mu}) \le Var(\hat{\mu}_0)$ uniformly in $(\mu, \sigma_1^2, \sigma_2^2)$. This settles the claim. QED.

3. A CLASS OF ADMISSIBLE ESTIMATORS IN C_1 FOR UNEQUAL SAMPLE SIZES

The main result of this section is given in the following theorem which generalises a theorem of SM(1982) for unequal sample sizes.

Theorem 3.1. Any estimator of the form $\hat{\mu} = \bar{X} + D\phi$ where $\phi = \frac{\{(n_1-1)s_1^2+\lambda_1\}/n_1}{\{(n_2-1)s_1^2+\lambda_1\}/n_1+\{(n_2-1)s_2^2+\lambda_2\}/n_2}$, $\lambda_1 > 0, \lambda_2 > 0$ is admissible in *C1*.

Proof. Using (2.1) for $\hat{\mu} \in C_1$,

$$Var(\hat{\mu}) = \frac{\sigma_1^2\sigma_2^2}{n_2\sigma_1^2 + n_1\sigma_2^2} + \frac{1}{n_1}(\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2)E(\phi - \frac{\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2})^2 \qquad (3.1)$$

Following arguments of SM(1982)[(3.1)], given a probability measure H for (σ_1^2, σ_2^2), $\hat{\mu}_H = \bar{X} + D\phi_H$ is admissible in C_1 where

$$\phi_H(s_1^2, s_2^2) = \frac{E(\frac{\sigma_1^2}{n_1}|s_1^2, s_2^2)}{E(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}|s_1^2, s_2^2)}. \qquad (3.2)$$

Here the expression (3.2) refers to the posterior expectation of $\frac{\sigma_1^2/n_1}{\sigma_1^2/n_1+\sigma_2^2/n_2}$. Taking H as product of Inverted Gamma priors given by

$$H(d\sigma_1^2, d\sigma_2^2) \propto \prod_{i=1}^{2} e^{-\frac{\lambda_i}{2\sigma_i^2}} (\frac{1}{\sigma_i^2})^{p-1} d(\frac{1}{\sigma_i^2}),\ p > 0$$

we get the result by computing (3.2).The details are omitted. QED.

4. A CLASS OF ADMISSIBLE ESTIMATORS IN C_2

We begin with the following lemma which is an extension of a lemma by Zacks(1970) where it is proved for the case $n_1 = n_2 = n$.

Lemma 4.1. Let H be a prior assigning positive probability to any interval. Then $\hat{\mu}_H = \bar{X} + D\phi_H$ is admissible in C_2 where $\phi_H(\frac{s_1^2}{D^2}, \frac{s_2^2}{D^2}) = \phi_H^*(U, V) =$

$$\frac{\int_0^\infty \eta^{-\frac{1}{2}(n_2-1)}(1+\frac{n_1}{n_2}\eta)^{\frac{1}{2}(n_1+n_2)-1}[1+(1+\frac{n_1}{n_2}\eta)U+(\frac{1}{\eta})(1+\frac{n_1}{n_2}\eta)V]^{-\frac{1}{2}(n_1+n_2+1)}H(d\eta)}{\int_0^\infty \eta^{-\frac{1}{2}(n_2-1)}(1+\frac{n_1}{n_2}\eta)^{\frac{1}{2}(n_1+n_2)}[1+(1+\frac{n_1}{n_2}\eta)U+(\frac{1}{\eta})(1+\frac{n_1}{n_2}\eta)V]^{-\frac{1}{2}(n_1+n_2+1)}H(d\eta)} \tag{4.1}$$

where $U = \frac{s_1^2}{n_1 D^2}, V = \frac{s_2^2}{n_2 D^2}$.

Proof. For $\hat{\mu} \in C_2, \hat{\mu} = \bar{X} + D\phi(V_1, V_2)$, where $V_i = \frac{s_i^2}{D^2}$, $i = 1, 2$. Following the arguments of Lemma 2.1 and using

$$E[D^2|V_1, V_2] = \sigma^2(1-\frac{n_2-1}{n_1})(1+\frac{n_1}{n_2}\eta)[1+\frac{1}{n_1}(1+\frac{n_1}{n_2}\eta)V_1+\frac{1}{n_2\eta}(1+\frac{n_1}{n_2}\eta)V_2]^{-1}$$

we get

$$Var(\hat{\mu}) = \frac{\sigma^2}{n_1} + \sigma^2(1+\frac{n_2-1}{n_1})(1+\frac{n_1}{n_2}\eta)\ A$$

where

$$A = E[\{\phi^2 - 2(1+\frac{n_1}{n_2}\eta)^{-1}\phi\}\{1+\frac{1}{n_1}(1+\frac{n_1}{n_2}\eta)V_1+\frac{1}{n_1\eta}(1+\frac{n_1}{n_2}\eta)V_2\}^{-1}]$$

and $\sigma^2 = \sigma_1^2, \eta = \frac{\sigma_2^2}{\sigma_1^2}$. Given the prior H one calculates the prior risk function. Applying Fubini's Theorem, $\hat{\mu}$ is admissible if the corresponding ϕ minimizes

$$\int_0^\infty H(d\eta|V_1, V_2)[\phi - (1+\frac{n_1}{n_2}\eta)^{-1}]^2[(1+\frac{n_1}{n_2}\eta)^{-1} + \frac{1}{n_1}V_1 + \frac{1}{n_1\eta}V_2]^{-1}$$

where $H(\eta|V_1, V_2)$ designates the posterior distribution of η given V_1, V_2. See Zacks(1970). The minimizer is given by

$$\phi(V_1, V_2) = \frac{E_{\eta|V_1,V_2}[1+\frac{1}{n_1}(1+\frac{n_1}{n_2}\eta)V_1+\frac{1}{n_1\eta}(1+\frac{n_1}{n_2}\eta)V_2]^{-1}}{E_{\eta|V_1,V_2}(1+\frac{n_1}{n_2}\eta)[1+\frac{1}{n_1}(1+\frac{n_1}{n_2}\eta)V_1+\frac{1}{n_1\eta}(1+\frac{n_1}{n_2}\eta)V_2]^{-1}} \tag{4.2}$$

where $E_{\eta|V_1,V_2}$ designates the posterior expectation given (V_1, V_2).

Calculating the joint density of (U,V), we note that $H(d\eta|V_1, V_2) \propto H(d\eta|U, V) \propto$

$$\eta^{-\frac{1}{2}(n_2-1)}(1+\frac{n_1}{n_2}\eta)^{\frac{1}{2}(n_1+n_2-2)}[1+(1+\frac{n_1}{n_2}\eta)U+\frac{1}{\eta}(1+\frac{n_1}{n_2}\eta)V]^{-\frac{1}{2}(n_1+n_2-1)}H(d\eta) \tag{4.3}$$

Using (4.3) in (4.2) we get the result. QED.

We now develop the class of admissible estimators in C_2 by the following steps:

$$I = \int_0^\infty \frac{dx}{(ax^2+2bx+c)^{\frac{3}{2}}} = \frac{1}{\sqrt{c}(\sqrt{ac}+b)} \tag{4.4}$$

where $a \geq 0,\ c > 0,\ \sqrt{ac} + b > 0$. Vide, Gradstein and Ryshik (1980).

$$I_q = \int_0^\infty \frac{x^{n+q}dx}{(ax^2+2bx+c)^{n+\frac{3}{2}}} = \frac{(-1)^n 2^q}{(2n+1)!!}\frac{\partial^n}{\partial a^q \partial b^{n-q}}(I) \tag{4.5}$$

where $(2n+1)!! = 1.3.5\cdots(2n+1),\ q = 0, 1, 2, \cdots, n+1$. Expression (4.6) is obtained by taking derivatives of (4.5).

$$\frac{\partial^q}{\partial b^q}(I) = (-1)p!\frac{1}{\sqrt{c}(\sqrt{ac}+b)^{q+1}} \tag{4.6}$$

$q = 0, 1, 2, \cdots$.

$$I_0 = \frac{n!}{(2n+1)!!}\frac{1}{\sqrt{c}(\sqrt{ac}+b)^{n+1}} \tag{4.7}$$

$$I_1 = \frac{n!}{(2n+1)!!}\frac{1}{\sqrt{a}(\sqrt{ac}+b)^{n+1}} \tag{4.8}$$

$$I_2 = \frac{n!}{(2n+1)!!}\frac{\sqrt{ac}+(\sqrt{ac}+b)n^{-1}}{a^{\frac{3}{2}}(\sqrt{ac}+b)^{n+1}} \tag{4.9}$$

Theorem 4.1. $\hat{\mu} = \bar{X} + D\phi_\lambda,\ \lambda \geq 0$, is an admissible class of estimators in C_2 where

$$\phi_\lambda = [1 + \frac{n_1(I_1+\lambda I_2)}{n_2(I_0+\lambda I_1)}]^{-1}. \tag{4.10}$$

In the above, $n = \frac{n_1+n_2}{2} - 1,\ a = \frac{n_1}{n_2}U,\ 2b = 1 + U + \frac{n_1}{n_2}V,\ c = V,\ U = \frac{s_1^2}{n_1 D^2}$, and $V = \frac{s_2^2}{n_2 D^2}$, where I_0, I_1, and I_2 are defined in (4.7) - (4.9) respectively.

Proof. Let us consider the following family of priors for $\eta > 0$:

$$(1+\lambda\eta)\eta^{\frac{n_2}{2}-2}(1+\frac{n_1}{n_2}\eta)^{-\frac{n_1+n_2}{2}+1},\ \lambda \geq 0 \tag{4.11}$$

It is a family of proper priors essentially in the form of F-distribution since the F_{ν_1,ν_2} distribution has all the r moments, $r = 0, 1, \cdots, [\frac{\nu_2}{2}],\ \nu_2 > 4$. Vide for example, Johnson and Kotz (1994). Using this family of priors in (4.1) we get the result in the following steps.

$$\begin{aligned}
\text{Numerator of (4.1)} &= \int_0^\infty \frac{(1+\lambda\eta)\eta^{-\frac{3}{2}}}{[1+(\frac{n_1}{n_2}\eta)U + \frac{1}{\eta}(1+\frac{n_1}{n_2}\eta)V]^{\frac{n_1+n_2+1}{2}}} d\eta \\
&= \int_0^\infty \frac{(1+\lambda\eta)\eta^{n}}{[\frac{n_1}{n_2}U\eta^2 + (1+U+\frac{n_1}{n_2}V)\eta + V]^{n+\frac{3}{2}}} d\eta \\
&= \int_0^\infty \frac{(1+\lambda\eta)\eta^{n}}{[a\eta^2 + 2b\eta + c]^{n+\frac{3}{2}}} d\eta \\
&= I_0 + \lambda I_1
\end{aligned}$$

$$\begin{aligned}
\text{Denominator of (4.1)} &= \int_0^\infty \frac{(1+\lambda\eta)(1+\frac{n_1}{n_2}\eta)\eta^{-\frac{3}{2}}}{[1+(\frac{n_1}{n_2}\eta)U + \frac{1}{\eta}(1+\frac{n_1}{n_2}\eta)V]^{\frac{n_1+n_2+1}{2}}} d\eta \\
&= \int_0^\infty \frac{(1+\lambda\eta)(1+\frac{n_1}{n_2}\eta)\eta^{n}}{[\frac{n_1}{n_2}U\eta^2 + (1+U+\frac{n_1}{n_2}V)\eta + V]^{n+\frac{3}{2}}} d\eta \\
&= \int_0^\infty \frac{(1+\lambda\eta)(1+\frac{n_1}{n_2}\eta)\eta^{n}}{[a\eta^2 + 2b\eta + c]^{n+\frac{3}{2}}} d\eta \\
&= I_0 + (\lambda + \frac{n_1}{n_2})I_1 + \lambda\frac{n_1}{n_2}I_2
\end{aligned}$$

So from (4.1)

$$\begin{aligned}
\phi_\lambda &= \frac{I_0 + \lambda I_1}{I_0 + (\lambda + \frac{n_1}{n_2}) I_1 + \lambda \frac{n_1}{n_2} I_2} \\
&= [\frac{I_0 + (\lambda + \frac{n_1}{n_2}) I_1 + \lambda \frac{n_1}{n_2} I_2}{I_0 + \lambda I_1}]^{-1} \\
&= [1 + \frac{n_1(I_1 + \lambda I_2)}{n_2(I_0 + \lambda I_1)}]^{-1}
\end{aligned}$$

as is given in (4.10). QED.

In particular for $n_1 = n_2 = n$ say, the estimator in (4.10) reduces to $\hat{\mu}_\lambda = \bar{X} + D\phi_\lambda(\frac{s_1^2}{D^2}, \frac{s_2^2}{D^2})$ where

$$0 \leq \phi_\lambda(\frac{s_1^2}{D^2}, \frac{s_2^2}{D^2}) = [1 + \frac{\frac{s_2}{s_1} + \lambda \frac{s_2}{s_1}\{\frac{s_2}{s_1} + \frac{1}{2n}\frac{(s_2+s_2)^2+(n+1)D^2}{s_1^2}\}}{1 + \lambda \frac{s_2}{s_1}}]^{-1} \leq 1$$

This result has been reported in De(1995).

If further $\lambda = 0$, we get $\hat{\mu} = \bar{X} + D\frac{s_1}{s_1+s_2}$ and the admissibility of $\hat{\mu}$ in C_2 has been proved in SM(1982).

CHAPTER 2

ESTIMATION OF THE COMMON MEAN OF TWO UNIVARIATE NORMAL POPULATIONS USING TWO-STAGE SAMPLING

In this chapter we consider the problem of estimating the common mean of two univariate normal populations with unequal variances, under a two-stage sampling scheme and exhibit the superiority of Graybill-Deal estimator over a few other unbiased estimators. We show that two-stage sampling is preferable to one stage sampling even when the cost of sampling per unit is (moderately) larger in the second stage than in the first stage. Two possible extensions of the two-stage procedure have also been discussed.

1. INTRODUCTION

In Chapter 1, we considered the problem of estimation of common mean μ of two independent univariate normal populations distributed as $\mathrm{N}(\mu, \sigma_1^2)$ and $\mathrm{N}(\mu, \sigma_2^2)$, with μ, σ_1^2 and σ_2^2 unknown. For two random samples of fixed sizes n_1 and n_2 collected from these two populations, the Graybill-Deal Estimator of μ is given by

$$\hat{\mu}_{GD} = \frac{\frac{n_1}{s_1^2}\bar{X} + \frac{n_2}{s_2^2}\bar{Y}}{\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}} \tag{1.1}$$

where $\bar{X}, \bar{Y}, s_1^2, s_2^2$ are as defined in (1.2) of Chapter 1.

Note that the Graybill-Deal estimator is a weighted average of the sample means with weights proportional to the inverse of the squared standard errors of the respective sample means. With equal samples of size $n/2$ from each of the populations $\mathrm{N}(\mu, \sigma_1^2)$ and $\mathrm{N}(\mu, \sigma_2^2)$, the usual Graybill-Deal estimator simplifies to

$$\frac{s_2^2\bar{X} + s_1^2\bar{Y}}{s_1^2 + s_2^2} \tag{1.2}$$

We may refer to the setup described above as 'One-stage sampling' procedure with equal / unequal sample sizes. For this one-stage sampling procedure we have obtained in the preceding chapter some admissibility results for a few unbiased estimators of μ including the Graybill-Deal estimator $\hat{\mu}_{GD}$.

In this chapter we consider the problem of estimation of the common mean μ under a two-stage sampling scheme according to which a total of n observations is taken as follows: In the first stage, $m(> 1)$ observations are taken from each population, where $n > 2m$. The first-stage sample variances are denoted by s_{1m}^2 and s_{1m}^2 where $s_{1m}^2 = \sum(X_i - \bar{X})^2/(m-1), s_{2m}^2 = \sum(Y_i - \bar{Y})^2/(m-1)$. If $s_{1m}^2 \leq s_{2m}^2$, then in the second stage, $n - 2m$ additional observations are taken from population 1; otherwise $n - 2m$ additional observations are taken from population 2. At the conclusion of the second stage of sampling, the sample sizes are

$$N_1 = (n-m)\, I(R \leq 1) + m\, I(R > 1) \;\; \textit{and} \;\; N_2 = n - N_1$$

where $R = s_{1m}^2/s_{2m}^2$ and I is the indicator function.

Two-stage estimators of the common mean were introduced by Richter (1960). Govindarajulu (1975, section 4) studied their relative efficiencies (exact and asymptotic risks). Among other things, it has been shown that the simple average of the two sample means as an estimator for the common mean can be improved upon by using a two-stage estimator. However, other sequential aspects of the problem have not yet been fully explored.

In this chapter we compare various two-stage estimators with the Graybill-Deal estimator in terms of their variances and establish the superiority of the Graybill-Deal estimator compared to others. We also demonstrate that the two-stage sampling provides a better estimator of μ than under one-stage sampling even when the second stage sampling cost per unit is (moderately) larger than that of the first stage. We further address the problem of estimation of the variance of the Graybill-Deal estimator based on two-stage sampling. We discuss two modifications of the two-stage sampling scheme, viz. (1) use of a fully sequential method and (2) inclusion of an indifference zone where equal observations are drawn from the two populations.

2. TWO-STAGE ESTIMATORS OF COMMON MEAN

The two-stage estimators proposed by Richter (1960) are

$$\hat{\mu}_1 = I(R \leq 1)\, \bar{X}_{N_1} + I(R > 1)\, \bar{Y}_{N_2} \tag{2.1}$$

and

$$\hat{\mu}_2 = \frac{N_1}{N_1 + RN_2} \bar{X}_{N_1} + \frac{RN_2}{N_1 + RN_2} \bar{Y}_{N_2} \tag{2.2}$$

It may be noted that $\hat{\mu}_2$ is not a proper Graybill-Deal estimator when $n > 2m$, because it includes sample variances based on the first stage data only. So $\hat{\mu}_2$ can possibly be improved by the two-stage Graybill-Deal estimator

$$\hat{\mu}_2^* = \frac{N_1}{N_1 + R^* N_2}\bar{X}_{N_1} + \frac{R^* N_2}{N_1 + R^* N_2}\bar{Y}_{N_2} \tag{2.3}$$

where $R^* = s_{1N_1}^2 / s_{2N_2}^2$ is the ratio of the sample variances at the conclusion of second stage sampling; that is, it incorporates excess data in the second stage from the population with smaller first-stage sample variance. Notice that the Graybill-Deal estimator $\hat{\mu}_2^*$ does not take into account the nature of the sampling scheme, it simply imitates the estimator $\hat{\mu}_{GD}$ in (1.1) for the fixed sample size case.

By taking iterated expectation (conditioning on R) it is straightforward to see that $\bar{X}_{N_1}, \bar{Y}_{N_2}$ $\hat{\mu}_1, \hat{\mu}_2$ and $\hat{\mu}_2^*$ are all unbiased estimators for μ. However, the evaluation of their variances poses varying degrees of difficulty as described below. Exact expressions are available for the variances of $\bar{X}_{N_1}, \bar{Y}_{N_2}$ and $\hat{\mu}_1$. The variance of $\hat{\mu}_2$ can be expressed as an integral, and therefore, can be evaluated numerically. But the variance of $\hat{\mu}_2^*$ can at best be expressed as a triple integral, and therefore, is difficult to evaluate.

Suppose that m denotes the first stage sample size for each population and $n = 2m + k$ is the total sample size. Defining $\eta \equiv \sigma_2^2/\sigma_1^2$, $g = g(\eta) \equiv P(F_{m-1,m-1} < \eta)$ and $A \equiv \{R \leq 1\}$ we have (by conditioning on R)

$$V(\bar{X}_{N_1}) = E\left[\frac{\sigma_1^2}{N_1}\right] = \sigma_1^2\left[\frac{g}{m+k} + \frac{1-g}{m}\right], \tag{2.4}$$

$$V(\bar{Y}_{N_2}) = E\left[\frac{\sigma_2^2}{N_2}\right] = \sigma_2^2\left[\frac{g}{m} + \frac{1-g}{m+k}\right] \tag{2.5}$$

and

$$V(\hat{\mu}_1) = E\left[\frac{\sigma_1^2}{N_1}I_A + \frac{\sigma_2^2}{N_2}I_{A^c}\right] = \sigma_1^2\left[\frac{g + \eta(1-g)}{m+k}\right]. \tag{2.6}$$

Along the same line of reasoning we also have

$$V(\hat{\mu}_2) = E\left[\frac{N_1\sigma_1^2 + R^2 N_2 \sigma_2^2}{(N_1 + R\, N_2)^2}\right]. \tag{2.7}$$

To evaluate (2.7), define $U = (m-1)s_{1m}^2/\sigma_1^2$ and $V = (m-1)s_{2m}^2/\sigma_2^2$. Then $\eta R = U/V$, being the ratio of two independent chi-square variables with degrees of freedom $m-1$ each, has an $F(m-1, m-1)$ distribution. Furthermore, $A = \{R \leq 1\} = \{U \leq \eta V\}$. Hence, from (2.7) we have

$$\begin{aligned} V(\hat{\mu}_2) &= \sigma_1^2 \, E\left[\frac{(m+k) + m\eta^{-1}(\frac{U}{V})^2}{\{(m+k) + m\eta^{-1}\frac{U}{V}\}^2} \, I_A + \frac{m + (m+k)\eta^{-1}(\frac{U}{V})^2}{\{m + (m+k)\eta^{-1}\frac{U}{V}\}^2} \, I_{A^c}\right] \\ &= \sigma_1^2 \int_0^\eta \frac{(m+k) + m\eta^{-1}z^2}{\{(m+k) + m\eta^{-1}z\}^2} \, f(z)\, dz \\ &\quad + \sigma_1^2 \int_\eta^\infty \frac{m + (m+k)\eta^{-1}z^2}{\{m + (m+k)\eta^{-1}z\}^2} \, f(z)\, dz \end{aligned}$$

where f(z) is the density of an $F(m-1, m-1)$ distribution. Thus, V($\hat{\mu}_2$) can either be evaluated by numerical integration, or be estimated by simulation. In fact, we have done both and presented the results in Table 2.1 below. The close agreement between the simulated values and numerically integrated values of $\hat{\mu}_2$ lends support to the reliability of simulated values based on 1,000 repetitions.

Finally, to evaluate V($\hat{\mu}_2^*$), define $W_1 = \frac{1}{\sigma_1^2}(N_1 - 1)s_{1N_1}^2 - U$ and $W_2 = \frac{1}{\sigma_2^2}(N_2 - 1)s_{2N_2}^2 - V$.

Then, conditional on $N_1 (> m)$, $W_1 \sim \chi^2_{N_1 - m}$ and W_1 is independent of U. This can be established by referring to an orthogonal transformation $\mathbf{x}^* = Q\mathbf{x}$ from $\mathbf{x} = (x_1, \cdots, x_{N_1})$ to $\mathbf{x}^* = (x_1^*, \cdots, x_{N_1}^*)$ using the orthogonal matrix $Q = \begin{pmatrix} \frac{1}{\sqrt{N_1}} & \cdots & \frac{1}{\sqrt{N_1}} \\ P & | & O \\ & R & \end{pmatrix}$ where $\begin{pmatrix} \frac{1}{\sqrt{m}} & \cdots & \frac{1}{\sqrt{m}} \\ & P & \\ & & \end{pmatrix}$ is an $m \times m$ orthogonal matrix, O denotes a null matrix of order $m \times (N_1 - m)$ and R represents the lower submatrix of Q of order $(N_1 - m) \times N_1$. It is now evident that $W_1 = \frac{1}{\sigma_1^2} \sum_{m+1}^{N_1} x_i^{*2}$. The rest of the argument is routine.

Similarly, it also follows that conditional on $N_2 (> m)$, $W_2 \sim \chi^2_{N_2 - m}$ and W_2 is independent of V.

Note that

$$\eta R^* = \frac{m-1}{m+k-1} \frac{U + W_1}{V} \, I_A + \frac{m+k-1}{m-1} \frac{U}{V + W_2} \, I_{A^c}$$

and (by conditioning on U, V, W_1 and W_2),

$$V(\hat{\mu}_2^*) = E\left[\frac{N_1\sigma_1^2 + R^{*2} N_2 \sigma_2^2}{(N_1 + R^* \, N_2)^2}\right] \tag{2.8}$$

which can be expressed as

$$V(\hat{\mu}_2^*) = \sigma_1^2 \, E\left[\frac{(m+k) + m\eta^{-1}\left\{\frac{(m-1)(U+W_1)}{(m+k-1)V}\right\}^2}{\{(m+k) + m\eta^{-1}\frac{(m-1)(U+W_1)}{(m+k-1)V}\}^2}\, I_A\right] + \sigma_1^2 \, E\left[\frac{m + (m+k)\eta^{-1}\left\{\frac{(m+k-1)U}{(m-1)(V+W_2)}\right\}^2}{\{m + (m+k)\eta^{-1}\frac{(m+k-1)U}{(m-1)(V+W_2)}\}^2}\, I_{A^c}\right] \qquad (2.9)$$

The evaluation of (2.9) rests on the joint distribution (U, V, W_1) over $\{I_A = 1\}$ and on the joint distribution of (U, V, W_2) over $\{I_{A^c} = 1\}$. In other words, the evaluation of V($\hat{\mu}_2^*$) involves a triple integral. Hence we have estimated V($\hat{\mu}_2^*$) by simulation based on 1,000 repetitions. The results, given in Table 2.1 below, show that the variance of $\hat{\mu}_2^*$ is smaller than that of $\bar{X}_{N_1}, \bar{Y}_{N_2}, \hat{\mu}_1$ *or* $\hat{\mu}_2$, for any value of k.

It should be pointed out that for each of the above-mentioned estimators, if η is replaced by η^{-1} then the variance is multiplied by η^{-1}. Therefore, we may restrict our attention to $\eta \geq 1$. The numerical integration was done using MAPLE V and all simulations were done using FORTRAN on VAX.

Table 2.1

Variance of two-stage estimators of μ. Here m=5, $\sigma_1^2 = 1, \sigma_2^2 = \eta$

	k	exact $V(\bar{X}_{N_1})$	exact $V(\bar{Y}_{N_2})$	exact V($\hat{\mu}_1$)	integrated $V(\hat{\mu}_2)$	simulated $V(\hat{\mu}_2)$	simulated $V(\hat{\mu}_2^*)$
$\eta = 1$	0	.200000	.200000	.200000	.120000	.120865	†
	1	.183333	.183333	.183333	.107028	.107347	.104512
	2	.171429	.171429	.171429	.096660	.096672	.093730
	3	.162500	.162500	.162500	.088101	.088233	.085044
	4	.155556	.155556	.155556	.080939	.081371	.078284
	5	.150000	.150000	.150000	.074857	.075092	.073178
	6	.145455	.145455	.145455	.069628	.069636	.067826
	7	.141667	.141667	.141667	.065083	.065307	.063395
	8	.138462	.138462	.138462	.061096	.061034	.059880
	9	.135714	.135714	.135714	.057570	.057542	.056520
	10	.133333	.133333	.133333	.054430	.054539	.053020

† When k=0, $\hat{\mu}_2^* = \hat{\mu}_2$.

Table 2.1 (Contd.)

Variance of two-stage estimators of μ. Here m=5, $\sigma_1^2 = 1, \sigma_2^2 = \eta$

	k	exact $V(\bar{X}_{N_1})$	exact $V(\bar{Y}_{N_2})$	exact $\mathrm{V}(\hat{\mu}_1)$	integrated $V(\hat{\mu}_2)$	simulated $V(\hat{\mu}_2)$	simulated $V(\hat{\mu}_2^*)$
$\eta = 2$	0	.200000	.400000	.251852	.160810	.160471	†
	1	.175309	.382716	.227160	.142333	.142109	.138100
	2	.157672	.370370	.209524	.127707	.126596	.122303
	3	.144444	.361111	.196296	.115840	.118453	.111627
	4	.134156	.353909	.186008	.106500	.106034	.102885
	5	.125926	.348148	.177778	.097707	.098171	.093363
	6	.119192	.343434	.171044	.091875	.091128	.086758
	7	.113580	.339506	.165432	.084513	.084115	.080267
	8	.108832	.336182	.160684	.079166	.079782	.075556
	9	.104762	.333333	.156614	.074467	.075066	.070901
	10	.101235	.330864	.153086	.070296	.071562	.067586
$\eta = 5$	0	.200000	1.000000	.444444	.202984	.203672	†
	1	.169136	.987657	.413580	.175161	.174743	.168441
	2	.147090	.978842	.391534	.154104	.153755	.148847
	3	.130556	.972224	.375000	.137598	.140368	.129272
	4	.117695	.967081	.362140	.124305	.122137	.117817
	5	.107407	.962963	.351852	.113366	.111843	.104952
	6	.098990	.959602	.343434	.104204	.102336	.095850
	7	.091975	.956798	.336420	.096418	.095999	.091158
	8	.086040	.954423	.330484	.089718	.088899	.083712
	9	.080952	.952381	.325397	.083892	.087712	.077787
	10	.076543	.950623	.320988	.078778	.076433	.072670

† When k=0, $\hat{\mu}_2^* = \hat{\mu}_2$.

3. AN ALTERNATIVE GRAYBILL-DEAL TYPE ESTIMATOR

For two-stage sampling we can define alternative Graybill-Deal type estimator as follows. Suppose first that the variance ratio $\eta = \sigma_2^2/\sigma_1^2$ is known.

Then for samples of fixed sizes N_1 and N_2 from the two populations the UMVUE is given by

$$\hat{\hat{\mu}} = \frac{\frac{N_1}{\sigma_1^2}\bar{X}_{N_1} + \frac{N_2}{\sigma_2^2}\bar{Y}_{N_2}}{\frac{N_1}{\sigma_1^2} + \frac{N_2}{\sigma_2^2}} = \frac{\eta N_1 \bar{X}_{N_1} + N_2 \bar{Y}_{N_2}}{\eta N_1 + N_2}. \tag{3.1}$$

This expression $\hat{\hat{\mu}}$ is identical to $\hat{\mu}_{UMVUE}$ in (1.4) of Chapter I. Here $\hat{\hat{\mu}}$ can be interpreted as a linear combination of $\bar{X}_{N_1}$ and $\bar{Y}_{N_2}$ with weights proportional to the inverse of their respective variances. In two-stage sampling N_1, N_2 are random but the estimator $\hat{\hat{\mu}}$ is still unbiased which may be seen by taking conditional expectation given N_1, N_2. However, in two-stage sampling scheme, one may propose an alternative unbiased estimator, which has minimum variance in the class of all linear combinations of $\bar{X}_{N_1}$ and $\bar{Y}_{N_2}$ with non-random weights, as follows.

Let

$$\tilde{\tilde{\mu}} = \frac{\bar{\sigma}_1^{-2}\ \bar{X}_{N_1} + \bar{\sigma}_2^{-2}\ \bar{Y}_{N_2}}{\bar{\sigma}_1^{-2} + \bar{\sigma}_2^{-2}} = \frac{\bar{\sigma}_2^2\ \bar{X}_{N_1} + \bar{\sigma}_1^2\ \bar{Y}_{N_2}}{\bar{\sigma}_2^2 + \bar{\sigma}_1^2} \tag{3.2}$$

where $\bar{\sigma}_1^2 = V(\bar{X}_{N_1})$ and $\bar{\sigma}_2^2 = V(\bar{Y}_{N_2})$. Note that, under two-stage sampling, these variances are no longer σ_i^2/N_i, i=1,2. The exact expressions for these variances are given in (2.4) and (2.5). Thus $\tilde{\tilde{\mu}}$ may be written as

$$\tilde{\tilde{\mu}} = \frac{\eta(\frac{g}{m} + \frac{1-g}{m+k})\bar{X}_{N_1} + (\frac{g}{m+k} + \frac{1-g}{m})\bar{Y}_{N_2}}{\eta(\frac{g}{m} + \frac{1-g}{m+k}) + (\frac{g}{m+k} + \frac{1-g}{m})} \tag{3.3}$$

where g=g(η)=P($F_{m-1,m-1} < \eta$).

Notice that in two-stage scenario, the estimator $\hat{\hat{\mu}}$ in (3.1) is also a linear combination of $\bar{X}_{N_1}$ and $\bar{Y}_{N_2}$, but with random weights. When η is known, it is evident that the estimator $\hat{\hat{\mu}}$ is better than the alternative estimator $\tilde{\tilde{\mu}}$ given in (3.3). We provide a proof for the sake of completeness.

Theorem 3.1 For known η, the estimator $\hat{\hat{\mu}}$ in (3.1) has smaller variance than the estimator $\tilde{\tilde{\mu}}$ in (3.3).

Proof. First note that both $\hat{\hat{\mu}}$ and $\tilde{\tilde{\mu}}$ are unbiased estimators of μ. Next notice that their variances are

$$\begin{aligned} V(\hat{\mu}) &= E\left[\frac{\eta^2 N_1\sigma_1^2 + N_2\sigma_2^2}{(\eta N_1 + N_2)^2}\right] = E\left[\frac{\eta\sigma_1^2}{\eta N_1 + N_2}\right] \\ &= \eta\sigma_1^2\left[\frac{g}{\eta(m+k)+m} + \frac{1-g}{\eta m + (m+k)}\right] \\ &= \eta\sigma_1^2\frac{[\eta(m+k)+m+(1-\eta)gk]}{[\eta(m+k)+m]\,[\eta m+(m+k)]} \end{aligned}$$

and, using (2.4) and (2.5),

$$\begin{aligned} V(\tilde{\tilde{\mu}}) &= \frac{\eta^2\,(\frac{g}{m}+\frac{1-g}{m+k})^2\,\sigma_1^2(\frac{g}{m+k}+\frac{1-g}{m}) + (\frac{g}{m+k}+\frac{1-g}{m})^2\,\sigma_2^2\,(\frac{g}{m}+\frac{1-g}{m+k})}{[\eta(\frac{g}{m}+\frac{1-g}{m+k}) + (\frac{g}{m+k}+\frac{1-g}{m})]^2} \\ &= \eta\sigma_1^2\,\frac{(\frac{g}{m}+\frac{1-g}{m+k})\,(\frac{g}{m+k}+\frac{1-g}{m})}{\eta(\frac{g}{m}+\frac{1-g}{m+k}) + (\frac{g}{m+k}+\frac{1-g}{m})} \\ &= \eta\sigma_1^2\,\frac{(m+gk)(m+k-gk)}{m(m+k)[\eta(m+gk)+(m+k-gk)]}. \end{aligned}$$

Hence, $V(\hat{\mu}) < V(\tilde{\tilde{\mu}})$ if and only if

$$\begin{aligned} &\frac{[\eta(m+k)+m+(1-\eta)gk]\,[\eta m+(m+k)-(1-\eta)gk]}{[\eta(m+k)+m]\,[\eta m+(m+k)]} \\ < \;&\frac{(m+gk)(m+k-gk)}{m(m+k)} \end{aligned}$$

or equivalently, (subtracting 1 from both sides), if and only if

$$\frac{(1-\eta)^2\,g(1-g)k^2}{(1+\eta)^2(m+k)m+\eta k^2} < \frac{g(1-g)k^2}{m(m+k)}.$$

Since $0 < g < 1$, this condition reduces to

$$(1-\eta)^2 m(m+k) < (1+\eta)^2 m(m+k) + \eta k^2$$

which is clearly true for all $\eta > 0$. QED.

Now consider the case when η is unknown. In this case, one may replace η by $\hat{\eta} = R^{*-1}$ and g by $\hat{g} = P[F_{m-1,m-1} < \hat{\eta}]$ and the estimators in (3.1) and (3.3) are modified to,

$$\hat{\hat{\mu}}^* = \frac{\hat{\eta}N_1\bar{X}_{N_1} + N_2\bar{Y}_{N_2}}{\hat{\eta}N_1 + N_2} = \frac{N_1\bar{X}_{N_1} + R^*N_2\bar{Y}_{N_2}}{N_1 + R^*N_2} \tag{4.4}$$

which is the same as the Graybill-Deal estimator $\hat{\mu}_2^*$ in (2.5), and

$$\tilde{\tilde{\mu}}^* = \frac{\hat{\eta}(\frac{\hat{g}}{m} + \frac{1-\hat{g}}{m+k})\bar{X}_{N_1} + (\frac{\hat{g}}{m+k} + \frac{1-\hat{g}}{m})\bar{Y}_{N_2}}{\hat{\eta}(\frac{\hat{g}}{m} + \frac{1-\hat{g}}{m+k}) + (\frac{\hat{g}}{m+k} + \frac{1-\hat{g}}{m})}. \tag{4.5}$$

As in section 3, let U, V, W_1 and W_2 be distributed as independent chi-square variables with degrees of freedom m–1, m–1, k and k respectively. Then

$$\hat{\eta} = R^{*-1} = \eta\left[\frac{m+k-1}{m-1}\frac{V}{U+W_1}\ I_A + \frac{m-1}{m+k-1}\frac{V+W_2}{U}\ I_{A^c}\right]$$

and (by conditioning on U, V, W_1 and W_2), V($\tilde{\tilde{\mu}}^*$) can be expressed as

$$\begin{aligned} V(\tilde{\tilde{\mu}}^*) \quad &= \quad \sigma_1^2\ E\left[\frac{\hat{\eta}^2(\frac{\hat{g}}{m} + \frac{1-\hat{g}}{m+k})^2/(m+k) + \eta(\frac{\hat{g}}{m+k} + \frac{1-\hat{g}}{m})^2/m}{\{\hat{\eta}(\frac{\hat{g}}{m} + \frac{1-\hat{g}}{m+k}) + (\frac{\hat{g}}{m+k} + \frac{1-\hat{g}}{m})\}^2}\ I_A\right] \\ &\quad + \sigma_1^2\ E\left[\frac{\hat{\eta}^2(\frac{\hat{g}}{m} + \frac{1-\hat{g}}{m+k})^2/m + \eta(\frac{\hat{g}}{m+k} + \frac{1-\hat{g}}{m})^2/(m+k)}{\{\hat{\eta}(\frac{\hat{g}}{m} + \frac{1-\hat{g}}{m+k}) + (\frac{\hat{g}}{m+k} + \frac{1-\hat{g}}{m})\}^2}\ I_{A^c}\right]. \end{aligned}$$

This time the comparison between $V(\hat{\hat{\mu}}^*)$ and $V(\tilde{\tilde{\mu}}^*)$ is not evident. We have carried out simulation towards this effect. The simulation results in Tables 3.1 and 3.2 indicate that the Graybill-Deal estimator $\hat{\hat{\mu}}^*$ performs better than the alternative Graybill-Deal type estimator $\tilde{\tilde{\mu}}^*$ for all η unknown. Furthermore, the results in Table 3.2 also indicate that, if one intends to utilize the Graybill-Deal estimator, then for a given total sample size n, the choice of m = 5 for the first-stage sample size is pretty good. We shall address the question of the optimal choice of m further in section 5.

Table 3.1

($10^5\times$) Variance of two-stage estimators with m=5, $\sigma_1^2 = 1, \sigma_2^2 = \eta$

	$\eta = 1$		$\eta = 2$		$\eta = 5$	
k	V($\hat{\hat{\mu}}^*$)	V($\tilde{\tilde{\mu}}^*$)	V($\hat{\hat{\mu}}^*$)	V($\tilde{\tilde{\mu}}^*$)	V($\hat{\hat{\mu}}^*$)	V($\tilde{\tilde{\mu}}^*$)
1	10451	10475	13810	13928	16844	17164
2	9373	9428	12230	12478	14885	15006
3	8504	8660	11163	11373	12927	13359
4	7828	8148	10288	10615	11782	12228
5	7318	7610	9336	10020	10495	11432
6	6783	7358	8676	9521	9585	10794
7	6339	7108	8027	9182	9116	10095
8	5988	6815	7556	8957	8371	9565
9	5652	6627	7090	8430	7779	8976
10	5302	6483	6759	8255	7267	8752

Table 3.2

($10^5\times$) Variance of two-stage estimators with $\sigma_1^2 = 1, \sigma_2^2 = \eta$
for various m and k

			$\eta = 1$		$\eta = 2$		$\eta = 5$	
n	m	k	V($\hat{\hat{\mu}}^*$)	V($\tilde{\tilde{\mu}}^*$)	V($\hat{\hat{\mu}}^*$)	V($\tilde{\tilde{\mu}}^*$)	V($\hat{\hat{\mu}}^*$)	V($\tilde{\tilde{\mu}}^*$)
20	5	10	5318	6468	6735	8282	7441	8684
30	5	20	3484	5680	4411	7108	4619	6569
30	10	10	3491	3670	4381	4599	4786	4959
40	5	30	2580	5486	3318	6421	3305	5953
40	10	20	2589	3132	3136	3652	3275	3395
40	15	10	2599	2640	3265	3337	3690	3697
50	5	40	2061	5317	2586	6438	2696	5272
50	10	30	2052	2823	2459	3168	2478	2681
50	15	20	2057	2257	2512	2687	2707	2719
50	20	10	2068	2079	2615	2644	3015	3021

4. CHOICE OF INITIAL SAMPLE SIZE AND COMPARISON WITH SINGLE STAGE PROCEDURE

In this section we consider the problem of choice of the initial sample size m and a comparison of the two-stage procedure with the single-stage procedure. It is anticipated that β, the second stage sampling cost per unit, is larger than that of the first stage cost which is taken to be unity. Clearly, if $\beta < 1$, setting aside a larger sample size for samples to be collected in the second stage looks appealing. Even when β is moderately larger than 1, two-stage sampling may be justified rather than sampling in one-stage. Of particular interest to us is the largest value of β for which it is still beneficial to conduct sampling in two stages.

Let $\alpha(=\frac{n-2m}{n})$ denote the proportion of the total sample size that we plan to collect in the second stage. Then $\alpha = 0$ corresponds to one-stage sampling scheme.

In Table 4.1, we display the numerical computations for variance of G-D estimator $\hat{\mu}_2^*$ in (2.9) for selected values of $\eta(=\frac{\sigma_2^2}{\sigma_1^*})$ and α. The bold figures in the table represent least values of the variance for each selected value of η. The corresponding value of α may be denoted by α_{opt}. It turns out that α_{opt} roughly ranges between 0.4 and 0.7. In case $\beta = 1$, the two stages of sampling are equally costly and, hence, α_{opt} indicates optimal splitting of the total sample size in two stages.

Table 4.1

Simulated ($10^5\times$) variance of the Graybill-Deal estimator for n=50 and various α and η

α	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 5$	$\eta = 10$
0.0	2095	2794	3186	3244	3295	3668
0.1	2109	2639	3317	3034	3201	3226
0.2	2078	2582	3191	2859	3133	3491
0.3	2468	2632	2824	2705	2777	3189
0.4	2976	**2522**	2820	2567	2462	2599
0.5	2104	2536	2592	2598	2607	2720
0.6	**2045**	2529	2721	**2505**	2572	2585
0.7	2136	2539	**2580**	2658	**2460**	**2462**
0.8	2210	2670	3364	2958	2837	2911
0.9	3118	4504	5067	5412	6020	8334

If, however $\beta > 1$, then there is a differential cost in two stages and the total cost will be expressed as $n[1 + (\beta - 1)\alpha]$. Therefore efficiency per unit cost is given by $\frac{1/V(\hat{\mu}_2^*)}{n[1+(\beta-1)\alpha]}$ and this is maximised when $n[1 + (\beta - 1)\alpha]V(\hat{\mu}_2^*)$ is minimised for given n, the total sample size. Once more, we calculate α_{opt}, but this time for selected values of η and β. These values are displayed in the upper block of Table 4.2. The resulting variance values are denoted by $V_{\alpha_{opt}(\beta,\eta)} = V_{II,\alpha_{opt}(\beta,\eta)}$ say. It may be noted that
(a) for fixed η, α_{opt} decreases as β increases.
(b) $V_{II,\alpha_{opt}(\beta,\eta)}$ is increasing in β for every fixed η.

For every value of $\eta > 1$, we determine the maximum value of β so that the corresponding α_{opt} is positive and the proportional reduction in the variance of the Graybill-Deal estimator (compared against the one stage sampling scheme) is approximately some prespecified quantity γ (say, $\gamma = .05$ or $\gamma = .01$). To be specific, let for fixed n,

$$V_I \equiv V(\hat{\mu}_2^*) \text{ under one stage sampling}$$

and

$$V_{II}(\beta) \equiv V_{II,\alpha_{opt}(\beta,\eta)}$$

Then we obtain β_{max} such that $\alpha_{opt}(\beta_{max}, \eta) > 0$ and
$[V_I - V_{II}(\beta_{max})]/V_I \approx \gamma$. We have appended β_{max} at the bottom of Table 4.2. For instance, when $\eta = 4$, even if the cost per unit sample in the second stage is 50% larger than that in the first stage, the variance of the Graybill-Deal estimator under two-stage sampling is 5% below that under one-stage sampling.

Table 4.2

(a) α_{opt} (upper block) and
(b) β_{max} for which second stage sampling in beneficial (lower block)

β	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 5$	$\eta = 10$
1.0	.6	.4	.7	.6	.7	.7
1.1	.2	.4	.5	.6	.4	.7
1.2	0	.2	.5	.4	.4	.7
1.3		.1	.5	.4	.4	.4
1.4		.1	.5	.4	.4	.4
1.5		.1	0	.4	.4	.4
1.6		0		.4	.4	.4
1.7				0	.4	.4
1.8					.4	.4
1.9					0	.1
2.0						.1
3.0						0
$\gamma = .05$	1.00	1.14	1.34	1.50	1.68	1.85
$\gamma = 0.1$	1.02	1.48	1.43	1.63	1.81	2.26

All unspecified entries are 0.

5. ESTIMATION OF THE VARIANCE OF THE GRAYBILL-DEAL ESTIMATOR

In this section we consider the problem of estimation of the variance of the Graybill-Deal estimator $\hat{\mu}_2^*$. The exact expression of $V(\hat{\mu}_2^*)$, given in (2.5), may also be written as

$$V(\hat{\mu}_2^*) = \sum_{i=1}^{2} E\left[\frac{N_i\sigma_i^2/s_i^4}{(\sum N_i/s_i^2)^2}\right]. \tag{5.1}$$

As noted in Section 3, the evaluation of $V(\hat{\mu}_2^*)$ is extremely difficult as it involves a triple integral. Therefore, it is not possible to provide an explicit unbiased estimator of $V(\hat{\mu}_2^*)$ by simply substituting appropriate statistics in place of the parameters in the expression of $V(\hat{\mu}_2^*)$. Hence, there is necessity to resort to other methods of estimating $V(\hat{\mu}_2^*)$ in order to construct confidence limits for the common mean. For example, one may construct an initial plug-in estimator as

$$\hat{V}_0 = \left(\frac{N_1}{s_1^2} + \frac{N_2}{s_2^2}\right)^{-1} \tag{5.2}$$

by substituting s_i^2 in place of σ_i^2 inside the expectation in (5.1). However, the sampling properties of $\hat{V}_0$ are still difficult to obtain. In particular, $\hat{V}_0$ is not unbiased.

In order to obtain possibly better estimate of $V(\hat{\mu}_2^*)$, we first note that the expression in (5.1) is at least in form similar to that of the variance of the Graybill-Deal estimator based on one stage sampling. We, therefore, imitate the construction of an unbiased estimator of $V(\hat{\mu}_2^*)$ in the one stage case.

For fixed sample sizes n_1 and n_2, if one obtains real-valued functions $\psi_i \equiv \psi_i(s_1^2, s_2^2)$ such that

$$E[\psi_i] = \sigma_i^2 E[s_i^2 \textstyle\sum(n_i/s_i^2)]^{-2}$$

then an unbiased estimator of the variance of the Graybill-Deal estimator $\hat{\mu}_0$ viz. $V(\hat{\mu}_0)$ is given by

$$\hat{V}(\hat{\mu}_0) = n_1\psi_1 + n_2\psi_2. \tag{5.3}$$

Using Haff's (1974) Wishart identity for the univariate case, Sinha (1985) derived ψ_i as

$$\psi_i(s_1^2, s_2^2) = \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}\right)^{-2} \sum_{l=0}^{\infty} \frac{2^l(l+1)!}{(n_i+1)^{[l]}} s_i^{-2} \left(1 - \frac{n_i/s_i^2}{n_1/s_1^2 + n_2/s_2^2}\right)^l \quad (5.4)$$

where $(n_i+1)^{[0]} = 1$ and $(n_i+1)^{[l]} = (n_i+1)(n_i+3)\cdots(n_i+2l-1)$ for $l \geq 1$. The following Theorem is due to Sinha (1985).

Theorem 5.1.

$$|\hat{V}(\hat{\mu}_0) - E[\sum_{i=1}^{2} n_i \psi_{i,m}(s_1^2, s_2^2)]| \leq O(\sum_{i=1}^{2} \sigma_i^2 / n_i^{m+1})$$

where $\psi_{i,m}(s_1^2, s_2^2)$ is the sum of first m terms in $\psi_i(s_1^2, s_2^2$ in (5.4).

Putting (5.4) in (5.3) and simplifying, a first order approximation to $\hat{V}(\hat{\mu}_0)$ is obtained as

$$\hat{V}_1 = (\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2})^{-1} + 4\left(\frac{1}{n_1+1} + \frac{1}{n_2+1}\right) \frac{n_1}{s_1^2} \frac{n_2}{s_2^2} \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}\right)^{-3} \quad (5.5)$$

and a second order approximation to $\hat{V}(\hat{\mu}_0)$ is obtained as

$$\hat{V}_2 = \hat{V}_1 + 24\left(\frac{n_2/s_2^2}{(n_1+1)(n_1+3)} + \frac{n_1/s_1^2}{(n_2+1)(n_2+3)}\right) \frac{n_1}{s_1^2} \frac{n_2}{s_2^2} \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}\right)^{-4} \quad (5.6)$$

For the two-stage scheme, the sample sizes N_1 and N_2 are random. But still one can suggest $\hat{V}_1$ and $\hat{V}_2$, given by (5.5) and (5.6) with n_1, n_2 replaced by N_1, N_2, as estimators of $V(\hat{\mu}_2^*)$.

The following numerical study exhibits the adequacy of the first order approximation $\hat{V}_1$ to $\hat{V}(\hat{\mu}_2^*)$ for choices of the values of $\alpha = \frac{n-2m}{n}$ and $\eta = \frac{\sigma_2^2}{\sigma_1^2}$.

Table 5.1

Simulated $(10^5\times)$ $V(\hat{\mu}_2^*)$ and its estimates

α		$\eta=1$	$\eta=2$	$\eta=3$	$\eta=4$	$\eta=5$	$\eta=10$
0.0	$V(\hat{\mu}_2^*)$	2004	2782	3241	3320	3411	3812
	$\hat{V}_0$	1916	2571	2881	3112	3273	3585
	$\hat{V}_1$	2058	2745	3047	3269	3418	3685
	$\hat{V}_2$	2073	2763	3065	3285	3434	3696
0.2	$V(\hat{\mu}_2^*)$	2107	2604	2789	2942	3143	3225
	$\hat{V}_0$	1914	2432	2655	2794	2884	3105
	$\hat{V}_1$	2048	2579	2786	2910	2987	3171
	$\hat{V}_2$	2062	2593	2798	2920	2996	3176
0.4	$V(\hat{\mu}_2^*)$	2018	2552	2469	2701	2660	2664
	$\hat{V}_0$	1930	2372	2493	2547	2570	2694
	$\hat{V}_1$	2060	2504	2601	2639	2646	2739
	$\hat{V}_2$	2072	2516	2610	2646	2651	2742
0.6	$V(\hat{\mu}_2^*)$	2171	2564	2698	2588	2579	2409
	$\hat{V}_0$	2030	2444	2481	2509	2492	2522
	$\hat{V}_1$	2163	2585	2593	2600	2567	2562
	$\hat{V}_2$	2176	2598	2603	2607	2572	2564
0.8	$V(\hat{\mu}_2^*)$	2148	2907	2931	3043	3174	2861
	$\hat{V}_0$	2010	2603	2716	2774	2838	2826
	$\hat{V}_1$	2136	2767	2881	2938	3014	2987
	$\hat{V}_2$	2148	2786	2901	2961	3043	3023

6. PURELY SEQUENTIAL SAMPLING

We have been studying the performances of different estimators under two-stage sampling. We have exhibited that under the two stage sampling scheme, the variance of the Graybill-Deal estimator is smaller than the variance of other two-stage estimators of Richter as well as the variance of the alternative Graybill-Deal type estimator of μ. The performance of the Graybill-Deal estimator can possibly be improved under a purely sequential sampling procedure.

In two-stage sampling two first stage samples of size m are taken from two populations, sample variances are calculated and second stage sample of size $n-2m$ is taken from the population with smaller sample variance at the first stage. Then μ is estimated by the Graybill-Deal estimator where sample means and sample variances are calculated based on the final sample. In sequential sampling, equal samples of size m are taken from each population at the first stage. The sample variances are then computed and next observation is taken from the population with smaller sample variance. At each stage, the sample variances are updated before collecting one more observation from the population with smaller updated sample variance. Sampling continues until the total sample size becomes n.

In both the cases the estimator used is the Graybill-Deal estimator $\hat{\mu}_2^*$ with updated sample means and sample variances. For the purely sequential sampling procedure the Graybill-Deal estimator is denoted by $\hat{\mu}_s$.

We have conducted a simulation study (based on 1,000 repetitions) to compare the performances of $\hat{\mu}_s$ and $\hat{\mu}_2^*$ (the Graybill-Deal estimator under two-stage sampling). The results are summarized in Table 6.1. The simulation study shows that the estimator $\hat{\mu}_s$ under purely sequential sampling scheme performs better than $\hat{\mu}_2^*$ under two-stage sampling scheme both in terms of the variance and ASN$\times$Variance which is the reciprocal of efficiency per unit sample. This is more obvious either for η away from 1 or for k suitably large. The first stage sample size m is taken to be 5. The total sample size is $2m+k$ where k runs from 0 to 10. Studies are done for different values of $\eta = \sigma_2^2/\sigma_1^2$.

This simulation study also shows that the estimators are progressively better (in terms of both the variance and ASN$\times$Variance) as k increases.

Table 6.1

($10^5\times$) $V(\hat{\mu}_2^*),\ V(\hat{\mu}_s),\ ASN\times V(\hat{\mu}_2^*)$ and $ASN\times V(\hat{\mu}_s)$.
Here m=5, $\sigma_1^2=1, \sigma_2^2=\eta$.

	k	$V(\hat{\mu}_2^*)$	$V(\hat{\mu}_s)$	$ASN\times V(\hat{\mu}_2^*)$	$ASN\times V(\hat{\mu}_s)$
$\eta=1$	0	11969	11969	119690	119690
	1	10527	10527	115800	115800
	2	9494	8372	113930	100460
	3	8575	7828	111470	101760
	4	7856	7311	109980	102360
	5	7333	6949	110000	104240
	10	5342	5118	106840	102370
$\eta=2$	0	15955	15955	159550	159550
	1	13642	13642	150060	150060
	2	12013	10561	144160	126730
	3	10956	9677	142430	125800
	4	9978	8986	139690	125800
	5	9099	8397	136480	125950
	10	6649	6103	132990	122060
$\eta=5$	0	20271	20271	202710	202710
	1	16931	16931	186240	186240
	2	14570	12540	174840	150480
	3	12814	11398	166580	148170
	4	11585	10381	162190	145330
	5	10485	9304	157270	139560
	10	7178	6366	145370	127330

7. TWO-STAGE SAMPLING WITH AN INDIFFERENCE ZONE

Another modification of the two-stage sampling scheme is to include an indifference zone. Here the following type of sampling is considered: first samples of size m are taken from two populations. The sample variances $s_1^2(m), s_2^2(m)$, and $R = s_1^2(m)/s_2^2(m)$ are calculated. $a > 1$ is fixed. If $R < a^{-1}$, all the remaining $k = n-2m$ observations are taken from population I. If $R > a$, all the remaining $k = n - 2m$ observations are taken from population II. If $a^{-1} < R < a$, $k = \frac{n-2m}{2}$ observations are taken from each of the population I and II. The two-stage sampling scheme with no indifference zone corresponds to the case a=1. The Graybill-Deal estimator of μ is defined as before. The Graybill-Deal estimator under the modified sampling scheme is denoted by $\hat{\mu}_a$.

Table 7.1 presents the simulation study showing variances of the Graybill-Deal estimators $\hat{\mu}_2^*$ (under two-stage sampling) and $\hat{\mu}_a$ (with indifference zone) for different values of a and η. Here we take m=5.

Erratic behaviour of $\hat{\mu}_a$ for various values of (a, η) needs further study. Across any column or any row values are oscillating.

Thus unlike under purely sequential sampling the performance of the Graybill-Deal estimator does not necessarily improve under inclusion of indifference zone as compared to two-stage sampling. In other words, there is no guaranteed reduction in the variance of the Graybill-Deal estimator with the inclusion of the indifference zone.

Table 7.1

$(10^5\times)$ $V(\hat{\mu}_2^*)$, and $V(\hat{\mu}_a)$. Here m=5, $\sigma_1^2 = 1, \sigma_2^2 = \eta$.

	k	$V(\hat{\mu}_2^*)$ a=1	$V(\hat{\mu}_a)$ a=2	$V(\hat{\mu}_a)$ a=3	$V(\hat{\mu}_a)$ a=4	$V(\hat{\mu}_a)$ a=5
$\eta = 1$	2	9415	9462	9344	9512	9559
	4	7854	7923	7874	7967	7984
	6	6758	6830	6836	6860	6875
	8	5955	6003	6000	6052	6092
	10	5342	5358	5411	5399	5414
$\eta = 2$	2	12353	12500	12616	12512	12693
	4	10146	10317	10361	10456	10464
	6	8651	8696	8976	8987	9035
	8	7619	7692	7744	7921	7910
	10	6863	6795	6923	6989	7071
$\eta = 5$	2	14647	14917	15040	15253	15453
	4	11713	11704	12013	12122	12519
	6	9643	9733	9996	10221	10406
	8	8365	8400	8595	8860	8973
	10	7348	7325	7450	7698	7856

CHAPTER 3

ESTIMATION OF THE COMMON MEAN OF NORMAL POPULATIONS IN THE PRESENCE OF A REGRESSOR UNDER HETEROSCEDASTIC VARIANCES

Suppose the common mean of several independent normal populations with possibly different variances is 'drifted' linearly by influence of a continuous regressor. Two estimators of the variance components — (a) Grubbs' estimator and (b) Rao's MINQUE (with invariance) — are derived and then used to construct (i) a Graybill-Deal estimator and (ii) a weighted least squares estimator for the common mean, after eliminating the effect of the regressor. The properties of these estimators are studied through simulation.

1. Introduction

In the previous chapters we have considered the problem of estimation of the common mean of two independent univariate normal populations with different and unknown variances. The problem of estimation of the common mean of several normal populations with different and unknown variances has also attracted attention of many researchers. Let $\{y_{ij} : 1 \le j \le n_i; n_i \ge 2\}$ be a random sample from the i-th population with distribution $N(\mu, \sigma_i^2)$, $(i = 1, 2, \cdots, I,\ I \ge 2)$. Let $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ and $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ be respectively the mean and the variance based on the sample from i-th population. Then, if σ_i^2's are known the minimum variance unbiased estimator of μ is given by

$$\hat{\mu}_{MVUE} = \frac{\sum n_i \bar{y}_i \, \sigma_i^{-2}}{\sum n_i \sigma_i^{-2}}. \tag{1.1}$$

If σ_i^2's are unknown, the Graybill-Deal estimator of μ is defined as

$$\hat{\mu}_{GD} = \frac{\sum n_i \bar{y}_i \, s_i^{-2}}{\sum n_i s_i^{-2}}. \tag{1.2}$$

In general terms, $\hat{\mu}_{GD}$ is unbiased since its conditional mean given s_i^2 and hence the unconditional mean is μ. Some of the results on the properties of $\hat{\mu}_{GD}$ or on comparison with other estimators for $I = 2$ have also been extended for general $I(\geq 2)$. As for example Norwood and Hinklemann (1977) proved that the estimator $\hat{\mu}_{GD}$ is uniformly better than the individual sample means if and only if $n_i \geq 11 \; \forall i$ or $n_i = 0$ for some i, $n_j \geq 19 \; \forall j \neq i$ [See Pal and Sinha (1996) for an excellent review of these results].

So far the populations are assumed to be normal with a common mean but heterogeneous with respect to variability. In this chapter, we contemplate the situation wherein the (common) mean is 'drifted' by influence of a covariate arising in a natural manner in application areas. See Introduction (Part I) for reference. We incorporate one concomitant variable (also known as covariate) in the form of a linear regressor taking values over a continuum. Suppose that data on the response variable Y as well as on a covariate X are available from I different sources. We assume the following regression model with unequal error variances:

$$y_{ij} = \mu + \gamma \; x_{ij} + \epsilon_{ij}; \quad j = 1, \ldots, n_i; \;\; i = 1, \ldots, I \tag{1.3}$$

where μ is the true mean response, γ is the 'drift' factor acting on X, both being unknown and ϵ_{ij}'s are independent random error variables distributed as $N(0, \sigma_i^2)$. The central problem is to estimate the common mean μ based on the data arising out of the model (1.3), treating the regression parameter γ and the variance components as unknown. The common mean μ may be estimated either by (i) a Graybill-Deal, or by (ii) a weighted least squares estimator (WLS), provided we have first estimated the variance components.

In Section 2 we derive a Grubbs' (1948) estimator, and in Section 3 we construct Rao's (1970) minimum quadratic unbiased estimator (MINQUE) with invariance for the variance components in our model (1.3). Section 4 presents estimation of the common mean using these variance components estimates. Section 5 concludes with a simulation study of the performance of these estimators.

2. Grubbs' estimation of variance components

Grubbs (1948) suggested a method of unbiased estimation of heterogeneous variances in a two-way complete layout with one observation per cell i.e., for the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij},\ i = 1, 2, \cdots, b,\ j = 1, 2, \cdots, v \tag{2.1}$$

where α_i's and β_j's are fixed effects and ϵ_{ij}'s are independent random variables distributed as $N(0, \sigma_i^2)$. The technique is to form usual 'within block' residual sum of squares

$$S_i^2 = \sum_{j=1}^{v} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2,\ i = 1, 2, \cdots, b \tag{2.2}$$

and equate these sums of squares to their expectations under the model (2.1). This ultimately results in the equations

$$\underline{S}^2 = A\underline{\sigma}^2$$

whence $\underline{\sigma}^2$ is estimated as

$$\hat{\underline{\sigma}}^2 = A^{-1}\underline{S}^2 \tag{2.3}$$

It may be noted that in a two-way complete layout as considered above, the matrix A is positive definite (assuming $b \geq 3$) and it involves only the design parameters b and v. Several generalisations of this concept are available in the literature. See for example Brindley and Bradley (1985), Meloney (1973), Ellenberg (1977), Brindley and Bradley (1986), Bradley and Rollier (1988) and Das, Sinha and Sinha (1993).

For our model (1.3), the sample-wise residual sums of squares are

$$S_i^2 = \sum_{j=1}^{n_i} \{y_{ij} - \bar{y}_i - \hat{\gamma}^{(i)}\ (x_{ij} - \bar{x}_i)\}^2$$

where $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij},\ \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ and

$$\hat{\gamma}^{(i)} = \frac{\sum_{j=1}^{n_i} y_{ij}\ (x_{ij} - \bar{x}_i)}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}. \tag{2.4}$$

A naive sample-wise estimator of σ_i^2 is given by $\hat{\sigma}_{i\,N}^2 = S_i^2/(n_i - 2)$. To improve this naive estimator, we begin by choosing an appropriate estimator of γ by pooling those in (2.4). One natural choice is the Graybill-Deal type estimator of γ given by

$$\hat{\gamma}_{GD} = \frac{\sum_{i=1}^{I} \hat{\gamma}^{(i)}\, S_{xi}^2\, (n_i - 2) \;/\; S_i^2}{\sum_{t=1}^{I} S_{xt}^2\, (n_t - 2) \;/\; S_t^2}$$

where $S_{xi}^2 = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$. However, the variance of $\hat{\gamma}_{GD}$ is intractable. Therefore, we choose another ad-hoc unbiased pooled estimator of γ viz.

$$\hat{\hat{\gamma}} = \frac{\sum_{i=1}^{I} \hat{\gamma}^{(i)}\, S_{xi}^2}{\sum_{t=1}^{I} S_{xt}^2}. \tag{2.5}$$

Using the estimator in (2.4), we define the (adjusted) residual sums of squares as

$$\tilde{S}_i^2 = \sum_{j=1}^{n_i} \{y_{ij} - \bar{y}_i - \hat{\hat{\gamma}}\, (x_{ij} - \bar{x}_i)\}^2 \tag{2.6}$$

for $i = 1, 2, \ldots, I$. To obtain Grubbs' estimators of variance components we must find expressions of the expected values $E[\tilde{S}_i^2]$, and then solve the system of equations $E[\tilde{S}_i^2] = \tilde{S}_i^2$, $i = 1, \ldots, I$. Towards that end, note that

$$var(\hat{\gamma}^{(i)}) = \frac{\sigma_i^2}{S_{xi}^2}, \quad var(\hat{\hat{\gamma}}) = \frac{\sum_{i=1}^{I} \sigma_i^2\, S_{xi}^2}{\left(\sum_{t=1}^{I} S_{xt}^2\right)^2} \tag{2.7}$$

and

$$cov(\hat{\hat{\gamma}}, \hat{\gamma}^{(i)}) = \frac{\sigma_i^2}{\sum_{t=1}^{I} S_{xt}^2}. \tag{2.8}$$

Hence, writing (2.6) as

$$\tilde{S}_i^2 = \sum_{j=1}^{n_i} \{[y_{ij} - \bar{y}_i - \hat{\gamma}^{(i)}\, (x_{ij} - \bar{x}_i)] - (\hat{\hat{\gamma}} - \hat{\gamma}^{(i)})\, (x_{ij} - \bar{x}_i)\}^2 \tag{2.9}$$

after some algebraic simplifications using (2.7) and (2.8), it can be seen that

$$
\begin{aligned}
E[\tilde{S}_i^2] &= (n_i-2)\sigma_i^2 + S_{xi}^2 \; var\;(\hat{\hat{\gamma}} - \hat{\gamma}^{(i)}) \\
&= (n_i-2)\sigma_i^2 + S_{xi}^2 \left[\frac{\sum_{j=1}^{I} \sigma_j^2 \, S_{xj}^2}{\left(\sum_{t=1}^{I} S_{xt}^2\right)^2} + \frac{\sigma_i^2}{S_{xi}^2} - 2\,\frac{\sigma_i^2}{\sum_{t=1}^{I} S_{xt}^2} \right] \\
&= \left(n_i - 1 - 2\,\frac{S_{xi}^2}{\sum_{t=1}^{I} S_{xt}^2} \right) \sigma_i^2 + \frac{S_{xi}^2}{\sum_{t=1}^{I} S_{xt}^2} \sum_{j=1}^{I} \sigma_j^2 \, \frac{S_{xj}^2}{\sum_{t=1}^{I} S_{xt}^2} \\
&= (n_i - 1 - 2\alpha_i)\sigma_i^2 + \alpha_i \sum_{j=1}^{I} \sigma_j^2 \alpha_j
\end{aligned}
\tag{2.10}
$$

where

$$\alpha_i = \frac{S_{xi}^2}{\sum_{t=1}^{I} S_{xt}^2} \tag{2.11}$$

for $i = 1, \ldots, I$. Using the vector notation, $\underline{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_I^2)'$, $\underline{\tilde{S}}^2 = (\tilde{S}_1^2, \ldots, \tilde{S}_I^2)'$, $\underline{\alpha} = (\alpha_1, \ldots, \alpha_I)'$ we note that

$$E[\underline{\tilde{S}}^2] = [\Delta + \underline{\alpha}\,\underline{\alpha}']\,\underline{\sigma}^2 \tag{2.12}$$

where Δ is an $I \times I$ diagonal matrix with i-th diagonal element

$$\delta_i = n_i - 1 - 2\,\alpha_i > 0.$$

Note that $[\Delta + \underline{\alpha}\,\underline{\alpha}']$ is invertible. The expression for the inverse matrix is given in Rao (1973), for example. From (2.12), Grubbs' estimators of the variance components in model (1.3) are given by

$$\hat{\underline{\sigma}^2}_G = \left[\Delta^{-1} - \frac{\Delta^{-1}\underline{\alpha}\;\underline{\alpha}'\Delta^{-1}}{1 + \underline{\alpha}'\Delta^{-1}\underline{\alpha}} \right] \underline{\tilde{S}}^2. \tag{2.13}$$

The issue of non-negativity of the estimators in (2.13) remains unsettled. As in Das, Sinha and Sinha (1993) we expect these estimators to be nonnegative with high probability whenever $\sigma_{max}^2/\sigma_{min}^2$ is at least moderately large. Here $\sigma_{max}^2 = max_{1\le i\le I}\;\sigma_i^2$ and $\sigma_{min}^2 = min_{1\le i\le I}\;\sigma_i^2$.

3. MINQUE (with invariance) of the variance components

In this section we adopt Rao's (1970) MINQUE theory to our model (1.3) to obtain estimates of the variance components σ_i^2's. Below we give a synopsis of the theory. For more details we refer to Khatri and Srivastava (1979), Chapter 5.

Suppose that Z is a known $n \times p$ design matrix. $U = [U_1|\cdots|U_I]$ is a known $n \times n$ matrix, and $\eta' = (\eta_1', \ldots, \eta_I')$ where $\eta_i' = (\eta_{i1}, \ldots, \eta_{in_i})$ is the unobservable random error vector associated with the i-th sample with dispersion matrix $D(\eta_i) = \sigma_i^2 I_{n_i}$. Consider a more general model

$$y = Z\beta + U\eta. \tag{3.1}$$

Note that $D(\eta) = D_\sigma = Diag(D(\eta_1), \ldots, D(\eta_I))$ and $D(U\eta) = UD_\sigma U' = \sum_{i=1}^I \sigma_i^2 V_i$ where $V_i = U_i U_i',\ i = 1, 2, \cdots, I$.

In order to estimate $f = \sum_{i=1}^I f_i \, \sigma_i^2$, a linear combination of the variance components, consider a quadratic estimator of the form $q = y'Ay$, where A is a symmetric positive definite matrix. Unbiasedness (that is, $E[q] = f$) requires that $Z'AZ = 0$ and trace $(A\ V_i) = f_i$. Furthermore, location invariance of q (that is, $y'Ay = (y - Z\beta)'A(y - Z\beta)$ for all β) demands that $A\ Z = 0$. However, there may be several A matrices (hence, quadratic estimators) satisfying $A\ Z = 0$ and trace $(A\ V_i) = f_i$. The choice of a solution is therefore motivated by optimization with respect to some other desirable criterion. Rao (1970) proposed minimizing a suitable norm (usually Euclidean) of $U'AU - F$ where

$$F = Diag\left(\frac{f_1}{n_1} I_{n_1}, \ldots, \frac{f_I}{n_I} I_{n_I}\right). \tag{3.2}$$

The motivation for this criterion is as follows. If η_i's were known then a natural estimator of σ_i^2 would have been $\eta_i'\eta_i/n_i$. Hence, a naive estimator of $\underline{f}$ would be $\underline{\hat{f}} = \eta' F \eta$. However, $y'Ay = \eta' U'AU\eta$ as a consequence of $A\ Z = 0$. Thus one should attempt to minimize $||U'AU - F||$.

The solution (with invariance) which minimizes $||U'AU - F||$ turns out to be

$$A = \sum_{i=1}^I \lambda_i B_i \tag{3.3}$$

where $B_i = (I - W^{-1}P)W^{-1}V_iW^{-1}(I - PW^{-1})$ with $W = UU'$ and $P = Z(Z'W^{-1}Z)^{-1}Z'$, and where λ_i's are obtained from

$$\sum_{i=1}^{I} \lambda_i tr(B_iV_k) = f_k; \quad for \quad 1 \le k \le I. \tag{3.4}$$

Specializing Rao's solution to our model (1.3) we obtain the MINQUE (with invariance) of the variance components. Note that in our model, $U = I$. Hence, $W = I, P = Z(Z'Z)^{-1}Z'$ and

$$\begin{aligned} B_i &= (I-P)V_i(I-P) \quad with \quad V_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_n & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad and \\ & \sum_{i=1}^{I} \lambda_i trace\ [B_iV_j] = f_j. \end{aligned}$$

In other words, (3.4) simplifies to

$$\underline{\lambda} \equiv \underline{\lambda}(f) = G^{-1}\underline{f} \tag{3.5}$$

where $G = ((g_{ij}))$ with

$$\begin{aligned} g_{ij} &= trace\ (B_iV_j) = trace\ (I-P)V_i(I-P)V_j \\ &= trace\ [(I-P)_{ji}(I-P)_{ij}] \\ &= \begin{cases} trace\ C_iC_j, & if\ \ i \ne j \\ n_i - 2trace\ C_i + trace\ C_i^2, & if\ \ i = j \end{cases} \end{aligned}$$

and $C_i = (Z'Z)^{-1}(Z_i'Z_i)$. Note also that, in our model (1.3), $Z = [\underline{1} \quad \underline{x}]$. Hence,

$$Z'Z_{2\times 2} = \begin{bmatrix} n & \underline{1}'\underline{x} \\ \underline{x}'\underline{1} & \underline{x}'\underline{x} \end{bmatrix}, \quad and \quad (Z_i'Z_i)_{2\times 2} = \begin{bmatrix} n_i & \underline{1}'\underline{x}_i \\ \underline{x}_i'\underline{1} & \underline{x}_i'\underline{x}_i \end{bmatrix}.$$

Hence, $C_{i2\times 2}$ and $G_{I\times I}$ can be easily calculated.

Choosing $\underline{f} = \underline{f}^{(k)} \stackrel{def}{=} (0, \ldots, 0, 1, 0, \ldots, 0)$ (here, 1 appears in the k-th coordinate) we have from (3.5), $\underline{\lambda} = \underline{\lambda}^{(k)} = (G^{-1})_{*k}$, the k-th column of G^{-1}. Hence, from (3.3), $\hat{\sigma_k^2} = y'A^{(k)}y = \sum_{i=1}^{I} \lambda_i^{(k)}(y'B_iy) = \sum_{i=1}^{I} g^{ik}(y'B_iy)$. Therefore, the MINQUE (with invariance) of $\underline{\sigma}^2$ is

$$\hat{\underline{\sigma}^2}_M = G^{-1}\underline{t} \tag{3.6}$$

where

$$\underline{t}' = (y'B_1y, \ldots, y'B_Iy) = (e_1'e_1, \ldots, e_I'e_I)$$

with $\underline{e}' = (\underline{e}_1, \ldots, \underline{e}_I) = (I-P)y$ being the vector of residuals of the ordinary least square model, assuming homoscedasticity.

4. Estimation of the common mean

Having obtained the estimators $\hat{\underline{\sigma}^2}$ of the variance components — either by Grubbs' method or by Rao's method, we can now obtain an estimator for the true population mean μ in the model (1.3) using either (i) the Graybill-Deal estimator or (ii) the weighted least squares estimator.

Note that, based on the i-th sample alone, the population mean is unbiasedly estimated by $\hat{\mu}^{(i)} = \bar{y}_i - \hat{\gamma}^{(i)}\bar{x}_i$ with variance

$$var(\hat{\mu}^{(i)}) = \sigma_i^2 \left[\frac{1}{n_i} + \frac{\bar{x}_i^2}{S_{xi}^2}\right].$$

Therefore, the Graybill-Deal estimator of μ is given by

$$\hat{\mu}_{GD} = \frac{\sum_{i=1}^{I}(\bar{y}_i - \hat{\gamma}^{(i)}\,\bar{x}_i)\,\hat{\sigma_i}^{-2}\left[\frac{1}{n_i} + \frac{\bar{x}_i^2}{S_{xi}^2}\right]^{-1}}{\sum_{i=1}^{I}\hat{\sigma_i}^{-2}\left[\frac{1}{n_i} + \frac{\bar{x}_i^2}{S_{xi}^2}\right]^{-1}} \tag{4.1}$$

The weighted least squares estimator of μ is obtained from the regression model (1.3) assuming that the independent random error variables ϵ_{ij}'s are distributed as $N(0, \sigma_i^2)$. If it is assumed that σ_i^2's are known, $D(\epsilon) = \Sigma$ which is a known block diagonal matrix with $\{\sigma_i^2\, I_{n_i}, i = 1, 2, \cdots, I\}$ on the diagonal. Therefore, recalling that $Z = [\underline{1} \;\; \underline{x}]$, the weighted least square estimate of μ is

$$\begin{aligned}\hat{\mu}_{WLS} &= (1,0)\,(Z'\Sigma^{-1}Z)^{-1}\,Z'\Sigma^{-1}y \\ &= \frac{(\sum_i \frac{SS_i}{\sigma_i^2})(\sum_i \frac{G_i}{\sigma_i^2}) - (\sum_i \frac{T_i}{\sigma_i^2})(\sum_i \frac{CP_i}{\sigma_i^2})}{(\sum_i \frac{n_i}{\sigma_i^2})(\sum_i \frac{SS_i}{\sigma_i^2}) - (\sum_i \frac{T_i}{\sigma_i^2})^2}\end{aligned} \tag{4.2}$$

where $T_i = \sum_j x_{ij},\ SS_i = \sum_j x_{ij}^2,\ G_i = \sum_j y_{ij},\ CP_i = \sum_j x_{ij}y_{ij}$. As σ_i^2's are unknown, we may replace σ_i^2 in (4.2) by $\hat{\sigma}_i^2$ to obtain estimate of μ.

5. Simulation Study

The following simulation study summarizes some statistical properties of the variance component estimators and the estimators of the common mean.

For $I = 2$, we have taken $n_1 = 10$ and $n_2 = 20$ observations from the two populations. The covariate values are

$$x_{1j} = 0.1, 0.2, \ldots, 1.0 \quad and \quad x_{2j} = 0.1, 0.2, \ldots, 2.0.$$

We have chosen $\mu = 5, \gamma = 3$. Since the behavior of the Graybill-Deal estimator of μ depends on σ_1^2 and σ_2^2 only through their ratio, we have chosen $\sigma_2 = 4$ and $\sigma_1 = 1, 2, 3$. Based on 10,000 repetitions, the mean, standard deviation, skewness and kurtosis of $\hat{\sigma_1}^2$ and $\hat{\sigma_2}^2$ are given in Table 1 below, and those of $\hat{\mu}_{GD}$ and $\hat{\mu}_{WLS}$ are given in Table 2 below. From Table 1, it is evident that the two methods of estimating the variances are comparable. Although MINQUE slightly underestimates σ_2^2, this has no bearing on the estimation of the common mean, as evident from Table 2.

Also the following observations are evident from Table 2. The statistical properties of $\hat{\mu}_{GD}$ and $\hat{\mu}_{WLS}$ are unaffected by the choice of variance component estimator. Under both MINQUE and Grubbs' estimation of the variance components,

- estimators $\hat{\mu}_{GD}$ and $\hat{\mu}_{WLS}$ have comparable small biases (due to simulation error); however, $\hat{\mu}_{WLS}$ has a slightly better precision than $\hat{\mu}_{GD}$
- for both $\hat{\mu}_{GD}$ and $\hat{\mu}_{WLS}$, the standard deviation increases and the already small bias decreases further as σ_1/σ_2 increases to 1
- the small values of skewness and kurtosis indicate that both $\hat{\mu}_{GD}$ and $\hat{\mu}_{WLS}$ have approximate normal distributions.

Table 1.
Properties of the Grubb's estimator and MINQUE of the variances ($\sigma_2 = 4$)

		Variance components estimated by			
		Grubbs' method		MINQUE method	
σ_1		$\hat{\sigma_1}^2$	$\hat{\sigma_2}^2$	$\hat{\sigma_1}^2$	$\hat{\sigma_2}^2$
1	mean	0.9969	15.9975	1.0208	15.1904
	standard dev.	0.5973	5.4108	0.8354	5.0823
	skewness	1.2397	0.7314	1.4215	0.7200
	kurtosis	2.7041	0.9478	3.5242	0.9138
2	mean	3.9953	15.9974	4.0009	15.1911
	standard dev.	1.9717	5.4110	2.0103	5.0838
	skewness	0.9962	0.7315	0.9038	0.7182
	kurtosis	1.3672	0.9474	1.2778	0.9057
3	mean	8.9941	15.9972	8.9677	15.1920
	standard dev.	4.3144	5.4114	4.2250	5.0888
	skewness	0.9852	0.7309	0.9238	0.7144
	kurtosis	1.3193	0.9499	1.3445	0.8993

Table 2.
Properties of two estimators of the common mean ($\sigma_2 = 4$)

		Variance components estimated by			
		Grubbs' method		MINQUE method	
σ_1		$\hat{\mu}_{GD}$	$\hat{\mu}_{WLS}$	$\hat{\mu}_{GD}$	$\hat{\mu}_{WLS}$
1	mean	5.0046	5.0042	5.0035	5.0035
	standard dev.	0.6527	0.6026	0.7153	0.6221
	skewness	0.0495	0.0131	0.0804	0.0253
	kurtosis	0.1466	0.1354	0.7428	0.3977
2	mean	5.0033	5.0043	5.0017	5.0032
	standard dev.	1.1166	0.9697	1.1493	0.9772
	skewness	0.0583	0.0291	0.0615	0.0290
	kurtosis	0.0455	0.0727	0.1341	0.1070
3	mean	4.9999	5.0019	4.9986	5.0010
	standard dev.	1.3982	1.2356	1.4124	1.2360
	skewness	0.0549	0.0300	0.0532	0.0297
	kurtosis	0.0488	0.0376	0.0790	0.0487

CHAPTER 4

OPTIMAL SAMPLING SCHEME FOR ESTIMATING THE COMMON MEAN OF A BIVARIATE NORMAL POPULATION UNDER DIFFERENTIAL SAMPLING COSTS

In the context of estimating the common mean of a bivariate normal distribution when the two component variables have different sampling costs per unit, consider the problem of allocating a fixed sampling budget among the first variable, the second variable and the bivariate data. For the known dispersion matrix case, the optimal strategy is shown to be the one which allocates the entire budget to exactly one of these three types of data depending on the values of (a) the parameters of the bivariate normal distribution, (b) the cost components. For the case of unknown dispersion matrix, a two-stage sampling strategy involving bivariate data in the first stage and optimal type of data in the second stage is considered. Two estimators of the common mean are proposed and their accuracy is studied through simulation. Furthermore the question of choice of optimal first-stage sample size is also addressed.

1. Introduction

In the previous Chapters we have studied some aspects of the problem of estimation of common mean of two or more univariate normal populations with unknown unequal variances. In Chapter 1 we have studied the admissibility property of some estimators of common mean of two normal populations. In Chapter 2 we have studied the performances of various two-stage estimators of common mean of two normal populations. In Chapter 3 we considered the problem of estimation of common mean of several normal populations incorporating covariate information. However in all these studies very little attention has been paid to the sampling cost aspect of the problem.

An interesting example involving differential costs of sampling from the two populations is found in Sun (1999). The Environment Protection Agency monitors gasoline quality based on Reid Vapor Pressure (RVP) which is measured using two methods: a cheap and quick but crude measure obtained on site, and an expensive laboratory analysis yielding a measure of presumably higher precision. Sun (1999) studies the inference problem involving RVP by appropriately combining information from the bivariate data and additional cheap data exclusively from the quick method, but makes no attempt to study the allocation of sampling budget to the two types of data.

In this chapter we study the optimal sampling strategy in order to estimate with highest precision the common mean of a bivariate normal population under differential sampling costs per sampling unit subject to budget constraints.

Consider a random vector (X_1, X_2) following a bivariate normal distribution with mean vector $\binom{\mu}{\mu}$ and dispersion matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\,\sigma_1\,\sigma_2 \\ \rho\,\sigma_1\,\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{1.1}$$

We assume that $\sigma_{11} \gg \sigma_{22}$ and $c_1 \ll c_2$, where c_1 is the sampling cost per unit of X_1 and c_2 is the sampling cost per unit of X_2. The sampling cost of a bivariate observation (X_1, X_2) is $c_1 + c_2$. The objective of this study is to determine how to allocate a fixed budget C among the three types of data: exclusive data from X_1, exclusive data from X_2 and bivariate data (X_1, X_2), in order to obtain the most precise point estimator of the common mean μ.

In Section 2 we solve the budget allocation problem for the case of known Σ. We show that the entire budget should be allocated to exactly one type of data depending on the parameters $c_2/c_1, \sigma_1/\sigma_2$ and ρ. In Section 3 we consider the case of unknown Σ and propose a two-stage sampling scheme. Based on the first stage data analysis, we decide on the nature of optimal allocation of second stage sample among the three types discussed in the preceding paragraph. At the conclusion of the second-stage sampling, the common mean may be estimated either using the second-stage data only, or augmenting it with the appropriate part of the first-stage data. Standard errors of the estimators are obtained through a simulation study. Furthermore, the choice of the optimal first-stage sample size is studied through simulation.

2. Optimal Sampling Scheme when Σ is Known

Suppose that the sampling strategy is to take n_1^* observations exclusively on X_1, n_2^* observations exclusively on X_2, and n_{12} observations on the bivariate data (X_1, X_2). Assume that these three sets of observations are independent. Let $n_1 = n_1^* + n_{12}$ and $n_2 = n_2^* + n_{12}$. Suppose C denotes the total cost of sampling, and C_0 is the overhead cost irrespective of the type(s) or number of observations. We are interested in the optimal choice of (n_1^*, n_2^*, n_{12}) in order to estimate μ with the highest precision, subject to

$$C = C_0 + c_1 n_1 + c_2 n_2. \tag{2.1}$$

From the three sets of independent data, the estimators of μ are

$$\hat{\mu}_1 = \bar{X}_1, \quad \hat{\mu}_2 = \bar{X}_2, \quad \hat{\mu}_c = l_1 \bar{X}_1^* + l_2 \bar{X}_2^* \tag{2.2}$$

where $\bar{X}_i$ and $\bar{X}_i^*$ are respectively the sample means based on n_i^* observations exclusively on X_i and sample mean based on n_{12} observations on the i-th component of the bivariate data (X_1, X_2), i=1,2. Also the choice of l_1 and l_2 given by

$$\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \frac{\Sigma^{-1}1}{1'\Sigma^{-1}1} = \frac{1}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}} \begin{pmatrix} \sigma_{22} - \sigma_{12} \\ \sigma_{11} - \sigma_{12} \end{pmatrix} \tag{2.3}$$

provides the BLUE of μ as a function of $\bar{X}_1^*$ and $\bar{X}_2^*$ [vide Searle (1971)]. Since the three estimators in (2.2) are independent, the dispersion matrix for the estimator vector $(\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_c)'$ is given by

$$W = \begin{bmatrix} \sigma_{11}/n_1^* & 0 & 0 \\ 0 & \sigma_{22}/n_2^* & 0 \\ 0 & 0 & (1'\Sigma^{-1}1)^{-1}/n_{12} \end{bmatrix}. \tag{2.4}$$

The BLUE of μ using all three sets of data together, is given by

$$\hat{\hat{\mu}} = w_1 \hat{\mu}_1 + w_2 \hat{\mu}_2 + w_3 \hat{\mu}_c$$

where

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \frac{W^{-1}1}{1'W^{-1}1}$$

with

$$var(\hat{\mu}) = (1'W^{-1}1)^{-1} = \left[\frac{n_1^*}{\sigma_{11}} + \frac{n_2^*}{\sigma_{22}} + n_{12}(1'\Sigma^{-1}1)\right]^{-1}$$

Thus, the allocation problem amounts to **maximizing**

$$\frac{n_1^*}{\sigma_{11}} + \frac{n_2^*}{\sigma_{22}} + n_{12}(1'\Sigma^{-1}1) \tag{2.5}$$

subject to constraint (2.1).

Note that (2.5) can be rewritten as

$$\frac{c_1 n_1^*}{c_1\sigma_{11}} + \frac{c_2 n_2^*}{c_2\sigma_{22}} + \frac{(c_1+c_2)n_{12}}{(c_1+c_2)/1'\Sigma^{-1}1} \tag{2.6}$$

If sampling is **exclusively** restricted to X_1 alone, then the choice of n_1^* is dictated by $C = C_0 + c_1 n_1^*$ from (2.1) i.e., $c_1 n_1^* = C - C_0$. In that case, (2.6) has contribution only from the first term i.e., (2.6) reduces to $\frac{C-C_0}{c_1\sigma_{11}}$. Similarly, in case of exclusive drawing from X_2 population, (2.6) reduces to $\frac{C-C_0}{c_2\sigma_{22}}$. Lastly, in case of combined drawing, (2.6) reduces to $\frac{C-C_0}{(c_1+c_2)/1'\Sigma^{-1}1}$. Further, the maximisation problem can be equivalently stated as corresponding minimisation problem with the constant factor $(C - C_0)$ ignored. Hence, the optimal choice of (n_1^*, n_2^*, n_{12}) involves sampling exactly one type of data corresponding to the **minimum** of

$$c_1\sigma_{11}, \quad c_2\sigma_{22}, \quad \frac{c_1+c_2}{1'\Sigma^{-1}1} = \frac{(c_1+c_2)\,(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}{(\sigma_{11} + \sigma_{22} - 2\sigma_{12})}. \tag{2.7}$$

To solve this allocation problem, in the case of known Σ, let $\alpha^2 = \sigma_{11}/\sigma_{22}$ and $\gamma = c_2/c_1$. By assumption we have $\alpha, \gamma \gg 1$. To determine which type of data one must sample, let us consider the following cases separately.

Case I: $\alpha^2 < \gamma$. In this case, $c_1\sigma_{11} < c_2\sigma_{22}$. Hence, from (2.7), an optimal choice requires that $n_2^* = 0$. Furthermore, one must sample from exclusive X_1 data if

$$c_1\sigma_{11} < \frac{(c_1+c_2)\,(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}{(\sigma_{11} + \sigma_{22} - 2\sigma_{12})}$$

or equivalently, if

$$\frac{(1+\gamma)(1-\rho^2)}{1+\alpha^2 - 2\alpha\rho} > 1. \tag{2.8}$$

Otherwise, one must sample from bivariate data only.

Case II: $\alpha^2 > \gamma$. Proceeding in the same way as in Case I, we note that in this case, $n_1^* = 0$, and one must sample from exclusive X_2 data if

$$\frac{(1+\gamma)(1-\rho^2)}{1+\alpha^2-2\alpha\rho} > \frac{\gamma}{\alpha^2}. \tag{2.9}$$

Otherwise, one must sample from bivariate data only.

Case III: $\alpha^2 = \gamma$. In this case, one is indifferent between exclusive X_1 data and exclusive X_2 data. But one must decide whether the bivariate data is preferable to any one of these. Note that, in this case, the conditions in (2.8) and (2.9) are identical. If these equivalent conditions fail, then one must sample from the bivariate data.

Therefore, the optimal sampling scheme may be summarized as below.

Theorem 1. Suppose that Σ in (1.1) is known. For any $\gamma \gg 1$, the optimal sampling scheme that minimizes the variance of the point estimator $\hat{\mu}$ of μ subject to constraint (2.1) consists of sampling exclusively from X_1, exclusively from X_2, or from bivariate (X_1, X_2) data according as (α, ρ) belongs to R_1, R_2 or R_3 respectively, where

$$R_1 = \{(\alpha,\rho) : \alpha^2 < \gamma, \rho < \rho^*(\alpha)\},$$

$$R_2 = \{(\alpha,\rho) : \alpha^2 \geq \gamma, \rho < \rho^*(\alpha)\},$$

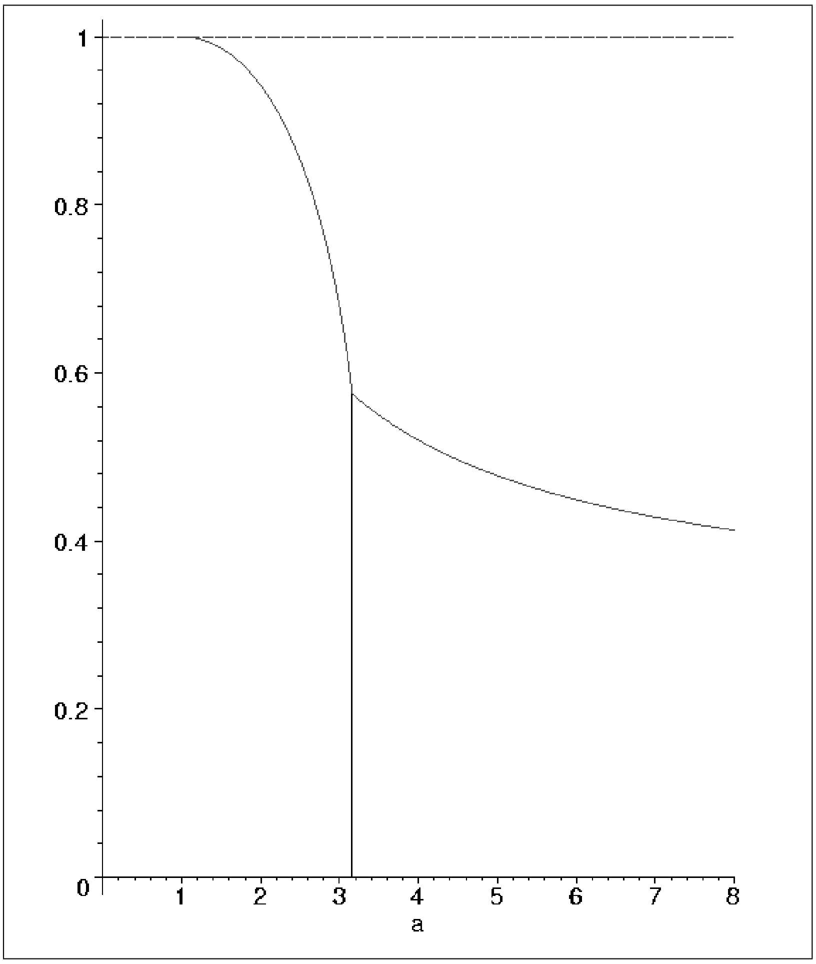
$$R_3 = \{(\alpha,\rho) : \rho \geq \rho^*(\alpha)\}$$

with

$$\begin{aligned} \rho^*(\alpha) &\equiv \frac{\alpha + \sqrt{\gamma(1+\gamma-\alpha^2)}}{1+\gamma}, \quad \alpha^2 < \gamma \\ &\equiv \frac{\gamma + \sqrt{\alpha^2(1+\gamma)-\gamma}}{\alpha(1+\gamma)}, \quad \alpha^2 \geq \gamma \end{aligned} \tag{2.10}$$

It is easily verified that for each fixed γ, $\rho^*(\alpha)$ is a decreasing function of α, with $\rho^*(1) = 1, \rho^*(\sqrt{\gamma}) = 2\sqrt{\gamma}/(1+\gamma)$ and $\rho^*(\alpha) \to 1/\sqrt{1+\gamma}$ as $\alpha \to \infty$. Furthermore, $\rho^*(\alpha)$ is concave in $(1, \sqrt{\gamma})$ and convex in $(\sqrt{\gamma}, \infty)$. These regions are shown for $\gamma = 10$ in Fig. 1 below.

Fig. 1. The three regions R_1, R_2, R_3 in the (α, ρ) plane where $\gamma = 10$.



Let $I_j = I[(\alpha, \rho) \in R_j]$ for $j = 1, 2, 3$, where I is the indicator function. Since the optimal sampling scheme involves sampling only one type of data, from (2.2) and (2.3), the estimator of μ simplifies to

$$\hat{\mu} = \bar{X}_1 \ I_1 + \bar{X}_2 \ I_2 + [l_1 \bar{X}_1^* + (1 - l_1)\bar{X}_2^*)] \ I_3 \qquad (2.11)$$

where $l_1 = (\sigma_{22} - \sigma_{12})/(\sigma_{11} + \sigma_{22} - 2\sigma_{12})$, and its variance, from (2.4), simplifies to

$$var(\hat{\mu}) = \frac{c_1 \sigma_{11}}{C - C_0} \ I_1 + \frac{c_2 \sigma_{22}}{C - C_0} \ I_2 + \frac{(c_1 + c_2)(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}{(C - C_0)(\sigma_{11} + \sigma_{22} - 2\sigma_{12})} \ I_3 \qquad (2.12)$$

where C, C_0, c_1, c_2 are as in (2.1).

3. Sampling Scheme when Σ is Unknown

When Σ is unknown the sampling scheme of the preceding section is not implementable. Instead, we may suggest a two-stage sampling scheme as follows. One may split the total sampling budget into two components: first-stage sampling budget $C^{(1)}$ and second-stage sampling budget $C^{(2)}$. Thus

$$C = C_0 + C^{(1)} + C^{(2)} \tag{3.1}$$

where C_0 is the overhead cost. The unit sampling costs are assumed to remain the same in both stages. Based on the first-stage bivariate sample of size m costing $C^{(1)}$, Σ is first estimated. This estimate is then used to determine the data-driven optimal type of data to be obtained at second stage, and such data is sampled using the remaining budget $C^{(2)}$. Note that if the first-stage budget is too small, the estimate of Σ will have high degree of error, resulting in a high probability of suboptimal sampling in the second stage. On the other hand, if the first-stage budget is too large, then the second-stage sample size will be small, resulting in inefficient estimate of μ at the conclusion of the two-stage sampling procedure. Thus, care must be taken to determine the ideal number of bivariate observations in the first stage. We address this issue in Subsection 3.4.

3.1 Sampling strategy, estimation of common mean and its accuracy

Suppose that during the first stage m observations are taken on bivariate data (X_1, X_2). We compute $(\hat{\alpha}, \hat{\rho})$ based on the first stage sample. Specifically, $\hat{\alpha}^2 = \hat{\sigma}_{11}^{(1)}/\hat{\sigma}_{22}^{(1)}$ is the ratio of first-stage sample variances and $\hat{\rho} = r^{(1)}$ is the first-stage sample correlation coefficient. Then we imitate the optimal sampling scheme of the previous section, replacing (α, ρ) by $(\hat{\alpha}, \hat{\rho})$, to determine which type of data should be taken in the second stage. We denote the second stage sample size by $n_1^{(2)}$ or $n_2^{(2)}$ or $n_{12}^{(2)}$ according as exclusive X_1 data, exclusive X_2 data or bivariate data are to be selected in the second stage.

The estimate of μ based on the two-stage sampling may be obtained in one of the following two ways:

(A) Using only the second-stage data obtain an estimate of μ, say $\hat{\mu}^{(2)}$, as in (2.11).

(B) Using the data of the second-stage together with the appropriate marginal/bivariate data of the first stage obtain an estimate of μ, say $\hat{\mu}^{(t)}$ as in (2.11).

Thus for approach (A), the estimate of common mean μ is

$$\hat{\mu}^{(2)} = \bar{X}_1^{(2)}\ \hat{I}_1 + \bar{X}_2^{(2)}\ \hat{I}_2 + [\hat{l}_1 \bar{X}_1^{*(2)} + (1 - \hat{l}_1)\bar{X}_2^{*(2)}]\ \hat{I}_3 \tag{3.2}$$

where $\bar{X}_i^{(2)}[\bar{X}_i^{*(2)}]$ denotes the sampling mean of X_i[the i-th component of (X_1, X_2)] based on second-stage data exclusively on X_i[bivariate data (X_1, X_2)], i=1,2. Further $\hat{I}_i = I[(\hat{\alpha}, \hat{\rho}) \in R_i]$ for $i = 1, 2, 3$, and $\hat{l}_1 = (\hat{\sigma}_{22}^{(1)} - \hat{\sigma}_{12}^{(1)})/(\hat{\sigma}_{11}^{(1)} + \hat{\sigma}_{22}^{(1)} - 2\hat{\sigma}_{12}^{(1)})$. Since given $\hat{\Sigma}$, the conditional expectation of $\hat{\mu}^{(2)}$ is μ, we have $E[\hat{\mu}^{(2)}] = \mu$ and

$$var(\hat{\mu}^{(2)}) = P_1\ \frac{\sigma_{11}}{n_1^{(2)}} + P_2\ \frac{\sigma_{22}}{n_2^{(2)}} + P_3\ \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{(\sigma_{11} + \sigma_{22} - 2\sigma_{12})\ n_{12}^{(2)}}. \tag{3.3}$$

where P_1 =Prob$[\hat{I}_1 = 1], P_2$ =Prob$[\hat{I}_2 = 1]$ and $P_3 = 1 - P_1 - P_2$.

Likewise, for approach (B), the estimate of μ is

$$\hat{\mu}^{(t)} = \bar{X}_1^{(t)}\ \hat{I}_1 + \bar{X}_2^{(t)}\ \hat{I}_2 + [\hat{l}_1 \bar{X}_1^{(t)} + (1 - \hat{l}_1)\bar{X}_2^{(t)}]\ \hat{I}_3 \tag{3.4}$$

where $\bar{X}_i^{(t)}[\bar{X}_i^{*(t)}]$ denotes the sample mean of X_i [the i-th component of (X_1, X_2)] based on the data of the second stage together with the appropriate marginal [bivariate] data of the first stage. By the same conditional argument it follows that $E[\hat{\mu}^{(t)}] = \mu$ and

$$var(\hat{\mu}^{(t)}) = P_1 \frac{\sigma_{11}}{m + n_1^{(2)}} + P_2 \frac{\sigma_{22}}{m + n_2^{(2)}} + P_3 \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{(\sigma_{11} + \sigma_{22} - 2\sigma_{12})\ (m + n_{12}^{(2)})}. \tag{3.5}$$

Clearly, as anticipated, $\mathrm{var}(\hat{\mu}^{(t)}) < \mathrm{var}(\hat{\mu}^{(2)})$. To evaluate the standard error of the estimates $\hat{\mu}^{(2)}$ and $\hat{\mu}^{(t)}$ using (3.3) and (3.5) respectively, we may resort to estimating the P_i's through simulation. In Example 1 of Subsection 3.3 we give such a numerical illustration.

3.2 Expected penalty of the two-stage strategy

When (α, ρ) is unknown, we estimate it by $(\hat{\alpha}, \hat{\rho})$ and check whether this estimate belongs to R_1, R_2 or R_3, in order to determine which type of data should be sampled in the second stage. When in reality $(\alpha, \rho) \in R_j$ but due to randomness in the first-stage data we observe $(\hat{\alpha}, \hat{\rho}) \in R_i$, the conditional penalty, for not knowing Σ (and hence using the two-stage strategy), is defined to be the ratio

$$L_{i|j} = \frac{var(\hat{\mu}^{(2)} | (\hat{\alpha}, \hat{\rho}) \in R_i)}{var(\hat{\mu}^{(2)} | (\alpha, \rho) \in R_j)}. \tag{3.6}$$

Clearly, $L_{j|j} = 1$ for $j = 1, 2, 3$. Next, notice from (3.3) that

$$var(\hat{\mu}^{(2)}) = \left[P_1 \frac{\alpha^2}{1+\gamma} + P_2 \frac{\gamma}{1+\gamma} + P_3 \frac{\alpha^2(1-\rho)^2}{1+\alpha^2-2\alpha\rho} \right] \frac{\sigma_{22}}{n_{12}^{(2)}}. \tag{3.7}$$

Hence, from (3.6) and (3.7), the expressions for $L_{i|j}$'s simplify as follows:

$L_{i\|j}$	$j = 1$	$j = 2$	$j = 3$
$i = 1$	1	$\frac{\alpha^2}{\gamma}$	$\frac{1+\alpha^2-2\alpha\rho}{(1+\gamma)(1-\rho^2)}$
$i = 2$	$\frac{\gamma}{\alpha^2}$	1	$\frac{\gamma(1+\alpha^2-2\alpha\rho)}{\alpha^2(1+\gamma)(1-\rho^2)}$
$i = 3$	$\frac{(1+\gamma)(1-\rho^2)}{1+\alpha^2-2\alpha\rho}$	$\frac{\alpha^2(1+\gamma)(1-\rho^2)}{\gamma(1+\alpha^2-2\alpha\rho)}$	1

When $(\alpha, \rho) \in R_j$, the expected penalty, for not knowing Σ, is defined by $P_1 L_{1|j} + P_2 L_{2|j} + P_3 L_{3|j}$, for $j = 1, 2, 3$. This expected penalty may be evaluated by estimating the classification probabilities $P_i, i = 1, 2, 3$, through simulation. See Example 1 of Subsection 3.3 for an illustration.

3.3 Simulation study

In order to obtain the standard errors of estimates of the common mean, or to evaluate the expected penalty for not knowing Σ, we need to estimate the classification probabilities $P_i, i = 1, 2, 3$, which depend on the joint distribution of $(\hat{\alpha}, \hat{\rho})$ and are analytically intractable. Therefore, we resort to a simulation study, as described below.

Simulation 1. (Estimating classification probabilities)
Given γ and m, we consider a grid of values of (α, ρ) such that α^2 ranges over $(\gamma/3, 3\gamma)$ and $\rho = .1(.1).9$. For each (α, ρ) combination considered, we estimate P_i by first repeating 1000 times steps (a)–(c) below, and then computing the proportion of times $(\hat{\alpha}, \hat{\rho})$ falling into R_i for $i = 1, 2, 3$.

(a) Draw m bivariate normal data with mean $\binom{5}{5}$ and dispersion $\begin{bmatrix} \alpha^2 & \rho\alpha \\ \rho\alpha & 1 \end{bmatrix}$

(b) Calculate $\hat{\alpha} = s_1/s_2$, the ratio of sample standard deviations and $\hat{\rho} = r$, the sample correlation

(c) Determine whether $(\hat{\alpha}, \hat{\rho})$ falls into R_1, R_2 or R_3.

In Example 1 below, we illustrate the estimation of classification probabilities, expected penalties and the standard errors of $\hat{\mu}^{(2)}$ and $\hat{\mu}^{(t)}$.

Example 1. Suppose that $\gamma = 10$ and that there is enough budget to take 50 bivariate observations in the two stages combined. Suppose that $m = 10$ bivariate observations are taken in the first stage, followed by 440 additional samples from X_1, or 44 additional samples from X_2, or 40 additional samples from bivariate data according as $(\hat{\alpha}, \hat{\rho})$ belongs to R_1, R_2 or R_3. Table 1 below records the estimated probabilities of classification, estimated expected penalty and the standard errors of $\hat{\mu}^{(2)}$ and $\hat{\mu}^{(t)}$ for some selected (α, ρ) combinations.

Table 1.
Estimated[a] classification probabilities, estimated expected penalty and standard errors of estimates of the common mean when $\gamma = 10, m = 10$.

α	ρ	$(\alpha,\rho) \in$	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{E}$[penalty]	SE($\hat{\mu}^{(2)}$)	SE($\hat{\mu}^{(t)}$)
2	.9	R_1	.649	.000	.351	1.173	.103	.098
	.8	R_1	.836	.005	.159	1.198	.104	.100
	.7	R_1	.882	.011	.107	1.182	.104	.100
	.6	R_1	.889	.025	.086	1.184	.104	.100
	.5	R_1	.910	.048	.042	1.146	.102	.099
	.4	R_1	.924	.051	.025	1.119	.101	.098
	.3	R_1	.909	.072	.019	1.139	.102	.099
	.2	R_1	.896	.089	.015	1.156	.103	.099
	.1	R_1	.890	.098	.012	1.163	.103	.100
4	.9	R_3	.014	.006	.980	1.063	.091	.082
	.8	R_3	.065	.053	.882	1.145	.125	.113
	.7	R_3	.121	.115	.764	1.156	.144	.132
	.6	R_3	.169	.244	.587	1.144	.155	.143
	.5	R_2	.179	.353	.468	1.115	.159	.147
	.4	R_2	.234	.459	.307	1.162	.163	.151
	.3	R_2	.207	.590	.203	1.144	.161	.149
	.2	R_2	.245	.617	.138	1.160	.162	.151
	.1	R_2	.230	.681	.089	1.145	.161	.150
6	.9	R_3	.000	.004	.996	1.010	.081	.073
	.8	R_3	.001	.040	.959	1.043	.111	.099
	.7	R_3	.008	.138	.854	1.090	.132	.119
	.6	R_3	.014	.240	.746	1.087	.145	.131
	.5	R_3	.020	.360	.620	1.071	.153	.138
	.4	R_2	.020	.542	.438	1.066	.156	.141
	.3	R_2	.028	.645	.327	1.099	.158	.144
	.2	R_2	.022	.746	.232	1.080	.157	.142
	.1	R_2	.040	.791	.169	1.120	.160	.146

[a]Based on 1000 repetitions.

3.4 Determining the first-stage sample size

Given γ and some preselected $\delta > 0$, we define the optimal first-stage sample size $m^* = m^*(\gamma, \delta)$ to be the smallest m such that the supremum of the expected penalty, of not knowing Σ, is bounded above by $1 + \delta$. Here, the supremum is taken over all (α, ρ).

Given γ and $\delta > 0$, to determine the optimal first-stage sample size, we do the following computations for each selected value of m: We consider a grid of values of (α, ρ) such that α^2 ranges over $(\gamma/3, 3\gamma)$ and $\rho = .1(.1).9$. For each (α, ρ) combination considered, we calculate the expected penalty as described in Subsection 3.2, when the first stage sample size is m. Then we obtain the maximum expected penalty as (α, ρ) ranges over the chosen grid.

Table 2 below records the maximum expected penalty as α^2 ranges over $(\gamma/3, 3\gamma)$ and $\rho = .1(.1).9$ for various values of γ and m.

Table 2.

Maximum expected penalty for various γ and m.

	γ					
m	2	5	10	25	50	100
5	1.4219	1.3870	1.4962	1.5785	1.4599	1.5369
10	1.1690	1.1726	1.1902	1.2214	1.1941	1.2401
15	1.1206	1.1157	1.1444	1.1504	1.1408	1.1553
20	1.0930	1.1011	1.1088	1.1196	1.1184	1.1251
25	1.0818	1.0846	1.0965	1.0983	1.0994	1.1115
50	1.0546	1.0585	1.0614	1.0606	1.0555	1.0680
100	1.0361	1.0394	1.0384	1.0391	1.0405	1.0407

Table 2 is used to approximate the optimal first-stage sample size. For example, if $\gamma = 10$ and $\delta = .20$ then it suffices to take $m^* = 10$. However, if $\gamma = 10$ and $\delta = .10$ then we must take $m^* = 25$. Note that for each fixed γ, as δ decreases the optimal first-stage sample size increases. Also for each fixed δ, as γ changes, the optimal first-stage sample size remains more or less stable.

The choice of δ in the above determination of the first-stage sample size is a subjective issue. Alternatively, one may choose the optimal first-stage sample size by minimizing the standard error of $\hat{\mu}^{(2)}$ using (3.3) (or that of $\hat{\mu}^{(t)}$ using (3.5)) over all (α, ρ) combinations. However, one must recognize that because of random variations in estimating the classification probabilities, a strict uniform minimization may not be always possible. In such cases, the maximum standard error over all (α, ρ) combinations may be minimized.

Table 3 below records the standard errors of $\hat{\mu}^{(2)}$ and $\hat{\mu}^{(t)}$, when $\gamma = 10$ and there is enough budget to obtain 50 bivariate observations in the two stages combined, for various values of m. Note that for $m = 50$, the entire budget is used up in one stage on bivariate data and there is no second-stage data. Hence, in this case, there is no $\hat{\mu}^{(2)}$.

Table 3.
Standard error of estimates of the common mean when
$\gamma = 10, m + n_{12}^{(2)} = 50$.

			$10^3\times$ SE($\hat{\mu}^{(2)}$)				$10^3\times$ SE($\hat{\mu}^{(t)}$)				
			m				m				
α	ρ	$(\alpha,\rho)\in$	5	10	15	25	5	10	15	25	50
2	.9	R_1	100	103	109	126	97	98	102	114	272
	.5	R_1	106	102	108	122	104	99	104	116	283
	.1	R_1	104	103	106	122	101	100	103	116	283
4	.9	R_3	94	91	94	111	90	82	79	79	79
	.5	R_2	153	159	170	198	148	147	148	150	170
	.1	R_2	157	161	173	199	152	150	153	153	190
6	.9	R_3	82	81	86	102	78	73	72	72	72
	.5	R_3	155	153	161	189	149	138	136	135	135
	.1	R_2	166	160	167	193	160	146	141	139	141

From Table 3 we conclude that for $\gamma = 10$, if there is budget for 50 bivariate data then the optimal first-stage sample size is $m^* = 10$, no matter whether $\hat{\mu}^{(2)}$ or $\hat{\mu}^{(t)}$ is used to estimate the common mean. In particular, therefore, Table 3 shows that it is not optimal to use up the entire budget by taking all bivariate data in the first stage, unless *a priori* it is known that $(\alpha, \rho) \in R_3$.

Similar calculations show that when there is budget for only 30 bivariate data $m^* = 5$, and when there is budget for 500 bivariate data m^* can be any number between 10 and 25. Details are not shown for the sake of brevity.

3.5 Properties of estimators under the two-stage strategy

The sampling distribution of various estimators under the two-stage sampling strategy are studied through a simulation. Table 4 below presents some selected results for $\gamma = 10, m^* = 10, n_{12}^{(2)} = 40, n_1^{(2)} = 440, n_2^{(2)} = 44$. The unbiasedness of $\hat{\mu}^{(2)}$ and $\hat{\mu}^{(t)}$ and their variance formulas (3.3) and (3.5) are supported by the simulation results.

Table 4.
Mean and standard error[b] (within parentheses) of estimates when $\gamma = 10, m^* = 10, n_{12}^{(2)} = 40, n_1^{(2)} = 440, n_2^{(2)} = 44$.

α	ρ	$(\alpha, \rho) \in$	$\hat{\sigma}_{22}^{(1)}$	$\hat{\alpha}^{(1)}$	$\hat{\rho}^{(1)}$	$\hat{\mu}^{(2)}$	$\hat{\mu}^{(t)}$
2	.9	R_1	0.975	2.020	0.891	4.997	4.993
			(0.207)	(0.373)	(0.066)	(0.114)	(0.101)
	.5	R_1	1.005	2.021	0.488	5.010	5.011
			(0.221)	(0.569)	(0.272)	(0.104)	(0.108)
	.1	R_1	1.011	2.152	0.047	4.999	4.996
			(0.240)	(0.857)	(0.341)	(0.099)	(0.095)
4	.9	R_3	0.970	4.039	0.902	4.999	4.999
			(0.218)	(0.607)	(0.086)	(0.097)	(0.085)
	.5	R_2	0.958	4.300	0.434	4.980	4.984
			(0.253)	(1.235)	(0.299)	(0.162)	(0.153)
	.1	R_2	0.954	4.303	0.067	4.984	4.984
			(0.242)	(1.493)	(0.340)	(0.181)	(0.179)
6	.9	R_3	0.928	6.023	0.880	5.006	5.007
			(0.228)	(0.951)	(0.083)	(0.083)	(0.071)
	.5	R_3	0.924	6.122	0.441	5.016	5.014
			(0.219)	(1.719)	(0.284)	(0.161)	(0.152)
	.1	R_2	0.986	6.384	0.125	4.984	4.981
			(0.251)	(1.836)	(0.310)	(0.152)	(0.136)

[b]Based on 100 repetitions.

4. Concluding Section

Motivated by a practical example, we have addressed the question of optimal allocation of resources in the context of most efficient estimation of the common mean of two continuous outputs from a bivariate normal population. The differential sampling costs make the problem intriguing. We have attempted to present our initial thoughts on this fascinating problem.

PART II : ESTIMATION OF THE SIZE OF A FINITE POPULATION

CHAPTER 5

UNBIASED ESTIMATION OF THE SIZE OF A FINITE POPULATION USING CAPTURE-MARK-RELEASE-RECAPTURE SAMPLING SCHEME I

The problem considered is that of unbiased estimation of the size (N) of a finite closed population under *Capture-Mark-Release-Recapture*(CMRR) sequential sampling procedure. The existing results are supplemented with various other results and the CMRR procedure is compared with the Negative Binomial and the Negative Hypergeometric sampling schemes in terms of the ASN and the variance of the UMVUE of the population size. The maximum likelihood estimator of population size N is also derived and its performance is compared with that of UMVUE of the population size N.

1. INTRODUCTION

The problem of estimation of the population size (N) of a finite closed population is known to be of great importance. Well known problems of this kind are the estimation of the total number of fish in a lake, the estimation of total number of wild animals in a forest etc. Several authors have already considered the problem in the past and suggested different methods of sampling with associated estimation procedures (see Boswell et al. (1988), Seber (1982) and the references therein). The basic procedure is to initially catch, mark and release k population units into the target population and then to recatch units randomly from the population in one or more samples.

For unbiased estimation of N, a simple procedure (to be called procedure I) is to recatch and release units one by one until m $(\leq k)$ of k initially marked units are recaptured. If S_m denotes the number of trials required, then S_m follows a Negative Binomial distribution with success probability $\frac{k}{N}$ and the uniformly minimum variance unbiased estimator (UMVUE) of N is obtained from the well known results on Negative Binomial distribution (see Johnson and Kotz (1969), page 126) as $\hat{N}_I = \frac{kS_m}{m}$ with variance $V(\hat{N}_I) = \frac{N(N-k)}{m}$. Also the expected number of trials of the procedure is $ASN(I) = E(S_m) = \frac{mN}{k}$.

If units are sampled one by one without being replaced into the population until m ($\leq k$) of the k initially marked units are recaptured (to be called procedure II), then S_m, the number of trials required, follows a Negative Hypergeometric distribution and in this case the UMVUE of N is given by $\hat{N}_{II} = \frac{(k+1)S_m}{m} - 1$ with $V(\hat{N}_{II}) = \frac{(N+1)(N-k)(k+1-m)}{m(k+2)}$ and $ASN(II) = E(S_m) = \frac{m(N+1)}{k+1}$ (see Johnson and Kotz (1969), page 157).

A simple modification of the procedure I (to be called procedure III) is also suggested in the literature as follows: initially k population units are marked and released into the target population and then units are sampled at random, marked and released one by one until m marked units are recaptured. The procedure is a special case of a more general procedure suggested in Goodman (1953) and is termed as capture-mark-release-recapture (CMRR) sampling scheme. Using more general methods, Goodman (1953) obtained the UMVUE of N for this procedure as the quotient of two determinants and gave some simplified expression for $k = 1$. Darroch (1958) had shown that Goodman's estimator, for $k = 1$, can also be expressed as the ratio of two Sterling numbers or differences of zero. Hossain (1995) had considered the special case of $k = m = 1$ in which case the UMVUE of N is $\binom{S_1+1}{2}$, where S_1 is the number of trials required. Maximum likelihood estimation of N had been discussed in Darroch (1985) and Samuel (1968). Some other aspects of the procedure have been studied among others by Ahmad et. al (1993,1999), Ahmad et. al (1999) and Ahmad and Chaubay (2000).

The purpose of the Chapter is to supplement these studies with various other results and to compare procedure III with procedures I and II in terms of the ASN and the variance of the UMVUE of N. It is demonstrated in section 4 that the procedure III is always better than the procedure I and also appears to be better than procedure II when N is considerably large. The supplementary results are discussed in section 2 and 3.

2. ESTIMATION OF POPULATION SIZE UNDER CMRR PROCEDURE

Let us consider the procedure III and let for $j = 1, \cdots, m,\ S_j$ denote the number of trials required to recapture j marked units and $S_j^* = S_j - j$. It is easy to verify that

$$\begin{aligned}
& P[S_1^* = s_1^*, \cdots, S_m^* = s_m^*] \\
= \ & (1 - \frac{k}{N})(1 - \frac{k+1}{N}) \cdots (1 - \frac{s_m^* + k - 1}{N}) \frac{s_1^* + k}{N} \cdots \frac{s_m^* + k}{N} \\
= \ & \binom{N-k}{s_m^*} \frac{s_m^*!}{N^{s_m^*+m}} (s_1^* + k) \cdots (s_m^* + k)
\end{aligned}$$

$$0 \leq s_1^* \leq \cdots \leq s_m^* \leq N - k \tag{2.1}$$

whence the probability distribution of S_m^* is obtained as

$$\begin{aligned}
& P[S_m^* = s_m^*] \\
= \ & \binom{N-k}{s_m^*} \frac{s_m^* + k}{N^{s_m^*+m}} s_m^*! \sum \cdots \sum\nolimits_{0 \leq s_1^* \leq \cdots \leq s_{m-1}^* \leq s_m^*} (s_1^* + k) \cdots (s_{m-1}^* + k) \\
= \ & \binom{N-k}{s_m^*} \frac{s_m^* + k}{N^{s_m^*+m}} s_m^*! \sum \cdots \sum\nolimits_{k \leq s_1^* + k \leq \cdots \leq s_{m-1}^* + k \leq s_m^* + k} (s_1^* + k) \cdots (s_{m-1}^* + k) \\
= \ & \binom{N-k}{s_m^*} \frac{s_m^* + k}{N^{s_m^*+m}} \ g_{m-1}(k, s_m^* + k)
\end{aligned}$$

$$0 \leq s_m^* \leq N - k \tag{2.2}$$

where for non-negative integers $a, b(\geq a), c,\ g_c(a, b)$ is defined as

$$g_c(a, b) = (b - a)! \sum \cdots \sum\nolimits_{a \leq i_1 \leq \cdots \leq i_c \leq b} i_1 \cdots i_c \tag{2.3}$$

with $g_0(a, b) = (b - a)!$.

The function $g_c(a,b)$ can also be expressed as

$$\begin{aligned}
g_c(a,b) &= \Delta^{b-a} a^{b-a+c} \\
&= \Delta^{b-a} x^{b-a+c}|_{x=a} \\
&= (E-1)^{b-a} x^{b-a+c}|_{x=a} \\
&= \{E^{b-a} - \binom{b-a}{1} E^{b-a-1} + \binom{b-a}{2} E^{b-a-2} - \cdots \\
&\quad +(-1)^{b-a}\} x^{b-a+c}|_{x=a} \\
&= (x+b-a)^{b-a+c} - \binom{b-a}{1}(x+b-a-1)^{b-a+c} \\
&\quad + \binom{b-a}{2}(x+b-a-2)^{b-a+c} - \cdots + (-1)^{b-a} x^{b-a+c}|_{x=a} \\
&= b^{b-a+c} - \binom{b-a}{1}(b-1)^{b-a+c} \\
&\quad + \binom{b-a}{2}(b-2)^{b-a+c} - \cdots + (-1)^{b-a} a^{b-a+c}
\end{aligned} \tag{2.4}$$

and this has been proved in Appendix A.
A useful identity follows from (2.2) viz.

$$\sum_{s_m^*=0}^{N-k} \binom{N-k}{s_m^*} \frac{s_m^*+k}{N^{s_m^*}} g_{m-1}(k, s_m^*+k) = N^m \tag{2.5}$$

From (2.1) and (2.2) it follows that

$$\begin{aligned}
P_N[S_1^* = s_1^*, \cdots, S_m^* = s_m^* | S_m^* = s_m^*] &= \frac{P_N[S_1^* = s_1^*, \cdots, S_m^* = s_m^*]}{P_N[S_m^* = s_m^*]} \\
&= \frac{s_m^*!}{g_{m-1}(k, s_m^*+k)} (s_1^*+k) \cdots (s_{m-1}^*+k)
\end{aligned}$$

is independent of N. This implies that S_m^* is sufficient. Also S_m^* is complete which is proved in Appendix B. Thus S_m or equivalently S_m^* is a complete sufficient statistic for the parameter space $\{N \geq k\}$. It follows, therefore, that the unbiased estimator of N based on S_m^* is also the UMVUE which is obtained in the following theorem.

Theorem 2.1 : The UMVUE of N for the suggested procedure is given by $\hat{N}_{III} = \frac{g_m(k,S_m^*+k)}{g_{m-1}(k,S_m^*+k)} = \frac{\Delta^{S_m^*} k^{S_m^*+m}}{\Delta^{S_m^*} k^{S_m^*+m-1}}$.

Proof. It is enough to prove that $\hat{N}_{III}$ is an unbiased estimator of N. Now,

$$
\begin{aligned}
& E(\hat{N}_{III}(S_m^*)) \\
= & \sum_{s_m^*=0}^{N-k} \frac{g_m(k, s_m^*+k)}{g_{m-1}(k, s_m^*+k)}\ P_N(S_m^* = s_m^*) \\
= & \sum_{s_m^*=0}^{N-k} \frac{g_m(k, s_m^*+k)}{g_{m-1}(k, s_m^*+k)} \binom{N-k}{s_m^*} \frac{s_m^*+k}{N^{s_m^*+m}}\ g_{m-1}(k, s_m^*+k),\ from\ (2.2) \\
= & \sum_{s_m^*=0}^{N-k} \binom{N-k}{s_m^*} \frac{s_m^*+k}{N^{s_m^*+m}}\ g_m(k, s_m^*+k) \\
= & \frac{N^{m+1}}{N^m},\ from\ (2.5) \\
= & N
\end{aligned}
$$

QED.

In particular when $k = 1$, the UMVUE of N, as obtained in Goodman(1953) and Darroch(1958), is

$$
\frac{g_m(1, S_m^*+1)}{g_{m-1}(1, S_m^*+1)} = \frac{\Delta^{S_m^*} 1^{S_m^*+m}}{\Delta^{S_m^*} 1^{S_m^*+m-1}} = \frac{\Delta^{S_m^*+1} O^{S_m^*+m+1}}{\Delta^{S_m^*+1} O^{S_m^*+m}}
$$

using the identity

$$
\begin{aligned}
\frac{1}{\nu}\Delta^{\nu} O^{\nu+m} &= \frac{1}{\nu}\Delta^{\nu} x^{\nu+m}|_{x=0} \\
&= \frac{1}{\nu}(E-1)^{\nu} x^{\nu+m}|_{x=0} \\
&= \frac{1}{\nu}[\nu^{\nu+m} - \binom{\nu}{1}(\nu-1)^{\nu+m} + \binom{\nu}{2}(\nu-2)^{\nu+m} - \cdots + (-1)^{\nu-1}\nu] \\
&= \nu^{\nu+m-1} - \binom{\nu-1}{1}(\nu-1)^{\nu+m-1} + \binom{\nu-2}{2}(\nu-2)^{\nu+m-1} - \cdots + (-1)^{\nu-1} \\
&= (E-1)^{\nu-1} x^{\nu+m-1}|_{x=1} \\
&= \Delta^{\nu-1} x^{\nu+m-1}|_{x=1} \\
&= \Delta^{\nu-1} 1^{\nu+m-1}
\end{aligned}
$$

If, further, $m = 1$, the UMVUE of N is

$$\frac{g_1(1, S_1^* + 1)}{g_0(1, S_1^* + 1)} = \sum_{i=1}^{S_1^*+1} i = \binom{S_1^* + 2}{2} = \binom{S_1 + 1}{2} \tag{2.6}$$

as given in Hossain(1995).

It follows similarly that the UMVUE of N^2 is $\frac{g_{m+1}(k, S_m^* + k)}{g_{m-1}(k, S_m^* + k)}$ and hence, the UMVUE of $V(\hat{N}_{III})$ is

$$\begin{aligned}
\hat{V}(\hat{N}_{III}) &= \hat{N}_{III}^2 - \frac{g_{m+1}(k, S_m^* + k)}{g_{m-1}(k, S_m^* + k)} \\
&= \frac{g_m^2(k, S_m^* + k) - g_{m-1}(k, S_m^* + k) g_{m+1}(k, S_m^* + k)}{g_{m-1}^2(k, S_m^* + k)} \\
&= \frac{(\Delta^{S_m^*} k^{S_m^* + m})^2 - \Delta^{S_m^*} k^{S_m^* + m - 1} \Delta^{S_m^*} k^{S_m^* + m + 1}}{(\Delta^{S_m^*} k^{S_m^* + m - 1})^2}
\end{aligned}$$

It can be proved by induction on b for fixed a that

$$g_c^2(a, b) \geq g_{c-1}(a, b) g_{c+1}(a, b) \tag{2.7}$$

noting that (2.7) is trivially true for b=a whatever be c (see Hardy *et al.*(1952), page 52). The inequality (2.7) ensures that $\hat{V}(\hat{N}_{III})$ is uniformly non-negative.

3. THE SPECIAL CASE OF m=1

In particular when $m = 1$, the probability distribution of S_1^* simplifies to

$$P[S_1^* = s_1^*] = (1 - \frac{k}{N})(1 - \frac{k+1}{N}) \cdots (1 - \frac{s_1^* + k - 1}{N})\frac{s_1^* + k}{N}$$

$$0 \le s_1^* \le N - k \tag{3.1}$$

and the estimator reduces to

$$\begin{aligned}
\hat{N}_{III} &= \sum_{k}^{S_1^*+k} i = \binom{S_1^* + k + 1}{2} - \binom{k}{2} \\
\hat{V}(\hat{N}_{III}) &= (\sum_{k}^{S_1^*+k} i)^2 - \sum\sum\nolimits_{k \le i_1 \le i_2 \le S_1^*+k} i_1 i_2 \\
&= \sum\sum\nolimits_{k \le i_1 \le i_2 \le S_1^*+k} i_1 i_2 - \sum_{k}^{S_1^*+k} i^2 \\
&= \sum_{k}^{S_1^*+k} i\{\binom{i+1}{2} - \binom{k}{2}\} - \sum_{k}^{S_1^*+k} i^2 \\
&= \sum_{k}^{S_1^*+k} i\{\binom{i}{2} - \binom{k}{2}\} \\
&= \frac{(S_1^* + k - 1)(S_1^* + k)(S_1^* + k + 1)(3S_1^* + 3k + 2)}{24} \\
&\quad - \binom{k}{2}\binom{S_1^* + k + 1}{2} + \frac{(k-1)k(k+1)(3k-2)}{24}
\end{aligned}$$

For $m = 1$, we derive a closed expression for $V(\hat{N}_{III})$ in theorem 3.1 stated below. The derivation is based on the following lemmas.

Lemma 3.1 : For every $i = 0, 1, \cdots, N - k$,

$$\sum_{s_1^*=i}^{N-k} P[S_1^* = s_1^*] = \frac{N}{i+k} P[S_1^* = i]$$

Proof. It is easy to verify from (3.1)that

$$\begin{aligned} P[S_1^* = j+1] &= P[S_1^* = j](1 - \frac{j+k}{N})\frac{j+k+1}{j+k}, \\ j &= 0, 1, \cdots, N-k-1 \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{s_1^*=N-k-1}^{N-k} P[S_1^* = s_1^*] &= P[S_1^* = N-k-1][1 + (1 - \frac{N-1}{N})\frac{N}{N-1}] \\ &= \frac{N}{N-1} P[S_1^* = N-k-1] \\ \sum_{s_1^*=N-k-2}^{N-k} P[S_1^* = s_1^*] &= P[S_1^* = N-k-2][1 + \frac{N}{N-1}(1 - \frac{N-2}{N})\frac{N-1}{N-2}] \\ &= \frac{N}{N-2} P[S_1^* = N-k-2] \\ &\vdots \\ \sum_{s_1^*=i}^{N-k} P[S_1^* = s_1^*] &= \frac{N}{i+k} P[S_1^* = i] \end{aligned}$$

QED.

Lemma 3.2 : For any given function f, $E\sum_{i=k}^{S_1^*+k} if(i) = NEf(S_1^*+k)$.

Proof:

$$\begin{aligned} E\sum_{i=k}^{S_1^*+k} if(i) &= \sum_{s_1^*=0}^{N-k} P[S_1^*=s_1^*]\sum_{i=k}^{s_1^*+k} if(i) \\ &= \sum_{i=k}^{N} if(i)\sum_{s_1^*=i-k}^{N-k} P[S_1^*=s_1^*] \\ &= \sum_{i=0}^{N-k}(i+k)f(i+k)\sum_{s_1^*=i}^{N-k} P[S_1^*=s_1^*] \\ &= N\sum_{i=0}^{N-k} f(i+k)P[S_1^*=i] \text{ (by lemma 3.1)} \\ &= NEf(S_1^*+k) \end{aligned}$$

QED.

Lemma 3.3 :

$$\begin{aligned} E(S_1^*+k) &= N\ E(\frac{1}{S_1^*+k}) + (k-1) \\ &= \frac{(N-k)!}{N^{N-k}}\{1+\frac{N}{1!}+\frac{N^2}{2!}+\cdots+\frac{N^{N-k}}{(N-k)!}\} + (k-1) \end{aligned}$$

Proof. By lemma 3.2,

$$E(S_1^*+k) = E\sum_{i=k}^{S_1^*+k}\frac{i}{i} + (k-1) = N\ E(\frac{1}{S_1^*+k}) + (k-1)$$

Also by (3.1),

$$\begin{aligned} N\ E(\frac{1}{S_1^*+k}) &= \sum_{s_1^*=0}^{N-k}(1-\frac{k}{N})\cdots(1-\frac{s_1^*+k-1}{N}) \\ &= \frac{(N-k)!}{N^{N-k}}\{1+\frac{N}{1!}+\frac{N^2}{2!}+\cdots+\frac{N^{N-k}}{(N-k)!}\} \end{aligned}$$

Hence, follows the proof. QED.

Theorem 3.1 : For $m = 1$,

$$\begin{aligned} V(\hat{N}_{III}) &= N^2 - N\ E(S_1^* + k) \\ &= N^2 - \frac{(N-k)!}{N^{N-k-1}}\{1 + \frac{N}{1!} + \frac{N^2}{2!} + \cdots + \frac{N^{N-k}}{(N-k)!}\} - N(k-1) \end{aligned}$$

Proof. For $m = 1$,

$$\begin{aligned} V(\hat{N}_{III}) &= E\hat{V}(\hat{N}_{III}) \\ &= E \sum_{i=k}^{S_1^*+k} i\{\binom{i}{2} - \binom{k}{2}\} \\ &= N\ E[\binom{S_1^*+k}{2} - \binom{k}{2}],\ \text{by lemma 3.2} \\ &= N\ E(\hat{N}_{III}) - N\ E(S_1^* + k) \\ &= N^2 - N\ E(S_1^* + k) \end{aligned}$$

whence the theorem follows by lemma 3.3. QED.

In the following Theorem we study the effect of k on $V(\hat{N}_{III}), E(S_1)$ or equivalently $E(S_1^*)$ and $E(S_1 + k)$ or equivalently $E(S_1^* + k)$.

Theorem 3.2 :
(a) $E(S_1^*)$ is decreasing in k,
(b) $E(S_1^* + k)$ is increasing in k,
(c) $V(\hat{N}_{III})$ is decreasing in k.

Proof. We write S_1^* as S_{1k}^* to denotes its dependence on k.

Proof (a). By Lemma (3.3),
$E(S_{1k+1}^* - S_{1k}^*) = N[E(\frac{1}{S_{1k+1}^*+k+1}) - E(\frac{1}{S_{1k}^*+k})]$ and

$$NE(\frac{1}{S_{1k+1}^* + k + 1}) = \frac{(N-k-1)!}{N^{N-k-1}} \sum_{x=0}^{N-k-1} \frac{N^{N-x}}{x!} - \frac{(N-k)!}{N^{N-k}} \sum_{x=0}^{N-k} \frac{N^{N-x}}{x!}$$

$$\begin{aligned} &= \frac{N}{N-k}\left[\frac{(N-k)!}{N^{N-k}}\sum_{x=0}^{N-k}\frac{N^{N-x}}{x!} - \frac{(N-k)!}{N^{N-k}}\ \frac{N^{N-k}}{(N-k)!}\right] \\ &= \frac{N}{N-k}\left[N\ E\left(\frac{1}{S_{1k}^*+k}\right)-1\right] \end{aligned}$$

Hence, $E(S_{1k+1}^*) - E(S_{1k}^*) = \frac{Nk}{N-k}E(\frac{1}{S_{1k}^*+k}) - \frac{N}{N-k} < 0$
as $E(\frac{1}{S_{1k}^*+k}) < \frac{1}{k}$. QED.

Proof (b). As in proof of (a)

$$\begin{aligned} [E(S_{1k+1}^*) + (k+1)] - [E(S_{1k}^*) + k] &= E(S_{1k+1}^* - S_{1k}^*) + 1 \\ &= \frac{Nk}{N-k}E\left(\frac{1}{S_{1k}^*+k}\right) - \frac{N}{N-k} + 1 \\ &= \frac{Nk}{N-k}\left[E\left(\frac{1}{S_{1k}^*+k}\right) - \frac{1}{N}\right] \\ &> 0 \end{aligned}$$

as $E(\frac{1}{S_{1k}^*+k}) > \frac{1}{N}$. QED.

Proof (c). The proof follows from (b) since by Theorem (3.1),
$V(\hat{N}_{III}) = N^2 - NE(S_1^* + k)$. QED.

Thus as the value of k is increased, although $V(\hat{N}_{III})$ decreases, the average number of units to be caught viz. $E(S_1^* + k)$ is increased. In Table 3.1 given below we compute numerically the effect of k on $E(S_1^* + k) \times V(\hat{N}_{III})$ which is the inverse of efficiency per unit to be caught. It is observed that this is increasing in k upto a certain value and then decreasing.

Table 3.1

$E(S_1^* + k) \times V(\hat{N}_{III})$ for different values of N and k

N	k	$E(S_1^* + k)$	$V(\hat{N}_{III})$	$E(S_1^* + k) \times V(\hat{N}_{III})$
10	1	3.6602	63.3978	232.0498
	2	3.9558	60.4421	239.0964
	3	4.4447	55.5526	246.9169
	4	5.0639	49.3608	249.9591
	5	5.7732	42.2680	244.0216
	6	6.5464	34.5360	226.0865
	7	7.3660	26.3400	194.0204
	8	8.2200	17.8000	146.3160
	9	9.1000	9.0000	81.9000
	10	10.0000	0.0000	0.0000
20	1	5.2936	294.1283	1556.9930
	3	5.9106	281.7875	1665.5400
	5	7.0303	259.3934	1823.6210
	8	9.2134	215.7315	1987.6260
	9	10.0224	199.5525	1999.9900
	10	10.8589	182.8227	1985.2470
	12	12.5950	148.1010	1865.3250
	14	14.3925	112.1503	1614.1210
	16	16.2332	75.3370	1222.9570
	18	18.1050	37.9000	686.1795
30	1	6.5495	703.5162	4607.6520
	4	7.5327	674.0183	5077.1940
	8	10.0839	597.4827	6024.9610
	13	14.1399	475.8037	6727.8060
	14	15.0116	449.6535	6749.9960
	15	15.8967	423.1004	6725.8800
	19	19.5382	313.8552	6132.1540
	23	23.2889	201.3322	4688.8110
	26	26.1476	115.5724	3021.9400
	29	29.0333	29.0000	841.9667

Table 3.1 (Contd.)

$E(S_1^* + k) \times V(\hat{N}_{III})$ for different values of N and k

N	k	$E(S_1^* + k)$	$V(\hat{N}_{III})$	$E(S_1^* + k) \times V(\hat{N}_{III})$
50	1	8.5431	2072.8440	17708.5700
	4	9.3576	2032.1220	19015.7100
	8	11.5975	1920.1260	22268.6400
	15	16.9847	1650.7670	28037.7100
	23	24.0826	1295.8700	31207.9200
	24	25.0048	1249.7600	31250.0000
	25	25.9323	1203.3850	31206.5400
	33	33.4937	825.3160	27642.8700
	40	40.2427	487.8663	19633.0400
	49	49.0200	49.0000	2401.9800
100	1	12.2100	8779.0040	107191.3000
	15	19.3577	8064.2280	156105.1000
	30	32.1280	6787.1970	218059.3000
	48	49.0407	5095.9290	249908.0000
	49	50.0014	4999.8630	250000.0000
	50	50.9635	4903.6540	249907.2000
	60	60.6493	3935.0730	238659.3000
	70	70.4202	2957.9790	208301.5000
	80	80.2462	1975.3780	158516.6000
	99	99.0100	99.0000	9801.9900
200	1	17.3984	36520.3100	635396.6000
	50	52.7997	29440.0700	1554425.0000
	80	81.4565	23708.7000	1931228.0000
	98	99.0202	20195.9600	1999808.0000
	99	100.0004	19999.9300	2000000.0000
	100	100.9809	19803.8100	1999808.0000
	130	130.5323	13893.5500	1813556.0000
	150	150.3304	9933.9140	1493370.0000
	170	170.1753	5964.9470	1015086.0000
	199	199.0050	199.0000	39602.0000

4. COMPARISON WITH NEGATIVE BINOMIAL AND NEGATIVE HYPERGEOMETRIC SAMPLING SCHEMES

We study below the performance of the procedure III in terms of the ASN i.e. $E(S_m)$ and the variance of the estimator of N as compared to the procedures I and II.

We first consider the special case of $m = 1$ and prove the following theorem.

Theorem 4.1 : for $m = 1$,
(a) $ASN(III) \leq ASN(II) \leq ASN(I)$,
(b) $V(\hat{N}_{II}) \leq V(\hat{N}_{III}) \leq V(\hat{N}_I)$.

Proof. (a) By (3.1) and lemma 3.1 and 3.3

$$\begin{aligned}
E(S_1) &= E(S_1^* + k) - (k-1) - N\ E(\frac{1}{S_1^* + k}) \\
&\leq N[k^{-1}P(S_1^* = 0) + (k+1)^{-1}\sum_{s_1^*=1}^{N-k} P(S_1^* = s_1^*)] \\
&= N[k^{-1}P(S_1^* = 0) + N(k+1)^{-2}P(S_1^* = 1)] \\
&= \frac{N+1}{k+1}
\end{aligned}$$

with equality iff $N = k$ or $k+1$. This proves that, for $m = 1$, $ASN(III) \leq ASN(II)$. Also by simple comparison of ASN(I) and ASN(II), it follows that $ASN(II) \leq ASN(I)$ with equality iff $N = k$. Hence follows the proof of (a). QED.

Proof. **(b)** By theorem 3.1, we have, for $m = 1$, $V(\hat{N}_{III}) = N^2 - N\ E(S_1^* + k)$ which is clearly $\leq N(N-k)$ with equality iff N=k. This proves that, for $m = 1$, $V(\hat{N}_{III}) \leq V(\hat{N}_I)$. Also by (a), $E(S_1^* + k) = E(S_1) + (k-1) \leq \frac{N+1}{k+1} + (k-1)$ implying that, for $m = 1$, $V(\hat{N}_{III}) \geq \frac{Nk(N-k)}{k+1}$. Hence for $m = 1$,

$$\begin{aligned} V(\hat{N}_{III}) - V(\hat{N}_{II}) &= V(\hat{N}_{III}) - \frac{(N+1)k(N-k)}{k+2} \\ &\geq \frac{Nk(N-k)}{k+1} - \frac{(N+1)k(N-k)}{k+2} \\ &= \frac{k(N-k)(N-k-1)}{(k+1)(k+2)} \\ &\geq 0 \end{aligned}$$

with equality iff $N = k$ or $k + 1$. This proves the theorem. QED.

Thus, for $m = 1$, the procedure III is generally better than the procedure I both in terms of ASN and the variance of the estimator of N. However, $V(\hat{N}_{III})$ is generally larger than $V(\hat{N}_{II})$, although ASN(III) is smaller than ASN(II). In table 4.1 given below we try to compare numerically the procedures II and III for some selected values of k in terms of $ASN \times V(\hat{N})$ and $(ASN + k) \times V(\hat{N})$ which are the inverses of efficiency per unit sample.

Table 4.1

$ASN \times V(\hat{N})$ and $(ASN + k) \times V(\hat{N})$ for Procedures II and III for $m = 1$

		$ASN \times Variance$		$(ASN + k) \times Variance$	
k	N	Procedure II	Procedure III	Procedure II	Procedure III
1	10	181.5000	168.6520	214.5000	232.0498
	20	1396.5000	1262.8647	1529.5000	1556.9930
	30	4644.8333	3904.1358	4944.5000	4607.6520
	50	21241.5000	15635.7260	22074.5000	17708.5700
	100	168316.5000	98412.2960	171649.5000	107191.3000
	200	1339966.5000	598876.2900	1353299.5000	635396.6000
2	10	161.3333	118.2123	249.3333	239.0964
	20	1323.0000	1019.2966	1701.0000	1598.5140
	30	4484.6667	3308.0282	5352.6667	4703.5790
	50	20808.0000	13830.4260	23256.0000	17960.7200
	100	166616.3333	90510.4400	176514.3333	108045.8000
	200	1333233.0000	565110.2400	1373031.0000	638117.9000
5	10	72.0238	32.6816	268.4524	244.0216
	20	787.5000	526.6540	1912.5000	1823.6210
	30	2860.1190	2023.2710	5627.9762	5311.8380
	50	13933.9286	9689.4600	22130.3571	19733.6000
	100	115368.4524	71149.2000	149636.3095	114550.9000
	200	937880.3571	478028.7000	1077862.5000	659885.7000
10	10	0.0000	0.0000	0.0000	0.0000
	20	334.0909	157.0200	2084.0909	1985.2470
	30	1456.0606	898.7900	6622.7273	6409.4930
	50	7881.8182	5555.6500	24881.8182	24053.9900
	100	69552.2727	49186.5400	145302.2727	133341.8000
	200	581529.5455	369313.1000	899779.5455	728764.4000

The computational study indicates that for N considerably large the procedure III is better than procedure II in terms of $ASN \times V(\hat{N})$, but for smaller values of N the procedure II is better. However, $(ASN + k) \times V(\hat{N})$ is generally smaller for procedure II. We now consider the case of general m $(\leq k)$ and extend partially Theorem 4.1.

Theorem 4.2 : For $m \leq k$
(a) $ASN(III) \leq ASN(II) \leq ASN(I)$,
(b) $V(\hat{N}_{III}) \leq V(\hat{N}_I)$.

Proof. **(a)** For procedure III,

$$E(S_m) = E(S_1) + E(S_2 - S_1) + \cdots + E(S_m - S_{m-1}).$$

Now for $j = 2, \cdots, m$, we have by (2.1)

$$\begin{aligned}
& P[S_j^* - S_{j-1}^* = s_j^* - s_{j-1}^* | S_{j-1}^* = s_{j-1}^*] \\
= \;& \frac{P[S_{j-1}^* = s_{j-1}^*, S_j^* = s_j^*]}{P[S_{j-1}^* = s_{j-1}^*]} \\
= \;& \frac{\sum \cdots \sum_{0 \leq s_1^* \leq \cdots \leq s_{j-2}^* \leq s_{j-1}^*} P[S_1^* = s_1^*, \cdots, S_j^* = s_j^*]}{\sum \cdots \sum_{0 \leq s_1^* \leq \cdots \leq s_{j-2}^* \leq s_{j-1}^*} P[S_1^* = s_1^*, \cdots, S_{j-1}^* = s_{j-1}^*]} \\
= \;& (1 - \frac{s_{j-1}^* + k}{N}) \cdots (1 - \frac{s_j^* - s_{j-1}^* + s_{j-1}^* + k - 1}{N})(s_j^* - s_{j-1}^* + s_{j-1}^* + k) \\
0 \leq \;& s_j^* - s_{j-1}^* \leq N - k - s_{j-1}^*
\end{aligned}$$

which is of the form (3.1) with k replaced by $s_{j-1}^* + k$. This means that, for $j = 2, \cdots, m$, the conditional probability distribution of $S_j^* - S_{j-1}^*$ given S_{j-1}^* is same as the unconditional probability distribution of S_1^* with k replaced by $S_{j-1}^* + k$. Hence by part (a) of theorem 4.1,

$$\begin{aligned}
E(S_j - S_{j-1}) &= EE(S_j - S_{j-1} | S_{j-1}) \\
&= EE(S_j^* - S_{j-1}^* + 1 | S_{j-1}^*) \\
&\leq E(\frac{N+1}{k + S_{j-1}^* + 1}) \\
&\leq \frac{N+1}{k+1}
\end{aligned}$$

for every $j = 1, \cdots, m$ with $S_0 = 0$. This implies that $ASN(III) \leq \frac{(N+1)m}{k+1} = ASN(II)$. Also, since by simple comparison $ASN(II) \leq ASN(I)$, the part (a) of the theorem follows. QED.

Proof. **(b)** We first prove that $V(\hat{N}^*_{III}) \le V(\hat{N}_I)$, where $\hat{N}^*_{III}$ is an unbiased estimator of N defined as $\hat{N}^*_{III} = \frac{1}{m}\sum_{j=1}^{m}\hat{N}(j)$, where, for $j = 1, \cdots, m$, $\hat{N}(j) = \sum_{S^*_{j-1}+k}^{S^*_j+k} i$ with $S_0 = 0$. Since, as shown in the proof of part (a), the conditional probability distribution of $S^*_j - S^*_{j-1}$ given S^*_{j-1} is of the form (3.1) with k replaced by $S^*_{j-1} + k$, as for m = 1, $\hat{N}(j)$ is conditionally unbiased and, hence, an unconditionally unbiased estimator of N for each j which implies the unbiasedness of $\hat{N}^*_{III}$. It follows by similar argument that, for $j' < j$, the conditional probability distribution of $S^*_j - S^*_{j-1}$ given $S^*_{j'-1}, \cdots, S^*_{j-1}$ is same as the conditional probability distribution of $S^*_j - S^*_{j-1}$ given S^*_{j-1}. Hence, for $j' < j$,

$$\begin{aligned} E(\hat{N}(j')\hat{N}(j)) &= E(\hat{N}(j'))E(\hat{N}(j)|S^*_{j'-1}, \cdots, S^*_{j-1}) \\ &= E(\hat{N}(j'))E(\hat{N}(j)|S^*_{j-1}) \\ &= N\,E(\hat{N}(j')) \\ &= N^2 \end{aligned}$$

Thus $\hat{N}(j), j = 1, \cdots, m$ are uncorrelated so that

$$V(\hat{N}^*_{III}) = \frac{1}{m^2}\sum_{j=1}^{m} V(\hat{N}(j)),$$

where for $j = 1, \cdots, m$,

$$\begin{aligned} V(\hat{N}(j)) &= EV(\hat{N}(j)|S^*_{j-1}) \\ &= N^2 - N\,EE(S^*_j + k|S^*_{j-1}) \\ &= N^2 - NE(S^*_j + k) \end{aligned}$$

$V(\hat{N}(j)|S^*_{j-1})$ being given by Theorem 3.1 with S^*_1 replaced by $S^*_j - S^*_{j-1}$ and k replaced by $S^*_{j-1} + k$.

Now $V(\hat{N}(j)) \le N^2 - NE(S^*_1 + k) = V(\hat{N}(1))$ for every $j = 1, \cdots, m$, and consequently, by part (b) of Theorem 4.1, $V(\hat{N}^*_{III}) \le \frac{V(\hat{N}(1))}{m} \le \frac{N(N-k)}{m} = V(\hat{N}_I)$. Since $\hat{N}_{III}$ is the UMVUE of N for procedure III, $V(\hat{N}_{III}) \le V(\hat{N}^*_{III}) \le V(\hat{N}_I)$, and the part (b) of theorem follows. QED.

We also believe that $V(\hat{N}_{II}) \leq V(\hat{N}_{III})$, although a completely satisfactory proof of this claim, for $m > 1$, has so far eluded us. In table 4.2 we present some numerical computations in this regard for some selected values of m and k. In table 4.3 we also compute the values of $ASN \times V(\hat{N})$ and $(ASN+k) \times V(\hat{N})$ for procedures II and III. The computational study again indicates that in terms of $ASN \times V(\hat{N})$ the procedure III is better than procedure II when N is considerably large. However, in terms of $(ASN+k) \times V(\hat{N})$ procedure II is generally better.

Table 4.2

$V(\hat{N})$ and $ASN \times V(\hat{N})$ for Procedures II and III for $m > 1$

			$Variance$		$ASN \times Variance$	
m	k	N	Procedure II	Procedure III	Procedure II	Procedure III
2	3	30	167.4000	323.6262	2594.7000	2786.8200
		35	230.4000	453.4350	4147.2000	4269.7000
		40	303.4000	605.7967	6219.7000	6159.6600
		45	386.4000	780.8748	8887.2000	8492.2370
		50	479.4000	978.8585	12224.7000	11301.5200
	5	20	90.0000	121.7214	630.0000	673.1599
		25	148.5714	205.5523	1287.6190	1321.7240
		30	221.4286	311.6173	2288.0950	2260.9680
		35	308.5714	440.1366	3702.8570	3531.1440
		40	410.0000	591.2849	5603.3330	5169.4930
3	7	30	132.0370	188.2415	1534.9310	1633.0330
		35	186.6667	269.7184	2520.0000	2594.6260
		40	250.5556	366.0504	3852.2920	3846.2890
		45	323.7037	477.3474	5583.8890	5416.8470
		50	406.1111	603.7205	7766.8750	7333.7120
4	8	45	212.7500	340.7283	4349.5560	4578.4400
		50	267.7500	432.6468	6069.0000	6218.9910
		55	329.0000	534.4419	8188.4440	8160.7680
		60	396.5000	645.5012	10749.5600	10412.5200

Table 4.3

$ASN \times V(\hat{N})$ and $(ASN + k) \times V(\hat{N})$ for Procedures II and III for $m > 1$

			$ASN \times Variance$		$(ASN + k) \times Variance$	
m	k	N	Procedure II	Procedure III	Procedure II	Procedure III
2	3	30	2594.7000	2786.8200	3096.9000	3757.6986
		35	4147.2000	4269.7000	4838.4000	5630.0050
		40	6219.7000	6159.6600	7129.9000	7977.0501
		45	8887.2000	8492.2370	10046.4000	10834.8614
		50	12224.7000	11301.5200	13662.9000	14238.0955
	5	20	630.0000	673.1599	1080.0000	1281.7669
		25	1287.6190	1321.7240	2030.4760	2349.4855
		30	2288.0950	2260.9680	3395.2380	3819.0545
		35	3702.8570	3531.1440	5245.7140	5731.8270
		40	5603.3330	5169.4930	7653.3330	8125.9175
3	7	30	1534.9310	1633.0330	2459.1900	2950.7235
		35	2520.0000	2594.6260	3826.6669	4482.6548
		40	3852.2920	3846.2890	5606.1812	6408.6418
		45	5583.8890	5416.8470	7849.8149	8758.2788
		50	7766.8750	7333.7120	10609.6527	11559.7555
4	8	45	4349.5560	4578.4400	6051.5560	7304.2664
		50	6069.0000	6218.9910	8211.0000	9680.1654
		55	8188.4440	8160.7680	10820.4440	12436.3032
		60	10749.5600	10412.5200	13921.5600	15576.5296

5. MAXIMUM LIKELIHOOD ESTIMATION OF N UNDER CMRR SAMPLING

In this section we consider the problem of estimation of the population size (N) under CMRR sampling scheme using the maximum likelihood method of estimation. Note that for $S_m^* = s_m^*$, the likelihood function is defined as

$$\begin{aligned} L(N|\, s_m^*) &= P_N(S_m^* = s_m^*) \\ &\propto \frac{(N-k)!}{(N-k-s_m^*)!}\frac{1}{N^{s_m^*+m}}, \\ N &\geq s_m^* + k \end{aligned} \tag{5.1}$$

For $s_m^* = 0$, L(N) is strictly decreasing in N implying that the maximum likelihood estimate (m.l.e.) $\hat{N}_{MLE}$ is k. Consider now $s_m^* \geq 1$ and let

$$A(N|\, s_m^*) = A(N) = \frac{L(N)}{L(N-1)} = \frac{N-k}{N-k-s_m^*}(\frac{N-1}{N})^{s_m^*+m} \tag{5.2}$$

Let us first assume that $A(N)$ is a continuous function of N, $k + s_m^* < N < \infty$. Then

$$\begin{aligned} \frac{d\, log\, A(N)}{d\, N} &= \frac{1}{N-k} - \frac{1}{N-k-s_m^*} + (s_m^*+m)(\frac{1}{N-1} - \frac{1}{N}) \\ &= -\frac{s_m^*}{(N-k)(N-k-s_m^*)} + \frac{s_m^*+m}{N(N-1)} \end{aligned}$$

so that $\frac{d\, log\, A(N)}{d\, N} > 0 \Leftrightarrow B(N) < \frac{s_m^*}{s_m^*+m}$ where

$$B(N) = (1 - \frac{k}{N})(1 - \frac{k+s_m^*-1}{N-1})$$

Note that $B(N)$ is decreasing in N and tends to 0 as $N \to k + s_m^*$ and tends to 1 as $N \to \infty$. Thus there exists a single finite root, say N_1, of $B(N) = \frac{s_m^*}{s_m^*+m}$. Thus $log A(N)$ and hence $A(N)$ is decreasing in N for $N < N_1$ and is increasing in N for $N > N_1$. Also $A(N) \to \infty$ as $N \to k + s_m^*$ and $A(N) \to 1$ as $N \to \infty$. This implies that $A(N) = 1$ will have a single finite root, say N_0, greater than $k + s_m^*$ and

$$\begin{aligned} A(N) &> 1 \; for \; N < N_0 \\ &= 1 \; for \; N = N_0 \\ &< 1 \; for \; N > N_0 \end{aligned} \tag{5.3}$$

Now, for integral values of N, (5.3) implies that

$$\begin{aligned} A(N) &\geq 1 \; for \; N \leq [N_0] \\ &< 1 \; for \; N > [N_0] \end{aligned} \tag{5.4}$$

where $[N_0]$ is the largest integer contained in N_0. (5.4) means that $L(N)$ is non-decreasing for $N \leq [N_0]$ and is decreasing for $N > [N_0]$, so that $[N_0]$ is the m.l.e. of N. In fact $[N_0]$ is the unique m.l.e. if N_0 is not an integer and in case N_0 is an integer, both N_0 and $N_0 - 1$ are m.l.e.'s of N.

Thus m.l.e. $\hat{N}_{MLE}$ of N is given by

$$\begin{aligned} \hat{N}_{MLE} &= k \; for \; s_m^* = 0 \\ &= [N_0] \; for \; s_m^* \geq 1 \end{aligned} \tag{5.5}$$

where $[N_0]$ is the largest integer contained in N_0 which is the unique finite root of

$$(N-k)(N-1)^{s_m^*+m} = (N-k-s_m^*)N^{s_m^*+m} \tag{5.6}$$

greater than $k + s_m^*$.

We now study the nature of bias of $\hat{N}_{MLE}$ for $m = 1$. Note that for $s_1^* = 0$, $\hat{N}_{III} = \binom{k+1}{2} - \binom{k}{2} = k = \hat{N}_{MLE}$. In the following theorem we prove that, for $s_1^* \geq 1, \hat{N}_{MLE} < \hat{N}_{III}$. This implies that $\hat{N}_{MLE}$ is always negatively biased, i.e. $\hat{N}_{MLE}$ always underestimates N.

Theorem 5.1: For $m = 1$ and $s_1^* \geq 1$, $\hat{N}_{III} > \hat{N}_{MLE}$.

Proof: Let $m = 1$ and $s_1^* \geq 1$ and A(N) be defined by (5.2). In view of the derivation of $\hat{N}_{MLE}$ it is sufficient to show that

$$A(N)|_{N=\hat{N}_{III}} < 1 \tag{5.7}$$

Now, $\hat{N}_{III} = \binom{s_1^*+k+1}{2} - \binom{k}{2} = \frac{(s_1^*+1)(s_1^*+2k)}{2}$.Hence, using the inequality $(1-x)^n < \frac{1}{1+nx+\binom{n+1}{2}}$ for $0 < x < 1$.

$$\begin{aligned} A(N)|_{N=\hat{N}_{III}} &= \frac{\hat{N}_{III}-k}{\hat{N}_{III}-k-s_1^*}(\frac{\hat{N}_{III}-1}{\hat{N}_{III}})^{s_1^*+1} \\ &= \frac{s_1^*+2k+1}{s_1^*+2k-1}[1-\frac{2}{(s_1^*+1)(s_1^*+2k)}]^{s_1^*+1} \\ &< \frac{\frac{s_1^*+2k+1}{s_1^*+2k-1}}{1+\frac{2}{s_1^*+2k}+\frac{2(s_1^*+2)}{(s_1^*+1)(s_1^*+2k)^2}} \\ &= \frac{u}{u+4(k-1)} \\ &\leq 1 \end{aligned}$$

where $u = (s_1^*+1)(s_1^*+2k)^2(s_1^*+2k+1)$.
This proves (5.7) and the theorem. QED.

In Table 5.1 we present the numerical values of the relative absolute bias of $\hat{N}_{MLE}$, viz. $\frac{N-E_N(\hat{N}_{MLE})}{N} \times 100$ for $m = 1$ and for different values of N and k. It is observed that the relative absolute bias, for fixed k, is first increasing and then decreasing in N and is small for large values of N.

Table 5.1

Relative bias of $\hat{N}_{MLE}$

	N					
k	10	20	30	50	100	200
1	14.6484	10.2675	8.2897	6.3266	4.3944	3.0642
2	11.5616	8.7875	7.3061	5.7362	4.1258	2.9656
3	9.3520	8.1639	7.0612	5.6775	4.0980	2.9191
5	5.6120	6.0753	5.6517	4.8784	3.7577	2.7851
7	3.0600	4.5981	4.6347	4.2549	3.4433	2.6375
10	0.0000	2.9996	3.3856	3.3836	2.9596	2.3904

We also compare empirically $\hat{N}_{MLE}$ with the UMVUE $\hat{N}_{III}$ for $m = 1$ and different values of k. Table 5.2 shows the values of $V(\hat{N}_{III}), MSE(\hat{N}_{MLE})$ and the relative gain in efficiency (RGE) of $\hat{N}_{MLE}$ over $\hat{N}_{III}$ defined as

$$RGE = \frac{V(\hat{N}_{III}) - MSE(\hat{N}_{MLE})}{MSE(\hat{N}_{MLE})} \times 100 \tag{5.6}$$

for different values of N. It is clear that $\hat{N}_{MLE}$ always performs better than $\hat{N}_{III}$. However, RGE decreases as N increases and the gain is not very significant when N is very large.

Table 5.2

Comparison of $\hat{N}_{III}$ and $\hat{N}_{MLE}$ for m=1 and different values of k

k	N	$V(\hat{N}_{III})$	$MSE(\hat{N}_{MLE})$	RGE
1	10	63.3978	55.6940	13.8325
1	20	294.1283	269.0078	9.3382
1	30	703.5162	653.8655	7.5934
1	50	2072.8440	1957.6180	5.8860
1	100	8779.0040	8427.3370	4.1729
1	200	36520.3100	35471.8700	2.9557
2	10	60.4421	52.0142	16.2030
2	20	289.6087	263.6353	9.8521
2	30	697.7754	647.3004	7.7978
2	50	2065.1470	1948.2990	5.9974
2	100	8767.6800	8401.5670	4.3577
2	200	36503.8300	35410.3600	3.0880
3	10	55.5526	46.5086	19.4459
3	20	281.7875	252.2592	11.7055
3	30	687.6165	632.3671	8.7369
3	50	2051.1940	1929.8680	6.2868
3	100	8746.6130	8389.9380	4.2512
3	200	36472.5600	35422.1700	2.9653
5	10	42.2680	35.7100	18.3646
5	20	259.3934	232.5601	11.5382
5	30	657.7134	603.4845	8.9860
5	50	2008.8280	1882.9760	6.6837
5	100	8680.3400	8308.2980	4.4780
5	200	36371.4000	35277.5000	3.1009

Table 5.2 (Contd.)

Comparison of $\hat{N}_{III}$ and $\hat{N}_{MLE}$ for m=1 and different values of k

k	N	$V(\hat{N}_{III})$	$MSE(\hat{N}_{MLE})$	RGE
7	10	26.3400	22.1700	18.8092
7	20	231.2255	208.0996	11.1129
7	30	619.0701	568.0366	8.9842
7	50	1952.3080	1830.2730	6.6676
7	100	8588.5110	8216.6250	4.5260
7	200	36226.8100	35122.4100	3.1444
10	20	182.8227	166.4832	9.8145
10	30	551.0703	510.1518	8.0208
10	50	1849.8340	1740.7110	6.2689
10	100	8415.5260	8056.0850	4.4617
10	200	35945.1300	34843.9200	3.1604

For general m, $\hat{N}_{III} = \frac{k^m}{k^{m-1}} = k = \hat{N}_{MLE}$ for $s_m^* = 0$. It has also been observed empirically that $\hat{N}_{III} > \hat{N}_{MLE}$ for $s_m^* \geq 1$, although we have not been able to prove it analytically. This means that $\hat{N}_{MLE}$ is always negatively biased. In Table 5.3 we present, for different values of m and k, the values of relative absolute bias (RAB) of $\hat{N}_{MLE}$ viz. $\frac{N-E_N(\hat{N}_{MLE})}{N} \times 100$ and the relative gain in efficiency (RGE) of $\hat{N}_{MLE}$ over $\hat{N}_{III}$ viz. $\frac{V(\hat{N}_{III})-MSE(\hat{N}_{MLE})}{MSE(\hat{N}_{MLE})} \times 100$ for different values of N. The table clearly indicates that in most situations, particularly for large N, $\hat{N}_{MLE}$ performs better than $\hat{N}_{III}$ in terms of MSE. Also, for large N, The gain decreases with N.

Table 5.3
Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
2	1	10	11.9202	8.5894
2	1	20	7.8766	3.8932
2	1	30	6.0719	3.9130
2	1	50	4.5843	4.0864
2	1	100	3.2177	3.0047
2	1	200	2.2580	2.1569
2	2	10	10.4479	4.1444
2	2	20	7.4405	7.3203
2	2	30	6.2103	5.8649
2	2	50	4.7521	3.9299
2	2	100	3.2676	2.9294
2	2	200	2.2754	2.1367
2	3	10	9.5094	12.5679
2	3	20	7.0106	6.5976
2	3	30	5.8676	5.9572
2	3	50	4.5892	4.1373
2	3	100	3.1897	2.9555
2	3	200	2.2531	2.2157
2	4	10	8.1283	10.8001
2	4	20	6.5999	7.7828
2	4	30	5.6005	5.6620
2	4	50	4.3682	4.1015
2	4	100	3.1145	3.1057
2	4	200	2.2237	2.2018
2	5	10	5.7500	22.2257
2	5	20	5.5706	7.2697
2	5	30	4.9243	6.4165
2	5	50	4.0871	4.6643
2	5	100	3.0185	3.1836
2	5	200	2.1835	2.2400

Table 5.3 (Contd.)

Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
2	6	10	5.5600	17.1126
2	6	20	5.5458	9.2789
2	6	30	5.0161	6.3518
2	6	50	4.0877	4.4630
2	6	100	3.0143	3.1982
2	6	200	2.1792	2.2167
2	7	10	2.5800	22.7260
2	7	20	4.6375	9.4670
2	7	30	4.3910	6.6243
2	7	50	3.7703	4.8423
2	7	100	2.9079	3.3638
2	7	200	2.1483	2.2769
2	8	10	2.9000	30.4357
2	8	20	4.3838	8.5718
2	8	30	4.2166	6.7364
2	8	50	3.6538	4.6797
2	8	100	2.8037	3.2380
2	8	200	2.0809	2.2612
2	9	10	0.5000	10.7314
2	9	20	3.5668	10.5913
2	9	30	3.8164	7.6331
2	9	50	3.5126	5.0411
2	9	100	2.7585	3.2654
2	9	200	2.0575	2.2506
2	10	20	3.5137	8.1220
2	10	30	3.5527	6.5383
2	10	50	3.2536	4.8630
2	10	100	2.6408	3.3779
2	10	200	2.0167	2.3185

Table 5.3 (Contd.)

Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
3	1	10	7.0783	5.6902
3	1	20	5.2621	4.8374
3	1	30	4.3860	4.1620
3	1	50	3.4657	3.3625
3	1	100	2.4984	2.4607
3	1	200	1.7902	1.7758
3	2	10	11.4917	-0.4353
3	2	20	7.3599	2.9760
3	2	30	5.7410	3.2454
3	2	50	4.2497	2.9785
3	2	100	2.8747	2.3381
3	2	200	1.9725	1.7353
3	3	10	9.8924	1.9766
3	3	20	6.4439	3.1594
3	3	30	5.0650	3.2583
3	3	50	3.7977	2.9534
3	3	100	2.6219	2.3195
3	3	200	1.8355	1.7264
3	4	10	8.6550	3.9175
3	4	20	6.2588	5.7591
3	4	30	5.2439	4.5221
3	4	50	4.0122	3.0943
3	4	100	2.7284	2.2428
3	4	200	1.8776	1.6858
3	5	10	7.3176	11.6371
3	5	20	5.7933	5.8952
3	5	30	4.8693	4.2195
3	5	50	3.7164	2.8873
3	5	100	2.5377	2.1624
3	5	200	1.7652	1.6540

Table 5.3 (Contd.)
Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
3	6	10	6.1856	15.7751
3	6	20	4.8230	3.0914
3	6	30	3.9753	4.3122
3	6	50	3.4211	4.5255
3	6	100	2.6215	2.7624
3	6	200	1.9033	1.8869
3	7	10	5.4990	11.0849
3	7	20	5.2834	6.0630
3	7	30	4.3981	3.3835
3	7	50	3.3198	2.8432
3	7	100	2.4931	2.9497
3	7	200	1.8528	1.9081
3	8	10	2.4380	31.7297
3	8	20	5.1161	6.8766
3	8	30	4.2508	3.0118
3	8	50	3.2143	3.0017
3	8	100	2.4532	2.9266
3	8	200	1.8408	1.9347
3	9	10	1.8700	48.3749
3	9	20	3.9826	5.5802
3	9	30	3.7371	5.8535
3	9	50	3.2834	4.0509
3	9	100	2.4569	2.4069
3	9	200	1.7427	1.7532
3	10	20	3.9937	6.9152
3	10	30	3.6358	4.0967
3	10	50	3.0013	3.7155
3	10	100	2.3899	2.8854
3	10	200	1.7612	1.7191

Table 5.3 (Contd.)

Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
4	1	10	8.7429	6.4107
4	1	20	6.0361	2.4223
4	1	30	4.8269	2.9301
4	1	50	3.4457	1.4689
4	1	100	2.3278	2.3094
4	1	200	1.6667	1.4999
4	2	10	7.3281	7.7856
4	2	20	5.7821	4.6871
4	2	30	4.7138	3.1877
4	2	50	3.6217	2.8239
4	2	100	2.4169	1.7572
4	2	200	1.6737	1.5406
4	3	10	6.7860	0.5653
4	3	20	5.0648	2.7323
4	3	30	4.0703	3.2320
4	3	50	3.3112	3.6594
4	3	100	2.4406	2.0835
4	3	200	1.6619	1.4679
4	4	10	8.9819	5.5832
4	4	20	5.5921	1.0257
4	4	30	4.4431	3.3768
4	4	50	3.5108	2.9604
4	4	100	2.3437	1.7198
4	4	200	1.6530	1.6485
4	5	10	8.9180	-2.4339
4	5	20	6.2438	2.7258
4	5	30	4.9372	2.2171
4	5	50	3.4432	1.2631
4	5	100	2.2534	2.0779
4	5	200	1.6376	1.6067

Table 5.3 (Contd.)

Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
4	6	10	5.4944	35.0230
4	6	20	6.1626	3.7405
4	6	30	4.9618	2.7445
4	6	50	3.5163	1.1145
4	6	100	2.2058	1.8607
4	6	200	1.6154	1.6967
4	7	10	5.5062	33.8111
4	7	20	5.2827	2.4771
4	7	30	4.2382	2.8932
4	7	50	3.3027	3.0203
4	7	100	2.3594	1.8875
4	7	200	1.6039	1.4935
4	8	10	3.7910	25.5932
4	8	20	4.2843	3.4896
4	8	30	3.5280	2.9109
4	8	50	2.8132	2.9264
4	8	100	2.1519	2.4804
4	8	200	1.6086	1.6487
4	9	10	3.1220	90.7822
4	9	20	4.6205	2.6347
4	9	30	3.7801	3.3230
4	9	50	3.1366	3.5038
4	9	100	2.3022	2.0930
4	9	200	1.6228	1.4671
4	10	20	4.1010	10.4316
4	10	30	3.7784	3.3900
4	10	50	3.0242	3.2158
4	10	100	2.2539	2.1474
4	10	200	1.6119	1.5284

Table 5.3 (Contd.)

Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
5	1	10	9.2651	-5.0477
5	1	20	6.2878	0.4671
5	1	30	4.4262	-0.6820
5	1	50	3.1146	2.3244
5	1	100	2.1808	1.7484
5	1	200	1.5161	1.3453
5	2	10	8.2185	-3.9660
5	2	20	5.6517	0.7481
5	2	30	4.0252	0.2100
5	2	50	3.0387	2.9170
5	2	100	2.1489	1.7012
5	2	200	1.4956	1.3367
5	3	10	6.2358	3.2164
5	3	20	5.6197	2.9037
5	3	30	4.5044	1.9677
5	3	50	3.1365	1.3182
5	3	100	2.1489	1.7682
5	3	200	1.4872	1.3141
5	4	10	8.1207	-0.7731
5	4	20	4.2806	1.4183
5	4	30	3.7186	4.0684
5	4	50	3.1016	3.1526
5	4	100	2.1774	1.6148
5	4	200	1.5089	1.4042
5	5	10	6.4607	0.1009
5	5	20	4.9727	4.8216
5	5	30	4.2238	2.6158
5	5	50	3.0486	1.5240
5	5	100	2.0960	1.9527
5	5	200	1.4998	1.4069

Table 5.3 (Contd.)

Relative absolute bias of $\hat{N}_{MLE}$ and relative gain in efficiency of $\hat{N}_{MLE}$ over $\hat{N}_{III}$

m	k	N	RAB	RGE
5	6	10	8.9299	-1.6579
5	6	20	5.6312	-2.4758
5	6	30	3.8943	0.0885
5	6	50	2.8575	2.5509
5	6	100	2.1161	1.8802
5	6	200	1.4846	1.4032
5	7	10	5.9767	-7.8737
5	7	20	4.7296	2.4186
5	7	30	3.4588	0.5150
5	7	50	2.7090	3.2672
5	7	100	2.1200	2.0179
5	7	200	1.4792	1.3461
5	8	10	2.5684	32.6204
5	8	20	4.4025	3.3208
5	8	30	3.6979	2.7683
5	8	50	2.7460	1.7129
5	8	100	1.9924	2.1014
5	8	200	1.4391	1.4335
5	9	10	1.8098	44.1943
5	9	20	5.0271	2.0237
5	9	30	4.0934	2.6772
5	9	50	3.0678	1.5821
5	9	100	2.0187	1.6648
5	9	200	1.4611	1.4378
5	10	20	3.5757	6.0916
5	10	30	3.1793	3.0312
5	10	50	2.6558	3.0069
5	10	100	2.0250	2.0607
5	10	200	1.4477	1.3639

APPENDIX A

Proof of the identity

$$g_c(a,b) = (b-a)!\sum\cdots\sum_{a\le i_1\le\cdots\le i_c\le b} i_1\cdots i_c = \sum_{y=0}^{b-a}(-1)^y\binom{b-a}{y}(b-y)^{b-a+c} \tag{A.1}$$

We prove (A.1) by induction on b for fixed a. Note that for $b=a$, LHS of (A.1) $= a^c =$ RHS of (A.1), so that (A.1) is true for $b=a$ whatever be c. Assume that (A.1) is true for $b=a+t$ whatever be c. Then for $b=a+t+1$,

$$\begin{aligned}
& LHS\ of\ (A.1)\\
= \ & (t+1)!\sum\cdots\sum_{a\le i_1\le\cdots\le i_c\le a+t+1} i_1\cdots i_c\\
= \ & (t+1)![\sum\cdots\sum_{a\le i_1\le\cdots\le i_c\le a+t} i_1\cdots i_c\\
& +(a+t+1)\sum\cdots\sum_{a\le i_1\le\cdots\le i_c\le a+t} i_1\cdots i_{c-1}+\cdots+(a+t+1)^c]\\
= \ & (t+1)\sum_{x=0}^{c}(a+t+1)^x g_{c-x}(a,a+t)\\
= \ & (t+1)\sum_{x=0}^{c}(a+t+1)^x\sum_{y=0}^{t}(-1)^y\binom{t}{y}(a+t-y)^{t+c-x}\\
= \ & (t+1)\sum_{y=0}^{t}(-1)^y\binom{t}{y}(a+t-y)^t\sum_{x=0}^{c}(a+t+1)^x(a+t-y)^{c-x}\\
= \ & (t+1)\sum_{y=0}^{t}(-1)^y\binom{t}{y}\frac{(a+t-y)^t}{y+1}[(a+t+1)^{c+1}-(a+t-y)^{c+1}]\\
= \ & \sum_{y=0}^{t}(-1)^y\binom{t+1}{y+1}(a+t-y)^t[(a+t+1)^{c+1}-(a+t-y)^{c+1}]\\
= \ & \sum_{y=1}^{t+1}(-1)^{y-1}\binom{t+1}{y}(a+t+1-y)^t[(a+t+1)^{c+1}-(a+t+1-y)^{c+1}]\\
= \ & (a+t+1)^{c+1}\sum_{y=1}^{t+1}(-1)^{y-1}\binom{t+1}{y}(a+t+1-y)^t\\
& +\sum_{y=1}^{t+1}(-1)^y\binom{t+1}{y}(a+t+1-y)^{t+c+1}
\end{aligned}$$

$$
\begin{aligned}
&= (a+t+1)^{c+1} + \sum_{y=1}^{t+1}(-1)^y \binom{t+1}{y}(a+t+1-y)^{t+c+1} \\
&\quad -(a+t+1)^{c+1} \sum_{y=0}^{t+1}(-1)^y \binom{t+1}{y}(a+t+1-y)^{t} \\
&= \sum_{y=0}^{t+1}(-1)^y \binom{t+1}{y}(a+t+1-y)^{t+c+1} \\
&= RHS\ of\ (A.1)
\end{aligned}
$$

since $\sum_{y=0}^{t+1}(-1)^y\binom{t+1}{y}(a+t+1-y)^t = \Delta^{t+1}a^t = 0$. Thus (A.1) is true for $b = a+t+1$. Since (A.1) is true for $b = a$, it is true for $b \geq a$. QED.

APPENDIX B

Proof of Completeness of S_m^*

Let $f(S_m^*)$ be such that

$$\begin{aligned} 0 &= E_N[f(S_m^*)] \\ &= \sum_{s_m^*=0}^{N-k} f(s_m^*) P_N[S_m^* = s_m^*] \\ &= \sum_{s_m^*=0}^{N-k} f(s_m^*) \binom{N-k}{s_m^*} \frac{s_m^*+k}{N^{s_m^*+m}}\, g_{m-1}(k, s_m^*+k) \end{aligned}$$

$$\forall N \geq k \tag{$B.1$}$$

For $N = k$, (B.1) implies that

$$\begin{aligned} & f(0)\frac{k}{N^m}\, g_{m-1}(k,k) = 0 \\ \Leftrightarrow \quad & f(0) = 0 \end{aligned} \tag{B.2}$$

since $\frac{k}{N^m}\, g_{m-1}(k,k) \neq 0$.

For $N = k+1$, (B.1) and (B.2) imply that

$$\begin{aligned} & f(1)\frac{k+1}{N^{m+1}}\, g_{m-1}(k,k+1) = 0 \\ \Leftrightarrow \quad & f(1) = 0 \end{aligned} \tag{B.3}$$

since $\frac{k+1}{N^{m+1}}\, g_{m-1}(k,k+1) \neq 0$.

Similarly, putting successively $N = k, k+1, \cdots$ we get, from (B.1), $f(s_m^*) = 0,\ \forall s_m^* = 0, \cdots, N-k, \forall N \geq k$. This proves the completeness of S_m^*. QED.

CHAPTER 6

UNBIASED ESTIMATION OF THE SIZE OF A FINITE POPULATION USING CAPTURE-MARK-RELEASE-RECAPTURE SAMPLING SCHEME II

The problem considered is that of unbiased estimation of the size (N) of a finite closed population under *Capture-Mark-Release-Recapture*(CMRR) sequential sampling procedure. Borrowing ideas from the Bernoulli sequential estimation and using the notions of 'closed' and 'pushed-up' sampling plans, we provide here an unified approach to the problem of unbiased estimation of N and, in particular, give a necessary and sufficient condition for unbiased estimability under an arbitrary stopping rule. The ideas are illustrated with several examples.

1. INTRODUCTION

In Chapter 5 we have considered unbiased and maximum likelihood estimation of size (N) of a finite population under a sequential sampling procedure (termed as *Capture-Mark-Release-Recapture* or CMRR sampling) described as follows. Initially k population units are marked and released into the target population and then units are sampled at random, marked and released one by one until the sampling is stopped according to a given stopping rule. We have studied some properties of the minimum variance unbiased estimator of N and have studied the relative performances of the maximum likelihood and the minimum variance unbiased estimators in terms of their bias and mean square errors under a specific stopping rule viz. when a specific number of marked units are recaptured and have also compared the performance of the CMRR sampling scheme with some other procedures. Other stopping rules had been considered in the literature e.g. in Samuel (1968) and Berg (1987).

The purpose of the present Chapter is to study CMRR sampling in the standard framework of the notion of Bernoulli sequential sampling plans introduced in the literature for unbiased sequential estimation of powers of the Bernoulli parameter (p) under the set-up of independent and identical Bernoulli trials. Of particular importance and relevance here are the papers of Gupta (1967), Sinha and Sinha (1975) and Sinha and Bose (1985) wherein the problem of sequential estimation of p^{-1} had been discussed and a characterisation of sampling plans admitting unbiased estimators of p^{-1} had been provided in terms of the 'closure' of the corresponding 'pushed-up' plans. A review article in this direction is Sinha and Sinha (1992). Borrowing the same ideas, we provide here an unified approach to the problem of unbiased estimation of N under CMRR sampling and, in particular, give a necessary and sufficient condition for unbiased estimability under an arbitrary stopping rule. The ideas are also illustrated with several examples.

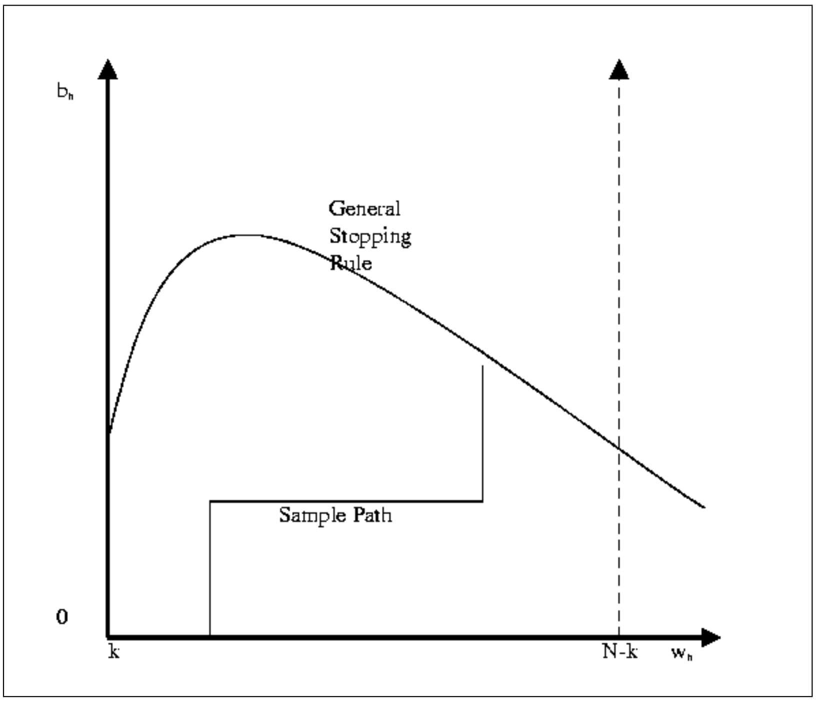
2. GENERALIZED BERNOULLI SEQUENTIAL SAMPLING

We recall the set-up of CMRR sampling for unbiased estimation of N, the (unknown) size of a finite closed population. In urn sampling terminology, we have initially an urn containing k black and $N-k$ white balls where k is known and N is unknown. Balls are drawn at random one at a time, each time examining the colour of the ball drawn. At any stage, whenever a white ball is drawn, its colour is changed to black and is replaced into the urn. However, if a black ball is drawn it is replaced into the urn without any change of colour. Thus after each draw, whatever be the colour of the ball drawn, one black ball is replaced into the urn unless the sampling is stopped. The experiment provides dependent Bernoulli trials, the nature of dependence being manifested in the conditional probabilities of drawing black balls at various draws. All these probabilities are of course functions of N. A general stopping rule relevant to this situation is described below.

Let w_n and b_n be the numbers of white and black balls, respectively, observed upto the n-th trial. Clearly, $0 \leq w_n \leq N-k$, $\quad b_n \geq 0$ and $w_n + b_n = n$. Note that these refer to the scenario before change of colour, if any, at the end of each draw.

We may plot $(w_n, b_n), n = 0, 1, 2, \cdots$ as a point with non-negative integral co-ordinates in the XY-plane with $(w_0, b_0) = (0, 0)$. A sampling plan is given by a stopping rule specified by a set B of 'boundary' points, a typical point being denoted by $\alpha = (x, y)$. The sampling is terminated when the successive outcomes of the trials lead to such a point. Other than boundary points, there are also points which are 'accessible', meaning thereby that these points can be generated through the outcomes of the experiment. Of course, there are also points which are 'inaccessible' in the sense that these points are not reachable in view of the existence of the boundary points. An 'accessible path' leading to a boundary point α is a succession of accessible points with α as the ultimate boundary point and with no other boundary point in between α and the origin $(0, 0)$.

Fig. 1.
Plotting the samples and the general boundary.



The probability that a certain accessible path will originate at the origin and will reach an accessible (or boundary) point $\alpha = (x, y)$ is

$$\frac{(N-k)(N-k-1)\cdots(N-k-x+1)}{N^{x+y}}\Pi_{i=0}^{x}(k+i)^{t_i}, \tag{1.1}$$

where t_i is the magnitude of jump (always upwards) at $(i, t_0 + t_1 + \cdots + t_{i-1})$, $t_0 + \cdots + t_x = y$, $0^0 = 1$.

Hence, the probability of reaching an accessible (or boundary) point $\alpha = (x, y)$ is $P_N(\alpha) = P_N(\alpha|B)$

$$= \frac{(N-k)(N-k-1)\cdots(N-k-x+1)}{N^{x+y}} \sum_{accessible\ paths} \Pi_{i=0}^{x}(k+i)^{t_i}$$

$$= \frac{(N-k)(N-k-1)\cdots(N-k-x+1)}{N^{x+y}} \phi(\alpha, B), \quad say \tag{1.2}$$

Without loss of generality, we can ignore accessible (or boundary) points α for which $P_N(\alpha) = 0 \;\; \forall N \geq k$. So from now α is accessible (or $\alpha \in B$) will mean $P_N(\alpha) > 0$ for at least one N. Clearly, B can have at most one boundary point on the Y-axis.

The sampling plan is said to be 'closed' iff $\sum_{\alpha \in B} P_N(\alpha) = 1 \;\; \forall N \geq k$, i.e. iff there does not exist any non-terminating sampling path which is accessible (having positive probability for some N). It is clear that only closed plans are of relevance for the purpose of inference on N and throughout the paper, every sampling plan is assumed to be closed.

It follows from (1.1) and (1.2) that the conditional probability of a certain accessible path upto a boundary point α given that α is reached is independent of N. This proves that boundary points are sufficient for the parameter space $\{N \geq k\}$. It follows, therefore, that for unbiased estimation of N, we can restrict to estimators $f(\alpha)$ defined only at the boundary points $\alpha \in B$. An unbiased estimator $f(\alpha)$ will also be the uniformly minimum variance unbiased estimator (UMVUE) if the set of boundary points is complete.

3. CHARACTERIZATION OF CLOSED SAMPLING PLANS

Let us consider an arbitrary sampling plan given by a set of boundary points B. We first provide a characterization of a closed plan given by the following theorem.

Theorem 3.1. A sampling plan given by a set of boundary points B is closed iff
(i)$\exists$ a point $(0, y_0) \in B$,
(ii) for any accessible point $(i, y_i^*) \notin B, \exists$ a point $(i+1, y_{i+1}) \in B \ni y_{i+1} \geq y_i^*$.

Proof. Fix $N \geq$k and note that for any accessible point $\alpha = (x, y), 0 \leq x \leq N - k$.

Suppose that B does not satisfy (i). Then, for $N = k$, $\sum_{\alpha \in B} P_K(\alpha) = \sum_{\alpha=(0,y)\in B} P_K(\alpha) = 0$ implying that the plan is not closed. This proves the necessity of (i).

Suppose next that B satisfies (i) but not (ii). If $B = \{(0,0)\}$, the condition (ii) is vacuous. Otherwise, for $N = k + i + 1$, the probability of the non-terminating sampling path $(i, y_i^*), (i+1, y_i^*), (i+1, y_i^*+1), (i+1, y_i^*+2), \cdots$ is $\frac{P_{k+i+1}(i, y_i^*)}{k+i+1}$ which is positive. So B is not closed implying the necessity of (ii).

Suppose now that B satisfies (i) and (ii). Consider first $N = k$. Then $\sum_{\alpha \in B} P_K(\alpha) = \sum_{\alpha=(0,y)\in B} P_K(\alpha) = P_k(0, y_0) = 1$.

Further, considering successively $N = k + 1, k + 2, \cdots$, it follows from the structure of B that, $\nexists$ any non-terminating sampling path with positive probability for any given N. Hence B is closed. QED.

Remark 3.1: Gupta (1967) and Sinha and Sinha (1975) examined necessary and sufficient conditions for closure of sequential sampling plans for estimation of Bernoulli parameters. The set-up here is, however, different from that of Bernoulli parameter estimation. Although conditions of Theorem 3.1 are not necessary for closure of a sequential sampling plan for estimation of a Bernoulli parameter, it is the discrete nature of N that enables us to deduce the above theorem.

Remark 3.2: It may be noted that a closed sampling plan has exactly one boundary point of the type $(0, y_0)$ but may have more than one boundary point of the type (x, y_x) for $x > 0$.

4. COMPLETENESS OF THE SET OF BOUNDARY POINTS

For a closed sampling plan with the set of boundary points B, the following theorem gives a necessary and sufficient condition for B to be complete.

Theorem 4.1. Consider a closed sampling plan with the set of boundary points B. Then B is complete iff $\alpha = (x, y), \alpha' = (x', y') \in B, x = x' \Rightarrow y = y'$, i.e. iff B contains at most one point on the line X=i $\forall i \geq 0$.

Proof. Let B satisfy the condition of the theorem. Consider an arbitrary function f on $B \ni E_N(f) = 0 \ \ \forall N \geq k$. Then

$\sum_{\alpha=(x,y_x)\in B} f(\alpha)P_N(\alpha) = 0 \forall N \geq k$, or

$$\sum_{x=0}^{N-k} f(x,y_x)\frac{(N-k)(N-k-1)\cdots(N-k-x+1)}{N^{x+y}}\phi(\alpha,B) = 0 \ \ \forall N \geq k.$$

Here we use (1.2) and the facts that B satisfies the condition of the theorem and that $\alpha = (x,y) \in B \Rightarrow 0 \leq x \leq N-k$.

Taking $N = k, k+1, \cdots$ successively, we get $f(\alpha) = 0 \ \ \forall \alpha \in B$ and this proves that B is complete.

Conversely, suppose $\exists (i,y_i'), (i,y_i'') \in B$ for some i ≥ 0 with $y_i' \neq y_i''$. Let a function f on B be defined as follows:

$$\begin{aligned} f(\alpha) &= 0, & \forall \alpha = (j,y) \in B, & \quad j < i \\ &= t(\neq 0), & \forall \alpha = (i,y) \in B, & \quad y \neq y_i' \\ &= t', & \alpha = (i,y_i') \in B, & \end{aligned}$$

where, t, t' are so determined as to have $E_N(f) = 0$ for $N = k+i$. We can then successively, for increasing values of j, construct f on $\alpha = (j,y) \in B$ with $j > i \ni E_N(f) = 0 \ \ \forall N \geq k$. This implies that B is not complete. Hence, the necessity of the condition follows. QED.

5. NECESSARY AND SUFFICIENT CONDITION FOR UNBIASED ESTIMATION OF N

We can now give a necessary and sufficient condition for a closed sampling plan to admit unbiased estimation of N. This is in terms of the closure property of the corresponding 'pushed-up' plan.

For a sampling plan with the set of boundary points B, the corresponding 'pushed-up' plan is defined to be one having the set of boundary points B_+ defined as $B_+ = \{\alpha_+ = (x, y+1) : \alpha = (x,y) \in B\}$.

The following lemmas are useful for the study of the existence of an unbiased estimator of N for an arbitrary closed plan.

Lemma 5.1. A sampling plan is closed if the corresponding pushed-up plan is closed.

Proof: Let B and B_+ denote, respectively, the set of boundary points for the original and corresponding pushed-up plans. Clearly, $\exists$ a point $(0, y_0) \in B_+$ implies that $\exists$ a point $(0, y_0 - 1) \in B$. Consider now a point $(i, y_i^*) \notin B$ which is accessible (w.r.t. B). Then $(i, y_i^* + 1) \notin B_+$ and is accessible (w.r.t. B_+). If B_+ satisfies the the condition (ii) of Theorem 3.1, then $\exists$ a point

$(i+1, y_{i+1}) \in B_+ \ni y_{i+1} \geq y_i^* + 1$. This in turn implies that $\exists$ a point $(i+1, y_{i+1}-1) \in B \ni y_{i+1} - 1 \geq y_i^*$. Thus B satisfies the conditions of Theorem 3.1 if so does B_+. Hence, by Theorem 3.1, the lemma follows.QED.

Lemma 5.2 Consider a closed sampling plan with the set of boundary points B. The corresponding pushed-up plan is closed iff B contains infinitely many points.

Proof: Let B_+ be the set of boundary points for the pushed-up plan and suppose B contains infinitely many points. Since the sampling plan is closed, B satisfies the conditions of Theorem 3.1. Hence, successively for $i = 0, 1, \cdots$, we can uniquely define $B_i = \{(i, y_i), (i, y_i - 1), \cdots, (i, y_i - s_i)\} \subset B$ with $s_0 = 0 \ni y_{i+1} \geq y_i - s_i - 1$, where $y_i - s_i - 1$ is the largest Y-coordinate of the accessible points $\notin B$ on the line X=i. We note that for each $i \geq 0, \exists$ at least one accessible point $\notin B$ on the line X= i, since, otherwise B is finite. Also it may be seen, by induction on i, that $y_i - s_i$ is the maximum Y-coordinate of the accessible (w.r.t. B_+) points $\notin B_+$ on the line X=i. Hence, $(0, y_0 + 1) \in B_+$ and for any $(i, y_i^*) \not\in B_+$ which is accessible (w.r.t.B_+), $\exists (i+1, y_{i+1}+1) \in B_+ \ni y_{i+1} + 1 \geq y_i - s_i \geq y_i^*$. This means B_+ satisfies the conditions of Theorem 3.1 and, hence, the corresponding pushed-up plan is closed.

Conversely, if B is finite, $\exists$ an $i \geq 0 \ni$ no point on the line X= i belongs to B and hence, to B_+. So the path $(0,0), \cdots, (i,0), (i,1), (i,2), \cdots$ is accessible w.r.t. B_+ and is non-terminating implying that B_+ is not closed. QED.

We now prove in the following theorem that a closed sampling plan admits unbiased estimation of N iff the corresponding pushed-up plan is closed.

Theorem 5.1 For a closed sampling plan, N is unbiasedly estimable iff the corresponding pushed-up plan is closed.

Proof: Let B and B_+ denote, respectively, the sets of boundary points for the original and the pushed-up plan.

Suppose first that the pushed-up plan is closed.
Then $\sum_{\alpha_+ \in B_+} P_N(\alpha_+|B_+) = 1 \;\; \forall N \geq k$, implying, by (1.2) and the definition of the pushed-up plan,

$$\sum_{\alpha \in B} \frac{\phi(\alpha_+, B_+)}{\phi(\alpha, B)} P_N(\alpha|B) = N \;\; \forall N \geq k.$$

Thus $\hat{N}(\alpha) = \frac{\phi(\alpha_+, B_+)}{\phi(\alpha, B)}$ is an unbiased estimator of N. This is also the UMVUE if B satisfies the condition of Theorem 4.1.

Suppose now that the pushed-up plan is not closed. If possible, let $f(\alpha)$ be an unbiased estimator of N, i.e.

$$\sum_{\alpha \in B} f(\alpha) P_N(\alpha) = N \;\; \forall N \geq k \tag{5.1}$$

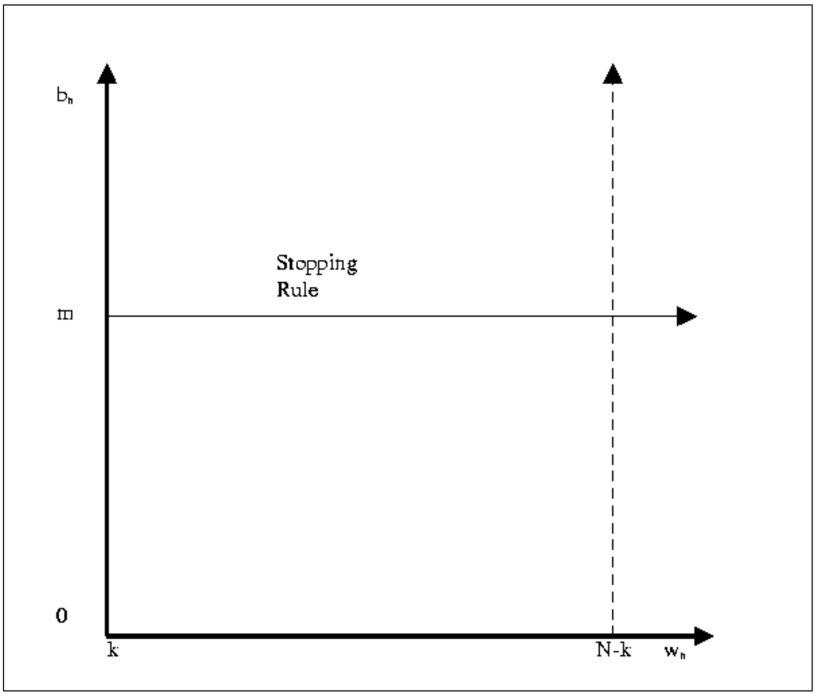
Now, since the pushed-up plan is not closed, B is finite by Lemma 5.2. Hence, the L.H.S. of (5.1) is bounded as $N \to \infty$. This shows that (5.1) is impossible, i.e. there can not exist an unbiased estimator of N. QED.

For a positive integer c, it follows by the same arguments as of Theorem 5.1, that, for a closed sampling plan, N^c is unbiasedly estimable if and only if N is unbiasedly estimable. Here we use the fact that for a closed plan, the closure of the pushed-up plan implies the closure of the c-step pushed-up plan by Lemma 5.2, where the c-step pushed-up plan is defined to be one having the set of boundary points $B_{+c} = \{\alpha_{+c} = (x, y+c) : \alpha = (x, y) \in B\}$. Hence, if the pushed-up plan is closed, $\hat{N^c}(\alpha) = \frac{\phi(\alpha_{+c}, B_{+c})}{\phi(\alpha, B)}$ is an unbiased estimator of N^c. In particular, for a sampling plan admitting unbiased estimator $\hat{N}$ of N, N^2 and hence, $V(\hat{N})$ are unbiasedly estimable and an unbiased estimator of $V(\hat{N}) = (\hat{N})^2 - \hat{N^2}$ where $\hat{N^2}(\alpha) = \frac{\phi(\alpha_{+2}, B_{+2})}{\phi(\alpha, B)}$. When B satisfies the conditions of Theorem 4.1, we end up with UMVUE.

6. ILLUSTRATIVE EXAMPLES

The ideas developed above can be illustrated by the following examples.
Example 6.1.

Fig. 2.
Example 6.1.



Consider a sampling plan given by the the set of boundary points B={(x,m),x=0,1,$\cdots$}, where m is a given positive integer. Clearly B and B_+ are closed and B satisfies the condition of Theorem 4.1. Also it may be seen that, for $\alpha = (x, m)$,

$$\phi(\alpha, B) = (k+x)\sum \cdots \sum_{k \le i_1 \le \cdots \le i_{m-1} \le k+x} i_1 \cdots i_{m-1}$$

$$= \frac{(k+x)}{x!} \Delta^x k^{x+m-1},$$

where $\Delta^a b^d = \Delta^a t^d|_{t=b}$ (see Chapter 5).

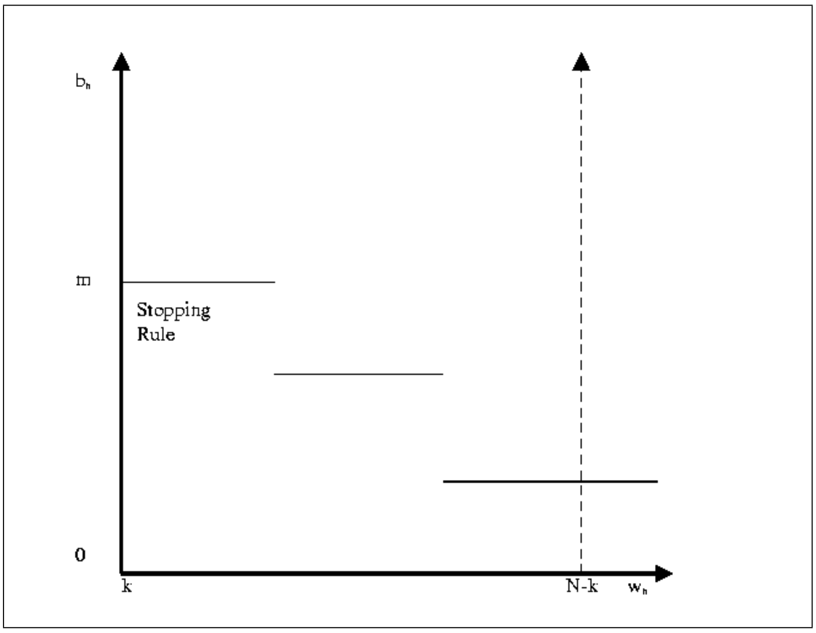
Similarly, for any positive integer c and $\alpha_{+c} = (x, m+c)$

$$\phi(\alpha_{+c}, B_{+c}) = \frac{(k+x)}{x!}\Delta^x k^{x+m+c-1}.$$

So that the UMVUE of N is $\hat{N} = \frac{\Delta^x k^{x+m}}{\Delta^x k^{x+m-1}}$ and the UMVUE of $V(\hat{N})$ is $\hat{V}(\hat{N}) = (\hat{N})^2 - \hat{N^2}$, where $\hat{N^2} = \frac{\Delta^x k^{x+m+1}}{\Delta^x k^{x+m-1}}$. Here x denotes the number of white balls observed.

Example 6.2.

Fig. 3.
Example 6.2.



Consider a sampling plan given by the set of boundary points $B = \{(0, m+d), (1, m+d), \cdots, (M_1, m+d), (M_1+1, m+d-1), \cdots, (M_2, m+d-1), \cdots, (M_{d-1}+1, m+1), \cdots, (M_d, m+1), (x, m), x \geq M_d + 1\}$ where m and d are given positive integers and $0 \leq M_1 < M_2 < \cdots < M_d$.

Clearly, B and B_+ are closed and B satisfies the condition of Theorem 4.1. Also, as in the previous example, it may be seen that for $\alpha = (x, y) \in B$,

$$
\begin{aligned}
\phi(\alpha, B) &= (k+x)\sum\cdots\sum\nolimits_{k\le i_1\le\cdots\le i_{m+d-1}\le k+x} i_1\cdots i_{m+d-1} \\
&\quad \text{for } x = 0, 1, \cdots, M_1 \\
&= (k+x)\sum\cdots\sum\nolimits_{k\le i_1\le\cdots\le i_{m+d-b-1}\le k+x} i_1\cdots i_{m+d-b-1} \\
&\quad \text{for } x = M_b+2, \cdots, M_{b+1}, b = 1, \cdots, d-1 \\
&= (k+x)\sum\cdots\sum\nolimits_{k\le i_1\le\cdots\le I_{m-1}\le k+x} i_1\cdots i_{m-1} \\
&\quad \text{for } x \ge M_d+2 \\
&= \sum\cdots\sum\nolimits_{k\le i_1\le\cdots\le i_{m+d-b}\le k+x} i_1\cdots 1_{m+d-b} \\
&\quad \text{for } x = M_b+1, b = 1, \cdots, d
\end{aligned}
$$

or,

$$
\begin{aligned}
\phi(\alpha, B) &= \tfrac{(k+x)}{x!}\Delta^x k^{x+m+d-1}, \\
&\quad \text{for } x = 0, 1, \cdots, M_1 \\
&= \tfrac{(k+x)}{x!}\Delta^x k^{x+m+d-b-1}, \\
&\quad \text{for } x = M_b+2, \cdots, M_{b+1}, b = 1, \cdots, d-1 \\
&= \tfrac{(k+x)}{x!}\Delta^x k^{x+m-1}, \\
&\quad \text{for } x \ge M_d+2 \\
&= \tfrac{1}{x!}\Delta^x k^{x+m+d-b}, \\
&\quad \text{for } x = M_b+1, b = 1, \cdots, d
\end{aligned}
$$

Similarly, for a positive integer c and $\alpha_{+c} = (x, y) \in B_{+c}$,

$$
\begin{aligned}
\phi(\alpha_{+c}, B_{+c}) &= \tfrac{(k+x)}{x!}\Delta^x k^{x+m+d+c-1}, && \text{for } x = 0, 1, \cdots, M_1 \\
&= \tfrac{(k+x)}{x!}\Delta^x k^{x+m+d+c-b-1}, && \text{for } x = M_b+2, .., M_{b+1}, \\
& && b = 1, \cdots, d-1 \\
&= \tfrac{(k+x)}{x!}\Delta^x k^{x+m+c-1}, && \text{for } x \ge M_d+2 \\
&= \tfrac{1}{x!}\Delta^x k^{x+m+d+c-b}, && \text{for } x = M_b+1, b = 1, \cdots, d
\end{aligned}
$$

Hence, for $\alpha = (x, y) \in B$, the UMVUE of N is given by,

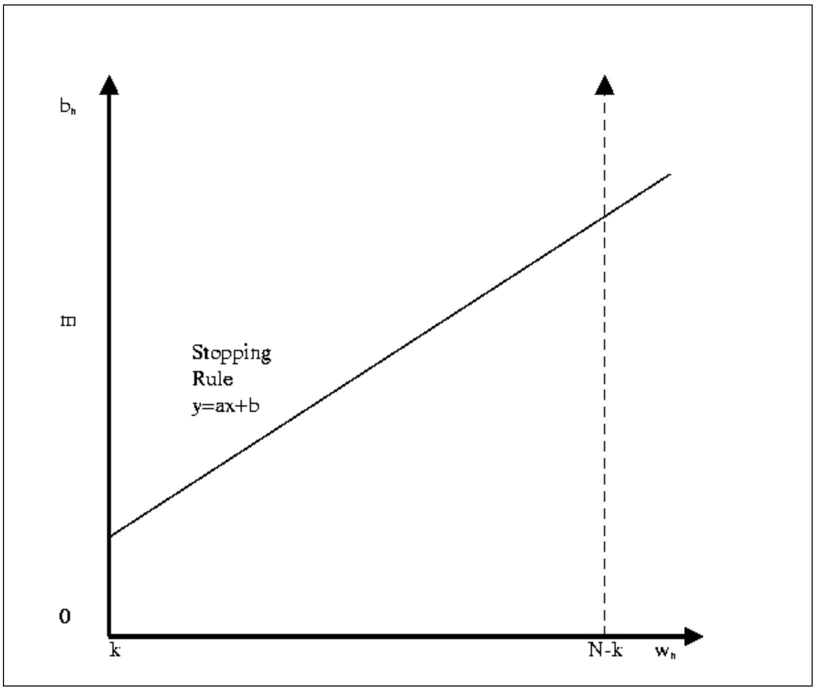
$$
\begin{aligned}
\hat{N}(\alpha) &= \tfrac{\Delta^x k^{x+m+d}}{\Delta^x k^{x+m+d-1}}, && \text{for } x = 0, 1, \cdots, M_1+1 \\
&= \tfrac{\Delta^x k^{x+m+d-b}}{\Delta^x k^{x+m+d-b-1}}, && \text{for } x = M_b+2, \cdots, M_{b+1}+1, b = 1, .., d-1 \\
&= \tfrac{\Delta^x k^{x+m}}{\Delta^x k^{x+m-1}}, && \text{for } x \ge M_d+2
\end{aligned}
$$

The UMVUE of $V(\hat{N})$ is $(\hat{N})^2 - \hat{N^2}$, where, for $\alpha = (x, y) \in B$,

$$
\begin{aligned}
\hat{N^2}(\alpha) &= \frac{\Delta^x k^{x+m+d+1}}{\Delta^x k^{x+m+d-1}}, \quad \text{for } x = 0, 1, \cdots, M_1+1 \\
&= \frac{\Delta^x k^{x+m+d-b+1}}{\Delta^x k^{x+m+d-b-1}}, \quad \text{for } x = M_b+2, \cdots, M_{b+1}+1, b = 1, \cdots, d-1 \\
&= \frac{\Delta^x k^{x+m+1}}{\Delta^x k^{x+m-1}}, \quad \text{for } x \ge M_d+2.
\end{aligned}
$$

Example 6.3.

Fig. 4.
Example 6.3.



Consider a sampling plan given by the set of boumdary points B={(x,ax+b),x=0,1,$\cdots$}, where a,b are given positive integers. B and B_+ are closed and B satisfies the condition of Theorem 4.1. For $\alpha = (x, ax+b)$,

$$\phi(\alpha, B) = \sum_{j=0}^{x-1} \sum_{i_j=0}^{aj+b-i_0-i_1-\cdots-i_{j-1}-1} \Pi_{t=0}^{x-1}(k+t)^{i_t}(k+x)^{ax+b-i_0-i_1-\cdots-i_{x-1}}$$

$$= (k+x)^{a+1} S^{a,b}_{x,a(x-1)+b-1,k}, \quad say$$

where, $S^{a,b}_{x,a(x-1)+b-1,k} =$

$$\sum_{j=0}^{x-1} \sum_{i_j=0}^{aj+b-i_0-i_1-\cdots-i_{j-1}-1} \Pi_{t=0}^{x-1}(k+t)^{i_t}(k+x)^{a(x-1)+b-i_0-i_1-\cdots-i_{x-1}-1}$$

which may also be defined by the following recurrence relation (see Berg (1987)):

$S^{a,b}_{x,y,k}$

$= (k+x)S^{a,b}_{x,y-1,k} + S^{a,b}_{x-1,y,k}, \quad for \;\; y < a(x-1)+b$

$= (k+x)S^{a,b}_{x,y-1,k}, \quad for \;\; a(x-1)+b \leq y \leq ax+b$

$= 0, \quad for \;\; y > ax+b$

with $S^{a,b}_{0,0,k} = 1$.

Similarly, for any positive integer c and for $\alpha_{+c} = (x, ax+b+c)$,

$$\phi(\alpha_{+c}, B_{+c}) = (k+x)^{a+1} S^{a,b+c}_{x,a(x-1)+b+c-1,k}.$$

So that the UMVUE of N is $\hat{N} = \frac{S^{a,b+1}_{x,a(x-1)+b,k}}{S^{a,b}_{x,a(x-1)+b-1,k}}$ and the UMVUE of $V(\hat{N})$ is $\hat{V}(\hat{N}) = (\hat{N})^2 - \hat{N^2}$, where $\hat{N^2} = \frac{S^{a,b+2}_{x,a(x-1)+b+1,k}}{S^{a,b}_{x,a(x-1)+b-1,k}}$. Here x denotes the number of white balls observed.

REFERENCES

REFERENCES: Estimation of Common Mean

Bhattacharya, C. G. (1978): Yates type estimators of common mean, *Ann. Inst. Stat. Math.***30,** 407-414.

Bhattacharya, C. G. (1979): A note on estimating the common mean of k normal populations, *Sankhya B,***40,** 272-275.

Bhattacharya, C. G. (1980): Estimation of common mean and recovery of interblock information,*Ann. Statist.***8,** 205-211.

Bhattacharya, C. G. (1981): Estimation of common location,*Common. Stat. A Theo. Meth.***10,** 955-961.

Bhattacharya, C. G. (1988): On the Cohen-Sakrowitz estimator of a common mean, *Statistics,***19,** 493-501.

Bradley, R. A. and Brindley, D. A. (1988): An approximate test of variance homogeneity based on Grubbs' estimators, *Commun. Statist. - Simul.,* **15,** 27-34.

Bradley, R. A. and Rollier, D. A. (1988): The extension of Grubbs estimators for classes of incomplete block designs, *Essays in honour of F. A. Graybill (Ed. J. N. Srivastava),* Elsevier Sc. Publishers (North-Holland), 43-59.

Brindley, D. A. and Bradley, R. A. (1985): Some new results on Grubbs' estimators, *J. Amer. Statist. Assoc.,* **80,** 711-714.

Das, K., Sinha, B. K. and Sinha, B. K. (1993): Variance components estimation - on the status of Grubbs' estimator, *Quality through engineering design* Ed. W. Kuo, Elsevier Science Publishers B. V., 141-157.

De, M. (1995): On a class of admissible estimators of common mean of two univariate normal populations, *C. S. A. Bulletin,***45(177-178)103-109.**

De, M. and Sarkar, J. (1998): Estimation of common mean of two univariate normal populations using two stage sampling,*C. S. A. Bulletin,***48(189-190),73-81.**

De, M., Sarkar, J. and Sinha, B. K. (1999): Estimating the variance components and the common mean of normal populations in the presence of covariates. *Paper presented at National Level Conference held in the Department of Statistics, Calcutta University in December, 1999.*

De, M. and Sarkar, J. and Sinha, B. K. (2004): Optimal sampling scheme for estimating the common mean of a bivariate normal population under differential sampling costs. *To appear in Journal of Applied Statistical Sciences.*

Ellenberg, J. H. (1977): The joint distribution of the standardized row sums of squares from a balanced two-way layout, *J. Amer. Statist. Assoc.*, **72,** 407-411.

Govindarajulu, Z. (1975): *Sequential Statistical Procedures* Academic Press, New York.

Gradstein, I. S. and Ryshik, I. M. (1980): Tables of series, products and integrals, *Academic Press,* New York.

Graybill, F. A. and Deal, R. B. (1959): Combining unbiased estimators, *Biometrics,***15,**543-550.

Grubbs, F. E. (1948): On estimating precision of measuring instruments and product variability, *J. Am. Stat. Assoc.***43,**243-264.

Johnson, N. L. , Kotz, S. and Balakrishnan, N. (1994): Continuous univariate distribution, Vol 2, 2nd ed. *John Wiley,* New York.

Khatri, C. G. and Srivastava, M. S. (1979): *Introduction to Multivariate Statistics,*New York: North-Holland.

Lehmann, E. L. (1983): Theory of Point Estimation,*John Wiley and Sons.*

Maloney, C. J. (1973): Grubbs's estimators to date, *Proc. 18th Conf. on Design of Experiments in Army Res. Devel. and testing***Part 2,** 523-547.

Norwood, T. E. and Hinklemann K. (1977): Estimating the common mean of several normal populations, *Annals of Statistics***5,** 1047-1050.

Pal N. and Sinha B. K. (1996): Estimation of common mean of several normal populations: A review, *Far East J. Math. Sci.***Special Vol. I,** 97-110.

Ritcher, D. (1960): Two-Stage experiments for estimating a common mean, *Ann. Math. Statist.***31,**1164-1173.

Searle. (1987): Linear Models for Unbalanced Data,*John Wiley and Sons*

Sinha B. K. (1985): Unbiased estimation of the variance of the G-D estimator of the common mean of the several normal populations,*Canad. J.Statist.***13(3),**243-247.

Sinha B. K. and Mouqadem O. (1982): Estimation of the Common Mean of two univariate normal populations, *Comm. Statist. Theo. Meth.* **11(14),**1603-1614.

Sun, Y. (1999): *Statistical Meta Analysis with Applications,*PhD. dissertation. University of Maryland, Baltimore County, USA.

Zacks S. (1970): Bayes and fiducial equivariant estimators of the common mean of two normal distributions, *Ann. Math. Stat.***41,**59-69.

REFERENCES: Estimation of the Size of a Finite Population

Ahmad, M., Alalouf, S. and Chaubay, Y. P. (1993): A comparison of three sampling strategies for estimating population size,*Quaderni di Statistica e Matematica,* XV, 2-11.

Ahmad, M., Alalouf, S. and Chaubay, Y. P. (2000): Estimation of population total when population size is unknown, *Statistics and Probability Letters,* **49,** 211-213.

Ahmad, M., Chaubay, Y. P. and Sen, P. K.(2000): On a sequential sampling design with multiple repeats for estimating population size, *Journal of Statistical Research,* **34,** 19-29.

Ahmad, M., and Chaubay, Y. P. (2000): Bayesian estimation of the size of a population under a sequential sampling design,*American Journal of Mathematical and Management Sciences,* **20,** 1&2, 131-143.

Berg,S.(1987).Estimation of the size of a finite population in sequential sampling: Truncated combinatorial numbers. *Sequential Analysis,* **6,** 165-177.

Boswell, M. T., Burnham, K. P., Patil, G. P. (1988): Role and use of composite sampling and capture-recapture sampling in echological studies,*Handbook of Statistics,*Vol. 6, Eds. P. R. Krishnaiah and C. R. Rao, North Holland, Amsterdam, 469-488.

Darroch, J. N. (1958): The multiple recapture census: I Estimation of a closed population,*Biometrika,***45,**343-359.

De, M. and Sengupta, S. (2003): Bernoulli Sequential Estimation of the size of a finite population, *Sequential Analysis,***22,95-106.**

Goodman, L. A. (1953): Sequential sampling tagging for population size problems,*Ann. Math. Statist.,* **24,**56-69.

Gupta, M. K. (1967): Unbiased estimate of 1/p,*Ann. Inst. Stat. Math.* **19,**413-416.

Hossain, M. F. (1995): Unknown population size estimation: an urn model approach, *Journal of Statistical Studies,***15,**89-94.

Johnson, N. L. and Kotz, S. (1969): Distribution in Statistics: Discrete Distribution, *John Wiley,* New York.

Samuel, E.(1968). Sequential maximum likelihood estimation of the size of a population.*Ann. Math. Statist.,***39,**1057-1068.

Seber, G. A. F. (1982): *The Estimation of Animal Abundance and Related Parameters,*2nd edition, Macmillan, New York.

Sengupata, S. and De, M. (1997): On the estimation of size of a finite population,*Sankhya, Series B,***9,66-75.**

Sinha, B. K. and Bose, A. (1985): Unbiased sequential estimation of 1/p: Settlement of a conjecture, *Ann. Inst. Statist. Math.,***37,** 455-460.

Sinha, B. K. and Sinha, B. K. (1975): Some problems in unbiased sequential binomial estimation, *Ann. Inst. Statist. Math.***27,**245-258.