# Two-Stage and Alternative Fixed-Width Interval Estimation Procedures for the Difference of Normal Means

N. Mukhopadhyay* Basil M de Silva† and Vasant B. Waikar‡

Mathematics and Statistics Research Report 2006/05

July 2006

**Department of Mathematics and Statistics**
**RMIT University**
GPO Box 2476V, Melbourne VIC 3001
AUSTRALIA

2000 AMS Classification: Primary 62L12; Secondary 62L05; 62J15.
ISBN 10 : 1-921166-40-1
ISBN 13 : 978-1-921166-40-2

*Department of Statistics, UBox 4120 CLAS Building, University of Connecticut, 215 Glenbrook Road, Storrs CT 06269-4120 U.S.A., email: nitis.mukhopadhyay@uconn.edu

†email: desilva@rmit.edu.au

‡Department of Mathematics and Statistics, Miami University, Oxford, Ohio 45056, U.S.A., email: waikarvb@muohio.edu

## Abstract

We consider two independent $N(\mu_1, \sigma^2)$ and $N(\mu_2, a^2\sigma^2)$ populations where $\mu_1, \mu_2, \sigma^2$ are unknown parameters with $-\infty < \mu_1, \mu_2 < \infty, 0 < \sigma < \infty$. We assume that $a(> 0)$ is known! The problem is one of estimating $\Delta = \mu_1 - \mu_2$ by some appropriately constructed fixed-width $(2d)$ confidence interval with the confidence coefficient at least $1 - \alpha$. Here, $d(> 0)$ and $0 < \alpha < 1$ are two preassigned numbers.

First, a two-stage procedure $\mathcal{P}_1$ is designed in the spirit of Stein (1945). Then, another two-stage procedure $\mathcal{P}_2$ is tried in the spirit of Chapman (1950). We report that the Stein-type two-stage procedure $\mathcal{P}_1$ performed better. Next, a new alternative sampling design $\mathcal{P}_3$ is proposed where we have incorporated a two-stage sampling technique only from one population, followed by a single-stage sampling strategy with a random sample size from the other population. We show that this new alternative estimation strategy $\mathcal{P}_3$ performs better than the Stein-type two-stage estimation strategy $\mathcal{P}_1$. The findings are validated by both theory and extensive sets of computer simulations. In the end, we exhibit the superiority of the new procedure $\mathcal{P}_3$ with the help of an example using a real dataset from horticulture (Mukhopadhyay et al., 2004).

**Key Words and Phrases:** Behrens-Fisher problem; Chapman procedure; Fixed-width interval; Operational simplicity; Proportional variance; Sampling; Single-stage sampling; Stein-type procedure; Superior alternative methodology; Two-stage sampling.

# 1. Introduction

We suppose that $X_{1,1}, X_{1,2}, ..., X_{1,n_1}, ...$ denote independent and identically distributed (i.i.d.) random variables having the common $N(\mu_1, \sigma^2)$ distribution and $X_{2,1}, X_{2,2}, ...,$ $X_{2,n_2}, ...$ denote i.i.d. random variables having the common distribution $N(\mu_2, a^2\sigma^2)$ where $\mu_1, \mu_2, \sigma^2$ are unknown parameters with $-\infty < \mu_1, \mu_2 < \infty, 0 < \sigma < \infty$. We assume that $a(> 0)$ is known! We also suppose that the $X_1$'s are independent of the $X_2$'s.

The problem is one of estimating $\Delta = \mu_1 - \mu_2$ by some appropriately constructed *fixed-width* confidence interval. For motivation, one may refer, for example, to Hochberg and Tamhane (1987), Mukhopadhyay and Solanky (1994), Hsu (1996), Bechhofer et al. (1995), Ghosh et al. (1997). Liu (1995), Mukhopadhyay and de Silva (1997), Aoshima (2001), Aoshima and Mukhopadhyay (2002), and de Silva and Waikar (2006) have cited other important sources too. When $a = 1$, Stein's (1945) two-stage procedure immediately applies. In the literature, one finds methods in the case when $a = 1$ or when $a(> 0)$ is unknown (Chapman, 1950).

Now, having observed the data $\mathbf{X_n} = \{X_{i,1}, X_{i,2}, ..., X_{i,n_i}; i = 1, 2\}$, we consider estimating $\Delta$ by the fixed-width confidence interval

$$J_{\mathbf{n}} = [U_{\mathbf{n}} - d, U_{\mathbf{n}} + d] \text{ with preassigned } d(> 0) \ , \tag{1.1}$$

where we denote

$$\overline{X}_{i,n_i} = n_i^{-1}\Sigma_{j=1}^{n_i} X_{i,j}, i = 1, 2 \text{ and } U_{\mathbf{n}} = \overline{X}_{1,n_1} - \overline{X}_{2,n_2}, \mathbf{n} = (n_1, n_2).$$

Let us denote the sample variance based on the data recorded from the $i^{th}$ population and the pooled sample variance by

$$\begin{aligned} S_{i,n_i}^2 &= (n_i - 1)^{-1}\Sigma_{j=1}^{n_i}(X_{i,j} - \overline{X}_{i,n_i})^2, n_i \geq 2, i = 1, 2, \text{ and} \\ V_{\mathbf{n}}^2 &= \left\{(n_1 - 1)S_{1,n_1}^2 + (n_2 - 1)a^{-2}S_{2,n_2}^2\right\}/(n_1 + n_2 - 2), \end{aligned} \tag{1.2}$$

respectively.

One should note that the traditional confidence intervals have random widths which can be fairly large with positive probability even if the population variances are small. An experimenter will choose a "small" positive number $d$ to begin with so that a proposed confidence interval of width $2d$ would not be construed too wide for practical uses.

We additionally require the confidence coefficient associated with $J_{\mathbf{n}}$ to be at least $1 - \alpha$ where $0 < \alpha < 1$ is also preassigned. That is, we must have

$$P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{n}}\} \geq 1 - \alpha \text{ for all fixed } \boldsymbol{\mu}, \sigma, d, a, \text{ and } \alpha, \tag{1.3}$$

where we denote $\boldsymbol{\mu} = (\mu_1, \mu_2)$. It is easy to check that (1.3) holds if we determine

$$\begin{aligned} n_1 \text{ to be the smallest integer} \geq 2z_{\alpha/2}^2\sigma^2/d^2 = n_1^*, \text{ say, and} \\ n_2 \text{ is the smallest integer} \geq 2z_{\alpha/2}^2 a^2\sigma^2/d^2 = n_2^*, \text{ say.} \end{aligned} \tag{1.4}$$

where $z_{\alpha/2}$ is the upper $50\alpha\%$ point of the standard normal distribution. It is clear that the magnitude of $n_1^*$, and hence that of $n_2^*(\equiv a^2 n_1^*)$, remains unknown since $\sigma^2$ is unknown. Let us denote the (unknown) total fixed-sample-size by $n^*$, that is one has

$$n^* = n_1^* + n_2^* = (1 + a^2)n_1^*. \tag{1.5}$$

It is well-known that no fixed-sample-size procedure would provide a solution for the problem on hand (Dantzig, 1940; Ghosh et al., 1997, Section 3.7). So, we opt for sampling in two steps.

## 2. A Stein-Type Two-Stage Procedure

Stein's (1945,1949) original two-stage procedure will apply immediately when $a = 1$. In what follows, we extend the Stein two-stage procedure when $a$ is known but $a$ is not necessarily one. Let $\langle u \rangle$ stand for the largest integer $< u$.

**The Procedure** $\mathcal{P}_1$: Now, one starts with the pilot observations $\mathbf{X}_m = \{X_{i,1}, X_{i,2}, ..., X_{i,m}; i = 1, 2\}$ where $m \geq 2$. Then, one defines the stopping variables

$$N_1 = \max\left\{m, \langle 2q_m^2 V_m^2/d^2 \rangle + 1\right\}, N_2 = \max\left\{m, \langle a^2 N_1 \rangle + 1\right\} \tag{2.1}$$

where the pooled estimator $V_m^2$ of $\sigma^2$ from (1.2) reduces to

$$V_m^2 = \tfrac{1}{2}\left\{S_{1,m}^2 + a^{-2}S_{2,m}^2\right\}.$$

We emphasize that $q_m \equiv q_m(\alpha) > 0$ has to be appropriately determined in the sequel.

The positive integer valued random variables $N_1, N_2$ respectively estimate the fixed-sample-sizes $n_1^*, n_2^*$ defined in (1.4). Thus, we would treat $N = \Sigma_{i=1}^2 N_i$ as a natural estimator of $n^*$ from (1.5).

The two-stage procedure $\mathcal{P}_1$ from (2.1) is implemented as follows. On the basis of the pilot observations of size $m$ from the $i^{th}$ population, we first determine $N_i$. If $N_i = m$ is observed, then no more observations are taken from the $i^{th}$ population, because the pilot sample size itself overestimates $n_i^*, i = 1, 2$. But, if one observes $N_i > m$, then $N_i - m$ additional observations are taken from the $i^{th}$ population in the second stage, $i = 1, 2$. The combined data from the pilot and the second stage from both populations is then $\mathbf{X_N} = \{X_{i1}, X_{i2}, ..., X_{iN_i}; i = 1, 2\}$ where we denote $\mathbf{N} = (N_1, N_2)$. Finally, the fixed-width confidence interval

$$J_{\mathbf{N}} = [U_{\mathbf{N}} - d, U_{\mathbf{N}} + d] \text{ where } d(> 0) \text{ is preassigned,} \tag{2.2}$$

is then proposed for $\Delta$ in the light of (1.1).

At this point, the main concern ought to be the *appropriate determination* of the design constant $q_m(\alpha)$. In what follows, we pursue this concern.

**Theorem 2.1** *Under two-stage sampling procedure $\mathcal{P}_1$ from (2.1), for all fixed $\mu_1, \mu_2, \sigma, d, a$, and $\alpha$, we have*

$$P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{N}}\} \geq 1 - \alpha,$$

*where $\mathbf{N} = (N_1, N_2)$, if $q_m \equiv q_m(\alpha)$ is chosen as the upper $50\alpha\%$ point $t_{2m-2,\alpha/2}$ for the Student's $t$ distribution with $2(m-1)$ degrees of freedom.*

*Proof*: First, we observe some simple facts. The random variable $I(N_1 = n_1, N_2 = n_2)$ depends only on $V_m^2$, that is, on the random variable $\left(S_{1,m}^2, S_{2,m}^2\right)$ only for every fixed $n_1 \geq m, n_2 \geq m$. Thus, the random variable $I(N_1 = n_1, N_2 = n_2)$ must be independent of $U_{\mathbf{n}}$ for every fixed $n_1 \geq m, n_2 \geq m$. Also, $2(m-1)V_m^2/\sigma^{-2}$ is distributed as $\chi^2_{2m-2}$. Then, we have

$$\begin{aligned} &P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{N}}\} \\ &\quad = E_\sigma\left[2\Phi\left(d\sigma^{-1}\left\{N_1^{-1} + a^2 N_2^{-1}\right\}^{-1/2}\right) - 1\right] \\ &\quad \geq E_\sigma\left[2\Phi\left(q_m V_m \sigma^{-1}\right) - 1\right], \end{aligned} \tag{2.3}$$

since $N_1 \geq 2q_m^2 V_m^2/d^2$ so that $N_2 \geq 2a^2 q_m^2 V_m^2/d^2$ with probability one. The last expression in (2.3) is obviously exactly $1 - \alpha$ if $q_m \equiv q_m(\alpha)$ is chosen as the upper $50\alpha\%$ point of the Student's $t$ distribution with $2(m-1)$ degrees of freedom. □

# 3. The Chapman Two-Stage Procedure

Chapman (1950) had proposed a two-stage fixed-width confidence interval procedure for estimating $\Delta = \mu_1 - \mu_2$ in the case of two independent $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ populations where $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ are all unknown parameters with $-\infty < \mu_1, \mu_2 < \infty, 0 < \sigma_1^2, \sigma_2^2 < \infty$. One may also refer to Ghosh et al. (1997, p. 186). The Chapman procedure was further investigated by Ghosh (1975a,b). Aoshima and Mukhopadhyay (2002) generalized this methodology with much emphasis on practical implementation. de Silva and Waikar (2006) have proposed ways to make the original methodology more efficient by appropriately incorporating some interesting techniques via bootstrapping.

We first show that Chapman's methodology can still apply in our problem for estimating $\Delta = \mu_1 - \mu_2$ in the case of two independent $N(\mu_1, \sigma^2)$ and $N(\mu_2, a^2\sigma^2)$ populations where $\mu_1, \mu_2, \sigma^2$ are all unknown parameters with $-\infty < \mu_1, \mu_2 < \infty, 0 < \sigma < \infty$. Recall that $\langle u \rangle$ continues to stand for the largest integer $< u$.

**The Procedure** $\mathcal{P}_2$: One again starts with the pilot observations $\mathbf{X}_m = \{X_{i,1}, X_{i,2}, ..., X_{i,m}; i = 1, 2\}$ where $m \geq 2$. Then, one defines the stopping variables

$$N_i = \max\left\{m, \left\langle h_m^2 S_{i,m}^2 / d^2 \right\rangle + 1\right\}, i = 1, 2 \tag{3.1}$$

where $h_m \equiv h_m(\alpha) > 0$ is determined from the following equation:

$$P\left\{T_1 - T_2 \leq h_m\right\} = 1 - \tfrac{1}{2}\alpha \tag{3.2}$$

with $T_i$'s independent random variables, each having the Student's $t$ distribution with $m-1$ degrees of freedom, $i = 1, 2$. Chapman (1950) included a table giving the required $h_m$ values whereas Ghosh (1975a,b) derived the Cornish-Fisher expansion of $h_m$. Referring to Chapman's (1950) $h_m$ values, Ghosh (1975b, p. 463) justifiably noted that "many of his values are incorrect." We may add that Ghosh's (1975b) approximations and tables have proven to be immensely useful. The Cornish-Fisher expansion given by Aoshima and Mukhopadhyay (2002) turned out to be more versatile in practical implementation of the corresponding estimation technique.

The two-stage procedure $\mathcal{P}_2$ from (3.1)-(3.2) is implemented as follows. On the basis of the pilot sample of size $m$ from the $i^{th}$ population, one first determines $N_i$. If $N_i = m$ is observed, then no more observations are taken from the $i^{th}$ population, because the pilot sample size itself is an overestimate of $n_i^*, i = 1, 2$. But, if one observes $N_i > m$, then $N_i - m$ additional observations are taken from the $i^{th}$ population in the second stage, $i = 1, 2$. The positive integer valued random variable $N_i$ estimates the fixed-sample-size $n_i^*$, defined in (1.4), from the $i^{th}$ population, $i = 1, 2$. We would treat $N = \Sigma_{i=1}^2 N_i$ as a natural estimator of $n^*$. The combined data from the pilot and the second stage happens to be $\mathbf{X_N} = \{X_{i,1}, X_{i,2}, ..., X_{i,N_i}; i = 1, 2\}$ where $\mathbf{N} = (N_1, N_2)$. Finally, in the light of (1.1), the fixed-width confidence interval

$$J_{\mathbf{N}} = [U_{\mathbf{N}} - d, U_{\mathbf{N}} + d] \tag{3.3}$$

is then proposed for $\Delta$ under two-stage sampling procedure $\mathcal{P}_2$ from (3.1)-(3.2).

Then, we may note that

$$\begin{aligned} & P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{N}}\} \\ & \quad = E_\sigma\left[2\Phi\left(d\sigma^{-1}\left\{N_1^{-1} + a^2 N_2^{-1}\right\}^{-1/2}\right) - 1\right] \\ & \quad \geq E_\sigma\left[2\Phi\left(h_m\left\{\sigma^2 S_{1,m}^2 + a^2\sigma^2 S_{2,m}^2\right\}^{-1/2}\right) - 1\right], \end{aligned} \tag{3.4}$$

since $N_i \geq h_m^2 S_{i,m}^2/d^2, i = 1, 2$, with probability one. The last expression in (3.4) is obviously exactly $1 - \alpha$ if $h_m \equiv h_m(\alpha)$ is chosen satisfying (3.2). This may be verified along the line of Chapman (1950) or by invoking Theorem 6.7.1 in Ghosh et al. (1997, p. 186).

**Remark 3.1** In a fully Behrens-Fisher situation, that is if $\sigma_1^2$, $\sigma_2^2$ were completely unknown, purely sequential methodologies were introduced by Robbins et al. (1967) and Srivastava (1970) in order to construct a fixed-width confidence interval for $\Delta = \mu_1 - \mu_2$. Since then, these methodologies have been generalized and applied in many other types of problems. One will find some of these sources listed in Ghosh et al. (1997) and in other places including Aoshima (2001), Aoshima and Mukhopadhyay (2002), de Silva and Waikar (2006), Mukhopadhyay and de Silva (1997), Mukhopadhyay et al. (2004). In this note, we do not consider purely sequential or other multi-stage sampling techniques for at least two important reasons: These other sampling designs (1) are operationally more complex than two-stage sampling designs, and (2) would let us conclude only that "the achieved confidence coefficient is *asymptotically* $1 - \alpha$" rather than "the achieved confidence coefficient is *at least* $1 - \alpha$".

## 3.1. Comparing the Stein-Type and Chapman Procedures

Note that $E_\sigma [N_1 + N_2 \mid \mathcal{P}_1] \approx 2q_m^2(1 + a^2)\sigma^2/d^2$ for the Stein-type procedure (2.1), but on the other hand we have $E_\sigma [N_1 + N_2 \mid \mathcal{P}_2] \approx h_m^2(1+a^2)\sigma^2/d^2$ for the Chapman procedure (3.1)-(3.2). Recall from (1.5) that $n^* = 2z_{\alpha/2}^2(1+a^2)\sigma^2/d^2$. We can surely argue that $q_m \equiv t_{2m-2,\alpha/2}, h_m$ would be "close" to $z_{\alpha/2}, \sqrt{2}z_{\alpha/2}$ respectively for "large" $m$. Thus, we may expect the following

$$\begin{aligned} E_\sigma [N_1 + N_2 \mid \mathcal{P}_1]/n^* &\approx q_m^2/z_{\alpha/2}^2 \equiv \tau_1, \text{ say}, \\ E_\sigma [N_1 + N_2 \mid \mathcal{P}_2]/n^* &\approx \tfrac{1}{2}h_m^2/z_{\alpha/2}^2 \equiv \tau_2, \text{ say}, \end{aligned} \tag{3.5}$$

to hold approximately.

In Table 1, we first provide the values of $q_m$ and $2^{-1/2}h_m$ when $m$ = 5, 10, 15, 20, 25, 30, 40, 50 and $\alpha = 0.05, 0.01$. The $q_m$ values were obtained using MAPLE whereas $2^{-1/2}h_m$ values are taken from Table 1 in Aoshima and Mukhopadhyay (2002). Then, the efficiency ratios $\tau_1$ and $\tau_2$ from (3.5) were calculated.

The values of $\tau_1$ and $\tau_2$ both exceed one which clearly indicate that the two-stage procedures $\mathcal{P}_1$ from (2.1) and $\mathcal{P}_2$ from (3.1)-(3.2) would be respectively expected to oversample compared with the *optimal* (had $\sigma^2$ been known) fixed sample size $n^*$. This should not be surprising because $n_1^*$, $n_2^*$, and $n^*$ are unknown in the first place. Also, when $\alpha$ is held fixed, one notes that $\tau_1$ is smaller than $\tau_2$ which means that the Stein-type procedure $\mathcal{P}_1$ is expected to oversample less than the Chapman procedure

$\mathcal{P}_2$. Again, we should expect to see this feature because the Chapman procedure $\mathcal{P}_2$ is supposed to be used if the two population variances were unequal. The procedure $\mathcal{P}_2$ surely did not take into account the fact that we had assumed $\sigma_1^2 = \sigma^2, \sigma_2^2 = a^2\sigma^2$ with $a(> 0)$ known! On the other hand, the Stein-type procedure $\mathcal{P}_1$ fully utilized this special structure regarding the variances! So, it seems natural for us to expect that the Stein-type procedure $\mathcal{P}_1$ from (2.1) should perform better and it does. Given this, what may be surprising is that the efficiency ratios $\tau_1, \tau_2$ become nearly the same when the pilot sample size $m$ is chosen to be a little large (40 or more).

**Table 1.** Values of $\tau_1$ and $\tau_2$ from (3.5)

| $m$ | $q_m$ | $2^{-1/2}h_m$ | $\tau_1$ | $\tau_2$ |
|---|---|---|---|---|
| | | $\boldsymbol{\alpha = 0.05, z_{\alpha/2} = 1.96}$ | | |
| 5 | 2.3060 | 2.779 | 1.3842 | 2.0103 |
| 10 | 2.1009 | 2.249 | 1.1489 | 1.3166 |
| 15 | 2.0484 | 2.134 | 1.0922 | 1.1854 |
| 20 | 2.0244 | 2.084 | 1.0668 | 1.1305 |
| 25 | 2.0106 | 2.056 | 1.0523 | 1.1004 |
| 30 | 2.0017 | 2.039 | 1.0430 | 1.0822 |
| 40 | 1.9908 | 2.018 | 1.0317 | 1.0601 |
| 50 | 1.9845 | 2.006 | 1.0252 | 1.0475 |
| | | $\boldsymbol{\alpha = 0.01, z_{\alpha/2} = 2.5758}$ | | |
| 5 | 3.3554 | 4.278 | 1.6969 | 2.7584 |
| 10 | 2.8784 | 3.114 | 1.2488 | 1.4615 |
| 15 | 2.7633 | 2.891 | 1.1509 | 1.2597 |
| 20 | 2.7116 | 2.798 | 1.1082 | 1.1800 |
| 25 | 2.6822 | 2.747 | 1.0843 | 1.1373 |
| 30 | 2.6633 | 2.715 | 1.0691 | 1.1110 |
| 40 | 2.6403 | 2.677 | 1.0507 | 1.0801 |
| 50 | 2.6269 | 2.656 | 1.0401 | 1.0632 |

We conclude that given a choice between these two procedures $\mathcal{P}_1$ and $\mathcal{P}_2$, one ought to use the Stein-type procedure $\mathcal{P}_1$ from (2.1) if some known value of $a$ can be confidently postulated. But, if there is some hesitation on the part of an experimenter about a possible known value for $a$, a practical advice is then to implement the Chapman procedure $\mathcal{P}_2$ from (3.1)-(3.2) with a pilot sample size $m$ that is 40 or more.

Section 5 includes more elaborate data analysis and comparisons of the two procedures $\mathcal{P}_1$ from (2.1) and $\mathcal{P}_2$ from (3.1)-(3.2).

# 4. A New Alternative Procedure

Let us rethink through the Stein procedure (2.1). The expressions of $n_1^*, n_2^*$ involve the same unknown variance $\sigma^2$. Hence, instead of recording $m$ pilot observations from both populations first, we could conceivably start by taking $2m$ pilot observations from population 1 *alone* in order to estimate the fixed-sample-size $n_1^*$. The required estimated sample size from population 2 would then be determined easily by appropriately adjusting the previously found estimate of $n_1^*$. In other words, a single-stage sampling design of appropriate size can be carried out from population 2 after the required number of observations have been gathered from population 1 in two steps.

## 4.1. The Motivation

We may cite the following scenario. In a greenhouse, suppose that a horticulturist wants to compare two varieties of seeds of some flower and record the time a variety takes for the first bud to appear. Now, the Stein-type procedure (2.1) requires that the horticulturist prepare two distinct sets of $m$ experimental pots for planting seeds of each variety. The two sets of experimental pots must be labeled, monitored, fed, watered, and kept separated from each other. One must bear some expenses for this added effort and attention that will be obviously essential. In some cases, there may not be enough available space or required help to prepare and monitor two distinct sets of $m$ experimental pots with seeds of each variety in one location under one roof. One will find an example of this kind and some data analysis in Mukhopadhyay et al. (2004). Our illustration included in Section 6 is based on some of the data that were reported and analyzed in Mukhopadhyay et al. (2004).

We may also record a subtle point here. One observes that the distribution of the stopped sample sizes arising from the Stein-type procedure (2.1) involves only $\sigma^2$ and not $\mu_1, \mu_2$. Given this, why should one at all take pilot observations from both seed varieties to estimate $\sigma^2$? Instead, we may simply record pilot observations of size $2m$ from the first variety alone. This approach will save significant expenses due to logistics, experimental setup, and monitoring.

Last, but not least, we cite the most attractive practical advantage of the alternative Stein-type procedure $\mathcal{P}_3$ from (4.1). We will show that the new alternative procedure $\mathcal{P}_3$ would require fewer total number of observations on an average compared with the Stein-type procedure (2.1). So, the operational rationality behind the modification proposed in (4.1) and its advantages over the Stein-type procedure (2.1) should be both obvious and appealing in some designed experiments.

One may criticize this alternative methodology because it would start with a pilot sample of size $2m$ from, say, $X_1$ *only* whereas the Stein-type procedure (2.1) had instead started with $m$ pilot observations from $X_1$. We may emphasize that the Stein-type procedure (2.1) *also* required $m$ pilot observations from $X_2$. But, one notices no differential cost structure involved here. So, why not allocate the same resources (budgetary, personnel, or otherwise) that would have been spent on gathering $m$ pilot observations from $X_2$ toward gathering the same number of additional data from $X_1$? Suppose an experimenter was preparing to implement the Stein-type procedure (2.1) with some *pre-fixed pilot size* $m$ from both populations 1 and 2. The new alternative methodology would ask the experimenter to begin with a pilot of size $2m$ from one population only. This change of one's mind-set will not require new resources for sampling or a new specification of $m$.

## 4.2. A New Alternative Procedure $\mathcal{P}_3$

We decide to start with $X_{1,1}, X_{1,2}, ..., X_{1,2m}$, the pilot observations from the first population *alone* with $m \geq 1$, and then define the stopping variables

$$N_1 = \max\left\{2m, \left\langle 2b_m^2 S_{1,2m}^2/d^2\right\rangle + 1\right\}, N_2 = \langle a^2 N_1\rangle + 1. \tag{4.1}$$

Here, we have used the pilot sample variance $S_{1,2m}^2$ as the sole estimator of $\sigma^2$ unlike what we had proposed to do in (2.1). It should be clear that $b_m \equiv b_m(\alpha) > 0$ has to be appropriately determined in the sequel.

The sampling procedure (4.1) is implemented as follows. On the basis of the pilot observations of size $2m$ from the first population, we first determine $N_1$. If $N_1 = 2m$ is observed, then no more observations are taken from this population. But, if one observes $N_1 > 2m$, then $N_1 - 2m$ additional observations are taken from the first population in the second stage. That is, we first obtain $N_1$ observations from population 1 alone in two stages. Next, we determine $N_2$ and thereby record $N_2$ observations from the second population in a *single* step. The combined data from both populations is then $\mathbf{X_N} = \{X_{i,1}, X_{i,2}, ..., X_{i,N_i}; i = 1, 2\}$ where we again denote $\mathbf{N} = (N_1, N_2)$. Finally, the fixed-width confidence interval

$$J_{\mathbf{N}} = [U_{\mathbf{N}} - d, U_{\mathbf{N}} + d] \text{ where } d(>0) \text{ is preassigned,} \tag{4.2}$$

is then proposed for $\Delta$ in light of (1.1).

At this point, again the main concern is the *appropriate determination* of the design constant $b_m(\alpha)$. In what follows, we pursue this concern.

**Theorem 4.1** *Under the alternative two-stage sampling procedure $\mathcal{P}_3$ from (4.1), for all fixed $\mu_1, \mu_2, \sigma, d, a$, and $\alpha$, we have*

$$P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{N}}\} \geq 1 - \alpha,$$

*where $\mathbf{N} = (N_1, N_2)$, if $b_m \equiv b_m(\alpha)$ is chosen as the upper $50\alpha\%$ point $t_{2m-1,\alpha/2}$ for the Student's $t$ distribution with $2m-1$ degrees of freedom.*

*Proof*: Along the lines of our proof for Theorem 2.1, we can write

$$\begin{aligned} & P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{N}}\} \\ & \quad = E_\sigma \left[ 2\Phi \left( d\sigma^{-1} \left\{ N_1^{-1} + a^2 N_2^{-1} \right\}^{-1/2} \right) - 1 \right] \\ & \quad \geq E_\sigma \left[ 2\Phi \left( b_m S_{1,2m} \sigma^{-1} \right) - 1 \right], \end{aligned} \tag{4.3}$$

since we again have $N_1 \geq 2b_m^2 S_{1,2m}^2/d^2$ so that $N_2 \geq 2a^2 b_m^2 S_{1,2m}^2/d^2$ with probability one. Now, we note that $(2m-1)S_{1,2m}^2\sigma^{-2}$ is distributed as $\chi_{2m-1}^2$. Thus, the last expression in (4.3) is obviously exactly $1-\alpha$ if $b_m \equiv b_m(\alpha)$ is chosen as the upper $50\alpha\%$ point $t_{2m-1,\alpha/2}$ of the Student's $t$ distribution with $2m-1$ degrees of freedom. □

## 4.3. Comparing the Stein-Type and the New Alternative Procedures

There are subtle differences between how the two methodologies (2.1) and (4.1) are implemented. We had explained these points in Section 4.1. The Stein-type methodology (2.1) consists of sampling in two-stages with pilot observations of size $m(\geq 2)$ from both populations. But, the alternative methodology (4.1) consists of sampling in two-stages with pilot observations of size $2m$ from the first population alone that is followed by batch sampling from the second population in a single step. In other words, the alternative methodology (4.1) is operationally more convenient than the original methodology (2.1).

We may note that $E_\sigma[N_1 + N_2 \mid \mathcal{P}_3] \approx 2b_m^2(1+a^2)\sigma^2/d^2$ for the alternative methodology (4.1) where $b_m \equiv t_{2m-1,\alpha/2}$ would be "close" to $z_{\alpha/2}$ for "large" $m$. So, one may expect the following

$$E_\sigma[N_1 + N_2 \mid \mathcal{P}_3]/n^* \approx b_m^2/z_{\alpha/2}^2 \equiv \tau_3, \text{ say,} \tag{4.4}$$

to hold approximately.

**Table 2.** Values of $\tau_3$ from (4.4)

| | $m$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 15 | 20 | 25 | 30 | 40 | 50 |
| | **$\boldsymbol{\alpha} = 0.05, \mathbf{z}_{\alpha/2} = 1.96$** | | | | | | | |
| $b_m$ | 2.2622 | 2.0930 | 2.0452 | 2.0227 | 2.0096 | 2.0010 | 1.9905 | 1.9842 |
| $\tau_3$ | 1.3321 | 1.1403 | 1.0888 | 1.0650 | 1.0513 | 1.0423 | 1.0314 | 1.0248 |
| | **$\boldsymbol{\alpha} = 0.01, \mathbf{z}_{\alpha/2} = 2.5758$** | | | | | | | |
| $b_m$ | 3.2498 | 2.8609 | 2.7564 | 2.7079 | 2.6800 | 2.6618 | 2.6395 | 2.6264 |
| $\tau_3$ | 1.5918 | 1.2336 | 1.1451 | 1.1052 | 1.0825 | 1.0679 | 1.0501 | 1.0397 |

Recall that $q_m \equiv q_m(\alpha)$ and $b_m \equiv b_m(\alpha)$ are chosen as the upper $50\alpha\%$ point of the Student's $t$ distribution with $2m-2$ and $2m-1$ degrees of freedom respectively. But, since $q_m$ will surely exceed $b_m$, the discrepancy ratio $\tau_1$ from (3.5) will necessarily exceed the discrepancy ratio $\tau_3$ from (4.4). In other words, the alternative Stein-type methodology (4.1) will require fewer observations than the original Stein methodology (2.1). So, the alternative methodology (4.1) wins on two fronts: Operationally, the new methodology is more attractive and also it will require fewer observations on an average!

A quick glance at the corresponding values of $\tau_1$ (from Table 1) and $\tau_3$ validates the fact that the alternative methodology (4.1) requires fewer observations. For small pilot sample sizes $m$, the new methodology can save substantial number of observations whereas this edge diminishes as $m$ grows. But, the moot point is that the new methodology will require fewer observations on an average whatever be $m$!

**Remark 4.1** We have consistently argued in favor of the newly proposed procedure (4.1), and yet one may consider implementing the same methodology with a pilot sample of size $m$ (instead of $2m$) from population 1. One may decide to experiment with this idea especially since the Stein-type procedure (2.1) required $m$ pilot observations from both populations. But, in the sense of having less oversampling, the methodology $\mathcal{P}_3$ as stated in (4.1) will be superior to a similar methodology with a pilot sample size $m$ from population 1 followed by sampling from population 2 in one step.

**Remark 4.2** While implementing the methodology $\mathcal{P}_3$, we began with a pilot sample of size $2m$ from population 1 followed by sampling from population 2 in one step. Alternatively, one could surely begin with pilot observations $X_{2,1}, X_{2,2}, ..., X_{2,2m}$ of size $2m$ from population 2 followed by sampling from population 1 in one step. Then, one would define the stopping variables

$$Q_2 = \max\left\{2m, \left\langle 2b_m^2 S_{2,2m}^2/d^2\right\rangle + 1\right\}, Q_1 = \langle a^{-2}Q_2\rangle + 1 \tag{4.5}$$

where the pilot sample variance $S_{2,2m}^2$ is used as the sole estimator of $a^2\sigma^2$ unlike what we had proposed to do in (4.1). Again, $b_m \equiv b_m(\alpha)$ is the upper $50\alpha\%$ point $t_{2m-1,\alpha/2}$ for the Student's $t_{2m-1}$ distribution. Finally, based on the combined data

$$\mathbf{X_Q} = \{X_{i,1}, X_{i,2}, ..., X_{i,Q_i}; i = 1, 2\}, \mathbf{Q} = (Q_1, Q_2)$$

from both populations, the fixed-width confidence interval $J_{\mathbf{Q}} = [U_{\mathbf{Q}} - d, U_{\mathbf{Q}} + d]$ would be proposed for $\Delta$ in the light of (1.1). One would then conclude that $P_{\boldsymbol{\mu},\sigma}\{\Delta \in J_{\mathbf{Q}}\} \geq 1-\alpha$. It is rather obvious that the random variables $Q_1, Q_2$, and $\mathbf{Q}$ from (4.5) and the random variables $N_1, N_2$, and $\mathbf{N}$ from (4.1) have respectively identical distributions. That is, the methodology $\mathcal{P}_3$ from (4.1) would be equivalent to the methodology described via (4.5). In other words, there is no so called identifiability issue involved here. From a practical perspective, however, if the implementation of two-stage sampling may be logistically simpler when carried out from population 1(2), then one should clearly opt for double sampling with $2m$ pilot observations from population 1(2) followed by sampling from population 2(1) in one step.

## 5. Comparing the Performances with Simulations

In this section, we proceed to investigate how the three procedures $\mathcal{P}_1$-$\mathcal{P}_3$ compare with each other with the help of computer simulations. We generated pseudo-random samples independently from $N(\mu_1, 2^2)$ and $N(\mu_2, 2^2a^2)$ populations. We fixed the set of values $\mu_1 = 8, \mu_2 = 4$ plus five possible choices for the values of $m$ and $a$, namely, $m = 5, 10, 15, 25, 30$ and $a = 0.25, 0.50, 0.75, 1.0, 2.0$.

We proceed to construct 95% confidence intervals for the difference of two means, $\Delta = \mu_1 - \mu_2$, with the fixed-width $2d$. Table 3 provides the values of $d$ with the corresponding values of $n^*(\equiv n_1^* + n_2^*)$ from (1.5). This table shows that our simulations were designed to include a wide range of values for sample sizes.

**Table 3.** Optimal Sample Size $n^*$ for Given $d$, $\alpha = 0.05, \sigma = 2$, and $a$

| | $n^* \equiv n^*(d, \alpha, a)$ | | | | |
|---|---|---|---|---|---|
| $d$ | $a = 0.25$ | $a = 0.50$ | $a = 0.75$ | $a = 1$ | $a = 2$ |
| 1.0 | 32.7 | 38.4 | 48.0 | 61.5 | 153.7 |
| 0.9 | 40.3 | 47.4 | 59.3 | 75.9 | 189.7 |
| 0.8 | 51.0 | 60.0 | 75.0 | 96.0 | 240.1 |
| 0.7 | 66.6 | 78.4 | 98.0 | 125.4 | 313.6 |
| 0.6 | 90.7 | 106.7 | 133.4 | 170.7 | 426.8 |
| 0.5 | 130.6 | 153.7 | 192.1 | 245.9 | 614.6 |
| 0.4 | 204.1 | 240.1 | 300.1 | 384.1 | 960.4 |
| 0.3 | 362.8 | 426.8 | 533.5 | 682.9 | 1707.3 |
| 0.2 | 816.3 | 960.4 | 1200.5 | 1536.6 | 3841.5 |
| 0.1 | 3265.2 | 3841.5 | 4801.8 | 6146.3 | 15365.8 |

Under each configuration of the values of $d$, $a$ and $m$ with fixed values of $\mu_1 = 8, \mu_2 = 4, \sigma = 2$, we replicated $10,000$ independent simulations. The summary results presented in this section are based on averages obtained from $10,000$ independent runs in each case.

We recorded the average sample sizes, *generically denoted* by $\overline{n}_1, \overline{n}_2$ from populations 1 and 2 respectively and the average total sample size by $\overline{n}$, under all three estimation methodologies $\mathcal{P}_1$-$\mathcal{P}_3$. We noted that $\overline{n}_1, \overline{n}_2$ and $\overline{n}$ exceeded the corresponding optimal sample sizes $n_1^*, n_2^*$ and $n^*$ in (1.4) and (1.5) under the procedures $\mathcal{P}_1$-$\mathcal{P}_3$ as expected. Further, the Chapman procedure consistently came out with the largest discrepancies between the average simulated sample sizes $\overline{n}_1, \overline{n}_2, \overline{n}$ and the optimal sample sizes $n_1^*, n_2^*, n^*$. The new alternative procedure $\mathcal{P}_3$ was associated with the smallest discrepancy between the average simulated sample sizes $\overline{n}_1, \overline{n}_2, \overline{n}$ and the optimal sample sizes $n_1^*, n_2^*, n^*$. This performance is clearly illustrated by Figure 1.

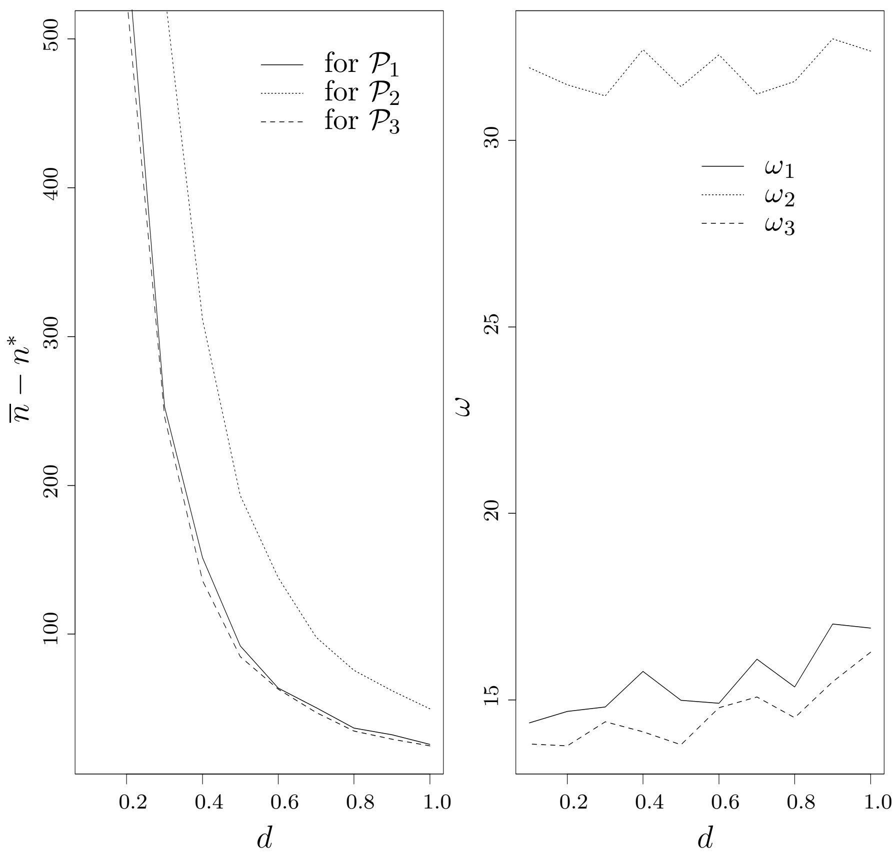


Fig 1.a Fig 1.b

Figure 1. *Average Oversampling* $\overline{n} - n^*$ *and the Percentage Oversampling* $\omega_1, \omega_2$, *and* $\omega_3$ from (5.1) *versus* $d$ *for Procedures* $\mathcal{P}_1$-$\mathcal{P}_3$ *when* $a = 2, m = 10$

We may quantify the percentage oversampling associated with the procedure $\mathcal{P}_i$, as follows:

$$\omega_i = 100(\overline{n} - n^*)/n^*, i = 1, 2, 3. \tag{5.1}$$

Figure 1.a plots the values of $\overline{n} - n^*$ against $d$ for the three procedures whereas Figure 1.b shows the plots of $\omega_i$ against $d$ when $a = 2$ and $m = 10$. Similar performances

were noted for other values of $a$ and $m$ as long as $2m < n_1^*$, but $\mathcal{P}_3$ tended to oversample more compared with the other two procedures when $m < n_1^* < 2m$.

Table 4 shows the estimated coverage probabilities of the simulated fixed-with confidence intervals obtained via all three procedures under consideration. Overall, the estimated coverage probabilities appear higher than the preset confidence coefficient 0.95 regardless of which procedure is implemented from among $\mathcal{P}_1$-$\mathcal{P}_3$. We noticed another interesting feature. The estimated coverage probabilities were hardly influenced by the choice of the value of $a$ and $m(< \frac{1}{2}n_1^*)$, and this was seen across the board for all three procedures. Table 4 expresses the same sentiment even though we have included the values $a = 1, 2$ only. The estimated standard errors did not appreciably influence our overall assessment of the big picture and hence the standard error entries are not provided in Figure 1 or Table 4.

**Table 4.** Coverage Probabilities for Given Values of $d$, $a$ and $m = 10$

| | Coverage Probability | | | | | |
|---|---|---|---|---|---|---|
| | Stein-type: $\mathcal{P}_1$ | | Chapman: $\mathcal{P}_2$ | | Alternative: $\mathcal{P}_3$ | |
| $d$ | $a = 1$ | $a = 2$ | $a = 1$ | $a = 2$ | $a = 1$ | $a = 2$ |
| 1.0 | 0.948 | 0.953 | 0.951 | 0.952 | 0.959 | 0.955 |
| 0.9 | 0.954 | 0.955 | 0.950 | 0.954 | 0.956 | 0.951 |
| 0.8 | 0.948 | 0.952 | 0.953 | 0.956 | 0.952 | 0.953 |
| 0.7 | 0.950 | 0.951 | 0.950 | 0.951 | 0.950 | 0.950 |
| 0.6 | 0.952 | 0.950 | 0.950 | 0.952 | 0.950 | 0.950 |
| 0.5 | 0.952 | 0.948 | 0.950 | 0.948 | 0.952 | 0.952 |
| 0.4 | 0.950 | 0.949 | 0.950 | 0.949 | 0.951 | 0.951 |
| 0.3 | 0.950 | 0.948 | 0.947 | 0.946 | 0.948 | 0.951 |
| 0.2 | 0.951 | 0.950 | 0.954 | 0.951 | 0.947 | 0.949 |
| 0.1 | 0.950 | 0.948 | 0.951 | 0.945 | 0.951 | 0.951 |

Let us now recall the discrepancy ratios $\tau_1, \tau_2$ from (3.5) and $\tau_3$ from (4.4). We may reiterate that these discrepancy ratios represented *approximately* the ratio of the average total sample size and $n^*$ under the estimation methodology $\mathcal{P}_i, i = 1, 2, 3$. But, the level of accuracy of this approximation is not fully known. Hence, we decided to look at the estimated values of $\rho_i \equiv E_\sigma[N_1 + N_2 \mid \mathcal{P}_i]/n^*, i = 1, 2, 3$, namely,

$$\widehat{\rho}_i = \text{estimate of } E_\sigma[N_1 + N_2 \mid \mathcal{P}_i]/n^*, i = 1, 2, 3. \tag{5.2}$$

Further, let $s(\widehat{\rho}_i)$ be the estimated standard error of $\widehat{\rho}_i$, obtained from 10,000 independent replications. Even though we ran simulations with

$$a = 0.25, 0.50, 0.75, 1.0, 2.0 \text{ and } m = 5, 10, 15, 25, 30,$$

in Table 5 we summarize our findings with regard to $\widehat{\rho}_i$ and $s(\widehat{\rho}_i)$ only when $m = 10$ and $a = 1, 2$. The performances in general were very similar in the other cases. The estimated values $\widehat{\rho}_i$ came very close to its approximate theoretical values $\tau_i$ with relatively small estimated standard errors $s(\widehat{\rho}_i)$ under each estimation methodology $\mathcal{P}_i, i = 1, 2, 3$.

From Table 5, one immediately finds that $\widehat{\rho}_3$ associated with the new alternative methodology $\mathcal{P}_3$ from (4.1) is generally the lowest across the board. It is also clear from Table 5 that the standard error of $\widehat{\rho}_3$ is smaller than those associated with $\mathcal{P}_1$ or $\mathcal{P}_2$. We also note that the estimated standard errors $s(\widehat{\rho}_i)$ appear to stay practically unaffected by the proposed confidence interval's preassigned half-width $d$ under any of the three procedures $\mathcal{P}_1$-$\mathcal{P}_3$.

**Table 5.** Estimated Efficiency Ratio $\widehat{\rho}_i$ from (5.2) and $s(\widehat{\rho}_i)$ Under the Estimation Methodology $\mathcal{P}_i, i = 1, 2, 3$ and $m = 10$

| | | Stein-type: $\mathcal{P}_1$ | | Chapman: $\mathcal{P}_2$ | | Alternative: $\mathcal{P}_3$ | |
|---|---|---|---|---|---|---|---|
| $d$ | $n^*$ | $\widehat{\rho}_1$ | $s(\widehat{\rho}_1)$ | $\widehat{\rho}_2$ | $s(\widehat{\rho}_2)$ | $\widehat{\rho}_3$ | $s(\widehat{\rho}_3)$ |
| | | $\tau_1 = 1.1489$ | | $\tau_2 = 1.3166$ | | $\tau_3 = 1.1403$ | |
| | | | | **a = 1** | | | |
| 1.0 | 61.5 | 1.1601 | 0.0038 | 1.3377 | 0.0044 | 1.1624 | 0.0036 |
| 0.9 | 75.9 | 1.1573 | 0.0038 | 1.3283 | 0.0044 | 1.1533 | 0.0036 |
| 0.8 | 96.0 | 1.1557 | 0.0039 | 1.3313 | 0.0044 | 1.1474 | 0.0037 |
| 0.7 | 125.4 | 1.1603 | 0.0039 | 1.3238 | 0.0044 | 1.1490 | 0.0037 |
| 0.6 | 170.7 | 1.1612 | 0.0039 | 1.3210 | 0.0044 | 1.1487 | 0.0037 |
| 0.5 | 245.9 | 1.1577 | 0.0039 | 1.3224 | 0.0044 | 1.1476 | 0.0037 |
| 0.4 | 384.1 | 1.1498 | 0.0039 | 1.3146 | 0.0043 | 1.1420 | 0.0037 |
| 0.3 | 682.9 | 1.1553 | 0.0038 | 1.3231 | 0.0044 | 1.1422 | 0.0037 |
| 0.2 | 1536.6 | 1.1498 | 0.0038 | 1.3172 | 0.0044 | 1.1450 | 0.0037 |
| 0.1 | 6146.3 | 1.1544 | 0.0038 | 1.3161 | 0.0044 | 1.1380 | 0.0037 |
| | | | | **a = 2** | | | |
| 1.0 | 153.7 | 1.1692 | 0.0039 | 1.3239 | 0.0051 | 1.1628 | 0.0036 |
| 0.9 | 189.7 | 1.1703 | 0.0038 | 1.3272 | 0.0052 | 1.1549 | 0.0036 |
| 0.8 | 240.1 | 1.1535 | 0.0038 | 1.3158 | 0.0051 | 1.1452 | 0.0037 |
| 0.7 | 313.6 | 1.1609 | 0.0038 | 1.3124 | 0.0051 | 1.1508 | 0.0037 |
| 0.6 | 426.8 | 1.1491 | 0.0039 | 1.3230 | 0.0051 | 1.1479 | 0.0037 |
| 0.5 | 614.6 | 1.1499 | 0.0038 | 1.3144 | 0.0051 | 1.1380 | 0.0037 |
| 0.4 | 960.4 | 1.1576 | 0.0038 | 1.3243 | 0.0051 | 1.1415 | 0.0037 |
| 0.3 | 1707.3 | 1.1481 | 0.0039 | 1.3119 | 0.0051 | 1.1442 | 0.0037 |
| 0.2 | 3841.5 | 1.1470 | 0.0038 | 1.3149 | 0.0052 | 1.1377 | 0.0037 |
| 0.1 | 15365.8 | 1.1439 | 0.0039 | 1.3194 | 0.0051 | 1.1382 | 0.0037 |

Figure 2 shows the simulated values of $\widehat{\rho}_1$ and $\widehat{\rho}_2$ and $\widehat{\rho}_3$. Figure 2.a shows the discrepancies among three procedures when $a = 2$. From this figure, it is clear that the Chapman procedure has the highest $\widehat{\rho}$ value whereas the alternative procedure (4.1) has the lowest $\widehat{\rho}$ value. Figure 2.b shows the discrepancy ratio estimates only for the alternative procedure (4.1) when $a = 0.5, 1, 2$. This figure shows that the discrepancy ratio estimates for the alternative procedure $\mathcal{P}_3$ remain nearly independent of the choice of $a$ and that they approach $\tau_3$ as $d$ becomes small.

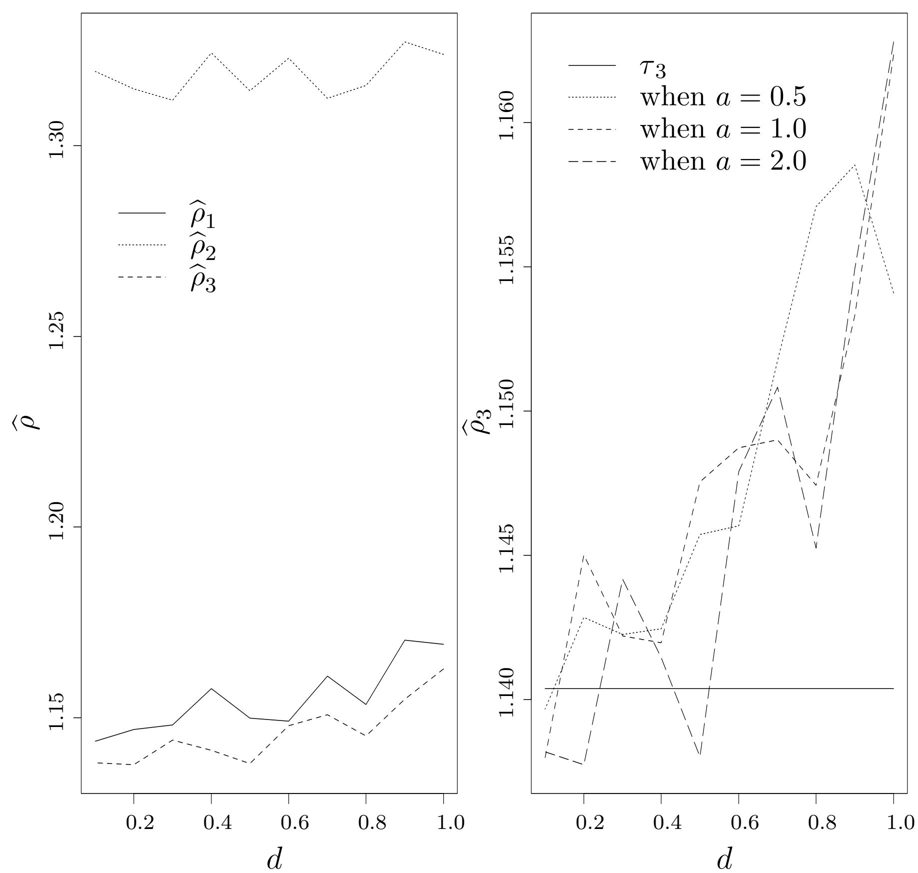


Fig 2.a Fig 2.b

Figure 2. *Estimated Efficiency Ratios from Simulations versus $d$ for Procedures $\mathcal{P}_1$-$\mathcal{P}_3$ when $a = 2$.*

However, it seems evident to us that the new alternative procedure $\mathcal{P}_3$ performed better than the other two as exhibited by the following broad set of observations:

1. The procedure $\mathcal{P}_3$ came up with the minimum oversampling and the minimum percentage oversampling rate.

2. The procedure $\mathcal{P}_3$ was operationally simpler than the other two procedures since we started with the pilot observations only from one population.

3. The procedure $\mathcal{P}_3$ maintained the coverage probabilities as tightly around the preset confidence coefficient $1 - \alpha$ as did the other two procedures.

# 6. An Example with Real Data

The setup for this example is borrowed from Mukhopadhyay et al. (2004). A designed experiment was developed, implemented, reported, and analyzed for a problem involving multiple comparisons in Mukhopadhyay et al. (2004). For detailed explanations and validations regarding the dataset under consideration, one is referred to the original source.

We only give an example of a two-sample problem although Mukhopadhyay et al. (2004) reported datasets for comparing three varieties of marigold seeds. In a similar setup, we may be interested in comparing marigold seeds from a local variety (#1) and a hybrid variety (#2) supplied by a vendor by observing the number of days ($X$) each variety took from planting seeds to reach a point when the first bud appeared on a plant. The response times (namely, $X_1$ and $X_2$ in days) from seeding to first-time budding for each variety was assumed normal (see Mukhopadhyay et al., 2004) with the averages $\mu_1$ and $\mu_2$ days for variety #1 and #2 respectively.

From what was reported in Mukhopadhyay et al. (2004), it seems reasonable to postulate that $a = 1.5$ and we may like to construct a 95% confidence interval for $\Delta = \mu_2 - \mu_1$ with a fixed-width of 2 days, that is, $d = 1.0$ day. The response variables were observed within one-half day of possible first-time budding because inspections of the experimental units could be carried out at 6:00 AM and again at 6:00 PM. As soon as a bud was discovered on a plant, it was moved to an adjoining greenhouse. We fixed $m = 15$, the pilot sample size.

From what has been explained in Sections 2 through 5, it should be clear that only the Stein-type procedure $\mathcal{P}_1$ from (2.1) and the new alternative procedure $\mathcal{P}_3$ from (4.1) should be included in this present discourse. We deliberately bypass the Chapman procedure (3.1)-(3.2). We do so because we had already reported in Sections 4 and 5 that it did not perform as well as the other two procedures during the large-scale computer simulations that we had undertaken.

**The Stein-Type Procedure (2.1):**

Fifteen pilot observations were drawn randomly from the respective first-stage samples that were reported in Table 1 of Mukhopadhyay et al. (2004). We recorded the following pilot observations:

$$\begin{aligned}&\textbf{Variety \#1:}\\&29.5, 28.0, 30.5, 28.5, 30.0, 25.5, 28.5, 32.5, 31.5, 28.5, 30.0,\\&27.5, 30.0, 29.5, 28.5 \Rightarrow \overline{x}_{1,15} = 29.233, S_{1,15} = 1.659\\&\textbf{Variety \#2:}\\&42.0, 41.5, 38.5, 38.0, 33.0, 31.0, 45.0, 36.0, 31.0, 30.5, 37.5,\\&35.0, 35.0, 38.5, 36.0 \Rightarrow \overline{x}_{2,15} = 36.57, S_{2,15} = 4.25\end{aligned} \tag{6.1}$$

So, $V_{15}^2 \equiv \frac{1}{2}\left(S_{1,15}^2 + a^{-2}S_{2,15}^2\right) = \frac{1}{2}\left((1.659)^2 + (1.5)^{-2}(4.25)^2\right) \approx 5.39$. With 28 degrees of freedom, we found $q_{15} = 2.0484$ so that we had

$$2q_{15}^2V_{15}^2/d^2 = 2(2.0484)^2(5.39)/(1.0)^2 = 45.232.$$

From (2.1), we obtained $n_1 = 46$ and $n_2 = 104$. Thus, in the second stage, we recorded 31 more observations from variety #1 and 89 more observations from variety #2. We obtained these additional observations randomly from the second-stage data that were reported in Table 2 of Mukhopadhyay et al. (2004). From the full datasets so obtained, that is after combining the observations from both stages of sampling for each variety, we found $\overline{x}_{1,46} = 29.466, \overline{x}_{2,104} = 35.394$.

Now, from (2.2), we may write down the 95% fixed-width confidence interval for $\Delta = \mu_2 - \mu_1$ as follows:

$$[\overline{x}_{2,104} - \overline{x}_{1,46} \pm d] = [35.394 - 29.466 \pm 1] = [4.928, 6.928]. \tag{6.2}$$

In other words, the average budding time for variety #2 seeds exceeded that for variety #1 seeds by nearly 5 days to 7 days! ▲

**The Alternative Procedure (4.1):**

Now, thirty pilot observations were drawn randomly from the first-stage samples for variety #1 seeds only that were reported in Table 1 of Mukhopadhyay et al. (2004). We opted for sampling from variety #1 because its responses had a smaller variance. We recorded the following pilot observations:

$$\begin{aligned}&\textbf{Variety \#1:}\\&27.5, 27.5, 30.0, 32.5, 28.5, 31.5, 31.5, 32.5, 34.0,\\&33.5, 28.5, 32.0, 32.5, 30.5, 28.5, 31.5, 29.5, 32.5,\\&34.0, 31.0, 28.0, 29.0, 30.5, 33.0, 30.0, 31.5, 31.0,\\&30.0, 31.5, 28.5 \Rightarrow \overline{x}_{1,30} = 30.75, S_{1,30} = 1.911\end{aligned} \tag{6.3}$$

With 29 degrees of freedom, we found $b_{15} = 2.0452$ so that we had

$$2b_{15}^2S_{1,30}^2/d^2 = 2(2.0452)^2(1.911)^2/(1.0)^2 = 30.551.$$

From (4.1), we obtained $n_1 = 31$ and $n_2 = 70$. Thus, in the second stage, we recorded 1 more observation from variety #1, and then 70 observations from variety #2, all in one single batch. We obtained these observations randomly from the second-stage data that were reported in Table 2 of Mukhopadhyay et al. (2004). From the full datasets so obtained, that is after combining the observations from both stages of sampling for variety #1, we found $\overline{x}_{1,31} = 30.661$, and the observations for variety #2 gave $\overline{x}_{2,70} = 35.071, S^2_{2,70} = 15.936$.

Now, from (4.2), we may write down the 95% fixed-width confidence interval for $\Delta = \mu_2 - \mu_1$ as follows:

$$[\overline{x}_{2,70} - \overline{x}_{1,31} \pm d] = [35.071 - 30.661 \pm 1] = [3.41, 5.41]. \tag{6.4}$$

In other words, the average budding time for variety #2 seeds exceeded that for variety #1 seeds by nearly 3.5 days to 5.5 days! ▲

**Concluding Thoughts:**

In the implementation of the alternative procedure, one could reuse the 15 pilot observations from variety #1 found in (6.1) and then augment this with another set of randomly chosen 15 observations from Table 1 in Mukhopadhyay et al. (2004). After we did just that, we had observed $\overline{x}_{1,30} = 29.75, S_{1,30} = 1.804$, thereby obtaining $n_1 = 30$ and $n_2 = 68$. Thus, in the second stage, we did not require any more observations from variety #1. Then, we recorded 68 observations from variety #2, all in one single batch, randomly from the second-stage data that were reported in Table 2 of Mukhopadhyay et al. (2004). From the full datasets so obtained, we had found $\overline{x}_{1,30} = 29.75$, and the observations for variety #2 gave $\overline{x}_{2,68} = 35.051, S^2_{2,68} = 18.499$. Thus, the 95% fixed-width confidence interval for $\Delta = \mu_2 - \mu_1$ came out to be

$$[\overline{x}_{2,68} - \overline{x}_{1,30} \pm d] = [35.051 - 29.75 \pm 1] = [4.301, 6.301]. \tag{6.5}$$

The point to be noted here is that the answer found in (6.5) is not drastically different from what was found in (6.2) or (6.4).

But, before we wrap up this example, let us add that the sample mean and the sample variance for the varieties #1 and #2 for the full datasets found in Mukhopadhyay et al. (2004) were respectively as follows:

**Variety #1:** 30.068 and 4.41
**Variety #2:** 35.135 and 16.265

The confidence intervals from (6.4) and (6.5) for the parameter $\Delta = \mu_2 - \mu_1$ which were obtained by using our alternative methodology (4.1) required 101 and 98 total number of observations respectively compared with a total 150 observations required by the Stein-type procedure (2.1). Practical merits of the alternative methodology (4.1) should be rather obvious at this point.

# References

1. Aoshima, M. (2001). Sample size determination for multiple comparisons with components of a linear function of mean vectors. *Commun. Statist. Theor. Meth.* 30:1773-1788.

2. Aoshima, M. and Mukhopadhyay, N. (2002). Two-stage estimation of a linear function of normal means with second-order approximations. *Sequential Anal.* 21:109-144.

3. Bechhofer, R. E., Santner, T. J., and Goldsman, D. M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons.* New York: Wiley.

4. Chapman, D. G. (1950). Some two-sample tests. *Ann. Math. Statist.* 21:601-606.

5. Dantzig, G. B. (1940). On the non-existence of tests of Student's hypothesis having power functions independent of $\sigma$. *Ann. Math. Statist.* 11:186-192.

6. de Silva, B. M. and Waikar, V. B. (2006). A sequential approach to Behrens-Fisher problem. *Sequential Anal., Milton Sobel Memorial Issue, Part II*, 25: 311-326.

7. Ghosh, B. K. (1975a). A two-stage procedure for the Behrens-Fisher problem. *J. Amer. Statist. Assoc.* 70:457-462.

8. Ghosh, B. K. (1975b). On the distribution of the difference of two t-variables. *J. Amer. Statist. Assoc.* 70:463-467.

9. Ghosh, M., Mukhopadhyay, N., and Sen, P. K. (1997). *Sequential Estimation.* New York: Wiley.

10. Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedure.* New York: Wiley.

11. Hsu, J. C. (1996). *Multiple Comparisons.* New York: Chapman & Hall.

12. Liu, W. (1995). Fixed-width simultaneous confidence intervals for all pairwise comparisons. *Comp. Statist. Data Anal.* 20:35-44.

13. Mukhopadhyay, N. and de Silva, B. M. (1997). Multistage fixed-width confidence intervals in the two-sample problem: The normal case. *J. Statist. Res.* 31:1-20.

14. Mukhopadhyay, N. and Solanky, T. K. S. (1994). *Multistage Selection and Ranking Procedures*. New York: Marcel Dekker.

15. Mukhopadhyay, N., Son, M. S., and Ko, Y. C. (2004). A new two-stage sampling design for estimating the maximum average time to flower. *J. Agric. Biol. Environ. Statist.* 9:479-499.

16. Robbins, H., Simons, G., and Starr, N. (1967). A sequential analogue of the Behrens-Fisher problem. *Ann. Math. Statist.* 38:1384-1391.

17. Srivastava, M. S. (1970). On a sequential analogue of the Behrens-Fisher problem. *J. Roy. Statist. Soc., B* 32:144-148.

18. Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* 16:243-258.

19. Stein, C. (1949). Some problems in sequential estimation (abstract). *Econometrica* 17:77-78.

Nitis Mukhopadhyay
Department of Statistics
University of Connecticut
215 Glenbrook Road, U 4120
Storrs, CT 06269-4120, U.S.A.
E-mail: nitis.mukhopadhyay@uconn.edu

Basil M. de Silva
School of Mathematical & Geospatial Sciences
RMIT University, GPO Box 2476V
Melbourne, Vic 3001, Australia
E-mail: desilva@rmit.edu.au

Vasant B. Waikar
Department of Mathematics & Statistics
Miami University
Oxforda, Ohio 45056, U.S.A.
E-mail: waikarvb@muohio.edu