The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 297-312

Estimating the Difference of Two Normal Means

M. Aimahmeed[1], M. S. Son[2], H. I. Hamdy[3]

Abstract

A three stage sampling procedure designed to estimate the difference between two normal means is proposed and evaluated within a unified decision-theoretic framework. Both point and fixed-width confidence interval estimation are combined in a single decision rule to make full use of the available data. Adjustments to previous solutions focusing on only one of the latter objectives are indicated. The sensitivity of the confidence interval for detecting shifts in the true mean difference is also investigated. Numerical and simulation studies are presented to supplement the theoretical results.

1. Introduction

The literature devoted to estimating the difference between tow normal means (the well known Behrens - Fisher problem) is immense and there does not appear to be a satisfactory solution within the classical theory of statistical inference, see for example Barnard (1984) and Nel and Van de Merwe (1986). There is also some controversy within the Bayesian framework for comparing two normal means when the variances are completely unspecified, as reported by Box and Tiao (1973) and Patil (1964). Sequential analytic contributions to the problem have also been limited, since they were developed based on Stein's (1945) two stage group sampling technique or though one-by-one sequential sampling with a switching decision rule to determine which population to sample next. Our proposed procedure for the Behrens - Fisher problem is inspired by Hall's (1981) triple sampling scheme. We refer to it as an integrated approach since it incorporates point and fixed width confidence interval estimation, as well as a sensitivity analysis.

Here, we begin with a brief description of Hall's elegant three-stage procedure to construct a fixed width confidence interval for a single normal mean μ when the variance σ^2 is unknown, and we summarize its main asymptotic properties. In addition, we extend his results to obtain the mean of a continuously differentiable and bounded function of the stopping time. These results for one population will be applied in the context of the two-population case in

1) Kuwait University
2) University of Vermont
3) Kuwait Univerity & University of Vermont

subsequent sections. In particular, in Section 2 we discuss some limitations of previous multistage sampling approaches to the Behrens-Fisher problem. We also motivate our proposed approach based on more integrated objectives and suggest an adjustment which combines both point and interval estimation in one decision rule in Section 2.1. In Section 3, we outline our proposed triple sampling approach. Moreover, in Section 4 we investigate the sensitivity and the coverage of the fixed width confidence interval for the mean difference, and in Section 5 we assess the asymptotic point estimation regret of the estimates used in the fixed width confidence interval. In Section 6 we supplement our theoretical asymptotic results with some simulations similar to those performed by Hall(1981).

Assume we have observed a random sample of size $t_1, t_2, \cdots, t_r$, for $r \geq 2$ from a normal population with unknown mean μ and unknown variance σ^2. The sample mean $\overline{T_r} = r^{-1}\sum t_i$ and the sample variance $S_r^2 = (r-1)^{-1}\sum(t_i - \overline{T_r})^2$ are the UMVUE's for μ and σ^2, respectively.

Assume further that for two predetermined constants, $d > 0$ and $0 < a < 1$, we wish to construct a fixed width confidence interval for μ whose width is $2d$ and whose coverage probability is at least $100(1-a)\%$. The required interval is assumed to bc of the form $I_r = \overline{T_r} \pm d$ It is not hard to show that the optimal sample size required to establish such an interval to satisfy the above conditions is $r^* = a^2\sigma^2/d^2$, where a is the $100(1-a/2)\%$ percentile of $N(0,1)$. Since r^* depends on the unknown nuisance parameter σ^2, the fixed sample size procedure is impractical. Instead, Hall (1981) utilized the idea of group sampling in three stages to estimate μ via estimation of r^*.

Specifically, he proposed three consecutive stages to complete the estimation process, namely the pilot study, the main study, and the fine tuning study. During the pilot study phase an initial sample of size $r_0 \geq 2$ is selected to compute $\overline{T}_{r_0}$ and $S^2_{r_0}$ to initiate the process. Consequently, the second stage sample size is determined according to the decision rule:

$$R_1 = \max\{r_0, [\gamma a^2 d^{-2} S^2_{r_0}] + 1\} \tag{1.1}$$

Where $0 < \gamma < 1$ is a design factor which represents the proportion of r^* to be estimated by the second stage information, and $[\cdot]$ is the largest integer function. If sampling is to be continued in the main study phase, a second random sample of size $R_1 - r_0$ is selected and combined with the previously collected r_0 observations to update the corresponding sample measures based on R_1 observations. Entering the fine tuning phase with $\overline{T}_{R_1}$ and $S^2_{R_1}$, we determine the third stage sample size by the decision rule:

$$R = \max\{R_1, [a^2 d^{-2} S^2_{R_1}] + 1\} \tag{1.2}$$

If necessary, a third batch of size $R - R_1$ is randomly selected and augmented with the

previously collected R_1 observations to compute $\overline{T}_R$ and the interval $I_R = \overline{T}_R \pm d$ for μ.

We highlight several desirable features of the triple sampling estimation procedure. First, the sampling scheme is designed in such a way that the sample sizes can be closely monitored and accordingly revised by the next decision rule to reduce the possibility of over sampling. Second, the idea of group sampling in each stage expedites the process so decisions can be reached quickly, an obvious advantage when time and money are of concern. Third, by estimating only a fraction γ of r^* during the second stage, we force the procedure to enter the fine tuning phase to mend any possible deficiency, particularly when r^* is large.

Unlike Stein's (1945) two-stage sampling technique, the triple sampling procedure is more robust against the possibility that the pilot sample happens to be much smaller than the optimal sample size had σ^2 been known. Clearly, the procedure is designed to combine all the asymptotic properties of the Anscombe (1953), Chow and Robbins (1965) one-by-one sequential sampling and the operational savings made possible by applying Stein's method of group sampling.

Hall's (1981) results concerning the asymptotic characteristics of the triple sampling procedure can be summarized in the following theorem.

Theorem 1. In the context of the triple sampling procedure (1.1) - (1.2), we have as $r_0 \to \infty$

(i) $E(R) = r^* + 1/2 - 2\gamma^{-1} + o(1)$

(ii) $Var(R) = 2\gamma^{-1} r^* + o(d^{-2})$

(iii) $E|(R - r^*)^3| = o(d^{-4})$

(iv) $P(\mu \in I_R) = (1 - a) - a\phi(a)(a^2 - \gamma + 5)(2\gamma r^*)^{-1} + o(d^2)$.

Proof of Theorem 1. I can be found in Hall (1981), Hamdy (1988) and Hamdy et al. (1996), thus we omit any further details.

In Theorem 2 we extend the results in Theorem 1 to obtain a second order asymptotic expansion of the expectation of a continuously differentiable and bounded function g of R.

Theorem 2. For R defined by (1.2), let g be a continuously differentiable real valued function in a neighborhood around r^* such that $\text{Sup}_{r \geq r_0} |g'''(r)| = O(g'''(r^*))$, then

$$E(g(R)) = g(r^*) + (2\gamma)^{-1}\{(\gamma - 4)g'(r^*) + 2r^* g''(r^*)\} + o(d^{-4}|g'''(r^*)|)$$

Proof of Theorem 2. A Taylor expansion of $g(R)$ around r^* yields

$$g(R) = g(r^*) + (R - r^*)g'(r^*) + \frac{1}{2}(R - r^*)^2 g''(r^*) + F_R, \tag{1.3}$$

where F_R is the remainder term in Taylor's expansion. Taking the expectation throughout (1.3) and making use of (i) and (ii) of Theorem 1 we obtain

$$E(g(R)) = g(r^*) + (2\gamma)^{-1}\{(\gamma-4)g'(r^*) + 2r^*g''(r^*)\} + E(F_R).$$

It remains to evaluate $E(F_R)$ as follows, $E(F_R) \leq \frac{1}{6} E|(R-r^*)^3 g'''(v)|$, where v is a random variable lying between R and r^*. Thus,

$$E(F_R) \leq \frac{1}{6} \text{Sup}_{r \geq r_0} |g'''(r)| E|(R-r)^3| = o(d^{-4}|g'''(r^*)|),$$

where, we have used (iii) of Theorem 1 and the assumption that g''' is bounded.

2. Multistage procedures for the difference between two normal means

Sequential and two stage procedures for the Behrens-Fisher problem have been discussed by many authors over the years. A brief list of key contributions includes Chapman (1950), Ghosh (1975, a, b), Ghosh and Mukhopadhyay (1980), Ghurye and Robbins (1954), Hayre (1983), Louis (1975), O'Neill and Rohatgi (1973) Robbins (1959), Robbins, Simons and Starr (1967), Scheffe (1943, 1970), and Srivastava (1970), as well as the more recent work of Eisele (1990)and Hamdy et al. (1996).

Let $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ be two independent sequences of independent random variables having normal distributions with unknown mean μ_i and unknown variances σ_i^2 for $i=1,2$. The initial focus will be on estimating the difference $\delta=(\mu_1-\mu_2)$ by a fixed width confidence interval such that the coverage probability is at least the nominal value $100(1-\alpha)\%$.

Having observed $X_1, X_2, \cdots, X_n, n \geq 2$ form the first population and $Y_1, Y_2, \cdots, Y_m, m \geq 2$ from the second population, we employ $\overline{X}_n, \overline{Y}_m, S_n^2, S_m^2$, the usual estimates for $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, respectively, to construct the interval $I_{n,m} = (\overline{X}_n - \overline{Y}_m \pm d)$ for δ.

To guarantee $100(1-\alpha)\%$ coverage probability that the fixed width confidence interval captures δ, the condition $P(\delta \in I_{n,m}) \geq (1-\alpha)$ should be satisfied. This requirement implies that $2\Phi(d(\sigma_1^2 n^{-1} + \sigma_2^2 m^{-1})^{-1/2}) - 1 \geq 2\Phi(a) - 1$. Therefore the optimal sample sizes must satisfy the following inequality

$$\sigma_1^2 n^{-1} + \sigma_2^2 m^{-1} \leq d^2 a^{-2}, \tag{2.1}$$

Minimization of the total sample size ($(n+m)$ subject to (2.1) to explicitly determine the optimal sample sizes n^* (in terms of σ_1^2 and independent of σ_2^2) and m^*(in terms of σ_2^2

and independent of σ_1^2) has remained a troublesome problem. Several "solutions" have been proposed to define one or more convenient decision rules to achieve the goal of constructing the required fixed width interval. Here, we present two approaches that may be considered fundamental in this context.

First, Chapman (1950) and Ghosh (1975, a, b) suggested two stage rules based on the difference between two independent Student variables. Such procedures ignore the sample variance of the second stage. In addition, it has been shown that the two stage sampling scheme, in general, leads to substantial oversampling if the pilot sample size is chosen much smaller than the optimal sample size, as noted by Ghosh and Mukhopadhyay (1981).

A second approach was proposed by Robbins, Simons and Starr (1967). Where, the allocation of $n^*$and m^* was carried out by minimizing the total sample size (n^*+m^*) under the restriction in (2.1) to obtain

$$n^* = a^2 d^{-2}(\sigma_1^2+\sigma_1\sigma_2),\quad m^* = a^2 d^{-2}(\sigma_2^2+\sigma_1\sigma_2),\text{ and }\quad n^* m^{*-1} = \sigma_1^2\sigma_2^{-2} \tag{2.2}$$

They proposed one-by-one sequential rules that mimic (2.2) to determine the random variables N and M. Their procedure consists of (i) a switching scheme that is used to determine whether to take the next observation on X or Y at each stage (ii) a stopping rule which determines N and M.

It is clear that the procedure is difficult to apply in real life situations. Also, although the original populations are independent, the way sampling is performed creates dependence between N and M. Thus, it would be mathematically inconvenient to thoroughly investigate the asymptotic characteristics of the proposed procedure. Robbins, Simons and Starr (1967) reported, only, limiting results concerning, N, M, and the coverage probability. More recently, Eisele (1990) provided a one-by-one sequential "solution" to the problem using a randomization criterion. However, his decision rules are practically the same as those given above. Ghosh and Mukhopadhyay (1980) treated the point estimation case for the same problem under a squared error loss function with a linear sampling cost. Their proposed one-by-one sequential sampling procedures are also similar to those given above.

Since the foundation of the sequential decision approach by Wald (1947), sequential analysts have adopted two main methodologies to tackle estimation problems: (i) point estimation, where a specific cost function is assumed to assess the risk, and (ii) fixed width confidence interval estimation to attain a given nominal coverage value. Usually, these two methodologies are applied separately in a given estimation problem. No attempt has been made (to the best of our knowledge) to design sampling procedures that combine both point and interval estimation in a unified decision theory framework. For example, assuming that a multistage fixed width confidence interval estimation procedure is conducted to estimate the normal mean μ , would it be appropriate to use the same data to provide a point estimate for μ? If so, how good is such an estimate in reference to the asymptotic risk efficiency and regret as

defined by Robbins (1959), or vice versa? Some answers to such questions are provided in later sections.

Our approach incorporates four main objectives in addition to satisfying the requirements for fixed width confidence intervals. These objectives are to define a sampling technique (i) that makes maximal use of all the sampling information; (ii) that is simple to apply and also saves time in reaching decisions; (iii) that maintains the independence between the random variables N and M so rigorous mathematical analysis can be performed to evaluate the performance of the procedure; (iv) that can be used to tackle both point as well as confidence interval estimation problems.

The key to our subsequent developments lies in the explicit determination of the optimal sample sizes n^* and m^* from (2.1). To achieve the above stated objectives we impose an additional (but reasonable) restriction while sampling from each population. The restriction involves forcing the cost functions that control the risks associated with estimation of μ_i by the corresponding sample measures from the i^{th} population for $i=1,2$, to share common weighting constant. Specifically, we assume that the costs incurred in estimating μ_1 and μ_2 are respectively,

$$L_n(A)=A^2\sigma_1^2(\overline{X}_n-\mu_1)^2+n \tag{2.3}$$

and

$$L_m(V)=A^2\sigma_2^2(\overline{Y}_m-\mu_2)^2+m \tag{2.4}$$

Cost functions similar to (2.3) and (2.4) were considered by Chow and Martinsek (1982), Chow and Yu (1981), Martinsek (1988), and Woodroofe (1985, 1987). In decision theory the constant A is treated as a "known" weight whose assignment is usually left to the decision maker. None of the existing published literature discusses how the decision maker can determine A. In this study, however, we specify A precisely.

The risks associated with (2.3) and (2.4) are given by

$$E(L_n(A))=A^2\sigma_1^2E(\overline{X}_n-\mu_1)^2+n=A^2\sigma_1^4n^{-1}+n \tag{2.5}$$

$$E(L_m(A))=A^2\sigma_2^2E(\overline{X}_m-\mu_2)^2+m=A^2\sigma_2^4m^{-1}+m \tag{2.6}$$

Minimization of the risks in (2.5) and (2.6) provides the optimal sample sizes are $n^*=A\sigma_1^2$ and $m^*=A\sigma_2^2$. Hence, the associated optimal risks are $E(L_{n^*}(A))=2n^*$ and $E(L_{m^*}(A))=2m^*$. If n^* and m^* are used to construct a fixed $2d$ width confidence interval for δ with at least $100(1-\alpha)\%$ coverage, the constant A should be chosen such that (2.1) is satisfied. Therefore, we obtain $A\geq 2a^2d^{-2}$. Obviously, as $d\to 0, A\to\infty$. In other words, as the width of the interval gets smaller we expect larger estimation risk.

Thus the optimal sample sizes are respectively,

$$n^* = 2a^2 d^{-2}\sigma_1^2 \text{ and } m^* = 2d^{-2}\sigma_2^2 \tag{2.7}$$

It is clear that n^* and m^* provide the minimum risks in estimating μ_1 and μ_2 by the corresponding sample measures $\overline{X}_{n\cdot}$ and $\overline{Y}_{m\cdot}$ within the quadratic loss function structures and collectively they ensure at least a $100(1-\alpha)\%$ fixed width confidence interval for δ.

2.1 Adjusting cost Functions to construct confidence intervals and point estimates

Ghosh and Mukhopadhyay (1980) assumed the following cost function to obtain a point estimate for δ,

$$L_{n,m}(A) = A(\overline{X}_n - \overline{Y}_m - \delta)^2 + n + m, \tag{2.8}$$

where $A > 0$ is a "known" weight constant. (Here we assume that the cost of sampling is a unit cost.). They proposed a one-by-one sequential procedure with a switching rule and provide $(\overline{X}_N - \overline{Y}_M)$ as a point estimate for δ. The question is, how might one adjust their cost function to be able to use the available data to construct a fixed $2d$ width confidence interval for δ with at least $100(1-\alpha)\%$ coverage? This can be accomplished easily by linking the constant A that appears in (2.8) with the confidence interval requirements. It follows from (2.8) that the risk is

$$E(L_{n,m}(A)) = A(\sigma_1^2 n^{-1} + \sigma_2^2 m^{-1})^2 + n + m$$

Hence, the optimal sample sizes are

$$n^* = \sqrt{A}\sigma_1 \text{ and } m^* = \sqrt{A}\sigma_2 \tag{2.9}$$

To establish the required confidence interval, n^* and m^* in (2.9) should satisfy (2.1). Therefore we conclude that A should be chosen such that $A \geq \sigma^4(V)d^{-4}$. It follows that the optimal sample sizes are $n^* = a^2(\sigma_1^2 + \sigma_1\sigma_2)d^{-2}$ and $m^* = a^2(V)d^{-2}$, which are the same optimal sample sizes used in Robbins, Simons and Starr (1967).

Conversely, Robbins, Simons and Starr (1967) could have used their one-by-one sequential procedure for a fixed width confidence interval to provide a point estimate for δ with squared error loss function and a linear sampling cost. Their loss function would have had to have the form

$$L_{n,m}(A) = a^4 d^{-4}(\sigma_1 + \sigma_2)^2(\overline{X}_n - \overline{Y}_m - \delta)^2 + n + m,$$

Thus far we have shown that both the point and fixed width confidence interval estimation methodologies can be treated by one decision rule to achieve both goals. And while it appears from the above determinations that, the constant A could be unknown, contrary to what has been claimed in the sequential point estimation literature, a portion of A is indeed knowable, as we have demonstrated.

In the following section we present a triple sampling procedure following the work of Hall (1981), Mukhopadhyay (1985), Mukhopadhyay et al. (1987), Hamdy et al. (1987), Woodroofe (1987), Hamdy (1988) and Hamdy et al. (1996), to estimate μ_1 and μ_2 though estimation of n^* and m^*. Consequently, we construct the required fixed width confidence interval for δ.

3. Triple sampling schemes

Since n^* and m^* are explicitly defined in (2.7) in terms of their own variances, we implement the triple sampling procedure on each population separately.

In the first population, we start by taking an initial random sample of size $n_0 \geq 2$ and choose an appropriate $\gamma \in (0,1)$ to determine the second stage sample size by the rule

$$N_1 = \max\{n_0, (2\gamma a^2 S^2_{n_0} d^{-2}) + 1\} \tag{3.1}$$

and the final stage sample size from the rule

$$N = \max\{N_1, (2a^2 S^2_{N_1} d^{-2}) + 1\} \tag{3.2}$$

Similarly, we take $m_0 \geq 2$ random samples from the second population and define M_1 and M by the rules

$$M_1 = \max\{m_0, (2\gamma a^2 S^2_{m_0} d^{-2}) + 1\} \tag{3.3}$$

$$M = \max\{M_1, (2a^2 S^2_{M_1} d^{-2}) + 1\} \tag{3.4}$$

Once we determine N by (2.3) and M from (3.4), we propose $\overline{X}_N$ for μ_1 and $\overline{Y}_M$ for μ_2 and the interval $I_{N,M} = (\overline{X}_N - \overline{Y}_M \pm d)$ for δ.

Results like those in (i), (ii) and (iii) of Theorem 1 are still valid in this context. Regarding (iv) of Theorem 1, we devote the following section to examining the sensitivity of the confidence interval $I_{N,M}$ to possible shifts in δ.

4. Measuring the sensitivity of $I_{N,M}$ to departures in δ

Fixed width confidence intervals are designed primarily to control both the precision of estimation d and the coverage probability. Basically, the performance of the interval depends

on three factors; (i) the sample sizes N and M; (ii) the validity of the normal models; (iii) the robustness of the corresponding sample measures of δ, σ_1^2 and σ_2^2.

Now assume that a triple sampling fixed width confidence interval $I_{N,M}$ has been established to monitor δ, for example in quality control for a manufacturing process. It is of interest to explore the ability of $I_{N,M}$ to signify possible shifts in δ when it is thought that these shifts never took place.

Detection of departures is quite important since confidence intervals provide a range of parameter values that would not be rejected if they were hypothesized as null values. To elaborate further on this issue we consider the following two hypotheses. The null hypothesis is $H_0: \delta = \mu_1 - \mu_2$, which asserts that no shifts are recognized by the interval $I_{N,M}$. The alternative hypothesis is $H_0: \delta = \mu_1 - \mu_2 + kd$, which asserts that the parameter value departs away by a magnitude of k measured in united of d. The probability of not detecting such a "k-shift" in the parameter value can be expressed by the corresponding type II error. Let β_t be the type II error associated with triple sampling confidence interval $I_{N,M}$. Then,

$$\begin{aligned}\beta_t &= P\{|\overline{X}_N - \overline{Y}_M - \delta| \le d \mid H_1\} \\ &= E\Phi(d(1-k)(\sigma_1^2 N^{-1} + \sigma_2^2 M^{-1})^{-1/2}) - E\Phi(-d(1+k)(\sigma_1^2 N^{-1} + \sigma_2^2 M^{-1})^{-1/2}),\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$. A bivariate Taylor expansion of $E\Phi(\cdot)$ and (i), (ii) and (iii) of Theorem 1 provide

$$\beta_t = \Phi(a(1-k)) - \Phi(-a(1+k)) - a(16\gamma)^{-1} Q(a, k, \gamma)(n^{*-1} + m^{*-1}) + o(d^2), \tag{4.1}$$

where

$$Q(a, k, \gamma) = (1-k)\phi(a(1-k))\{a^2(1-k)^2 - \gamma + 13\} + (1+k)\phi(-a(1+k))\{a^2(1+k)^2 - \gamma + 13\}.$$

The special case of $k = 0$ in (4.1) yields the coverage probability of $I_{N,M}$ as

$$P(\delta \in I_{N,M}) = (1-a) - a(16\gamma)^{-1}\phi(a)\{a^2 - \gamma + 13\}(n^{*-1} + m^{*-1}) + o(d^2). \tag{4.2}$$

The quantity $\{(a^2 - \gamma + 13)(16\gamma)^{-1}\}$ is known as the cost of ignoring σ_1^2 and σ_2^2. It is clear from (4.2) that the coverage probability is less than the nominal value $(1-a)$. It has been suggested by Simons (1968) and Hall (1981) to add an extra sample to the final stage to correct the coverage. In this case we add a sample of size $[(a^2 - \gamma + 13)(16\gamma)^{-1}]$ to each of the final stage sample sizes.

Similarly, had σ_1^2 and σ_2^2 been known and the fixed sample size confidence interval employed the type II error β_f would be

$$\beta_f = \Phi(a(1-k)) - \Phi(-a(1+k)). \tag{4.3}$$

We performed some computations based on equations (4.1) and (4.3) to provide a feel concerning the ability of the interval $I_{N,M}$ to detect departures from the true parameter δ. We set the pair (n^*, m^*)= (24, 384), (43, 246), (61, 171), (76, 125), (96, 96) by analogy to Hall's (1981) choices. We also set $\alpha = 0.05$ and $\gamma = 0.5$. The departure parameter was permitted to vary over the range $0 \leq k \leq 3$ by 0.1. Table 1 presents both the triple sampling type II error, β_t, and the corresponding fixed sample size type II error, β_f.

Inspecting the results in Table I, we come to the following conclusions. If the departure is within the interval, i.e. $0 \leq k \leq 1$, then the triple sampling type II error $\beta_t \leq \beta_f$, the corresponding fixed sample size type II error. This indicates that the triple sampling fixed width confidence interval is more sensitive to departures occurring within the interval than the fixed sample size confidence interval, especially for the smaller values of (n^*, m^*). However, $\beta_t \rightarrow \beta_f$ for large values of (n^*, m^*). The case of $k=1$ indicates that $\beta_t = \beta_f$ independent of (n^*, m^*). The previous finding is reversed outside the interval $I_{N,M}$, i.e., for $k>1$. In this case the fixed sample size confidence interval becomes more sensitive to shifts occurring outside the interval than the triple sampling confidence intervals. As a general pattern, as k increases both β_t and β_f decrease and approach 0 as $k \rightarrow 3$. Figure I provides the operating characteristic curves for results presented in Table I.

5. The asymptotic regret of estimating μ_1 and μ_2

Robbins (1959) defined the asymptotic regret as ω the measure of the opportunity cost of using multistage techniques instead of the fixed sample size procedures had the nuisance parameter(s) been known. Define the triple sampling risks by

$$E(L_N(A)) = A^2 \sigma_1^2 E(\overline{X}_N - \mu_1)^2 + E(N) \tag{5.1}$$

$$E(L_M(A)) = A^2 \sigma_2^2 E(\overline{Y}_M - \mu_2)^2 + E(M) \tag{5.2}$$

We then make use of the independence between the random variables $N=n, n+1, \cdots$ and $\overline{X}_N$ for all n and $M=m, m+1, \cdots$ and $\overline{Y}_M$ for all m and write

$$E(L_N(A)) = A^2 \sigma_1^4 E(N^{-1}) + E(N)$$

$$E(L_M(A)) = A^2 \sigma_2^4 E(M^{-1}) + E(M)$$

We then apply (i) of Theorem 1 and Theorem 2 with $r(t)=t^{-1}, g'(t)=-t^{-2}, g''(t)=2t^{-3}$ and the remainder term will be of order $o(d^2)$. Thus, we obtain

$$E(L_N(A))=2n^*+2\gamma^{-1}+o(1)$$

and

$$E(L_M(A))=2m^*+2\gamma^{-1}+o(1)$$

Therefore the asymptotic regrets of $\overline{X}_N$ and $\overline{Y}_M$ are as follows

$$\omega_1=E(L_N(A))-E(L_{n^*}(A))=2\gamma^{-1}+o(1) \tag{5.3}$$

and

$$\omega_2=E(L_M(A))-E(L_{m^*}(A))=2\gamma^{-1}+o(1) \tag{5.4}$$

It is clear from (5.3) and (5.4) that the triple sampling asymptotic regret is equivalent to the loss of $(2\gamma^{-1})$ observations from each population had σ_1^2 and σ_2^2 been known and the fixed size sampling procedure utilized. It has been recommended by Hall (1981) and others to set $\gamma=1/2$ for practical purposes.

Since our theoretical developments are asymptotic in nature, we present simulation studies to explore the small to moderate sample size performance of our proposed procedure in the next section.

Table I. Triple Sampling & Fixed Sample Type II Errors $\gamma=0.5, \alpha=0.05$

k	$n^*=24$	$n^*=43$	$n^*=61$	$n^*=76$	$n^*=96$	fixed	k	$n^*=24$	$n^*=43$	$n^*=61$	$n^*=76$	$n^*=96$	fixed
	$m^*=384$	$m^*=246$	$m^*=171$	$m^*=125$	$m^*=96$			$m^*=384$	$m^*=246$	$m^*=171$	$m^*=125$	$m^*=96$	
0	0.929	0.937	0.94	0.94	0.94	0.95	1.5	0.182	0.175	0.173	0.172	0.172	0.164
0.1	0.925	0.933	0.935	0.936	0.936	0.946	1.6	0.138	0.131	0.129	0.128	0.128	0.12
0.2	0.911	0.919	0.922	0.922	0.922	0.932	1.7	0.102	0.096	0.094	0.093	0.093	0.085
0.3	0.888	0.896	0.899	0.899	0.9	0.91	1.8	0.074	0.068	0.066	0.066	0.066	0.058
0.4	0.856	0.964	0.867	0.867	0.867	0.877	1.9	0.052	0.047	0.045	0.045	0.045	0.039
0.5	0.815	0.823	0.825	0.825	0.825	0.835	2.0	0.035	0.031	0.03	0.03	0.03	0.025
0.6	0.765	0.772	0.774	0.774	0.774	0.783	2.1	0.024	0.02	0.02	0.019	0.019	0.016
0.7	0.707	0.712	0.714	0.714	0.714	0.721	2.2	0.015	0.013	0.012	0.012	0.012	0.009
0.8	0.642	0.646	0.647	0.647	0.647	0.652	2.3	0.01	0.008	0.008	0.007	0.007	0.005
0.9	0.572	0.574	0.575	0.575	0.575	0.578	2.4	0.006	0.005	0.004	0.004	0.004	0.003
1.0	0.5	0.5	0.5	0.5	0.5	0.5	2.5	0.003	0.003	0.003	0.003	0.002	0.002
1.1	0.428	0.426	0.425	0.425	0.425	0.422	2.6	0.002	0.002	0.001	0.001	0.001	0.001
1.2	0.358	0.354	0.353	0.352	0.352	0.348	2.7	0.001	0.001	0.001	0.001	0.001	0
1.3	0.292	0.287	0.285	0.285	0.285	0.278	2.8	0.001	0	0	0	0	0
1.4	0.233	0.227	0.225	0.224	0 .224	0.217	2.9	0	0	0	0	0	0

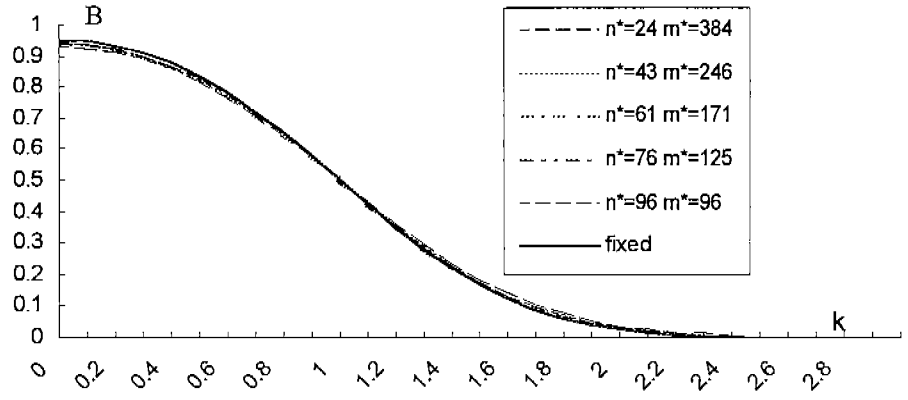

Figure I

6. Simulation results

A series of 5000 simulation were performed to evaluate the performance of the triple sampling procedure equations (3.1) and (3.3)-(3.4) under small to moderate sample sizes. The moderate optimal sample sizes (n^*, m^*) were set to (24, 384), (43, 246), (6, 171), (76, 125), (96, 96), (125, 76), (246, 43), (171, 61), (384, 24). These choices are the same as those used by Hall (1981) in his simulations. We also took $\mu_1 = 0, \mu_2 = 0, \sigma_2 = 1$, and $\sigma_1 = \sqrt{n^* m^{*-1}}$. The initial sample sizes (n_0, m_0) were set al (5,5), (10,10) ad (15,15) to study their impact on the coverage probability in general and on N and M in particular. The design factor γ was set at 0.3, 0.5 and 0.8 and the confidence coefficient was set at 0.90, 0.95 and 0.99. Results for smaller choices of (n^*, m^*) showed similar patterns and are omitted for brevity.

Detailed findings for some selected cases are presented in Table II. In each row of a given case, we provide summary statistics corresponding to (n^*, m^*), their estimates $(\overline{N}, \overline{M})$, their standard errors $(s.e.\overline{N}, s.e.\overline{M})$, the observed regret $\overline{\omega}$, and finally the estimated coverage $\hat{P}$

We notice that the estimated coverage $\hat{P}$ is larger than or equal to the nominal value $(1-\alpha)$ in all cases. However, such performance depends on the choice of both (n_0, m_0) and γ. Of the cases examined the best performance is achieved when $\gamma = 0.5$ and $(n_0, m_0) = (10, 10)$. In this case the procedure seems to have no tendency to oversample, as is evident by comparing (n^*, m^*) to the corresponding estimates $(\overline{N}, \overline{M})$. The estimated regret $\overline{\omega}$ is bounded as expected from our theoretical results in (5.3), even for small and moderate optimal sample sizes. The estimated value $\overline{\delta}$ of δ is always 0 for all cases (data

not shown).

To compare our results with those of Hall (1981), we find that the estimated coverage $\hat{P}$ is larger than Hall's without adding extra samples to the final stages sample sizes. This remark is supported by considering our cost of ignorance $\{(a^2-\gamma+13)(16\gamma)^{-1}\}$, which is less than Hall's (1981) cost of ignorance, $\{(a^2-\gamma+5)(2\gamma)^{-1}\}$

Table II. Triple Sampling Simulation

$n_0=m_0=5,\ \alpha=0.10,\ \gamma=0.3$

n^*	$\bar{N}$	$se_{..}\bar{N}$	m^*	$\bar{M}$	$se_{..}\bar{M}$	$\bar{\omega}$	$\hat{P}$
24	20 1	0 153	384	372.9	1.026	2.67	0 9232
43	36	0 262	246	234.4	0.824	2.36	0 9148
61	52.2	0 348	171	159.7	0.671	-0.66	0.9168
76	66.6	0.4	125	114	0.567	10.64	0.9102
96	85.5	0.487	96	86 5	0.474	0.44	0.9136
125	114.5	0.561	76	66 1	0 401	3.42	0.9184
171	159.2	0.68	61	52 8	0 352	7 61	0.9172
246	235.1	0.807	43	63.5	0 263	13.98	0.9172
384	372.7	1.027	24	20.6	0 155	-7.04	0 9204

$n_0=m_0=15,\ \alpha=0.10,\ \gamma=0.3$

n^*	$\bar{N}$	$se_{..}\bar{N}$	m^*	$\bar{M}$	$se_{..}\bar{M}$	$\bar{\omega}$	$\hat{P}$
24	24.7	0.119	384	377.4	0.791	-5.31	0 9446
43	41.9	0.195	246	238.9	0.645	-1.48	0 9404
61	57.3	0.257	171	163.7	0.534	-5.90	0.9398
76	70.7	0 313	125	118.6	0.453	-2 99	0.9352
96	90	0.372	96	89.1	0.377	-2 53	0.9382
125	118 1	0.452	76	70.7	0.31	-4.88	0.9416
171	163.8	0.526	61	57.1	0.257	2.44	0.9306
246	237.2	0.639	43	42	0.194	2.86	0.9420
384	377.7	0.790	24	24.8	0.119	5.25	0.9456

$n_0=m_0=15,\ \alpha=0.01,\ \gamma=0.8$

n^*	$\bar{N}$	$se\ \bar{N}$	m^*	$\bar{M}$	$se_{..}\bar{M}$	$\bar{\omega}$	$\hat{P}$
24	25	0 121	384	378	0.79	20.46	0 9952
43	41 9	0 189	246	238.5	0.627	11.54	0.9920
61	57	0.255	171	162.8	0.533	3.91	0.9898
76	70.6	0.314	125	117.7	0.445	-0.91	0.9890
96	89.6	0.378	96	89.6	0.382	2.84	0.9886
125	118.3	0.448	76	71 3	0 31	-3 38	0.9920
171	161	0.521	61	57 7	0 259	-2 36	0.9900
246	239.3	0.632	43	41.8	0.193	2.77	0.9918
384	400.1	0.871	24	24.8	0.118	9.46	0.9938

$n_0=m_0=15,\ \alpha=0.05,\ \gamma=0.5$

n^*	$\bar{N}$	$se_{..}\bar{N}$	m^*	$\bar{M}$	$se_{..}\bar{M}$	$\bar{\omega}$	$\hat{P}$
24	23.6	0.099	384	400.8	0.877	8 37	0 9726
43	42.1	0.177	246	256	0.604	-2.86	0.9688
61	60 7	0.225	171	176.6	0.455	-3.65	0.9736
76	77 1	0.261	125	129.3	0.364	-2 36	0.9756
96	98.7	0.317	96	97.7	0.298	1.75	0.9714
125	128.2	0.369	76	76.8	0.26	2.76	0.9688
171	176.9	0.466	61	61.2	0.228	8.88	0.9694
246	256.2	0.628	43	42.6	0.179	-0.17	0.9706
384	380.9	0.607	24	23.7	0 1	0.17	0.9750

$n_0=m_0=10,\ \alpha=0.05,\ \gamma=0.5$

n^*	$\bar{N}$	$se_{..}\bar{N}$	m^*	$\bar{M}$	$se_{..}\bar{M}$	$\bar{\omega}$	$\hat{P}$
24	22.6	0.119	384	381.8	0.645	2.23	0.9676
43	39.1	0.199	246	242 6	0 527	-0 41	0 9660
61	56.7	0.255	171	167	0.446	1.16	0 9658
76	71.8	0.288	125	120.4	0.38	0.84	0.9698
96	91.6	0.325	96	91.1	0.332	1.19	0.9654
125	121	0.378	76	71.6	0.285	1.41	0.9646
171	167.9	0.444	61	57	0.256	-0.25	0.9656
246	242.8	0.53	43	39 1	0.198	3.80	0.9648
384	382.9	0.680	24	22 3	0.117	-7.40	0.9670

$n_0=m_0=10,\ \alpha=0.01,\ \gamma=0.8$

n^*	$\bar{N}$	$se_{..}\bar{N}$	m^*	$\bar{M}$	$se_{..}\bar{M}$	$\bar{\omega}$	$\hat{P}$
24	22.1	0.116	384	380.6	0.681	0.47	0.9910
43	39.2	0.199	246	242.9	0.534	4.76	0.9872
61	56 7	0.253	171	167.4	0.441	-1.37	0.9916
76	71.8	0.296	125	120.5	0.375	3.63	0.9896
96	91 8	0.322	96	91.3	0.329	-2.71	0.9908
125	121.2	0.383	76	71.6	0.29	1 74	0.9914
171	166.7	0.438	61	57.1	0.249	4.92	0.9896
246	242.4	0.53	43	39.7	0.199	-1.83	0.9884
384	411.6	1.156	24	22.1	0.116	-0.50	0.9918

Reference

[1] Anscombe, F. J. (1953). Sequential estimation, *Journal of the Royal Statistics Society, B*, Vol. 15, 1-12.

[2] Barnard, G.A. (1984). Comparing the mean of two independent samples, *Applies Statistics*, Vol. 33, 266-271.

[3] Box, G. E.P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, Reading, Massachusetts. Addison-Wesley Publishing Co.

[4] Chapman, G. G. (1950). Some two-sample tests, *Annals of Mathematical Statistics*, Vol. 21, 601-606.

[5] Chow, Y. S. and Martinsek, A. T. (1982). Bounded regret of a sequential procedure for estimation of the mean, *Annals of Statistics*, Vol. 10, 909-914.

[6] Chow, Y. S. and robbins, H. (1965). On the asymptotics theory of fixed width sequential confidence intervals for the mean, *Annals of Mathematical Statistics*, Vol. 36, 457-462.

[7] Chow, Y. S. and Yu, K. F. (1981). On the performance of a sequential procedure for the estimation of the mean, *Annals of Statistics*, Vol. 9, 184-189.

[8] Eisele, J. R. (1990). An adaptive biased coin design for the Behrens-Fisher problem, *Sequential Analysis*, Vol. 9, 335-342.

[9] Ghosh, B.K. (1975a). A two-stage procedure for the Behrens-Fisher problem, *Journal of the American Statistics Association*, Vol. 70, 457-462.

[10] Ghosh, B.K. (1975b). On the distribution of the difference of two t-variables, *Journal of the American Statistics Association*, Vol. 70, 463-467.

[11] Ghosh, M. and Mukhopadhyay, N. (1980). Sequential point estimation of the difference of two normal means. *Annals of Statistics*, Vol. 8, 221-225.

[12] Ghosh, M. and Mukhopadhyay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential estimation procedures, *Sankhya, A*, Vol. 43, 220-227.

[13] Ghurye, S. G. and Robbins, H. (1954). Two-stage proceudres for estimating the difference between means, *Biometrika*, Vol. 41, 146-152.

[14] Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean, *Annals of Statistics*, Vol. 9, 1229-1238.

[15] Hamdy, H. I. (1988). Remarks on the asymptotic theory of triple stage estimation of the normal mean, *Scandinavian Journal of the Statistics*, Vol. 15, 303-310.

[16] Hamdy, H. I., Mukhopadhyay, N., Costanza, M. C. and Son, M. (1988). Triple stage point estimation for the exponential location parameter, *Annals of Institute Statistical Mathematics*, Vol. 40, 785-797.

[17] Hamdy, H. I., Almahmeed, M. and Al-Zalzala, Y. (1996). A certain Accelerating sequential procedure to construct simultaneous confidence region: The exponential case, *Statistics,* Vol. 28, 57-71.

[18] Hayre, L. S. (1983). Sequential estimation of the difference between the means of two normal populations, *Metrika*, Vol. 30, 101-107.

[19] Louis, T. A. (1975). Optimal allocation in sequential tests comparing means of two Gaussian populations, *Biometrika*, Vol. 62, 359-369.

[20] Martinsek, A. T. (1988). Negative regret, optimal stopping, and the elimination of outliers, *Journal of the American Statistical Association*, Vol. 83, 160-163.

[21] Mukhopadhyay, N. (1985). A ote on three stage and sequential point estimation procedures for a normal mean, *Sequential Analysis*, Vol. 4, 311-319.

[22] Mukhopadhyay, N., Hamdy, H. I., Al-Mahmeed, M. and Costanza, M. C. (1987). Three-stage point estimation procedures a normal mean, *Sequential Analysis*, Vol. 6, 21-36.

[23] Nel, D. G. and Van der Merwe, C. A. (1986). A solution to the Behrens-Fisher problem, *Communication Statistics*, Vol. 15, 3719-3736.

[24] O'Neill, R. T. and Rohtgi, V. K. (1973). A two-stage proceudre for estimating the difference between the mean vectors of two multivariate normal distributions, *Trabajas Estadis. Investigation Oper.*, Vol. 24, 123-130.

[25] Patil, V. H. (1964). The Behrens-Fisher problem and its Bayesian solution, *Journal of the India. Statistical Association*, Vol. 2.

[26] Robbins, H. (1959). *Sequential estimation of the mean of a normal population. In Probability and Statistics* (Harold Cramer Volume), 235-245. Almquist and Wiksell, Uppsala.

[27] Robbins, H., Simons, G., and Starr, N. (1967). A sequential analogue of the Behrens-Fisher problem, *Annals of Mathematical Statistics*, Vol. 38, 1384-1391.

[28] Scheffe, H. (1943). On solutions of the Behrens-Fisher problem based on the t-distribution, *Annals of Mathematical Statistics,* Vol. 14, 35-44.

[29] Scheffe, H. (1970). Practical solutions of the Behrens-Fisher problem, *Journal of the American Statistical Association*, Vol. 65, 1501-1508.

[30] Simons, G. (1968). On the cost of not knowing the variance when making a fixed width confidence interval for the mean, *Annals of Mathematical Statistics*, Vol. 39, 1946-1952.

[31] Srivastava, M. S. (1970). On a sequential analogue of the Behrens-Fisher problem, *Journal of the Royal Statistical Society, B*, Vol. 32, 144-148.

[32] Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent

of the variance, *Annals of Mathematical Statistics, Vol. 16, 243-258.*

[33] Wald, A. (1947). *Sequential Analysis*. Wiley, New York: Reprinted by Dover Publications Inc. (1973).

[34] Woodroofe, M. (1985). Asymptotic local minimxity in sequential point estimation, *Annals of Statistics*, Vol. 13, 676-688.

[35] Woodroofe, M. (1987). *Asymptotically optimal sequential point estimation in three stages. In New Perspectives in Theoretical and Applied Statistics*, 397-411. Wiley, New York.