專題研究計畫成果報告中英文摘要

計畫名稱：兩常態母群體共同平均數估計式 MLE 和 GDE 之比較

1. 計畫中文摘要：假設兩常態母群體具有相同平均數的條件底下，在過去幾十年當中，被用估計這個共同平均數的估計式，除了最大概似估計式(MLE)之外就屬 Graybill-Deal 估計式(GDE) 最受矚目了。然而這兩個估計式之間在不同準則底下的統計性質哪一個表現較佳，在一般文獻中卻鮮少被討論，原因可能是最大概似估計式(MLE)較為複雜，而且在母體變異數未知的情況底下無封閉解，所以造成理論發展的阻礙。本研究已變異數來比較 MLE 和 GDE 這兩個估計式的表現，另外，在探討 GDE 的漸近變異數時，本研究一更新了 GDE 變異數的區間範圍。

關鍵詞：變異數、漸近變異數、均方差、PNC、SDC

2. 計畫英文摘要：For estimating the common mean of two normal populations with unknown and possibly unequal variances the well known Graybill-Deal estimator (GDE) has been a motivating factor for research over the last five decades. Surprisingly the literature doesn’t have much to show when it comes to the maximum likelihood estimator (MLE) and its properties compared to those of the GDE. The purpose of this note is to shed some light on the structure of the MLE, and compare it with the GDE. While studying the asymptotic variance of the GDE, we provide an upgraded set of bounds for its variance. A massive simulation study has been carried out with very high level of accuracy to compare the variances of the above two estimators results of which are quite interesting.

Keywords: Admissibility, Inadmissibility, Asymptotic Variance

行政院國家科學委員會專題研究計畫成果報告

計畫編號：**NSC 96-2118-M-032-004-**

執行期限：**96** 年 **08** 月 **01** 日至 **97** 年 **07** 月 **31** 日

主持人：林志娟 淡江大學統計系專任副教授

計畫參與人員：碩士班研究生吳旻珊 詹凱婷 王孝欽 高欣如 張杰楷

博士班研究生林亭妤

Abstract

For estimating the common mean of two normal populations with unknown and possibly unequal variances the well known Graybill-Deal estimator (GDE) has been a motivating factor for research over the last five decades. Surprisingly the literature doesn't have much to show when it comes to the maximum likelihood estimator (MLE) and its properties compared to those of the GDE. The purpose of this note is to shed some light on the structure of the MLE, and compare it with the GDE. While studying the asymptotic variance of the GDE, we provide an upgraded set of bounds for its variance. A massive simulation study has been carried out with very high level of accuracy to compare the variances of the above two estimators results of which are quite interesting.

Keywords: Admissibility, Inadmissibility, Asymptotic Variance

1. Introduction

One of the oldest and most interesting problems in statistical inference, which has dogged the researchers over the last five decades, is the estimation of a common mean of two normal populations with unknown and possibly unequal variances.

To be specific, let us assume that we have *iid* observations

$X_{i1}, \cdots, X_{in_i}$ from $N(\mu, \sigma_i^2), \quad i = 1, 2.$

Define $\overline{X}_i$ and S_i as

$$\overline{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i, \qquad S_i = \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2, \qquad (1.1)$$

where

$$\overline{X}_i \sim N(\mu, \sigma_i^2 / n_i), \quad S_i \sim \sigma_i^2 \chi^2_{(n_i-1)} \quad (i = 1, 2),$$

and these four statistics are mutually independent. Throughout this note it is assumed that $n_i \geq 2 \quad (i = 1, 2)$ unless mentioned otherwise.

Note that $(\bar{X}_1, \bar{X}_2, S_1, S_2)$ is minimal sufficient for $(\mu, \sigma_1^2, \sigma_2^2)$ but not complete. As a result, one can not get the UMVUE (if it exists) using the standard Rao-Blackwell theorem on an unbiased estimator for estimating the common mean μ.

The motivation of this problem (i.e., estimation of μ) comes from a balanced incomplete block design (BIBD) with uncorrelated random block effects. For the ℓ^{th} treatment effect (say, τ_ℓ) one has two estimates - namely, the intra-block estimate and the inter-block estimate (say, $\hat{\tau}_\ell$ and $\hat{\tau}_\ell^*$ respectively). Under the usual design assumptions, $\hat{\tau}_\ell$ and $\hat{\tau}_\ell^*$ are independent, have normal distributions with the common mean τ_ℓ but with unknown and possibly unequal variances. The problem thus boils down to derive an efficient estimate of τ_ℓ on the basis of $\hat{\tau}_\ell$, $\hat{\tau}_\ell^*$ and their variance estimates.

Coming back to our original model (1.1), if the population variances $(\sigma_i^2, i = 1, 2)$ are completely known, then the optimal estimator of μ is

$$\hat{\mu} = \sum_{i=1}^{2} (n_i / \sigma_i^2) \bar{X}_i / \sum_{i=1}^{2} (n_i / \sigma_i^2) \quad (1.2)$$

which is the UMVUE, BLUE as well as the MLE. For the case of equal sample sizes, one just needs to know the ratio $(\sigma_1^2 / \sigma_2^2)$, apart from $\bar{X}_1$ and $\bar{X}_2$, to obtain (1.2).

In our present problem, where $\sigma_i^2, i = 1, 2$, are unknown and possibly unequal, the most appealing unbiased estimator of μ has been the Graybill-Deal estimator (GDE) given as

$$\hat{\mu}_{GD} = \sum_{i=1}^{2} (n_i / s_i^2) \bar{X}_i / \sum_{i=1}^{2} (n_i / s_i^2) \quad (1.3)$$

where $s_i^2 = S_i / (n_i - 1), \quad i = 1, 2.$

Graybill and Deal (1959) obtained conditions on n_1 and n_2 for which $\hat{\mu}_{GD}$ has a smaller variance than $\bar{X}_i$, $i = 1, 2$.

Even though Graybill and Deal pioneered the research on common mean estimation, it is probably due to Zack's (1966, 1970) that many other researchers paid attention to this interesting problem and its real-life applications. In Zack's own words - "⋯ In 1963 I was approached by a soil engineer. He wanted to estimate the common mean of two populations and he didn't know anything about the variances. But, apriori from his theory he said that the means should be same, and here are the two samples from two different soils. So I thought about this problem a little bit and I started to investigate. I realized that there is room for innovation ⋯" (see Kempthorne et al. (1991)).

For other applications of the common mean problem, especially in clinical trials, see Kelleher (1996).

Most of the research so far on the estimation of a common mean has taken place on three fonts: (i) comparing the GDE with the individual sample means (i.e., $\bar{X}_i$'s); (ii) studying the optimality of GDE and its natural generalizations in suitable classes of estimators; and (iii) studying the performance of Bayes and preliminary test-based estimators with that of the GDE. For a good review of the literature on this problem and other generalizations one can see Kubokawa (1987, 1991) and other references therein. Among some interesting results pertaining to the GDE, Sinha (1985) obtained an unbiased estimate of the variance of the GDE in the form of an infinite series which can be truncated suitably to get an approximate unbiased estimate up to any desired order. This result is helpful because the studentized version $[(\hat{\mu}_{GD}-\mu)/\hat{Var}(\hat{\mu}_{GD})\}^{1/2}]$, which follows $N(0,1)$ asymptotically, can be used for testing as well as for interval estimation of μ.

Quite surprisingly there hasn't been any discussion about the MLE and its performance relative to the other estimators, especially the GDE. It should be pointed out that the GDE (in (1.3)) is not the MLE, contrary to the statement made by Kelleher (1996) or Sinha (1979).

The purpose of this note is to focus on the MLE and its properties which have long been neglected in the literature. Even with the availability of affordable and efficient computing resources no comparison has been made so far to see how the GDE performs relative to the MLE. An important component of our study has been to see how the variances of the GDE and the MLE depend on the parameters as well as the sample sizes. The numerical results that are reported in the literature didn't take this aspect seriously. As a result, the reported numerical studies have been haphazard, or incomplete at best.

In Section 2 we study the structure of the MLE and provide a useful representation. Also we find its bias (exact) and variance expressions. Further, we upgrade the existing results to obtain tighter bounds for the variance of the GDE. In Section 3 we report the results of our extensive numerical study comparing the variances of the GDE and the MLE. A large number of replications has been used to ensure a very high level of accuracy of our results. Also, the numerical results provide some interesting and useful trends.

Before going to the next section we clarify some of the notations which have been used heavily in the rest of the paper. The other notations will be mentioned later as they are adopted.

Notations: Define

$$\gamma = \sigma_1^2 / \sigma_2^2, \qquad \alpha = \gamma / (k_0 k_1), \quad (1.4)$$

where

$k_0 = (n_1 / n_2),\ k_1 = (n_1 - 1)/(n_2 - 1).$

Also assuming $n_i > 3, i = 1, 2,$

$k_1^* = (n_2 - 3)/(n_1 - 1)$ and

$k_2^* = (n_2 - 1)/(n_1 - 3)$ (1.5)

2. The MLE of the Common Mean

The reason why the MLE of μ has eluded the interest of many researchers is probably its complicated structure. It doesn't have any closed expression, and as a result the exact sampling distribution is impossible to derive.

The log-likelihood function of the minimal sufficient statistics (1.1) is

$$L^* = \sum_{i=1}^{2} [\text{constant} - (n_i / 2)\ ln(\sigma_i^2)\ . - \{S_i + n_i(\overline{X}_i - \mu)^2\} / (2\sigma_i^2)] \quad (2.1)$$

Differentiations of L^* *w.r.t.* μ, σ_1^2 and σ_2^2, and setting them equal to 0 yield the MLEs $\hat{\mu}_{ML}, \hat{\sigma}^2_{1(ML)}, \hat{\sigma}^2_{2(ML)}$ which must satisfy

$$\hat{\sigma}^2_{1(ML)} = (S_1 / n_1) + \{n_2 \hat{\sigma}^2_{1(ML)} / (n_2 \hat{\sigma}^2_{1(ML)} + n_1 \hat{\sigma}^2_{2(ML)})\}^2 D^2, \quad (2.2)$$

$$\hat{\sigma}^2_{2(ML)} = (S_2 / n_2) + \{n_1 \hat{\sigma}^2_{2(ML)} / (n_2 \hat{\sigma}^2_{1(ML)} + n_1 \hat{\sigma}^2_{2(ML)})\}^2 D^2, \quad (2.3)$$

$$\hat{\mu}_{ML} = \{\sum_{i=1}^{2} (n_i / \hat{\sigma}^2_{i(ML)}) \overline{X}_i\} / \{\sum_{i=1}^{2} (n_i / \hat{\sigma}^2_{i(ML)})\}, \quad (2.4)$$

where $D = (\overline{X}_1 - \overline{X}_2)$.

Notice that both $\hat{\sigma}^2_{1(ML)}$ and $\hat{\sigma}^2_{2(ML)}$ are functions of S_1, S_2 and $D^2 = (\overline{X}_1 - \overline{X}_2)^2$. Thus it is easy to write $\hat{\mu}_{ML}$ as

$$\hat{\mu}_{ML} = \overline{X}_1 - D\hat{\phi}_{ML}, \quad (2.5)$$

where

$$\hat{\phi}_{ML} = \hat{\phi}_{ML}(S_1, S_2, D^2) = (n_2 / \hat{\sigma}^2_{2(ML)}) / \{(n_1 / \hat{\sigma}^2_{1(ML)}) + (n_2 / \hat{\sigma}^2_{2(ML)})\}.$$

(2.6)

Using the facts that D^2 is independent of (S_1, S_2), and the conditional distribution of $D \mid D^2$ is centered at 0, it is easy to obtain the following result.

Theorem 2.1. The MLE $\hat{\mu}_{ML}$ (in (2.4)) is unbiased for the common mean μ.

The fact that the MLE is unbiased is not surprising. Also, $\hat{\mu}_{ML}$ is a member of the class of affine (i.e., location and scale) transformations given as

$$C = \{ \hat{\mu} \mid \hat{\mu} = \overline{X}_1 - D\hat{\phi}, \quad 0 \le \hat{\phi} = \hat{\phi}(S_1, S_2, D^2) \le 1 \}, \tag{2.7}$$

which is also a class of unbiased estimators. The admissibility (or otherwise) of the GDE (in (1.3)), under the squared error loss function, in C has been an open problem for a long time. However, Sinha and Mouqadem (1982) considered the special case of $n_1 = n_2 = n$ (say), and showed that $\hat{\mu}_{GD}$ is extended admissible in C for $n \ge 5$, i.e., there doesn't exist any $\hat{\mu} \in C$ such that $Var(\hat{\mu}) \le Var(\hat{\mu}_{GD}) - \varepsilon$ for all (σ_1^2, σ_2^2) and $\varepsilon > 0$.

In the rest of the paper we'll study the variances of $\hat{\mu}_{ML}$ and $\hat{\mu}_{GD}$ which, among other things, enable us to compare them comprehensively. We write the variance expression of each estimator in a standardized form as $Var(\sqrt{n_1}\,\hat{\mu}/\sigma_1)$, and that too through a simple representation which allows us to see how each expression depends on (σ_1^2, σ_2^2) as well as (n_1, n_2). This also helps us in our numerical study to compare the variances effectively.

Characterize the minimal sufficient statistics as

$$S_i = \sigma_i^2 V_i \text{ and } \overline{X}_i = \mu + (\sigma_i / \sqrt{n_i}\,)\, Z_i, \quad i = 1, 2 \tag{2.8}$$

where $V_i \sim \chi^2_{(n_i - 1)}$, $Z_i \sim N(0,1)$, $i = 1, 2$, and they are independent. Further, we write

$$S_2 = \sigma_1^2 (1/\gamma) V_2 \quad \text{and} \quad \overline{X}_2 = \mu + (\sigma_1 / \sqrt{n_1}\,)\sqrt{(k_0/\gamma)}\, Z_2. \tag{2.9}$$

Now, define the scaled versions of $\hat{\sigma}^2_{1(ML)}$ and $\hat{\sigma}^2_{2(ML)}$ as $\hat{\sigma}^2_{10} = (n_1/\sigma_1^2)\,\hat{\sigma}^2_{1(ML)}$ and

$$\hat{\sigma}^2_{20} = (n_2/\sigma_2^2)\,\hat{\sigma}^2_{2(ML)} = (n_1/\sigma_1^2)\,(\gamma/k_0)\,\hat{\sigma}^2_{2(ML)}. \tag{2.10}$$

Using (2.8)–(2.10), the equations (2.2) and (2.3) can be rewritten as (with details omitted)

$$\hat{\sigma}^2_{10} = V_1 + \{\hat{\sigma}^2_{10} / (\hat{\sigma}^2_{10} + (k_0^2/\gamma)\hat{\sigma}^2_{20})\}^2 \times (Z_1 - \sqrt{(k_0/\gamma)}\, Z_2)^2 \tag{2.11}$$

$$\hat{\sigma}_{20}^2 = V_2 + (k_0/\gamma)\{k_0\hat{\sigma}_{20}^2/(\hat{\sigma}_{10}^2 + (k_0^2/\gamma)\hat{\sigma}_{20}^2)\}^2 \times (Z_1 - \sqrt{(k_0/\gamma)}\, Z_2)^2. \quad (2.12)$$

Thus, to obtain the MLEs of σ_1^2 and σ_2^2 for given (σ_1^2, σ_2^2) and (V_1, V_2, Z_1, Z_2), one first needs to solve (2.11) and (2.12) to get $\hat{\sigma}_{10}^2$ and $\hat{\sigma}_{20}^2$, and then rescale them according to (2.10) resulting into $\hat{\sigma}_{1(ML)}^2$ and $\hat{\sigma}_{2(ML)}^2$.

Using (2.8) — (2.10) in (2.4), the estimator $\hat{\mu}_{ML}$ is represented as (with details omitted)

$$\begin{aligned}\sqrt{n_1}\,(\hat{\mu}_{ML} - \mu)/\sigma_1 &= \{k_0^2\,\hat{\sigma}_{20}^2\,Z_1 + \sqrt{\gamma k_0}\,\hat{\sigma}_{10}^2\,Z_2\} /\{k_0^2\,\hat{\sigma}_{20}^2 + \gamma\,\hat{\sigma}_{10}^2\} \\ &= H_{ML}(V_1, V_2, Z_1, Z_2 \mid \gamma) = H_{ML} \text{ (say).} \quad (2.13)\end{aligned}$$

The following result is now trivial.

Theorem 2.2. $Var(\sqrt{n_1}\hat{\mu}_{ML}/\sigma_1) = E(H_{ML}^2)$, where H_{ML} is given in (2.13).

From (2.13) and the above theorem it is now clear that $Var(\sqrt{n_1}\,\hat{\mu}_{ML}/\sigma_1)$ depends on the parameters solely through $\gamma = (\sigma_1^2/\sigma_2^2)$. In fact, this is true for any $\hat{\mu}$ in C (see (2.7)).

Remark 2.1. The expression (2.13) comes handy to study $Var(\sqrt{n_1}\,\hat{\mu}_{ML}/\sigma_1)$ numerically through simulation. For specific γ and (n_1, n_2), the random vector $T = (V_1, V_2, Z_1, Z_2)$ is generated a large number of times, say Q times. Get $H_{ML}^{(\ell)} = H_{ML}^{(\ell)}(T^{(\ell)} \mid \gamma)$, where $T^{(\ell)}$ is the ℓ^{th} replication of T. Note that $V_i \sim \chi^2_{(n_i-1)}$, $Z_i \sim N(0,1), i = 1, 2.$ Then $E(H_{ML}^2)$ is approximated by $\{\sum_{\ell=1}^{Q}(H_{ML}^{(\ell)})^2/Q\}$.

Though any further simplification of (2.13) seems unlikely, the asymptotic variance of $\hat{\mu}_{ML}$ is easy to get. The (1.1) element of I^{-1}, where $I = I(\mu, \sigma_1^2, \sigma_2^2)$ is the (3×3) information matrix, is $((n_1/\sigma_1^2) + (n_2/\sigma_2^2))^{-1}$. The following result is immediate.

Theorem 2.3. $AsympVar(\sqrt{n_1}\hat{\mu}_{ML}/\sigma_1)_0 = \{1 + (\gamma/k_0)\}^{-1} = V$ (say).

The above asymptotic variance of $\hat{\mu}_{ML}$, which has a very simple closed form, will come useful to see how close the simulated variance (in Theorem 2.2) gets to its asymptotic counterpart. Though k_0 has been defined as (n_1/n_2), for asymptotic purposes it can be treated as a limiting value of (n_1 / n_2).

We now turn our attention to the age-old GDE for comparison to the MLE. From (1.3) it is trivial to get

$$Var(\hat{\mu}_{GD}) = (\sigma_1^2 / n_1) E(\hat{\phi}_{GD}^2) + (\sigma_2^2 / n_2) E((1-\hat{\phi}_{GD})^2), \quad (2.14)$$

where

$$\hat{\phi}_{GD} = (n_1 / s_1^2) / \{(n_1 / s_1^2) + (n_2 / s_2^2)\},$$
$$s_i^2 = S_i / (n_i - 1), i = 1, 2.$$

Using the notations in (1.4), (1.5), and the representation in (2.8), it is seen that (with details omitted)

$$Var(\sqrt{n_1}\, \hat{\mu}_{GD} / \sigma_1) = E\{(1+\alpha F_1)^{-2}\} + \beta E\{(\alpha + F_2)^{-2}\}$$
$$= E\{H_{GD}^2 (V_1, V_2 \mid \gamma)\} = E\{H_{GD}^2\} \quad \text{(say)}, \quad (2.15)$$

where $\alpha = \gamma / (k_0\, k_1)$, $\beta = \gamma / (k_0\, k_1^2)$, $F_1 = V_1 / V_2$ and $F_2 = 1/F_1$ and $H_{GD} = \{(1+\alpha F_1)^{-2} + \beta(\alpha + F_2)^{-2}\}^{1/2}$.

Remark 2.2. Again, the expression (2.15) will come handy for computing $Var(\sqrt{n_1}\, \hat{\mu}_{GD} / \sigma_1)$ through simulation. The variance will be approximated by $(\sum_{\ell=1}^{Q} (H_{GD}^{(\ell)})^2 / Q)$, after Q replications $H_{GD}^{(\ell)}$'s of H_{GD}.

An observation can be made about the lower bound of $Var(\sqrt{n_1}\, \hat{\mu}_{GD} / \sigma_1)$ from (2.15). The two terms of H_{GD} in (2.15) are convex in F_1 and F_2 respectively. Using Jensen's inequality and using $E(F_1) = 1/k_1^*$, $E(F_2) = k_2^*$, we have

$$Var(\sqrt{n_1}\hat{\mu}_{GD} / \sigma_1) \geq \{1 + \alpha / k_1^*\}^{-2} + \beta\{\alpha + k_2^*\}^{-2}$$
$$= \{1 + (\gamma / k_0)(n_2 - 1)/(n_2 - 3)\}^{-2}$$
$$+ (\gamma k_0)\{\gamma + k_0 (n_1 - 1)/(n_1 - 3)\}^{-2}$$
$$= A_*(\gamma), \quad \text{(say)}. \quad (2.16)$$

Nanayakkara and Cressie (1991) provided an upper bound of $Var(\sqrt{n_1}\, \hat{\mu}_{GD} / \sigma_1)$ which simplifies, after rearranging the terms in their inequality (2.30), as follows:

$$Var(\sqrt{n_1}\, \hat{\mu}_{GD} / \sigma_1) \leq \{(1 + \sqrt{k_2^* / k_1^*}\,)/2\}\{1 + (\gamma / k_0)\}^{-1}$$
$$= A_{NC}^*(\gamma), \quad \text{(say)}. \quad (2.17)$$

It is easy to see that as n_1 and n_2 go to infinity, taking k_0 as the limiting value of (n_1 / n_2), both $A_*(\gamma)$ and $A_{NC}^*(\gamma)$ converge to $\{1 + (\gamma / k_0)\}^{-1}$ (defined as V_0 in Theorem 2.3). Therefore, the following result is obvious, i.e., the GDE is first order efficient.

Theorem 2.4. $AsympVar(\sqrt{n_1}\,\hat{\mu}_{GD}/\sigma_1)$.

$= \{1+(\gamma/k_0)\}^{-1} = V_0$

Another set of bounds for $Var(\sqrt{n_1}\,\hat{\mu}_{GD}/\sigma_1)$ has been suggested by Korwar (1985) which is presented below in a much convenient form.

Theorem 2.5. (Korwar (1985)) With α as in (1.4), define

$$b_0 = (n_2-1)\{(3n_1-7)\alpha+(3n_2-1)\},$$
$$b_2 = 2(n_2-1)^2,$$
$$b_1 = \{(n_1-1)\alpha+(n_2-3)\} \times\{(n_1-3)\alpha+(n_2+1)\} + 2(n_2+1)(1-\alpha),$$
$$\mathrm{B}_1 = \mathrm{B}_1(n_1, n_2, \alpha) = \{b_0 - \sqrt{b_0^2-4b_1b_2}\}/(2b_1),$$
$$\mathrm{B}_2 = \mathrm{B}_2(n_1, n_2, \alpha) = 1-\mathrm{B}_1(n_2, n_1, 1/\alpha),$$
$$C_0 = (n_2-1)\{\alpha(n_1+1) + (n_2-3)\}/\{2\alpha(1-\alpha)(n_1-1)\},$$

and

$$C_1 = [\{(n_1-1)\alpha+(n_2-1)\} \times\{(n_1-3)\alpha+(n_2+1)\} - 4(1-\alpha)(n_2-1)]/\{2\alpha(1-\alpha)(n_1-1)\}.$$

Then,

$$A_{*K}(\gamma) \le Var(\sqrt{n_1}\,\hat{\mu}_{GD}/\sigma_1) \le A_K^*(\gamma),$$

where

(a) for $0<\alpha<1$, $A_{*K}(\gamma) = C_1\mathrm{B}_1 - C_0$ and $A_K^*(\gamma) = C_1\mathrm{B}_2 - C_0$;

(b) for $1<\alpha<\infty$, $A_{*K}(\gamma) = C_1\mathrm{B}_2 - C_0$ and $A_K^*(\gamma) = C_1\mathrm{B}_1 - C_0$.

Remark 2.3. Korwar's (1985) original result is a bit cumbersome, and discusses only the case $0<\alpha<1$. Also, the bounds don't exist for $\alpha=1$. However, for $\alpha=1$, exact variance expression can be derived (from his expression (2.1), page-357) as

$$Var(\sqrt{n_1}\,\hat{\mu}_{GD}/\sigma_1)\Big|_{\alpha=1} = \{(n_2-1)(n_1+n_2+2)\} /\{(n_1+n_2)(n_1+n_2-2)\}.$$

Combining all the bounds discussed above, a set of tighter bounds are obtained as

$$max\{A_*(\gamma), A_{*K}(\gamma)\} \le Var(\sqrt{n_1}\hat{\mu}_{GD}/n_1) \le min\{A_{NC}^*(\gamma), A_K^*(\gamma)\}. \quad (2.18)$$

Remark 2.4. The above upgraded bounds in (2.18) are the best (to our knowledge) so far. Our extensive numerical computations show that the bounds in (2.18) can be tighter than those in Theorem 2.5, especially when γ is near zero or too large.

3. Numerical Comparison of the Variances

The expressions in (2.13) and (2.15)

are used to simulate the variances of $\hat{\mu}_{ML}$ and $\hat{\mu}_{GD}$ in a comprehensive manner some of which are reported below. The simulation was massive in the sense that $Q = 10^5$ replications have been used for obtaining each simulated variance value, and this is done to achieve a high level of accuracy. Also, the asymptotic variance V_0 of the MLE as well as the GDE is reported here as a benchmark which is calculated directly from the formula in Theorem 2.3. For convenience the following notations are used in the subsequent tables.

$$V_{ML} = Var(\sqrt{n_1}\hat{\mu}_{ML} / \sigma_1)$$

$$V_{GD} = Var(\sqrt{n_1}\hat{\mu}_{GD} / \sigma_1).$$

The values of V_{GD} are more stable than the V_{ML} values in the sense that while the standard error (SE) of the simulation varied from 0.0051 to 0.00001 for the V_{ML} values, the range of the SE for the V_{GD} values has been 0.0005 to 0.00001. This is expected because $\hat{\mu}_{ML}$ is obtained after solving a system of nonlinear equations which may add a component of computational error, however small, to the overall sampling variation.

The overall picture that emerges from the simulation study is quite interesting, and challenges the conjecture that $\hat{\mu}_{GD}$ might be admissible.

First of all, V_{ML} and V_{GD} are found to be very close to each other, indicating that probably there isn't much difference between these two estimators' performance. The overall picture shows **three broad trends** as presented in the following tables. Apart from the simulated variance, the SE of the simulation is provided in brackets under each value.

<u>Case-I.</u> For equal sample sizes (i.e., $k_0 = n_1 / n_2 = 1$), we observe two subtrends.

(a) When $n_1 = n_2 = n$ (say) < 25, the $\hat{\mu}_{ML}$ is better than $\hat{\mu}_{GD}$ for extreme values of γ (i.e. γ close to 0 or ∞). For γ near or around 1, $\hat{\mu}_{GD}$ has better performance than $\hat{\mu}_{ML}$. The Table 3.1 (a) shows this subtrend.

(b) For $n_1 = n_2 = n$ (say) ≥ 25, $\hat{\mu}_{ML}$ appears to be better than $\hat{\mu}_{GD}$ uniformly over γ. However, for γ near 1, the dominance of $\hat{\mu}_{ML}$ over $\hat{\mu}_{GD}$ doesn't seem to be statistically significant, since the difference that we see between the simulated V_{ML} and V_{GD} is well within 2SE of the difference. (A simple two sample normal or *t*-test would strongly indicate that V_{GD} and

V_{ML} are equal for γ near 1.) The following Table 3.1 (b) testifies this subtrend. Even though asymptotically $V_{ML} = V_{GD} = V_0$, it appears that V_{GD} is converging a bit slowly.

Case-II. For n_1 and n_2 unequal, but not drastically different (i.e., $k_0 \neq 1$, but roughly $0.2 < k_0 < 5.0$), the variance curves of $\hat{\mu}_{ML}$ and $\hat{\mu}_{GD}$ cross each other only once (from small values of γ to large, or vice-versa). In this case none dominates the other uniformly. However, as k_0 moves away from 1, the trend of Case-III (discussed later) slowly emerges. This is shown in the following Table 3.2.

Case-III. When n_1 and n_2 are drastically different from each other (i.e., k_0 is roughly >5.0 or <0.2), the MLE seems to outperform the GDE uniformly. In some cases, for γ too small or too large, there may not be any statistical difference (taking the SE into consideration) between the two simulated variances; but for γ in the middle the MLE is certainly better than the GDE. This is reported in the following Table 3.3.

Remark 3.1. Under the Subcase-I (a) and the Case-II, the MLE and the GDE doesn't dominate each other uniformly. If one performs better on one part of the parameter space, then the other does so on the remaining parameter space. But under the Subcase-I (b) and the Case-III, the MLE does show superior performance than the GDE uniformly, though this may be marginal occasionally. This appears to be something new since no other estimator of the common mean in the literature has been reported to be performing better than the GDE uniformly. Considering all the three cases discussed above, the MLE's overall performance seems more appealing than the GDE, barring the fact that no explicit expression is available for the MLE.

4. Concluding Remark and Comments

The work focuses on the performance of the MLE of a common mean with unknown and possibly unequal variances which has been long neglected in the literature. Our surprising finding is that the MLE has better overall performance than the popular GDE, and at least for the heavily unbalanced case and the asymptotic balanced case the MLE seems to outperform the GDE uniformly. It is hoped that this paper will stimulate further research in studying the MLE of the common mean. Last but not least, a copy of our program and/or further simulation results will be made available to any interested reader.

References

1. Graybill, F. A. and Deal, R. B. (1959). Combining Unbiased Estimators. *Biometrics*, 15, 543 - 550.
2. Kelleher, T. (1996). A Bayes Method for Estimating the Common Mean of Two Normal Populations. *Communications in Statistics-Theory and Methods*, 25(9), 2141 - 2157.
3. Kempthorne, O., Mukhopadhyay, N., Sen, P. K. and Zacks, S. (1991). Research - How To Do It: A Panel Discussion. *Statistical Science*, 6(2), 149 - 162.
4. Korwar, R. M. (1985). Bounds for the Variance of the Graybill-Deal Estimator of the Common Mean of Two Normal Distributions. *Communications in Statistics - Theory and Methods*, 14(2), 355 - 361.
5. Kubokawa, T. (1987). Admissible Minimax Estimation of a Common Mean of Two Normal Populations. *The Annals of Statistics*, 15, 1245 - 1256.
6. Kubokawa, T. (1991). Estimation of the Common Mean and Its Applications. *Sugaku Expositions* (Translated and Published by the American Mathematical Society), 4(1), 97 - 110. [The original article appeared in Japanese in *Sugaku* (1990), Vol. 42, 121 - 130.]
7. Pal, N. and Lim, W. K. (1997). A note on Second Order Admissibility of the Graybill-Deal Estimator of a common Mean of Several Normal Populations. *Journal of Statistical Planning and Inference*, 63, 71-78.
8. Sinha, B. K. (1979). Is the Maximum Likelihood Estimate of the Common Mean of Several Normal Populations Admissible? *Sankhyā : The Indian Journal of Statistics*, B, 40, 192 - 196.
9. Sinha, B. K. (1985). Unbiased Estimation of the Variance of the Graybill-Deal Estimator of the Common Mean of Several Normal Populations. *Canadian Journal of Statistics*, 13, 243 - 247.
10. Sinha, B. K. and Mouqadem, O. (1982). Estimation of the Common Mean of Two Univariate Normal Populations. *Communications in Statistics - Theory and Methods*, 11, 1603 - 1614.
11. Nanayakkara, N. and Cressie, N.

(1991). Combining Two Unbiased Estimators of a Common Mean of Two Normal Populations. *Australian Journal of Statistics* (Currently, *Australian and New Zealand Journal of Statistics*), 33(1), 43 - 56.

12. Zacks, S. (1966). Unbiased Estimation of the Common Mean of Two Normal Distributions Based on Small Samples of Equal Size. *Journal of the American Statistical Association*, 61, 467 - 476.
13. Zacks, S. (1970). Bayes and Fiducial Equivariant Estimators of the Common Mean of Two Normal Distributions. *The Annals of the Mathematical Statistics*, 41, 59 - 69.

Table 3.1 (a). Variances of the MLE and the GDE for Case-I $(k_0 = 1)$ with $n_1 = n_2 = n < 25$.

(n_1, n_2)	Variance	γ						
		0.1	0.2	0.5	1.0	2.0	5.0	10.0
(5, 5)	V_{ML}	1.0199	1.0023	0.8518	0.6449	0.4274	0.1999	0.1019
		(0.0051)	(0.0050)	(0.0040)	(0.0030)	(0.0020)	(0.0010)	(0.0005)
	V_{GD}	1.0985	1.0135	0.8033	0.5997	0.4016	0.2023	0.1098
		(0.0018)	(0.0013)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909
(10, 10)	V_{ML}	0.9431	0.8997	0.7428	0.5627	0.3718	0.1792	0.0948
		(0.0042)	(0.0041)	(0.0034)	(0.0026)	(0.0017)	(0.0008)	(0.0004)
	V_{GD}	0.9641	0.9006	0.7310	0.5503	0.3658	0.1802	0.0966
		(0.0005)	(0.0005)	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909
(15, 15)	V_{ML}	0.9286	0.8630	0.7088	0.5380	0.3555	0.1720	0.0927
		(0.0041)	(0.0039)	(0.0032)	(0.0024)	(0.0016)	(0.0008)	(0.0004)
	V_{GD}	0.9391	0.8729	0.7085	0.5331	0.3544	0.1747	0.0939
		(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0000)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909
(20, 20)	V_{ML}	0.9172	0.8534	0.6984	0.5311	0.3496	0.1692	0.0918
		(0.0041)	(0.0038)	(0.0031)	(0.0024)	(0.0016)	(0.0008)	(0.0004)
	V_{GD}	0.9295	0.8613	0.6977	0.5249	0.3489	0.1722	0.0929
		(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909

\

Table 3.1 (b). Variances of the MLE and the GDE for Case-I $(k_0 = 1)$ with $n_1 = n_2 = n \geq 25$.

(n_1, n_2)	Variance	γ						
		0.1	0.2	0.5	1.0	2.0	5.0	10.0
(25, 25)	V_{ML}	0.9141	0.8432	0.6868	0.5187	0.3423	0.1696	0.0915
		(0.0041)	(0.0038)	(0.0031)	(0.0023)	(0.0015)	(0.0008)	(0.0004)
	V_{GD}	0.9245	0.8550	0.6911	0.5201	0.3458	0.1710	0.0924
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909
(50, 50)	V_{ML}	0.9077	0.8383	0.6746	0.5065	0.3369	0.1670	0.0908
		(0.0041)	(0.0038)	(0.0030)	(0.0023)	(0.0015)	(0.0007)	(0.0004)
	V_{GD}	0.9159	0.8433	0.6788	0.5100	0.3394	0.1687	0.0916
		(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909
(100, 100)	V_{ML}	0.8999	0.8315	0.6687	0.5015	0.3334	0.1664	0.0906
		(0.0040)	(0.0037)	(0.0030)	(0.0023)	(0.0015)	(0.0007)	(0.0004)
	V_{GD}	0.9123	0.8382	0.6727	0.5050	0.3363	0.1676	0.0912
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909
(500, 500)	V_{ML}	0.9072	0.8288	0.6684	0.4955	0.3315	0.1666	0.0908
		(0.0041)	(0.0037)	(0.0030)	(0.0022)	(0.0015)	(0.0007)	(0.0004)
	V_{GD}	0.9097	0.8343	0.6679	0.5010	0.3339	0.1669	0.0910
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	V_0	0.9091	0.8333	0.6667	0.5000	0.3333	0.1667	0.0909

Table 3.2. Variances of the MLE and the GDE for Case-II $(k_0 \neq 1, 0.2 \leq k_0 \leq 5.0)$.

(n_1, n_2)	Variance	γ						
		0.1	0.2	0.5	1.0	2.0	5.0	10.0
(10, 25)	V_{ML}	0.8397	0.7156	0.4866	0.3062	0.1750	0.0757	0.0387
		(0.0038)	(0.0032)	(0.0022)	(0.0014)	(0.0008)	(0.0003)	(0.0002)
	V_{GD}	0.8390	0.7081	0.4775	0.3067	0.1773	0.0772	0.0395
		(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
	V_0	0.8000	0.6667	0.4444	0.2857	0.1667	0.0741	0.0385
(25, 10)	V_{ML}	0.9762	0.9508	0.8757	0.7664	0.6059	0.3576	0.2104
		(0.0044)	(0.0042)	(0.0039)	(0.0035)	(0.0027)	(0.0016)	(0.0009)
	V_{GD}	0.9879	0.9649	0.8869	0.7668	0.5965	0.3540	0.2097
		(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0001)	(0.0000)
	V_0	0.9615	0.9259	0.8333	0.7143	0.5556	0.3333	0.2000
(50, 10)	V_{ML}	0.9788	0.9649	0.9304	0.8686	0.7511	0.5390	0.3525
		(0.0044)	(0.0043)	(0.0042)	(0.0039)	(0.0034)	(0.0024)	(0.0016)
	V_{GD}	0.9942	0.9847	0.9491	0.8822	0.7619	0.5322	0.3512
		(0.0002)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0001)	(0.0001)
	V_0	0.9804	0.9615	0.9091	0.8333	0.7143	0.5000	0.3333
(10, 50)	V_{ML}	0.7100	0.5339	0.3046	0.1730	0.0933	0.0390	0.0198
		(0.0032)	(0.0024)	(0.0014)	(0.0008)	(0.0004)	(0.0002)	(0.0001)
	V_{GD}	0.7027	0.5320	0.3049	0.1765	0.0949	0.0394	0.0199
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
	V_0	0.6667	0.5000	0.2857	0.1667	0.0909	0.0385	0.0196

Table 3.3. Variances of the MLE and the GDE for Case-III ($k_0 < 0.2$ or $k_0 > 5$).

(n_1, n_2)	Variance	γ						
		0.1	0.2	0.5	1.0	2.0	5.0	10.0
(10, 75)	V_{ML}	0.6040 (0.0027)	0.4235 (0.0019)	0.2194 (0.0010)	0.1208 (0.0005)	0.0634 (0.0003)	0.0260 (0.0001)	0.0132 (0.0001)
	V_{GD}	0.6044 (0.0001)	0.4261 (0.0001)	0.2236 (0.0001)	0.1233 (0.0000)	0.0647 (0.0000)	0.0264 (0.0000)	0.0133 (0.0000)
	V_0	0.5714	0.4000	0.2105	0.1176	0.0625	0.0260	0.0132
(75, 10)	V_{ML}	0.9888 (0.0044)	0.9856 (0.0044)	0.9520 (0.0043)	0.9136 (0.0041)	0.8211 (0.0037)	0.6384 (0.0029)	0.4525 (0.0021)
	V_{GD}	0.9964 (0.0002)	0.9907 (0.0002)	0.9691 (0.0003)	0.9250 (0.0004)	0.8379 (0.0003)	0.6390 (0.0002)	0.4534 (0.0001)
	V_0	0.9868	0.9740	0.9375	0.8824	0.7895	0.6000	0.4286
(10, 100)	V_{ML}	0.5282 (0.0024)	0.3518 (0.0016)	0.1725 (0.0008)	0.0926 (0.0004)	0.0484 (0.0002)	0.0197 (0.0001)	0.0099 (0.0000)
	V_{GD}	0.5302 (0.0001)	0.3550 (0.0001)	0.1760 (0.0001)	0.0947 (0.0000)	0.0489 (0.0000)	0.0199 (0.0000)	0.0100 (0.0000)
	V_0	0.5000	0.3333	0.1667	0.0909	0.0476	0.0196	0.0099
(100, 10)	V_{ML}	0.9911 (0.0044)	0.9822 (0.0044)	0.9650 (0.0043)	0.9306 (0.0042)	0.8686 (0.0039)	0.7064 (0.0032)	0.5278 (0.0024)
	V_{GD}	0.9970 (0.0001)	0.9937 (0.0002)	0.9781 (0.0003)	0.9469 (0.0004)	0.8794 (0.0003)	0.7099 (0.0002)	0.5303 (0.0001)
	V_0	0.9901	0.9804	0.9524	0.9091	0.8333	0.6667	0.5000