SHRINKAGE ESTIMATORS FOR THE MEAN OF A STRATIFIED NORMAL POPULATION

Richard E. Auer, Loyola College of Baltimore
Chien-Pai Han, University of Texas at Arlington

1. Introduction

The problem of estimating the mean vector of a multivariate normal population has been studied through the Bayesian framework, the James-Stein framework and the empirical Bayes framework with the result of obtaining estimators with smaller risk than the classical estimator (see Albert (1979), James and Stein (1961), Thisted (1976), James Thompson (1968)). Thompson studied the problem for the univariate case and did so from a still different viewpoint. His result was an estimator which possesses smaller mean square error (MSE) for a portion of the parameter space. A commonality of all these estimators is that each of them can be expressed as a shrinkage estimator which shrinks from a classical estimator toward a prior guess of the parameter value or else some data based value.

The purpose of this paper is not to simply extend Thompson's approach to the multivariate case. Rather, we assume a stratified normal population where the aim is to find estimators of the stratified mean of the population which offer some improvement over the usual estimator. Once these shrinkage estimators are found, their structural properties will be noted and, through a simulation study, their mean square errors will be studied. It should be noted that the most complete results and the simulation study deal with the case of two strata, however generalizations are given in most cases.

We assume that each of the p strata can be modelled by a normal distribution. Hence X_{ij} denotes the jth observation from the ith stratum where X_{ij} comes from a normal population with mean μ_i and variance σ_i^2. Furthermore, we assume the stratum weights, the proportions that the strata represent of the total population, are known and equal to π_i for $i=1,2,\dots,p$. The stratified mean, which is to be estimated, is then $\bar{\mu} = \sum^{p} \pi_i\mu_i$. The usual estimator of $\bar{\mu}$ is $\bar{X} = \sum^{p} \pi_i\bar{X}_i$ where $\bar{X}_i = \frac{1}{n_i}\sum^{n_i} X_{ij}$ and n_i is the sample size from the ith stratum. We will use the notation V_i to stand for σ_i^2/n_i, the variance of $\bar{X}_i$, and $V\bar{X}$, the variance of $\bar{X}$.

In order to estimate μ_i, the component part of $\bar{\mu}$, we consider estimators of the following forms:

a) $(1-c_i)(\bar{X}_i-\theta_i)+\theta_i$

b) $(1-c_i)(\bar{X}_i-\bar{X})+\bar{X}$. (1)

In the first case, θ_i is some pre-conceived value believed to be near the true value μ_i . In the second case, we shrink toward the overall stratified mean. In any case, the value of the shrinking factor c_i determines whether the estimator is nearly the same as $\bar{X}_i$ or if its value is nearer the value to which $\bar{X}_i$ is shrunk towards. With this flexibility built into the estimator of μ_i one hopes to find the MSE of our estimators to be less than the variance of $\bar{X}$.

An added restriction could be put on these estimators by requiring the shrinking factors to take on a common value, i.e., $c_i=c$ for $i=1,2,\dots,p$. Such a restriction leads to quite different shrinking factors from those of before and the resulting estimators of $\bar{\mu}$ can prove valuable in different settings than the estimators in (1).

One should note that the values of the shrinking factors will depend upon unknown μ_i's and possibly unknown V_i's. The performance of the estimators, when we must substitute estimates of these unknown parameters, will be analyzed in section 5 by the use of a Monte Carlo study. In sections 3 and 4, the estimators are studied in their pre-substitution form.

2. The Development of the Estimators

Before proceeding with finding these estimators, we must decide on a criterion upon which the estimators will be judged. Since we are generally dealing with biased estimators of $\bar{\mu}$ and μ_i , we attempt to minimize the MSE. We must find the value of c or else of c_i that insures such a minimization. This is done by differentiating the MSE with respect to the shrinking factor, setting the result equal to zero and solving for the factor.

If we define δ_i as the subsequent estimator of μ_i , we are dealing with $\Sigma\pi_i\delta_i$ as the estimator of $\bar{\mu}$. We can develop the estimator δ_i which minimizes the component MSE of δ_i or else we can choose to minimize the MSE of $\Sigma\pi_i\delta_i$ directly. If we note that

$$\text{MSE}\left(\sum_{i=1}^{p} \pi_i\delta_i\right)$$

$$= \sum_{i=1}^{p} \pi_i^2 \text{MSE}\delta_i + \sum_{i=1}^{p}\sum_{j\neq i}^{p} \pi_i\pi_j E(\delta_i-\mu_i)(\delta_j-\mu_j), \quad (2)$$

we can see that minimizing the component MSE's only assures us of minimizing the first sum on the right-hand side of (2). But minimizing the MSE of $\Sigma\pi_i\delta_i$ assures us of minimizing the entire right-hand side of (2).

Both of these possible criteria are pursued and comparisons of the results are made. When dealing with minimizing the component MSE's, the resulting estimators are called the component-wise estimators.

3. Component-wise Estimators

Considering the two forms in (1), we now determine the shrinking factors that minimize the component MSE's. As we are considering each component separately, there is no need to consider a common factor c.

The first estimator of $\bar{\mu}$ to be considered is given by $\delta_{\theta C}= \Sigma\, \pi_i\left((1-c_i)(\bar{X}_i-\theta_i)+\theta_i\right)$. The subscript C denotes componentwise minimization and the subsequent δ_i is the Thompson estimator. Thompson found the minimizing value of $V_i/\left(V_i+(\mu_i-\theta_i)^2\right)$ and found that this δ_i uniformly beats $\bar{X}_i$ in estimating μ_i. This does not mean, however, that $\Sigma\pi_i\delta_i=\delta_{\theta C}$ beats $\bar{X}$ in terms of overall MSE.

In order to show that $\delta_{\theta C}$ does often improve upon $\bar{X}$ note that,

defining $R_i= \dfrac{1}{V_i+(\mu_i-\theta_i)^2}$,

$$V\bar{X}\text{-MSE } \delta_{\theta C} = \sum_{i=1}^{p} \pi_i^2 V_i^2 R_i\left(1-(p-1)R_i(\mu_i-\theta_i)^2\right)$$

$$+ \sum_{i<j} \left(\pi_i V_i R_i(\mu_i-\theta_i)-\pi_j V_j R_j(\mu_j-\theta_j)\right)^2.$$

From this, one can see that $\delta_{\theta C}$ is better than $\bar{X}$ if one of the following conditions is met:

a) $p=2$,

b) $|\theta_i-\mu_i|<\sqrt{V_i/(p-2)}$ for $i=1,2,\ldots,p$,

meaning that our prior guess for μ_i is good or else the variation within a stratum is quite large when θ_i is not close to μ_i,

c) $\pi_i^2/\left(V_i+(\mu_i-\theta_i)^2\right)$ is especially small for the strata where $|\theta_i-\mu_i|>\sqrt{V_i/(p-2)}$.

Now consider $\delta_{\bar{X}C}= \Sigma\, \pi_i\left((1-c_i)(\bar{X}_i-\bar{X})+\bar{X}\right)$. With this estimator no prior guesses are needed in order to estimate $\bar{\mu}$. If all the μ_i's are nearly equal, one would expect $\delta_{\bar{X}C}$ to do well as it shrinks to a common value $\bar{X}$. Should the μ_i's be widely dispersed, it may still be better to shrink to $\bar{X}$ than an arbitrary set of θ_i's if these θ_i's are poorly chosen.

The superiority of $\delta_{\bar{X}C}$ to $\bar{X}$ will be shown explicitly when an equivalent estimator is developed later.

Minimizing the component MSE leads to a shrinking factor

$$c_i=(1-\pi_i)V_i/\left((\mu_i-\bar{\mu})^2+V(\bar{X}_i-\bar{X})\right) \quad . \qquad (3)$$

When the difference $\bar{X}_i-\bar{X}$ has a large variance or when the variance of $\bar{X}_i$ is small or when μ_i is not near $\bar{\mu}$, the value of c_i will be nearly equal to zero which, in turn, makes the estimator of μ_i shrink very little to $\bar{X}$. This is an expected result because in all three cases the value of $\bar{X}_i$ is likely to be a better estimator of μ_i than $\bar{X}$.

The value of c_i is always greater than zero and usually less than one, but, to be sure we are dealing with a true shrinkage estimator, we could further define c_i to be

$$\min\left(1,(1-\pi_i)V_i/\left((\mu_i-\bar{\mu})^2+V(\bar{X}_i-\bar{X})\right)\right) .$$

This precaution is taken in the simulation studies.

4. Estimators Minimizing MSE $\left(\sum_{i=1}^{p}\pi_i\delta_i\right)$

4.1. The Common Shrinking Factor

Let us find shrinkage estimators of the form $\Sigma\pi_i\delta_i$ that not only minimize the total MSE, but also assume a common shrinking factor, c. Notationally let the two forms in (1) be δ_θ and $\delta_{\bar{X}}$, respectively.

Considering the MSE of δ_θ, we find

$$\text{MSE}\delta_\theta= \text{MSE}\left(\sum_{i=1}^{p} \pi_i\left((1-c)(\bar{X}_i-\theta_i)+\theta_i\right)\right)$$

$$=E\left((1-c)(\bar{X}-\bar{\theta})+\bar{\theta}-\bar{\mu}\right)^2$$

where $\bar{\theta}= \Sigma\, \pi_i\theta_i$. But this is simply a special case of the univariate Thompson estimator where $c=V\bar{X}/\left(V\bar{X}+(\bar{\mu}-\bar{\theta})^2\right)$.

The MSE of $\delta_{\bar{X}}$ gives

$$\text{MSE}\left(\sum_{i=1}^{p} \pi_i\left((1-c)(\bar{X}_i-\bar{X})+\bar{X}\right)\right)$$

$$=\text{MSE}\left((1-c)\sum_{i=1}^{p} \pi_i(\bar{X}_i-\bar{X})+\bar{X}\right)$$

$$=\text{MSE}(\bar{X}) = V\bar{X} .$$

Hence any constant value of c leads to $\bar{X}$ itself showing that we are unable to improve upon $\bar{X}$ with the form $\delta_{\bar{X}}$.

4.2 Varying Shrinking Factors

Again we minimize the MSE of $\Sigma\,\pi_i\delta_i$, as in the previous section, but in allowing the shrinking factors to differ from stratum to stratum the results are not nearly as simple and automatic. It is of interest to compare the results of this section with those dealing with the component-wise estimators as both sets of estimators utilize shrinking factors that vary across the strata.

Notationally, we define $\delta_{\theta i}$ and $\delta_{\bar{X}i}$ as the estimators for the two forms in (1) respectively, where the subscript i denotes the fact that the shrinking factors, c_i, differ for each stratum.

Beginning with $\delta_{\theta i}$, we must minimize MSE $\delta_{\theta i}$ where

$$\text{MSE}(\delta_{\theta i}) =$$

$$E\left(\sum_{i=1}^{p} \pi_i\left((1-c_i)(\bar{X}_i-\theta_i)+\theta_i-\mu_i\right)^2\right) .$$

We find that $1-c_i$ equals

$$\frac{\pi_i\left((\mu_i-\theta_i)(\bar{\mu}-\bar{\theta})-(\mu_i-\theta_i)\sum_{j\neq i}^{p}\pi_j(\mu_j-\theta_j)(1-c_j)\right)}{\pi_i^2\left(V_i+(\mu_i-\theta_i)^2\right)}$$

for $i=1,2,\ldots,p$,

creating a system of p equations and p unknowns.

To clearly see the form of the shrinking factors, we study the case where p=2 and solve for $1-c_i$:

$$1-c_i=\frac{V_j\pi_j(\mu_j-\theta_j)(\bar{\mu}-\bar{\theta})}{\pi_j^2(V_iV_j+V_i(\mu_j-\theta_j)^2+V_j(\mu_i-\theta_i)^2)} \quad \text{for}$$

$(i,j)=(1,2)$ or $(i,j)=(2,1)$.

When $\bar{\mu}=\bar{\theta}$, giving our prior guess the mark of perfection, we find $1-c_i$ to be equal to zero for i=1 and 2. Thus $\delta_{\theta i}$ becomes simply $\bar{\theta}$.

Since $\delta_{\theta i}$ minimizes the total MSE while $\delta_{\theta C}$ only minimizes the sum of the component MSE's, we should expect the former to obviously dominate the latter. With p=2 and $\theta_1=\theta_2=0$, we find

$$\mathrm{MSE}\delta_{\theta C}-\mathrm{MSE}\delta_{\theta i}$$

$$=\frac{\mu_1^2\mu_2^2((\pi_2V_2\mu_1-\pi_1V_1\mu_2)^2+V_1V_2V\bar{X})}{(V_1V_2+V_1\mu_2^2+V_2\mu_1^2+\mu_1^2\mu_2^2)(V_1V_2+V_1\mu_2^2+V_2\mu_1^2)}$$

$$> 0.$$

To show that both $\delta_{\theta i}$ and $\delta_{\theta C}$ perform better than $\bar{X}$ for p=2, we consider

$$V\bar{X}-\mathrm{MSE}\delta_{\theta C}=\frac{(\pi_1\mu_2V_1-\pi_2\mu_1V_2)^2+V_1V_2V\bar{X}}{V_1V_2+V_1\mu_2^2+V_2\mu_1^2+\mu_1^2\mu_2^2}>0.$$

Because of the extensive algebraic difficulties in dealing with the estimator $\delta_{\bar{X}i}$, we initially consider the estimator for the case p=2. Generalizations will be commented upon for p>2 later in the paper.

The problem is simplified when one considers the following:

$$\mathrm{MSE}(\delta_{\bar{X}i})=E\big(\pi_1((1-c_1)(\bar{X}_1-\bar{X})+\bar{X})$$
$$+\pi_2((1-c_2)(\bar{X}_2-\bar{X})+\bar{X})-\bar{\mu}\big)^2$$
$$=E\big(\pi_1\pi_2(c_2-c_1)(\bar{X}_1-\bar{X}_2)+(\bar{X}-\bar{\mu})\big)^2.$$

Here we see that only the minimizing value of c_2-c_1 is needed. Any selection of c_2 and c_1 which gives the minimizing value for the difference would make $\delta_{\bar{X}i}$ the optimal estimator. If the solutions for c_1 and c_2 were found by differentiating the MSE with respect to c_1 and c_2 individually, a system of two equations would lead to the same result as considering the difference c_2-c_1.

The optimal value of c_2-c_1 subsequently becomes $(\pi_2V_2-\pi_1V_1)/\big(\pi_1\pi_2(V_1+V_2+(\mu_1-\mu_2)^2)\big)$ and the minimized MSE can be expressed as $\mathrm{MSE}\delta_{\bar{X}i}=V\bar{X}-(\pi_2V_2-\pi_1V_1)^2/\big(V_1+V_2+(\mu_1-\mu_2)^2\big)$.

This expression not only shows the obvious superiority of $\delta_{\bar{X}i}$ to $\bar{X}$, but also expresses the amount of superiority. The improvement is maximized with respect to the stratum means if $\mu_1=\mu_2$, an intuitively pleasing result as this is when $\bar{X}$ would seem to be the best expression to shrink toward. But the question of improving upon $\bar{X}$ versus not improving lies in the values of π_1, π_2, V_1, and V_2. These are the terms that, in a sense, dictate the relative weighting of the two strata.

To demonstrate how the structure of $\delta_{\bar{X}i}$ dictates how it works, let us recall that the form of $\delta_{\bar{X}i}$ was shown to be $\pi_1\pi_2(c_2-c_1)(\bar{X}_1-\bar{X}_2)+\bar{X}$ or more particularly

$$\frac{\pi_2V_2-\pi_1V_1}{(V_1+V_2+(\mu_1-\mu_2)^2)}(\bar{X}_1-\bar{X}_2)+\bar{X}\,. \qquad (4)$$

Should we want to express $\delta_{\bar{X}i}$ as an estimator of $\bar{\mu}$ that is shrinking between $\bar{X}_1$ and $\bar{X}_2$, it is a simple result to find that $\delta_{\bar{X}i}$ equals $(1-K)\,\bar{X}_1 + K\bar{X}_2$ where

$$K=\frac{V_1+\pi_2(\mu_1-\mu_2)^2}{V_1+V_2+(\mu_1-\mu_2)^2} \text{ and } 0<K<1. \qquad (5)$$

Each of the preceding forms gives us reasonable insight into $\delta_{\bar{X}i}$.

From (5) we see that $\delta_{\bar{X}i}$ shrinks more heavily toward $\bar{X}_2$ if V_1 is relatively larger than V_2 or if π_2 is larger than π_1. This is intuitively pleasing because we would expect smaller error when shrinking toward the sample stratum mean with smaller variance. Furthermore, since we are shrinking toward $\bar{X}$ in $\delta_{\bar{X}i}$, it also makes sense that we are shrinking more heavily toward the stratum mean with larger stratum weight as this tends to reduce the bias in estimating $\bar{\mu}$. Also noteworthy is that, as the difference between μ_1 and μ_2 expands to infinity, $\delta_{\bar{X}i}$ goes to $\bar{X}$. This demonstrates that the sensitive relationships of stratum variances and weights no longer matters in the face of such different stratum means.

We also study the structure of $\delta_{\bar{X}i}$ with respect to how its value compares with $\bar{X}$ by using expression (4). If we assume, without loss of generality, that $\bar{X}_1$ is less than $\bar{X}_2$, then the difference $\bar{X}_1-\bar{X}_2$ is negative. In the case of equal stratum weights, $\delta_{\bar{X}i}$ takes a value between $\bar{X}_1$ and $\bar{X}$ when V_2 is larger than V_1. Again this is a reasonable result as $\bar{X}_1$ will contribute less variance to the final measurement of MSE.

If we now assume equal values of V_1 and V_2 and a value of π_2 greater than $\frac{1}{2}$, this will force $\bar{X}$ closer to $\bar{X}_2$ than $\bar{X}_1$. When we inspected (5) we saw where such a value of π_2 would also force $\delta_{\bar{X}i}$ closer to $\bar{X}_2$ than $\bar{X}_1$. But here we use (4) to determine on which side of $\bar{X}$ the value of $\delta_{\bar{X}i}$ will fall. Although $\delta_{\bar{X}i}$ is nearer $\bar{X}_2$ than $\bar{X}_1$, we also see that $\delta_{\bar{X}i}$ is on the $\bar{X}_1$ side of $\bar{X}$. The estimator $\delta_{\bar{X}i}$ in attempting to reduce MSE by being less biased, also attempts to minimize MSE by shrinking to the side of $\bar{X}$ where the smaller stratum weight will reduce variance. So $\delta_{\bar{X}i}$ is balancing variance and bias in its attempt to minimize MSE.

The comparison of $\delta_{\bar{X}C}$ and $\delta_{\bar{X}i}$ is now a relevant issue. When p=2, it can be shown from (3) that $c_i= V_i/\big(\pi_j(V_1+V_2+(\mu_1-\mu_2)^2)\big)$ where $(i,j)=(1,2)$ or $(i,j)=(2,1)$. With these values for the two shrinking factors, we find the value of c_2-c_1 to be $(\pi_2V_2-\pi_1V_1)/\left(\pi_1\pi_2(V_1+V_2+(\mu_1-\mu_2)^2)\right)$. But this is the minimizing value for c_2-c_1 for the estimator $\delta_{\overline{X}i}$. This implies that, for p=2, the component-wise values of c_1 and c_2 not only minimize the sum of the component MSE's but the sum of the cross product terms in (2) as well.

To study the case of $\delta_{\overline{X}i}$ for $p>2$ and to learn more about the relationship between $\delta_{\overline{X}i}$ and $\delta_{\overline{X}C}$ we now consider:

$$\text{MSE}\delta_{\overline{X}i} = E\left(\sum_{i=1}^{p} \pi_i\left((1-c_i)(\overline{X}_i-\overline{X})+\overline{X}-\mu_i\right) \right)^2$$

$$=E\left(\sum_{i=1}^{p} \sum_{j\neq i}^{p} \pi_i\pi_j(c_j-c_i)(\overline{X}_i-\overline{X}_j)+(\overline{X}-\overline{\mu})\right)^2.$$

Again we must concentrate on solving for the minimizing values of the differences of shrinking factors. After differentiating, we get a system of $\binom{p}{2}$ equations from which we hope to find unique solutions for the p - 1 unknown values of c_i-c_j (i=j+1, j=1,2,...,p-1).

A typical equation in the set results from differentiating with respect to c_2-c_1:

$$2\pi_1^2\pi_2^2(c_2-c_1)E(\overline{X}_1-\overline{X}_2)^2$$

$$+\ 2\pi_1\pi_2E(\overline{X}_1-\overline{X}_2)(\overline{X}-\overline{\mu})$$

$$+\ 2\pi_1\pi_2\Big(\sum_{\substack{k\neq 1\\ k\neq 2}} \pi_1\pi_k(c_k-c_1)E(\overline{X}_1-\overline{X}_2)(\overline{X}_1-\overline{X}_k)\Big)$$

$$+\ 2\pi_1\pi_2\Big(\sum_{\substack{k\neq 1\\ k\neq 2}} \pi_2\pi_k(c_k-c_2)E(\overline{X}_1-\overline{X}_2)(\overline{X}_2-\overline{X}_k)\Big)$$

$$+\ 2\pi_1\pi_2\Big(\sum_{\substack{k<1\\ k,1\neq 1\\ k,1\neq 2}} \pi_k\pi_1(c_1-c_k)E(\overline{X}_1-\overline{X}_2)(\overline{X}_k-\overline{X}_1)\Big)=0.$$

Simple examples can be constructed to show that subsequent values of c_i-c_j (i=j+1) can be solved for and that they do not agree with the differences of the factors obtained from component-wise minimization. So the perfect relationship between the estimators $\delta_{\overline{X}i}$ and $\delta_{\overline{X}C}$ ceases for p>2.

In our previous discussion it was pointed out how $\delta_{\overline{X}i}$ uses the relationships of the different strata in determining the values of the differences of the shrinking factors. But when p=2, the expression $(1-c_i)(\overline{X}_i-\overline{X})+\overline{X}$ in $\delta_{\overline{X}C}$ immediately involves the special relationship between the two strata and hence $\delta_{\overline{X}C}$ is able to account for the cross-product terms in (2) in this way. But $(1-c_i)(\overline{X}_i-\overline{X})+\overline{X}$ does not sufficiently weigh the relationships between all the strata when p>2 , and hence there is a discrepancy between $\delta_{\overline{X}C}$ and $\delta_{\overline{X}i}$ for p>2 .

To utilize the above system of equations in solving for c_i-c_j (i=j+1), we must, in practice, substitute estimates for the unknown parameters and use a computer routine to solve for the factors. From this solution, the estimator $\delta_{\overline{X}i}$ can then be formed.

5. Monte Carlo Results for p=2

So far, we have discussed shrinkage estimators of $\overline{\mu}$ that perform better than the usual stratified sample mean. These estimators, however, depend upon unknown population parameters. In fact, they depend upon the stratum means which are essentially what we hope to estimate. In order to use these estimators in practice, sample statistics must be substituted for the unknown parameters. The ensuing estimators are not necessarily going to maintain all the good properties that we have discussed, but will still show merit themselves. Because the distribution of these more practical estimators are difficult to derive, we must depend upon a simulation study in order to judge performance.

Five thousand sets of sample stratum means and variances are generated (for the case p=2), from which the expected values of the estimators and their MSE's are approximated. We find that the simulated MSE's of the estimators, when stratum variances are assumed unknown, are only from 1 to 4 percent larger than the MSE's when only the stratum means are assumed unknown.

Let us first inspect the performance of $\delta_{\overline{X}i}$. First we check its behavior for different settings of stratum variances and weights. Then, from these results, we compare $\delta_{\overline{X}i}$ to both $\delta_{\theta C}$ and $\delta_{\theta i}$, where we assume (V_1,V_2) is known, $\theta_1=\theta_2=0$ and $n_1=n_2=10$. To do this we use the information on Table 1.

When $(\mu_1,\mu_2)=(0,2)$, the value of $\overline{X}_1$ will clearly be less than $\overline{X}_2$ most of the time, so the setting of the discussion on (4) and (5) exists. If we let the location of the star (*) on the graph show where the average simulation value of $\delta_{\overline{X}i}$ lies, we can see that this location is on the left side of the respective value of $\overline{\mu}$ when $\pi_2V_2-\pi_1V_1$ is positive and on the right otherwise. This agrees with our discussion involving (4).

Also note that the starred location is closer to 2 than 0 whenever π_2 is larger than 1/2. Furthermore, this location seems to be tugged to the left when V_1 is small and to the right when V_2 is small. This shows the estimator is favoring, to some degree, the stratum with smaller variance. On the fifth line the starred location is at .973 which is very reasonable considering that $\overline{\mu}$=1 and that the strata show balanced weighting with respect to the π_i's and the V_i's. The comments of this paragraph agree with our discussion involving expression (5).

When $(\pi_2V_2-\pi_1V_1)^2$ is large, $\delta_{\overline{X}i}$ does its best improving upon $V\overline{X}$ since $\text{MSE}\delta_{\overline{X}i}=V\overline{X}-(\pi_2V_2-\pi_1V_1)^2/\left(V_1+V_2+(\mu_1-\mu_2)^2\right)$. This is upheld in the simulation studies as columns 5 and 6 in Table 1 show. Note that, even under total balance of the strata, $\delta_{\overline{X}i}$ dominates $\overline{X}$.

Also note that the dotted lines give some indication of the bias of the estimator $\delta_{\overline{X}i}$.

To compare $\delta_{\overline{X}i}$ to the estimators $\delta_{\theta C}$ and $\delta_{\theta i}$, we see from columns 7 and 8 of Table 1 that $\delta_{\overline{X}i}$ usually performs worse than either of the other two. This result is not surprising considering that $\theta_1=\theta_2=0$ was used and that the true means $(\mu_1,\mu_2)=(2,0)$ are reasonably close to (0,0).

When the values of the means are not close to (0,0) relative to the variation within the strata, we can expect $\delta_{\theta C}$ and $\delta_{\theta i}$ to lose their dominance over $\delta_{\overline{X}i}$.

If we set $(\mu_1,\mu_2)=(3,2)$, $(V_1,V_2)=(.3,.6)$, $(\pi_1,\pi_2)=(\frac{1}{2},\frac{1}{2})$, $(\theta_1,\theta_2)=(0,0)$ and $(n_1,n_2)=(10,10)$, we find

$$\frac{\text{MSE}\delta_{\overline{X}i}}{\text{MSE}\delta_{\theta C}} = .8090 \text{ and } \frac{\text{MSE}\delta_{\overline{X}i}}{\text{MSE}\delta_{\theta i}} = .7526 .$$

In relation to (V_1,V_2), the stratum means are far from (θ_1,θ_2) and subsequently $\delta_{\overline{X}i}$ not only dominates $\overline{X}$ but also dominates $\delta_{\theta C}$ and $\delta_{\theta i}$.

If we set $(\mu_1,\mu_2) = (4,2)$ and $(V_1,V_2) = (48,60)$ while holding the other values the same as before, we find

$$\frac{\text{MSE}\delta_{\overline{X}i}}{\text{MSE}\delta_{\theta C}} = 1.7947 \text{ and } \frac{\text{MSE}\delta_{\overline{X}i}}{\text{MSE}\delta_{\theta i}} = 1.5841 .$$

Here the means are more distant from (θ_1,θ_2) than before, yet they are much closer relative to the stratum variances.

To more graphically represent the performance of these estimators, Figure 1 considers the case of equal stratum means, $(V_1,V_2) = (3.75,5)$, $(\pi_1,\pi_2) = (1/2,1/2)$ and $(n_1,n_2) = (8,12)$. If we choose to utilize Neyman allocation for these choices of (V_1,V_2) and (π_1,π_2) , we would find (8,12) to be the optimal choice of (π_1,π_2) . Therefore, $\overline{X}$ not only represents the classical stratified estimator, but also the optimal choice of $\overline{X}$ based upon Neyman allocation. From this graph we now can make clear statements on the performances of the estimators.

It was known previous to the simulation work that $\delta_{\overline{X}i}$ dominates $\overline{X}$ and that $\delta_{\theta i}$ dominates $\delta_{\theta C}$ which dominates $\overline{X}$. From Figure 1 it is clear that the estimators perform much worse when (μ_1,μ_2) is assumed unknown. But also assuming unknown stratum variances adds very little further loss.

As to be expected, $\delta_{\theta i}$ and $\delta_{\theta C}$ perform well when (μ_1,μ_2) is near the value $(\theta_1,\theta_2) = (0,0)$.

Also note that the MSE $\delta_{\overline{X}i}$ does not change at all as the common value of μ changes.

Since $\text{MSE}(\delta_{\overline{X}i}) = V(\overline{X}) - \dfrac{(\pi_2V_2-\pi_1V_1)^2}{V_1+V_2+(\mu_1-\mu_2)^2}$ and Figure 1 assumes $\mu_1=\mu_2$ and constant values for the stratum variances and weights, this result is expected.

Finally, note that the simple Thompson estimator $(1-c)(\overline{X}-\overline{\theta})+\overline{\theta}$ does as well or better than many of the estimators in this paper for many values of the common mean μ . Again it must be pointed out that the stratum weights and variances are either equal or near equal. The advantage of these estimators over Thompson's is that a different shrinking factor is employed for every stratum. So one would expect the best improvement when the strata are quite different in weights and variances. Columns 7,8,and 9 on Table 1 show clearly that the Thompson estimator of $\overline{\mu}$ is consistently beaten when the strata are reasonably different in this way.

6. Conclusion

Our estimators do improve upon $\overline{X}$ for at least a portion of the parameter space. The estimators $\delta_{\theta i}$ and $\delta_{\theta C}$ perform well if the value of (θ_1,θ_2) is relatively near the actual (μ_1,μ_2) . The estimators also fare well against a Thompson-type estimator of $\overline{\mu}$ and even the estimator $\overline{X}$ under Neyman allocation.

7. References

1 Albert, J.H.(1979) Robust Bayes estimation. Mimeograph Series #79-9. Department of Statistics, Purdue University.

2 James, W. and Stein, C.(1961) Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob 1 361-379. Univ. of California Press.

3 Thisted, R.A.(1976) Ridge regression minimax estimation, and empirical Bayes methods. Technical Report No. 28, Public Health Service Grant 5R01GM21215-02. Department of Biostatistics, Stanford University.

4 Thompson, J.R.(1968) Some shrinking techniques for estimating the mean. Journal of the American Statistical Association 63, 113-122.

TABLE 1 MSE OF SHRINKAGE ESTIMATORS RELATIVE TO THE USUAL ESTIMATOR UNDER DIFFERENT CONDITIONS

(μ_1,μ_2)	(V_1,V_2)	(π_1,π_2)	$\bar{\mu}$	$(\pi_2V_2-\pi_1V_1)^2$	$\frac{MSE\delta_{\bar{X}i}}{V\bar{X}}$	$\frac{MSE\delta_{\theta i}}{V\bar{X}}$	$\frac{MSE\delta_{\theta C}}{V\bar{X}}$	$\frac{MSE\delta_{\theta}}{V\bar{X}}$
(0,2)	(5.0,1.0)	$(\frac{1}{4},\frac{3}{4})$	1.5	$\frac{4}{16}$	.9589	.9841	.9238	1.0520
(0,2)	(3.0,3.0)	$(\frac{1}{4},\frac{3}{4})$	1.5	$\frac{36}{16}$	.9211	.7925	.8123	.8842
(0,2)	(1.0,5.0)	$(\frac{1}{4},\frac{3}{4})$	1.5	$\frac{196}{16}$	.7708	.6619	.7140	.7543
(0,2)	(1.0,5.0)	$(\frac{1}{2},\frac{1}{2})$	1	$\frac{64}{16}$	.8487	.6727	.6774	.7025
(0,2)	(3.0,3.0)	$(\frac{1}{2},\frac{1}{2})$	1	$\frac{0}{16}$	.0750	.7004	.6577	.7025

LINE GRAPH

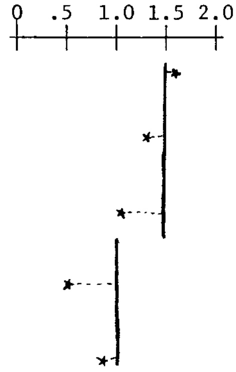

FIGURE 1
MSE OF SHRINKAGE ESTIMATORS FOR THE STRATIFIED MEAN for $\mu_1=\mu_2=\mu$ and $V_1=3.75$, $V_2=5$