DOCUMENT RESUME

ED 079 400 TM 003 003

AUTHOR Stroud, T. W. F.
TITLE Combining Unbiased Estimates of a Parameter Known to be Positive.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-47
PUB DATE Jun 73
NOTE 28p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Calculation; Mathematical Applications; *Mathematical Models; Measurement Techniques; *Statistical Analysis; *Statistical Bias; Statistics

ABSTRACT

The statistician has n independent estimates of a parameter he knows is positive, but, as is the case in components-of-variance problems, some of the estimates may be negative. If the n estimates are to be combined into a single number, we compare the obvious rule, that of averaging the n values and taking the positive part of the result, with that of averaging the positive parts. Although the estimator generated by the second rule is not consistent, it is shown by numerical calculation that for small n it has a smaller mean square error than the first over a considerable region of the parameter space, and that for $n = 2$ or 3 the second is minimax relative to the first over a region consisting of almost the whole parameter space. The distribution of each of the n estimates is assumed to be either Gaussian or the distribution of a weighted difference of two independent chi-squares with known degrees of freedom, as in one-way components of variance. Some other simply calculated estimators, including the positive part of the median, are studied for the chi-square difference case with (2,2) degrees of freedom and $n = 3$. (Author)

RB-73-47

RESEARCH BULLETIN

COMBINING UNBIASED ESTIMATES OF A PARAMETER KNOWN TO BE POSITIVE

T. W. F. Stroud
Queen's University
and
Educational Testing Service

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
June 1973

COMBINING UNBIASED ESTIMATES OF A PARAMETER KNOWN TO BE POSITIVE

T. W. F. Stroud*

ABSTRACT

The statistician has n independent estimates of a parameter he knows is positive but, as is the case in components-of-variance problems, some of the estimates may be negative. If the n estimates are to be combined into a single number, we compare the obvious rule, that of averaging the n values and taking the positive part of the result, with that of averaging the positive parts. Although the estimator generated by the second rule is not consistent, it is shown by numerical calculation that for small n it has a smaller mean square error than the first over a considerable region of the parameter space, and that for $n = 2$ or 3 the second is minimax relative to the first over a region consisting of almost the whole parameter space. The distribution of each of the n estimates is assumed to be either Gaussian or the distribution of a weighted difference of two independent chi-squares with known degrees of freedom, as in one-way components of variance. Some other simply calculated estimators, including the positive part of the median, are studied for the chi-square difference case with (2,2) degrees of freedom and $n = 3$.

COMBINING UNBIASED ESTIMATES OF A PARAMETER KNOWN TO BE POSITIVE

1. INTRODUCTION

Sometimes, most notably in components-of-variance problems, a statistician is trying to estimate a parameter he knows is positive, but in using standard unbiased estimation techniques he obtains an estimate which is negative. When this happens, the statistician would usually estimate the parameter by zero, since in so doing he is coming closer to the true parameter value than the original estimate.

Occasionally the statistician may have several unbiased estimates of the same parameter arising from independent sources. The statistician will want to combine the raw data sets and treat them as one data set; however, this may not always be possible. There may be nuisance parameters which vary from one source to another. Or the computational procedure may require too much memory to accommodate all the data at once. This may happen in using Rao's MINQUE technique [8], where matrices the size of the data set are handled, or Henderson's third method [4,9] with a large number of groups. (For a problem where either or both of these techniques are called for, see [11]). Finally, the statistician may not be able to combine the observations due to not being supplied with the raw data.

Let us assume that from each source the statistician has nothing more than an unbiased estimate (which can be positive or negative) of a common unknown parameter value which is necessarily positive. We study alternative procedures for combining the estimates into a single number. The estimates are assumed to be identically distributed. (This does not cover the situation of nuisance parameters varying over experiments, but, when caution is used, some features of the results reported here may still apply.)

The main part of this paper is devoted to studying two simple techniques for combining the estimates. A few other procedures are treated in a limited way in the last section. We study the performances of the two techniques when the distribution of the estimates is normal and when it is the weighted difference of two independent chi-squares with known degrees of freedom. The latter model is the correct one for one-way components of variance, and may be considered as a prototype for more complicated problems. As the degrees of freedom vary, a range of distributional shapes is obtained. The normal distribution is, of course, the limiting case as both degrees of freedom become large.

2. THE TWO ESTIMATORS AND THEIR MEAN SQUARE ERRORS

Let $X_1, X_2, \ldots, X_n$ be a set of independent random variables, each member representing the estimate of the unknown parameter μ based on one source. The X_i are assumed to be identically distributed with expectation $\mu = E(X_i)$ and one nuisance parameter which is taken to be the standard deviation $\sigma = \sigma(X_i)$. We know that μ is greater than or equal to zero, although any X_i has a positive probability of being negative.

Denote by X_i^+ the positive part of X_i , i.e., the random variable which equals X_i when $X_i > 0$ and which equals zero otherwise. When $n = 1$, X_1^+ is the obvious estimator of μ . When $n > 1$, one can either average the X_i values and then take the positive part or take the positive parts before averaging. Denote $(\bar{X})^+$, the estimator obtained by the first method, by $\hat{\mu}$, since it is the maximum-likelihood estimator under normality. We denote the second estimator $(\Sigma X_i^+)/n$ by $\bar{\mu}$, since it is the arithmetic mean of a set of quantities. These two estimators were considered by Sirotnik [10] in a mental testing application.

We note that $\bar{\mu}$ is not consistent, since it converges in probability to $E(X_i^+)$ which is always greater than μ, i.e., $E(X_i^+) =$
$P(X_i \geq 0)E(X_i|X_i \geq 0) > P(X_i \geq 0)E(X_i|X_i \geq 0) + P(X_i < 0)E(X_i|X_i < 0) = E(X_i)$.
One's first reaction to this fact is that $\hat{\mu}$ has better properties and is thus the preferred estimator. This is certainly true when n is large, or even moderate. The reason for presenting this article is that, for small n, $\bar{\mu}$ performs better over a reasonably large portion of the parameter space, assuming as we do that the X_i are either Gaussian or the weighted difference of two independent chi-squares.

The comparison of the estimators' performances is based on mean square error (MSE), which is defined as the expected squared difference between the parameter and its finally estimated value, and which equals the variance plus the square of the bias. It will be seen that formulas for the MSE contain a proportionality factor of σ^2; for this reason the comparisons to be presented are in terms of MSE/σ^2.

Insert Table 1 about here

Table 1 is a summary of the relative performance of $\bar{\mu}$ and $\hat{\mu}$, as measured by MSE/σ^2. When $\mu/\sigma > 1/2$, $\bar{\mu}$ is generally better, but the advantage becomes minimal as n increases, and the critical value of μ/σ such that $\bar{\mu}$ is as good as $\hat{\mu}$ is increasing in n and eventually exceeds 1/2. The value of MSE/σ^2 for both $\hat{\mu}$ and $\bar{\mu}$ increases to $1/n$ as μ/σ approaches its theoretical upper bound[1] M. However, if we restrict consideration to the region $\mu/\sigma \leq K$, for any $K < M$, the maximum MSE/σ^2 of $\bar{\mu}$ over this region is less than that of $\hat{\mu}$ when $n = 2$ or 3 and, depending on the distribution of X_i, sometimes also when $n = 4$ or 5.

The relative advantage of $\bar{\mu}$ over $\hat{\mu}$ when $\mu/\sigma > 1/2$ also depends on the distribution of the X_i. In the next section some numerical results are presented for the normal case and for the chi-square difference with certain specified degrees of freedom. From these results we shall see in detail under what conditions $\bar{\mu}$ is the preferred estimator.

3. NUMERICAL CALCULATIONS AND GRAPHS

When n is large, $\bar{\mu}$ is clearly a poor competitor to $\hat{\mu}$, since in this case $\hat{\mu}$ will be close to μ and $\bar{\mu}$ will be close to a number known to be larger than μ. Let us therefore see what happens for n ranging from 2 to 5, and to start with assume the X_i to be normally distributed. The calculations are obtained from formulas (4.3) and (4.4). We note that for both $\hat{\mu}$ and $\bar{\mu}$ the MSE equals σ^2 times a function of the standardized mean $m = \mu/\sigma$ (the reciprocal of the coefficient of variation). For this reason we graph MSE/σ^2 as a function of m ; we use m rather than $1/m = \sigma/\mu$ to display results for values of m close to zero, since the behavior as $m \to \infty$ is obvious.

Insert Figures A and B about here

To separate the curves, those for $n = 2$ and 4 are shown in Figure A and for $n = 3$ and 5 in Figure B.

It is seen that $\bar{\mu}$ enjoys a healthy advantage over $\hat{\mu}$ in the region $1/2 < \mu/\sigma < 2$. When μ is much smaller than σ the bias of $\bar{\mu}$ has a noticeable effect, and when μ is much larger the chances of a negative X_i are slight so that the values of both $\hat{\mu}$ and $\bar{\mu}$ will usually coincide with $\bar{X}$. Note that for the estimator $\bar{X}$ the value of MSE/σ^2 is the constant $1/n$. This value is approached by the curves for both $\hat{\mu}$ and $\bar{\mu}$ as m gets large, but the $\hat{\mu}$ curve gets there faster.

Insert Table 2 about here

Table 2 indicates the behavior of the MSE functions for larger values of n . It is clear that for n as large as 9, the MSE/σ^2 of $\bar{\mu}$ is so much greater at $m = 0$ than the MSE/σ^2 of $\hat{\mu}$ is anywhere that one would want to avoid using $\bar{\mu}$ for $n \geq 9$ even though it is better than $\hat{\mu}$ for some values of m .

This gives the general picture when the X_i are normal; for very small n there are certain advantages to $\bar{\mu}$, but as n increases these are outweighed by a very high MSE/σ^2 in the neighborhood of $m = 0$. The next question is, do the small-sample advantages of $\bar{\mu}$ persist when the X_i are distributed other than normally?

In the one-way components of variance problem, the usual unbiased estimator X of the main effect variance component has the distribution of $\theta_1\chi_1^2 - \theta_2\chi_2^2$, where χ_1^2 and χ_2^2 are independent chi-square random variables with known degrees of freedom f_1 and f_2, respectively, and θ_1 and θ_2 are positive numbers related to the unknown variance components and which satisfy the inequality $E(X) = \theta_1 f_1 - \theta_2 f_2 \geq 0$. There are four cases to consider: low f_1 and low f_2, low f_1 and high f_2, high f_1 and low f_2, and high f_1 and high f_2. The low and high degrees of freedom have been chosen as 2 and 20. Other even values are easily treated using the formulas of the next section. Odd-numbered values are much more difficult mathematically, but by continuity we expect the characteristics to be similar to the adjacent even values.

Insert Figures C and D about here

Figures C and D show, respectively, some MSE/σ^2 curves for $f_1 = 2$, $f_2 = 20$ and for $f_1 = 20$, $f_2 = 2$. The curves for $f_1 = f_2 = 2$ have been examined and their appearance is similar to Figure C, with the $\bar{\mu}$ curve having less upsweep near $m = 0$. Some values were also calculated for $f_1 = f_2 = 20$, and they were reasonably close to the values for the normal case (see Table 3). The conclusion may be drawn that when f_1 is high the estimator $\bar{\mu}$ is considerably better than $\hat{\mu}$ for moderately high values of m, but that when f_1 is low the advantage is so meager that it does not seem worth the risk near $m = 0$.

Insert Table 3 about here

To summarize, in problems where the X_i can be considered to be a weighted difference of two independent chi-squares, we would recommend $\bar{\mu}$ over $\hat{\mu}$ if n is quite small, the degrees of freedom of the first chi-square is fairly large, and it is thought likely that $m \geq 1/2$, i.e., that the mean of X_i is at least half its standard deviation.

In a one-way variance components model where it is desired to combine estimates, from n independent sources, of the common variance components σ_b^2 (between) and σ_w^2 (within), this means we recommend that negative estimates $\hat{\sigma}_b^2$ of σ_b^2 be replaced by zero before averaging provided n is small, the number of levels k in the one-way classification is large, and the ratio $\sigma_b^2/(\mathrm{var}\ \hat{\sigma}_b^2)^{\frac{1}{2}}$ is at least 1/2. From [2, page 322, formula (5.7)], the latter condition in the balanced case is equivalent to $\sigma_b^2/\sigma_w^2 \geq [\ell^2 + \ell\{\ell^2 + (kr - 1)/(kr - k)\}^{\frac{1}{2}}]/r$, where r is the number of observations per level (within one source) and $\ell = (2k - 3)^{-\frac{1}{2}}$. A more simply calculated quantity which exceeds the above lower bound for σ_b^2/σ_w^2, provided $k \geq 3$ and $r \geq 2$, is $2\ell/r$. If $k = 26$ and $r = 10$, for example, $\bar{\mu}$ is preferable to $\hat{\mu}$ when $\sigma_b^2 \geq \sigma_w^2/35$ (and n is small).

4. DERIVATION OF FORMULAS FOR MEAN SQUARE ERROR

We now indicate how the values for the curves discussed in Section 2 were calculated.

As before, denote by μ and σ^2 the mean and variance, respectively, of the X_i, and for any U denote $U^+ = \max\{U,0\}$. We wish to derive formulas for the MSE of $\hat{\mu} = \bar{X}^+$ and $\bar{\mu} = (\Sigma X_i^+)/n$. These can be obtained from expressions for the mean and variance of $\bar{X}^+$ and of X_i^+, since the MSE of any estimator of μ with mean ν and variance τ^2 is given by $\tau^2 + (\nu - \mu)^2$.

If F is the c.d.f. of X_i, the mean ν and variance τ^2 of $\bar{\mu}$ are given by

$$\nu = E(X_i^+) = \int_0^\infty x\,dF(x) \quad , \tag{4.1}$$

$$n\tau^2 + \nu^2 = E\{(X_i^+)^2\} = \int_0^\infty x^2\,dF(x) \quad . \tag{4.2}$$

For the normal case, let $z = (x - \mu)/\sigma$; then

$$\int_0^\infty x\,dF(x) = \int_{-\mu/\sigma}^\infty (\mu + \sigma z)\,d\Phi(z) \quad ,$$

$$\int_0^\infty x^2\,dF(x) = \int_{-\mu/\sigma}^\infty (\mu^2 + 2\mu\sigma z + \sigma^2 z^2)\,d\Phi(z) \quad ,$$

where Φ is the standard normal c.d.f. The integration is straightforward, using integration by parts on $z^2 d\Phi(z)$. Hence $\nu_2 = \sigma[m\Phi(m) + \phi(m)]$ and $n\tau^2 = \sigma^2\lambda_m$, where $\phi(m) = (2\pi)^{-\frac{1}{2}} \exp(-m^2/2)$ and $\lambda_m = m^2\Phi(m)\{1 - \Phi(m)\} - m\phi(m)\{2\Phi(m) - 1\} + \Phi(m) - \{\phi(m)\}^2$. Finally the MSE for $\bar{\mu}$ is

$$\sigma^2\{(\lambda_m/n) + \delta_m^2\} \quad , \tag{4.3}$$

where $\delta_m = \phi(m) - m\{1 - \Phi(m)\}$. Since $\bar{X}$ has a normal distribution, the MSE for $\hat{\mu}$ may be derived similarly; its value is

$$(\sigma^2/n)\{\lambda_{m\sqrt{n}} + \delta^2_{m\sqrt{n}}\} \quad . \tag{4.4}$$

The distribution of a weighted difference of two independent chi-squares with even degrees of freedom can be represented as a finite mixture of positive and negative chi-squares. This result has been derived variously by Box [1], Mantel and Pasternack [6], and Jayachandran and Barr [5]. Using the notation of Section 3 with $p = f_1/2$ and $q = f_2/2$ as integers the density of $X = \Theta_1\chi_1^2 - \Theta_2\chi_2^2$ can be written as

$$f_X(x) = \sum_{j=1}^{p} \binom{p + q - j - 1}{q - 1} \frac{r^{p-j}}{(1 + r)^{p+q-j}} g_j(x)$$

$$+ \sum_{k=1}^{q} \binom{p + q - k - 1}{p - 1} \frac{r^p}{(1 + r)^{p+q-k}} h_k(x) \tag{4.5}$$

$$= \Sigma_j c_j(r;p,q)g_j(x) + \Sigma_k d_k(r;p,q)h_k(x) \quad ,$$

say, where $r = \Theta_2/\Theta_1$, $g_j(x)$ is the density of Θ_1 times a chi-square with $2j$ degrees of freedom and $h_k(x)$ is the density of $(-\Theta_2)$

times a chi-square with $2k$ degrees of freedom. Note that the $g_j(x)$ are nonzero only when x is positive and the $h_k(x)$ are nonzero only when x is negative. Since (4.1) and (4.2) involve integration over only positive values of x, we need consider only the terms containing $g_j(x)$. The expectation of X_i^+ is therefore the weighted sum of expectations of $\theta_1 \chi_{2j}^2$ random variables with the weights given by the coefficients of the $g_j(x)$ in (4.5), and the second moment of X_i^+ is a similar weighted sum of second moments of these $\theta_1 \chi_{2j}^2$. Using the fact that $\sigma^2 = 4\theta_1^2(p + r^2 q)$,

$$E(X_i^+) = \sum_{j=1}^{p} c_j(r;p,q)(2j\theta_1)$$

$$= \sigma(p + r^2 q)^{-\frac{1}{2}} \sum_{j=1}^{p} j c_j(r;p,q)$$

and

$$E\{(X_i^+)^2\} = [\sigma^2/(p + r^2 q)] \sum_{j=1}^{p} j(j + 1)c_j(r;p,q) \quad .$$

These formulas were used to compute MSE/σ^2 for the estimator $\bar{\mu}$ as a function of r. The quantity MSE/σ^2 was then plotted as a function of m by obtaining m from the monotonically decreasing relation $m = (p - rq)(p + r^2 q)^{-\frac{1}{2}}$ which follows from $\mu = 2\theta_1(p - rq)$ and $\sigma^2 = 4\theta_1^2(p + r^2 q)$. Since $\mu \geq 0$, the maximum value of r allowed is p/q; also $r > 0$ implies that $m < p^{\frac{1}{2}} = (f_1/2)^{\frac{1}{2}}$.

Note that the distribution of $\bar{X}$ has the representation $\theta_3 \chi_3^2 - \theta_4 \chi_4^2$, where $\theta_3 = \theta_1/n$, $\theta_4 = \theta_2/n$, and χ_3^2 and χ_4^2 have nf_1 and nf_2 degrees of freedom, respectively. Thus MSE/σ^2 may be calculated for $\hat{\mu}$ in the same manner as for $\bar{\mu}$, with the appropriate substitutions.

5. OTHER POSSIBLE ESTIMATORS

In this section we look at a few other naturally suggested estimators of simple form. Motivated by the results just stated, the author decided to seek a simply calculated compromise between $\hat{\mu}$ and $\bar{\mu}$, one that does better than $\hat{\mu}$ for larger values of $m = \mu/\sigma$ but does not have the characteristically high MSE that $\bar{\mu}$ has for values of m near zero. It is appealing to try to use the sample information to get some idea of m and then choose $\hat{\mu}$ or $\bar{\mu}$ accordingly. An easy rule to use would be based on the proportion of X_i which are negative. If this proportion is high, it indicates a greater likelihood of low values of μ/σ, and we would perhaps be predisposed to use the rule that performs best for these low values.

For brevity we restrict ourselves to the model $\theta_1 X_1^2 - \theta_2 X_2^2$ where both X_1^2 and X_2^2 have just two degrees of freedom. This turns out to be the easiest case mathematically. From (4.5),

$$f_x(x) = \{\theta_1 g_1(x) + \theta_2 h_1(x)\}/(\theta_1 + \theta_2) \quad .$$

Here X_i is distributed as a mixture of a $\theta_1 X^2$ and a $-\theta_2 X^2$ random variable, each with 2 degrees of freedom, with weights proportional to θ_1 and θ_2 respectively. Regard the sampling as coming from an urn with "positive" and "negative" balls in this proportion. For a small sample it is easy to write the probabilities of observing J negative balls (J taking values from 0 to n) and to compute the mean square error for any rule which chooses $\hat{\mu}$ or $\bar{\mu}$ according to the value of J. Letting Y stand for the resulting estimator,

$$\mathrm{MSE} = E(Y - \mu)^2 = \Sigma_{j=0}^{n} \Pr\{J = j\}E\{(Y - \mu)^2 | J = j\} \quad .$$

Taking as an example $n = 3$, by symmetry we can write

$$E\{(Y - \mu)^2 | J = 1\} = E\{(Y - \mu)^2 | X_1 < 0,\ X_2 > 0,\ X_3 > 0\} \quad .$$

When $X_1 < 0$, $X_2 > 0$, and $X_3 > 0$ we have $X_1^+ = 0$, $X_2^+ = X_2$, and $X_3^+ = X_3$, so that here $\bar{\mu} = (X_2 + X_3)/3$, where X_2 and X_3 have the conditional distribution of two independent $\theta_1 \chi^2$ random variables, each with 2 degrees of freedom. Hence $E\{\bar{\mu} | J = 1\} = 4\theta_1/3$ and $\mathrm{Var}\{\bar{\mu} | J = 1\} = 8\theta_1^2/9$. The formula for $E\{(\bar{\mu} - \mu)^2 | J = 1\}$ follows directly. The conditional distribution of $\bar{X}$, given that $J = 1$, is that of $(\theta_1/3)\chi_a^2 - (\theta_2/3)\chi_b^2$, where χ_a^2 and χ_b^2 are independent chi-squares with 4 and 2 degrees of freedom, respectively. Thus the first and second moments of $\hat{\mu}$ may be calculated along similar lines to those indicated in Section 4. The $J = 2$ case is similar, and the $J = 0$ and $J = 3$ cases are trivial.

If χ_1^2 or χ_2^2 have an even number of degrees of freedom more than two, the same general principles may be used, but the distributions, given whether positive or negative, are no longer pure chi-squares but mixtures of chi-squares. The computations for $\hat{\mu}$ then become rather involved.

When $n = 3$ there are just 4 possible rules of the type described above, based on choice of $\hat{\mu}$ or $\bar{\mu}$ when $J = 1$ and when $J = 2$, two of which are simply $\hat{\mu}$ and $\bar{\mu}$. (When $J = 0$ or 3, $\hat{\mu}$ and $\bar{\mu}$ yield the same value.) Let $\bar{\mu}_1$ be the estimator obtained by choosing $\bar{\mu}$ when $J = 1$ and $\hat{\mu}$ when $J = 2$. (This is the rule based on the heuristic likelihood argument of the first paragraph.) Let $\bar{\mu}_2$ be the estimator obtained by choosing $\bar{\mu}$ when $J = 2$ and $\hat{\mu}$ when $J = 1$.

Insert Table 4 about here

For selected values of $r = \theta_2/\theta_1$, Table 4 gives the corresponding value of $m = \mu/\sigma$ and MSE/σ^2 for the rules $\hat{\mu}$, $\bar{\mu}$, $\bar{\mu}_1$ and $\bar{\mu}_2$, in the case $f_1 = f_2 = 2$. Surprisingly, the rule $\bar{\mu}_1$ which chooses $\hat{\mu}$ or $\bar{\mu}$ according to the relative likelihood of low or high values of m performs badly, and the rule $\bar{\mu}_2$ which reverses the choice does relatively well. $\bar{\mu}_2$ has a lower MSE than $\hat{\mu}$ for all m greater than about .28 and a lower MSE than either $\hat{\mu}$ or $\bar{\mu}$ for m between .28 and .55. At $m = 0$, where $\bar{\mu}$ loses to $\hat{\mu}$ by .083, the compromise estimator $\bar{\mu}_2$ is worse than $\hat{\mu}$ by .031. $\bar{\mu}_2$ is thus a much more acceptable alternative to $\hat{\mu}$ than is $\bar{\mu}$.

We present a possible explanation for the fact that the MSE for $\bar{\mu}_2$ is usually lower than for $\bar{\mu}_1$. When two out of three of the X_i are negative, we know that these X_i are less than μ, and hence that $\bar{X}$ is probably an underestimate of μ. The estimator $\hat{\mu} = \bar{X}^+$ compensates for this to some extent by frequently yielding zero in these cases. This is good if μ is close to zero. But for larger μ (i.e., $\mu > .28\sigma$), the estimator $\bar{\mu}$ which is strictly positive in these cases seems to perform better.

When only one X_i is negative, on the other hand, $\bar{X}$ will usually be positive and as such will usually be closer to μ than will the biased $\bar{\mu}$. This holds unless μ/σ is quite high (> .55), where now even one negative X_i is unlikely and when it occurs it is evidence that $\bar{X}$ will underestimate μ, so again the estimate should be raised.

It should be pointed out that this article is written from the frequentist point of view, according to which expectations are based on repeated sampling with the same parameter values. This viewpoint has been challenged by many statisticians (see, e.g., [7]), and alternative criteria of performance might conceivably yield different results.

Also included in Table 4 are results for $\tilde{\mu}$, the positive part of the sample median. Its mean and variance are easily calculated in the $f_1 = f_2 = 2$ case since here $\theta_1\chi_1^2$ is exponentially distributed, and using the no-memory property the order statistics are directly expressed as convolutions of exponential random variables; see, e.g., [3, p. 55, Prop. 3]. $\tilde{\mu}$ performs better than its competitors when m is small and worse when m is large. In the case of large m the distribution of X_i is markedly skewed, and we would thus expect the sample median to be centered around the population median, but not the population mean. As $m \to 0$ the distribution of $X_{i\beta}$ approaches the double exponential, for the case $f_1 = f_2 = 2$, for which the sample median is the maximum-likelihood estimator of the center of symmetry and is known to have good properties. This feature would not be expected to be present throughout all the distributions for X_i studied here.

6. CONCLUSION

We have compared two simple rules for the problem of combining independent, identically distributed, unbiased estimates of a parameter value known to be positive when the estimates may be negative. The first rule $\hat{\mu}$ (the positive part of the average of the estimates) is consistent and the second rule $\bar{\mu}$ (the average of the positive parts) is inconsistent. In large samples there is of course no difficulty in choosing between the two. We have seen that when n is very small the relative performance depends both on the ratio μ/σ (when $\mu/\sigma > 1/2$, $\bar{\mu}$ is generally better; otherwise $\hat{\mu}$ is better) and on the underlying distribution. A rule of thumb for when to use $\bar{\mu}$ in practice is suggested in the last two paragraphs of Section 3; however, this requires a prior idea of μ/σ and of the underlying family of distributions. A good practical rule is not obvious for situations where no such prior knowledge exists.

For the case of the chi-square difference with (2,2) degrees of freedom and $n = 3$, the positive part of the sample median performs very well when μ/σ is small, but badly when μ/σ is large. The estimator $\bar{\mu}_2$ (choose $\bar{\mu}$ if two of the independent unbiased estimates are negative) performs better than the others in the intermediate region and also reasonably well for large μ/σ.

All estimators studied here have a positive probability of being zero. It would be desirable to have a simply calculated estimator with the property of always being (strictly) positive, or at least of being positive whenever one or more of the X_i is positive. To date no reasonable method of obtaining such an estimator seems to be available.

FOOTNOTES

*T. W. F. Stroud is assistant professor, Department of Mathematics, Queen's University, Kingston, Ontario, Canada. This research was performed while the author was visiting research fellow, Psychometric Group, Educational Testing Service, Princeton, New Jersey. The author is grateful for the comments of Donald B. Rubin and Robert I. Jennrich. The author wishes to thank Michael W. Browne for suggesting the estimator $\bar{\mu}$, and Frederic M. Lord for some of the references.

[1]For the normal case $M = +\infty$, and for the weighted difference of two chi-squares M is the square root of half the degrees of freedom of the positive chi-square (see Section 4, second last paragraph).

REFERENCES

[1] Box, G. E. P., "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification," Annals of Mathematical Statistics, 25 (June 1954), 290-302.

[2] Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering, 2nd ed., New York: John Wiley & Sons, Inc., 1965.

[3] Chernoff, H., Gastwirth, J. L. and Johns, M. V., Jr., "Asymptotic Distribution of Linear Combinations of Functions of Order Statistics with Applications to Estimation," Annals of Mathematical Statistics, 38, (February 1967), 52-72.

[4] Henderson, C. R., "Estimation of Variance and Covariance Components," Biometrics, 9 (June 1953), 226-52.

[5] Jayachandran, Toke and Barr, D. R., "On the Distribution of a Difference of Two Scaled Chi-Square Random Variables," American Statistician, 24 (December 1970), 29-30.

[6] Mantel, Nathan and Pasternack, Bernard S., "Light Bulb Statistics," Journal of the American Statistical Association, 61 (September 1966), 633-39.

[7] Pierce, Donald A., "On Some Difficulties in a Frequency Theory of Inference," Annals of Statistics, 1 (March 1973), 241-50.

[8] Rao, C. Radhakrishna, "Estimation of Variance and Covariance Components in Linear Models," Journal of the American Statistical Association, 67 (March 1972), 112-15.

[9] Searle, Shayle R., Linear Models, New York: John Wiley & Sons, Inc., 1971.

[10] Sirotnik, Ken, "An Analysis of Variance Framework for Matrix Sampling," Educational and Psychological Measurement, 30 (Winter 1970), 891-908.

[11] Stroud, Thomas W. F., "Forecasting a Regression Relationship Which Varies over a Large Number of Subpopulations," Research Bulletin 73-00. Princeton, N.J.: Educational Testing Service, 1973.

1. SUMMARY OF PERFORMANCE OF $\bar{\mu}$ RELATIVE TO $\hat{\mu}$

	n			
	2,3	4,5	6-8	≥ 9
$\frac{\mu}{\sigma} > \frac{1}{2}$	$\bar{\mu}$ better	$\bar{\mu}$ better	$\bar{\mu}$ better	very close
$\frac{\mu}{\sigma} < \frac{1}{2}$	$\bar{\mu}$ worse	$\bar{\mu}$ worse	$\bar{\mu}$ worse	$\bar{\mu}$ much worse
estimator with lower maximum[a]	$\bar{\mu}$	depends on distribution of X_i	$\hat{\mu}$	$\hat{\mu}$

[a]"Maximum" refers to maximum MSE/σ^2 over any region of the form $0 \leq \mu/\sigma \leq K < M$, where M is the largest possible value of μ/σ.

2. COMPARISON OF MSE/σ^2 FOR $\hat{\mu}$ AND $\bar{\mu}$ (NORMAL CASE, $\sigma^2 = 1$)

	n = 6		n = 9		n = 12	
m	$\hat{\mu}$	$\bar{\mu}$	$\hat{\mu}$	$\bar{\mu}$	$\hat{\mu}$	$\bar{\mu}$
0	.0833	.2160	.0556	.1970	.0417	.1876
0.25	.1047	.1562	.0750	.1314	.0595	.1191
0.50	.1374	.1314	.0988	.1006	.0774	.0852
0.75	.1572	.1268	.1087	.0903	.0826	.0720
1.00	.1645	.1321	.1108	.0904	.0833	.0695
1.25	.1663	.1407	.1111	.0946	.0833	.0716
1.50	.1666	.1489	.1111	.0996	.0833	.0749
2.00	.1667	.1601	.1111	.1068	.0833	.0801

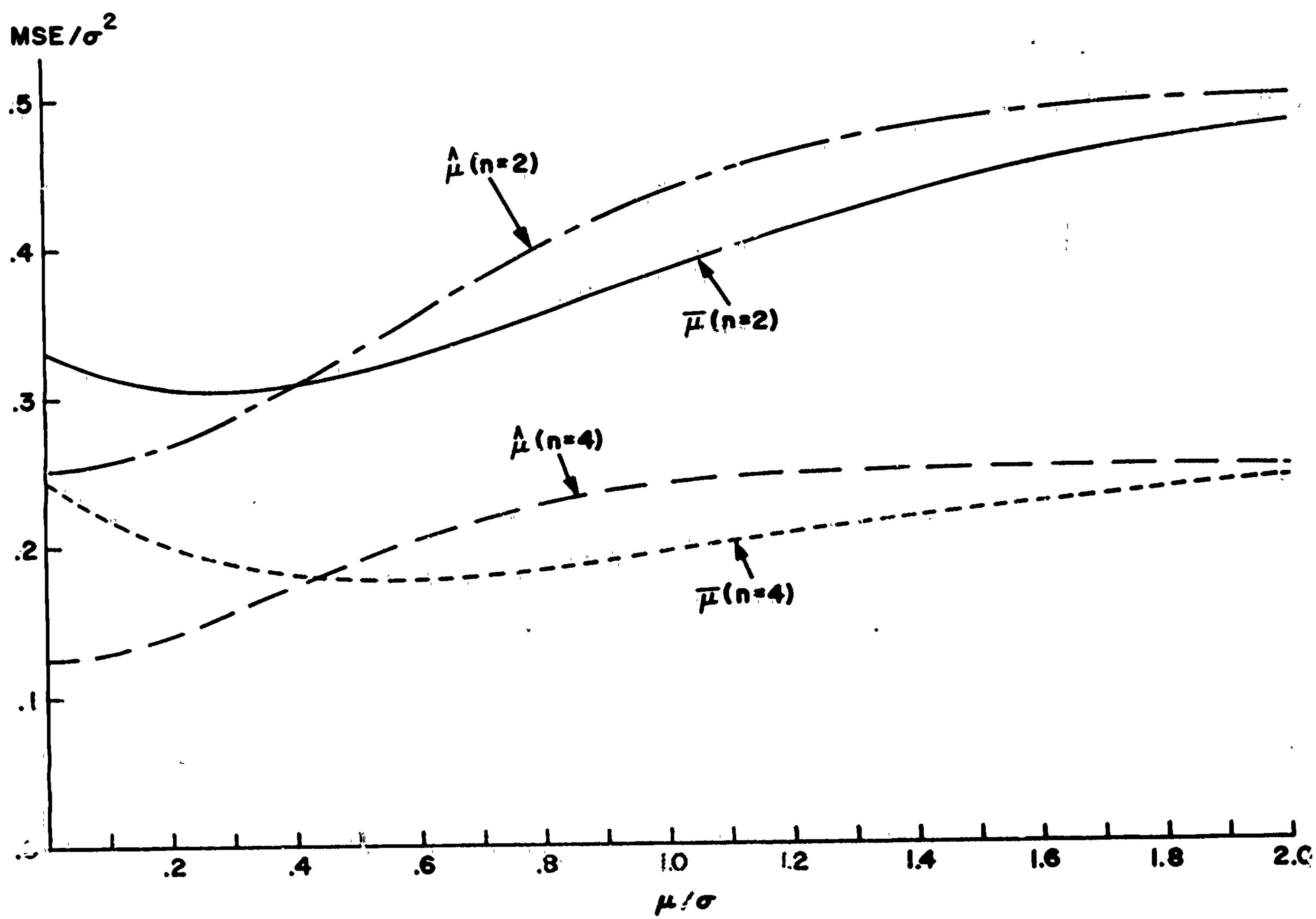
3. COMPARISON OF NORMAL CASE WITH CHI-SQUARE DIFFERENCE ($\sigma^2 = 1$)

n	m	$f_1 = f_2 = 2$		$f_1 = f_2 = 20$		Normal	
		$\hat{\mu}$	$\bar{\mu}$	$\hat{\mu}$	$\bar{\mu}$	$\hat{\mu}$	$\bar{\mu}$
3	0	.167	.250	.167	.270	.167	.273
	0.4	.248	.253	.221	.227	.219	.223
	0.8	.329	.315	.294	.249	.288	.240
	1.2	.333[a]	.333[a]	.327	.288	.322	.275
	1.6	.333[a]	.333[a]	.333	.318	.332	.303
	2.0	.333[a]	.333[a]	.333	.330	.333	.320
6	0	.083	.188	.083	.213	.083	.216
	0.4	.136	.141	.127	.139	.125	.138
	0.8	.166	.158	.161	.131	.159	.127
	1.2	.167[a]	.167[a]	.166	.146	.166	.139
	1.6	.167[a]	.167[a]	.167	.159	.167	.152
	2.0	.167[a]	.167[a]	.167	.165	.167	.160

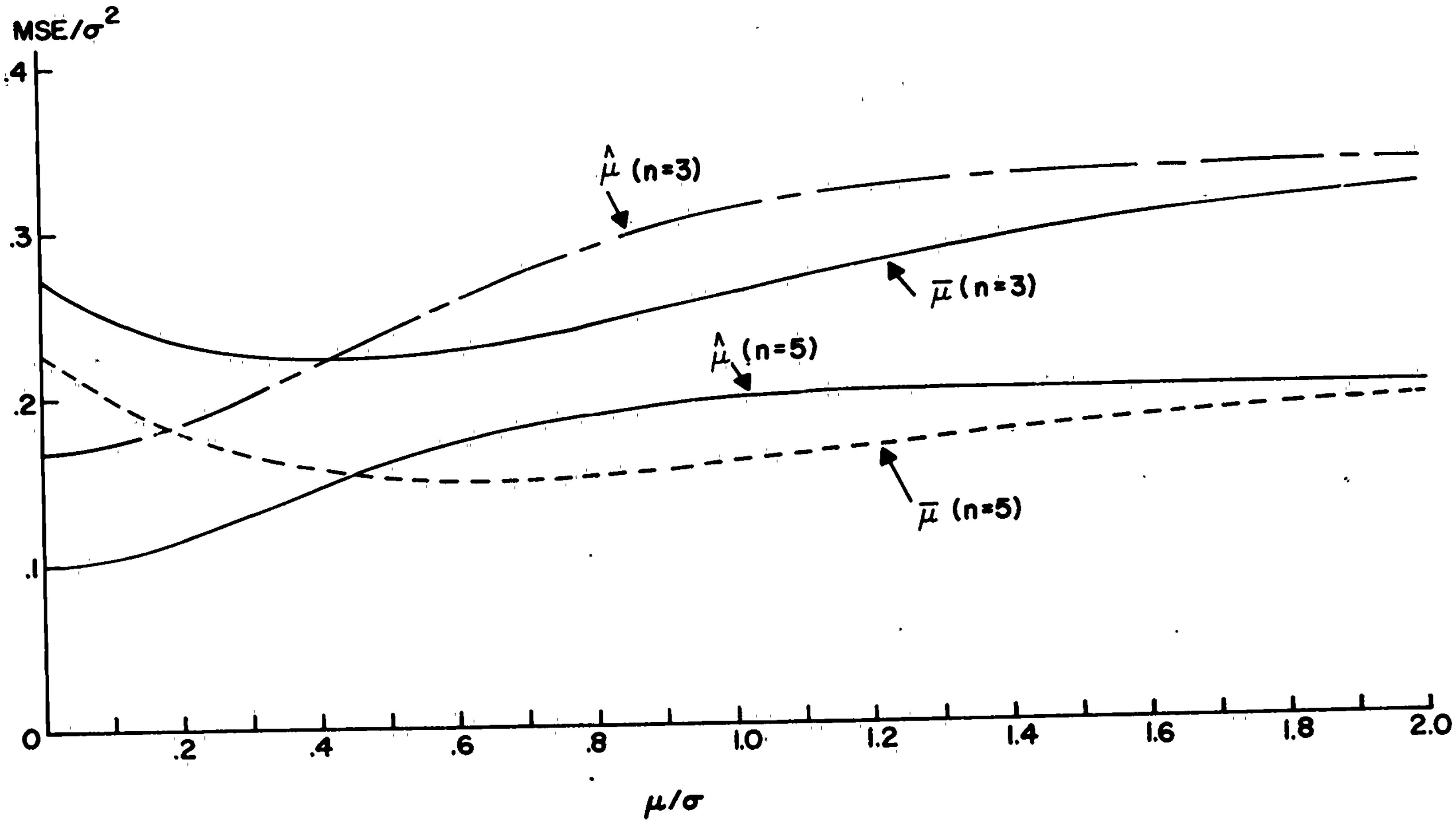
[a]Values have been extrapolated mathematically; $m > 1$ cannot exist when $f_1 = f_2 = 2$.

4. VALUES OF MSE/σ^2 FOR $\hat{\mu}$, $\bar{\mu}$, $\bar{\mu}_1$, $\bar{\mu}_2$ AND $\tilde{\mu}$ WHEN $n = 3$ AND $f_1 = f_2 = 2$

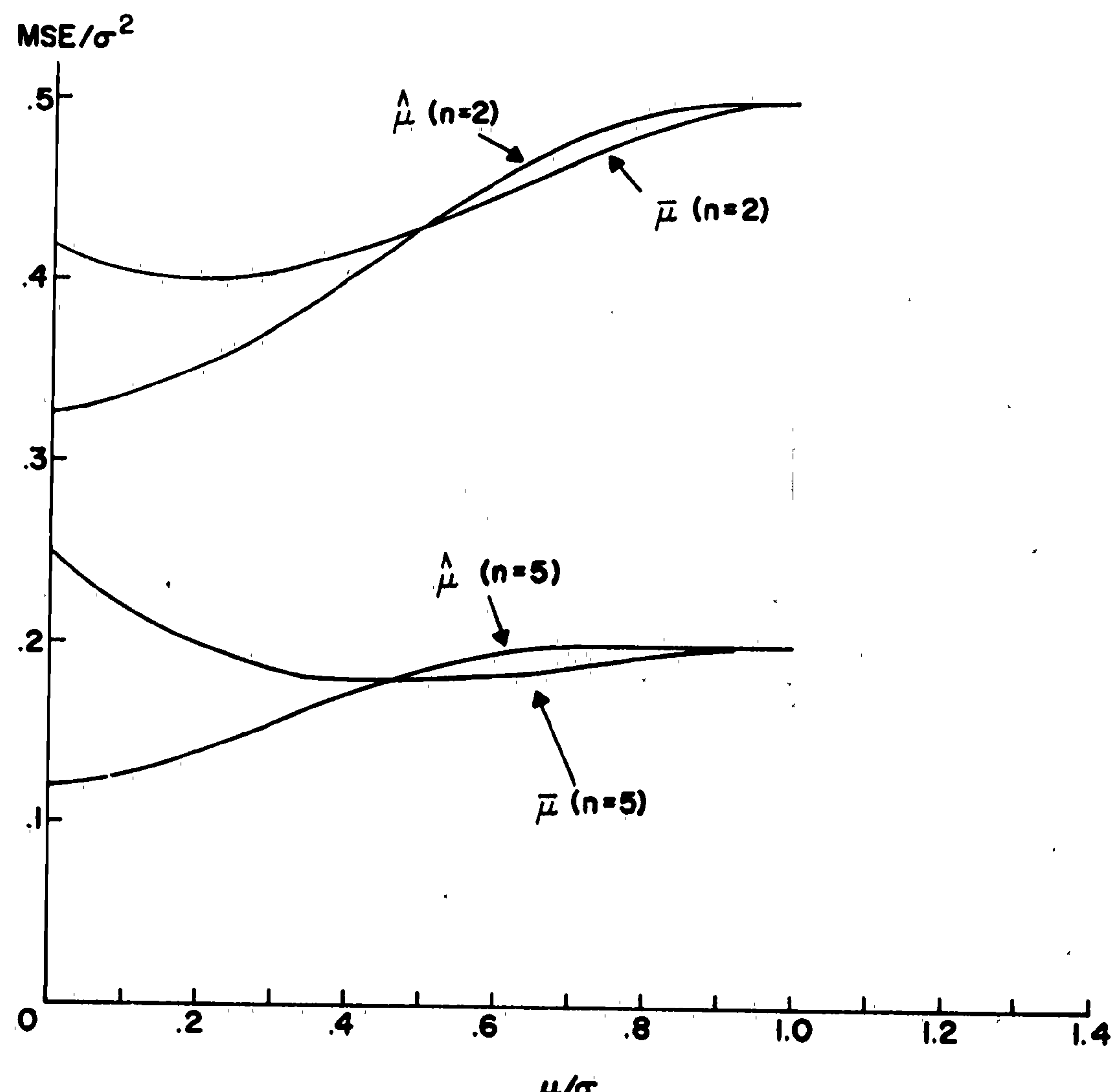
		MSE/σ^2				
$r = \theta_2/\theta_1$	m	$\hat{\mu}$	$\bar{\mu}$	$\bar{\mu}_1$	$\bar{\mu}_2$	$\tilde{\mu}$
0	1.000	.333	.333	.333	.333	.389
0.2	0.784	.327	.313	.319	.321	.361
0.4	0.557	.289	.275	.288	.276	.290
0.5	0.447	.261	.259	.270	.250	.253
0.6	0.343	.234	.248	.253	.228	.220
0.7	0.246	.209	.242	.239	.212	.194
0.8	0.156	.189	.240	.228	.201	.175
1.0	0	.167	.250	.219	.198	.160



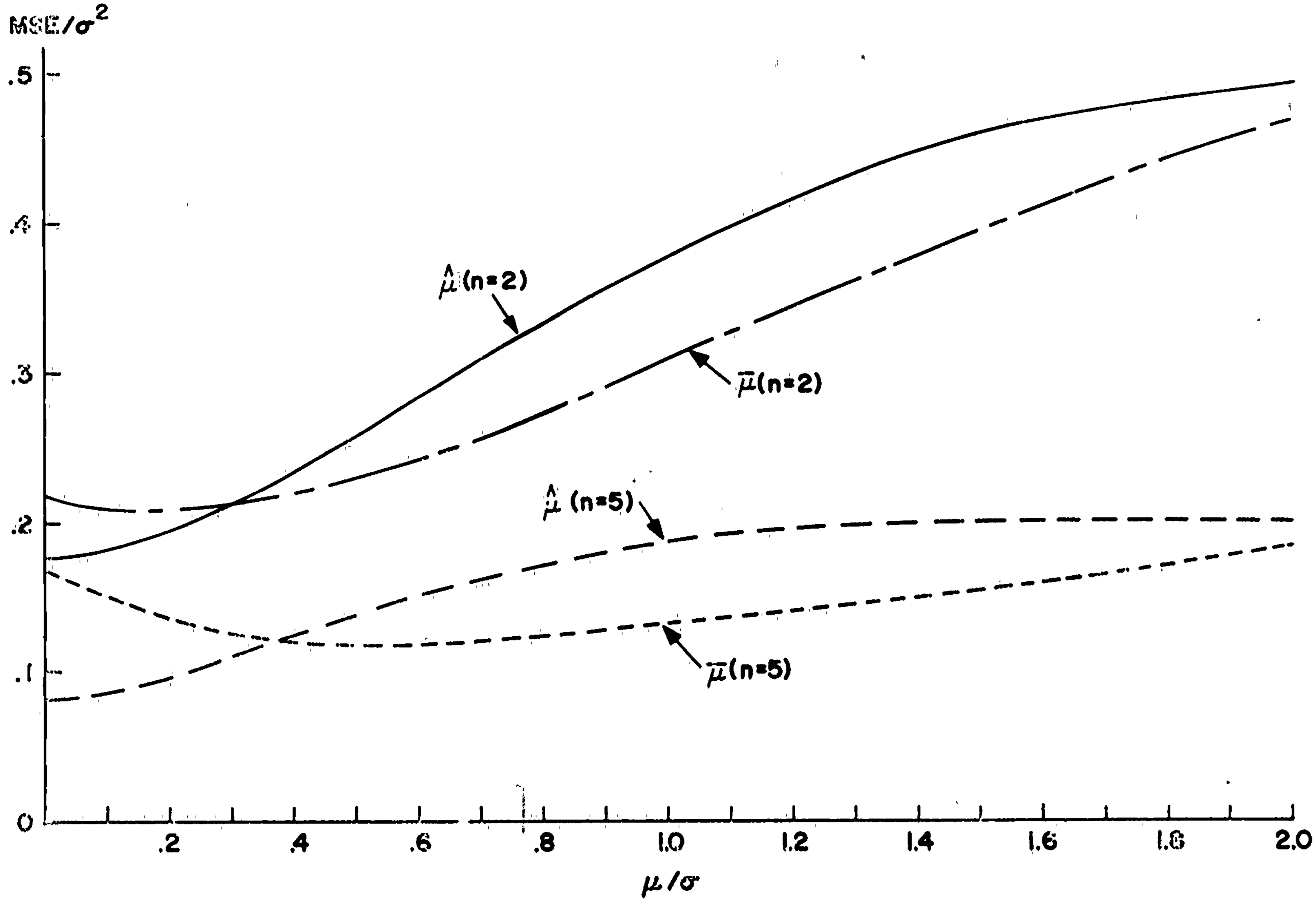
A. MSE comparison for $n = 2$ and 4 (normal case).



B. MSE comparison for $n = 3$ and 5 (normal case).



C. MSE comparison for chi-square difference ($f_1 = 2$, $f_2 = 20$).



D. MSE comparison for chi-square difference ($f_1 = 20$, $f_2 = 2$).