Department of Statistics

University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT No. 175

October 1968

Combinations of Unbiased Estimators of the Mean which Consider Inequality of Unknown Variances

by

J. S. Mehta and John Gurland

University of Wisconsin

This research was supported by the Wisconsin Alumni Research Foundation and the National Science Foundation.

$$\hat{\mu}(F) = (s_2^2 \bar{x} / s_1^2 + \bar{y}) / (s_2^2 / s_1^2 + 1) = \frac{(F\bar{x}+\bar{y})}{(F+1)} \tag{3}$$

where

$$(N-1)\, s_1^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2;\ (N-1)\, s_2^2 = \sum_{i=1}^{N} (y_i - \bar{y})^2$$
$$F = s_2^2 / s_1^2. \tag{4}$$

They prove a theorem according to which $\hat{\mu}(F)$ has smaller variance than that of $\bar{x}$ and $\bar{y}$ for all values of k if and only if $N > 10$. The present paper constitutes an attempt to improve upon the estimator $\hat{\mu}(F)$ when $N \leq 10$ and when it is known that $k \geq 1$. This corresponds to situations where it is known that the variance in one specific sample exceeds that in the other. For this case we present estimators for sample sizes 3, 5, 7, 9 and 11 which for all values of $k \geq 1$ are more efficient than the corresponding estimator $\hat{\mu}(F)$. We prove that, in fact, for all sample sizes there exist estimators belonging to the class defined below which have efficiency greater than that of $\hat{\mu}(F)$ for all $k \geq 1$. This should not at all appear surprising, however, because $\hat{\mu}(F)$ was presented as an estimator of μ for all values of $k > 0$, whereas the estimators presented here have been formulated to behave well only in the restricted range $k \geq 1$. The purpose of comparing the behavior of our estimators with that of $\hat{\mu}(F)$ over the range $k \geq 1$ is merely to indicate what advantage can be gained by utilizing such a priori knowledge about k.

As far as the two sided case of k is concerned, i.e. $k > 0$, Zacks [2] has developed a class of estimators for sample size $N = 3$ which for a certain finite interval of k values, symmetric around $k = 1$,

iii) The class $T_3 = \hat{\mu}(\phi)$ where $\phi(F) = \frac{(C+F)^{\frac{1}{2}}}{(C+F)^{\frac{1}{2}}+A}$ (8)

It should be noted here that the constants A,C and k^o which appear in the classes of estimators defined above are to be interpreted as general constants, that is to say not necessarily the same values of these constants would be employed in different classes. For example the constants A and C which appear in T_1 need not have the same values as A and C in T_3.

2.2 Rationale for the choice of the above estimators.

The class T_1 includes $\hat{\mu}(F)$ as a special case when $C = 0$ and $A = 1$. The constant C in this class has the effect of changing the origin of the F distribution and the constant A simply changes the weight given to $\bar{y}$. On intuitive grounds T_1 has the possibility of surpassing $\hat{\mu}(F)$ in efficiency by a prudent choice of constants C and A which would increase the weight given to $\bar{x}$ and decrease that given to $\bar{y}$ for $k > 1$. We shall prove in section 2.4 below the existence of a set of constants for which efficiency of T_1 is greater than that of $\hat{\mu}(F)$ for all values of $k \geq 1$ and for all sample sizes.

In the class of estimators T_2 a preliminary test is employed on the basis of an F-statistic in order to decide whether to use the average of $\bar{x}$ and $\bar{y}$ or to use a statistic of the form T_1 with $C=0$. There is no loss of generality in taking $C = 0$ here for otherwise the constant k^o in the preliminary test would merely be adjusted accordingly. This preliminary test is of the hypothesis H_o: $k=1$; consequently if it is accepted, it is reasonable to use the statistic $\frac{\bar{x}+\bar{y}}{2}$ and when it is rejected, to use a statistic of the form T_1.

It may be pointed out that other functions ϕ of F could also be employed to generate further classes of estimators. In particular, we explored the following two functions

$$\phi(F) = \frac{C + F^{\frac{1}{2}}}{C + F^{\frac{1}{2}} + A} \tag{17}$$

$$\phi(F) = \begin{cases} \dfrac{F^{\frac{1}{2}}}{F^{\frac{1}{2}} + 1} & \text{for } F < k^{o} \\ \dfrac{F^{\frac{1}{2}}}{F^{\frac{1}{2}} + A} & \text{for } F \geq k^{o} \end{cases} \tag{18}$$

The estimators of μ obtained thus do not posses any interesting properties which T_1, T_2 and T_3 do not already posses, and consequently will be left out from further consideration here.

2.3 Efficiency of the three classes of estimators for sample size N= 3

A natural way of comparing the behaviour of the estimators presented here is to examine their variances. Furthermore all of these should be compared with the best linear unbiased estimate available when the value of k is known, viz, $\hat{\mu}_o$. For this purpose we define the efficiency of an estimator in the usual manner, viz, the ratio of variance of $\hat{\mu}_o$ to the variance of the estimator in question.

Although it is possible to obtain expressions for the efficiencies of the estimators T_1, T_2 and T_3 for any specific value of N, these are generally very cumbersome and really of little practical value for sample sizes N>3 . In the present section we derive the expressions for efficiency for the case N = 3. Further consideration of the efficiency for sample sizes N > 3 will be given in section 3 . Now we may write

$$\text{eff } T_2 = (k+1)^{-1} \left[\frac{k+1}{4} \left\{ \frac{1}{k} - \frac{1}{k+k^o} \right\} + \frac{\gamma_2}{A+k^o} + \frac{\gamma_3}{k+k^o} - \gamma_1 \log\left\{(A+k^o)/(k+k^o)\right\} \right]^{-1} \quad (26)$$

where

$$\gamma_2 = \frac{(k+1)A^2}{(k-A)^2} \; ; \; \gamma_3 = \frac{k(k+A^2)}{(k-A)^2} \; ; \; \gamma_1 = \frac{\gamma_2 k^2 + \gamma_3 A^2 - kA^2}{k\,A(A-k)} \quad (27)$$

also if we let k tend to infinity in (26) we obtain

$$\lim_{k \to \infty} \text{eff } T_2 = \left[\frac{k^o}{4} + \frac{A^2}{k^o + A} + 1 \right]^{-1} \quad (28)$$

The variance of the class of estimators T_3 is given by

$$V(T_3) = \frac{\sigma_1^2}{3} k \int_0^{\infty} \frac{(C+F) + kA^2}{[(C+F)^{\frac{1}{2}} + A]^2} \frac{dF}{(k+F)^2} \quad (29)$$

Application of the transformation $F = t^2 - C$ so that $dF = 2tdt$ immediately gives

$$V(T_3) = \frac{2\sigma_1^2}{3} k \left[Y(3) + k A^2 Y(1) \right] \quad (30)$$

where

$$Y(i) = \int_{\sqrt{C}}^{\infty} \frac{t^i dt}{(A+t)^2 (k-C+t^2)^2} \quad \text{for } i = 1, 3. \quad (31)$$

The integrals Y(i) can be evaluated by employing the standard reduction formulae. Consequently we obtain

$$\text{eff } T_3 = 0.5\,(k+1)^{-1} \left[Y(3) + kA^2\, Y(1) \right]^{-1} \quad (32)$$

2.4 Some properties of the estimators:

In this section we prove certain results which depict some important properties of the estimators T_1, T_2 and T_3.

Since

$$\frac{1}{(k\,F+1)^3} > \frac{1}{(F+k)^3} > \frac{F^2}{(F+k)^3}$$

in the region $0 < F < 1$ and $k > 1$, the integrand in the above integral is negative, and hence

$$\left[\frac{d}{dC} \quad V(T_1) \right]_{C=0} < 0 \tag{38}$$

consequently there exists a value of $C > 0$ such that $V(T_1) < V[\hat{\mu}(F)]$.

(2). An optimal choice of constants A and k^o in T_2

Let us consider the estimator T_2 for which the variance expression for any N is given by

$$V(T_2) = \frac{\sigma_1^2}{N} \frac{k^{\frac{N-1}{2}}}{B(\frac{N-1}{2}, \frac{N-1}{2})} \left[\frac{k+1}{4} \int_0^{k^o} \frac{F^{\frac{N-3}{2}} dF}{(k+F)^{N-1}} + \int_{k^o}^{\infty} \frac{F^2 + k A^2}{(F+A)^2} \frac{F^{\frac{N-3}{2}}}{(k+F)^{N-1}} \right] \tag{39}$$

On taking the first derivative of $V(T_2)$ with respect to k^o we obtain

$$\frac{d}{dk^o} V(T_2) = \frac{\sigma_1^2}{N} \frac{k^{\frac{N-1}{2}}}{B(\frac{N-1}{2}, \frac{N-1}{2})} \left[\frac{k+1}{4} - \frac{k^{o^2} + kA^2}{(k^o + A)^2} \right] \frac{k^{o^{\frac{N-3}{2}}}}{(k+k^o)^{N-1}} \tag{40}$$

It is easy to see that the above expression vanishes for all k if we set $k^o = A$. Further, it can be shown that the second derivative with respect to k^o is positive at $k^o = A$ whenever $k > 1$. Consequently for value of A the estimator T_2 has minimum variances when $k^o = A$. This simplifies the choice of constants for seeking estimators in the class T_2.

In Table 3 we have depicted the behavior of the efficiency of T_2 for N = 3 at k = 1 and k = ∞ for values of $k^o = A$ lying in the range $0 < k^o = A < \infty$. At k = 1 the efficiency of T_2 increases from 0.5 at $k^o = 0$ to 1.0 at $k^o = \infty$. On the other hand at $k = \infty$ the efficiency of T_2 decreases from 1.0 at $k^o = 0$ to 0 at $k^o = \infty$. The behavior of T_2 for intermediate values of k^o has been presented in Figure 3. Let R_1 be the value of $\log_e k^o$ for which the efficiency of T_2 corresponding to k = 1 assumes the value of 0.75. Let R_2 be the value of $\log_e k^o$ for which the efficiency of T_2 corresponding to $k = \infty$ assumes the value 0.5. Then for $R_1 < \log_e k^o < R_2$ i.e. $0.35 < k^o < 1.33$ all the estimators T_2 with k^o in this range and $A = k^o$ are more efficient than $\hat{\mu}(F)$ for all $k \geq 1$.

(3) A property of the estimator T_3

Let us now consider the behaviour of $V(T_3)$ as k tends to infinity. We can write down the variance of T_3 for any N as

$$V(T_3) = \frac{\sigma_1^2}{N} \frac{k^{\frac{N-1}{2}}}{B(\frac{N-1}{2}, \frac{N-1}{2})} \int_0^\infty \frac{C + F + k A^2}{[(C+F)^{\frac{1}{2}}+A]^2} \frac{F^{\frac{N-3}{2}}\, dF}{(k + F)^{N-1}} \tag{41}$$

Setting F = kt so that dF = kdt we get

$$V(T_3) = \frac{\sigma_1^2}{N} \frac{1}{B(\frac{N-1}{2}, \frac{N-1}{2})} \left[\int_0^\infty \frac{t^{\frac{N-3}{2}}\, dt}{[1+ \frac{A}{(C+kt)^{\frac{1}{2}}}]^2 (1+t)^{N-1}} + \int_0^\infty \frac{A^2 t^{\frac{N-3}{2}}\, dt}{[(C/k+t)^{\frac{1}{2}} + A/\sqrt{k}\,]^2 (1+t)^{N-1}} \right] \tag{42}$$

Taking the limit as $k \to \infty$ and applying the Lebesque dominated convergence theorem, we note that

$$\lim_{k\to\infty} V(T_3) = \frac{\sigma_1^2}{N} \frac{1}{B(\frac{N-1}{2}, \frac{N-1}{2})} \left[\int_0^\infty \frac{t^{\frac{N-3}{2}} dt}{(1+t)^{N-1}} + A^2 \int_0^\infty \frac{t^{\frac{N-3}{2}-1} dt}{(1+t)^{N-1}} \right] = \frac{\sigma_1^2}{N} \left[1+\frac{(N-1)}{(N-3)} A^2 \right] \quad (43)$$

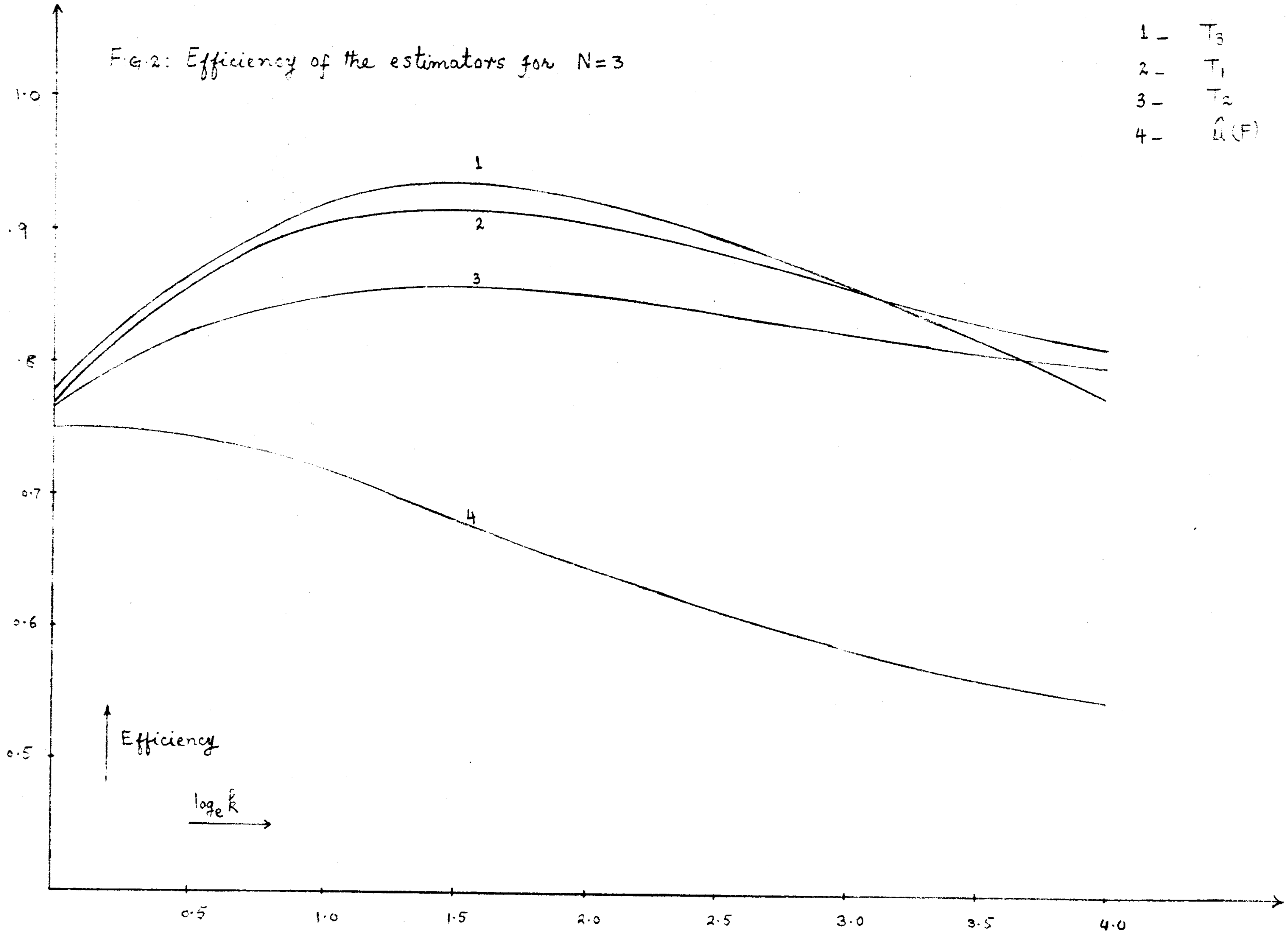
We note that for N = 3 the limiting variance of T_3 at $k = \infty$ becomes infinite, that is to say the efficiency of T_3 as defined earlier becomes zero. Nevertheless as will be pointed out in section 2.5 these estimators can be extremely useful for a wide range of k. In fact in Table 2 which considers the range $1 \leq k \leq 8103$ the efficiency of these estimators surpasses that of $\hat{\mu}(F)$ and furthermore surpasses that of the other estimators over a certain interval of k values.

2.5 Comparison of efficiencies of T_1, T_2, T_3 for the case N = 3

In this section we compare the efficiencies of the three classes of estimators when N=3. Two particular sets of constants A, C, k^o have been selected to illustrate that these estimators posses certain desirable properties. The efficiencies of these two sets of estimators are presented in Tables 1 and 2 respectively while their corresponding graphs are given in Figures 1 and 2 respectively. All the computations were carried out on a CDC 1604 computer.

The estimators in Table 1 are such that all of them have a high efficiency around k = 1 and at the same time they are more efficient than $\hat{\mu}(F)$ for the range of k values considered here, namely, $1 \leq k \leq 54.6$. In fact the estimators T_1 and T_2 have a higher efficiency than $\hat{\mu}(F)$ for $1 \leq k < \infty$. At $\log_e k = 0$ (i.e. k=1) the efficiencies are in the neighborhood of 0.86 but their behaviour beyond this neighborhood differs. The efficiency of T_3 rises to a peak of about 0.94 around $\log_e k = 1.0$ (i.e. k=2.72)

Fig. 2: Efficiency of the estimators for N=3



but then drops rapidly. On the other hand the efficiencies of T_1 and T_2 rise gently to a peak of about 0.90 at $\log_e k$ between 0.25 and 0.50 (i.e. at k between 1.28 and 1.65) but do not drop nearly as rapidly as that of T_3. In fact T_1, for $\log_e k > 3.50$ (i.e. $k > 33.12$) and T_2, for $\log_e k > 3.75$ (i.e. $k > 42.52$) have efficiencies which are higher than that of T_3.

In Table 2 we present constants A, C and k^o for which the estimators have higher efficiencies than $\hat{\mu}(F)$ for a wide range of k values, namely, $1 \leq k \leq 8103$. In fact the minimum efficiency of T_1 and T_2 for the range $1 \leq k < \infty$ is greater than the maximum efficiency of $\hat{\mu}(F)$ which is 0.75 at $k = 1$. Estimator T_3 has efficiency which exceeds 0.85 for the range $0.5 \leq \log_e k \leq 3.0$ (i.e. $1.65 \leq k \leq 20.09$) and is in fact higher than that of T_1 as well. But for values of $k > 8103$ the efficiency of T_3 is below 0.5.

It may be noted that the estimator T_1 which employs a continuous weight function $\phi(F)$ performs better than T_2 which employes a discontinuous weight function. We can therefore conclude that although all T_1, T_2, and T_3 perform better than $\hat{\mu}(F)$ yet T_1 is the best of them all in view of the fact that one can select constants A and C as in Table 1 or Table 2 such that it has higher efficiency than that of $\hat{\mu}(F)$ for <u>all</u> $1 \leq k < \infty$. If we are interested in a particular interval of k values, however, for example $1 < k < 20$ then as is evident from Tables 1 and 2 and Figures 1,2 the estimator T_3 is the best.

An explicit expression for the efficiency of T_1 for any (odd) sample size can be given as follows

$$\text{eff } T_1(N) = \frac{B(\frac{N-1}{2}, \frac{N-1}{2})}{k^{\frac{N-3}{2}}(k+1) \; [A_1 \log \frac{k}{A+C} + \frac{A_2}{A+C} + \frac{B_2}{k} + \frac{B_3}{2k^2} + \ldots + \frac{B_{N-1}}{(N-2)k^{N-2}}]} \tag{44}$$

where if we write $\psi_1(F)$ and $\psi_2(F)$ as follows

$$\psi_1(F) = \frac{F^{\frac{N-3}{2}} [(C+F)^2 + k A^2]}{(k+F)^{N-1}} \tag{45}$$

$$\psi_2(F) = \frac{F^{\frac{N-3}{2}} [(C+F)^2 + k A^2]}{(C+A+F)^2} \tag{46}$$

then

$$\left. \begin{aligned} A_2 &= [\psi_1(F)]_{F = -(C+A)} \\ A_1 &= [\frac{d}{dF} \psi_1(F)]_{F = -(C+A)} \end{aligned} \right] \tag{47}$$

and

$$\left. \begin{aligned} B_{N-1} &= [\psi_2(F)]_{F = -k} \\ B_{N-1-i} &= [\frac{1}{(i-1)!} \frac{d^i}{dF^i} \psi_2(F)]_{F = -k} \quad \text{for } i = 1,2,\ldots,N-2 \end{aligned} \right] \tag{48}$$

For even sample size the expression for efficiency involves an infinite series; consequently we confine our attention to odd sample size. Now we shall prove that the limiting efficiency of T_1 as k tends to infinity is unity and this is true for all values of $A > 0$ and $C \geq 0$ provided $N \geq 5$. Since

$$\lim_{k \to \infty} V(\hat{\mu}_o) = \frac{\sigma_1^2}{N}$$

it suffices to prove that

$$\lim_{k \to \infty} V(T_1) = \sigma_1^2/N \qquad \text{for all } N \geq 5.$$

Now we have

$$V(T_1) = \frac{\sigma_1^2}{N} \; E\left[\frac{(C+F)^2 + kA^2}{(C+F+A)^2} \right]$$

$$= \frac{\sigma_1^2}{N} \; \frac{1}{B(\frac{N-1}{2}, \frac{N-1}{2})} \int_o^\infty \frac{(C+x)^2 + k A^2}{k^{\frac{N-1}{2}} (C+x+A)^2} \; \frac{x^{\frac{N-3}{2}} dx}{(1+x/k)^{N-1}} \tag{49}$$

Application of the transformation x = ky immediately gives

$$V(T_1) = \frac{\sigma_1^2}{NB(\frac{N-1}{2}, \frac{N-1}{2})} \int_o^\infty \frac{(C+ky)^2 + kA^2}{(C+ky+A)^2} \; \frac{y^{\frac{N-3}{2}} dy}{(1+y)^{N-1}} \tag{50}$$

Hence

$$\lim_{k \to \infty} V(T_1) = \frac{\sigma_1^2}{N} \left[\lim_{k \to \infty} \int_o^\infty \frac{(C+ky)^2}{(C+ky+A)^2} \; dH_{N-1,N-1}(y) + A^2 \lim_{k \to \infty} \frac{1}{k} \int_o^\infty \frac{dH_{N-1,N-1}(y)}{[y+\frac{C+A}{k}]^2} \right] \tag{51}$$

where $H_{N-1,N-1}(y)$ is the cumulative distribution of Snedecor's F with N-1, N-1 degrees of freedom. Now by the Lebesgue dominated convergence theorem

$$\lim_{k \to \infty} \int_o^\infty \frac{(C+ky)^2}{(C+ky+A)^2} \; dH_{N-1,N-1}(y) = \int_o^\infty dH_{N-1,N-1}(y) = 1 \text{ for all } N \geq 1 \tag{52}$$

and

$$\lim_{k \to \infty} \int_o^\infty \frac{dH_{N-1,N-1}(y)}{[y+\frac{C+A}{k}]^2} = \int_o^\infty \frac{dH_{N-1,N-1}(y)}{y^2} < \infty \text{ for all } N \geq 5 \tag{53}$$

Hence

References

[1] F.A. Graybill and R.B. Deal. "Combining unbiased estimators" Biometrics 15, (1959), 543-50.

[2] S. Zacks. "Unbiased estimation of the common mean of two normal distributions based on small samples of equal size" J. of Amer. Statist. Assocn. 61 (1966) 467-76.