TWO-STAGE EXPERIMENTS FOR ESTIMATING A COMMON MEAN[1]

By Donald Richter

University of North Carolina[2]

Summary. Let π_1, π_2 be two normal populations with common mean μ and variances σ_1^2, σ_2^2, where the parameter values are unknown. Suppose that it is desired to estimate μ, and that the experimental procedure is to take m observations from each population, compute variance estimates, and then take $n - 2m$ observations from that population with the smaller observed variance, where n has been fixed beforehand. Let $R_n(\theta, m) = V_0^{-1}E(\hat{\mu} - \mu)^2$ be the risk of the estimator $\hat{\mu}$, where $V_0 = n^{-1} \min(\sigma_1^2, \sigma_2^2)$ and where $\theta = \sigma_2^2/\sigma_1^2$. For a class of "best" estimators, it is shown in this paper that $\sup_\theta R_n(\theta, m) \to 1$ as $n \to \infty$ if and only if $m/n \to 0$ and $m \to \infty$ as $n \to \infty$; that $\min_m \sup_\theta R_n(\theta, m) \sim 1 + Cn^{-\frac{1}{2}}$ as $n \to \infty$; and that the minimax sample size is $m \sim Cn^{\frac{1}{2}}$ as $n \to \infty$.

1. Introduction. This investigation treats the problem of estimating the common mean μ of two populations using a fixed number n of observations. If the population variances were known, the most efficient procedure would be to take all n observations from that population with the smaller variance. When prior information about the variances is lacking or is too vague to be quantified, it is natural to consider the procedure which consists of taking a preliminary sample of size m from each population, computing estimates of the variances, and then taking the remaining $n - 2m$ observations from that population with the apparently smaller variance. Since, if m is chosen too large or too small, the advantage of the two-stage sampling scheme over the procedure of simply taking $n/2$ observations from each population will be lost, the problem arises of determining for some good estimator an optimum choice of m as a function of n, not dependent on the unknown variances.

As an example, we may suppose that we have available two devices for measuring a physical constant, that each measurement is expensive or time consuming so that their total number is limited, and that we wish to estimate the constant as accurately as possible.

For related work on two-stage experiments with a fixed total sample size, reference should be made to Ghurye and Robbins [2], where it is shown that the ratio of the variance of a certain two-stage estimator for the difference of two means to the minimum variance tends to unity as the sample size increases, and

Received August 3, 1959; revised July 25, 1960.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Present address: University of Minnesota, Minneapolis.

to Putter [3], where an analogous result, among others, is obtained for a double sampling rule for estimating the mean of a stratified population. In neither paper is any indication given as to how the first-stage sample size might be chosen as a function of the total sample size. For an introduction to the problems of sequential experimentation, see Robbins [4].

In Section 2 the problem will be formulated explicitly, and a suitable risk function will be defined. In the remaining Sections, where the populations are assumed normal, necessary and sufficient conditions for the risk to converge to unity will be obtained, and the asymptotic minimax value of m will be derived; these results will be seen to be, in a certain sense, estimator-free.

2. Formulation of the problem. In this section, the problem will be formulated in an explicit and convenient manner.

Let $X_1, X_2, X_3, \cdots$ and $Y_1, Y_2, Y_3, \cdots$ be mutually independent random variables with common mean μ and Var $X_i = \sigma_1^2$ and Var $Y_i = \sigma_2^2$; write $\theta = \sigma_2^2/\sigma_1^2$. Let

$$R = \frac{S_1^2}{S_2^2} = \frac{\sum_1^m X_i^2 - \left(\sum_1^m X_i\right)^2 / m}{\sum_1^m Y_i^2 - \left(\sum_1^m Y_1\right)^2 / m},$$

so that $1/R$ is the usual estimator of θ based on $2m$ observations. Then our procedure is first to observe $X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_m$, and secondly to observe $X_{m+1}, \cdots, X_{n-m}$ if $R < 1$, $Y_{m+1}, \cdots, Y_{n-m}$ otherwise.

Writing $\bar{X}_{N_1} = \sum_1^{N_1} X_i/N_1$ and $\bar{Y}_{N_2} = \sum_1^{N_2} Y_i/N_2$, we will consider estimators $\hat{\mu}$ of μ which can be written in the form

$$\hat{\mu} = A\bar{X}_{N_1} + B\bar{Y}_{N_2}, \tag{1}$$

where N_1, N_2, A, B are random variables such that $N_1 = n - m$ if $R < 1$, $N_1 = m$ if $R \geqq 1$, $N_1 + N_2 = n$, $0 \leqq A \leqq 1$, $0 \leqq B \leqq 1$, and $A + B = 1$ with probability one; and, in addition, where A and B are such that

$$(1') \quad E_H\bar{X}_k = E\bar{X}_k, \quad E_H\bar{Y}_l = E\bar{Y}_l, \quad E_H\bar{X}_k\bar{Y}_l = E_H\bar{X}_k \cdot E_H\bar{Y}_l,$$
$$E_H\bar{X}_k^2 = E\bar{X}_k^2, \quad E_H\bar{Y}_l^2 = E\bar{Y}_l^2 \quad \text{for all } k, l,$$

where $E_H(\cdot) = E(\cdot \mid H)$ and $H = (A, B, N_1, N_2)$. If the X_i and the Y_i are assumed to be normally distributed, then, recalling that the sample mean and variance are independent for normal populations, Assumption (1′) may be replaced by the assumption that A and B are functions of the sample variances only. Restriction to estimators of form (1) seems reasonable since, if observations are available on $X_1, \cdots, X_{n_1}, Y_1, \cdots, Y_{n_2}$, the variables are normal, and θ is known, $a\bar{X}_{n_1} + b\bar{Y}_{n_2}$ is the uniformly minimum variance unbiased estimator of μ, where $a = n_1\theta/(n_1\theta + n_2)$, $a + b = 1$.

Next, let $V_0 = (1/n) \min (\sigma_1^2, \sigma_2^2)$, which is the variance of the standard estimator of μ for the case when sgn $(\sigma_2^2 - \sigma_1^2)$ is known beforehand, and define

$R_n(\theta, m) = V_0^{-1}E(\hat{\mu} - \mu)^2$ to be the risk function associated with the estimator $\hat{\mu}$. That $R_n(\theta, m) = V_0^{-1}$ Var $\hat{\mu}$ follows from the first part of the following theorem.

THEOREM 1: *For any estimator of form* (1),

(a) $E\hat{\mu} = \mu$,

(b) $R_n(\theta, m) = n \max(1, 1/\theta)E\{A^2/N_1 + \theta B^2/N_2\}$,

(c) $R_n(\theta, m) \geqq n \max(1, \theta)E\{1/(N_2 + N_1\theta)\} \geqq 1$.

PROOF. Since $E_H\hat{\mu} = AE_H\bar{X}_{N_1} + BE_H\bar{Y}_{N_2} = \mu$, $E\hat{\mu} = \mu$. Next, $E_H(\hat{\mu} - \mu)^2 = E_H(A\bar{X}_{N_1} + B\bar{Y}_{N_2} - \mu)^2 = A^2\sigma_1^2/N_1 + B^2\sigma_2^2/N_2$ which proves (b) since $R_n(\theta, m) = V_0^{-1}E\{E_H(\hat{\mu} - \mu)^2\}$. Finally, $A^2/N_1 + \theta B^2/N_2$ has a unique minimum with respect to $A = 1 - B$ at $A = N_1\theta/(N_2 + N_1\theta)$ so that $E\{A^2/N_1 + \theta B^2/N_2\} \geqq \theta E\{1/(N_2 + N_1\theta)\}$ which proves the left-hand inequality of (c); since $N_2 + N_1\theta \leqq n \max(1, \theta)$, the proof is complete.

It will be instructive at this point to examine the risk function for the usual one-stage experiment for estimating μ, which would be to observe $n/2$ of the X_i and $n/2$ of the Y_i. If we confine attention to estimators μ' such that $E\mu' = \mu$ and assume the variables normally distributed, then Var $\mu' \geqq 2\sigma_2^2/n(1 + \theta)$, since $(\theta\bar{X} + \bar{Y})/(\theta + 1)$ is the minimum variance unbiased estimator with variance $2\sigma_2^2/n(1 + \theta)$ when the variances are known. Then, $R(\theta) = (\text{Var } \mu')/V_0 \geqq 2 \max(\theta, 1)/(1 + \theta)$ and $2 \max(\theta, 1)/(1 + \theta) \geqq 1$ with equality if and only if $\theta = 1$. Hence for each fixed $\theta \neq 1$, the risk function is bounded away from unity independent of the sample size. One would hope that the risk for the two-stage scheme would prove to be smaller, and we shall see–for large samples at least–that this is in fact the case if m is suitably chosen.

Returning to the two-stage experiment, it is clear that, once an estimator is specified, the only variable left at the statistician's disposal is the quantity m. Then, given an estimator of form (1), we may say that any real-valued function $m(n)$ which satisfies $4 \leqq 2\, m(n) < n$ for all $n \geqq 5$ is a solution to the problem. Can we find an optimum solution?

With respect to an estimator of form (1), we shall call $m(n)$ a uniformly consistent solution (u.c.sol.) if $\sup_\theta R_n(\theta, m(n)) \to 1$ as $n \to \infty$; where they exist, we shall restrict attention to such solutions. Further, if $\sup_\theta R_n(\theta, m) < \infty$, a solution which minimizes $\sup_\theta R_n(\theta, m)$ will be called a minimax solution (m.m.sol.). If there exists a u.c.sol., then a m.m.sol. is u.c. too. Hence, the minimax principle affords a means of choosing one solution from the class of u.c. solutions.

3. A simple estimator. In this section an asymptotic minimax solution will be derived for a particular unbiased estimator. In this and the following section, we shall assume that the X_i's and the Y_i's are normally distributed. Hence, $R\theta = S_1^2\sigma_2^2/S_2^2\sigma_1^2$ obeys the F-distribution with $m - 1, m - 1$ degrees of freedom, and we may write

$$I(\theta, m) = \Pr\{R \geqq 1\} = B\left(\frac{m-1}{2}, \frac{m-1}{2}\right)^{-1} \int_\theta^\infty x^{(m-3)/2}(1 + x)^{1-m}\, dx.$$

We now define $\hat{\mu}_1 = A_1\bar{X}_{N_1} + B_1\bar{Y}_{N_2}$, where $A_1 = 1$ or 0 according as $R < 1$ or $R \geqq 1$. This estimator has form (1) and, by Theorem 1, $\hat{\mu}_1$ is unbiased and

$$R_{1n}(\theta, m) = n \max(1, 1/\theta)\, E\{A_1^2/N_1 + B_1^2\theta/N_2\}$$
$$= n \max(1,1/\theta)\left[\frac{1 - I(\theta, m)}{n - m} + \frac{\theta I(\theta, m)}{n - m}\right]$$
$$= \max(1,1/\theta)(1 - m/n)^{-1}\,[1 + (\theta - 1)\, I(\theta, m)].$$

It is easy to show that $R_{1n}(\theta, m) = R_{1n}(1/\theta, m)$ by using the fact that $I(\theta, m) = 1 - I(1/\theta, m)$; thus, when considering $\sup_\theta R_{1n}(\theta, m)$, we may restrict ourselves to $\theta \geqq 1$. Before continuing with the study of the risk of the proposed estimator, we introduce some lemmas.

LEMMA 1: *For* $\theta \geqq 1$, $I(\theta, m) \leqq [2\theta^{\frac{1}{2}}/(1 + \theta)]^{m-1}$.

LEMMA 2: *Let* $\theta = 1 + 2\tau m^{-\frac{1}{2}}$. *For* $\tau \geqq 0$ *and* τ *bounded*, $I(\theta, m) - \Phi(-\tau) = O(m^{-\frac{1}{2}})$ *uniformly in* τ *as* $m \to \infty$.

PROOF OF THE LEMMAS: To begin with,

$$I(\theta, m) = \Pr\{R\theta \geqq \theta\} = \Pr\left\{\sum_1^{m-1} U_i/\sum_1^{m-1} V_i \geqq \theta\right\} = \Pr\{S_{m-1} \geqq 0\},$$

where $S_{m-1} = Z_1 + \cdots + Z_{m-1}$, $Z_i = U_i - \theta V_i$, and $U_1, U_2, \cdots, U_{m-1}, V_1, V_2, \cdots, V_{m-1}$ are independent random variables, each obeying the chi-square distribution with one degree of freedom. Then $Z_1, Z_2, \cdots, Z_{m-1}$ are independent and identically distributed random variables, $EZ_i = 1 - \theta \leqq 0$ since by hypothesis $\theta \geqq 1$, and $\operatorname{Var} Z_i = 2(1 + \theta^2) \geqq 4$.

The moment generating function of Z_i is

$$M(t) = Ee^{tZ_i} = [1 + 2(\theta - 1)t - 4\theta t^2]^{-\frac{1}{2}}.$$

Using the Chebyshev inequality, $I(\theta, m) = \Pr\{S_{m-1} \geqq 0\} \leqq Ee^{tS_{m-1}} = [M(t)]^{m-1}$, $t \geqq 0$. Since $M(t)$ is a minimum at $t = (\theta - 1)/4\theta$, we obtain Lemma 1.

Next, let $F_{m-1}(x)$ be the distribution function of the variable $(S_{m-1} - ES_{m-1})/(\operatorname{Var} S_{m-1})^{\frac{1}{2}}$ so that $1 - I(\theta, m) = F_{m-1}(x)$ when $x = (\theta - 1)(m - 1)^{\frac{1}{2}}[2(1 + \theta^2)]^{-\frac{1}{2}}$; if $\theta = 1 + 2\tau m^{-\frac{1}{2}}$, $\tau \geqq 0$ and $\tau = O(1)$, then $x = \tau + \epsilon(\tau, m)$ where $\epsilon(\tau, m) = O(m^{-\frac{1}{2}})$ uniformly in τ. Noting that τ bounded implies θ bounded, and that

$$|Z_i - EZ_i|^3 \leqq 8\{|Z_i|^3 + |EZ_i|^3\} \leqq 8\{(U_i + \theta V_i)^3 + (\theta - 1)^3\},$$

we observe that $E|Z_i - EZ_i|^3$ is bounded. Then the Berry-Esseen Theorem [1] states that

$$|F_{m-1}(x) - \Phi(x)| \leqq CE\{|Z_i - EZ_i|^3\}/(\operatorname{Var} Z_i)^{\frac{3}{2}}(m - 1)^{\frac{1}{2}} \tag{2}$$

for all x where C is a constant. Since $(\operatorname{Var} Z_i)^{\frac{3}{2}} \geqq 8$ and $(m - 1)^{\frac{1}{2}} > m^{\frac{1}{2}}/2$, we find from (2) that $F_{m-1}(x) - \Phi(x) = O(m^{-\frac{1}{2}})$ uniformly in τ. But $\Phi(x) - \Phi(\tau) = O(m^{-\frac{1}{2}})$ uniformly in τ since $|\Phi(x) - \Phi(\tau)| < |\epsilon(\tau, m)|$. Then $F_{m-1}(x) - \Phi(\tau) = O(m^{-\frac{1}{2}})$ uniformly in τ and, since $1 - I(\theta, m) = F_{m-1}(x)$, the proof of Lemma 2 is complete.

Now let $K(\theta, m) = (\theta - 1)I(\theta, m)$ for $\theta \geqq 1$ and take $m \geqq 4$. Then $K(\theta, m)$ is continuous in θ, $K(1, m) = 0$, and $K(\theta, m) > 0$ for $1 < \theta < \infty$. From Lemma 1, $K(\theta, m) < 4(2\theta^{\frac{1}{2}}/(1 + \theta))^{m-3}$; therefore $K(\theta, m) < 4$ and $K(\theta, m) \to 0$ as $\theta \to \infty$. Hence $K(\theta, m)$ has an absolute maximum with respect to θ, $1 \leqq \theta < \infty$. Using the integral representation of $I(\theta, m)$, straightforward differentiation yields

$$\frac{\partial^2 K(\theta, m)}{\partial \theta^2} = \frac{(m - 3)\, \theta^{(m-5)/2}(\theta^2 - 2\theta(m + 1)/(m - 3) + 1)}{2(\theta + 1)^m B((m - 1)/2, (m - 1)/2)} .$$

Hence,

$$\frac{\partial^2 K(\theta, m)}{\partial \theta^2} \begin{cases} < 0 \text{ if } 1 \leqq \theta < \theta_0 \\ = 0 \text{ if } \theta = \theta_0 \\ > 0 \text{ if } \theta > \theta_0 \end{cases}$$

where $\theta_0 = 1 + 2^{\frac{3}{2}}(m - 1)^{\frac{1}{2}}/(m - 3) + 4/(m - 3)$. It follows that $K(\theta, m)$ has exactly one maximum w.r.t. θ, and that the maximizing value of θ satisfies the inequality, $1 < \theta < \theta_0$. Next, let $\theta = 1 + 2\tau m^{-\frac{1}{2}}$, $\tau \geqq 0$, $\tau_0 = (\theta_0 - 1)m^{\frac{1}{2}}2^{-1} = O(1)$, so that $m^{\frac{1}{2}} \max_{\theta<\theta_0} K(\theta, m) = \max_{\tau<\tau_0} 2\tau I(\theta, m) = \max_{\tau<\tau_0} 2\tau\Phi(-\tau) + O(m^{-\frac{1}{2}})$ by Lemma 2. The root of the equation $\Phi(-\tau) - \tau\phi(\tau) = 0$ gives the τ-value which maximizes $\tau\Phi(-\tau)$; call this root τ' (approximately .75) and let $c = 2\tau'\Phi(-\tau')$ (approximately .6). We have proved the following result.

Lemma 3: *For $\theta \geqq 1$, $(\theta - 1)I(\theta, m)$ has a unique, absolute maximum with respect to θ at $\theta = 1 + 2\tau' m^{-\frac{1}{2}} + o(m^{-\frac{1}{2}})$, and* $\max_\theta(\theta - 1)I(\theta, m) = cm^{-\frac{1}{2}} + O(m^{-1})$, *where τ', c are positive constants defined above.*

Returning to the expression for the risk, we find that $\max_\theta R_{1n}(\theta, m) = \max_{\theta \geqq 1} R_{1n}(\theta, m) = (1 - m/n)^{-1}(1 + cm^{-\frac{1}{2}} + O(m^{-1}))$ by Lemma 3. Suppose now that $\max_\theta R_{1n}(\theta, m) \to 1$ as $n \to \infty$. Then of necessity $(1 - m/n)^{-1} \to 1$ and $cm^{-\frac{1}{2}} \to 0$ as $n \to \infty$, implying respectively that $m/n \to 0$ and $m \to \infty$ as $n \to \infty$. Since the converse is obviously true, we have proved that $m(n)$ is u.c. if and only if

$$(*) \qquad m(n)/n \to 0 \quad \text{and} \quad m(n) \to \infty \quad \text{as} \quad n \to \infty.$$

That is, (*) characterizes the class of u.c. solutions for $\hat{\mu}_1$. We can now determine the m.m. sol.

Theorem 2: *The minimax solution for $\hat{\mu}_1$ is $m(n) = (cn/2)^{\frac{2}{3}} + O(n^{\frac{1}{3}})$ and* $\min_m \max_\theta R_{1n}(\theta, m) = 1 + 3(c/2)^{\frac{2}{3}} n^{-\frac{1}{3}} + O(n^{-\frac{2}{3}})$.

Proof: For all u.c. solutions, and hence for the m.m. sol., $\max_\theta R_{1n}(\theta, m) = 1 + m/n + cm^{-\frac{1}{2}} + \epsilon$ where $m = m(n)$ and $\epsilon = cm^{\frac{1}{2}}/n + O(m^2/n^2) + O(m^{-1}) = o(m/n) + o(m^{-\frac{1}{2}})$. Both m/n and $m^{-\frac{1}{2}}$ converge to zero as $n \to \infty$, one being an increasing function of m, the other a decreasing function. Hence a necessary condition for a minimum is that m/n and $m^{-\frac{1}{2}}$ be of the same order of magnitude. Hence $\max_\theta R_{1n}(\theta, m) \sim 1 + m/n + cm^{-\frac{1}{2}}$ as $n \to \infty$. The expression

$m/n + cm^{-\frac{1}{2}}$ has a unique, absolute minimum at $m = (cn/2)^{\frac{2}{3}}$, and finally, taking into account the order of magnitude of ϵ, we obtain Theorem 2.

4. A class of estimators. The estimator $\hat{\mu}_1$, though undoubtedly satisfactory in some instances, is a relatively simple function of the observations, and one may therefore ask if better (in the sense of smaller risk) estimators exist and–if they do–if results like Theorem 2 can be found for such better estimators. We shall see that the answer to both questions is in the affirmative.

We begin by defining $\hat{\mu}_2 = (N_1\theta\bar{X}_{N_1} + N_2\bar{Y}_{N_2})/(N_1\theta + N_2)$, whose risk is $R_{2n}(\theta, m) = n \max(1, \theta)E\{1/(N_1\theta + N_2)\}$; then $R_{2n}(\theta, m)$ is a lower bound for the risk of all estimators of form (1) by Theorem 1(c). Though $\hat{\mu}_2$ has form (1), it is not in fact an estimator, except in the case when θ becomes known at the completion of sampling. If $\hat{\theta} \rightarrow \theta$ in probability, a natural way to obtain a bona fide estimator from $\hat{\mu}_2$ is to replace θ by $\hat{\theta}$. It is mathematically convenient to use a $\hat{\theta}$ based on the first stage only; we take $\hat{\theta} = 1/R$, and define $\hat{\mu}_3 = (N_1\bar{X}_{N_1} + N_2R\bar{Y}_{N_2})/(N_1 + N_2R)$.

The motivation for $\hat{\mu}_3$ may be seen in another way. As mentioned in Section 2, for a one-stage experiment with θ known, $\eta = (n_1\theta\bar{X}_{n_1} + n_2\bar{Y}_{n_2})/(n_1\theta + n_2)$ is the UMVU estimator of μ. When θ is unknown, it is customary in practice to use the estimator obtained from η by replacing θ by $\hat{\theta}$. If one takes $\hat{\theta}$ to be the usual estimator of θ but based on 2 min (n_1, n_2) observations, and if one replaces the numbers n_1, n_2 in η by their values in our procedure, i.e., by the random variables N_1, N_2, the result is $\hat{\mu}_3$.

Another estimator which might be considered is $\hat{\mu}_4$, the grand mean of all the observations: $\hat{\mu}_4 = (N_1\bar{X}_{N_1} + N_2\bar{Y}_{N_2})/n$. Then

$$\begin{aligned} R_{4n}(\theta, m) &= (1/n) \max(1, 1/\theta)E\{N_1 + N_2\theta\} \\ &= (1/n) \max(1, 1/\theta)[(n - m + m\theta)[1 - I(\theta, m)] + [m + (n - m)\theta]I(\theta, m)] \\ &= \max(1, 1/\theta)[1 + (\theta - 1)[m/n + (1 - 2m/n)I(\theta, m)]] \end{aligned}$$

so that $\sup_\theta R_{4n}(\theta, m) > (m/n) \sup_{\theta>1}(\theta - 1)$ which tends to ∞ as $\theta \rightarrow \infty$. We conclude that for $\hat{\mu}_4$ there exist no u.c. solutions and no non-trivial m.m. solutions. (Note that $\hat{\mu}_4$ would be worthy of consideration if it were known *a priori* that θ was close to unity, since $R_{4n}(1, m) = R_{2n}(1, m) = 1$ and since it can be shown that $R_{1n}(\theta, m) > R_{4n}(\theta, m)$ for $\frac{1}{2} \leqq \theta \leqq 2$.)

Before discussing the proposed estimator $\hat{\mu}_3$, it will be convenient to study first $\hat{\mu}_2$ and $R_{2n}(\theta, m)$. By evaluating the expectation,

$$\begin{aligned} &R_{2n}(\theta, m) \\ &\quad = n \max(1, \theta)[[1 - I(\theta, m)]/[(n - m)\theta + m] + I(\theta, m)/(m\theta + n - m)] \\ &\quad = \max(1, 1/\theta)[1 + r_1(\theta, m/n) + r_2(\theta, m/n)I(\theta, m)], \end{aligned}$$

where $r_1(\theta, m/n) = (m(\theta - 1)/n\theta)/(1 - m(\theta - 1)/n\theta)$ and $r_2(\theta, m/n) = (1 - 2m/n)(\theta - 1)/[1 + (m/n)(1 - m/n)(\theta - 1)^2/\theta]$. As for $R_{1n}(\theta, m)$, we have $R_{2n}(\theta, m) = R_{2n}(1/\theta, m)$.

Now let $\hat{\mu}$ be any estimator of form (1) such that $\sup_\theta R_n(\theta, m) \leqq \max_\theta R_{1n}(\theta, m) + o(1)$. If $m/n \to 0$ and $m \to \infty$ as $n \to \infty$, then $\max_\theta R_{1n}(\theta, m)$ and hence $\sup_\theta R_n(\theta, m)$ tend to unity as $n \to \infty$. If on the other hand $\sup_\theta R_n(\theta, m) \to 1$ as $n \to \infty$, then, since $R_{2n}(\theta, m) \leqq R_n(\theta, m)$, $\sup_\theta R_{2n}(\theta, m) \to 1$ as $n \to \infty$. But $\sup_\theta R_{2n}(\theta, m) = \sup_{\theta \geqq 1} R_{2n}(\theta, m)$ and, for $\theta \geqq 1$, $R_{2n}(\theta, m) = 1 + r_1 + r_2 I$ is the sum of three positive terms. Therefore, (i) $\sup_{\theta \geqq 1} r_1(\theta, m/n) \to 0$ as $n \to \infty$, and (ii) $\sup_{\theta \geqq 1} r_2(\theta, m/n) I(\theta, m) \to 0$ as $n \to \infty$. Since $r_1(\theta, m/n) \geqq m(\theta - 1)/n\theta$, (i) implies that $m/n \to 0$ as $n \to \infty$; for any fixed $\theta' > 1$, $r_2(\theta', m/n)$ is bounded away from zero, so that (ii) implies that $I(\theta', m) \to 0$ as $n \to \infty$, and so $m \to \infty$ as $n \to \infty$. We have proved the following result.

THEOREM 3: *For any estimator of form* (1) *whose risk satisfies* $\sup_\theta R_n(\theta, m) \leqq \max_\theta R_{1n}(\theta, m) + o(1)$ *as* $n \to \infty$, $m(n)$ *is* u.c. *if and only if condition* (*) *holds.*

At this point we introduce a lemma.

LEMMA 4: *Let* $\theta = 1 + 2\tau m^{-\frac{1}{2}}$ *and let* C *be a positive number. If* $2\tau > C$, *then* $(\theta - 1)I(\theta, m) < m^{-\frac{1}{2}} e^{-C^2/64}$ *for* m, C *sufficiently large.*

PROOF: Let $L(\theta, m) = m^{\frac{1}{2}}(\theta - 1)I(\theta, m) = 2\tau I(\theta, m) \leqq 2\tau(2\theta^{\frac{1}{2}}/(1 + \theta))^{m-1}$ by Lemma 1. Since $2\theta^{\frac{1}{2}}/(1 + \theta) = 1 - (1/8)(\theta - 1)^2 + O((\theta - 1)^3)$, there exists a number $\theta_1 > 1$ such that $\log (2\theta^{\frac{1}{2}}/(1 + \theta)) < -(\theta - 1)^2/16$ for $1 \leqq \theta \leqq \theta_1$. Then for $1 \leqq \theta \leqq \theta_1$,

$$\log L(\theta, m) \leqq \log 2\tau + (m - 1) \log (2\theta^{\frac{1}{2}}/(1 + \theta))$$
$$< \log 2\tau - (m - 1)(\theta - 1)^2/16 \leqq \log 2\tau - (2\tau)^2/32.$$

For $2\tau > C$ and C larger than some constant, $\log L(\theta, m) < -(2\tau)^2/64 < -C/64$ which proves the lemma for $\theta \leqq \theta_1$. If we now require $m^{\frac{1}{2}} > 2\tau'/(\theta_1 - 1)$, then, by Lemma 3, $(\theta - 1)I(\theta, m)$ is a strictly decreasing function of θ for $\theta \geqq \theta_1$, which completes the proof.

We want next to determine the minimax solution for $\hat{\mu}_2$. For this purpose, we may assume with no loss of generality that $\theta \geqq 1$ and (*) holds. Then, $R_{2n}(\theta, m) = 1 + r_1(\theta, m/n) + r_2(\theta, m/n)I(\theta, m)$, where

$$0 \leqq r_1(\theta, m/n) = \sum_{j=1}^{\infty} (m(\theta - 1)/n\theta)^j < 2m(\theta - 1)/n\theta < 1,$$

$r_2(\theta, m/n) \leqq \theta - 1$ for $\theta \geqq 1$ and $r_2(\theta, m/n) = (\theta - 1)[1 + O(m/n)]$ uniformly in any bounded θ interval. We will employ a four-fold dissection of the set of θ values in order to find $\sup_\theta R_{2n}(\theta, m)$.

Case 1: $\{\theta : 1 \leqq \theta \leqq 1 + Cm^{-\frac{1}{2}}\}$ where C is a large number. Since $r_1(\theta, m/n) < 2Cm^{\frac{1}{2}}/n$, one finds from Lemma 3 that

$$\max_\theta R_{2n}(\theta, m) = 1 + cm^{-\frac{1}{2}} + O(m^{-1}) + O(m^{\frac{1}{2}}/n) = \rho_1, \quad \text{say.}$$

Case 2: $\{\theta : \theta > 64n/m\}$. Since

$$R_{2n}(\theta, m) \to (1 - m/n)^{-1} = 1 + m/n + m^2/n^2 + \cdots = \rho_2,$$

say, as $\theta \to \infty$, and since $(1 - m/n)^{-1} - 1 - r_1(\theta, m/n) > m/n\theta$, we have $\rho_2 - R_{2n}(\theta, m) > m/n\theta - r_2(\theta, m/n)I(\theta, m)$. From Lemma 1,

$$r_2(\theta, m/n)I(\theta, m) < 4(4/\theta)^{(m-3)/2},$$

so that, for $m \geqq 7$, $\rho_2 - R_{2n}(\theta, m) > m/n\theta - 4(4/\theta)^2$ which is positive for $\theta > 64n/m$. Hence, $\sup_\theta R_{2n}(\theta, m) = \rho_2$.

Case 3: $\{\theta : 1 + Cm^{-\frac{1}{2}} < \theta < \theta'\}$ where θ' is a number satisfying $4 < \theta' < 128$. Then,

$$r_1(\theta, m/n) < r_1(\theta', m/n) = m(\theta' - 1)/n\theta' + (m(\theta' - 1)/n\theta')^2 + \cdots$$

and

$$r_2(\theta, m/n)I(\theta, m) < (\theta - 1)I(\theta, m) < m^{-\frac{1}{2}}e^{-C^2/64}$$

by Lemma 4. Suppose $m/n = o(m^{-\frac{1}{2}})$; then

$$R_{2n}(\theta, m) < 1 + m^{-\frac{1}{2}}e^{-C^2/64} + o(m^{-\frac{1}{2}})$$

which is smaller than ρ_1 for C, m large enough. Otherwise suppose $m^{-\frac{1}{2}} = O(m/n)$; then

$$R_{2n}(\theta, m) < 1 + m(\theta' - 1)/n\theta' + e^{-C^2/64}\, O(m/n) + O(m^2/n^2)$$

which is less than ρ_2 for C, m sufficiently large.

Case 4: $\{\theta : \theta' \leqq \theta \leqq 64n/m\}$. Then,

$$r_1(\theta, m/n) \leqq m/n + (63/64)m^2/n^2 + O(m^3/n^3)$$

and

$$r_2(\theta, m/n)I(\theta, m) < 4(4/\theta)^{(m-3)/2} \leqq 4\delta^{m-3}$$

where $\delta^2 = 4/\theta' < 1$. Suppose $\delta^{m-3} = o(m^2/n^2)$; then $R_{2n}(\theta, m) < 1 + m/n + (63/64)\, m^2/n^2 + o(m^2/n^2)$ which is smaller than ρ_2. Else suppose $m^2/n^2 = O(\delta^{m-3})$; then $R_{2n}(\theta, m) < 1 + O(\delta^{(m-3)/2})$ which is smaller than ρ_1.

Combining the results of the dissection, we have

$$\sup_\theta R_{2n}(\theta, m) = \max\,[\rho_1, \rho_2] \tag{3}$$

for sufficiently large m. To minimize (3), we equate ρ_1 and ρ_2, obtaining

$$1 + c^{\frac{2}{3}}n^{-\frac{1}{3}} + O(n^{-\frac{2}{3}}), \tag{4}$$

$$m(n) = (cn)^{\frac{2}{3}} + O(n^{\frac{1}{3}}) \tag{5}$$

as the m.m. risk and m.m. sol., respectively, for $R_{2n}(\theta, m)$. Moreover, if we let $\epsilon_n > 0$ be any function of n which is $O(n^{-\frac{2}{3}})$, the results (4), (5) will hold for any $R_n(\theta, m)$ such that $R_{2n}(\theta, m) \leqq R_n(\theta, m) \leqq R_{2n}(\theta, m) + \epsilon_n$; Theorem 3 will apply also. We summarize these results as follows.

THEOREM 4: *Let $\hat{\mu}$ be any estimator of form* (1) *whose risk satisfies $R_n(\theta, m) \leqq R_{2n}(\theta, m) + \epsilon_n$ where $\epsilon_n = O(n^{-\frac{2}{3}})$. Then for $\hat{\mu}$, the class of u.c. sols. is characterized by* (*), *and the m.m. risk and the m.m. sol. are given by* (4) *and* (5) *respectively.*

Let us return now to the estimator $\hat{\mu}_3 = A_3\bar{X}_{N_1} + B_3\bar{Y}_{N_2}$ where $A_3 =$

$N_1/(N_1 + N_2R)$, whose risk by Theorem 1 is

$$R_{3n}(\theta, m) = n \max(1, 1/\theta)E\{A_3^2/N_1 + \theta B_3^2/N_2\}$$
$$= n \max(1, 1/\theta)E\{(N_1 + N_2R^2\theta)/(N_1 + N_2R)^2\}.$$

Let $Z = R\theta$ so that Z obeys the F-distribution with $m - 1$, $m - 1$ d.f.; denote by $F(z)$ the distribution function of Z, so that $F(z) = 1 - I(z, m)$. Then, $R_{3n}(\theta, m) = \max(\theta, 1)E\{(N_1\theta/n + N_2Z^2/n)/(N_1\theta/n + N_2Z/n)^2\}$, or

$$R_{3n}(\theta, m) = \max(\theta, 1)\int_0^\infty T_z\, dF(z) \qquad \text{where}$$
(6)
$$T_z = ((1-\alpha)\theta + \alpha z^2)/((1-\alpha)\theta + \alpha z)^2, \qquad \alpha = \begin{cases} m/n \text{ if } z < \theta, \\ 1 - m/n \text{ if } z \geqq \theta, \end{cases}$$

or

$$R_{3n}(\theta, m) = \max(\theta, 1)\left[\int_0^\theta \frac{(1 - m/n)\theta + (m/n)z^2}{[(1 - m/n)\theta + (m/n)z]^2}\, dF(z)\right.$$
(7)
$$\left. + \int_\theta^\infty \frac{(m/n)\theta + (1 - m/n)z^2}{[(m/n)\theta + (1 - m/n)z]^2}\, dF(z)\right],$$

where $dF(z) = [B((m-1)/2, (m-1)/2]^{-1} z^{(m-3)/2}(1 + z)^{1-m}\, dz$. By substituting $w = 1/z$ in (7), using the fact that $dF(z) = -dF(1/z)$, and simplifying, one shows that $R_{3n}(\theta, m) = R_{3n}(1/\theta, m)$.

Next, write $T_1 = 1/((1-\alpha)\theta + \alpha)$ so that $R_{2n}(\theta, m) = \max(\theta, 1)\int_0^\infty T_1\, dF(z)$. Using (6), define $D_n(\theta, m) = R_{3n}(\theta, m) - R_{2n}(\theta, m) = \max(\theta, 1)\int_0^\infty U_z\, dF(z)$ where $U_z = T_z - T_1 = \alpha(1-\alpha)\theta(z-1)^2/[((1-\alpha)\theta + \alpha z)^2((1-\alpha)\theta + \alpha)]$. Obviously $D_n(\theta, m) \geqq 0$ and $D_n(\theta, m) = D_n(1/\theta, m)$ by the properties of the risk functions; moreover, $D_n(\theta, m) > 0$ for $0 < \theta < \infty$ since $U_z > 0$ except when $z = 1$, an event of measure zero. And since $U_z \leqq (m/n)(1 - m/n)^{-1}(z-1)^2/\theta$ for all z and for $\theta \geqq 1$, we have $D_n(\theta, m) \leqq (m/n)(1 - m/n)^{-1}\int_0^\infty (z-1)^2\, dF(z)$ for all θ. But $\int_0^\infty (z-1)^2\, dF(z) = 4(m+1)/(m^2 - 8m + 15) \leqq 56/m$ for $m \geqq 6$. Since $(1 - m/n)^{-1} < 2$, we have proved that

(8) $$D_n(\theta, m) < 112/n, \qquad \text{for all } \theta \text{ and } 6 \leqq m < n/2.$$

Therefore, the solution for the proposed estimator $\hat{\mu}_3$ is specified by Theorem 4.

The last result implies that the m.m. risk for $\hat{\mu}_3$ is $\sim 1 + c^{\frac{1}{2}}n^{-\frac{1}{2}}$ and from Theorem 2 the m.m. risk for $\hat{\mu}_1$ is $\sim 1 + 3(c/2)^{\frac{1}{3}}n^{-\frac{1}{3}}$. Since $c^{\frac{1}{2}} < 3(c/2)^{\frac{1}{3}}$, this provides one reason for preferring $\hat{\mu}_3$ to $\hat{\mu}_1$; an additional reason is given by the following result: Given any θ, $0 < \theta < \infty$, one can find an m_0 such that $R_{3n}(\theta, m) < R_{1n}(\theta, m)$ for $m \geqq m_0$. This is easy to see since $R_{1n}(\theta, m) - R_{2n}(\theta, m) = \int_0^\infty (W - \theta T_1)\, dF(z)$, where $W = (1 - m/n)^{-1}$ if $z < \theta$, $W = \theta(1 - m/n)^{-1}$ if $z \geqq \theta$, and where we assume with no loss of generality that $\theta \geqq 1$. Noting that $W - \theta T_1 > m/\theta n$ for all z, and introducing (8), one obtains the desired result.

We see then that the main result of this paper is embodied in Theorem 4 where it is shown that the solution (5) holds throughout a certain class of best estimators. Moreover, by the inequality (8), it is shown that $\hat{\mu}_3$ is a member of this class and that, if there exists an estimator of form (1) with uniformly smaller risk than that of $\hat{\mu}_3$, it has the same large sample solution.

The author is deeply indebted to Professor Wassily Hoeffding for suggesting this problem, and for his advice and encouragement throughout the course of the work. The author is grateful to editor W. Kruskal and the referee for their helpful comments.

REFERENCES

[1] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.*, Vol. 49 (1941), pp. 122–136.

[2] S. G. Ghurye and Herbert Robbins, "Two-stage procedures for estimating the difference between means," *Biometrika*, Vol. 41 (1954), pp. 146–152.

[3] Joseph Putter, "Sur une méthode de double échantillonnage pour estimer la moyenne d'une population laplacienne stratifiée," *Rev. Inst. Internat. Stat.*, Part 3 (1951), pp. 1–8.

[4] Herbert Robbins, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, Vol. 58 (1952), pp. 527–535.