The Annals of Statistics
1976, Vol. 4, No. 6, 1077–1087

SEQUENTIAL ESTIMATION OF THE LARGEST NORMAL MEAN WHEN THE VARIANCE IS KNOWN[1]

BY SAUL BLUMENTHAL

University of Kentucky

Given a procedure $\hat{\theta}^*$ for estimating the largest mean θ^* of k normal populations with common known variance, it is desired to choose the common sample size n so that the mean squared error (M.S.E.) of $\hat{\theta}^*$ does not exceed a given bound, r, regardless of the configuration of values of the k means.

Let $\Delta_1 \geqq \cdots \geqq \Delta_k = 0$ be the ordered values of $(\theta^* - \theta_i)$, where $\theta_1, \cdots, \theta_k$ are the unknown means. The M.S.E. depends on the Δ's and the conservative approach chooses a sample size n^* to hold M.S.E. $\leqq r$ for all Δ's. Sequential and multisample procedures are considered which use sample information about the Δ's to reduce sample sizes. Asymptotic properties of the sample size and M.S.E. of the resulting estimates are developed. Improvements over using n^* are possible, but with limitations. The sample size behavior of any $\hat{\theta}^*$ depends on the limiting variance of the estimator as all of $\Delta_1, \cdots, \Delta_{k-1}$ become infinite.

1. Intoduction. Let $\theta_1, \cdots, \theta_k$ be the unknown means of k normal populations with common known variance σ^2 which will be taken as unity henceforth. Let $\bar{X}_1, \cdots, \bar{X}_k$ be the sample means of n observations taken from each population. Let $\theta_{[1]} \leqq \cdots \leqq \theta_{[k]}$ and $\bar{X}_{[1]} \leqq \cdots \leqq \bar{X}_{[k]}$ be the ordered population and sample means respectively. Let $\Delta_i = \theta_{[k]} - \theta_{[i]}$ and $\hat{\Delta}_i$ be the strongly consistent estimates $\bar{X}_{[k]} - \bar{X}_{[i]}$ $(1 \leqq i \leqq k)$. Define θ^* as $\theta_{[k]}$ and X^* as $X_{[k]}$. Consider estimation procedures $\hat{\theta}^*$ for θ^* which have the form

$$\hat{\theta}^* = X^* - (1/n^{\frac{1}{2}})\gamma(\hat{\Delta}_1 n^{\frac{1}{2}}, \cdots, \hat{\Delta}_{k-1} n^{\frac{1}{2}}) \tag{1.1}$$

for some real valued function $\gamma(\cdot)$ defined on E^{k-1}. Note that all the estimators examined in [3], [5] and [6] are of this form. The mean squared error function of $\hat{\theta}^*$ is given by

$$R(\hat{\theta}^*; \theta_1, \cdots, \theta_k) = (1/n)H_{\hat{\theta}^*}(n^{\frac{1}{2}}\Delta_1, \cdots, n^{\frac{1}{2}}\Delta_{k-1}) , \tag{1.2}$$

Received October 1974; revised March 1976.

[1] Some of the work in this paper was performed while the author was a visiting professor at Cornell University on leave from New York University, and was supported in part by Office of Naval Research Contract ONR N00014-67-A-0077-0020 at the College of Engineering, Cornell University, by National Science Foundation Grant GK 14073 at the School of Engineering and Science, New York University, and by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-75-2841. The United States government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

AMS 1970 *subject classifications.* Primary 62F07; Secondary 62L12.

Key words and phrases. Sequential estimation, estimation, largest mean, ordered parameters, ranking and selection.

where

$$H_{\hat\theta^*}(x_1, \cdots, x_{k-1}) \\ = \int (y + \gamma(u_1, \cdots, u_{k-1}))^2 \sum_{\beta \in S_k} \prod_{i=1}^{k} \phi(y - u_{\beta(i)} + x_i)\, dy \prod_{i=1}^{k-1} du_i ,$$

S_k is the group of permutations of k elements, $\phi(\cdot)$ is the standard normal density, $u_k = x_k = 0$ in the argument of $\phi(\cdot)$, and the region of integration is $0 \leqq y \leqq \infty$, $0 \leqq u_{k-1} \leqq \cdots \leqq u_1 < \infty$. For a given desired level of risk r and a given estimation procedure, it is desired to choose n such that

$$R(\hat\theta^*; \theta_1, \cdots, \theta_k) \leqq r , \qquad \text{all} \quad (\theta_1, \cdots, \theta_k) . \tag{1.3}$$

Achievement of this goal is complicated by the presence of the unknown Δ's in the risk function. The conservative approach would be to use $n = n^*$ where

$$n^* = (1/r) H_{\hat\theta^*, \max} , \tag{1.4}$$

and

$$H_{\hat\theta^*, \max} = \sup_{x_1, \cdots, x_{k-1}} H_{\hat\theta^*}(x_1, \cdots, x_{k-1}) .$$

We consider multiple-stage or sequential procedures which attempt to reduce the sample size below n^* by using the information available about the Δ_i. Let

$$n_\Delta = \inf \{ n : (1/n) H_{\hat\theta^*}(n^{\frac12}\Delta_1, \cdots, n^{\frac12}\Delta_{k-1}) \leqq r \} . \tag{1.5}$$

The form of the rules is: after obtaining t observations from each population, compute estimates $\hat\Delta_i$ and compute $\hat n(t) = n_{\hat\Delta}$ by using the $\hat\Delta$'s in (1.5).

A multistage procedure which has taken t_j observations up to stage j will observe on the next stage: max $(0, \hat n(t_j) - n_j)$ additional values from each population. A sequential rule simply observes the $(t+1)$st value if $\hat n(t) > t$, and stops otherwise. Let N denote the sample size at stopping for any of these rules.

Throughout, we require that any sequential or multistage estimator take at least n_0 observations initially, where

$$n_0 = [(1/r) H_{\hat\theta^*, \min}] , \\ H_{\hat\theta^*, \min} = \inf_{\infty > x_{k-1} \geqq \cdots \geqq x_1 > 0} H_{\hat\theta^*}(x_1, \cdots, x_{k-1}) \tag{1.6}$$

and $[x]$ denotes the smallest integer not less than x. Clearly, $\hat n(t) \geqq n_0$ so that any sequential procedure must eventually reach the n_0th observation and the computations made at earlier stages serve no purpose; multistage procedures might start with fewer than n_0 observations, then jump to more than n_0, but intuitively it seems more efficient to take as large a first sample as possible when it is known that the total sample must exceed n_0 in any case.

The main result of this paper (Theorem 1) is that if we force n_Δ to be large by making r small, then (N/n_Δ) is near unity. Our sequential procedure is designed to try to achieve equality in (1.3) for the risk of the estimator based on the random sample size. In the case $k = 2$, Theorem 2 affirms that this goal is achieved for two-stage sampling. The Δ's are assumed to be fixed and positive in these theorems. If the Δ_i are all zero, or are proportional to $r^{\frac12}$ when r is small, then Theorem 3 and Corollary 1 show that (N/n_Δ) does not have an almost sure limit. The ratio does have a limiting distribution in that case and this is examined in Appendix 2. In Section 3, some numerical results are obtained for the mean squared error actually achieved by the two-stage procedure for two populations and the estimator X^*.

2. Fixed Δ's as r becomes small. First, (N/n_Δ) is examined. Note that for two stage procedures,

$$N = \hat{n}(n_0)\,. \tag{2.1}$$

For multistage procedures, the method of choosing the sample sizes implies

$$N \geqq \hat{n}(n_0)\,. \tag{2.2}$$

THEOREM 1. *Given the strongly consistent estimates* $\hat{\Delta}_{i,n} = \bar{X}_{[k]} - \bar{X}_{[i]}$ *of the* Δ_i. *Let all* $\Delta_i (1 \leqq i \leqq k-1)$ *be* > 0. *Then*

$$\lim_{r\to 0} rN = \lim_{r\to 0} r\hat{n}(N) = H_{\hat{\theta}^*,\infty} \quad \text{a.s.,} \tag{2.3}$$

$$\lim_{r\to 0} E(rN)^j = \lim_{r\to 0} E(r\hat{n}(N))^j = (H_{\hat{\theta}^*,\infty})^j\,, \quad \textit{for any} \tag{2.4}$$

integer j (*positive or negative*) *where*

$$H_{\hat{\theta}^*,\infty} = \lim_{x_i\to\infty, i=1,\cdots,k-1} H_{\hat{\theta}^*}(x_1, \cdots, x_{k-1})\,. \tag{2.5}$$

PROOF. By the strong consistency of the $\hat{\Delta}_{i,n}$, for given $\varepsilon > 0$, there is an M such that

$$P\{\hat{\Delta}_{i,n} > (\Delta_i/2),\ 1 \leqq i \leqq k, \ \ \textit{all} \ \ n > M\} > 1 - \varepsilon\,. \tag{2.6}$$

From (1.6), it is seen that for $r \leqq r_0$ (say), $n_0 \geqq M$. From (2.6), the fact that $\hat{n} > n_0$, and (1.6) we see that r_1 can be chosen so that for any arbitrarily large T,

$$P\{\hat{\Delta}_{i,n}(\hat{n}(n))^{\frac{1}{2}} > T,\ 1 \leqq i \leqq k-1, \ \ \text{all} \ \ n \geqq n_0\} > 1 - \varepsilon\,, \quad \text{if} \quad r < r_1\,. \tag{2.7}$$

Since $\lim H_{\hat{\theta}^*}(x_1, \cdots, x_{k-1})$ exists as all the x_1 increase, and is $H_{\hat{\theta}^*,\infty}$ from (2.7) and the fact that T is arbitrary, we conclude from (1.5) that for arbitrary $\delta > 0$,

$$P\{H_{\hat{\theta}^*,\infty} - \delta < r\hat{n}(n) < H_{\hat{\theta}^*,\infty} + \delta, \ \ \text{all} \ \ n \geqq n_0\} > 1 - \varepsilon \quad \text{if} \quad r < r_1\,. \tag{2.8}$$

This completes the proof of (2.3) for $\hat{n}(N)$.

For two-stage procedures, (2.3) for N is immediate from (2.1). The additional steps for sequential and multistage procedures will be omitted. The dominated convergence theorem gives (2.4), noting that N is bounded both above (for positive j) and below (for negative j). This completes the proof of the theorem.

To see that (2.3) really implies

$$\lim_{r\to 0} (N/n_\Delta) = 1 \quad \text{a.s.} \tag{2.9}$$

we use the following:

$$\lim_{r\to 0} rn_\Delta = H_{\hat{\theta}^*,\infty}\,. \tag{2.10}$$

To obtain (2.10), write n_Δ as $[\rho^2/r]$ where ρ^2 is the smallest solution of

$$\rho^2 = H_{\hat{\theta}^*}((\rho\Delta_1/r^{\frac{1}{2}}), \cdots, (\rho\Delta_{k-1}/r^{\frac{1}{2}})) \,. \tag{2.11}$$

If $\rho(= \rho(r))$ has a limit as r decreases, and if all Δ_i are > 0, the limit will satisfy (2.11) with r replaced by zero on the right side, and (2.10) follows directly.

We examine the asymptotic moments of $\hat{\theta}_N{}^*$ to see whether the goal of having $(1/r)E(\hat{\theta}_N{}^* - \theta^*)^2$ approach unity is achieved. We have not answered this question in full generality, but for $k = 2$ and two-stage sampling, Theorem 2 gives an affirmative answer.

Additional structure will be imposed on the form of $\hat{\theta}_\cdot^*$, namely

$$\hat{\theta}^* = X^* - (\hat{\Delta}/2)[1 - \psi(n^{\frac{1}{2}}\hat{\Delta})] + (1/n^{\frac{1}{2}})\lambda(n^{\frac{1}{2}}\hat{\Delta}) \tag{2.12}$$

where $\hat{\Delta} \equiv \hat{\Delta}_1$, and both $\psi(\cdot)$ and $\lambda(\cdot)$ are bounded functions having, as their arguments increase, finite limits denoted as ψ_∞ and λ_∞ respectively. (Note that for X^*, $\psi = 1$, $\lambda = 0$, and all of the other estimators in [3] can be expressed this way).

The following two lemmas are proved in Appendix 1.

LEMMA 1. *Let* $\hat{\theta}^*$ *be given by* (2.12) *and* $\Delta \equiv \Delta_1$, *then*

$$\lim_{r\to 0} \hat{\theta}_N{}^* = \theta^* - (\Delta/2)(1 - \psi_\infty) \,. \tag{2.13}$$

Thus $\hat{\theta}_N{}^*$ *is consistent* iff $\psi_\infty = 1$.

LEMMA 2. *Let* $\hat{\theta}^*$ *be given by* (2.12). *Then*

$$\lim_{r\to 0} (1/r^{\frac{1}{2}})E(\hat{\theta}_N{}^* - \theta^*) = (1/(H_{\hat{\theta}^*,\infty})^{\frac{1}{2}})(L_\infty + \lambda_\infty) \,, \tag{2.14}$$

$$\lim_{r\to 0} (1/r)E(\hat{\theta}_N{}^* - \theta^*)^2 = (1/H_{\hat{\theta}^*,\infty})[1 + (L_\infty + \lambda_\infty)^2] \,, \tag{2.15}$$

$$H_{\hat{\theta}^*,\infty} = 1 + (L_\infty + \lambda_\infty)^2 \,, \tag{2.16}$$

where $L_\infty = \lim_{x\to\infty} x(\psi(x) - 1)$. *If* L_∞ *is infinite, the limits on the left in* (2.14), (2.15) *and* (2.16) *are infinite.*

Note that many of the estimators studied in [3] (such as maximum likelihood) satisfy $\lambda_\infty = 0$ and $L_\infty = 0$. For instance, the generalized Bayes estimator has $(\psi(x) - 1) = 2\Phi(-x/2^{\frac{1}{2}})$, and $\lambda(x) = 2^{\frac{1}{2}}\phi(x/2^{\frac{1}{2}})$. Now we give our main result on the moments of $\theta_N{}^*$.

THEOREM 2. *For* $k = 2$, *two-stage sampling, and* $\hat{\theta}^*$ *of form* (2.12), *with* $L_\infty < \infty$,

$$\lim_{r\to 0} (1/r^{\frac{1}{2}})E(\hat{\theta}_N{}^* - \theta^*) = (L_\infty + \lambda_\infty)/(1 + (L_\infty + \lambda_\infty)^2)^{\frac{1}{2}} \tag{2.17}$$

$$\lim_{r\to 0} (1/r)E(\hat{\theta}_N{}^* - \theta^*)^2 = 1 \,. \tag{2.18}$$

The estimator is strongly asymptotically unbiased iff $H_{\hat{\theta}^*,\infty} = 1$.

PROOF. From (2.16), we see that H_∞ is finite iff L is finite, and (2.17) comes from (2.14) with (2.16), while (2.18) uses (2.15) with (2.16).

3. Varying Δ's, as r becomes small. Even though the function H may have a minimum value considerably lower than H_∞ or $H_{\max}$ (for a specified $\hat{\theta}^*$), Theorem 1 indicates that the N which takes advantage of it is elusive. This section explores to what extent this N can be achieved.

Define n_x as

$$n_x = (1/r)H_{\hat{\theta}^*}(x_1, \cdots, x_{k-1}) . \tag{3.1}$$

The Δ values which correspond to n_x are given by $x_i = \Delta_i(n_x)^{\frac{1}{2}}$ or

$$\Delta_i = \Delta_i(x) = x_i r^{\frac{1}{2}}(H_{\hat{\theta}^*}(x_1, \cdots, x_{k-1}))^{-\frac{1}{2}} . \tag{3.2}$$

Thus, the parameter value which leads to large potential savings is proportional to $r^{\frac{1}{2}}$, and the question is whether, if we fix $x = (x_1, \cdots, x_{k-1})$ and look at $\Delta_i(x)$, does (N/n_x) approach unity almost surely.

THEOREM 3. *Let x be fixed and the Δ_i given by* (3.2). *Let the estimates $\hat{\Delta}_i$ be $X^* - \bar{X}_{[i]}$ $(i = 1, \cdots, k-1)$. Define*

$$S(r, \delta) = \{\Delta = (\Delta_1, \cdots, \Delta_{k-1}) \mid n_\Delta > (\delta/r)H_{\hat{\theta}^*, \max}\} . \tag{3.3}$$

Let cA denote the set obtained by multiplying each element of the set A by the scalar c. Assume there is some $\delta_0 < 1$ such that for any $\delta_0 < \delta < 1$, there exists a set B_δ of positive Lebesgue measure such that $r^{\frac{1}{2}}B_\delta \subset S(r, \delta)$. Then for any x such that $n_x < n^*$, there is an $\varepsilon > 0$ such that for multistage procedures

$$\lim_{r\to 0} P\{|(N/n_x) - 1| \geqq \varepsilon\} > 0 . \tag{3.4}$$

For any x such that $n_x > n_0$, (3.4) *holds for sequential procedures if $S(r, \delta)$ is defined by $n_\Delta < \delta n^*$, and B_δ exists for all $\delta < \delta_0 < 1$.*

PROOF. For brevity, only the proof for multistage procedures will be given. Let $n = (c/r)$ (any $c > 0$). Let $\hat{\Delta} = (\hat{\Delta}_1, \cdots, \hat{\Delta}_{k-1})$. Note that

$$\begin{aligned} P\{\hat{n}(n) \geqq \delta\} &= P\{\hat{\Delta} \in S(r, \delta)\} = P\{n^{\frac{1}{2}}\hat{\Delta} \in n^{\frac{1}{2}}S(r, \delta)\} \\ &= P\{n^{\frac{1}{2}}\hat{\Delta} \in (c/r^{\frac{1}{2}})S(r, \delta)\} \geqq P\{n^{\frac{1}{2}}\hat{\Delta} \in cB_\delta\} \end{aligned} \tag{3.5}$$

(the last inequality uses the assumption). Observe that the $\hat{\Delta}_i$ are the ordered values of the nonnegative $(X^* - \bar{X}_i)$, that $n^{\frac{1}{2}}(X^* - \bar{X}_i)$ is distributed as $(Y^* - Y_i)$ where the Y_i's are normal with means $-(n\Delta_i)^{\frac{1}{2}}$ and unit variance. If $n = (c/r)$ (any $c > 0$), and Δ_i is given by (3.2), these means are $(c^{\frac{1}{2}}x_i/(H_{\hat{\theta}^*}(x_1, \cdots, x_{k-1}))^{\frac{1}{2}})$, independently of r. For n of this form, the joint distribution of the positive values of $n^{\frac{1}{2}}\hat{\Delta}_i$ assigns positive weight to all measurable $(k-1)$ dimensional sets in the upper orthant having positive Lebesgue measure with weight independent of r. Thus, since (cB_δ) is such a set,

$$P\{\hat{n}_n \geqq \delta\} > 0 . \tag{3.6}$$

If n is taken to be n_0, then $c = H_{\hat{\theta}^*, \min}$, and (2.2) along with (3.6) implies that

$$P\{N \geqq \delta\} > P\{\hat{n}(n_0) \geqq \delta\} > 0 .$$

For given x, δ can be taken as $(1 + \varepsilon)(n_x/n^*)$ where ε is sufficiently large to assure $\delta > \delta_0$. This completes the proof.

REMARKS.

1. The conditions involving $S(r, \delta)$ and $B(\delta)$ can be verified if the function H satisfies mild continuity conditions near its maximum or minimum. For instance, when $\hat{\theta}^* = X^*$, these sets can be given explicitly and the conditions verified directly.

2. Setting $x = (0, \cdots, 0)$ shows that the result of Theorem 1 does not hold when all Δ_i's are zero.

3. The role of σ has been suppressed by taking $\sigma = 1$. If $\sigma \neq 1$, then (r/σ^2)

TABLE 1

Comparative mean squared errors and biases

x	$\Delta(x)(n^*)^{\frac{1}{2}}$	(n_x/n^*)	$\lim_{n^*\to\infty}(E(N)/n^*)$	$\lim_{n^*\to\infty}(1/r)E(X_N^* - \theta^*)^2$	$(1/r)^{\frac{1}{2}}E(X_{n_x}^* - \theta^*)$	$\lim_{n^*\to\infty}(1/r)^{\frac{1}{2}}E(X_N^* - \theta^*)$
.0	.0000	1.0000	.8662	1.1028	.5642	.5327
.1	.1037	.9298	.8664	1.0337	.5148	.4573
.2	.2135	.8772	.8671	.9786	.4634	.3836
.3	.3274	.8399	.8683	.9383	.4116	.3139
.4	.4429	.8156	.8702	.9121	.3608	.2503
.5	.5582	.8022	.8725	.8980	.3123	.1940
.6	.6719	.7976	.8753	.8936	.2671	.1454
.7	.7826	.8000	.8785	.8963	.2259	.1043
.8	.8901	.8077	.8820	.9041	.1892	.0703
.9	.9943	.8192	.8858	.9150	.1569	.0425
1.0	1.0945	.8334	.8897	.9279	.1291	.0200
1.1	1.1937	.8490	.8939	.9417	.1053	.0021
1.2	1.2900	.8654	.8982	.9559	.0853	−.0120
1.3	1.3845	.8816	.9026	.9699	.0686	−.0228
1.4	1.4780	.8973	.9072	.9835	.0547	−.0308
1.5	1.5706	.9121	.9118	.9963	.0434	−.0365
1.6	1.6630	.9256	.9165	1.0084	.0342	−.0402
1.7	1.7555	.9378	.9213	1.0194	.0267	−.0423
1.8	1.8481	.9486	.9261	1.0293	.0207	−.0430
1.9	1.9410	.9580	.9309	1.0380	.0160	−.0425
2.0	2.0349	.9660	.9357	1.0455	.0122	−.0411
2.5	2.5126	.9900	.9583	1.0643	.0028	−.0269
3.0	3.0035	.9977	.9764	1.0570	.0005	−.0120
3.5	3.5007	.9996	.9883	1.0398	.0001	−.0037
4.0	4.0001	.9999	.9949	1.0224	.0000	−.0005

replaces r and (Δ/σ), $(\hat{\Delta}/\sigma)$ replace Δ and $\hat{\Delta}$ respectively in all formulas. If r is fixed at r_0 in (1.3), n will be large if σ^2 is large, and we might ask what happens to (N/n_Δ) for fixed Δ as σ^2 increases. Since r is (r_0/σ^2), σ is $(r_0/r)^{\frac{1}{2}}$ so that the Δ's become proportional to $r^{\frac{1}{2}}$ (when (Δ/σ) is used). This allows Theorem 3 to be applied to give

COROLLARY 1. *Let $\Delta = (\Delta_1, \cdots, \Delta_{k-1})$ be fixed, and let $\hat{\Delta}_i = X^* - X_{[1]}$. Then for any Δ, such that $n_0 < n_\Delta < n^*$ there is an $\varepsilon > 0$ such that*

$$\lim_{\sigma^2\to\infty} P\{|(N/n_\Delta) - 1| \geqq \varepsilon\} > 0\,. \tag{3.7}$$

Although (N/n_x) does not have a probability limit, the proof of Theorem 3 suggests that it should have a distribution which has a limit as $r \to 0$. For the case $k = 2$ and two stage sampling, the limiting moments of N and the low order moments of $\hat{\theta}_N{}^*$ are obtained in Appendix 2 for general $\hat{\theta}^*$, and specialized to X^*. The formulas are quite complex, and numerical integrations have been performed to help ascertain the performance of N and $X_N{}^*$. In Table 1, x is an indexing variable, n_x is given by (3.1) and $\Delta(x)$ by (3.2). Columns two and three, respectively, give $\Delta(x)(n^*)^{\frac{1}{2}}$ and n_x/n^* (note $n^* = 1/r$) to eliminate r from the table. Column 4 (see (A2.15)) should be compared with column 3 to determine the sample size effectiveness of the two sample rule. Column 5 (see (A2.14)) should be compared to the ideal value $(1/r)E(X^*_{n_x} - \theta^*)^2$ which is unity. Column 6 gives the normalized bias of the perfect information procedure, namely $(1 - H(x))/x(2H(x))^{\frac{1}{2}}$ and column 7 (see (A2.11)) should be compared to this.

4. Conclusions. Although Theorem 1 indicates that our sequential procedure does not take full advantage of the possible sample size savings for known Δ's, it can improve on the conservative approach since $H_{\hat{\theta}^*,\infty} < H_{\hat{\theta}^*,\max}$ for many estimation procedures. Using our sequential plans, all estimators with the same $H_{\hat{\theta}^*,\infty}$ are equally desirable. Note that $H_{X^*,\infty} = 1$ and that heuristically, there is only one observation available to estimate θ^* when $(k - 1)$ of the Δ's are infinite, which leads to the conjecture that $H_{\hat{\theta}^*,\infty} \geqq 1$ for *any* $\hat{\theta}^*$. If this is true, then no procedure can really improve on X^* (in terms of sample size needed to achieve (1.3)), and none are as simple in form. The conjecture implies that the minimax value of $H_{\hat{\theta}^*}(\Delta_1, \cdots, \Delta_{k-1})$ is at least unity, so that any $\hat{\theta}^*$ with $H_{\hat{\theta}^*,\infty} = 1$ leads to an asymptotically minimax procedure in conjunction with our sequential approach.

It is conjectured that Theorem 2 holds also for sequential sampling and $k > 2$. If so, the various $\hat{\theta}^*$'s behave the same on the basis of mean squared error and should be compared according to sample size behavior as discussed above.

Theorems 1 and 3 indicate that our two-sample plan is asymptotically equivalent to the more complex plans in terms of sample size. Theorem 2 indicates that this sampling plan achieves the desired moment behavior of $\hat{\theta}_N{}^*$. Also, since $n_0 < \hat{n}(n) < n^*$ by construction, even our sequential plan is curtailed, and since (n_0/n^*) is a constant (nonzero), as n_0 and n^* increase the information about the

Δ_i's contained in the initial sample of n_0 is almost as great as the information available by looking at n observations for any $n_0 < n \leqq n^*$. Thus intuitively, the simple two-stage sampling scheme should be almost as efficient as any sequential rule as well as being much simpler to analyze, and to use. Table 1 shows that $(E(N)/n_x)$ is not too close to unity, and that for small values of Δ, M.S.E. (X_N^*) fails to achieve the goal of being constant at unity. Nonetheless, it is seen that the two-sample procedure does take some advantage of the possible savings available when Δ is known, and the M.S.E. curve does not rise very far above unity. Compared to the conservative procedure, about a 10% saving in sample size is achievable for moderate values of x. Thus the two-sample procedure may very well be of practial value.

APPENDIX 1

PROOFS OF LEMMAS 1 AND 2. Use the notation of [3], namely

$$Z_{0,n} = (X_{1,n} + X_{2,n})/2\,; \qquad Z_{1,n} = (X_{1,n} - X_{2,n})/2\,; \qquad Z_n = |Z_{1,n}| = (\hat{\Delta}/2)$$
$$\eta = (\theta_1 + \theta_2)/2\,; \qquad \nu = (\theta_1 - \theta_2)/2\,; \qquad \omega = |\nu| = (\Delta/2)\,; \qquad \theta^* = \eta + \omega$$
$$\hat{\theta}^* = Z_{0,n} + (1/n^{\frac{1}{2}})\xi(n^{\frac{1}{2}}Z_n)\,; \qquad \xi(x) = x - \gamma(2x)\,, \quad (\text{see } (1.1))\,.$$

Note that $Z_{0,n}$ and $Z_{1,n}$ are independent, normal with respective means η and ν, and variances $(1/2n)$. The risk (1.2) is a function of ω and the stopping rules depend only on Z_n. A straightforward consequence of the above characterizations and the independence of $Z_{0,n}$ from N is that for any stopping rule

$$E(\hat{\theta}_N^* - \theta^*)^p = \textstyle\sum_{i=0}^{p} \binom{p}{i} E(W^i) E\{[(1/N^{\frac{1}{2}})\xi(N^{\frac{1}{2}}Z_N) - \omega]^{p-i}(2N)^{-\frac{1}{2}}\}\,, \tag{A1.1}$$

where W is a standard normal variable, and p is any positive integer. To evaluate the term in { } in (A1.1), the two-sample rule will be assumed. Writing

$$Z_{1,n} = \frac{1}{N}[n_0 Z_{1,n_0} + (N - n_0)Z_{1,(N-n_0)}]\,, \tag{A1.2}$$

where $(N - n_0)Z_{1,(N-n_0)}$ is the sum of $(N - n_0)$ normal random variables, with mean ω and variance $\frac{1}{2}$, which are independent of Z_{1,n_0} and of one another, and writing

$$R_N = (H_{\hat{\theta}^*,\infty}/rN)\,, \qquad R_{0,N} = (n_0/N)\,, \tag{A1.3}$$

it is straightforward to arrive at the characterization for positive ν,

$$Z_{1,N} = (rR_N/2H_{\hat{\theta}^*,\infty})^{\frac{1}{2}}\{(R_{0,N})^{\frac{1}{2}}Y_1 + (1 - R_{0,N})^{\frac{1}{2}}Y_2 + (\omega(2H_{\hat{\theta}^*,\infty}/rR_N)^{\frac{1}{2}})\}\,, \tag{A1.4}$$

where Y_1 and Y_2 are independent standard normal variables, R_N and $R_{0,N}$ are functions of Y_1, and

$$\lim_{r\to 0} R_N = 1\,, \quad \text{a.s.}, \qquad \lim_{r\to 0} R_{0,N} = (H_{\hat{\theta}^*,\min}/H_{\hat{\theta}^*,\infty})\,, \quad \text{a.s.} \tag{A1.5}$$

For negative ν, replace ω by $(-\omega)$ in (A1.4). The characterization (2.12) becomes

$$\hat{\theta}^* = Z_{0,n} + Z_n \psi((n)^{\frac{1}{2}} Z_n) + (1/(n)^{\frac{1}{2}})\lambda((n)^{\frac{1}{2}} Z_n) . \tag{A1.6}$$

From (A1.4) and (A1.5) the a.s. convergence of Z_n to ω (as $r \to 0$) follows and that of $Z_{0,n}$ to η follows from independence of $Z_{0,N}$ and N. These convergences used in (A1.6) give Lemma 1.

Next, (A1.4) and (A1.5) are used to establish

$$\lim_{r\to 0} (1/(r)^{\frac{1}{2}}) E(Z_N - \omega) = 0 , \tag{A1.7}$$

$$\lim_{r\to 0} (1/(r)^{\frac{1}{2}}) E(X_N^* - \omega) = 0 , \tag{A1.8}$$

$$\lim_{r\to 0} (1/r) E(Z_N - \omega)^2 = (2H_{\hat{\theta}^*,\infty})^{-1} , \tag{A1.9}$$

$$\lim_{r\to 0} (1/r) E(X_N^* - \omega)^2 = (1/H_{\hat{\theta}^*,\infty}) . \tag{A1.10}$$

For instance, to get (A1.9), (A1.4) is used to obtain

$$\begin{aligned}(E(Z_N - \omega)^2/r) = (1/2H_\infty)E\{R_N R_{0,N} Y_1^2 + R_N(1 - R_{0,N}) \\ - (2(2)^{\frac{1}{2}}\omega/(rH_\infty)^{\frac{1}{2}})(B\phi(A/B) - A\Phi(-A/B))\} ,\end{aligned} \tag{A1.11}$$

where

$$A = (R_{0,N})^{\frac{1}{2}} Y_1 + \omega(2H_\infty/rR_N)^{\frac{1}{2}} , \qquad B = (1 - R_{0,N})^{\frac{1}{2}} . \tag{A1.12}$$

Take the limit as $r \to 0$ in (A1.11) using (A1.5) and dominated convergence for the first two terms. To see that the expectation of the third term (which has a $(1/(r)^{\frac{1}{2}})$ factor) approaches zero, note that for fixed Y_1, the tail approximation to the normal shows pointwise convergence, and the integral can be split into two parts—one where the pointwise limit can be taken, and one where A is small ($Y_1 \ll c/(r)^{\frac{1}{2}}$ for some c), in which case $(B\phi(\cdot) - A\Phi(\cdot))$ is bounded and the entire integral is small enough to give (A1.9).

Now (2.14) and (2.15) can be established. An outline for (2.15) follows. Find the limit of $E((Z\psi + \lambda - \omega)^2/r)$ then add $(1/2H_\infty)$. Rewrite $(1/r)(Z\psi + \lambda - \omega)^2$ as $(1/r)(Z-\omega)^2\psi + \psi(\psi - 1)((Z^2 - \omega^2)/r) + \lambda^2(1/rN) + 2\lambda(1/(rN)^{\frac{1}{2}})((Z-\omega)/(r)^{\frac{1}{2}})\psi + [\omega(\psi - 1)/(r)^{\frac{1}{2}}]^2 + 2\lambda(1/(rN)^{\frac{1}{2}})[\omega(\psi - 1)/(r)^{\frac{1}{2}}]$. Assume that $L_\infty < \infty$, which implies that $\psi_\infty = 1$. Dominated convergence can be used to justify all limit and expectation interchanges. The first term is handled with (A1.9), having limit $(1/2H_\infty)$, the third with (2.4), having limit $(\lambda_\infty^2/H_\infty)$, the fourth term has limit zero using (2.4) and (A1.7), the fifth term by strong convergence of Z_N and (2.4) has the limit of (L_∞^2/H_∞), the sixth by (2.4) has the limit of $(2\lambda_\infty L_\infty/H_\infty)$. To handle the second term, note that by (A1.11), and dominated convergence it will be $\lim\{\psi(\psi-1)[R_N R_{0,N} + (1 - R_{0,N})] + \psi(\omega/(r)^{\frac{1}{2}})(\psi - 1)(2R_N R_{0,N}/H_\infty)^{\frac{1}{2}}E(Y_1)\}$. Since the R's have finite limits, $(\psi - 1)$ has a zero limit, $(\omega/(r)^{\frac{1}{2}})(\psi - 1)$ has a finite limit, and $E(Y_1) = 0$, this term disappears. Noting that (2.15) is infinite if L_∞ is not finite completes the proof.

To get (2.16), use formulas (A.1) and (A.5) of [3] to characterize $H(\omega)$ as an integral, then take limits in the integral.

APPENDIX 2

Limiting moments. In Appendix 1 the quantities R_N and $R_{0,N}$ of (A1.3) had a.s. limits as $r \to 0$. If however ω is not fixed but is a function of r (as in Theorem 3) then R_N and $R_{0,N}$ have a limiting distribution. Using the notation of (A1.5), let

$$\tilde{\rho}^2 = H(\tilde{\rho} Z_{n_0}(2/r)^{\frac{1}{2}}) = H(\tilde{\rho}|(1/(rn_0)^{\frac{1}{2}})Y_1 + ((2/r)^{\frac{1}{2}})\omega|) \tag{A2.1}$$

$$N = \lceil \tilde{\rho}^2/r \rceil ; \qquad \beta = \omega(2H_{\hat{\theta}^*, \min}/r)^{\frac{1}{2}}) , \tag{A2.2}$$

$$\hat{\rho}^2(x) = H(x\hat{\rho}(x)/(H_{\min})^{\frac{1}{2}}) , \tag{A2.3}$$

where we omit the $\hat{\theta}^*$ subscript unless it is needed to avoid ambiguity. Using (1.6) and expanding (A2.1) in a Taylor series, if $H(\cdot)$ has a bounded derivative, $\tilde{\rho}^2 \to \hat{\rho}^2$ a.s., which with (A2.2) gives

$$\lim_{r\to 0} rN = \hat{\rho}^2(|Y_1 + \beta|) \quad \text{a.s.}, \tag{A2.4}$$

$$\lim_{r\to 0} E(rN)^j = E[\hat{\rho}^2(|Y_1 + \beta|)]^j , \qquad \text{any} \quad j \cdot \tag{A2.5}$$

Next, the distribution and moments of $\hat{\rho}^2$ will be given.

LEMMA *A. Let A_t be the set $((H_{\min}/t)^{\frac{1}{2}})H^{-1}(t)$. Let $A_T = \bigcup_{t \leq T} A_t$. Let $T \leqq H_{\max}$.*

$$P\{\hat{\rho}^2(|Y_1 + \beta|) \leqq T\} = \int_{A_T} (\phi(u + \beta) + \phi(u - \beta))\, du , \tag{A2.6}$$

$$\begin{aligned} &E(\hat{\rho}^2(|Y_1 + \beta|))^j \\ &\quad = (H_{\min})^{\frac{1}{2}} \int_0^\infty H^{j-\frac{3}{2}}(v)[H(v) + vH'(v)][\phi((v(H_{\min}/H(v))^{\frac{1}{2}}) + \beta) \\ &\qquad + \phi((v(H_{\min}/H(v))^{\frac{1}{2}}) - \beta)]\, dv . \end{aligned} \tag{A2.7}$$

PROOF. Let u_t be a solution of $H(u) = t$, i.e., $u_t \in H^{-1}(t)$. If $x\hat{\rho}(x)/(H_{\min})^{\frac{1}{2}} = u_t$, then $\hat{\rho}^2(x) = t$, so that $x = u_t(H_{\min}/t)^{\frac{1}{2}}$. Clearly, if $\hat{\rho}^2(x) = t$, then $x\hat{\rho}(x)/(H_{\min})^{\frac{1}{2}} = u_t$ for some $u_t \in H^{-1}(t)$, and $\{x : \hat{\rho}^2(x) \leqq T\} = A_T$. Using the density of $|Y_1 + \beta|$ where Y_1 is standard normal gives (A2.6). Next, (A2.7) follows from making the change of variables $v = u\hat{\rho}(u)/(H_{\min})^{\frac{1}{2}}$ in

$$E(\hat{\rho}^2(|Y_1 + \beta|))^j = \int_0^\infty (\hat{\rho}(u))^{2j}[\phi(u + \beta) + \phi(u - \beta)]\, du ,$$

which follows from (A2.3).

Finally, to get the second moment of $\theta_N^{*\prime}$ using (A1.1), we need

LEMMA *B. For estimators of the form* (A1.6),

$$\begin{aligned} &\lim_{r\to 0} (2H_{\min}/r)^{j/2} E[Z_N \psi((N)^{\frac{1}{2}} Z_N) + (1/(N)^{\frac{1}{2}})\lambda((N)^{\frac{1}{2}} Z_N) - \omega]^j \\ &\quad = E\{|Z(Y_1, Y_2, \beta)|\psi[|Z(Y_1, Y_2, \beta)|/(2)^{\frac{1}{2}}\gamma(|Y_1 + \beta|)] \\ &\qquad + (2)^{\frac{1}{2}}\gamma(|Y_1 + \beta|)\lambda[|Z(Y_1, Y_2\beta)|/(2)^{\frac{1}{2}}\gamma(|Y_1 + \beta|)] - \beta\}^j , \end{aligned} \tag{A2.8}$$

where the expectation is w.r.t. Y_1, Y_2 (*see* (A1.4)) *and*

$$\begin{aligned} Z(Y_1, Y_2, \beta) &= \gamma^2(|Y_1 + \beta|)Y_1 \\ &\quad + (\gamma^2(|Y_1 + \beta|)(1 - \gamma^2(|Y_1 + \beta|)))^{\frac{1}{2}} Y_2 + \beta , \end{aligned} \tag{A2.9}$$

$$\gamma^2(|Y_1 + \beta|) = H_{\min}/\hat{\rho}^2(|Y_1 + \beta|) . \tag{A2.10}$$

The proof uses (A1.3) and (A1.4) with (A2.4), (1.6) and (A2.2), and dominated convergence.

In the case of X^* more explicit forms for the expressions in Lemmas A and B can be given. These are summarized below:

$$\lim_{r\to0}(1/r)^{\frac{1}{2}}E(X_N^* - \theta^*) = ((2)^{\frac{1}{2}}/H_{\min}^{\frac{3}{2}})[I_4 - I_3]\,, \tag{A2.11}$$

$$\lim_{r\to0}(1/r)E(Z_{0N} - \eta)^2 = (1/2H_{\min}^2)I_2\,, \tag{A2.12}$$

$$\lim_{r\to0}(1/r)E(Z_N - \omega)^2 = (1/2H_{\min}^2)[4I_1 + I_2 + 4\beta I_3]\,, \tag{A2.13}$$

$$\lim_{r\to0}(1/r)E(X_N^* - \theta^*)^2 = (1/H_{\min}^2)[2I_1 + I_2 + 2\beta I_3]\,, \tag{A2.14}$$

$$\lim_{r\to0} rE(N) = I_5\,, \tag{A2.15}$$

where

$$I_1 = \int_0^\infty [\phi(v) - 2v\Phi(-v)]\gamma^6(v)[C(v)\phi(C(v)) + D(v)\phi(D(v))]\,dv\,, \tag{A2.16}$$

$$I_2 = \int_0^\infty \gamma^5(v)[1 - v\phi(v)][\phi(C(v)) + \phi(D(v))]\,dv\,, \tag{A2.17}$$

$$\begin{aligned} I_3 = \int_0^\infty \{&(\gamma^2(v)C(v) + \beta)\Phi(-(\gamma^2(v)C(v) + \beta)/B(v))\phi(C(v)) \\ &- (\gamma^2(v)D(v) - \beta)\Phi((\gamma^2(v)D(v) - \beta)/B(v))\Phi(D(v)) \\ &- 2B(v)\phi(v\gamma^2(v)/B(v))\phi(\beta/\gamma(v))\}\gamma^3(v)[1 - v\phi(v)]\,dv\,, \end{aligned} \tag{A2.18}$$

$$I_4 = \int_0^\infty [\phi(v) - 2v\Phi(-v)]\gamma^4(v)[\phi(C(v)) - \phi(D(v))]\,dv\,, \tag{A2.19}$$

$$I_5 = \int_0^\infty \gamma(v)[1 - v\phi(v)][\phi(C(v)) + \phi(D(v))]\,dv\,, \tag{A2.20}$$

$$\begin{aligned} &B^2(v) = \gamma^2(v)(1 - \gamma^2(v))\,; \qquad C(v) = v\gamma(v) - \beta\,; \\ &D(v) = v\gamma(v) + \beta\,; \qquad \gamma(v) = +(H_{\min}/(1 + 2F(v)))^{\frac{1}{2}}\,; \\ &F(v) = v^2\Phi(-v) - v\phi(v)\,. \end{aligned} \tag{A2.21}$$

REFERENCES

[1] BLUMENTHAL, S. (1975). Sequential estimation of the largest normal mean when the variance is unknown. *Comm. Statist.* **4** 655–669.

[2] BLUMENTHAL, S. and COHEN, A. (1968). Estimation of the larger translation parameter. *Ann. Math. Statist.* **39** 502–516.

[3] BLUMENTHAL, S. and COHEN, A. (1968). Estimation of the larger of two normal means. *J. Amer. Statist. Assoc.* **63** 861–876.

[4] DUDEWICZ, E. (1969). Estimation of ordered parameters. Technical Report 60, Dept. of Operations Research, Cornell Univ.

[5] DUDEWICZ, E. (1971). Maximum likelihood estimators for ranked means. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **19** 29–42.

[6] DUDEWICZ, E. (1972). Point estimation of ordered parameters: the general location parameter case. *Tamkang J. Math.* **3** 101–114.

[7] DUDEWICZ, E. (1973). Maximum probability estimators for ranked means. *Ann. Inst. Statist. Math.* **25** 467–477.

DEPARTMENT OF STATISTICS
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40506