UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

SOME SEQUENTIAL ANALOGS OF STEIN'S TWO-STAGE TEST

by

William Jackson Hall

September, 1961

This research was supported by the Office of Naval Research under Contract No. Nonr-855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 308

SOME SEQUENTIAL ANALOGS OF STEIN'S TWO-STAGE TEST

by

William Jackson Hall[1]

University of North Carolina

This paper presents several sequential analogs of Stein's two-stage test procedure for testing hypotheses about the mean of a normal population with unknown variance and with specified error probabilities. When sequential experimentation is feasible, they provide alternatives to the sequential normal test (variance known) or the sequential t-test. If the variance is assumed known, the procedures may still be recommended since the added cost may be only a very few additional observations on the average, and the performance of the tests does not depend on the validity of any assumption about the variance. Moreover, unlike the t-test, these procedures do not require that the alternative hypothesis be specified in standard deviation units.

1. INTRODUCTION

When sampling from a normal population with unknown mean μ and unknown variance σ^2, one may wish to test the composite hypotheses

$$H_0: \mu \leq 0, \sigma > 0 \qquad \text{vs.} \qquad H_1: \mu \geq \Delta \ (>0), \sigma > 0$$

with pre-assigned strength (α, β) (bounds on the error probabilities). It is a well-known fact that, unless at least bounds are placed on σ, no such non-sequential test exists. A common solution is to restate H_1 in (unknown) standard deviation units and use the t-test (non-sequen-

[1]This research was supported by the Office of Naval Research under contract No. Nonr-855(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.

tial or sequential), or, equivalently, allow β to be a function of the unknown σ. Neither of these reformulations may be completely satisfactory. The only known solution to the problem as stated is Stein's two-stage procedure (Stein, 1945, Moshman, 1958): a preliminary sample of fixed size m (> 1) is taken in order to estimate σ^2 and a second stage sample, of size depending on this first-stage estimate s_m^2, is then taken if necessary; since the first-stage sample mean and variance are statistically independent, the information from the first sample about the mean μ can be utilized, together with that from the second sample, in making the terminal decision. The size of the second sample depends only on s_m^2, so that its distribution depends only on σ^2, and not on μ.

A sequential analog of Stein's procedure is presented here. Again a first stage sample is used to estimate σ^2, but sampling is then continued, if at all, one observation at a time rather than in a non-sequential fashion. It is otherwise analogous to Stein's procedure, but, as one would expect, the distribution of the sample size now depends on μ as well as on σ^2.

This procedure, test T, may be described as a sequential probability ratio test (SPRT) which is not permitted to terminate before m observations and in which σ^2 is replaced in the probability ratio by the estimate s_m^2; the usual termination bounds A and B are modified by a method due to Paulson (1961) in order to achieve the required strength. An equivalent interpretation of the test T, useful for studying its properties, is that it is a conditional SPRT, given s_m and σ, with termination boundaries depending on s_m and σ. Its behavior can be studied by averaging (taking expectation) with respect to S_m. Thus approximations to its OC (operating characteristic) function and ASN (average sample number) function are obtained by averaging the corresponding approximations of Wald for the conditional SPRT; these approximations are valid if the test is not likely to terminate with the first stage. Stein's test can also be considered in this light. Some numerical comparisons of the power (or OC) and ASN functions, based on these approximations, for the test T, for Stein's two-stage test, and for the SPRT and fixed sample size test (FSST) assuming σ known, are presented. The approximation obtained for the ASN of T suggests that substantial savings, compared with Stein's procedure, are possible.

For the case $\beta = \alpha$, an alternative sequential test procedure T' is described, using the minimum probability ratio test (Hall, 1961) instead of the SPRT. Two-sided test procedures, analogous to T and T', are also briefly discussed.

If an estimate of σ^2 is available from previous experiments, the need for the first stage is eliminated; minor modifications of these procedures would make them applicable. In fact, this is the context in which Paulson's method was introduced.

In none of these procedures is any of the information about σ, other than from the first-stage sample, utilized, so that the tests do not depend on a sufficient sequence of statistics. Alternative sequential procedures, T_n and T'_n, using all available information about σ but without theoretical justification, are proposed. Some empirical evaluation of these procedures is planned.

Illustrative diagrams for carrying out these sequential tests are presented.

2. THE SEQUENTIAL TEST T.

Let $X_1, X_2, \ldots$ be independent $N(\mu, \sigma^2)$ random variables, $-\infty < \mu < \infty$, $0 < \sigma < \infty$. Let m be a specified integer exceeding one. Consider a SPRT of $\mu = 0$ vs. $\mu = \Delta$ (σ known) based on $\bar{X}_m, X_{m+1}, X_{m+2}, \ldots$, with termination boundaries A_m and B_m and with σ replaced by s_m where

$$\bar{x}_m = \Sigma_{i=1}^m x_i/m \quad , \qquad \nu = m - 1$$

$$s_m^2 = \Sigma_{i=1}^m (x_i - \bar{x}_m)^2/(m-1)$$

$$(1) \quad \begin{aligned} a_m &= \ln A_m = \nu(\alpha^{-2/\nu} - 1)/2 = (-\ln \alpha) [1 + (-\ln \alpha)/\nu + O(\nu^2)] \\ b_m &= \ln B_m = -\nu(\beta^{-2/\nu} - 1)/2 = -(-\ln \beta) [1 + (-\ln \beta)/\nu + O(\nu^2)]. \end{aligned}$$

We refer to this test as test T. Note that $A_m > A \equiv 1/\alpha$ and $B_m < B \equiv \beta$, with approximate equalities instead of inequalities if m is large, and A, B are the conservative termination bounds of Wald (1947, p. 42) appropriate if σ were known.

Denoting

$$(2) \qquad r_n(s_m) = \Delta \; \Sigma_{i=1}^n (x_i - \Delta/2)/s_m^2 \qquad (n \geq m) ,$$

T is found to be: observe $(X_1, \ldots, X_m)$ and then $X_{m+1}, X_{m+2}, \ldots$ successively and, for each $n \geq m$, after observing X_n,

stop sampling and make decision d_0 (accept H_0) if $r_n(s_m) \leq b_m$;

stop sampling and make decision d_1 (accept H_1) if $r_n(s_m) \geq a_m$;

continue sampling if $b_m < r_n(s_m) < a_m$.

3. THE STRENGTH OF TEST T.

For given (s_m, σ), consider the conditional SPRT, $T(s_m, \sigma)$, of $\mu = 0$ vs. $\mu = \Delta$ based on $\bar{X}_m, X_{m+1}, \ldots$ with termination bounds $\bar{A}_n$ and $\bar{B}_n$ where

(3) $\quad \bar{a}_n = \ell n\, \bar{A}_n = a_n\, s_m^2/\sigma^2, \quad \bar{b}_n = \ell n\, \bar{B}_n = b_n\, s_m^2/\sigma^2$.

Computing the relevant probability ratios, noting that $\bar{X}_m, X_{m+1}, \ldots$ are statistically independent of S_m, one finds that decisions are made according as $r_n(\sigma) \leq \bar{b}_m$, $r_n(\sigma) \geq \bar{a}_m$, or $\bar{b}_n < r_n(\sigma) < \bar{a}_n$. Using Wald's conservative bounds on the error probabilities of a SPRT, we have

$$\text{(4)}\quad \begin{aligned} &\Pr\left\{d_1 \text{ using } T(s_m,\sigma)\,|\,s_m, \sigma, \mu=0\right\} < 1/\bar{A}_n = \exp(-a_n s_m^2/\sigma^2) \\ &\Pr\left\{d_0 \text{ using } T(s_m,\sigma)\,|\,s_m, \sigma, \mu=\Delta\right\} < \bar{B}_n = \exp(b_n s_m^2/\sigma^2) \ . \end{aligned}$$

But $r_n(\sigma) = r_n(s_m)\, s_m^2/\sigma^2$, so that $T(s_m, \sigma)$ is seen to have precisely the same decision rule at each stage as does the test T, with s_m computed from the observed values of $X_1, \ldots, X_m$. Thus

$$\mathcal{E} \Pr\left\{d_i \text{ using } T(S_m,\sigma)\,|\,S_m,\sigma,\mu\right\} = \mathcal{E} \Pr\left\{d_i \text{ using } T \mid S_m, \sigma, \mu\right\}$$
$$= \Pr\left\{d_i \text{ using } T \mid \sigma,\mu\right\} \quad ;$$

and therefore, using (4) ,

(5) $\quad \Pr\left\{d_1 \text{ using } T \mid \sigma, 0\right\} < \mathcal{E} \exp(-a_n S_m^2/\sigma^2) = (1 + 2a_n/\nu)^{-\nu/2}$

for all σ since $\nu S_m^2/\sigma^2 = \chi_\nu^2$ and $\mathcal{E} \exp(t\, \chi_\nu^2) = (1 - 2t)^{-\nu/2}$.

Similarly,

(6) $\quad \Pr\left\{d_0 \text{ using } T \mid \sigma, \Delta\right\} < \mathcal{E} \exp(b_n S_m^2/\sigma^2) = (1 - 2b_n/\nu)^{-\nu/2}$

for all σ. Since the conditional test $T(S_m, \sigma)$ is a normal-mean SPRT, its OC function is monotone in μ (Wald, 1947), i.e., $\Pr\{d_1 \text{ using } T(s_m, \sigma) \mid s_m, \sigma, \mu\}$ is monotone in μ for every fixed s_m and σ. This, together with (5), (6) and (1), implies

$$(7) \quad \Pr\left\{d_1 \text{ using } T \mid H_0\right\} < \alpha, \quad \Pr\left\{d_0 \text{ using } T \mid H_1\right\} < \beta ,$$

and (7) states that T has strength (α, β).

Note also that since the SPRT $T(s_m, \sigma)$ terminates with certainty for every fixed S_m, the test T also terminates with certainty.

4. THE OC FUNCTION OF T.

For the conditional test $T(s_m, \sigma)$, Wald's approximation to the OC function may be used, namely:

$$(8) \quad \Pr\left\{d_0 \mid s_m, \sigma, \mu\right\} \overset{e}{=} (\bar{A}_m^h - 1)/(\bar{A}_m^h - B_m^h) \qquad (h \neq 0)$$

where $h(\mu) = 1 - 2\mu/\Delta$ and where the "e" implies neglect of excess over the boundaries; this excess should be small if the test is likely to have a sample number substantially larger than m. Taking expectations with respect to S_m in (8), using (3) and dropping the subscripts on a_m, b_m and S_m, we have

$$\Pr\left\{d_0 \mid \sigma, \mu\right\} \overset{e}{=} \mathcal{E}\, \frac{1 - \exp(-ahS^2/\sigma^2)}{1 - \exp(-\overline{a-b}hS^2/\sigma^2)}$$

$$= \mathcal{E}\left\{\left[1 - \exp(-ahS^2/\sigma^2)\right]\Sigma_{i=0}^{\infty}\exp(-i\overline{a-b}hS^2/\sigma^2)\right\}$$

$$(h > 0)$$

which does not depend on σ. The integrand on the RHS may be expressed as an alternating series with terms of decreasing magnitude so that successive partial sums give upper and lower bounds on it. Taking expectations term-by-term, the RHS thus yields successive upper and lower bounds on the approximation to $\Pr\left\{d_0 \mid \sigma, \mu\right\}$, which may be expressed as

$$1 - (1 + 2ah/\nu)^{-\nu/2} + (1 + 2\,\overline{a-b}\,h/\nu)^{-\nu/2}$$

$$- (1 + 2\,\overline{2a-b}\,h/\nu)^{-\nu/2} + (1 + 4\,\overline{a-b}\,h/\nu)^{-\nu/2} - \ldots$$

or, using (1), as

$$(9)\quad 1 - (1 - h + h\,\alpha^{-2/\nu})^{-\nu/2} + (1 - 2h + h\,\alpha^{-2/\nu} + h\,\beta^{-2/\nu})^{-\nu/2}$$

$$- (1 - 3h + 2h\,\alpha^{-2/\nu} + h\,\beta^{-2/\nu})^{-\nu/2} + (1 - 4h + 2h\alpha^{-2/\nu} + 2h\beta^{-2/\nu})^{-\nu/2}$$

$$- \ldots .$$

These are valid for $h > 0$, i.e., for $\mu < \Delta/2$. For $\mu > \Delta/2$ ($h < 0$), we obtain analogously

$$(10)\quad \Pr\left\{d_0 \mid \sigma, \mu\right\} \doteq 1 - \mathcal{E}\,\frac{1 - \exp(-bhS^2/\sigma^2)}{1 - \exp(-\overline{b-a}\,hS^2/\sigma^2)}$$

$$= (1 + 2bh/\nu)^{-\nu/2} - (1 + 2\,\overline{b-a}\,h/\nu)^{-\nu/2}$$

$$+ (1 + 2\,2b\text{-}a\,h/\nu)^{-2/\nu} - (1 + 4\,\overline{b-a}\,h/\nu)^{-\nu/2} + \ldots$$

$$= (1 + h - h\beta^{-2/\nu})^{-\nu/2} - (1 + 2h - h\beta^{-2/\nu}$$

$$- h\,\alpha^{-2/\nu})^{-\nu/2} + (1 + 3h - 2h\,\beta^{-2/\nu} - h\alpha^{-2/\nu})^{-\nu/2}$$

$$- (1 + 4h - 2h\beta^{-2/\nu} - 2h\alpha^{-2/\nu})^{-\nu/2} + \ldots .$$

For $\beta = \alpha$, we have

$$(11) \quad \Pr\left\{d_0 \middle| \sigma, \mu\right\} \; \underset{=}{e} \; \Sigma_{i=0}^{\infty} (-1)^i \left[1 + i\,|h|(\alpha^{-2/\nu} - 1)\right]^{-\nu/2} \quad (\mu \neq \Delta/2)$$

where again, as in (9) and (10), the partial sums provide successive upper and lower bounds.

At $\mu = \Delta/2$ ($h = 0$), we obtain by taking limits in (8) as $h \longrightarrow 0$ and using l'Hospital's rule:

$$\Pr\left\{d_0 \mid s, \sigma, \Delta/2\right\} \; \underset{=}{e} \; a/(a-b)$$

irrespective of s, so that

$$(12) \quad \Pr\left\{d_0 \mid \sigma, \Delta/2\right\} \; \underset{=}{e} \; a/(a-b) \; = \; (\alpha^{-2/\nu} - 1)/(\alpha^{-2/\nu} + \beta^{-2/\nu} - 2)$$

which equals $1/2$ if $\beta = \alpha$.

The series (9)-(11) converge reasonably fast except near $h = 0$ ($\mu = \Delta/2$). For example, with $\alpha = \beta = .05$ and $n = 16$ and 31, we find the following values for the successive approximations in (11) to the power function $\Pr\left\{d_1 \middle| \mu\right\} = 1 - \Pr\left\{d_0 \middle| \mu\right\}$:

μ/Δ	h	n							
1/4	1/2	16	.19,	.14,	.16,	.15,	.16,	.15,	.15
0	1	16	.050,	.044,	.045,	.045			
-1/4	3/2	16	.016,	.015,	.015				
-1/2	2	16	.0059,	.0056,	.0057,	.0057			
1/4	1/2	31	.21,	.16,	.17,	.17			
0	1	31	.050,	.046,	.046				
-1/4	3/2	31	.014,	.013,	.013				
-1/2	2	31	.0041,	.0040,	.0040				

The same speed of convergence was found for $\alpha = \beta = .01$. Calculation beyond two significant figures is actually unwarranted since these formulas ignore the excess.

5. THE ASN FUNCTION OF T.

For $\mu \neq \Delta/2$ $(h \neq 0)$ and for all σ, we have, using (2):

$$(13) \qquad \mathcal{E}\, r_N(\sigma) = \Delta\, \mathcal{E}\, \Sigma_{i=1}^{N} (X_i - \Delta/2)/\sigma^2$$
$$= \Delta\,(\mu - \Delta/2)\, \mathcal{E}\, N/\sigma^2 = -\frac{h}{2}\, \frac{\Delta^2}{\sigma^2}\, \mathcal{E}\, N .$$

Also, dropping the subscript n,

$$(14) \qquad \mathcal{E}\, r_N(\sigma) = \mathcal{E}\,\mathcal{E}\,[r_N(\sigma) \mid S] = \mathcal{E}\,(\Sigma_{i=0}^{1}\, \mathcal{E}\,[r_N(\sigma) \mid S, d_i]\; \Pr\left\{d_i \mid S\right\}) .$$

Now, still dropping the n's, and ignoring excess,

$$\mathcal{E}\,[r_n(\sigma) \mid s,\ d_1] \overset{e}{=} \bar{a} = a\ s^2/\sigma^2 ;$$

$$\mathcal{E}\,[r_n(\sigma) \mid s\ ,\ d_0] \overset{e}{=} \bar{b} = b\ s^2/\sigma^2 .$$

The excess should not be significant if N tends to be large relative to m. (14) thus leads to

$$(15) \qquad \mathcal{E}\, r_N(\sigma) \overset{e}{=} \mathcal{E}\,[b\ S^2 \Pr\left\{d_0 \mid S\right\} + a\ S^2\ \Pr\left\{d_1 \mid S\right\}]/\sigma^2$$
$$= a - (a-b)\ \mathcal{E}\,[S^2\ \Pr\left\{d_0 \mid S\right\}]/\sigma^2 .$$

Using (8) and proceeding as in the previous section, we obtain for $h > 0$

$$\mathcal{E}\, r_N(\sigma) \overset{e}{=} a - (a-b)\ \mathcal{E}\left\{\frac{S^2}{\sigma^2}\,[1-\exp(-ahS^2/\sigma^2)]\ \Sigma_{i=0}^{\infty}\ \exp(-i\ \overline{a-b}\ hS^2/\sigma^2)\right\}.$$

Noting that $\mathcal{E}[X^2 \exp(-t\, X^2)/\nu] = (1+2t)^{-1-\nu/2}$, we find after taking expectations term-by-term, and equating with (13), that for $h > 0$

$$(16)\quad (\Delta/\sigma)^2\, \mathcal{E}\, N \overset{e}{=} -\frac{2b}{h} - \frac{2(a-b)}{h}\Big[(1+2ah/\nu)^{-1-\nu/2} - (1+2\,\overline{a-b}\,h/\nu)^{-1-\nu/2}$$

$$+ (1+2\,\overline{2a-b}\,h/\nu)^{-1-\nu/2} - (1+4\,\overline{a-b}\,h/\nu)^{-1-\nu/2}$$

$$+ \ldots\Big]$$

$$= \frac{\nu}{h}\Big\{\beta^{-2/\nu} - 1 - (\alpha^{-2/\nu} + \beta^{-2/\nu} - 2)\Big[(1 - h +$$

$$h\,\alpha^{-2/\nu})^{-1-\nu/2} - (1 - 2h + h\,\alpha^{-2/\nu} + h\,\beta^{-2/\nu})^{-1-\nu/2}$$

$$+ (1 - 3h + 2h\,\alpha^{-2/\nu} + h\,\beta^{-2/\nu})^{-1-\nu/2}$$

$$- (1 - 4h + 2h\,\alpha^{-2/\nu} + 2h\,\beta^{-2/\nu})^{-1-\nu/2} + \ldots\Big]\Big\}.$$

For $h < 0$, the same formula holds with h replaced by $-h$ and with α and β interchanged. For $\beta = \alpha$,

$$(17)\quad (\Delta/\sigma)^2\, \mathcal{E}\, N \overset{e}{=} \frac{\nu}{|h|}(\alpha^{-2/\nu} - 1)\Big\{1 + 2\,\Sigma_{i=1}^{\infty}(-1)^i\Big[1 +$$

$$i\;h\,(\alpha^{-2/\nu} - 1)\Big]^{-1-\nu/2}\Big\} \quad (h \neq 0).$$

At $\mu = 0$ and $\mu = \Delta$, Wald's conservative bounds (4) on the error probabilities of the conditional test may be used instead of (8); thus (15) leads to

$$(\Delta/\sigma)^2\, \mathcal{E}[N \mid \mu = 0] \overset{e}{>} -b - (a-b)\,\mathcal{E}\, S^2 \exp(-aS^2/\sigma^2)/\sigma^2$$

$$= \nu\left[\beta^{-2/\nu} - 1 - \alpha(1 + \beta^{-2/\nu}\,\alpha^{2/\nu})\right]$$

and similarly for $(\Delta/\sigma)^2 \mathcal{E}[N \mid \mu = \Delta]$ with α and β interchanged.

If $\beta = \alpha$, these two relations reduce to

(18) $(\Delta/\sigma)^2 \mathcal{E} [N|\mu=0 \text{ or } \Delta] \overset{e}{>} \nu (\alpha^{-2/\nu} - 1 - 2\alpha)$.

For $h = 0$ $(\mu = \Delta/2)$, we have, analogous to (13)-(15),

(19) $\mathcal{E} [r_N(\sigma)]^2 = (\Delta/\sigma)^2 \mathcal{E} [N \mid \mu = \Delta/2]$

and

(20) $\mathcal{E} [r_N(\sigma)]^2 \overset{e}{=} a^2 \mathcal{E} s^4/\sigma^4 - (a^2 - b^2) \mathcal{E} [s^4 \Pr \{d_0 | s\}]/\sigma^4$.

Equating (19) and (20) and using (12),

(21) $(\Delta/\sigma)^2 \mathcal{E} [N \mid \mu = \Delta/2] \overset{e}{=} [a^2 - a(a+b)] \mathcal{E} \chi_\nu^4/\nu^2 = -ab(1 + \frac{2}{\nu})$

$$= \frac{\nu^2}{4} (1 + \frac{2}{\nu})(\alpha^{-2/\nu} - 1)(\beta^{-2/\nu} - 1).$$

Some successive approximations from (17) appear below $(\alpha = \beta = .05)$; evaluations from (18) and (21) also appear:

μ/Δ	h	n	
1/2	0	16	15.4 from (21)
1/4	1/2	16	10.2, 11.2, 10.9, 11.0, 10.9, 11.0, 11.0
0	1	16	6.9, 6.9; lower bound from (18) is 5.9
-1/4	3/2	16	4.8, 4.8
-1/2	2	16	3.7, 3.7
1/2	0	31	11.7 from (21)
1/4	1/2	31	8.3, 9.4, 9.1, 9.2, 9.2
0	1	31	6.1, 6.1; lower bound from (18) is 3.6
-1/4	3/2	31	4.3, 4.3
-1/2	2	31	3.3, 3.3

Convergence was slightly faster for $\alpha = \beta = .01$, and the lower bound from (18) was much closer to the approximation (17).

6. COMPARISON WITH THE OC AND ASN FUNCTIONS OF STEIN'S TWO-STAGE TEST

Let $t_{\nu,\alpha}$ be that number which is exceeded by a t-statistic with probability α (ν degrees of freedom). The total sample size N in Stein's (one-sided) procedure with initial sample size $m = \nu+1$ and error bounds (α, β) is given by

$$N = \max ([s_m^2 (t_{\nu,\alpha} + t_{\nu,\beta})^2 / \Delta^2] + 1, m) ,$$

where "[ ]" means "largest integer in", and

$$(22) \quad (\Delta/\sigma)^2 \mathcal{E} N \doteq (t_{\nu,\alpha} + t_{\nu,\beta})^2 ,$$

the approximation being valid if it implies $\mathcal{E} N$ is somewhat larger than m (Stein, 1945).

The terminal decision rule for Stein's test may be written

$$\Delta \Sigma_{i=1}^{N} (x_i - \Delta/2)/ s_m^2 \begin{matrix} > \\ < \end{matrix} (t_{\nu,\alpha}^2 - t_{\nu,\beta}^2)/2 \quad \begin{matrix} \text{decide } d_1 \\ \text{decide } d_0 \end{matrix} .$$

Approximating N by $s_m^2 (t_\alpha^2 + t_\beta^2)/\Delta^2$, the OC function is found to be

$$(23) \quad \Pr\left\{d_0 | \mu\right\} \doteq F_\nu [t_{\nu,\alpha} - \mu(t_{\nu,\alpha} + t_{\nu,\beta}) / \Delta]$$

where F_ν is the distribution function of a (central) t-statistic with ν degrees of freedom. The true OC function is presumably slightly steeper since the true sample size tends to be larger than the approximating value.

Tables 1 and 2 at the end of this paper present some comparisons of the approximate power and ASN functions of the sequential test T and Stein's two-stage test ($\alpha = \beta = .05$ and $.01$). Also included in the tables are the corresponding approximate values for the SPRT (with $A = \overline{1-\alpha}/\alpha = 1/B$) and the fixed sample size test (FSST) if σ were known (correctly). Of course, if the assumption about σ were incorrect, the power functions of the SPRT and FSST may be drastically altered.

It is of interest to note that the power functions of these tests become steeper as one moves from left to right in Table 1; thus, the test T discriminates best for intermediate μ-values and the FSST discriminates best for extreme μ-values.

These calculations suggest that substantial savings may be possible using the sequential test T -- at least if one of the hypotheses is correct. The comparison between T and Stein's test is analogous to the comparison between the SPRT and the FSST of the same strength.

Actually, the comparison would be in closer analogy if the SPRT with conservative boundaries ($A = 1/\alpha = 1/B$) were considered, since the test T uses conservative boundaries. If the boundaries of T were modified, by increasing α and β in (1), to achieve error probabilities equal to α, the ASN of T would be further reduced. (Calculations indicate that substitution of $\alpha/\overline{1-\beta}$ for α and $\beta/\overline{1-\alpha}$ for β in (1) still gives conservative bounds on the error probabilities.) The lack of knowledge about σ costs only a very few observations (perhaps two or three) on the average, and thus the test T or Stein's test may be recommended even if σ is thought to be known (if a one-stage test is not essential).

7. AN ALTERNATIVE SEQUENTIAL TEST T'

For the symmetric ($\alpha = \beta$) one-sided case, a minimum probability ratio test (MPRT), which has converging straight-line boundaries (Hall, 1961), can be adapted in the same manner as was the SPRT above. The MPRT is equivalent to one of Anderson's (1960) tests. We thus obtain the following test T' with decision rule:

stop sampling as soon as $\Delta \left| \Sigma_{i=1}^{n} (x_i - \Delta/2) \right| / s_m^2 \geq c_m - n\Delta^2/4s_m^2 \quad (n \geq m)$

and choose d_1 or d_0 according as $\Sigma(x_i - \Delta/2)$ is > or < 0 where

$$(24) \qquad c_m = \nu[(2\alpha)^{-2/\nu} - 1] = -2 \ln 2\alpha [1 + (-2 \ln 2\alpha)/\nu + O(\nu^2)] .$$

After m observations have been taken and s_m computed, an upper bound on the total sample size is $4c_m s_m^2/\Delta^2$.

No approximations to the OC or ASN functions have been obtained. Presumably T' compared with T would have a smaller ASN in the neighborhood of $\mu = \Delta/2$ at the cost of a slightly larger ASN at (and beyond) $\mu = 0$ and Δ. The comparison would be analogous to the comparisons of the SPRT and MPRT (σ known) given by Anderson (1960).

If $\alpha \neq \beta$, the test can still be used with 2α in (24) replaced by $\alpha + \beta$, but then one can only assert that the sum of the two error probabilities is less than $\alpha + \beta$ (Hall, 1961).

8. THE TWO-SIDED CASE

The two-sided normal test, variance known, based on the weight function method (Wald, 1947) or the invariance method (Hall, 1959), cannot be adapted to the case of unknown variance as was the one-sided test, since σ^2 does not factor out of the relevant probability ratios. However, the Sobel-Wald test procedure (Sobel and Wald, 1949), in which one in effect runs two one-sided tests simultaneously, is easily adapted. To test $\mu = 0$ against $|\mu| \geq \Delta$, one can run two T (or T') tests -- of $\mu = 0$ against $\mu = \Delta$ and $\mu = 0$ against $\mu = -\Delta$ -- simultaneously and continue sampling until both tests have terminated.

9. HEURISTIC TESTS T_n and T'_n

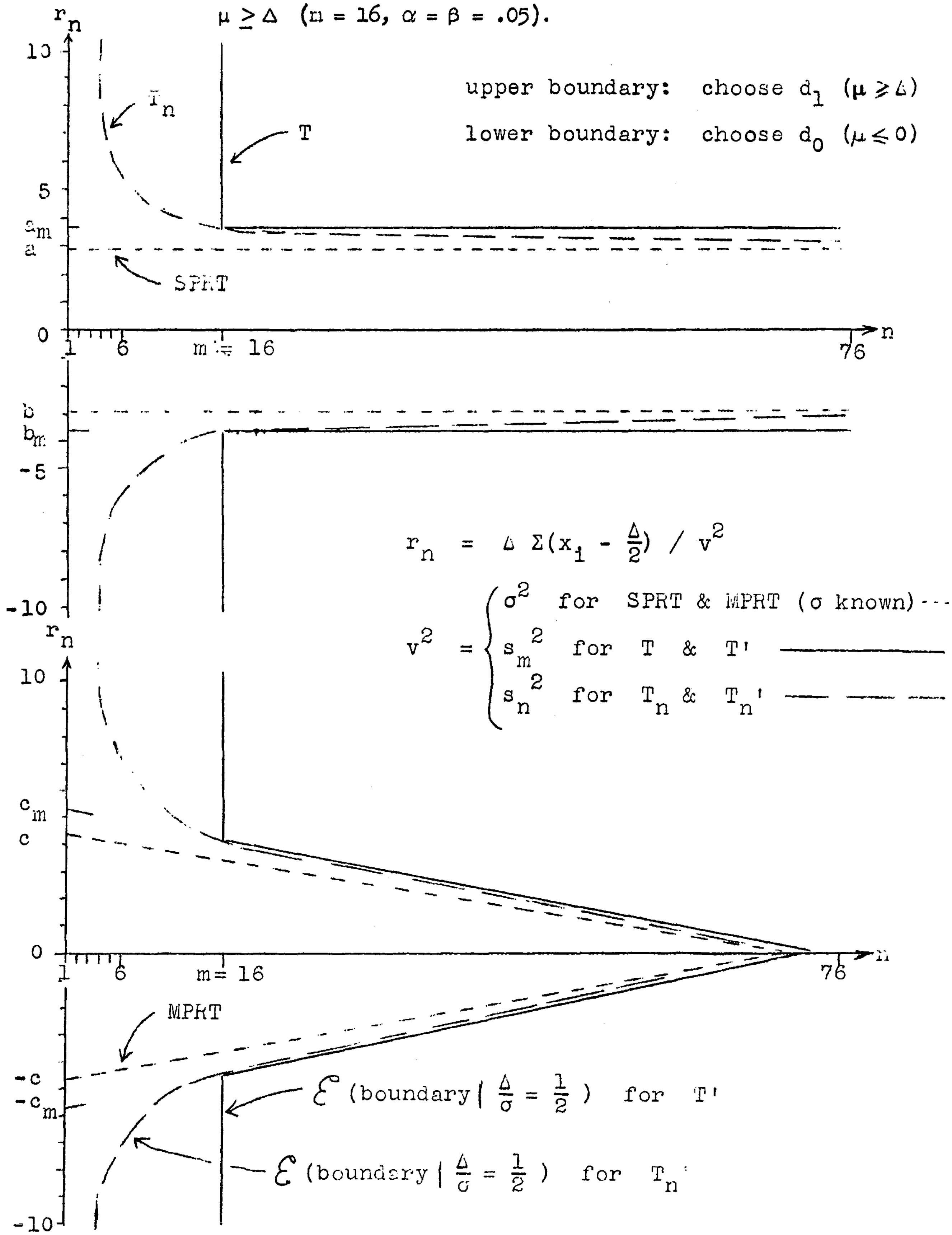
In discussing the sequential estimation of μ (σ unknown), Anscombe (1953) noted that, if the procedure were not allowed to terminate early, σ would essentially be known. Thus, if one uses a procedure requiring knowledge of σ but replaces it by an estimate, the properties of the procedure should not be greatly affected. The tests T and T' are like this; in fact, the test boundaries suitable if σ were known are widened to account for the fact that σ is estimated on $m-1$ degrees of freedom. However, σ is not re-estimated at each successive stage, and the choice of m seems arbitrary; in fact, if σ were much smaller than expected, a completely sequential procedure may terminate before the first stage of T is completed.

The following modification of T is proposed purely on intuitive grounds: re-estimate σ^2 by s_n^2 at each stage n and base the test on $r_n(s_n)$ $(n > 1)$ with boundaries (a_n, b_n) given by (1) with ν replaced by $n-1$. This test, T_n, is like a SPRT (σ known) with σ replaced by a new estimate at each stage, and with the boundaries widened in an attempt to account for the lack of knowledge about σ. The boundaries of T_n converge to Wald's conservative boundaries $(-\ell n\ \alpha, \ell n\ \beta)$ which are appropriate (though slightly conservative) if σ^2 is known. A diagram for carrying out this test is illustrated in Figure 1, together with diagrams for other sequential tests. (The SPRT in the diagram uses Wald's approximate boundaries, $a = \ell n\ (\overline{1 - \beta}/\alpha)$ and $b = \ell n\ (\beta/\overline{1 - \alpha})$.)

The alternative test T' can be modified analogously, obtaining T'_n with the roles of c_n and s_n replaced by c_n and s_n. Its boundaries depend on s_n^2 and thus cannot be graphed in advance (in the diagram, the expected values of the boundaries are graphed).

No theoretical evaluation of these procedures has been possible.

Figure 1. Diagrams for six sequential tests of $\mu \leq 0$ against $\mu \geq \Delta$ ($n = 16$, $\alpha = \beta = .05$).



REFERENCES

ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. Ann. Math. Stat. 31, 165-197.

ANSCOMBE, F. J. (1953). Sequential estimation. J. Roy. Stat. Soc. B, 15, 1-21.

HALL, W. J. (1959). The relationship between sufficiency and invariance with applications in sequential analysis, I. To appear in Ann. Math. Stat. (Institute of Statistics Mimeo Series, No. 228, Chapel Hill.)

HALL, W. J. (1961). Some sequential tests for symmetric problems. Invited address at Eastern Regional Meetings of Inst. Math. Stat., Ithaca, N. Y.

MOSHMAN, J. (1958). A method for selecting the size of the initial sample in Stein's two sample procedure. Ann. Math. Stat. 29, 1271-1275.

PAULSON, E. (1961). A sequential procedure for comparing several experimental categories with a control. Invited address at Eastern Regional Meetings of Inst. Math. Stat., Ithaca, N. Y.

PEARSON, E. S. and HARTLEY, H. O. (1954). Biometrika Tables for Statisticians, 1. Cambridge University Press.

SOBEL, M. and WALD, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. Ann. Math. Stat. 20, 502-522.

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Stat. 16, 243-258.

WALD, A. (1947). Sequential Analysis. New York: John Wiley and Sons.

APPENDIX

Explanation of Tables

Table 1: For the sequential test T, the power function $\Pr\{d_1 | \mu\}$ was calculated from (11) and (12). For the SPRT, it was calculated from Wald's approximation $(1 - e^{-ha})/(e^{ha} - e^{-ha})$ where $a = \ln(\overline{1 - \alpha}/\alpha)$. For the other tests, it was calculated from (23) which reduces to $F_\nu(-ht_{\nu,\alpha})$ where $F_\nu(t_{\nu,\alpha}) = 1 - \alpha$; for the fixed sample size test (FSST), $\nu = \infty$, i.e., F_ν is the standard normal d.f. The Pearson and Hartley (1954) tables of F_ν were used.

Table 2: For the test T, the ASN function was calculated from (17) and (21). For the SPRT, Wald's approximations were used, namely $(\Delta/\sigma)^2 \mathcal{E} N = 2a[1 - 2P(d_1)]/h$ if $h \neq 0$ and $= a^2$ if $h = 0$. For the other tests, it was calculated from (22), and is the same for all μ-values.

The terminal decision rule in every instance is the same, namely choose d_0 $(\mu \leq 0)$ if $\bar{x}_N < \Delta/2$ and choose d_1 $(\mu \geq \Delta)$ if $\bar{x}_N > \Delta/2$.

The approximations to the power function and ASN function ignore

a) excess over the boundaries in the sequential tests,

b) the restriction that N must be integral,

c) the restriction that $N \geq n$ for test T and Stein's test,

and are thus valid if N is large relative to m with high probability.

The author wishes to acknowledge the assistance of Mr. K. Fukushima in preparing these tables and those presented in sections 4 and 5.

TABLE 1

APPROXIMATE POWER FUNCTIONS, $\Pr\left\{d_1 \middle| \mu\right\}$, OF PROCEDURES

FOR TESTING $\mu \leq 0$ AGAINST $\mu \geq \Delta$ $(\alpha = \beta)$

α	μ/Δ	h	SEQUENTIAL TESTS			ONE- OR TWO-STAGE TESTS		
			Test T		SPRT	Stein's Test		FSST
			n = 16	n = 31	(σ known)	n = 16	n = 31	(σ known)
.05	.5	0	.5	.5	.5	.5	.5	.5
	.25	0.5	.154	.168	.187	.197	.201	.205
	.1	0.8	.0727	.0777	.0866	.0906	.0923	.0941
	0	1.0	$.04\bar{5}0$	.0463	.05	.05	.05	.05
	-.1	1.2	$.028\bar{5}$	.0278	.0284	.0264	.0253	.0242
	-.25	1.5	.0149	.0132	.0119	.0095	.0081	.0068
	-.5	2.0	.00565	$.0040\bar{5}$	.00276	.00159	.00097	.00050
.01	.5	0	.5	.5	.5	.5	.5	.5
	.25	0.5	.062	.075	.091	.106	.114	.122
	.1	0.8	.0192	.0216	.0247	.0275	.0293	.0313
	0	1.0	$.009\bar{5}$	.0097	.01	.01	.01	.01
	-.1	1.2	.00496	.00453	.00397	.00349	.00307	.00262
	-.25	1.5	.00206	.00154	.00101	.00071	$.0004\bar{5}$	.00019

TABLE 2

APPROXIMATE ASN FUNCTIONS, $(\Delta/\sigma)^2 \mathcal{E}\left\{N|\mu\right\}$, OF PROCEDURES

FOR TESTING $\mu \leq 0$ AGAINST $\mu \geq \Delta$ $(\alpha = \beta)$

α	μ/Δ	h	SEQUENTIAL TESTS			ONE- OR TWO-STAGE TESTS		
			Test T		SPRT	Stein's Test		FSST
			$m = 16$	$m = 31$	(σ known)	$m = 16$	$m = 31$	(σ known)
.05	.5	0	15.4	11.7	8.7	12.3	11.5	10.8
	.25	0.5	11.0	9.2	7.4			
	.1	0.8	8.2	7.2	6.1			
	0	1.0	6.9	6.1	5.3			
	-.1	1.2	5.9	5.3	4.6			
	-.25	1.5	4.8	4.3	3.8			
	-.5	2.0	3.7	3.3	2.9			
.01	.5	0	45.8	31.0	21.1	27.1	24.1	21.6
	.25	0.5	23.1	18.8	15.0			
	.1	0.8	15.5	13.0	10.9			
	0	1.0	12.6	10.6	9.0			
	-.1	1.2	10.5	8.9	7.6			
	-.25	1.5	$8.\overline{5}$	7.2	6.1			