# A SEQUENTIAL THREE HYPOTHESIS TEST FOR DETERMINING THE MEAN OF A NORMAL POPULATION WITH KNOWN VARIANCE[1]

By Gordon Simons[2]

University of Minnesota

**1. Summary.** This paper examines a sequential testing procedure for choosing one of three simple hypotheses concerning the unknown mean $\mu$ of the normal distribution when the variance is known. The test is conducted by plotting $S_n$, the sum of the observations, versus $n$, the current sample size, until the point $(n, S_n)$ is contained within one of three triangular regions. When this occurs, sampling is terminated and the region containing $(n, S_n)$ determines which state of nature is accepted. Although we shall formally view the problem as one with only three states of nature ($\mu = \mu_1$, $\mu_2$ or $\mu_3$), we shall proceed with the usual understanding that the performance of the test procedure should be evaluated for a wider class of states ($-\infty < \mu < \infty$). The test is approximated by a corresponding exact test for the Wiener process. Formulas are derived which approximate the operating characteristics (OC) and the average sample size (ASN) for all values of $\mu$. The ASN function is compared with theoretical lower bounds. The testing procedure is compared with a modification of a three hypothesis testing procedure proposed by Sobel and Wald [5].

**2. Introduction.** The study of three hypothesis tests is a natural first step in expanding from two hypothesis testing to the general multihypothesis problem. It presents an opportunity to examine in a particular case what might be disguised in the general case. However, three hypothesis tests are of interest in their own right. The familiar two-sided testing problem is often more naturally treated as a three hypothesis testing problem. In the two-sided test, the null hypothesis that, say, $\theta = \theta_0$ is sufficiently informative but its rejection probably is not. One would reasonably like to know whether $\theta < \theta_0$ or $\theta > \theta_0$ when $\theta$ is real valued. Also, the three hypothesis interpretation allows one to control more than just type I and type II errors.

The framework of our problem is as follows. Let $X_1$, $X_2$, $\cdots$ be a sequence of independent, normally distributed random variables with unknown common mean $\mu$ and known common variance $\sigma^2$. We desire to accept one of three simple hypotheses $H_1$, $H_2$, or $H_3$ where $H_i$ is the hypothesis $\mu = \mu_i$, $\mu_1 < \mu_2 < \mu_3$. The generalization of errors of type I and II is expressible in terms of a $3 \times 3$ "error matrix" $A = (\alpha_{ij})$ where $\alpha_{ij} = P_{\mu_i}$ [accepting $H_j$] for $i, j = 1, 2, 3$.

1365

We desire a testing procedure which adapts to any specified error matrix. Experience has indicated that this is not easily accomplished without randomization. The difficulty is that for reasonable test procedures $\alpha_{13}$ and $\alpha_{31}$ tend to be small in comparison to $\alpha_{23}$ and $\alpha_{21}$ respectively. This usually is an asset rather than a liability. In any event, the author's procedure is a non-randomized sequential procedure which allows for almost complete control of the error matrix beyond the stipulation that $\alpha_{13}$ and $\alpha_{31}$ must be very close to zero.

A brief digression will serve to introduce the author's procedure and at the same time place it in proper perspective with previous approaches. P. Armitage [2], with his generalization of Wald's sequential probability ratio test (SPRT) to multihypothesis testing, presents a procedure which can be expressed geometrically in terms of six lines $L_{ij}$, where $L_{ij}$ is the straightline passing through the points

$$(n, \sigma^2(\mu_i - \mu_j)^{-1} \log A_{ij} + \tfrac{1}{2}(\mu_i + \mu_j)^{-1} n)$$

for $n = 0, 1, \cdots$, and $i, j = 1, 2, 3, i \neq j$. The quantities $A_{ij}$ are arbitrary constants exceeding unity. One simply plots $S_n = \sum_1^n X_m$ versus $n$ until one of the following occurs:

   (i) The point $(n, S_n)$ lies below lines $L_{12}$ and $L_{13}$.

   (ii) The point $(n, S_n)$ lies above line $L_{21}$ and below $L_{23}$.

   (iii) The point $(n, S_n)$ lies above lines $L_{31}$ and $L_{32}$.

Accept $H_1$, $H_2$, or $H_3$ respectively.

It is frequently the case that in the right half plane line $L_{12}$ lies entirely below $L_{13}$ and line $L_{32}$ lies entirely above $L_{31}$. Such would be the case if $A_{ij} \equiv A$ for some $A > 1$. Then, (i) and (iii) above reduce to

   (i') The point $(n, S_n)$ lies below line $L_{12}$.

   (iii') The point $(n, S_n)$ lies above line $L_{32}$.

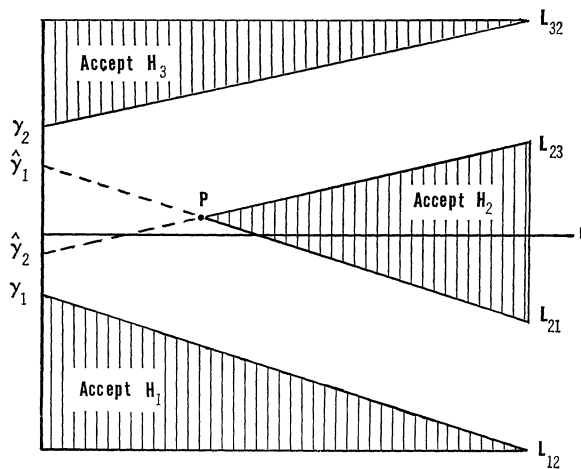The test is shown graphically in Figure I.



FIG. I

$S_n$ versus $n$ is plotted until $(n, S_n)$ is contained within one of the three cross-hatched regions. The hypothesis to be accepted is determined as indicated.

Armitage has shown that his test is closed (for $-\infty < \mu < \infty$) and the elements of the error matrix satisfy

$$(2.1) \qquad \alpha_{ij} \leqq A_{ji}^{-1} \quad \text{for} \quad i \neq j,$$

and

$$(2.2) \qquad \alpha_{ii} \geqq 1 - \sum_{j \neq i} A_{ji}^{-1}.$$

Unfortunately, (2.1) and (2.2) may not be close to equality. I.e., we understate our confidence in choosing the correct hypothesis and understate our protection against the various types of mistakes. Looked at from a different point of view, we probably can satisfy or improve upon the desired error matrix with fewer observations.

We will use the same procedure as the special Armitage procedure shown in Figure I. The essential difference is one of viewpoint. Whereas Armitage's analysis leads to (2.1) and (2.2) we are concerned with obtaining more accurate control of the error matrix. The geometry of Figure I may be summarized by six "geometrical parameters" $\gamma_1$, $\gamma_2$, $\delta_1$, $\delta_2$, $X$, and $T$ where $\gamma_1$ and $\gamma_2$ are the intercepts of $L_{12}$ and $L_{32}$, $\delta_1$ is the slope of $L_{12}$ and $L_{21}$, $\delta_2$ is the slope of $L_{23}$ and $L_{32}$, and $(T, X)$ are the coordinates of the point $P$.

Our primary concern in Section four will be with finding good approximations to the probability of accepting $H_1$, $H_2$, and $H_3$ as functions of $\mu$ and the six geometrical parameters. These probabilities can be approximated by passing from the discrete normal process $S_n \sim N(\mu n, \sigma^2 n)$ for $n = 0, 1, \cdots$ to the continuous Wiener process $X(t) \sim N(\mu t, \sigma^2 t)$, $t \geqq 0$. The corresponding test procedure is to graph $X(t)$ versus $t$ until $(t, X(t))$ is on the boundary of one of the cross-hatched regions in Figure I. (Note that $X(t)$ is almost surely continuous.)

This latter procedure is an example of what the author [3] has called boundary tests. His paper finds implicit methods for computing the OC functions, the average sample time (AST) and other moments of the sampling time. However, we shall derive our results directly from methods developed by T. W. Anderson [1]. It is a pleasant fact that these appoximate OC functions are quite good even for small sample sizes.

As a basis of comparison, it is informative to consider a modification of a three hypothesis, composite test suggested by Sobel and Wald [5], one adapted to our three states of nature problem. We explain this modification by referring again to Figure I:

Plot $(n, S_n)$ for $n = 0, 1, \cdots$ and define $n_1$ and $n_2$ as follows: Let $n_1$ be the smallest positive integer $n$ for which the point $(n, S_n)$ is simultaneously above or below lines $L_{12}$ and $L_{21}$. Define $n_2$ similarly with respect to lines $L_{23}$ and $L_{32}$.

(a) Accept $H_1$ if $n_1 \geqq n_2$ and $(n_1, S_{n_1})$ is below line $L_{12}$.

(b) Accept $H_3$ if $n_2 \geqq n_1$ and $(n_2, S_{n_2})$ is above line $L_{32}$.

(c) Accept $H_2$ otherwise.

For reasons of mathematical convenience, they insist that

(2.3)                          $\gamma_1 \leqq \hat{\gamma}_2 < 0 < \hat{\gamma}_1 \leqq \gamma_2$                          (See Figure I.)

Since the stopping time $N = \max (n_1, n_2)$ depends on both $n_1$ and $n_2$, the stopping rule and acceptance rule do not depend completely on the sufficient statistic $(N, S_N)$. Nevertheless this modified procedure represents a reasonable *ad hoc* procedure which exhibits easy execution and some degree of mathematical tractability.

**3. Results of numerical investigations.** Beyond this point we shall have occasion to refer to a "symmetric case", namely:

(i) $\mu_2 = (\mu_1 + \mu_3)/2 = 0$.   (ii) $\alpha_{ij} = \alpha_{(4-i),(4-j)}$ for $i, j = 1, 2, 3$.

Otherwise, we are in the "general case." It seems appropriate in the symmetric case to use a "symmetric procedure," namely:

(iii) $\gamma_2 = -\gamma_1 = \gamma$.      (iv) $\delta_2 = -\delta_1 = \delta$.      (v) $X = 0$.

In any event, symmetric procedures and "general procedures" must reasonably satisfy the obvious constraints:

(vi) $\gamma_1 < 0 < \gamma_2$.      (vii) $\delta_1 < \delta_2$.

(viii) $T > 0$.      (ix) $\gamma_1 + \delta_1 T < X < \gamma_2 + \delta_2 T$.

A. *Ability to satisfy error matrix requirements.* Usually, one desires the correct decision probabilities $\alpha_{11}$, $\alpha_{22}$ and $\alpha_{33}$ to be substantially larger than .5. In such situations we can only find procedures of the author's type (of the Sobel-Wald type also) which make $\alpha_{13}$ and $\alpha_{31}$ nearly zero. We indicate this briefly by $\alpha_{13} = 0^+$ and $\alpha_{31} = 0^+$. Then, of course, $\alpha_{12} = (1 - \alpha_{11})^-$ and $\alpha_{32} = (1 - \alpha_{33})^-$. It becomes apparent that in the general case we have at most four "degrees of freedom" in choosing an error matrix which can be satisfied. In the symmetric case the number is two. Thus, for instance, in the symmetric case we may desire to control $\alpha_{11}$ and $\alpha_{22}$. Indirectly we are controlling $\alpha_{21}$, $\alpha_{23}$, and $\alpha_{33}$ and virtually controlling $\alpha_{12}$, $\alpha_{13}$, $\alpha_{31}$, and $\alpha_{32}$. Numerical investigations have shown that we have great flexibility within our restricted degrees of freedom. For example, if we insist that $\alpha_{11} = \alpha_{33} = .95$ we may find symmetric procedures which allow $\alpha_{22}$ to range from about .5 to very nearly unity. Similar flexibility exists in the general case using general procedures. The wide degree of latitude that one has is exemplified by the extensive set of tables compiled by the author [3]. The modified Sobel-Wald procedure restricts ones choice of error matrix to a considerably greater degree. This is due to restriction (2.3). Nevertheless, there still remains considerable latitude within their class of procedures.

B. *Quasi-optimality.* In the general case we have, using general procedures, six geometrical parameters with which to "fit" four degrees of freedom in the error matrix. Actually one can typically fix $\delta_1$ and $\delta_2$, say, and make a fit with the remaining four geometrical parameters. A natural question to ask is whether there is some best way of fixing $\delta_1$ and $\delta_2$. For a given value of $\mu$, what values of $\delta_1$ and $\delta_2$ will minimize $E_\mu(\tau)$? ($E_\mu(\tau)$ denotes the expected sampling time for the Wiener process with drift parameter $\mu$.) The question is *nearly* answered for $\mu = \mu_1$, $\mu_2$

and $\mu_3$ : Let $\delta_1 = (\mu_1 + \mu_2)/2$ and $\delta_2 = (\mu_2 + \mu_3)/2$. That is, $E_\mu(\tau)$ is nearly simultaneously minimized for $\mu = \mu_1$, $\mu_2$, and $\mu_3$ with the same pair of $\delta$'s. The evidence is only of a numerical nature and does not appear to be a theoretical fact. (See [3].) The explanation of this phenomenon (called quasi-optimality) seems closely related to the fact that in the two states of nature problem, with $\mu = \mu_1$ and $\mu_2$, say, Wald's SPRT, when viewed geometrically, says to use slope $\delta = (\mu_1 + \mu_2)/2$ for the two parallel stopping lines. For fixed errors of types I an II, this slope minimizes $E_\mu(N)$ for $\mu = \mu_1$ and $\mu_2$ simultaneously. (See Wald and Wolfowitz [6].) Quasi-optimality appears to hold for the Sobel-Wald modified procedures also.

C. *Comparing the ASN with a theoretical lower bound.* In [4], the author compares two of his quasi-optimal tests with two different (general) lower bounds for ASN. These examples and several others in [3] demonstrate that the author's test does very well at $\mu = \mu_1$ or $\mu_3$ when compared with any other test satisfying the same constraints on the error matrix. For one of the examples in [4], the author's test looks quite good at $\mu = \mu_2$ also. In fairness, it should be pointed out that the entries in [4] are not precise for the author's test. The computations for the entries are actually based on the Wiener approximation and the correct entries would be somewhat larger. It is the author's judgment that this difference is slight. (See the next subsection on the Monte Carlo study.)

D. *Small sample size results.* T. W. Anderson [1] raises the question of how inaccurate are the OC functions and the ASN function for the discrete normal process if, in fact, they are computed (exactly) for the approximating Wiener process. Presumably the discrepancy should be most pronounced for small expected sample sizes. The following example shows that the situation is not serious.

EXAMPLE. Problem: $\mu_1 = -1$, $\mu_2 = 0$, $\mu_3 = 1$. Intended constraints (using Wiener process): $\alpha_{11} = \alpha_{22} = \alpha_{33} = .98$, $\alpha_{21} = \alpha_{23} = .01$.

The "actual" entries in Table 1 are each based on 10,000 Monte Carlo experiments. Note that the actual tests are conservative in that they increase the probabilities of making the correct decision when one of the hypothesis is true.

As one would predict, the actual values of the ASN are somewhat larger than the predicted values. (See Table 2.) This discrepancy is small when one of the three hypotheses holds. The situation is more serious at $\mu = -.5$, an intermediate value. Since the analysis developed by Sobel and Wald does not yield a precise ASN but only an upper and lower bound, the comparison for their test is less precise.

It seems intuitively reasonable that the increased "cost" in the ASN should be compensated for by improved OC functions, as we have seen. Nevertheless it should be pointed out that a particularly bad *ad hoc* procedure can do just the opposite. For instance, interchange the acceptance rules for $H_1$ and $H_2$.

**4. Computing the OC functions.** The probability of accepting $H_1$ is approximately equal to the probability of the Wiener process $X(t) \sim N(\mu t, \sigma^2 t)$ making

TABLE 1

| $\mu$ | $-1$ | $-.5$ | $0$ |
|---|---|---|---|
| Predicted[3] $P_\mu$ [accepting $H_1$] for author's test | .980 | .443 | .010 |
| Actual $P_\mu$ [accepting $H_1$] for author's test | .989 | .452 | .006 |
| Predicted $P_\mu$ [accepting $H_2$] for author's test | .020 | .557 | .980 |
| Actual $P_\mu$ [accepting $H_2$] for author's test | .011 | .548 | .988 |
| Predicted $P_\mu$ [accepting $H_1$] for Sobel-Wald test | .980 | .460 | .010 |
| Actual $P_\mu$ [accepting $H_1$] for Sobel-Wald test | .988 | .464 | .006 |
| Predicted $P_\mu$ [accepting $H_2$] for Sobel-Wald test | .020 | .540 | .980 |
| Actual $P_\mu$ [accepting $H_2$] for Sobel-Wald test | .012 | .536 | .988 |

TABLE 2

| $\mu$ | $-1 = \mu_1$ | $-.5$ | $0 = \mu_2$ |
|---|---|---|---|
| Predicted ASN for the author's test | 8.8 | 17.1 | 10.8 |
| Actual ASN for author's test | 10.4 | 22.2 | 12.5 |
| Predicted upper bound for ASN for Sobel-Wald Test | 9.2 | 19.7 | 12.0 |
| Predicted lower bound for ASN for Sobel-Wald test | 8.8 | 17.9 | 7.6 |
| Actual ASN for Sobel-Wald test | 10.4 | 23.7 | 13.2 |

contact with $L_{12}$ before any other part of the boundary. Then

$$(4.1) \quad P_\mu \text{ [accepting } H_1] \doteq P_{\mu 1} + \int_{\gamma_1 + \delta_1 T}^{X} P_{\mu 0}(x) Q_\mu(x) n(x \mid \mu T, \sigma^2 T) \, dx,$$

where

$$(4.2) \qquad P_{\mu 1} \equiv P_\mu[X(t) \text{ contacts } L_{12} \text{ before } L_{32} \text{ and before time } T],$$

$$(4.3) \quad P_{\mu 0}(x) \equiv P_\mu[X(t) \text{ contacts neither } L_{12} \text{ nor } L_{32} \text{ before time}$$

$$T \mid X(T) = x],$$

$$(4.4) \quad Q_\mu(x) \equiv P_\mu[X(t) \text{ contacts } L_{12} \text{ before } L_{21} \text{ after time } T \mid X(T) = x],$$

and $n(x \mid a, b)$ is the normal density with mean $a$ and variance $b$.

We shall set $\sigma^2 = 1$. There is really no loss in generality since the case $\sigma^2 \neq 1$ may be recovered by appropriate scaling of the geometrical parameters. (See section 6.)

COMPUTING $P_{\mu 1}$. Using Anderson's [1] Theorem 4.3, it is easy to show that

---

[3] The term "predicted" as used in this example means the value one computes using the Wiener approximation to the corresponding discrete process.

$$P_{\mu 1} = \sum_{r=0}^{\infty} \exp\{-2[(r+1)\gamma_1 - r\gamma_2][(r+1)(\delta_1 - \mu) - r(\delta_2 - \mu)]\}$$
$$\cdot \Phi([-(\delta_1 - \mu)T - 2r\gamma_2 + (2r+1)\gamma_1]T^{-\frac{1}{2}})$$
$$+ \exp\{-2[r^2\gamma_1(\delta_1 - \mu) + r^2\gamma_2(\delta_2 - \mu) - r(r+1)\gamma_1(\delta_2 - \mu)$$
$$- r(r-1)\gamma_2(\delta_1 - \mu)]\}\Phi([+(\delta_1 - \mu)T - 2r\gamma_2 + (2r+1)\gamma_1]T^{-\frac{1}{2}})$$

(4.5)
$$- \exp\{-2[(r+1)^2\gamma_1(\delta_1 - \mu) + (r+1)^2\gamma_2(\delta_2 - \mu)$$
$$- r(r+1)\gamma_1(\delta_2 - \mu) - (r+1)(r+2)\gamma_2(\delta_1 - \mu)]\}$$
$$\cdot \Phi([-(\delta_1 - \mu)T - 2(r+1)\gamma_2 + (2r+1)\gamma_1]T^{-\frac{1}{2}})$$
$$- \exp\{-2[r\gamma_1 - (r+1)\gamma_2][r(\delta_1 - \mu) - (r+1)(\delta_2 - \mu)]\}$$
$$\cdot \Phi([+(\delta_1 - \mu)T - 2(r+1)\gamma_2 + (2r+1)\gamma_1]T^{-\frac{1}{2}}),$$

where

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-u^2/2}\, du.$$

In the case of symmetric procedures, namely, when

(4.6) $\quad\quad \gamma_2 = -\gamma_1 = \gamma, \quad \delta_2 = -\delta_1 = \delta, \quad \text{and} \quad X = 0,$

(4.5) simplifies to

$$P_{\mu 1} = \sum_{r=0}^{\infty} \{h_{(2r+1)}(\gamma\mu, \gamma\delta) \cdot \Phi([(\delta + \mu)T - (4r+1)\gamma]T^{-\frac{1}{2}})$$
$$+ h_{-2r}(\gamma\mu, \gamma\delta) \cdot [1 - \Phi([(\delta + \mu)T + (4r+1)\gamma]T^{-\frac{1}{2}})]$$
$$- h_{2(r+1)}(\gamma\mu, \gamma\delta) \cdot \Phi([(\delta + \mu)T - (4r+3)\gamma]T^{-\frac{1}{2}})$$
$$- h_{-(2r+1)}(\gamma\mu, \gamma\delta) \cdot [1 - \Phi([(\delta + \mu)T + (4r+3)\gamma]T^{-\frac{1}{2}})]\},$$

where $h_s(u, v) \equiv \exp[-2us - 2vs^2]$. Further simplification yields

(4.7) $\quad P_{\mu 1} = \sum_{s=-\infty}^{\infty} (-1)^{s+1} h_s(\gamma\mu, \gamma\delta)\Phi([(\delta + \mu)T - (2s-1)\gamma]T^{-\frac{1}{2}})$
$$+ \sum_{s=-\infty}^{0} (-1)^s h_s(\gamma\mu, \gamma\delta).$$

COMPUTING $P_\mu(x)$. For $\gamma_1 + \delta_1 T \leq x \leq \gamma_2 + \delta_2 T$ (the only case we need), we have

$$P_{\mu 0}(x) = \sum_{r=-\infty}^{\infty} \{\exp\{-(2/T)[r^2\{\gamma_1(\gamma_1 + \delta_1 T - x) + \gamma_2(\gamma_2 + \delta_2 T - x)\}$$

(4.8)
$$- r(r-1)\gamma_1(\gamma_2 + \delta_2 T - x) - r(r+1)\gamma_2(\gamma_1 + \delta_1 T - x)\}$$
$$- \exp\{-(2/T)[r\gamma_1 - (r-1)\gamma_2][r(\gamma_1 + \delta_1 T - x)$$
$$- (r-1)(\gamma_2 + \delta_2 T - x)]\}\}.$$

This follows from Anderson's Theorem 4.2 and straightforward algebra.

The case of symmetric geometry becomes

$$P_{\mu 0}(x)$$

$$(4.9) \quad = \sum_{r=-\infty}^{\infty} \{h_{2r}(\gamma x/T, \ \gamma(\gamma + \delta T)/T) - h_{(2r-1)}[\gamma X/T, \ \gamma(\gamma + \delta T)/T\}$$

$$= \sum_{s=-\infty}^{\infty} (-1)^s h_s(\gamma x/T, \ \gamma(\gamma + \delta T)/T).$$

Notice that $P_{\mu 0}$ does not actually depend on $\mu$. This is because the conditional process $X(t)$ given $X(T)$ is independent of $\mu$ for $0 \leqq t \leqq T$.

COMPUTING $Q_\mu(x)$. Applying Anderson's Theorem 4.1,

$$Q_\mu(x) = (\exp [-2(X - x)(\delta_1 - \mu)] - 1)$$

$$(4.10) \qquad\qquad \cdot (\exp [-2(X - \gamma_1 - \delta_1 T)(\delta_1 - \mu)] - 1)^{-1} \quad \text{for} \quad \mu \neq \delta_1$$

$$= (X - x)/(X - \gamma_1 - \delta_1 T) \qquad\qquad \text{for} \quad \mu = \delta_1.$$

For symmetric procedures,

$$Q_\mu(x) = (\exp [-2x(\delta + \mu)] - 1)$$

$$(4.11) \qquad\qquad \cdot (\exp [2(\gamma + \delta T)(\delta + \mu)] - 1)^{-1} \quad \text{for} \quad \mu \neq -\delta$$

$$= -x/(\gamma + \delta T) \qquad\qquad \text{for} \quad \mu = -\delta.$$

We are now in a position to evaluate the probability of accepting $H_1$ using (4.1). The author has frequently found it convenient to evaluate the integral numerically for application purposes. (Considerable computing time may be saved by formally evaluating a representative integration of an expression which can be identified with a typical summand. Thus, the concept of "modular programming" can be used to avoid very complicated algebraic expressions.) However, it is possible to evaluate the integral formally. This is a straightforward but extremely tedious job. We shall be content with simply stating the result for symmetric procedures. The algebraic details may be found in the author's thesis [3].

$$P_\mu \text{ [accepting } H_1]$$

$$\doteq \exp [2(\gamma + \delta T)(\delta + \mu)](\exp [2(\gamma + \delta T)(\delta + \mu)] - 1)^{-1}$$

$$(4.12) \qquad\qquad \cdot \sum_{s=-\infty}^{\infty} (-1)^{s+1} h_s(\gamma\mu, \gamma\delta)\Phi([2(\gamma + \delta T) + \mu T - 2s\gamma]T^{-\frac{1}{2}})$$

$$+ (\exp [2\gamma + \delta T)(\delta + \mu)] - 1)^{-1}$$

$$\cdot \sum_{s=-\infty}^{\infty} (-1)^s h_s(\gamma\mu, \gamma\delta)\Phi((\mu T - 2s\gamma)T^{-\frac{1}{2}})$$

$$+ \sum_{s=-\infty}^{0} (-1)^s h_s(\gamma\mu, \gamma\delta) \qquad\qquad \mu \neq -\delta.$$

For $\mu = -\delta$,

$$P_\mu \text{ [accepting} \quad H_1] \doteq (\gamma + \delta T)^{-1}\{\sum_{s=1}^{\infty} (-1)^{s+1}[\delta T + 2s\gamma]h_s(-\gamma\delta, \gamma\delta)$$

$$(4.13) \qquad + \sum_{s=-\infty}^{\infty} (-1)^s[\delta T + 2s\gamma]h_s(-\gamma\delta, \gamma\delta)\Phi((\delta T + 2s\gamma)T^{-\frac{1}{2}})$$

$$+ (T/2\pi)^{\frac{1}{2}} \sum_{s=-\infty}^{\infty} (-1)^s h_s(-\gamma\delta, \gamma\delta) \exp [-(\delta T + 2s\gamma)^2/2T]\}.$$

The probability of accepting $H_3$ may be found in an analogous way. Then the probability of accepting $H_2$ may be found by subtraction.

**5. Computing the ASN.** The average sample size $E_\mu(N)$ is approximately equal to the average time required by the approximating Wiener process to contact one of the four boundary lines of Figure I. Let $\tau$ be the sample time for the Wiener process. Then

$$(5.1) \qquad E_\mu(N) \doteq E_{\mu 1} + \int_{\gamma_1 + \delta_1 T}^{\gamma_2 + \delta_2 T} P_{\mu 0}(x) F_\mu(x) n(x \mid \mu T, \sigma^2 T) \, dx,$$

where

$$(5.2) \quad E_{\mu 1} \equiv \int_0^T P_\mu[\tau > t] \, dt = \int_0^T \int_{\gamma_1 + \delta_1 t}^{\gamma_2 + \delta_2 t} P_\mu[\tau > t \mid X(t) = x] n(x \mid \mu t, \sigma^2 t) \, dx \, dt$$

and

$$(5.3) \qquad F_\mu(x) \equiv \int_T^\infty P_\mu[\tau > t \mid \tau > T, X(T) = x] \, dt.$$

$P_{\mu 0}(x)$ and $n(x \mid \mu T, \sigma^2 T)$ have been defined, previously in Section 4. Again, for convenience, we set $\sigma^2 = 1$.

COMPUTING $E_{\mu 1}$. $E_{\mu 1}$ may be computed from the double integral of (5.2). It will be observed that $P_\mu[\tau > T \mid X(T) = x] = P_{\mu 0}(x)$ which was previously evaluated. Identifying $t$ with $T$ in (4.8) and (4.9) we immediately get $P_\mu[\tau > t \mid X(t) = x]$. For symmetric procedures, the double integral evaluates as

$$E_{\mu 1} = T \sum_{s=-\infty}^\infty (-1)^s h_s(\gamma\mu, \gamma\delta)[\Phi([\gamma + (\delta + \mu)T - 2s\gamma]T^{-\frac{1}{2}})$$

$$+ \Phi([\gamma + (\delta - \mu)T + 2s\gamma]T^{-\frac{1}{2}}) - 1]$$

$$+ (\delta + \mu)^{-1} \sum_{s=-\infty}^\infty (-1)^s (\gamma - 2s\gamma) h_s(\gamma\mu, \gamma\delta) \Phi([\gamma + (\delta + \mu)T - 2s\gamma]T^{-\frac{1}{2}})$$

$$+ (\delta - \mu)^{-1} \sum_{s=-\infty}^\infty (-1)^s (\gamma + 2s\gamma) h_s(\gamma\mu, \gamma\delta) \Phi([\gamma + (\delta - \mu)T + 2s\gamma]T^{-\frac{1}{2}})$$

$$+ (\delta + \mu)^{-1} \sum_{s=-\infty}^0 (-1)^{s+1} (\gamma - 2s\gamma) h_s(\gamma\mu, \gamma\delta)$$

$$+ (\delta - \mu)^{-1} \sum_{s=0}^\infty (-1)^{s+1} (\gamma + 2s\gamma) h_s(\gamma\mu, \gamma\delta).$$

COMPUTING $F_\mu(x)$. $F_\mu(x)$ can be interpreted as an expectation, namely, $E_\mu(\tau - T \mid \tau > T, X(T) = x)$. Thus we are faced with the problem of computing the expected time for a Wiener process to contact one of two parallel lines. It follows (see [3], equation (2.4.40).), for $X \leq x \leq \gamma_2 + \delta_2 T$, that

$$F_\mu(x) = N_1(x)/D_1(x)$$

where

$$N_1(x) = \{[\gamma_2 + \delta_2 T - x][\exp [2(X - x)(\delta_2 - \mu)] - 1]$$

$$- [X - x][\exp [2(\gamma_2 + \delta_2 T - x)(\delta_2 - \mu)] - 1]\},$$

$$D_1(x) = \{[\delta_2 - \mu][\exp [2(\gamma_2 + \delta_2 T - x)(\delta_2 - \mu)] - \exp [2(X - x)(\delta_2 - \mu)]]\}.$$

For $\gamma_1 + \delta_1 T \leq x \leq X$,

$$F_\mu(x) = N_2(x)/D_2(x)$$

where

$$N_2(x) = \{[X - x][\exp[2(\gamma_1 + \delta_1 T - x)(\delta_1 - \mu)] - 1]$$

$$- [\gamma_1 + \delta_1 T - x][\exp[2(X - x)(\delta_1 - \mu)] - 1]\},$$

$$D_2(x) = \{[\delta_1 - \mu][\exp[2(X - x)(\delta_1 - \mu)] - \exp[2(\gamma_1 + \delta_1 T - x)(\delta_1 - \mu)]]\}.$$

We shall simply state the results of formal evaluation of (5.1) for symmetric procedures:

$$\begin{aligned}
E_\mu(N) &\doteq -(\gamma + \delta T)(\delta + \mu)^{-1} P_\mu \text{ [accepting } H_1] \\
&\quad - (\gamma + \delta T)(\delta - \mu)^{-1} P_\mu \text{ [accepting } H_3] + \delta T(\delta + \mu)^{-1} \\
&\quad + [(\delta + \mu)^{-1} + (\delta - \mu)^{-1}] \sum_{s=0}^{\infty} (-1)^{s+1}(-\mu T + 2s\gamma) h_s(\gamma\mu, \gamma\delta) \\
(5.4)\quad &\quad + [(\delta + \mu)^{-1} + (\delta - \mu)^{-1}] \sum_{s=-\infty}^{\infty} (-1)^s(-\mu T + 2s\gamma) h_s(\gamma\mu, \gamma\delta) \\
&\quad \cdot \Phi((-\mu T + 2s\gamma) T^{-\frac{1}{2}}) \\
&\quad + [(\delta + \mu)^{-1} + (\delta - \mu)^{-1}] (T/2\pi)^{\frac{1}{2}} \sum_{s=-\infty}^{\infty} (-1)^s h_s(\gamma\mu, \gamma\delta) \\
&\quad \cdot \exp[-(-\mu T + 2s\gamma)^2/2T].
\end{aligned}$$

It is not clear why the acceptance probabilities should fit so nicely into the right hand side of (5.4). It should be recalled that a similar thing happens with Wald's SPRT. There, the explanation is quite clear.

**6. Normalization.** Let $X^*(t^*) \sim N(\mu^* t^*, \sigma^{*2} t^*)$ be an arbitrary Wiener process and suppose that we have three simple hypotheses $H_1^*$, $H_2^*$, and $H_3^*$ where $H_i^*$ is the hypothesis that $\mu^* = \mu_i^*$ for $i = 1, 2, 3$ and $\mu_1^* < \mu_2^* < \mu_3^*$. We may normalize the problem with the following transformation:

$$(6.1) \qquad X(t) \equiv (\mu_2^* - \mu_1^*)(\sigma^*)^{-2}[X^*(t^*) - \mu_2^* t^*];$$

$$(6.2) \qquad t \equiv (\mu_2^* - \mu_1^*)^2(\sigma^*)^{-2} t^*;$$

$$(6.3) \qquad \mu \equiv (\mu^* - \mu_2^*)(\mu_2^* - \mu_1^*)^{-1}.$$

Then

$$(6.4) \qquad X(t) \sim N(\mu t, t),$$

and testing $H_i^*: \mu^* = \mu_i^*$, for $i = 1, 2, 3$, is equivalent to testing $H_i : \mu = \mu_i$, for $i = 1, 2, 3$, where

$$(6.5) \quad \mu_1 = -1, \qquad \mu_2 = 0, \qquad \text{and} \quad \mu_3 = (\mu_3^* - \mu_2^*)(\mu_2^* - \mu_1^*)^{-1}.$$

There is a one to one correspondence between the geometrical parameters $\gamma_1$, $\gamma_2$, $\delta_1$, $\delta_2$, $T$, and $X$ for the normalized problem and the geometrical parameters of the unnormalized problem:

$$(6.6) \qquad \gamma_i^* = \sigma^{*2}(\mu_2^* - \mu_1^*)^{-1}\gamma_i, \qquad\qquad i = 1, 2.$$

$$(6.7) \qquad \delta_i^* = (\mu_2^* - \mu_1^*)\delta_i + \mu_2^*, \qquad\qquad i = 1, 2.$$

$$(6.8) \qquad T^* = \sigma^{*2}(\mu_2{}^* - \mu_1{}^*)^{-2}T.$$

$$(6.9) \qquad X^* = (\sigma^*)^2(\mu_2{}^* - \mu_1{}^*)^{-1}X + \mu_2{}^*(\sigma^*)^2(\mu_2{}^* - \mu_1{}^*)^{-2}T.$$

Using, these transformed geometrical parameters, we find that

$$(6.10) \qquad P_{\mu*}^*[\text{accepting} \quad H_i{}^*] = P_\mu[\text{accepting} \quad H_i], \quad \text{for} \quad i = 1, 2, 3,$$

where $P_{\mu*}^*$ and $P_\mu$ are the corresponding probability measures.

Finally if $\tau$ and $\tau^*$ are the two sampling times for the normalized and unnormalized problems, we have

$$(6.11) \qquad E_{\mu*}^*(\tau^*) = (\sigma^*)^2(\mu_2{}^* - \mu_1{}^*)^{-2}E_\mu(\tau),$$

where $E_{\mu*}^*$ and $E_\mu$ are the two expectation operators.

The author [3] has constructed extensive tables for the normalized problem which extend to the unnormalized problem by using (6.5) through (6.11).

## REFERENCES

[1] ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

[2] ARMITAGE, P. (1950). Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *Suppl. J. Roy. Statist. Soc.* **12** 137–144.

[3] SIMONS, G. D. (1966). Multihypothesis testing. Ph.D. Thesis, Department of Statistics, University of Minnesota.

[4] SIMONS, G. D. (1967). Lower bounds for average sample number of sequential multihypothesis tests. *Ann. Math. Statist.* **28** 1343–1364.

[5] 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. Statist.* **20** 502–522.

[6] WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.