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

By

Gordon Simons

University of Minnesota [2]

Technical Report No. 84

1/ Research primarily sponsored by the Air Force Office of Aerospace Research, United States Air Force, under AFOSR Grant Nr 885-65, at the University of Minnesota. Partially sponsored by the National Science Foundation Grant GP-5705 at Stanford University.

2/ Currently at Stanford University.

1. Summary.

This paper examines a sequential testing procedure for choosing one of three simple hypotheses concerning the unknown mean μ 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 $(\mu = \mu_1, \mu_2 \text{ or } \mu_3)$, 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 (O.C.) and the average sample size (ASN) for all values of μ. 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 [4].

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, $\theta = \theta_o$ is sufficiently informative but its rejection probably is not. One would reasonably like to know whether $\theta < \theta_o$ or $\theta > \theta_o$ when θ 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 $X_1, X_2, \ldots$ be a sequence of independent, normally distributed random variables with unknown common mean μ and known common variance σ^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 3×3 "error matrix" $A = (\alpha_{ij})$ where $\alpha_{ij} = P_{\mu_i}$[accepting H_j] for $i,j=1,2,3$.

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 α_{13} and α_{31} tend to be small in comparison to α_{23} and α_{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 α_{13} and α_{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 straight line passing through the points

$$\left(n, \frac{\sigma^2}{\mu_i-\mu_j} \log A_{ij} + \frac{\mu_i+\mu_j}{2} n\right)$$

for $n = 0,1,\ldots$, 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 looks graphically as follows.

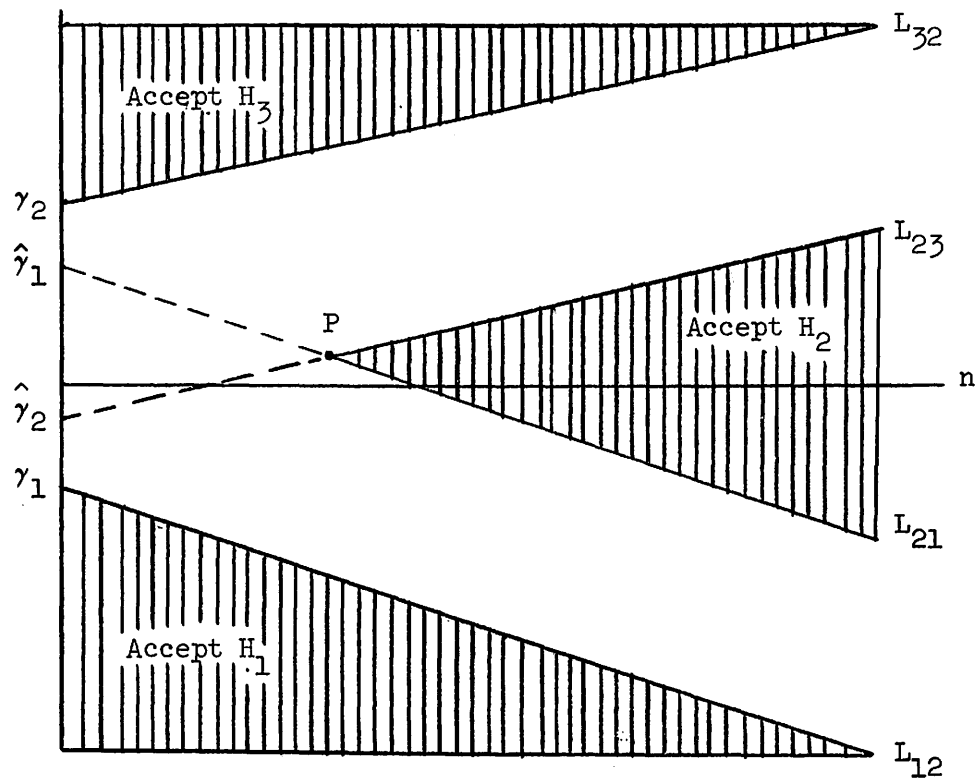

FIGURE I.

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

and

(2.2) $$\alpha_{ii} \geq 1 - \sum_{j \neq i} A^{-1}_{ji} .$$

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" γ_1, γ_2, δ_1, δ_2, X, and T where γ_1 and γ_2 are the intercepts of L_{12} and L_{32}, δ_1 is the slope of L_{12} and L_{21}, δ_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 μ 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 $n = 0,1,\ldots$ to the continuous Wiener process $X(t) \sim N(\mu t, \sigma^2 t)$, $t \geq 0$. 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 O.C. 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 approximate O.C. 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 [4], one adapted to our three states of nature problem. We explain this modification by referring again to Figure I:

Plot (n, S_n) for $n = 0,1,\ldots$ 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 $n_1 \geq n_2$ and (n_1, S_{n_1}) is below line L_{12}.

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

(c) Accept H_2 otherwise.

For reasons of mathematical convenience, they insist that

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

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 α_{11}, α_{22} and α_{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 α_{13} and α_{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 α_{11} and α_{22}. Indirectly we are controlling α_{21}, α_{23}, and α_{33} and virtually controlling α_{12}, α_{13}, α_{31}, and α_{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 α_{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 δ_1 and δ_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 δ_1 and δ_2. For a given value of μ, what values of δ_1 and δ_2 will minimize $E_\mu(\tau)$? ($E_\mu(\tau)$ denotes the expected sampling time for the Wiener process with drift parameter μ.)

The question is nearly answered for $\mu = \mu_1$, μ_2 and μ_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 $\mu = \mu_1$, μ_2, and μ_3 with the same pair of δ's. The evidence if 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 μ_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 and II, this slope minimizes $E_\mu(N)$ for $\mu = \mu_1$ and μ_2 simultaneously. (See Wald and Wolfowitz [5].) Quasi-optimality appears to hold for the Sobel-Wald modified procedures also.

C. Comparing the ASN with a theoretical lower bound.

In this subsection we are concerned with comparing the ASN for one of our procedures which satisfy certain error constraints with a theoretical lower bound to the ASN for all possible procedures satisfying those constraints. The author [3] has found two different theoretical lower bounds which will not be discussed here.

Example 1: Problem: $\mu_1 = -.1$, $\mu_2 = 0$, $\mu_3 = .2$

Constraints: $\alpha_{11} = \alpha_{22} = \alpha_{33} = .95$, $\alpha_{21} = \alpha_{23} = .025$

μ	-.2	-.1	-.05	0	.1	.2	.3
ASN for author's quasi-optimal test	242.5	661.8	1072	574.5	287.9	168.5	90.9
First lower bound for the ASN	87.6	661.4	787.3	561.0	196.8	165.4	48.2
Second lower bound for the ASN	104.0	335.0	883.3	335.0	220.8	83.8	43.0

Clearly the author's test does very well when $\mu = \mu_1, \mu_2$, or μ_3. Since the lower bounds are not actually attainable, it is impossible to say how much we can improve upon the test for other values of μ. Actually, the entries in the first row are based on the Wiener approximation so the true entries are larger than those given. It is the author's judgment that this difference is slight. (See the next subsection on the Monte Carlo study.) It is typical for the ASN to look good for $\mu = \mu_1$ and μ_3 but sometimes its goodness is not certain at $\mu = \mu_2$.

Example 2: Problem: $\mu_1 = -.1$, $\mu_2 = 0$, $\mu_3 = .1$

Constraints: $\alpha_{11} = \alpha_{22} = \alpha_{33} = .95$, $\alpha_{21} = 1/60$, $\alpha_{23} = 2/60$

μ	-.2	-.1	-.05	0	.05	.1	.2
ASN for author's quasi-optimal test	269.5	741.2	1167	803.3	972.7	609.8	223.3
First lower bound for the ASN	96.3	738.4	852.0	572.4	738.2	606.9	81.2
Second lower bound for the ASN	109.3	353.8	940.0	353.6	867.5	318.5	99.3

D. Small sample size results.

T. W. Anderson [1] raises the question of how inaccurate are the O.C. 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 3: Problem: $\mu_1 = -1$, $\mu_2 = 0$, $\mu_3 = 1$.

Intended constraints (using Wiener process): $\alpha_{11} = \alpha_{22} = \alpha_{33} = .98$, $\alpha_{21} = \alpha_{23} = .01$.

μ	-1	-.5	0
Predicted P_μ[accepting H_1] for author's test	.980	.443	.01
Actual P_μ[accepting H_1] for author's test	.989	.452	.006
Predicted P_μ[accepting H_2] for author's test	.020	.557	.980
Actual P_μ [accepting H_2] for author's test	.011	.548	.988
Predicted P_μ [accepting H_1] for Sobel-Wald test	.980	.460	.01
Actual P_μ [accepting H_1] for Sobel-Wald test	.988	.464	.006
Predicted P_μ [accepting H_2] for Sobel-Wald test	.020	.540	.980
Actual P_μ [accepting H_2] for Sobel-Wald test	.012	.536	.988

The "actual" entries are each based on 10,000 Monte Carol 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. 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.

μ	$-1=\mu_1$	$-.5$	$0=\mu_2$
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

It seems intuitively reasonable that the increased "cost" in the ASN should be compensated for by improved O.C. 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 O.C. 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 contact with L_{12}' before any other part of the boundary. Then

$$(4.1) \qquad P_\mu[\text{accepting } H_1] \doteq P_{\mu 1} + \int_{\gamma_1+\delta_1 T}^{X} P_{\mu o}(x) Q_\mu(x) \mathcal{N}(x|\mu T, \sigma^2 T)dx ,$$

where

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

(4.3) $P_{\mu o}(x) \equiv P_{\mu}[X(t)$ contacts neither L_{12} nor L_{32} before time $T | X(T) = x]$,

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

and $\mathfrak{N}(x|a,b)$ 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

$$(4.5) \quad P_{\mu 1} = \sum_{r=0}^{\infty} e^{-2[(r+1)\gamma_1 - r\gamma_2][(r+1)(\delta_1-\mu) - r(\delta_2-\mu)]} \Phi\left(\frac{-(\delta_1-\mu)T - 2r\gamma_2 + (2r+1)\gamma_1}{\sqrt{T}}\right)$$

$$+ e^{-2[r^2\gamma_1(\delta_1-\mu) + r^2\gamma_2(\delta_2-\mu) - r(r+1)\gamma_1(\delta_2-\mu) - r(r-1)\gamma_2(\delta_1-\mu)]}$$

$$\cdot \Phi\left(\frac{+(\delta_1-\mu)T - 2r\gamma_2 + (2r+1)\gamma_1}{\sqrt{T}}\right)$$

$$- e^{-2[(r+1)^2\gamma_1(\delta_1-\mu) + (r+1)^2\gamma_2(\delta_2-\mu) - r(r+1)\gamma_1(\delta_2-\mu) - (r+1)(r+2)\gamma_2(\delta_1-\mu)]}$$

$$\cdot \Phi\left(\frac{-(\delta_1-\mu)T - 2(r+1)\gamma_2 + (2r+1)\gamma_1}{\sqrt{T}}\right)$$

$$- e^{-2[r\gamma_1 - (r+1)\gamma_2][r(\delta_1-\mu) - (r+1)(\delta_2-\mu)]} \Phi\left(\frac{+(\delta_1-\mu)T - 2(r+1)\gamma_2 + (2r+1)\gamma_1}{\sqrt{T}}\right) ,$$

where

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

In the case of symmetric procedures, namely, when

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

(4.5) simplifies to

$$\begin{aligned} P_{\mu 1} = \sum_{r=0}^{\infty} \Bigg\{ & h_{(2r+1)}(\gamma\mu,\gamma\delta) \cdot \Phi\left(\frac{(\delta+\mu)T-(4r+1)\gamma}{\sqrt{T}}\right) \\ & + h_{-2r}(\gamma\mu,\gamma\delta) \cdot \left[1-\Phi\left(\frac{(\delta+\mu)T+(4r+1)\gamma}{\sqrt{T}}\right)\right] \\ & - h_{2(r+1)}(\gamma\mu,\gamma\delta) \cdot \Phi\left(\frac{(\delta+\mu)T-(4r+3)\gamma}{\sqrt{T}}\right) \\ & - h_{-(2r+1)}(\gamma\mu,\gamma\delta) \cdot \left[1-\Phi\left(\frac{(\delta+\mu)T+(4r+3)\gamma}{\sqrt{T}}\right)\right]\Bigg\} , \end{aligned}$$

where

$$h_s(u,v) \equiv e^{-2us-2vs^2} .$$

Further simplification yields

$$(4.7) \qquad P_{\mu 1} = \sum_{s=-\infty}^{\infty} (-1)^{s+1} h_s(\gamma\mu,\gamma\delta)\Phi\left(\frac{(\delta+\mu)T-(2s-1)\gamma}{\sqrt{T}}\right) + \sum_{s=-\infty}^{0} (-1)^s h_s(\gamma\mu,\gamma\delta).$$

Computing $P_\mu(x)$: For $\gamma_1+\delta_1 T \leqq x \leqq \gamma_2+\delta_2 T$ (the only case we need), we have

$$(4.8) \qquad P_{\mu o}(x) = \sum_{r=-\infty}^{\infty} \Big\{ e^{-(2/T)[r^2\{\gamma_1(\gamma_1+\delta_1 T-x)+\gamma_2(\gamma_2+\delta_2 T-x)\}-r(r-1)\gamma_1(\gamma_2+\delta_2 T-x)-r(r+1)\gamma_2(\gamma_1+\delta_1 T-x)]}$$

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

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

The case of symmetric geometry becomes

$$(4.9) \qquad P_{\mu o}(x) = \sum_{r=-\infty}^{\infty} \left\{ h_{2r}\left(\frac{\gamma x}{T}, \frac{\gamma(\gamma+\delta T)}{T}\right) - h_{(2r-1)}\left(\frac{\gamma x}{T}, \frac{\gamma(\gamma+\delta T)}{T}\right)\right\}$$

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

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

Computing $Q_\mu(x)$: Applying Anderson's theorem 4.1,

$$(4.10) \qquad Q_\mu(x) = \begin{cases} (e^{-2(X-x)(\delta_1-\mu)}-1)(e^{-2(X-\gamma_1-\delta_1 T)(\delta_1-\mu)}-1)^{-1} & \text{for } \mu \neq \delta_1 \\ \dfrac{X-x}{X-\gamma_1-\delta_1 T} & \text{for } \mu = \delta_1 . \end{cases}$$

For symmetric procedures,

$$(4.11) \quad Q_{\mu}(x) = \begin{cases} (e^{-2x(\delta+\mu)}-1)(e^{2(\gamma+\delta T)(\delta+\mu)}-1)^{-1} & \text{for } \mu \neq -\delta \\ \dfrac{-x}{\gamma+\delta T} & \text{for } \mu = -\delta \end{cases} .$$

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. 1/ 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].

1/ 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.

$$(4.12) \quad P_\mu[\text{accepting } H_1] \doteq$$

$$e^{2(\gamma+\delta T)(\delta+\mu)}(e^{2(\gamma+\delta T)(\delta+\mu)}-1)^{-1} \sum_{s=-\infty}^{\infty} (-1)^{s+1} h_s(\gamma\mu,\gamma\delta)\, \Phi\left(\frac{2(\gamma+\delta T)+\mu T-2s\gamma}{\sqrt{T}}\right)$$

$$+(e^{2(\gamma+\delta T)(\delta+\mu)}-1)^{-1} \sum_{s=-\infty}^{\infty} (-1)^{s} h_s(\gamma\mu,\gamma\delta)\, \Phi\left(\frac{\mu T-2s\gamma}{\sqrt{T}}\right)$$

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

For $\mu = -\delta$,

$$(4.13) \quad P_\mu[\text{accepting } H_1] \doteq$$

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

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

$$\left.+\sqrt{\frac{T}{2\pi}} \sum_{s=-\infty}^{\infty} (-1)^{s}\, h_s(-\gamma\delta,\gamma\delta)e^{-(\delta T+2s\gamma)^2/2T}\right\} .$$

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 τ be the sample time for the Wiener process. Then

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

where

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

and

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

$P_{\mu o}(x)$ and $\eta(x|\mu T, \sigma^2 T)$ 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 $P_\mu[\tau > T|X(T)=x] = P_{\mu o}(x)$ which was previously evaluated. Identifying t with T in (4.8) and (4.9), we immediately get $P_\mu[\tau > t|X(t)=x]$. For symmetric procedures, the double integral evaluates as

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

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

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

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

Computing $F_\mu(x)$: $F_\mu(x)$ can be interpreted as an expectation, namely, $E_\mu(\tau-T|\tau > T,\ X(T)=x)$. 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 $X \leqq x \leqq \gamma_2+\delta_2 T$, that

$$F_\mu(x) = \frac{[\gamma_2+\delta_2T-x]\left[e^{2(X-x)(\delta_2-\mu)}-1\right]-[X-x]\left[e^{2(\gamma_2+\delta_2T-x)(\delta_2-\mu)}-1\right]}{[\delta_2-\mu]\left[e^{2(\gamma_2+\delta_2T-x)(\delta_2-\mu)}-e^{2(X-x)(\delta_2-\mu)}\right]} .$$

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

$$F_\mu(x) = \frac{[X-x]\left[e^{2(\gamma_1+\delta_1T-x)(\delta_1-\mu)}-1\right]-\left[\gamma_1+\delta_1T-x\right]\left[e^{2(X-x)(\delta_1-\mu)}-1\right]}{[\delta_1-\mu]\left[e^{2(X-x)(\delta_1-\mu)}-e^{2(\gamma_1+\delta_1T-x)(\delta_1-\mu)}\right]} .$$

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

$$(5.4) \quad E_\mu(N) \doteq -\frac{\gamma+\delta T}{\delta+\mu} P_\mu[\text{accepting } H_1] - \frac{\gamma+\delta T}{\delta-\mu} P_\mu[\text{accepting } H_3] + \frac{\delta T}{\delta+\mu}$$

$$+ \left[\frac{1}{\delta+\mu} + \frac{1}{\delta-\mu}\right] \sum_{s=0}^{\infty} (-1)^{s+1}(-\mu T+2s\gamma)h_s(\gamma\mu,\gamma\delta)$$

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

$$+ \left[\frac{1}{\delta+\mu} + \frac{1}{\delta-\mu}\right] \sqrt{\frac{T}{2\pi}} \sum_{s=-\infty}^{\infty} (-1)^{s}h_s(\gamma\mu,\gamma\delta)e^{-(-\mu T+2s\gamma)^2/2T}$$

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:

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

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

(6.3)
$$\mu \equiv \frac{\mu^*-\mu_2^*}{\mu_2^*-\mu_1^*} \quad .$$

Then

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

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)
$$\mu_1 = -1 \ , \ \mu_2 = 0 \ , \text{ and } \ \mu_3 = \frac{\mu_3^*-\mu_2^*}{\mu_2^*-\mu_1^*} \ .$$

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

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

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

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

Using, these transformed geometrical parameters, we find that

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

where $P^*_{\mu^*}$ and P_μ are the corresponding probability measures.

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

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

where $E^*_{\mu^*}$ and E_μ 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 sequential analysis 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 descriminant function analysis." Supplement to the J. of the Roy. Statist. Soc., **12**, No. 1, 137-144.

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

[4] 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.

[5] Wald, A., and Wolfowitz, J. (1948). "Optimum character of the sequential probability ratio test." Ann. Math. Statist., **19**, 326-339.