MINIMAX SEQUENTIAL TESTS OF SOME COMPOSITE HYPOTHESES[1]

By M. H. DeGroot

Carnegie Institute of Technology

1. Introduction and summary. Let $X(t)$, $t \geqq 0$, be a Wiener process with unknown mean μ per unit time and unit variance per unit time. Thus, $X(0) = 0$ and for any $t_2 > t_1 \geqq 0$, $X(t_2) - X(t_1)$ is normally distributed with mean $(t_2 - t_1)\mu$ and variance $t_2 - t_1$. Furthermore, for any sequence

$$0 \leqq t_{11} < t_{12} \leqq t_{21} < t_{22} \leqq \cdots \leqq t_{k1} < t_{k2},$$

the random variables $X(t_{j2}) - X(t_{j1})$, $j = 1, \cdots, k$, are independent.

The process may be observed continuously beginning at $t = 0$ and the problem is to decide between the hypotheses that $\mu \leqq \mu_0$ and $\mu > \mu_0$, where μ_0 is a given number, which without loss of generality is taken as 0. Thus the hypotheses are

$$\begin{aligned} H_0 &: \quad \mu \leqq 0 \\ H_1 &: \quad \mu > 0. \end{aligned} \tag{1.1}$$

It is assumed that the cost of observing the process for a time t is bt, where $b > 0$, and that $W_i(\mu)$, the cost of accepting $H_i (i = 0, 1)$ when μ is the true mean, is of the form

$$\begin{aligned} W_0(\mu) &= \begin{cases} 0 & \text{for } \mu \leqq 0 \\ c\mu^r & \text{for } \mu > 0 \end{cases} \\ W_1(\mu) &= \begin{cases} c\,|\,\mu\,|^r & \text{for } \mu \leqq 0 \\ 0 & \text{for } \mu > 0 \end{cases} \end{aligned} \tag{1.2}$$

where $c > 0$ and $0 < r \leqq 2$.

The main result of this paper is that under these conditions the minimax decision procedure is a certain sequential probability ratio test (SPRT). The reason for restricting r to the interval $0 < r \leqq 2$ will be brought out in the derivation given in Section 3.

In Section 6, the analogous problem of testing the hypotheses (1.1) about the mean of a normal distribution is considered. The minimax procedure found for the Wiener process provides, in an obvious fashion, an approximation to the minimax procedure for this problem. Approximations of this type have been discussed in the literature. For $r = 1$, Moriguti [10] and Maurice [9] have found the approximate minimax procedure in a certain class of symmetric SPRT's. The same procedure is mentioned by Johnson in the discussion following [8].

Received January 12, 1960.

[1] This research was supported by the National Science Foundation under grant NSF-G9662.

Breakwell [2], [3], [4], has treated similar problems for the binomial and Poisson distributions. The work to be presented here not only puts all of this on a rigorous basis for the Wiener process but shows that, for the Wiener process, the minimax SPRT is in fact minimax among all decision procedures.

Finally, it is shown in Section 6 that if the cost per observation is large, the true minimax procedure for the normal decision problem is to take exactly one observation and then accept one of the hypotheses.

2. Loss functions and symmetric SPRT's. In this section I begin the discussion of the decision problem for the Wiener process. For any decision procedure δ let $P_i(\mu, \delta)$ denote the probability of accepting H_i, $i = 0, 1$, when μ is the true mean and let $T(\mu, \delta)$ denote the expected total observation time when μ is the true mean. Then the loss function for the decision procedure δ is

$$L(\mu, \delta) = \sum_{i=0}^{1} W_i(\mu)P_i(\mu, \delta) + bT(\mu, \delta), \tag{2.1}$$

and it follows from (1.2) that this can be written as

$$L(\mu, \delta) = \begin{cases} c\,|\,\mu\,|^r P_1(\mu, \delta) + bT(\mu, \delta) & \text{for } \mu \leqq 0 \\ c\,\mu^r P_0(\mu, \delta) + bT\,(\mu, \delta) & \text{for } \mu > 0. \end{cases} \tag{2.2}$$

The problem is to find a decision procedure δ^*, if one exists, such that

$$\max_\mu L(\mu, \delta^*) = \min_\delta \max_\mu L(\mu, \delta). \tag{2.3}$$

Of special importance is the class of symmetric SPRT's. A decision procedure belongs to this class if it satisfies the following conditions: (i) there exists a positive constant h such that the process is observed as long as $|\,X(t)| < h/2$; (ii) if at some t, $|\,X(t)| \geqq h/2$, then observation stops and either H_0 or H_1 is accepted, according as $X(t) \leqq -h/2$ or $X(t) \geqq h/2$.

The decision procedures δ_h belonging to this class are conveniently indexed by the positive constant h mentioned in the definition.

As is well-known [7],

$$\begin{aligned} P_0(\mu, \delta_h) &= 1/(e^{\mu h} + 1), \\ P_1(\mu, \delta_h) &= 1 - P_0(\mu, \delta_h), \\ T(\mu, \delta_h) &= h(e^{\mu h} - 1)/[2\mu(e^{\mu h} + 1)]. \end{aligned} \tag{2.4}$$

The singularity of $T(\mu, \delta_h)$ at $\mu = 0$ is removable and it is easily seen that $T(0, \delta_h) = h^2/4$.

Substituting these expressions in (2.2) gives, for any $h > 0$,

$$L(\mu, \delta_h) = \begin{cases} \dfrac{c\mu^r}{e^{\mu h} + 1} + \dfrac{bh(e^{\mu h} - 1)}{2\mu(e^{\mu h} + 1)} & \text{for } \mu \geqq 0 \\ L(-\,\mu, \delta_h) & \text{for } \mu < 0. \end{cases} \tag{2.5}$$

3. The minimax symmetric SPRT. In this section, values $h = h^*$ and $\mu = \mu^*$

will be found such that

$$L(\mu^*, \delta_{h^*}) = \min_h L(\mu^*, \delta_h) = \max_\mu L(\mu, \delta_{h^*}). \tag{3.1}$$

Thus, δ_{h^*} will be the minimax decision procedure in the class of all symmetric SPRT's.

It is convenient to make the following transformation of variables:

$$\begin{aligned} h &= (b^{-1}c)^{(r+2)^{-1}}\eta, \qquad \mu = (bc^{-1})^{(r+2)^{-1}}m, \\ \mathcal{L}(m, \eta) &= (b^r c^2)^{-(r+2)^{-1}} L(\mu(m), \delta_{h(\eta)}). \end{aligned} \tag{3.2}$$

Clearly, if values $\eta = \eta^*$ and $m = m^*$ can be found such that

$$\mathcal{L}(m^*, \eta^*) = \min_\eta \mathcal{L}(m^*, \eta) = \max_m \mathcal{L}(m, \eta^*), \tag{3.3}$$

then the corresponding values of h and μ will satisfy (3.1).

Making the substitutions (3.2) in (2.5) gives, for any $\eta > 0$,

$$\mathcal{L}(m, \eta) = \begin{cases} \dfrac{m^r}{e^{m\eta} + 1} + \dfrac{\eta(e^{m\eta} - 1)}{2m(e^{m\eta} + 1)} & \text{for } m \geqq 0 \\ \mathcal{L}(-m, \eta) & \text{for } m \leqq 0. \end{cases} \tag{3.4}$$

The convenience of the substitutions (3.2) is seen in the elimination of the constants b and c from (3.4). Furthermore, the symmetry exhibited in (3.4) makes it possible to restrict the search for values m^* and η^* that satisfy (3.3) to the region $m \geqq 0$. Finally, it should be remembered that $\mathcal{L}(m, \eta)$ is, for each fixed $\eta > 0$, continuous at $m = 0$.

Now fix $\eta > 0$. An elementary computation yields, for $m > 0$,

$$\begin{aligned} &\partial\mathcal{L}(m, \eta)/\partial m \lesseqgtr 0 \Leftrightarrow \\ &r(1 + e^{-m\eta}) - m\eta \lesseqgtr \eta^{r+2}[\sinh (m\eta)/((m\eta)^{r+1}) - (1/(m\eta)^r)] \Leftrightarrow \\ &r(1 + e^{-m\eta}) - m\eta \lesseqgtr \eta^{r+2}[((m\eta)^{2-r}/3!) + ((m\eta)^{4-r}/5!) + \cdots]. \end{aligned} \tag{3.5}$$

Denote the left-hand and right-hand sides of the final inequality in (3.5) by Φ and Ψ, respectively. Then Φ is a strictly decreasing function of m and

$$\lim_{m\to 0} \Phi = 2r, \qquad \lim_{m\to\infty} \Phi = -\infty. \tag{3.6}$$

For $0 < r < 2$, Ψ is a strictly increasing function of m and

$$\lim_{m\to 0} \Psi = 0, \qquad \lim_{m\to\infty} \Psi = \infty. \tag{3.7}$$

Hence, for each fixed $\eta > 0$, there exists a unique positive value of m at which $\Phi = \Psi$ and this value yields $\max_m \mathcal{L}(m, \eta)$.

Now consider the case when $r = 2$. Again, Ψ is a strictly increasing function of m, but now

$$\lim_{m\to 0} \Psi = \eta^4/6, \qquad \lim_{m\to\infty} \Psi = \infty. \tag{3.8}$$

A glance at (3.6) shows that if $\eta^4 < 24$ there will be a unique positive value of m at which $\Phi = \Psi$ and this value will again yield $\max_m \mathcal{L}(m, \eta)$. However, if $\eta^4 \geqq 24$, then $\max_m \mathcal{L}(m, \eta)$ occurs at $m = 0$. (It should be clear from (3.5) and this discussion why the values $r > 2$ are excluded from consideration.)

Similarly, for fixed $m > 0$, an easy computation gives

$$\begin{aligned} \partial\mathcal{L}(m, \eta)/\partial\eta \lesseqgtr 0 &\Leftrightarrow [\sinh(m\eta)/m\eta] + 1 \lesseqgtr m^{r+1}/\eta \\ &\Leftrightarrow 2 + [(m\eta)^2/3!] + [(m\eta)^4/5!] + \cdots \lesseqgtr m^{r+1}/\eta. \end{aligned} \tag{3.9}$$

It is clear from the final inequality in (3.9) that there is a unique positive value of η at which equality holds and this is the value that yields $\min_\eta \mathcal{L}(m, \eta)$, for each $m > 0$.

It follows from this discussion that if positive values $m = m^*$ and $\eta = \eta^*$ can be found that simultaneously satisfy the equations

$$\begin{cases} r(1 + e^{-m\eta}) - m\eta = (\eta^2/m^r)\,[\sinh(m\eta)/(m\eta) - 1] \\ [\sinh(m\eta)/m\eta] + 1 = m^{r+1}/\eta \end{cases} \tag{3.10}$$

and the added condition that when $r = 2$, $\eta^{*4} < 24$, then the values m^* and η^* will satisfy the minimax equation (3.3).

Setting $m = v/\eta$ in the second equation of (3.10) gives

$$\eta^{r+2} = v^{r+2}/(v + \sinh v), \tag{3.11}$$

and making these substitutions in the first equation of (3.10) yields

$$r(1 + e^{-v}) = 2v \sinh v/(v + \sinh v). \tag{3.12}$$

A routine analysis shows that there is a unique positive value of v satisfying (3.12) and, consequently, there exist unique values $m^* > 0$ and $\eta^* > 0$ satisfying (3.10). It remains to show that when $r = 2$, $\eta^{*4} < 24$. From (3.11), it follows that it is sufficient to show that

$$1/v^3 + (\sinh v/v^4) > \tfrac{1}{24} \tag{3.13}$$

for all $v > 0$. An examination of the first few terms in the series expansion of the left-hand side of (3.13) shows that this inequality holds.

Thus, the following result has been obtained. Let v^* be the unique positive solution of (3.12). Let

$$\eta^{*\,r+2} = v^{*\,r+2}/(v^* + \sinh v^*), \; m^* = v^*/\eta^*. \tag{3.14}$$

Then m^* and η^* satisfy (3.3) and, hence, the values

$$h^* = (b^{-1}c)^{(r+2)^{-1}}\eta^*, \qquad \mu^* = (bc^{-1})^{(r+2)^{-1}}m^*, \tag{3.15}$$

satisfy the minimax equation (3.1).

4. The minimax decision procedure. It will now be shown that the decision procedure δ_{h^*} derived in the preceding section as the minimax procedure in the class of symmetric SPRT's, is in fact minimax in the class of all decision procedures.

Consider the problem of finding the decision procedure that is Bayes against the *a priori* distribution that places probability $\frac{1}{2}$ at each of the two values $\mu = \mu^*$ and $\mu = -\mu^*$. That is, it is desired to find the procedure δ that minimizes

$$\rho(\delta) = \tfrac{1}{2}[L(\mu^*, \delta) + L(-\mu^*, \delta)]. \tag{4.1}$$

It is well-known [7] that a Bayes solution for this problem is either a procedure that makes an outright decision without any observation of the process, or else it is a procedure of the following form. The process is observed as long as

$$B < \frac{(\sqrt{2\pi t})^{-1}e^{-2t^{-1}(X(t)-\mu^*t)^2}}{(\sqrt{2\pi t})^{-1}e^{-2t^{-1}(X(t)+\mu^*t)^2}} < A, \tag{4.2}$$

where $B < 1 < A$ are constants, or equivalently, as long as

$$\ln B/(2\mu^*) < X(t) < \ln A/(2\mu^*). \tag{4.3}$$

Observation stops and the appropriate hypothesis is accepted as soon as either inequality in (4.3) is broken.

In the problem being considered here, the Bayes procedure cannot be an outright decision. This follows from the fact that for any procedure δ_0 specifying an outright decision,

$$\rho(\delta_0) = c\mu^{*r}/2 = \lim_{h\to 0} L(\mu^*, \delta_h) > L(\mu^*, \delta_{h^*}) = \rho(\delta_{h^*}). \tag{4.4}$$

Furthermore, it follows easily from the derivation given in [13] or [1] that, because of the symmetry of the *a priori* distribution and the cost function, it must be true that $\ln A = -\ln B$ in the Bayes procedure (4.3). (Expressed in other terms, if it is worthwhile to continue observation when the *a posteriori* probability that $\mu = \mu^*$ is α then it must also be worthwhile to continue observation when the *a posteriori* probability that $\mu = -\mu^*$ is α, and conversely.) Thus, the Bayes procedure is a symmetric SPRT. It obviously must be δ_{h^*}, since for any other symmetric SPRT, δ_h,

$$\rho(\delta_h) = L(\mu^*, \delta_h) > L(\mu^*, \delta_{h^*}) = \rho(\delta_{h^*}). \tag{4.5}$$

It may now be concluded that δ_{h^*} is minimax among all decision procedures. Indeed, for any decision procedure, δ,

$$\rho(\delta) = \tfrac{1}{2}[L(\mu^*, \delta) + L(-\mu^*, \delta)] \geqq \rho(\delta_{h^*}) = L(\mu^*, \delta_{h^*}). \tag{4.6}$$

Hence, either $L(\mu^*, \delta) \geqq L(\mu^*, \delta_{h^*})$ or $L(-\mu^*, \delta) \geqq L(\mu^*, \delta_{h^*})$. Since $L(\mu^*, \delta_{h^*})$ is the maximum value of $L(\mu, \delta_{h^*})$, the conclusion follows.

In the preceding development, the fact that the minimax decision procedure belongs to the class of symmetric SPRT's is a consequence of the existence of a pair, h^* and μ^*, satisfying (3.1). When $r > 2$, it cannot be concluded that the minimax procedure is a symmetric SPRT because the existence of such a pair has not been established.

5. Tests of hypotheses about the mean of a normal distribution. In this section I consider the analogous problem of testing hypotheses about the mean μ of a normal distribution with unit variance. Thus, suppose $X_1, X_2, \cdots$ is a sequential sample of independent observations, each with this distribution. It is desired to decide between the hypotheses (1.1) when the cost per observation is b and the cost of an incorrect decision is given by (1.2).

The similarities between the problem treated in the preceding sections and the one now being considered are clear. The symmetric SPRT's defined in Section 2 have obvious counterparts here, with $X(t)$ replaced by $\sum_{i=1}^{n} X_i$. The expressions given in (2.4) are the usual approximations, [5], [7], [12], for the OC and ASN functions of these tests (where $T(\mu, \delta_h)$ is now interpreted as the expected number of observations). Finally, the optimal property of the SPRT used in Section 4 is applicable to the problem now being considered [13].

It follows from these statements that the minimax procedure derived above for the Wiener process can serve as an "approximate" minimax procedure for the problem now being considered. However, it will now be shown that for sufficiently large values of b/c the actual minimax procedure is to take exactly one observation and then accept one of the hypotheses.

A decision procedure is said to be a generalized SPRT if it is of the following type: there are given two sequences $\{\alpha_n\}$ and $\{\beta_n\}$, with $\beta_n \leqq \alpha_n$ (either may be infinite) for $n = 1, 2, \cdots$; sampling continues as long as

$$\beta_n < \sum_{i=1}^{n} X_i < \alpha_n ; \tag{5.1}$$

sampling stops and the appropriate hypothesis is accepted as soon as either inequality is broken.

It is known, [6], [11], that the class of generalized SPRT's is essentially complete relative to the class of all decision procedures with bounded loss functions. It should be noted that since $W_0(\mu)$ and $W_1(\mu)$ are unbounded, any procedure with a bounded loss function must involve taking at least one observation.

Let δ^* be the decision procedure under which one observation, X_1, is taken and either H_0 or H_1, is accepted, according as $X_1 < 0$ or $X_1 > 0$. I will now show that δ^* is minimax if b/c is sufficiently large.

Clearly,

$$L(\mu, \delta^*) = \begin{cases} c\mu^r\Phi(-\mu) + b & \text{for } \mu \geqq 0 \\ L(-\mu, \delta^*) & \text{for } \mu < 0 \end{cases} \tag{5.2}$$

where

$$\Phi(y) = \int_{-\infty}^{y} (2\pi)^{-1/2} e^{-x^2/2}\, dx. \tag{5.3}$$

Since $L(\mu, \delta^*) = L(-\mu, \delta^*)$, the maximum value of $L(\mu, \delta^*)$ occurs at two points, say $\mu = \pm\mu_0$, with $\mu_0 > 0$.

Now let δ be any other generalized SPRT. If $0 \leqq \beta_1 \leqq \alpha_1$ then $L(\mu_0, \delta) \geqq L(\mu_0, \delta^*)$, since the probability of making an incorrect decision on the first observation is at least as large using δ as it is using δ^*. Similarly, if $\beta_1 \leqq \alpha_1 \leqq 0$ then $L(-\mu_0, \delta) \geqq L(-\mu_0, \delta^*)$.

Finally, suppose that $\beta_1 < 0 < \alpha_1$ and let $\Pr\{\beta_1 < X_1 < \alpha_1 \mid \mu_0\} = \xi > 0$. Then

$$\begin{aligned} L(\mu_0, \delta) &= c\mu_0^r \Pr\{\text{Acc. } H_0 \mid \mu_0, \delta\} + bE\{n \mid \mu_0, \delta\} \\ &> c\mu_0^r \Pr\{X_1 \leqq \beta_1 \mid \mu_0, \delta\} + b[(1 - \xi) + 2\xi] \\ &= c\mu_0^r\Phi(\beta_1 - \mu_0) + b + b\xi. \end{aligned} \tag{5.4}$$

But

$$\xi = \Phi(\alpha_1 - \mu_0) - \Phi(\beta_1 - \mu_0), \tag{5.5}$$

and hence

$$\Phi(\beta_1 - \mu_0) = \Phi(\alpha_1 - \mu_0) - \xi > \Phi(-\mu_0) - \xi. \tag{5.6}$$

Thus

$$L(\mu_0, \delta) > c\mu_0^r\Phi(-\mu_0) + b + \xi(b - c\mu_0^r). \tag{5.7}$$

It follows that if $b \geqq c\mu_0^r$, then $L(\mu_0, \delta) > L(\mu_0, \delta^*)$ and δ^* is minimax.

It is interesting to note that when $b \geqq c\mu_0^r$ there is no least favorable *a priori* distribution. In fact, the Bayes procedure against the *a priori* distribution that places probability $\frac{1}{2}$ at each of the values $\mu = \pm\mu_0$ is an outright decision.

6. Acknowledgment. I wish to thank L. J. Savage, who originally suggested problems of this type to me while I was at the University of Chicago.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954, Chap. 10.

[2] J. V. Breakwell, "The problem of testing for the fraction of defectives," *J. Op. Res. Soc. Amer.*, Vol. 2 (1954), pp. 59–69.

[3] J. V. Breakwell, "Minimax test for the parameter of a Poisson process," (abstract) *Ann. Math. Stat.*, Vol. 26 (1955), p. 768.

[4] J. V. Breakwell, "Economically optimum acceptance tests," *J. Amer. Stat. Assn.*, Vol. 51 (1956), pp. 243–256.

[5] D. A. Darling and A. J. F. Siegert, "The first passage problem for a continuous Markov process," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 624–639.

[6] M. H. De Groot, "The essential completeness of the class of generalized sequential probability ratio tests," submitted to *Ann. Math. Stat.*

[7] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Sequential decision problems for processes with continuous time parameter. Testing hypotheses," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 254–264.
[8] P. M. Grundy, M. J. R. Healy, and D. H. Rees, "Economic choice of the amount of experimentation," *J. Roy. Stat. Soc.*, Ser. B, Vol. 18 (1956), pp. 32–55.
[9] R. J. Maurice, "A minimax procedure for choosing between two populations using sequential sampling," *J. Roy. Stat. Soc.*, Ser. B, Vol. 19 (1957), pp. 255–261.
[10] S. Moriguti, "Notes on sampling inspection plans," *Rep. Stat. Appl. Res., JUSE*, Vol. 3 (1955), pp. 1–23.
[11] M. Sobel, "An essentially complete class of decision functions for certain standard sequential problems," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 319–337.
[12] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1947, Chap. 7.
[13] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–339.