# A MODIFICATION OF THE SEQUENTIAL PROBABILITY RATIO TEST TO REDUCE THE SAMPLE SIZE[1]

By T. W. Anderson

*Columbia University and*
*Center for Advanced Study in the Behavioral Sciences*

**1. Summary and introduction.** The sequential probability ratio test is constructed as a sequential test of one simple hypothesis against another. In many instances a parametric form is assumed for the density or (discrete) probability function, and the two simple hypotheses are specified by two values of the parameter. The sequential probability ratio test has an optimum property for these two hypotheses, namely, given such a test there is no other test with at least as low probabilities of Type I and Type II errors and with smaller expected sample sizes under either or both of the two hypotheses. Usually, however, one is interested in the performance of the procedure for more values of the parameter than these two. A disadvantage of the sequential probability ratio test is that in general the expected sample size is relatively large for values of the parameter between the two specified ones; that is, in cases in which one does not care greatly which decision is taken, a large number of observations is expected. The question is how to reduce the expected sample size for values of the parameter when this tends to be large.

In this paper we consider a special case of the problem, when the distribution is normal with known variance and the parameter of interest is the mean. The sequential probability ratio test in this case consists in taking observations sequentially and after each observation is taken comparing the sum of the observations (referred to a suitable origin) with two constants. In this study the two constants are replaced by two linear functions of the number of observations taken, and the taking of observations is truncated (Section 2). Approximations to the operating characteristic (or power function) and the average sample size number are given (Section 4 and 5). Computations for two cases of special interest show a considerable decrease in average sample size at parameter values between the two specified ones (Section 3).

The problem is studied by replacing the sum of observations by the Wiener stochastic process (of a continuous time parameter); this can be thought of intuitively as interpolating between observations in a manner consistent with the addition of independent random variables. For this procedure we calculate exactly the operating characteristic, the distribution of observation time, the expected observation time, and related probabilities.

---

Received July 24, 1959.

[1] This research was sponsored in part by the Office of Naval Research under Contract Number Nonr-266(33), Project Number NR 042-034. Reproduction in whole or in part is permitted for any purpose of the United States Government.

**2. The problem and the procedures studied.** Let $f(x, \theta)$ be a family of densities or (discrete) probability functions of a scalar random variable $X$ with $\theta$ a scalar parameter. Suppose that $\theta$ is unknown, and that we are going to take observations on $X$ to determine whether $\theta$ is large or small. One way of formalizing this problem is to say we are going to test the null hypothesis $H_0$ that $\theta = \theta_0$ against the alternative hypothesis $H_1$ that $\theta = \theta_1 > \theta_0$ where $\theta_0$ and $\theta_1$ are two suitably chosen numbers. The sequential probability ratio test is a procedure for this testing problem. Let

$$z(x) = \log \frac{f(x, \theta_1)}{f(x, \theta_0)}, \tag{2.1}$$

and choose two numbers $a$ and $b$ ($<a$). The procedure consists of taking observations $x_1, x_2, \cdots$ sequentially. At the $m$th step, if

$$b < \sum_1^m z(x_i) < a, \tag{2.2}$$

take another observation; if the sum is not greater than $b$, accept $H_0$ (equivalently reject $H_1$); and if the sum is not less than $a$, accept $H_1$ (equivalently reject $H_0$).

It is convenient to summarize the characteristics of this test (or any other sequential test) by two functions, namely, the operating characteristic

$$L(\theta) = \Pr\{\text{accepting } H_0 \mid \theta\} \tag{2.3}$$

(which is the complement of the power function) and the expected sample size, $\mathcal{E}_\theta n$, which is the expected number of observations when sampling from $f(x, \theta)$. In terms of these functions at $\theta_0$ and $\theta_1$ the sequential probability ratio test has an optimum property. If $L(\theta)$ and $\mathcal{E}_\theta n$ are the operating characteristic and expected sample size for a sequential probability ratio test and $L^*(\theta)$ and $\mathcal{E}_\theta^* n$ are the same functions for another test, then if

$$L^*(\theta_0) \geqq L(\theta_0), \; L^*(\theta_1) \leqq L(\theta_1),$$

it follows that $\mathcal{E}_{\theta_0}^* n \geqq \mathcal{E}_{\theta_0} n$ and $\mathcal{E}_{\theta_1}^* n \geqq \mathcal{E}_{\theta_1} n$. That is, if the second test is as good as the sequential probability ratio test with respect to the probabilities of decisions when sampling from $f(x, \theta_0)$ and $f(x, \theta_1)$ it cannot be better (and usually will be worse) with respect to the expected number of observations.

Usually, however, one is interested in the behavior of a procedure over a range of values of the parameter, not just a pair of values. In many situations one's desire to take a certain action increases as the parameter increases; the customary way of setting up a sequential probability ratio test requires an evaluation of the desirability of taking actions relative to values of parameters by specifying two values, $\theta_0$ and $\theta_1$ and the desired probabilities of actions at these two parameter values. This is a somewhat arbitrary way of formalizing the real life problem and is perhaps done mainly as a convenience for the theoretical statistician. However, it may be that these requirements on the operating characteristic for reasonable procedures control the operating characteristic enough so that it is satisfactory. On the other hand, controlling the expected sample size at these two values does not necessarily yield an expected sample size function that can be considered satisfactory over a range of parameter values. In particular, the sequential probability ratio test in many problems has an expected sample size function that is much higher for values of $\theta$ between $\theta_0$ and $\theta_1$ than for these two values. Between $\theta_0$ and $\theta_1$ presumably one is less interested in which of the two actions is taken, but here is where one has to take a large number of observations. For example, in one case considered below the expected number of observations goes up to about 5/3 of the number at $\theta_0$ and $\theta_1$; in another case it more than doubles. The question is whether there are other sequential procedures which will reduce the expected number of observations for parameter values in the middle of the range without increasing it much at $\theta_0$ and $\theta_1$.

Another difficulty with the sequential probability ratio test is that for most cases the number of observations is a random variable which is unbounded and has a positive probability of being greater than any given constant. Since it is awkward to provide for taking an arbitrarily large number of observations, frequently the sequential probability ratio test is truncated; that is, if a certain number of observations are taken, then the process is stopped and a decision is made. This modified procedure (with different numbers $a$ and $b$) may considerably increase the expected sample size at $\theta_0$ and $\theta_1$. We can also consider better methods of modifying the sequential probability ratio test so as to limit the number of observations that can be taken.

The sequential probability ratio test is defined in terms of $Z_m = \sum_1^m z(x_i)$. The procedure can be described graphically in the plane of $m$ and $Z$. There are two lines $Z = a$ and $Z = b$. Sampling is stopped as soon as the sequence $Z_1$, $Z_2$, $\cdots$ leaves the strip between the two lines, and the decision depends on which line is crossed. One might consider modifications of these boundaries to obtain other procedures (called generalized sequential probability ratio tests by Weiss). If the densities have the so called Koopman-Darmois form,

$$\exp[\alpha(\theta) + \beta(\theta)\gamma(x) + \delta(x)],$$

the probability ratio for any two values of $\theta$ depends on the observation only through $\gamma(x)$, and inference can be based on $\sum \gamma(x_i)$ which is equivalent to $Z$. To control the expected sample size it seems reasonable to put an upper bound on the number of observations that can be taken. One also expects that a good procedure should lead to decision after a small number of observations if $Z$ is either very large or very small. Another intuitive impression is that the boundaries should be smooth. (The truncated sequential probability ratio test seems inefficient because if $Z$ is large and the number of observations is near the truncation value a few additional observations will not permit much chance of rejecting $H_1$ whatever these observations are.)

There are many ways of formalizing the problem that has been discussed.

We can set $L(\theta_0)$ and $L(\theta_1)$ and ask for the procedure that minimizes $\sup_\theta \mathcal{E}_\theta n$ or that minimizes $\mathcal{E}_{\theta^*} n$ for some specified $\theta^*$ between $\theta_0$ and $\theta_1$. In some cases the former specification reduces to the latter (for example, if there is sufficient symmetry).

Kiefer and Weiss [8] have shown that the procedures which minimize $\mathcal{E}_{\theta^*} n$ for various specified pairs $L(\theta_0)$ and $L(\theta_1)$ are essentially the class of solutions to the Bayes problems, in which the three parameter values are given *a priori* probabilities, say $\xi_0$, $\xi_1$ and $\xi^*$, and

$$\xi_0[1 - L(\theta_0)] + \xi_1 L(\theta_1) + \xi^* \mathcal{E}_{\theta^*} n \tag{2.4}$$

is to be minimized. Furthermore, if $f(x, \theta)$ is of the Koopman-Darmois form[2] a Bayes procedure is defined in terms of the probability ratio (2.1) and has continuation regions

$$b_m < \sum_1^m z(x_i) < a_m \tag{2.5}$$

for $m = 1, \cdots, M$. At most $M$ observations are taken, and the two sequences of numbers usually satisfy $b_m < b_{m+1}$, $a_{m+1} < a_m$. In principle, the fact that at most a given number of observations is to be taken permits computation of the Bayes solution; the expected loss at the $M$th stage is a function of *a posteriori* probabilities; the best action for any such probabilities is clear. In turn, the *a posteriori* probabilities at the $M$th stage depend on the *a posteriori* probabilities at the $(M - 1)$st stage and the $M$th observation; this leads to computation of the best action or whether to continue sampling at the $(M - 1)$st stage. The computations can be carried back to the first observation. Unfortunately, even for the normal and binomial distributions these computations are laborious and the procedures do not seem to be easy to describe.

In this paper we consider a particular testing problem, namely, that of the mean of a normal distribution with known variance. The procedures studied consist of pairs of straight lines (not necessarily parallel) with possible truncation.

Observations are drawn sequentially from a normal distribution with mean $\mu$ and variance $\sigma^2$. We want to decide whether $\mu$ is large or small given knowledge of $\sigma^2$, and we want to keep the sample size down, particularly at moderate values of $\mu$. We can put the problem formally that we want to test the null hypothesis $\mu = \mu_0$ against the alternative $\mu = \mu_1$ $(\mu_1 > \mu_0)$ and wish a procedure to minimize $\mathcal{E}_\mu n$ at $\mu = \frac{1}{2}(\mu_0 + \mu_1)$ or alternatively to minimize the supremum of $\mathcal{E}_\mu n$. It is convenient to replace the observation $x$ by the transformed observation

[2] Assumption A of [8] essentially implies this for the three parameter values if $f(x, \theta_1)/f(x, \theta_0)$ takes on all values from 0 to $\infty$.

It might be helpful to readers of [8] to point out that for given *a priori* probabilities the *a posteriori* probabilities after $m$ observations, say $\xi_0(m)$, $\xi_1(m)$, $\xi^*(m)$, lie on a curve $\xi^*(m) = Ck^m \xi_0^c(m) \xi_1^{1-c}(m)$, where $C > 0$, $k > 1$, $0 < c < 1$; each such curve lies nearer the point $\xi^* = 1$ than the curves for observation numbers less than $m$.

$[x - \frac{1}{2}(\mu_0 + \mu_1)]/\sigma$ and call $\mu^* = \frac{1}{2}(\mu_1 - \mu_0)/\sigma$. Then the problem is to test the null hypothesis $\mu = -\mu^*(\mu^* > 0)$ against the alternative $\mu = \mu^*$ when sampling from $N(\mu, 1)$.

The probability ratio in this problem is $z(x) = 2\mu^* x$. The sequential probability ratio test has the continuation interval

$$b < 2\mu^* \sum_1^m x_i < a. \tag{2.6}$$

Equivalently, it can be represented as

$$\bar{b} < \sum_1^m x_i < \bar{a}, \tag{2.7}$$

where $\bar{b} = b/(2\mu^*)$ and $\bar{a} = a/(2\mu^*)$. The test can be described graphically in the plane of $m$ and $y = \sum x_i$. There are two parallel lines $y = \bar{b}$ and $y = \bar{a}$. One plots successively the points $(1, x_1)$, $(2, x_1 + x_2)$, $\cdots$. As soon as a point is obtained which is not between the lines, sampling is stopped; if the point is on or above $y = \bar{a}$, $H_0$ is rejected, and if the point is on or below $y = \bar{b}$, $H_0$ is accepted. In principle $\bar{a}$ and $\bar{b}$ are selected to attain the desired $L(\mu^*)$ and $L(-\mu^*)$, but in practice these are approximated.

In this paper we shall consider replacing the parallel straight lines $y = \bar{a}$ and $y = \bar{b}$ by arbitrary straight lines $y = c_1 + d_1 m$ and $y = c_2 + d_2 m$ with possibly truncation at $N$ (that is, a line $m = N$). Such a procedure is as follows: Take observations $x_1, x_2, \cdots$ sequentially. At the $m$th stage $(m < N)$, reject $H_0$ if

$$\sum_1^m x_i \geqq c_1 + d_1 m, \tag{2.8}$$

accept $H_0$ if

$$\sum_1^m x_i \leqq c_2 + d_2 m, \tag{2.9}$$

and take another observation if

$$c_2 + d_2 m < \sum_1^m x_i < c_1 + d_1 m. \tag{2.10}$$

If $N$ observations are taken, stop sampling and reject $H_0$ if $\sum_1^N x_i \geqq k$ and accept $H_0$ if $\sum_1^N x_i < k$. We take $c_1 > 0 > c_2$. To avoid redundancy in the definition (that is, intersection of the lines before $n = N$), we require

$$c_2 + d_2(N - 1) < c_1 + d_1(N - 1).$$

From the intuitive considerations mentioned above we might surmise that the desirable procedures of this type are those for which $d_1 < 0 < d_2$; that is, those for which the lines converge.

To calculate the probabilities and expected values that are of interest is extremely complicated and involved. However, we can calculate such quantities if we replace the sequence $\sum_1^m x_i$ ($m = 1, 2, \cdots$) by an analogous $X(t)$ ($0 \leqq t < \infty$). The random variable $\sum_1^m x_i$ is normally distributed with mean $m\mu$ and variance $m$; the increment from $m$ to $m'$, $\sum_1^{m'} x_i - \sum_1^m x_i = \sum_{m+1}^{m'} x_i$ is normally distributed independently of $\sum_1^m x_i$. These properties are retained in the Wiener stochastic process. Let $X(t)$ be a Gaussian (i.e., normally distributed) stochastic process with $\mathcal{E}X(t) = \mu t$, $\mathcal{E}[X(t) - \mu t]^2 = t$ and such that $X(s) - X(t)$ is distributed independently of $X(t)$ for $s > t$. Then $\sum_1^m x_i$ has the properties of $X(m)$. Throughout the paper we shall assume this process defined so that the probability is 1 that the path functions are continuous. Now for this family of processes we consider the problem of testing the hypothesis that $\mu = -\mu^*$ against the alternative $\mu = \mu^*$. The procedure is to observe $X(t)$ continuously as long as

$$c_2 + d_2 t < X(t) < c_1 + d_1 t \tag{2.11}$$

and $t \leqq T$. If $X(t) = c_1 + d_1 t$ or $X(t) = c_2 + d_2 t$, observation is stopped and in the first case $H_0$ is rejected and in the second case $H_0$ is accepted. If $X(t)$ remains between the two lines, observation is stopped at $t = T$ and $H_0$ is rejected if $X(T) \geqq k$ and accepted if $X(T) < k$. Since $X(t)$ is continuous with probability 1, the inequalities in (2.11) are violated directly by an equality. Here we require

$$c_2 + d_2 T \leqq k \leqq c_1 + d_1 T. \tag{2.12}$$

In this paper the probability of rejecting $H_0$ is computed as a function of $\mu$, and the expected length of time of observation is found. It is proposed here that one can select $c_1$, $c_2$, $d_1$, $d_2$, $T$ and $k$ so as to achieve a desirable procedure for $X(t)$; that is, obtain a specified significance level at $-\mu^*$ and a specified power at $\mu^*$ with some minimization of the expected time. Then the operating characteristic and expected time functions are approximations to the operating characteristic and expected sample number functions when observations are taken discretely.

It is difficult to ascertain how good these approximations are. One might hope they are as good as for the sequential probability ratio test, which is the special case of $d_1 = d_2 = 0$ and $T = \infty$. In principle this problem could be studied by considering the case of observations taken discretely as observing $X(t)$ at $t = 1, 2, \cdots$. It is clear that applying the procedure when $X(t)$ is only observed at discrete time points leads to decision later (or at least no earlier) than when observing it continuously. Hence, the expected time function underestimates the expected sample number function. In the case of the sequential probability ratio test the usual approximation underestimates the power function when it is over $\frac{1}{2}$ and overestimates it when it is under $\frac{1}{2}$ (that is, indicates poorer significance level and power at $\mu^*$ than is actually the case). Similarly here, at least when the lines converge to a point ($c_1 + d_1 T = c_2 + d_2 T$), the same is true for these procedures.

In the case of observations taken discretely, the properties of a procedure depend only on $\mu/\sigma$, and the scale can be changed so that $\sigma = 1$. In the case of the Wiener process we can change the scale on $x$ and the scale on $t$. Let $t = \beta s$ and define $X^*(s) = \alpha X(\beta s)$ $(\alpha > 0, \beta > 0)$. This has mean $\alpha\mu\beta s$ and variance $\alpha^2\beta s$. If $\alpha^2\beta = 1$ $[t = s/\alpha^2$ and $X^*(s) = \alpha X(s/\alpha^2)]$, the variance of $X^*(s)$ is $s$ and the mean is $(\mu/\alpha)s$. The region (2.11) goes into

$$\alpha c_2 + (d_2/\alpha)s < X^*(s) = \alpha X(s/\alpha^2) < \alpha c_1 + (d_1/\alpha)s, \tag{2.13}$$

the bound $t = T$ goes into $s = S = \alpha^2 T$, and the value $k$ goes into $\alpha k$. For any value of $\mu$ the probabilities for the procedure in terms of $X^*(s)$ are the same as for the related procedure in terms of $X(t)$, and the expected observation time in terms of $X^*(s)$ is $\alpha^2$ times the expected observation time in terms of $X(t)$. Thus a problem stated in terms of arbitrary $\mu_0$, $\mu_1$, and $\sigma^2$ is reduced to a problem in terms of $\mu_0 = -\mu^*$, $\mu_1 = \mu^*$ and $\sigma^2 = 1$, where $\mu^*$ is arbitrary. (The accuracy of these results for the continuous time parameter as an approximation for the case of observations taken discretely would depend on what the original parameters were.)

The problem of modifying sequential analysis to reduce the sample size has been considered by several statisticians. In the literature the case of a normal mean has been investigated. Armitage [2] proposed straight line boundaries for a two-sided test of $\mu = \mu_0$, converted to the Wiener process, but then only approximated the probabilities and expected time. Donnelly [3] has proposed straight line boundaries that meet and converted to the Wiener process; he obtained some results similar to those of this paper by a different method (that is, solutions of a partial differential equation satisfying certain boundary conditons).

**3. Numerical investigation of two cases.** As will be seen later, the operating characteristic and expected observation time are complicated functions of the parameter $\mu$ and the constants defining the procedure. It, therefore, seems hopeless to find analytically the optimum procedure within the class. Hence, two cases have been investigated computationally. The results given in Tables 1 and 2 show the advantage of the best procedures in these two cases.

TABLE 1

*Characteristics of Procedures with Probabilities of Types* I *and* II *Errors of 5% at* $\mu = -.1$ *and* $.1$

| Condition | $c$ | $T$ | Expected Time $\mu = 0$ | Expected Time $\mu = -.1$ and $.1$ |
|---|---|---|---|---|
| Fixed size | | 270.55 | 270.6 | 270.6 |
| SPRT | 14.722 | $\infty$ | 216.7 | 132.5 |
| $c + dT = 0$ | 19.905 | 600.25 | 192.2 | 139.2 |
| $c + dT = .1c$ | 20.083 | 529.00 | 192.2 | 139.3 |
| $c + dT = .2c$ | 20.340 | 457.10 | 192.2 | 139.8 |
| $c + dT = .3c$ | 20.025 | 416.16 | 192.4 | 139.4 |

TABLE 2

*Characteristics of Procedures with Probabilities of Types* I *and* II *Errors of 1% at* $\mu = -.1$ *and* $.1$

| Condition | $c$ | $T$ | Expected Time $\mu = 0$ | Expected Time $\mu = -.1$ and $.1$ |
|---|---|---|---|---|
| Fixed size | | 541.19 | 541.2 | 541.2 |
| SPRT | 22.976 | $\infty$ | 527.9 | 225.2 |
| $c + dT = 0$ | 35.52 | 870.26 | 402.2 | 249.4 |
| $c + dT = .1c$ | 35.52 | 783.24 | 402.1 | 249.4 |
| $c + dT = .2c$ | 35.52 | 699.63 | 402.8 | 249.8 |

The specified null hypothesis, $\mu = -\mu^*$, and alternative hypothesis, $\mu = \mu^*$, are symmetrically located about $\mu = 0$. We consider cases when the probabilities of Type I and Type II errors are equal; that is, $1 - L(-\mu^*) = L(\mu^*)$. For this probability fixed we want to find the procedure in our class that minimizes the expected observation time at $\mu = 0$. Since this problem is symmetric [that is, $X(t)$ can be replaced by $-X(t)$], it seems reasonable to consider only symmetric procedures; that is, $c_1 = -c_2 = c$ say, $d_1 = -d_2 = d$ say, and $k = 0$. (In fact, if an asymmetric procedure solved the problem, its mirror image would and so would a symmetric randomization between these two procedures, but one can argue that in this problem nonrandomized procedures form a complete class.) For symmetric procedures the maximum expected observation time[3] is at $\mu = 0$.

A symmetric procedure is defined by the constants $c$, $d$, and $T$. There is one condition imposed by specifying the (equal) probabilities of error, leaving two degrees of freedom in the constants. Subject to this condition the constants were varied to obtain the smallest expected observation time at $\mu = 0$. The most relevant results of these computations are given in Tables 1 and 2 for probabilities of equal Type I and II errors of .05 and .01, respectively.[4]

The line $x = c + dt$ has intercept $c$ at $t = 0$ and $c + dT$ at $t = T$; when $c + dT = 0$ the two straightline boundaries converge to a point. For each of several values of the ratio of these two intercepts $[(c + dT)/c = 0, .1, .2]$ the tables give the combination of $c$ and $T$ (and hence $d$) that approximately minimizes the expected observation time at $\mu = 0$.

The most interesting features of the numerical results are the comparisons of expected observation times between the sequential probability ratio test (SPRT) and the procedures for $c + dT = 0$ (lines converging to a point). The convergent line procedures show a considerable improvement over the sequential proba-

[3] This is a consequence of the fact that for a symmetric procedure the probability that the observation time exceeds a given value of $t$ is maximum for $\mu = 0$ (and is increasing for $\mu < 0$ and decreasing for $\mu > 0$). The latter follows from Corollary 5 of [1] applied to $X(t)/(c + dt)$. This demonstration was suggested by Hoeffding [7].

[4] I wish to acknowledge the computational assistance of Mrs. Judy Frankman and Mr. George Bump.

bility ratio tests at $\mu = 0$ with a moderate decrease in efficiency at $\mu = -.1$ and $\mu = .1$. At the 5% levels the expected time for convergent lines at $\mu = 0$ is 24.5 less than for the usual procedure and is 6.7 more at $\mu = \pm.1$ (a ratio of 3.7 to 1); at the 1% levels it is 125.7 less at $\mu = 0$ and 24.2 more at $\mu = \pm.1$ (a ratio of 5.2 to 1). Roughly speaking, we can say that when one operates at the 5% levels he is better off with the convergent lines procedure if intermediate values of $\mu$ occur at least $\frac{1}{5}$ of the time and when one operates at the 1% level if intermediate values occur at least $\frac{1}{6}$ of the time.

Hoeffding [7] has given a general lower bound for the expected sample size of a sequential procedure at one parameter value when the probabilities of error are specified at two other parameter values (assuming the variance of the sample size is finite). His lower bound for the expected time at $\mu = 0$ is 187.0 for the case in Table 1 and 388.3 for the case in Table 2. Thus the best procedure reported in Table 1 accomplishes at least 82% of the possible improvement over the sequential probability ratio test at the 5% level and at least 90% at the 1% level. The lower bound given by Hoeffding cannot be achieved. While it is unknown how much this lower bound underestimates the minimum expected sample size, the comparison between the bound and the results in the tables shows that the given bound does not underestimate the minimum by much and that the tests presented in this paper come close to yielding the minimum possible expected sample size.

A combination of $c$ and $T$ in the tables yields the required probabilities of Types I and II errors to 6 or 7 decimal places. The expected times are reported to one decimal place; there may be an error of .1 (or occasionally even .2) in these numbers. In particular in Table 2 the difference in the expected times at $\mu = 0$ between $c + dT = 0$ and $c + dT = .1c$ is not significant (402.17 compared to 402.147). For the values of $c$ given in the table one cannot distinguish between the cases $c + dT = 0$ and $c + dT = .1c$ because in the latter case the probability of reaching a decision at $t = T$ is almost 0.

As will be seen later, the probabilities and expected times can be given as infinite series of terms involving Mill's ratios. It was convenient to use tables [5], [9] (which were extended for these computations) where these are given to 5 decimal places, and this determined the eventual accuracy of the calculations. A good guess is that more accurate computation would not yield minimum expected times that differ from the figures in the tables by more than .1 or .2.

For a given ratio of $(c + dT)/c$ the value of $c$ that minimizes the expected time cannot be determined very accurately. For example, at the 1% levels at $c + dT = .1c$ the expected times at $\mu = 0$ were 402.21, $402.15^-$ and $402.25^-$ for $c = 34.695$, 35.52, and 36.345, respectively. Of course, since the functions are flat it is of no importance to obtain an accurate determination of where the minimum is. At the 1% levels the computations were done by setting $c$ and adjusting $T$ (and hence $d$); at the 5% levels they were done by setting $\sqrt{T}$ and adjusting[5] $c$. The variations in $c$ given in Table 1 are of no consequence; a more

[5] The former procedure is much preferable for comparing the different ratios of $(c + dT)/c$.

accurate determination of the $c$'s minimizing the expected times at $\mu = 0$ should find them much closer[6] than in Table 1 and not all equal as indicated in Table 2.

The computations have been done on the basis of the Wiener process and are considered as approximations to the problem of sampling discretely (as described in Section 2). It can be expected that the errors of approximation are greater than the errors of computation. Thus to the extent that one accepts the approximation one can consider the procedures given in the tables for $c + dT = 0$ and $c + dT = .1c$ as procedures nearly minimizing the expected sample size at $\mu = 0$.

## 4. Probabilities of error for the Wiener Process.

4.1. *Outline of derivations.* The process $X(t)$ with mean $\mathcal{E}X(t) = \mu t$ and variance $t$ depends on a single parameter $\mu$. The probability of accepting $H_0$ is the probability of the process $X(t)$ touching the lower boundary $x = c_2 + d_2 t$ before touching the upper boundary $x = c_1 + d_1 t$ and before $t = T$ plus the probability of the process staying between the boundaries to $t = T$ and $X(T) \leqq k$. This probability is the operating characteristic, which we shall denote by $L(\mu)$. Its complement $1 - L(\mu)$ is the power function.

Let $P_1(T)$ be the probability that the process touch the upper boundary before touching the lower boundary before $t = T$ and $P_2(T)$ be the probability that the process touch the lower boundary before touching the upper boundary before $t = T$. Then $P_0(T) = 1 - P_1(T) - P_2(T)$ is the probability that the process stay between the boundaries to $t = T$. In this section we shall find expressions for these various probabilities.

We can let

$$X(t) = Y(t) + \mu t, \tag{4.1}$$

where $Y(t)$ is a Wiener process with $\mathcal{E}Y(t) = 0$ and $\mathcal{E}Y^2(t) = t$. Then $X(t) = c_i + d_i t$ is equivalent to $Y(t) = c_i + (d_i - \mu)t$ and $X(T) \leqq k$ is equivalent to $Y(T) \leqq k - \mu T$. It will be convenient to obtain some of the results for $Y(t)$ and then convert them back to $X(t)$.

The (unconditional) distribution of $Y(T)$ is normal with mean 0 and variance $T$. Given $Y(T) = y$, the process $Y(t)$ is Gaussian (normal) with a certain expected value function and covariance function. We obtain $P_1(\mu, T)$ by finding the conditional probability of touching the upper boundary before the lower boundary and before $t = T$ given $Y(T) = y$ and then taking the expected value of this conditional probability relative to the marginal distribution of $Y(T)$. The process $Y(t)$ conditional on $Y(T)$ can be transformed into the Wiener process by a transformation which carries the original straightline boundaries into other straightline boundaries. The problem then becomes finding the

---

[6] The lack of monotonicity in the last column of Table 1 with respect to $(c + dT)/c$ appears to be due to $c = 20.340$ being larger than the other approximately minimizing values of $c$, which in turn seems due to variations in computing procedures. Since it is hoped to do a more extensive numerical investigation on a high-speed computing machine, it has not seemed worthwhile to carry out the present calculations more accurately.

probability that the Wiener process touches an upper straightline boundary before a lower one; this problem is handled first.

4.2. *Probability of going over one line first.* The first problem is to find the probability of touching one boundary before the other when the process can go on without limit ($T = \infty$). In this case the lines do not converge for then the process could not go on beyond the point of intersection of the two lines.

THEOREM 4.1. *If $Y(t)$ is the Wiener process with $\mathcal{E}Y(t) = 0$ and $\mathcal{E}Y^2(t) = t$, then for $\gamma_1 > 0$, $\gamma_2 < 0$ and $\delta_1 \geqq \delta_2$ (not $\delta_1 = \delta_2 = 0$) the probability that $Y(t) \geqq \gamma_1 + \delta_1 t$ for a smaller $t$ than any $t$ for which $Y(t) \leqq \gamma_2 + \delta_2 t$ is*

$$(4.2)\quad \begin{aligned} P_1 &= \sum_{r=1}^{\infty} \{e^{-2[r^2\gamma_1\delta_1+(r-1)^2\gamma_2\delta_2-r(r-1)(\gamma_1\delta_2+\gamma_2\delta_1)]} \\ &\qquad - e^{-2[r^2(\gamma_1\delta_1+\gamma_2\delta_2)-r(r-1)\gamma_1\delta_2-r(r+1)\gamma_2\delta_1]}\}, \qquad \delta_1 \geqq 0, \\ &= 1 - \sum_{r=1}^{\infty} \{e^{-2[(r-1)^2\gamma_1\delta_1+r^2\gamma_2\delta_2-r(r-1)(\gamma_1\delta_2+\gamma_2\delta_1)]} \\ &\qquad - e^{-2[r^2(\gamma_1\delta_1+\gamma_2\delta_2)-r(r+1)\gamma_1\delta_2-r(r-1)\gamma_2\delta_1]}\}, \qquad \delta_1 \leqq 0, \\ &= \frac{e^{-2\gamma_2\delta_1} - 1}{e^{2(\gamma_1-\gamma_2)\delta_1} - 1}, \qquad \delta_1 = \delta_2 \neq 0. \end{aligned}$$

PROOF. Let $A_i$ ($i = 1, 2, \cdots$) be the event of a path $y(t)$ touching (or going over) the upper line and then touching (or going below) the lower line and alternating touching the upper and lower lines until each has been touched $i - 1$ times followed by touching the upper line[7]; let $B_i$ be the event of touching the lower line and then touching the upper line and alternately touching the two lines until each has been touched $i$ times. Then the event whose probability we are finding, namely, touching the upper line before the lower line, is

$$(4.3)\quad A_1 - B_1 + A_2 - B_2 + \cdots.$$

For $\delta_1 > 0$, $\delta_2 < 0$ Doob [4] has shown that the probability of $A_r$ is

$$(4.4)\quad \begin{aligned} \alpha_r &= e^{-\frac{1}{2}[(2r)^2\gamma_1\delta_1+(2r-2)^2\gamma_2\delta_2-\{(2r-1)^2-1\}(\gamma_1\delta_2+\gamma_2\delta_1)]} \\ &= e^{-2[r^2\gamma_1\delta_1+(r-1)^2\gamma_2\delta_2-r(r-1)(\gamma_1\delta_2+\gamma_2\delta_1)]}, \end{aligned}$$

and the probability of $B_r$ is

$$(4.5)\quad \begin{aligned} \beta_r &= e^{-\frac{1}{2}[(2r)^2\gamma_1\delta_1+(2r)^2\gamma_2\delta_2-(2r)(2r-2)\gamma_1\delta_2-2r(2r+2)\gamma_2\delta_1]} \\ &= e^{-2[r^2(\gamma_1\delta_1+\gamma_2\delta_2)-r(r-1)\gamma_1\delta_2-r(r+1)\gamma_2\delta_1]}. \end{aligned}$$

We shall derive these results and show that they also hold when $0 \leqq \delta_2 < \delta_1$. Then the theorem follows directly for $0 < \delta_1$.

[7] In "touching" one line before "touching" the other line the path may contact and cross the first line several times before contacting the other line. For example, a path is in $A_2$ if it has contacted the upper line at some $t_1$, the lower line at some $t_2$ ($t_2 > t_1$), and the upper line at some $t_3$ ($t_3 > t_2$) regardless of other contacts with the lines.

The probabilities (4.4) and (4.5) result from a more general result given in Lemma 4.1. We consider a sequence of lines such that the odd-numbered lines are above the origin and each even-numbered line is below the origin and entirely below the odd-numbered line preceding it and the one following it in the sequence. We consider the event of a path touching the first line, then the second, $\cdots$, until $2r - 1$ are touched and also the event of touching the second, $\cdots$, until $2r - 2$ are touched.

LEMMA 4.1. *If* $L_i$ *is the line*

$$y = (-1)^{i+1}u_i t + (-1)^{i+1}v_i, \tag{4.6}$$

$v_i > 0$, $u_{2i-1} \geqq 0$, $-u_{2i} < u_{2i-1}$, $-u_{2i} < u_{2i+1}$, *the probability of the process* $Y(t)$ *touching* $L_1, L_2, \cdots, L_{2r-1}$ *in sequence is*

$$\alpha_r(u_1, v_1; \cdots; u_{2r-1}, v_{2r-1}) = \exp\left\{-2\left(\sum_{i=1}^{2r-1} u_i v_i + 2\sum_{i=2}^{2r-1}\sum_{j=1}^{i-1} u_i v_j\right)\right\}, \tag{4.7}$$

*and the probability of touching* $L_2, \cdots, L_{2r-1}$ *in sequence is*

$$\beta_{r-1}(u_2, v_2; \cdots; u_{2r-1}, v_{2r-1}) = \exp\left\{-2\left(\sum_{i=2}^{2r-1} u_i v_i + 2\sum_{i=3}^{2r-1}\sum_{j=2}^{i-1} u_i v_j\right)\right\}. \tag{4.8}$$

PROOF. For $v_1 > 0$, $u_1 \geqq 0$, the probability of reaching the one line $L_1$ is

$$\alpha_1(u_1, v_1) = e^{-2u_1 v_1}. \tag{4.9}$$

We shall reduce the other cases to this formula. Consider a path that touches $L_1, \cdots, L_{2r-2}$ in sequence and let $t_{2r-2}$ be the first value of $t$ for which the path touches $L_{2r-2}$ after touching $L_1, \cdots, L_{2r-3}$. The conditional probability of then touching $L_{2r-1}$ is

$$e^{-2u_{2r-1}[(u_{2r-1}+u_{2r-2})t_{2r-2}+v_{2r-1}+v_{2r-2}]} \tag{4.10}$$

for the line $L_{2r-1}$ has slope $u_{2r-1}$ and has intercept $(u_{2r-2}t_{2r-2} + v_{2r-2}) + (u_{2r-1}t_{2r-2} + v_{2r-1})$ when referred to $(t_{2r-2}, -u_{2r-2}t_{2r-2} - v_{2r-2})$ as origin. This conditional probability is the same as the conditional probability of touching the line with slope $u_{2r-1} + u_{2r-2}$ and which is at $t = t_{2r-2}$ a distance of

$$h = \frac{u_{2r-1}[(u_{2r-1} + u_{2r-2})t_{2r-2} + v_{2r-1} + v_{2r-2}]}{u_{2r-1} + u_{2r-2}} \tag{4.11}$$

above $L_{2r-2}$. This is also the same as the conditional probability of touching the line with slope $-(u_{2r-1} + u_{2r-2})$ and which is at $t = t_{2r-2}$ a distance of $h$ below $L_{2r-2}$ (since $Y(t)$ is symmetrically distributed and has independent increments). This last line is

$$y = -(u_{2r-2} + u_{2r-1})t - \frac{u_{2r-1}v_{2r-1} + u_{2r-2}v_{2r-2} + 2u_{2r-1}v_{2r-2}}{u_{2r-1} + u_{2r-2}}, \tag{4.12}$$

which does not depend on $t_{2r-2}$. Thus the probability of touching in sequence $L_1, \cdots, L_{2r-1}$ is the same as the probability of touching in sequence $L_1, \cdots,$

$L_{2r-2}$, and the last line (4.12). This last line, however, lies entirely below $L_{2r-2}$ and can be touched only if the path has touched $L_{2r-2}$ (which lies entirely below $L_{2r-3}$). Hence, this is the probability of touching in sequence $L_1, \cdots, L_{2r-3}$, and (4.12).

Now let us reduce this one step further. Let the line (4.12) be

$$L^*: y = -u^*t - v^*.$$

Note that $u^* > 0$ and $v^* > 0$. Let $t_{2r-3}$ be the first value of $t$ for which the path touches $L_{2r-3}$ after touching the preceding lines in sequence. Then the conditional probability of touching $L^*$ is

$$e^{-2u^*[(u^*+u_{2r-3})t_{2r-3}+v^*+v_{2r-3}]}. \tag{4.13}$$

By the same reasoning as before the probability sought for is

$$\alpha_{r-1}\left(u_1, v_1; \cdots; u_{2r-4}, v_{2r-4}; u_{2r-3} + u^*, \frac{u_{2r-3}v_{2r-3} + u^*v^* + 2u^*v_{2r-3}}{u_{2r-3} + u^*}\right). \tag{4.14}$$

From this it follows

$$\begin{aligned} &\alpha_r(u_1, v_1; \cdots; u_{2r-1}, v_{2r-1}) \\ &= \alpha_{r-1}\Big(u_1, v_1; \cdots; u_{2r-4}, v_{2r-4}; u_{2r-3} + u_{2r-2} + u_{2r-1}, \\ &\frac{u_{2r-3}v_{2r-3} + u_{2r-2}v_{2r-2} + u_{2r-1}v_{2r-1} + 2(u_{2r-2}v_{2r-3} + u_{2r-1}v_{2r-3} + u_{2r-1}v_{2r-2})}{u_{2r-3} + u_{2r-2} + u_{2r-1}}\Big). \end{aligned} \tag{4.15}$$

By carrying this procedure on, we arrive at

$$\alpha_r(u_1, v_1; \cdots; u_{2r-1}, v_{2r-1}) = \alpha_1\left(\sum u_i, \frac{\sum u_iv_i + 2\sum_{j<i} u_iv_j}{\sum u_i}\right), \tag{4.16}$$

which is (4.7).

The other part of the lemma follows similarly. It will be noted that the conditions for the lemma can be reduced further. Pairs of successive lines should not intersect; the slopes should satisfy $u_{2r-1} \geqq 0$, $u_{2r-1} + u_{2r-2} > 0$, $u_{2r-1} + u_{2r-2} + u_{2r-3} > 0, \cdots, \sum_1^{2r-1} u_i > 0$.

The probabilities used in Theorem 4.1 are

$$\alpha_r = \alpha_r(\delta_1, \gamma_1; -\delta_2, -\gamma_2; \delta_1, \gamma_1; \cdots; \delta_1, \gamma_1), \tag{4.17}$$

$$\beta_r = \beta_r(-\delta_2, -\gamma_2; \delta_1, \gamma_1; \cdots; \delta_1, \gamma_1), \tag{4.18}$$

for $\gamma_1 > 0$, $\gamma_2 < 0$, $\delta_1 \geqq 0$, $\delta_2 < \delta_1$.

Now consider the case $\delta_1 < 0$. Then the probability is 1 that the upper line is touched at least once. Hence, the probability that the upper line is touched first is simply 1 minus the probability that the lower line is touched before the upper. This latter probability corresponds to the first part of Theorem 4.1 with $(\gamma_1, \delta_1)$ and $(-\gamma_2, -\delta_2)$ interchanged.

Finally, the case $\delta_1 = \delta_2$ cannot be obtained directly from Lemma 4.1 but can be derived in a similar fashion. Suppose $\delta_1 = \delta_2 > 0$. Let $L_i$ be the line

$$y = \delta t + (-1)^{i+1} v_i, \qquad v_i > 0, \delta > 0.$$

Given that a path touches $L_{2r-2}$, the conditional probability of then touching $L_{2r-1}$ is $\exp\{-2\delta(v_{2r-2} + v_{2r-1})\}$ since the last line has slope $\delta$ and is $v_{2r-2} + v_{2r-1}$ above the next-to-last line. Given that a path touches $L_{2r-3}$ it must touch $L_{2r-2}$. Thus given that a path touches $L_{2r-3}$, the conditional probability of touching $L_{2r-2}$ and then $L_{2r-1}$ is $\exp\{-2\delta(v_{2r-2} + v_{2r-1})\}$ which is the conditional probability of touching the line $y = \delta t + (v_{2r-3} + v_{2r-2} + v_{2r-1})$. Thus

$$(4.19) \quad \alpha_r(\delta, v_1; -\delta, v_2; \delta, v_3; \cdots; \delta, v_{2r-1}) = \alpha_{r-1}(\delta, v_1; -\delta, v_2; \cdots; \delta, v_{2r-3} + v_{2r-2} + v_{2r-1}).$$

From this it follows that

$$(4.20) \quad \alpha_r(\delta, v_1; -\delta, v_2; \cdots; \delta, v_{2r-1}) = \alpha_1(\delta, v_1 + \cdots + v_{2r-1}) = e^{-2\delta \Sigma v_i}.$$

If $v_{2i-1} = \gamma_1$, $v_{2i} = -\gamma_2$, and $\delta = \delta_1$, we have

$$(4.21) \qquad \alpha_r = e^{-2\delta_1[r\gamma_1 - (r-1)\gamma_2]}.$$

Similarly

$$(4.22) \qquad \beta_r(-\delta, v_2; \delta, v_3; \cdots; \delta, v_{2r-1}) = e^{-2\delta \Sigma_2^{2r-1} v_i}$$

and

$$(4.23) \qquad \beta_{r-1} = e^{-2\delta_1((r-1)\gamma_1 - \gamma_2)}.$$

From these results, the third part of Theorem 4.1 follows.

4.3. *Conditional probability of going over one line first.* We now find the probability of touching one line first conditional on the path going through a certain point.

THEOREM 4.2. *If $Y(t)$ is the Wiener process with $\mathcal{E}Y(t) = 0$ $\mathcal{E}Y^2(t) = t$, and if $T, \gamma_1, \gamma_2, \delta_1, \delta_2$ are numbers such that $\gamma_1 > 0$, $\gamma_2 < 0$, $\gamma_1 + \delta_1 T \geqq \gamma_2 + \delta_2 T$, $T > 0$, the conditional probability that $Y(t) \geqq \gamma_1 + \delta_1 t$ for a smaller $t$ ($t \leqq T$) than any $t$ for which $Y(t) \leqq \gamma_2 + \delta_2 t$ given $Y(T) = y$ (not $\gamma_1 + \delta_1 T = \gamma_2 + \delta_2 T = y$) is*

$$
\begin{aligned}
P_1(T, y) &= \\
&\sum_{r=1}^{\infty} \{ e^{-(2/T)[r^2\gamma_1(\gamma_1+\delta_1 T-y)+(r-1)^2\gamma_2(\gamma_2+\delta_2 T-y)-r(r-1)\{\gamma_1(\gamma_2+\delta_2 T-y)+\gamma_2(\gamma_1+\delta_1 T-y)\}]} \\
&\quad - e^{-(2/T)[r^2\{\gamma_1(\gamma_1+\delta_1 T-y)+\gamma_2(\gamma_2+\delta_2 T-y)\}-r(r-1)\gamma_1(\gamma_2+\delta_2 T-y)-r(r+1)\gamma_2(\gamma_1+\delta_1 T-y)]} \}, \\
&\qquad\qquad \gamma_1 + \delta_1 T \geqq y, \\
&= 1 - \\
&\sum_{r=1}^{\infty} \{ e^{-(2/T)[(r-1)^2\gamma_1(\gamma_1+\delta_1 T-y)+r^2\gamma_2(\gamma_2+\delta_2 T-y)-r(r-1)\{\gamma_1(\gamma_2+\delta_2 T-y)+\gamma_2(\gamma_1+\delta_1 T-y)\}]} \\
&\quad - e^{-(2/T)[r^2\{\gamma_1(\gamma_1+\delta_1 T-y)+\gamma_2(\gamma_2+\delta_2 T-y)\}-r(r+1)\gamma_1(\gamma_2+\delta_2 T-y)-r(r-1)\gamma_2(\gamma_1+\delta_1 T-y)]} \}, \\
&\qquad\qquad \gamma_1 + \delta_1 T \leqq y, \\
&= \frac{e^{-(2/T)\gamma_2(\gamma_1+\delta_1 T-y)} - 1}{e^{(2/T)(\gamma_1-\gamma_2)(\gamma_1+\delta_1 T-y)} - 1}, \qquad \gamma_1 + \delta_1 T = \gamma_2 + \delta_2 T \neq y.
\end{aligned} \tag{4.24}
$$

PROOF. Let $Y(t \mid T, y)$ be the process $Y(t)$ given $Y(t) = y$. Then

$$\mathcal{E}Y(t \mid T, y) = \frac{t}{T} y, \qquad t \leqq T, \tag{4.25}$$

$$\operatorname{Var} Y(t \mid T, y) = t - \frac{t^2}{T} = t\left(1 - \frac{t}{T}\right), \qquad t \leqq T, \tag{4.26}$$

$$
\begin{aligned}
\operatorname{Cov}[Y(s \mid T, y), Y(t \mid T, y)] &= s - \frac{st}{T} \\
&= s\left(1 - \frac{t}{T}\right), \qquad s \leqq t \leqq T.
\end{aligned} \tag{4.27}
$$

Define a new process by

$$Z(u) = \frac{T+u}{T}\left[Y\left(\frac{Tu}{T+u} \,\middle|\, T, y\right) - \frac{u}{T+u} y\right], \qquad 0 \leqq u < \infty. \tag{4.28}$$

Then

$$\mathcal{E}Z(u) = 0, \tag{4.29}$$

$$\mathcal{E}Z^2(u) = u, \tag{4.30}$$

$$\mathcal{E}Z(v)Z(u) = v, \qquad v \leqq u. \tag{4.31}$$

Thus $Z(u)$ is the Wiener process. The event $Y(t \mid T, y) \geqq \gamma_1 + \delta_1 t$ is equivalent to

$$Z(u) \geqq \gamma_1 + \left(\frac{\gamma_1 - y}{T} + \delta_1\right) u.$$

We note that

$$\frac{\gamma_1 - y}{T} + \delta_1 \geqq \frac{\gamma_2 - y}{T} + \delta_2,$$

and

$$\frac{\gamma_1 - y}{T} + \delta_1 > 0$$

if $\gamma_1 + \delta_1 T > y$, that is, if $y$ is below the intersection of the upper line and $t = T$. Now Theorem 4.2 follows from Theorem 4.1.

4.4. *The probability of going over one line first in a fixed time.* Here we find $P_1(T)$, which is obtained from the following theorem.

THEOREM 4.3. *If $Y(t)$ is a Wiener process with $\mathcal{E}Y(t) = 0$ and $\mathcal{E}Y^2(t) = t$ and if $T, \gamma_1, \gamma_2, \delta_1, \delta_2$ are numbers such that $\gamma_1 > 0$, $\gamma_2 < 0$, $\gamma_1 + \delta_1 T \geqq \gamma_2 + \delta_2 T$, $T > 0$, the probability that $Y(t) \geqq \gamma_1 + \delta_1 t$ for a $t \leqq T$ which is smaller than any $t$ for which $Y(t) \leqq \gamma_2 + \delta_2 t$ is*

$$\begin{aligned} P_1(T) &= 1 - \Phi\left(\frac{\delta_1 T + \gamma_1}{\sqrt{T}}\right) \\ &+ \sum_{r=1}^{\infty} \left\{ e^{-2[r\gamma_1 - (r-1)\gamma_2][r\delta_1 - (r-1)\delta_2]} \Phi\left(\frac{\delta_1 T + 2(r-1)\gamma_2 - (2r-1)\gamma_1}{\sqrt{T}}\right) \right. \\ &- e^{-2[r^2(\gamma_1\delta_1 + \gamma_2\delta_2) - r(r-1)\gamma_1\delta_2 - r(r+1)\gamma_2\delta_1]} \Phi\left(\frac{\delta_1 T + 2r\gamma_2 - (2r-1)\gamma_1}{\sqrt{T}}\right) \\ &- e^{-2[(r-1)\gamma_1 - r\gamma_2][(r-1)\delta_1 - r\delta_2]} \left[1 - \Phi\left(\frac{\delta_1 T - 2r\gamma_2 + (2r-1)\gamma_1}{\sqrt{T}}\right)\right] \\ &+ e^{-2[r^2(\gamma_1\delta_1 + \gamma_2\delta_2) - r(r-1)\gamma_2\delta_1 - r(r+1)\gamma_1\delta_2]} \\ &\left. \cdot \left[1 - \Phi\left(\frac{\delta_1 T + (2r+1)\gamma_1 - 2r\gamma_2}{\sqrt{T}}\right)\right] \right\}, \end{aligned} \tag{4.32}$$

*where*

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-(u^2/2)} \, du. \tag{4.33}$$

PROOF. The density of $Y(t)$ at $t = T$ is $n(y \mid 0, T)$, the normal density with mean 0 and variance $T$. Then the probability of a path touching the upper line before the lower line for $t \leqq T$ is

$$\int_{-\infty}^{\infty} P_1(T, y) n(y \mid 0, T) \, dy. \tag{4.34}$$

The integration to $\gamma_1 + \delta_1 T$ is

$$\int_{-\infty}^{\gamma_1+\delta_1 T} \sum_{r=1}^{\infty} \{e^{-g_{1r}(y)} - e^{-g_{2r}(y)}\} n(y \mid 0, T)\, dy, \tag{4.35}$$

where

$$\begin{aligned} g_{1r}(y) &= \frac{2}{T}\,[y\{-r^2\gamma_1 - (r-1)^2\gamma_2 + r(r-1)\gamma_1 + r(r-1)\gamma_2\} \\ &\qquad + k_{1r}] \\ &= \frac{2}{T}\,[y\{-r\gamma_1 + (r-1)\gamma_2\} + k_{1r}], \end{aligned} \tag{4.36}$$

$$\begin{aligned} k_{1r} = r^2\gamma_1(\gamma_1 + \delta_1 T) + (r-1)^2\gamma_2(\gamma_2 + \delta_2 T) - r(r-1)\gamma_1(\gamma_2 + \delta_2 T) \\ - r(r-1)\gamma_2(\gamma_1 + \delta_1 T), \end{aligned} \tag{4.37}$$

$$\begin{aligned} g_{2r}(y) &= \frac{2}{T}\,[y\{-r^2\gamma_1 - r^2\gamma_2 + r(r-1)\gamma_1 + r(r+1)\gamma_2\} + k_{2r}] \\ &= \frac{2}{T}\,[(-r\gamma_1 + r\gamma_2)y + k_{2r}], \end{aligned} \tag{4.38}$$

$$\begin{aligned} k_{2r} = r^2\gamma_1(\gamma_1 + \delta_1 T) + r^2\gamma_2(\gamma_2 + \delta_2 T) - r(r-1)\gamma_1(\gamma_2 + \delta_2 T) \\ - r(r+1)\gamma_2(\gamma_1 + \delta_1 T). \end{aligned} \tag{4.39}$$

Since

$$e^{-g_{1r}(y)} \leqq e^{-(2/T)(r-1)[r\gamma_1-(r-1)\gamma_2][\gamma_1+\delta_1 T-(\gamma_2+\delta_2 T)]}, \tag{4.40}$$

$$e^{-g_{2r}(y)} \leqq e^{-(2/T)r\{[(r-1)\gamma_1-r\gamma_2][\gamma_1+\delta_1 T-(\gamma_2+\delta_2 T)]}\tag{4.41}$$

for $y \leqq \gamma_1 + \delta_1 T$ the series in $P_1(T, y)$ is bounded in absolute value by a series that converges when $\gamma_1 + \delta_1 T > \gamma_2 + \delta_2 T$ and hence the order of summation and integration in (4.35) can be reversed. Then

$$\begin{aligned} &\int_{-\infty}^{\gamma_1+\delta_1 T} e^{-g_{1r}(y)} n(y \mid 0, T)\, dy \\ &= \int_{-\infty}^{\gamma_1+\delta_1 T} \frac{1}{\sqrt{2\pi T}}\, e^{-[1/(2T)](y^2+4y\{-r\gamma_1+(r-1)\gamma_2\}+4k_{1r}}\, dy \\ &= e^{(2/T)[\{(r-1)\gamma_2-r\gamma_1\}^2-k_{1r}]} \int_{-\infty}^{\gamma_1+\delta_1 T} \frac{1}{\sqrt{2\pi T}}\, e^{-[1/(2T)](y+2\{(r-1)\gamma_2-r\gamma_1\})^2}\, dy \\ &= e^{-2[r\gamma_1-(r-1)\gamma_2][r\delta_1-(r-1)\delta_2]}\, \Phi\left(\frac{\delta_1 T + 2(r-1)\gamma_2 - (2r-1)\gamma_1}{\sqrt{T}}\right) \end{aligned} \tag{4.42}$$

Similarly

$$
\begin{aligned}
&\int_{-\infty}^{\gamma_1+\delta_1 T} e^{-g_{2r}(y)} n(y \mid 0, T)\, dy \\
(4.43)\qquad &= \int_{-\infty}^{\gamma_1+\delta_1 T} \frac{1}{\sqrt{2\pi T}} e^{-[1/(2T)](y^2+4yr(\gamma_2-\gamma_1)+4k_{2r})}\, dy \\
&= e^{(2/T)[r^2(\gamma_2-\gamma_1)^2-k_{2r}]} \int_{-\infty}^{\gamma_1+\delta_1 T} \frac{1}{\sqrt{2\pi T}} e^{-[1/(2T)](y+2r(\gamma_2-\gamma_1))^2} \\
&= e^{-2[r^2(\gamma_1\delta_1+\gamma_2\delta_2)-r(r-1)\gamma_1\delta_2-r(r+1)\gamma_2\delta_1]}\, \Phi\left(\frac{\delta_1 T + 2r\gamma_2 - (2r-1)\gamma_1}{\sqrt{T}}\right).
\end{aligned}
$$

The integration from $\gamma_1 + \delta_1 T$ to $\infty$ is

$$
\begin{aligned}
(4.44)\qquad &\int_{\gamma_1+\delta_1 T}^{\infty} \frac{1}{\sqrt{2\pi T}} e^{-[y^2/(2T)]}\, dy \\
&\qquad - \int_{\gamma_1+\delta_1 T}^{\infty} \sum_{r=1}^{\infty} \{e^{-g_{3r}(y)} - e^{-g_{4r}(y)}\} n(y \mid 0, T)\, dy,
\end{aligned}
$$

where

$$
\begin{aligned}
(4.45)\qquad g_{3r}(y) &= \frac{2}{T} [y\{-r^2\gamma_2 - (r-1)^2\gamma_1 + r(r-1)\gamma_1 + r(r-1)\gamma_2\} + k_{3r}] \\
&= \frac{2}{T} [y\{(r-1)\gamma_1 - r\gamma_2\} + k_{3r}],
\end{aligned}
$$

$$
\begin{aligned}
(4.46)\qquad k_{3r} &= r^2\gamma_2(\gamma_2 + \delta_2 T) + (r-1)^2\gamma_1(\gamma_1 + \delta_1 T) \\
&\qquad - r(r-1)\gamma_1(\gamma_2 + \delta_2 T) - r(r-1)\gamma_2(\gamma_1 + \delta_1 T),
\end{aligned}
$$

$$
\begin{aligned}
(4.47)\qquad g_{4r}(y) &= \frac{2}{T} [y\{-r^2\gamma_1 - r^2\gamma_2 + r(r-1)\gamma_2 + r(r+1)\gamma_1\} + k_{4r}] \\
&= \frac{2}{T} [y\{r\gamma_1 - r\gamma_2\} + k_{4r}],
\end{aligned}
$$

$$
\begin{aligned}
(4.48)\qquad k_{4r} &= r^2\gamma_1(\gamma_1 + \delta_1 T) + r^2\gamma_2(\gamma_2 + \delta_2 T) \\
&\qquad - r(r-1)\gamma_2(\gamma_1 + \delta_1 T) - r(r+1)\gamma_1(\gamma_2 + \delta_2 T).
\end{aligned}
$$

In this case

$$
(4.49)\qquad e^{-g_{3r}(y)} \leqq e^{-(2/T)r[(r-1)\gamma_1-r\gamma_2][\gamma_1+\delta_1 T-(\gamma_2+\delta_2 T)]},
$$

$$
(4.50)\qquad e^{-g_{4r}(y)} \leqq e^{-(2/T)r[(r+1)\gamma_1-r\gamma_2][\gamma_1+\delta_1 T-(\gamma_2+\delta_2 T)]}.
$$

Then

$$\int_{\gamma_1+\delta_1 T}^{\infty} e^{-g_{3r}(y)} n(y \mid 0, T)\, dy$$

(4.51)
$$= \int_{\gamma_1+\delta_1 T}^{\infty} \frac{1}{\sqrt{2\pi T}}\, e^{-[1/(2T)]\,(y^2+4y\{(r-1)\gamma_1-r\gamma_2\}+4k_{3r})}\, dy$$
$$= e^{(2/T)[\{(r-1)\gamma_1-r\gamma_2\}^2-k_{3r}]} \int_{\gamma_1+\delta_1 T}^{\infty} \frac{1}{\sqrt{2\pi T}}\, e^{-[1/(2T)]\,(y+2\{(r-1)\gamma_1-r\gamma_2\})^2}\, dy$$
$$= e^{-2[r\gamma_2-(r-1)\gamma_1][r\delta_2-(r-1)\delta_1]} \left[1 - \Phi\left(\frac{\delta_1 T - 2r\gamma_2 + (2r-1)\gamma_1}{\sqrt{T}}\right)\right].$$

Also

$$\int_{\gamma_1+\delta_1 T}^{\infty} e^{-g_{4r}(y)} n(y \mid 0, T)\, dy$$

(4.52)
$$= \int_{\gamma_1+\delta_1 T}^{\infty} \frac{1}{\sqrt{2\pi T}}\, e^{-[1/(2T)]\,(y^2+4y\{r\gamma_1-r\gamma_2\}+4k_{4r})}\, dy$$
$$= e^{-2[r^2(\gamma_1\delta_1+\gamma_2\delta_2)-r(r-1)\gamma_2\delta_1-r(r+1)\gamma_1\delta_2]}$$
$$\cdot \left[1 - \Phi\left(\frac{\delta_1 T + (2r+1)\gamma_1 - 2r\gamma_2}{\sqrt{T}}\right)\right].$$

Then $P_1(T)$ follows for $\gamma_1 + \delta_1 T > \gamma_2 + \delta_2 T$. In case $\gamma_1 + \delta_1 T = \gamma_2 + \delta_2 T$ we argue that $P_1(T) = \lim P_1(t)$, $t \to T$ and $t \leq T$ since $P_1(t) < P_1(T)$, $t < T$, and $P_1(T) - P_1(t)$ is less than the probability that $\gamma_2 + \delta_2 t < Y(t) < \gamma_1 + \delta_1 t$ which converges to 0. It will be seen in the discussion following Corollary 4.1 that when $\delta_1 T + \gamma_1 = \delta_2 T + \gamma_2$ the series (4.54) converges as $1/(ar^2 + br + c)$. Hence, for $t \leq T$, the series (4.54) for $P_1(t)$ can be majorized (uniformly in $t$) by a series $k/r^2$, which converges. This proves the theorem.

The probability of touching the lower line first and before $T = t$, say $P_2(T)$, is obtained from Theorem 4.3 by replacing $(\gamma_1, \delta_1)$ by $(-\gamma_2, -\delta_2)$.

The probability (4.32) can be written in different ways. One which is convenient for computing is to use Mill's Ratio

(4.53)
$$R(x) = \frac{1 - \Phi(x)}{\phi(x)},$$

where $\phi(x) = n(x \mid 0, 1)$.

COROLLARY 4.1.

$$
\begin{aligned}
P_1(T) &= 1 - \Phi\left(\frac{\delta_1 T + \delta_1}{\sqrt{T}}\right) + \phi\left(\frac{\delta_1 T + \delta_1}{\sqrt{T}}\right)\sum_{r=1}^{\infty} \\
&\quad \Bigg\{ e^{-[2(r-1)/T][r\gamma_1-(r-1)\gamma_2][\delta_1 T+\gamma_1-(\delta_2 T+\gamma_2)]} R\left[\frac{2(r\gamma_1-(r-1)\gamma_2)-(\delta_1 T+\gamma_1)}{\sqrt{T}}\right] \\
&\quad - e^{-(2r/T)[(r-1)\gamma_1-r\gamma_2][\delta_1 T+\gamma_1-(\delta_2 T+\gamma_2)]} R\left[\frac{2r(\gamma_1-\gamma_2)-(\delta_1 T+\gamma_1)}{\sqrt{T}}\right] \\
&\quad - e^{-(2r/T)[(r-1)\gamma_1-r\gamma_2][\delta_1 T+\gamma_1-(\delta_2 T+\gamma_2)]} R\left[\frac{2((r-1)\gamma_1-r\gamma_2)+(\delta_1 T+\gamma_1)}{\sqrt{T}}\right] \\
&\quad + e^{-(2r/T)[(r+1)\gamma_1-r\gamma_2][\delta_1 T+\gamma_1-(\delta_2 T+\gamma_2)]} R\left[\frac{2r(\gamma_1-\gamma_2)+(\delta_1 T+\gamma_1)}{\sqrt{T}}\right]\Bigg\} \qquad (4.54) \\
&= \phi\left(\frac{\delta_1 T+\gamma_1}{\sqrt{T}}\right)\sum_{r=0}^{\infty}\Bigg\{ e^{-(2r/T)[(r+1)\gamma_1-r\gamma_2][\delta_1 T+\gamma_1-(\delta_2 T+\gamma_2)]} \\
&\quad \cdot\left[R\left(\frac{2((r+1)\gamma_1-r\gamma_2)-(\delta_1 T+\gamma_1)}{\sqrt{T}}\right) + R\left(\frac{2r(\gamma_1-\gamma_2)+(\delta_1 T+\gamma_1)}{\sqrt{T}}\right)\right] \\
&\quad - e^{-[2(r+1)/T][r\gamma_1-(r+1)\gamma_2][\delta_1 T+\gamma_1-(\delta_2 T+\gamma_2)]} \\
&\quad \left[R\left(\frac{2(r+1)(\gamma_1-\gamma_2)-(\delta_1 T+\gamma_1)}{\sqrt{T}}\right)\right. \\
&\quad \left. + R\left(\frac{2(r\gamma_1-(r+1)\gamma_2)+(\delta_1 T+\gamma_1)}{\sqrt{T}}\right)\right]\Bigg\}.
\end{aligned}
$$

Mill's Ratio $R(x)$ for $x > 0$ satisfies the inequalities $R(x) < 1/x$,

$$
\begin{aligned}
&\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \frac{5\cdot 3}{x^7} + \cdots - \frac{(4k-3)\cdots 1}{x^{4k-1}} < R(x) \\
&\qquad < \frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \cdots + \frac{(4k-1)\cdots 1}{x^{4k+1}}, \quad k = 1, 2, \cdots.
\end{aligned} \qquad (4.55)
$$

Thus for large $x$ $R(x)$ behaves like $1/x$ (with an error of less than $1/x^3$). If $\delta_1 T + \gamma_1 = \delta_2 T + \gamma_2$, then for $r$ large, the $r$th term of (4.54) is approximately

$$\frac{\sqrt{T}}{(2r+1)\gamma_1 - 2r\gamma_2 - \delta_1 T} + \frac{\sqrt{T}}{(2r+1)\gamma_1 - 2r\gamma_2 + \delta_1 T}$$

$$(4.56) \quad - \left[\frac{\sqrt{T}}{(2r+1)\gamma_1 - 2(r+1)\gamma_2 - \delta_1 T} + \frac{\sqrt{T}}{(2r+1)\gamma_1 - 2(r+1)\gamma_2 + \delta_1 T}\right]$$

$$= 2\sqrt{T}\left\{\frac{(2r+1)\gamma_1 - 2r\gamma_2}{[(2r+1)\gamma_1 - 2r\gamma_2]^2 - \delta_1^2 T^2} - \frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2}{[(2r+1)\gamma_1 - 2(r+1)\gamma_2]^2 - \delta_1^2 T^2}\right\}$$

which is of the order $1/r^2$.

To express the probability of $X(t) = Y(t) + \mu t$ touching $x = c_1 + d_1 t$ before $x = c_2 + d_2 t$ we replace $\gamma_1$ by $c_1$, $\gamma_2$ by $c_2$, $\delta_1$ by $d_1 - \mu$, and $\delta_2$ by $d_2 - \mu$ in the preceding formulas. If the sequential procedure is symmetric, $c_1 = -c_2 = c$, say, and $d_1 = -d_2 = d$, say, and the formulas are simplified.

COROLLARY 4.2. *If $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$, and $\delta_2 = -d - \mu$, then the probability of touching the upper line first before $t = T$ is*

$$(4.57) \quad \begin{aligned} P_1(T) = 1 - \Phi\left(\frac{(d-\mu)T + c}{\sqrt{T}}\right) + \sum_{r=1}^{\infty}\Bigg\{ & e^{-2(2r-1)c[(2r-1)d-\mu]} \\ & \cdot \Phi\left(\frac{(d-\mu)T - (4r-3)c}{\sqrt{T}}\right) - e^{-2\cdot 2rc[2rd-\mu]}\Phi\left(\frac{(d-\mu)T - (4r-1)c}{\sqrt{T}}\right) \\ & - e^{-2(2r-1)c[(2r-1)d+\mu]}\left[1 - \Phi\left(\frac{(d-\mu)T + (4r-1)c}{\sqrt{T}}\right)\right] \\ & + e^{-2\cdot 2rc[2rd+\mu]}\left[1 - \Phi\left(\frac{(d-\mu)T + (4r+1)c}{\sqrt{T}}\right)\right]\Bigg\} \\ = 1 - \Phi\left(\frac{(d-\mu)T + c}{\sqrt{T}}\right) & \\ + \sum_{s=1}^{\infty}(-1)^{s+1}\Bigg\{ & e^{-2sc(sd-\mu)}\Phi\left(\frac{(d-\mu)T - (2s-1)c}{\sqrt{T}}\right) \\ & - e^{-2sc(sd+\mu)}\left[1 - \Phi\left(\frac{(d-\mu)T + (2s+1)c}{\sqrt{T}}\right)\right]\Bigg\}. \end{aligned}$$

*The probability of touching the lower line first before $t = T$ is*

$$(4.58) \quad \begin{aligned} P_2(T) = 1 - \Phi\left(\frac{(d+\mu)T + c}{\sqrt{T}}\right) & \\ + \sum_{s=1}^{\infty}(-1)^{s+1}\Bigg\{ & e^{-2sc(sd+\mu)}\Phi\left(\frac{(d+\mu)T - (2s-1)c}{\sqrt{T}}\right) \\ & - e^{-2sc(sd-\mu)}\left[1 - \Phi\left(\frac{(d+\mu)T + (2s+1)c}{\sqrt{T}}\right)\right]\Bigg\}. \end{aligned}$$

*The probability of touching one or both lines before $t = T$ is*

$$P_1(T) + P_2(T) = 1 - \left[\Phi\left(\frac{\mu T + dT + c}{\sqrt{T}}\right) - \Phi\left(\frac{\mu T - (dT + c)}{\sqrt{T}}\right)\right]$$

$$+ \sum_{j=1}^{\infty} (-1)^{s+1} \left\{ e^{-2sc(sd-\mu)} \left[\Phi\left(\frac{\mu T + 2sc + dT + c}{\sqrt{T}}\right)\right.\right.$$

$$\left. - \Phi\left(\frac{\mu T + 2sc - (dT + c)}{\sqrt{T}}\right)\right]$$

$$(4.59) \qquad + e^{-2sc(sd+\mu)} \left[\Phi\left(\frac{-\mu T + 2sc + dT + c}{\sqrt{T}}\right)\right.$$

$$\left.\left. - \Phi\left(\frac{-\mu T + 2sc - (dT + c)}{\sqrt{T}}\right)\right]\right\}$$

$$= 1 + \sum_{s=-\infty}^{\infty} (-1)^{s+1} e^{-2s^2cd+2sc\mu} \left[\Phi\left(\frac{(\mu + d)T + (2s + 1)c}{\sqrt{T}}\right)\right.$$

$$\left. - \Phi\left(\frac{(\mu - d)T + (2s - 1)c}{\sqrt{T}}\right)\right].$$

These results can also be expressed in terms of Mill's Ratio.

COROLLARY 4.3. *If $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$ and $\delta_2 = -d - \mu$, then the probability of touching the upper line first before $t = T$ is*

$$P_1(T) = \Phi\left(\frac{\mu T - (dT + c)}{\sqrt{T}}\right) + \phi\left(\frac{\mu T - (dT + c)}{\sqrt{T}}\right)$$

$$\cdot \left\{ \sum_{s=1}^{\infty} (-1)^{s+1} e^{-2(c/T)(dT+c)s(s-1)} R\left(\frac{2sc + \mu T - (dT + c)}{\sqrt{T}}\right)\right.$$

$$(4.60) \qquad \left. - \sum_{s=1}^{\infty} (-1)^{s+1} e^{-2(c/T)(dT+c)s(s+1)} R\left(\frac{2sc - \mu T + (dT + c)}{\sqrt{T}}\right)\right\}$$

$$= \phi\left(\frac{\mu T - (dT + c)}{\sqrt{T}}\right) \sum_{s=0}^{\infty} (-1)^s e^{-2(c/T)(dT+c)s(s+1)}$$

$$\cdot \left[R\left(\frac{2sc - (d - \mu)T + c}{\sqrt{T}}\right) + R\left(\frac{2sc + (d - \mu)T + c}{\sqrt{T}}\right)\right].$$

*The probability of touching the lower line first before $t = T$ is*

$$
\begin{aligned}
P_2(T) &= \Phi\left(\frac{-\mu T - (dT + c)}{\sqrt{T}}\right) + \phi\left(\frac{\mu T + (dT + c)}{\sqrt{T}}\right) \\
&\cdot \left\{\sum_{s=1}^{\infty} (-1)^{s+1} e^{-2(c/T)(dT+c)s(s-1)} R\left(\frac{2sc - \mu T - (dT + c)}{\sqrt{T}}\right)\right. \\
&\left. - \sum_{s=1}^{\infty} (-1)^{s+1} e^{-2(c/T)(dT+c)s(s+1)} R\left(\frac{2sc + \mu T + (dT + c)}{\sqrt{T}}\right)\right\} \\
&= \phi\left(\frac{\mu T + dT + c}{\sqrt{T}}\right) \sum_{s=0}^{\infty} (-1)^s e^{-2(c/T)(dT+c)s(s+1)} \\
&\cdot \left[R\left(\frac{2sc - (d + \mu)T + c}{\sqrt{T}}\right) + R\left(\frac{2sc + (d + \mu)T + c}{\sqrt{T}}\right)\right].
\end{aligned} \tag{4.61}
$$

The expressions in Corollary 4.3 were used for computation. If $c + dT > 0$ $(\gamma_1 + \delta_1 T > \gamma_2 + \delta_2 T)$, the exponential term in the sum

$$\exp[-2(c/T)(c + dT)s(s + 1)]$$

decreases rapidly and the series converges rapidly. It is an alternating series, and the last term used bounds the error. For large $x$ $R(x)$ behaves like $1/x$. If $dT + c = 0$, the convergence of the series for $P_1(T)$ and $P_2(T)$ is like

$$\sum (-1)^s/[(2s + 1)c \pm (d \pm \mu)T].$$

When $\gamma_1 + \delta_1 T = \gamma_2 + \delta_2 T$, we can use the third part of (4.24). In particular if $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$, $\delta_2 = -d - \mu$,

$$P_1(T, y) = \frac{e^{-2c(\mu T+y)/T} - 1}{e^{-4c(\mu T+y)/T} - 1} = \frac{1}{e^{-2c(\mu T+y)/T} + 1}. \tag{4.62}$$

COROLLARY 4.4. *If* $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$, $\delta_2 = -d - \mu$ *and* $c + dT = 0$, *then*

$$
\begin{aligned}
P_1(T) &= \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{\infty} \frac{e^{-[y^2/(2T)]}}{1 + e^{-(2c/T)y - 2\mu c}} \, dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(z^2/2)}}{1 + e^{-(2c/\sqrt{T})z - 2\mu c}} \, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-[(z-\mu\sqrt{T})^2/2]}}{1 + e^{-(2c/\sqrt{T})z}} \, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-\mu\sqrt{T})^2/2} \frac{e^{(2c/\sqrt{T})z}}{1 + e^{(2c/\sqrt{T})z}} \, dz.
\end{aligned} \tag{4.63}
$$

The expressions in Corollary 4.4 are not very useful since the integration cannot be evaluated in closed form. We can give an approximation which may be useful for some purposes.

COROLLARY 4.5. *If* $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$, $\delta_2 = -d - \mu$ *and* $c + dT = 0$, *then* $P_1(T)$ *is approximately*

$$\Phi\left(\frac{\mu\sqrt{T}}{\sqrt{\frac{\beta^2 T}{4c^2} + 1}}\right), \tag{4.64}$$

*where* $\beta = 1.702$.

PROOF. Haley [6] has shown that

$$\left|\Phi(x) - \frac{1}{1 + e^{-\beta x}}\right| < .01 \tag{4.65}$$

From Corollary 4.4 we have that $P_1(T)$ is approximately

$$\begin{aligned} &\int_{-\infty}^{\infty} \Phi\left[\frac{2c}{\beta T}(y + \mu T)\right] \frac{1}{\sqrt{2\pi T}} e^{-[y^2/(2T)]}\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{[2c/(\beta T)](y+\mu T)} \frac{1}{2\pi\sqrt{T}} e^{-\frac{1}{2}[u^2 + (y^2/T)]}\, du\, dy \\ &= \Pr\left\{U \leqq \frac{2c}{\beta T} Y + \frac{2c\mu}{\beta}\right\} = \Pr\left\{U - \frac{2c}{\beta T} Y \leqq \frac{2c\mu}{\beta}\right\}. \end{aligned} \tag{4.66}$$

But $U$ and $Y$ are independently normally distributed and $U - [2c/(\beta T)]Y = Z$ is normally distributed with mean zero and variance $1 + [4c^2/(\beta^2 T)]$. The above is

$$\Pr\left\{Z^* \leqq \frac{2c\mu/\beta}{\sqrt{1 + \frac{4c^2}{\beta^2 T}}}\right\} = \Pr\left\{Z^* \leqq \frac{\mu\sqrt{T}}{\sqrt{\frac{\beta^2 T}{4c^2} + 1}}\right\}, \tag{4.67}$$

where $Z^*$ is normally distributed with mean 0 and variance 1.

As an example of the accuracy of the approximation, let $\mu = .16449$, $T = 169$, and $c = 13.2$. Then $P_1(T) = .9453$. The argument in $\Phi$ here is 1.639 and the probability is .9495. The error is .004.

4.5. *The probability of accepting a hypothesis.* The sequential procedure is to accept the hypothesis that the mean is large if either the path touches the upper line before $t = T$ or it stays between the two lines to $t = T$ and at $t = T$ is above some value $k$. We now proceed to find expressions for this probability.

THEOREM 4.4. *If* $Y(t)$ *is the Wiener process with* $\mathcal{E}Y(t) = 0$ *and* $\mathcal{E}Y^2(t) = t$ *and if* $\gamma_1, \gamma_2, \delta_1, \delta_2, T$, *and* $\theta$ *are numbers such that* $\gamma_1 > 0$, $\gamma_2 < 0$, $T > 0$, $\gamma_1 + \delta_1 T \geqq \theta \geqq \gamma_2 + \delta_2 T$, *the probability of either* $Y(t) \geqq \gamma_1 + \delta_1 t$ *for a* $t(\leqq T)$ *smaller than any* $t$ *for which* $Y(t) \leqq \gamma_2 + \delta_2 t$ *or* $\gamma_2 + \delta_2 t < Y(t) < \gamma_1 + \delta_1 t$, $0 \leqq t \leqq T$, *and* $Y(T) > \theta$ *is*

$$A(T,\theta) = 1 - \Phi\left(\frac{\theta}{\sqrt{T}}\right)$$

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

$$(4.68) \qquad - e^{-2[r^2(\gamma_1\delta_1+\gamma_2\delta_2)-r(r-1)\gamma_1\delta_2-r(r+1)\gamma_2\delta_1]}\Phi\left(\frac{\theta + 2r(\gamma_2 - \gamma_1)}{\sqrt{T}}\right)$$

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

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

Proof. We have

$$(4.69) \qquad A(T,\theta) = \int_{-\infty}^{\theta} P_1(T,y)\frac{1}{\sqrt{2\pi T}}e^{-(y^2/2T)}\,dy + \int_{\theta}^{\infty}[1 - P_2(T,y)]\frac{1}{\sqrt{2\pi T}}e^{-(y^2/2T)}\,dy;$$

that is, if $Y(T) \leqq \theta$, the event could happen only by touching the upper line first and if $Y(T) \geqq \theta$ the event could only fail to happen by touching the lower line first. These two probabilities are evaluated as for Theorem 4.3. We find

$$\int_{-\infty}^{\theta} P_1(T,y)\frac{1}{\sqrt{2\pi T}}e^{-[y^2/(2T)]}\,dy$$

$$(4.70) \qquad = \sum_{r=1}^{\infty}\left\{e^{-2[r\gamma_1-(r-1)\gamma_2][r\delta_1-(r-1)\delta_2]}\Phi\left(\frac{\theta + 2[(r-1)\gamma_2 - r\gamma_1]}{\sqrt{T}}\right)\right.$$

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

$$\int_{\theta}^{\infty} P_2(T,y)\frac{1}{\sqrt{2\pi T}}e^{-[y^2/(2T)]}\,dy$$

$$(4.71) \qquad = \sum_{r=1}^{\infty}\left\{e^{-2[r\gamma_2-(r-1)\gamma_1][r\delta_2-(r-1)\delta_1]}\Phi\left(\frac{-\theta + 2[r\gamma_2 - (r-1)\gamma_1]}{\sqrt{T}}\right)\right.$$

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

The result of Theorem 4.4 can also be expressed using Mill's Ratio.

Corollary 4.6. *The probability of the process touching the upper line before* *the lower for $t \leqq T$ or staying between the lines to $t = T$ with $Y(T) > \theta$ is*

$$
\begin{aligned}
A(T, \theta) = 1 - \Phi\left(\frac{\theta}{\sqrt{T}}\right) &+ \phi\left(\frac{\theta}{\sqrt{T}}\right) \sum_{r=1}^{\infty} \\
\cdot \Big\{ & e^{-(2/T)[r\gamma_1 - (r-1)\gamma_2][r(\delta_1 T+\gamma_1)-(r-1)(\delta_2 T+\gamma_2)-\theta]} R\left(\frac{2[r\gamma_1 - (r-1)\gamma_2] - \theta}{\sqrt{T}}\right) \\
& - e^{-(2/T)r\{(\gamma_1-\gamma_2)[(\delta_1 T+\gamma_1)-(\delta_2 T+\gamma_2)]+\gamma_1(\delta_2 T+\gamma_2-\theta)-\gamma_2(\delta_1 T+\gamma_1-\theta)\}} \\
& \qquad \cdot R\left(\frac{2r(\gamma_1 - \gamma_2) - \theta}{\sqrt{T}}\right) \\
& - e^{-(2/T)[(r-1)\gamma_1 - r\gamma_2][(r-1)(\delta_1 T+\gamma_1)-r(\delta_2 T+\gamma_2)+\theta]} R\left(\frac{2[(r-1)\gamma_1 - r\gamma_2] + \theta}{\sqrt{T}}\right) \\
& + e^{-(2/T)r\{r(\gamma_1-\gamma_2)[(\delta_1 T+\gamma_1)-(\delta_2 T+\gamma_2)]-\gamma_1(\delta_2 T+\gamma_2-\theta)+\gamma_2(\delta_1 T+\gamma_1-\theta)\}} \\
& \qquad \cdot R\left(\frac{2r(\gamma_1 - \gamma_2) + \theta}{\sqrt{T}}\right)\Big\}.
\end{aligned} \tag{4.72}
$$

The sequential procedure involves $X(t) = Y(t) + \mu t$, the lines $x = c_1 + d_1 t$ and $x = c_2 + d_2 t$ and a value $k$ at $t = T$ ($c_2 + d_2 T \leqq k \leqq c_1 + d_1 T$). The probability of accepting the hypothesis that $\mu$ is small ($\mu = -\mu^*$) is given by the complement of the above probability when $\gamma_1 = c_1$, $\gamma_2 = c_2$, $\delta_1 = d_1 - \mu$, $\delta_2 = d_2 - \mu$, and $\theta = k - \mu T$; then the operating characteristic is $L(\mu) = 1 - A(T, k - \mu T)$.

A case of particular interest is when the sequential procedure is symmetric. Then $k = 0$ (as well as $c_1 = -c_2 = c$, say, and $d_1 = -d_2 = d$, say).

COROLLARY 4.7. *If $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$, $\delta_2 = -d - \mu$, $\theta = -\mu T$, then*

$$
\begin{aligned}
A(T, -\mu T) &= 1 - \Phi(-\mu\sqrt{T}) \\
&\quad + \sum_{r=1}^{\infty} \left\{ e^{-2(2r-1)c[(2r-1)d-\mu]} \Phi\left(\frac{-\mu T - 2(2r-1)c}{\sqrt{T}}\right) \right. \\
&\quad - e^{-2\cdot 2rc[2rd-\mu]} \Phi\left(\frac{-\mu T - 2\cdot 2rc}{\sqrt{T}}\right) \\
&\quad - e^{-2(2r-1)c[(2r-1)d+\mu]} \Phi\left(\frac{\mu T - 2(2r-1)c}{\sqrt{T}}\right) \\
&\quad \left. + e^{-2\cdot 2rc[2rd+\mu]} \Phi\left(\frac{\mu T - 2\cdot 2rc}{\sqrt{T}}\right) \right\} \\
&= \Phi(\mu\sqrt{T}) + \sum_{s=1}^{\infty} (-1)^{s+1} \left\{ e^{-2sc[sd-\mu]} \Phi\left(\frac{-\mu T - 2sc}{\sqrt{T}}\right) \right. \\
&\qquad \left. - e^{-2sc[sd+\mu]} \Phi\left(\frac{\mu T - 2sc}{\sqrt{T}}\right) \right\}.
\end{aligned} \tag{4.73}
$$

In terms of Mill's Ratio, we obtain the following:

COROLLARY 4.8. *If* $\gamma_1 = -\gamma_2 = c, \delta_1 = d - \mu, \delta_2 = -d-\mu$, *and* $\theta = -\mu T$,

$$A(T, -\mu T) = \Phi(\mu\sqrt{T}) + \phi(\mu\sqrt{T}) \sum_{s=1}^{\infty} (-1)^{s+1} \cdot e^{-2s^2(c/T)(c+dT)} \left[ R\left(\frac{2sc + \mu T}{\sqrt{T}}\right) - R\left(\frac{2sc - \mu T}{\sqrt{T}}\right) \right]. \tag{4.74}$$

As in the formula for $P_1(T)$ the convergence is rapid when $c + dT > 0$. If $c + dT = 0$ the convergence is of the order

$$\frac{\sqrt{T}}{2sc + \mu T} - \frac{\sqrt{T}}{2sc - \mu T} = -\frac{2\mu T\sqrt{T}}{4s^2c^2 - \mu^2T^2}. \tag{4.75}$$

The terms are paired differently here from Corollary 4.3.

**5. Expected time to decision.** The time to decision, say $\tau$, is a random variable. If a path touches either line at a time less than $T$, observation stops and $\tau$ is this time; if neither line is touched at a time less than $T$, observation is stopped at $T$. Thus the probability distribution of the time of observation is

$$\begin{aligned} \Pr\{\tau \leqq t\} &= P_1(t) + P_2(t), && 0 \leqq t < T, \\ &= 1, && T \leqq t. \end{aligned} \tag{5.1}$$

The expressions for $P_1(t)$ and $P_2(t)$ given in Section 4 are valid for $\gamma_1 + \delta_1 t < \gamma_2 + \delta_2 t$. If $\gamma_1 + \delta_1 T < \gamma_2 + \delta_2 T$, there is a positive probability that $\tau = T$, namely,

$$\Pr\{\tau = T\} = 1 - [P_1(T) + P_2(T)] = P_0(T). \tag{5.2}$$

For $t < T$, there is a density, namely $d[P_1(t) + P_2(t)]/dt$.

THEOREM 5.1. *If* $\gamma_1 + \delta_1 t > \gamma_2 + \delta_2 t$ $(t < T)$, *the density of the time of observation is the sum of*

$$\begin{aligned} \frac{dP_1(t)}{dt} = \frac{1}{t^{3/2}} \phi\left(\frac{\delta_1 t + \gamma_1}{\sqrt{t}}\right) \sum_{r=0}^{\infty} &\cdot \{e^{-(2r/t)[(r+1)\gamma_1 - r\gamma_2][\delta_1 t + \gamma_1 - (\delta_2 t + \gamma_2)]}[(2r+1)\gamma_1 - 2r\gamma_2] \\ &- e^{-[2(r+1)/t][r\gamma_1 - (r+1)\gamma_2][\delta_1 t + \gamma_1 - (\delta_2 t + \gamma_2)]}[(2r+1)\gamma_1 - 2(r+1)\gamma_2]\} \end{aligned} \tag{5.3}$$

*and*

$$\begin{aligned} \frac{dP_2(t)}{dt} = \frac{1}{t^{3/2}} \phi\left(\frac{\delta_2 t + \gamma_2}{\sqrt{t}}\right) \sum_{r=0}^{\infty} &\cdot \{e^{-(2r/t)[r\gamma_1 - (r+1)\gamma_2][\delta_1 t + \gamma_1 - (\delta_2 t + \gamma_2)]}[2r\gamma_1 - (2r+1)\gamma_2] \\ &- e^{-[2(r+1)/t][(r+1)\gamma_1 - r\gamma_2][\delta_1 t + \gamma_1 - (\delta_2 t + \gamma_2)]}[2(r+1)\gamma_1 - (2r+1)\gamma_2]\}. \end{aligned} \tag{5.4}$$

PROOF. It is convenient to write $P_1(T)$ as ($r$ replacing $r+1$ in these terms)

$$
\begin{aligned}
P_1(T) = \sum_{r=0}^{\infty}\Big\{ & e^{-2[(r+1)\gamma_1 - r\gamma_2][(r+1)\delta_1 - r\delta_2]}\Phi\left(\frac{\delta_1 T + 2r\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right) \\
& + e^{-2[r^2\gamma_1\delta_1 + r^2\gamma_2\delta_2 - r(r+1)\gamma_1\delta_2 - r(r-1)\gamma_2\delta_1]}\Phi\left(\frac{-\delta_1 T + 2r\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right) \\
& - e^{-2[(r+1)^2\gamma_1\delta_1 + (r+1)^2\gamma_2\delta_2 - r(r+1)\gamma_1\delta_2 - (r+1)(r+2)\gamma_2\delta_1]} \\
& \cdot\Phi\left(\frac{\delta_1 T + 2(r+1)\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right) \\
& - e^{-2[r\gamma_1 - (r+1)\gamma_2][r\delta_1 - (r+1)\delta_2]}\Phi\left(\frac{-\delta_1 T + 2(r+1)\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right)\Big\}.
\end{aligned} \tag{5.5}
$$

Then

$$
\begin{aligned}
\frac{dP_1(T)}{dT} = \sum_{r=0}^{\infty}\Big\{ & e^{-2[(r+1)\gamma_1 - r\gamma_2][(r+1)\delta_1 - r\delta_2]}\phi\left(\frac{\delta_1 T + 2r\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right) \\
& \cdot\frac{\delta_1 T + (2r+1)\gamma_1 - 2r\gamma_2}{2T^{3/2}} + e^{-2[r^2\gamma_1\delta_1 + r^2\gamma_2\delta_2 - r(r+1)\gamma_1\delta_2 - r(r-1)\gamma_2\delta_1]} \\
& \cdot\phi\left(\frac{-\delta_1 T + 2r\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right)\frac{-\delta_1 T + (2r+1)\gamma_1 - 2r\gamma_2}{2T^{3/2}} \\
& - e^{-2[(r+1)^2\gamma_1\delta_1 + (r+1)^2\gamma_2\delta_2 - r(r+1)\gamma_1\delta_2 - (r+1)(r+2)\gamma_2\delta_1]} \\
& \phi\left(\frac{\delta_1 T + 2(r+1)\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right)\frac{\delta_1 T + (2r+1)\gamma_1 - 2(r+1)\gamma_2}{2T^{3/2}} \\
& - e^{-2[r\gamma_1 - (r+1)\gamma_2][r\delta_1 - (r+1)\delta_2]}\phi\left(\frac{-\delta_1 T + 2(r+1)\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right) \\
& \cdot\frac{-\delta_1 T + (2r+1)\gamma_1 - 2(r+1)\gamma_2}{2T^{3/2}}\Big\},
\end{aligned} \tag{5.6}
$$

which leads to (5.3). By interchanging $(\gamma_1, \delta_1)$ with $(-\gamma_2, -\delta_2)$, we obtain (5.4).

A characteristic of the procedure is the expected length of time of observation

$$
\mathcal{E}_T = \int_0^T t\,\frac{dP_1(t)}{dt}\,dt + \int_0^T t\,\frac{dP_2(t)}{dt}\,dt + TP_0(T). \tag{5.7}
$$

This section will be devoted to evaluating these expressions.

THEOREM 5.2.

$$\mathcal{E}_1^* = \int_0^T t \frac{dP_1(t)}{dt} dt$$

(5.8)
$$= \frac{1}{\delta_1} \sum_{r=0}^{\infty} \Bigg\{ \Bigg[ e^{-2[(r+1)\gamma_1 - r\gamma_2][(r+1)\delta_1 - r\delta_2]} \Phi\left(\frac{\delta_1 T + 2r\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right)$$
$$- e^{-2[r^2\gamma_1\delta_1 + r^2\gamma_2\delta_2 - r(r+1)\gamma_1\delta_2 - r(r-1)\gamma_2\delta_1]} \Phi\left(\frac{-\delta_1 T + 2r\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right)\Bigg]$$
$$[(2r+1)\gamma_1 - 2r\gamma_2] - \Bigg[ e^{-2[(r+1)^2\gamma_1\delta_1 + (r+1)^2\gamma_2\delta_2 - r(r+1)\gamma_1\delta_2 - (r+1)(r+2)\gamma_2\delta_1]}$$
$$\cdot \Phi\left(\frac{\delta_1 T + 2(r+1)\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right) - e^{-2[r\gamma_1 - (r+1)\gamma_2][r\delta_1 - (r+1)\delta_2]}$$
$$\cdot \Phi\left(\frac{-\delta_1 T + 2(r+1)\gamma_2 - (2r+1)\gamma_1}{\sqrt{T}}\right)\Bigg] [(2r+1)\gamma_1 - 2(r+1)\gamma_2] \Bigg\}.$$

*When* $(\gamma_1, \delta_1)$ *and* $(-\gamma_2, -\delta_2)$ *are interchanged,* (5.8) *is* $\mathcal{E}_2^* = \int_0^T t[dP_2(t)/dt]\, dt$.

PROOF. The derivative of the right hand side of (5.8) with respect to $T$ is identical to the derivative of the left hand side of (5.8) as obtained from Theorem 5.1. The theorem follows from the observation that the right hand side of (5.8) is 0 for $T = 0$. (Each term is 0 because each argument of $\Phi(\quad)$ goes to $-\infty$.)

In Corollary 5.1 below the series is written in another way. It will be seen then that when $\delta_1 T + \gamma_1 > \delta_2 T + \gamma_2$, the exponential terms insure rapid convergence of the series (and justify integration and differentiation term by term). When $\delta_1 T + \gamma_1 = \delta_2 T + \gamma_2$, the convergence in (5.10) (and hence (5.8)) is like $1/r^2$ and in (5.11) is like $1/r^4$. Thus the series can be majorized by a convergent series uniformly in $T$; $\mathcal{E}_i^*(t) \to \mathcal{E}_i^*(T)$, $t \leqq T$, and the convergence is term by term.

It might be noted that the proof of Theorem 5.2 is essentially a verification and depends on the way the terms are paired. The theorem could be proved directly from the following lemma, which in turn can be verified in a similar manner or can be developed by transformation of the integral:

LEMMA 5.1.

(5.9)
$$\int_0^T t \frac{\partial \Phi}{\partial t}\left(\frac{At + B}{\sqrt{T}}\right) dt = \frac{2AB - 1}{2A^2}\left[1 - \Phi\left(\frac{AT + B}{\sqrt{T}}\right)\right]$$
$$+ \frac{1}{2A^2} e^{-2AB} \Phi\left(\frac{AT - B}{\sqrt{T}}\right) - \frac{\sqrt{T}}{A} \frac{1}{\sqrt{2\pi}} e^{-(AT+B)^2/(2T)}, \qquad B > 0,$$
$$= -\frac{2AB - 1}{2A^2} \Phi\left(\frac{AT + B}{\sqrt{T}}\right) - \frac{1}{2A^2} e^{-2AB}\left[1 - \Phi\left(\frac{AT - B}{\sqrt{T}}\right)\right]$$
$$- \frac{\sqrt{T}}{A} \frac{1}{\sqrt{2\pi}} e^{-(AT+B)^2/(2T)}, \qquad B < 0.$$

We can rewrite Theorem 5.2 in terms of Mill's Ratio.

COROLLARY 5.1.

$$
\begin{aligned}
\mathcal{E}_1^* &= \int_0^T t\,\frac{dP_1(t)}{dt}\,dt = \frac{1}{\delta_1}\,\phi\left(\frac{\delta_1 T + \gamma_1}{\sqrt{T}}\right)\sum_{s=0}^{\infty} \\
&\cdot\Bigg\{ e^{-(2r/T)[(r+1)\gamma_1 - r\gamma_2][\delta_1 T + \gamma_1 - (\delta_2 T + \gamma_2]}[(2r+1)\gamma_1 - 2r\gamma_2] \\
&\cdot\left[R\left(\frac{(2r+1)\gamma_1 - 2r\gamma_2 - \delta_1 T}{\sqrt{T}}\right) - R\left(\frac{(2r+1)\gamma_1 - 2r\gamma_2 + \delta_1 T}{\sqrt{T}}\right)\right] \\
&- e^{-[2(r+1)/T][r\gamma_1 - (r+1)\gamma_2][\delta_1 T + \gamma_1 - (\delta_2 T + \gamma_2)]}[(2r+1)\gamma_1 - 2(r+1)\gamma_2] \\
&\cdot\left[R\left(\frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2 - \delta_1 T}{\sqrt{T}}\right)\right. \\
&\left.- R\left(\frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2 + \delta_1 T}{\sqrt{T}}\right)\right]\Bigg\}.
\end{aligned}
\tag{5.10}
$$

*When* $(\gamma_1, \delta_1)$ *and* $(-\gamma_2, -\delta_2)$ *are interchanged,* (5.10) *is* $\mathcal{E}_2^* = \int_0^T t[dP_2(t)/dt]\,dt$.

COROLLARY 5.2.

$$
\begin{aligned}
\mathcal{E}_1^* &= TP_1(T) + \frac{\sqrt{T}}{\delta_1}\,\phi\left(\frac{\delta_1 T + \gamma_1}{\sqrt{T}}\right) \\
&\cdot\sum_{r=0}^{\infty}\Bigg\{ e^{-(2r/T)[(r+1)\gamma_1 - r\gamma_2][\delta_1 T + \gamma_1 - (\delta_2 T + \gamma_2)]} \\
&\cdot\left[\frac{(2r+1)\gamma_1 - 2r\gamma_2 - \delta_1 T}{\sqrt{T}}\,R\left(\frac{(2r+1)\gamma_1 - 2r\gamma_2 - \delta_1 T}{\sqrt{T}}\right)\right. \\
&\left.- \frac{(2r+1)\gamma_1 - 2r\gamma_2 + \delta_1 T}{\sqrt{T}}\,R\left(\frac{(2r+1)\gamma_1 - 2r\gamma_2 + \delta_1 T}{\sqrt{T}}\right)\right] \\
&- e^{-[2(r+1)/T][r\gamma_1 - (r+1)\gamma_2][\delta_1 T + \gamma_1 - (\delta_2 T + \gamma_2)]} \\
&\cdot\left[\frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2 - \delta_1 T}{\sqrt{T}}\,R\left(\frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2 - \delta_1 T}{\sqrt{T}}\right)\right. \\
&\left.- \frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2 + \delta_1 T}{\sqrt{T}}\,R\left(\frac{(2r+1)\gamma_1 - 2(r+1)\gamma_2 + \delta_1 T}{\sqrt{T}}\right)\right]\Bigg\}.
\end{aligned}
\tag{5.11}
$$

*When* $(\gamma_1, \delta_1)$ *and* $(-\gamma_2, -\delta_2)$ *are interchanged,* (5.11) *is* $\mathcal{E}_2^*$.

When we return to the formulation of the sequential procedure of the observed process, $X(t) = Y(t) + \mu t$ referred to the lines $x = c_1 + d_1 t$ and $x = c_2 + d_2 t$, we use the above expressions with $\gamma_1 = c_1$, $\gamma_2 = c_2$, $\delta_1 = d_1 - \mu$, and $\delta_2 = d_2 - \mu$. Again if the procedure is symmetric ($c_1 = -c_2 = c$, $d_1 = -d_2 = d$) the formulas are simplified.

COROLLARY 5.3. *If* $\gamma_1 = -\gamma_2 = c$, $\delta_1 = d - \mu$ *and* $\delta_2 = -d - \mu$ $(c + dT \geqq 0)$,

$$\begin{aligned}
\mathcal{E}_1^* &= \frac{c}{d-\mu} \sum_{s=0}^{\infty} (-1)^s (2s+1) \\
&\quad \cdot \left\{ e^{-2(s+1)c[(s+1)d-\mu]} \Phi\left( \frac{(d-\mu)T - (2s+1)c}{\sqrt{T}} \right) \right. \\
&\qquad \left. - e^{-2sc(sd+\mu)} \Phi\left( \frac{-(d-\mu)T - (2s+1)c}{\sqrt{T}} \right) \right\} \\
&= \frac{c}{d-\mu} \phi\left( \frac{dT + c - \mu T}{\sqrt{T}} \right) \sum_{s=0}^{\infty} (-1)^s (2s+1) e^{-2(c/T)(dT+c)s(s+1)} \\
&\quad \cdot \left[ R\left( \frac{(2s+1)c - (d-\mu)T}{\sqrt{T}} \right) - R\left( \frac{(2s+1)c + (d-\mu)T}{\sqrt{T}} \right) \right] \\
&= TP_1(T) + \frac{\sqrt{T}}{d-\mu} \phi\left( \frac{dT + c - \mu T}{\sqrt{T}} \right) \sum_{s=0}^{\infty} (-1)^s e^{-2(c/T)(dT+c)s(s+1)} \\
&\quad \cdot \left[ \frac{(2s+1)c - (d-\mu)T}{\sqrt{T}} R\left( \frac{(2s+1)c - (d-\mu)T}{\sqrt{T}} \right) \right. \\
&\qquad \left. - \frac{(2s+1)c + (d-\mu)T}{\sqrt{T}} R\left( \frac{(2s+1)c + (d-\mu)T}{\sqrt{T}} \right) \right].
\end{aligned} \tag{5.12}$$

*If* $\mu$ *is replaced by* $-\mu$, (5.12) *is* $\mathcal{E}_2^*$.

The second and third parts of (5.12) were used in computing. As long as $dT + c > 0$, the convergence is very rapid. When $dT + c = 0$, we have

$$\begin{aligned}
\mathcal{E}_1^* &= \frac{c}{d-\mu} \phi(\mu\sqrt{T}) \sum_{s=0}^{\infty} (-1)^s (2s+1) \\
&\quad \cdot \left[ R\left( \frac{2(s+1)c + \mu T}{\sqrt{T}} \right) - R\left( \frac{2sc - \mu T}{\sqrt{T}} \right) \right] \\
&= TP_1(T) + \frac{\sqrt{T}}{d-\mu} \phi(\mu\sqrt{T}) \sum_{s=0}^{\infty} (-1)^s \\
&\quad \cdot \left[ \frac{2(s+1)c + \mu T}{\sqrt{T}} R\left( \frac{2(s+1)c + \mu T}{\sqrt{T}} \right) - \frac{2sc - \mu T}{\sqrt{T}} R\left( \frac{2sc - \mu T}{\sqrt{T}} \right) \right].
\end{aligned} \tag{5.13}$$

The first series converges as

$$\begin{aligned}
&\sum (-1)^s (2s+1) \left[ \frac{\sqrt{T}}{2(s+1)c + \mu T} - \frac{\sqrt{T}}{2sc - \mu T} \right] \\
&\qquad = \sqrt{T} \sum (-1)^s \frac{-2(2s+1)(c + \mu T)}{(2s+1)^2 c^2 - (\mu T + c)^2};
\end{aligned} \tag{5.14}$$

the second series converges as

$$\sum (-1)^s \left[ -\left(\frac{\sqrt{T}}{2(s+1)c + \mu T}\right)^2 + \left(\frac{\sqrt{T}}{2sc - \mu T}\right)^2 \right] = \sum (-1)^s \frac{4(2s+1)c(\mu T + c)T}{[(2s+1)^2c^2 - (\mu T + c)^2]^2}. \tag{5.15}$$

In case $c + dT = 0$, we can manipulate the series (at least formally) to obtain

$$\mathcal{E}_1^* = \frac{c}{d - \mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2/T} \frac{e^{2c[(v/T)+\mu]} - e^{4c[(v/T)+\mu]}}{[1 + e^{2c[(v/T)+\mu]}]^2} dv. \tag{5.16}$$

However, this formula is of doubtful usefulness.

**6. Some remarks.** The methods used in this paper can be extended to treat more general procedures. For example, instead of an upper or lower boundary consisting of one line, we could consider the boundary consisting of two lines. Let each boundary consist of $x = c_i + d_i t$ $(i = 1, 2)$ for $0 \leqq t < T$ and $x = c_i^* + d_i^* t$ for $T \leqq t < T^*$. The process $X(t)$ [or $Y(t)$] at $t = T$ and $t = T^*$ has a bivariate normal distribution. Conditional on $X(T)$ and $X(T^*)$, the process $X(t)$ can be treated in the interval $0 \leqq t < T$ and $T \leqq t < T^*$ as in Section 4.3. Then the result can be integrated relative to the bivariate normal distribution of $X(T)$ and $X(T^*)$. Use of these boundaries would come closer to the optimum procedure.

The procedure that Armitage [2] suggested could also be studied by this method. To test $\mu = 0$ against two-sided alternatives a procedure is to reject the hypothesis if $X(t)$ touches $x = c_1 + d_1 t$ or $x = c_2 + d_2 t$ for $0 \leqq t \leqq T^*$, where $c_1 > 0$, $c_2 < 0$ and to accept the hypothesis if $X(t)$ touches $x = c_1^* + d_1^* t$ or $x = c_2^* + d_2^* t$ for $T \leqq t \leqq T^*$ where $d_1^* > 0$, $d_2^* < 0$ and these last two lines intersect at $t = T$. (The graph of the boundaries may look roughly like a reversed $\sum$.) Again we can consider the problem conditional on $X(T)$, $X(T^*)$ and then integrate the result.

For the procedures considered in this paper we can show the following:

THEOREM 6.1. *If* $\gamma_1 + \delta_1 T > \gamma_2 + \delta_2 T$ *and* $\delta_1 = d - \mu$, $\delta_2 = -d - \mu$,

$$\frac{\partial P_1(T)}{\partial \mu} = \gamma_1 P_1(T) + \delta_1 \mathcal{E}_1^*. \tag{6.1}$$

PROOF. This is verified by expressing the various functions in terms of Mill's Ratios and using the fact that $R'(x) = xR(x) - 1$.

*Acknowledgments.* The author wishes to thank Bernard Epstein and John Tukey for helpful suggestions.

REFERENCES

[1] T. W. ANDERSON, "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proc. Amer. Math. Soc.*, Vol. 6 (1955), pp. 170–176.

[2] P. ARMITAGE, "Restricted sequential procedures," *Biometrika*, Vol. 44 (1957), pp. 9–26.
[3] T. G. DONNELLY, "A family of truncated sequential tests," doctoral dissertation, University of North Carolina, 1957.
[4] J. L. DOOB, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.
[5] EDITORIAL, "The normal probability function: tables of certain area-ordinate ratios and of their reciprocals," *Biometrika*, Vol. 42 (1955), pp. 217–222.
[6] DAVID C. HALEY, "Estimation of the dosage mortality relationship when the dose is subject to error," Technical Report No. 15, August 29, 1952, Stanford University.
[7] WASSILY HOEFFDING, "Lower bounds for the expected sample size and the average risk of a sequential procedure," submitted to *Ann. Math. Stat.*
[8] J. KIEFER AND LIONEL WEISS, "Some properties of generalized sequential probability ratio tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 57–75.
[9] K. PEARSON, *Tables for Statisticians and Biometricians, Part I*, Cambridge University Press, Cambridge, 1924.