SEQUENTIAL TESTS FOR THE MEAN OF A NORMAL DISTRIBUTION

HERMAN CHERNOFF
STANFORD UNIVERSITY

1. Summary

The problem of sequentially testing whether the drift of a Wiener process is positive or negative, given an a priori normal distribution, is reduced to the solution of a free boundary problem involving a diffusion equation.

2. Introduction

This paper is concerned with an approach to obtaining an asymptotically optimal solution (as sampling cost approaches zero) of the problem of sequentially testing whether the unknown mean μ of a normal distribution with known variance is positive or negative. That is the sequential test of $H_1: \mu > 0$ versus $H_2: \mu \leqq 0$.

In a number of problems of varying degrees of generality the following procedure has been found to yield asymptotically optimal solutions [6], [7]. This procedure consists of selecting some nondegenerate a priori distribution on the unknown states of nature and of studying the limiting behavior of the corresponding Bayes solution. Therefore we shall investigate the sequential problem where the a priori distribution of the unknown parameter will be assumed to be normal with fixed mean μ_0 and variance σ_0^2.

The nature of the Bayes solution will depend in part on the loss function. The case where there is an indifference zone, that is, the regret associated with either terminal action is zero in some interval about $\mu = 0$, has been solved and generalized by G. Schwarz [13]. A case which seems of more interest is that where the loss functions corresponding to the two terminal actions are approximately linear in a neighborhood of the origin. Then, the regret due to taking the wrong action may be expressed by

$$r(\mu) = k|\mu| + o(1) \tag{2.1}$$

as $\mu \to 0$. We shall confine our attention to this case although our attack applies more generally, covering quadratic regret among others. Assuming that the cost of sampling is c per observation where $c \to 0$, a relatively simple-minded study will indicate that the main contribution of the Bayes risk is due to values of μ

This work was supported in part by the Office of Naval Research under Task 342-022 at Stanford University.

close to zero and that, with large probability, many observations will be taken when the Bayes test is applied.

This immediately suggests that the solution of our problem is approximately that of a corresponding problem involving the Wiener process with drift μ and variance σ^2 per unit time. There it is desired to test sequentially whether the drift is positive or negative where the regret function is

$$r(\mu) = k|\mu| \tag{2.2}$$

and μ has an a priori normal distribution with mean μ_0 and variance σ_0^2 and there is a cost of c per unit time of sampling.

The substance of this paper will deal with the "reduction" of this problem to the solution of a free boundary problem involving the diffusion equation.

3. Normalization

In this section we normalize the problem of testing for the drift of a Wiener process to the case where the constants k, σ^2, and c are replaced by 1.

Let

$$X_t = \mu t + Z_t, \tag{3.1}$$

where X_t is the observed Wiener process. Then Z_t is a continuous process of independent Gaussian increments with mean 0 and variance $E[Z_t]^2 = \sigma^2 t$.

Let

$$\begin{aligned} X^*_{t^*} &= \alpha X_t, \\ \mu^* &= \beta\mu, \\ t^* &= \gamma t. \end{aligned} \tag{3.2}$$

Then $X^*_{t^*}$ is a Wiener process with drift μ^* per unit time (measured in t^*) if

$$\frac{\alpha}{\beta\gamma} = 1. \tag{3.3}$$

Furthermore $X^*_{t^*}$ has unit variance per unit time if

$$\frac{\alpha^2\sigma^2}{\gamma} = 1. \tag{3.4}$$

Finally the Bayes risk for an arbitrary procedure in the original problem is

$$\mathcal{B} = E[cT + \epsilon(\mu)k|\mu|], \tag{3.5}$$

where T is the time of sampling, $\epsilon(\mu)$ is the probability of error when μ is the mean drift with the given a priori distribution. Then

$$\mathcal{B} = \frac{c}{\gamma}\mathcal{B}^*, \tag{3.6}$$

where

$$\begin{aligned} \mathcal{B}^* &= E\left[T^* + \epsilon^*(\mu^*)\frac{\gamma k}{c\beta}|\mu^*|\right], \\ T^* &= \gamma T, \qquad \epsilon^*(\mu^*) = \epsilon(\mu). \end{aligned} \tag{3.7}$$

If

$$\frac{\gamma k}{c\beta} = 1, \tag{3.8}$$

the problem of minimizing $\mathcal{B}$ has been reduced to that of solving a corresponding problem with the constants c, σ, and k replaced by ones. Note that the mean μ_0 and variance σ_0^2 of the a priori distribution are replaced by

$$\mu_0^* = \beta\mu_0, \qquad \sigma_0^{*2} = \beta^2\sigma_0^2. \tag{3.9}$$

The solution of equations (3.3), (3.4), and (3.8) yields

$$\begin{aligned} \alpha &= c^{1/3}k^{-1/3}\sigma^{-4/3}, \\ \beta &= c^{-1/3}k^{1/3}\sigma^{-2/3}, \\ \gamma &= c^{2/3}k^{-2/3}\sigma^{-2/3}, \end{aligned} \tag{3.10}$$

and consequently

$$\mathcal{B} = c^{1/3}k^{2/3}\sigma^{2/3}\mathcal{B}^*. \tag{3.11}$$

If c is small, ordinary values of μ^* in the normalized problem correspond to values of μ of the order of magnitude of $c^{1/3}$. Similarly ordinary values of t^* correspond to values of t of the order of magnitude of $c^{-2/3}$. The values of μ_0 and σ_0^2 are transformed to values of the order of magnitude $c^{-1/3}$ and $c^{-2/3}$ respectively. Thus our original problem has been reduced to a normalized one where the a priori distribution has large mean and standard deviation.

Hereafter we shall treat the normalized problem unless the contrary is specified.

4. A posteriori probability

Cameron and Martin [5] have shown that the Radon-Nikodym derivative for X_s, where $0 \leqq s \leqq t$, corresponding to two values of μ can be expressed in terms of X_t. Furthermore, the probability measures of X_s on $[0, t]$ corresponding to different values of μ are absolutely continuous with respect to one another. Consequently the a posteriori distribution of μ given X_s on $[0, t]$ may be computed in terms of X_t using the data

$$\mathcal{L}(X_t|\mu) = \mathfrak{N}(\mu t, t) \tag{4.1}$$

and

$$\mathcal{L}(\mu) = \mathfrak{N}(\mu_0, \sigma_0^2), \tag{4.2}$$

where $\mathcal{L}$ represents distribution law and $\mathfrak{N}(a, b)$ represents the normal distribution with mean a and variance b. Consequently a straightforward computation yields

$$\mathcal{L}(\mu|X_t = x) = \mathfrak{N}\left(\frac{x + \frac{\mu_0}{\sigma_0^2}}{t + \frac{1}{\sigma_0^2}}, \frac{1}{t + \frac{1}{\sigma_0^2}}\right). \tag{4.3}$$

It can be shown that an optimal Bayes procedure can be theoretically produced as follows. Using the a posteriori distribution at time t, compute $\mathcal{B}_{it}$, the conditional risk due to stopping and accepting H_i. Also compute $\mathcal{B}_t$, the infimum of risks corresponding to all sequential tests based on data $X_{t'} - X_t$, with $t' \geqq t$. Continue sampling if $\mathcal{B}_t < \min(\mathcal{B}_{1t}, \mathcal{B}_{2t})$. Otherwise stop and accept the hypothesis corresponding to the smaller stopping risk. (If $\mathcal{B}_{1t} = \mathcal{B}_{2t}$, any measurable procedure can be used to decide between H_1 and H_2.)

Because the conditional distribution depends on the past only through X_t, *this procedure can be represented by continuation and stopping sets in the* (x, t) *plane.* Consequently we can confine our attention to procedures which can be so represented and where the stopping rule does not involve past history. In fact, $X(t)$ is a sufficient statistic for $X(s)$, where $0 \leqq s \leqq t$. Presumably this can be used to prove that given any measurable sequential test procedure, there is one with equivalent operating characteristics which can be represented by continuation and stopping sets in the (x, t) plane.

Now let

$$X'_{t'} = X_t + \frac{\mu_0}{\sigma_0^2}, \qquad t' = t + \frac{1}{\sigma_0^2}. \tag{4.4}$$

Then

$$\mathcal{L}(\mu | X'_{t'} = x') = \mathfrak{N}\left(\frac{x'}{t'}, \frac{1}{t'}\right). \tag{4.5}$$

By this transformation, optimal procedures for all versions of our problem corresponding to normal a priori distributions can be represented by two stopping regions and a single continuation region in an (x', t') plane. Each version is distinguished by the point $(\mu_0/\sigma_0^2, 1/\sigma_0^2)$ from which the process starts. Thus the Bayes risk for the optimal procedure can be represented by some function $B_0(x', t')$ evaluated at $x' = \mu_0/\sigma_0^2$ and $t' = 1/\sigma_0^2$. Similarly the conditional Bayes risk given X_s, where $0 \leqq s \leqq t$, is equal to $B_0(X_t + \mu_0/\sigma_0^2, t + 1/\sigma_0^2)$.

Finally, the stopping regions and continuation regions in the (x', t') space can be regarded as determining the Bayes solution for the testing problem where the a priori distribution is replaced by a uniform distribution from $(-\infty, \infty)$.

Hereafter we shall treat the normalized problem in the (x', t') plane and delete the primes unless the contrary is specified. Corresponding to each point (x, t), the a posteriori distribution of μ will be

$$\mathcal{L}(\mu | X_t = x) = \mathfrak{N}\left(\frac{x}{t}, \frac{1}{t}\right). \tag{4.6}$$

The conditional risk of stopping at t and accepting H_1 given $X_t = x$ is

$$D_1(x, t) = \int_{-\infty}^{0} |\mu| \varphi\left(\mu; \frac{x}{t}, \frac{1}{t}\right) d\mu, \tag{4.7}$$

where $\varphi(y; a, b)$ is the normal density with mean a and variance b. Then

$$D_1(x, t) = \frac{1}{\sqrt{t}}\left\{\varphi\left(\frac{x}{\sqrt{t}}\right) - \frac{x}{\sqrt{t}}\left[1 - \Phi\left(\frac{x}{\sqrt{t}}\right)\right]\right\} \tag{4.8}$$

where $\varphi(y) = \varphi(y; 0, 1)$ is the standardized normal density and $\Phi(y)$ is the corresponding cdf. Thus it follows that for the specified loss function, $r(\mu) = |\mu|$, if one should stop at time t with $X_t = x$ he should select the hypothesis according to the sign of x, and his stopping risk is

$$D(x, t) = \frac{1}{\sqrt{t}} \psi\left(\frac{x}{\sqrt{t}}\right), \tag{4.9}$$

where

$$\psi(y) = \begin{cases} \varphi(y) - y[1 - \Phi(y)], & y > 0, \\ \varphi(y) - y\Phi(y), & y < 0. \end{cases} \tag{4.10}$$

Note that $\psi(y)$ is symmetric and well behaved except at $y = 0$, where the right and left derivatives are $-1/2$ and $+1/2$ respectively.

5. Diffusion equation

Let us consider an arbitrary measurable test where the continuation rule can be represented by an open set in the (x, t) plane and the conditional Bayes risk is continuous everywhere and sufficiently differentiable for every point in the continuation set. (This condition can be relaxed considerably. Boundedness of B will imply analyticity on the continuation set.) We shall assume that upon stopping the best terminal decision is made and hence $B(x, t) = D(x, t)$ on the stopping set.

Let $B(x, t)$ represent the conditional Bayes risk which takes into account the probability of error and the expected cost of sampling further. We shall now derive equation

$$1 + B_t + \frac{x}{t} B_x + \frac{B_{xx}}{2} = 0 \tag{5.1}$$

for points in the continuation set and the boundary condition

$$B = D. \tag{5.2}$$

First we note that if $X_t = x$,

$$X_{t+h} = x + h\mu + \sqrt{h}\eta_1, \tag{5.3}$$

where

$$\mathcal{L}(\eta_1) = \mathfrak{N}(0, 1). \tag{5.4}$$

Furthermore

$$\mu = \frac{x}{t} + \frac{1}{\sqrt{t}} \eta_2, \tag{5.5}$$

where η_2 is independent of η_1 and

$$\mathcal{L}(\eta_2) = \mathfrak{N}(0, 1). \tag{5.6}$$

Hence

$$X_{t+h} = x\left(1 + \frac{h}{t}\right) + \left[h\left(1 + \frac{h}{t}\right)\right]^{1/2}\eta, \tag{5.7}$$

where, as in the remainder of this paper, η denotes a standardized normal variate. Then the probability that $|X_s - x| > a$ for some s in $(t, t + h)$ is $o[\exp(-a^2/4h)]$. If (x, t) is an interior point of the continuation set, there is an a so that the probability of termination with δ units of time is $o[\exp(-a^2/4\delta)]$ as $\delta \to 0$ and

$$B(x, t) = \delta + \int_{X-a}^{X+a} B(u, t + \delta)\varphi\left[u; x\left(1 + \frac{\delta}{t}\right), \delta\left(1 + \frac{\delta}{t}\right)\right] du + o(e^{-a^2/4\delta}). \tag{5.8}$$

Hence

$$B(x, t) = \delta + E\left[B + \delta B_t + YB_x + \frac{1}{2}Y^2B_{xx}\right] + o(\delta), \tag{5.9}$$

where

$$Y = \frac{\delta x}{t} + \left[\delta\left(1 + \frac{\delta}{t}\right)\right]^{1/2}\eta \tag{5.10}$$

and the expectation is with respect to the normal distribution of η. Then

$$B = \delta + B + \delta B_t + \frac{\delta x}{t}B_x + \frac{\delta B_{xx}}{2} + o(\delta) \tag{5.11}$$

and equation (5.1) follows. Note that 1 would be replaced by $c(t)$ if $c(t)$ were the cost of observation per limit time at time t. The boundary condition follows from the continuity of B. Under weaker conditions this boundary condition may break down. In fact to prevent breakdown it is required that the boundary be well behaved. Thus the condition fails where the boundary has a vertical section.

It should be noted that this equation can be reduced to the heat equation by the transformation

$$y = \frac{x}{t}, \qquad s = \frac{1}{2t}, \qquad 0 < t < \infty. \tag{5.12}$$

Then, if

$$\tilde{B}(y, s) = B(x, t) + \frac{1}{2s}, \tag{5.13}$$

where $\tilde{B}$ represents the conditional risk including the amount already spent on sampling, we have

$$\tilde{B}_s = \tilde{B}_{xx} \tag{5.14}$$

with the boundary condition

$$\tilde{B} = D\left(\frac{y}{2s}, \frac{1}{2s}\right) + \frac{1}{2s} = \sqrt{2s}\,\psi\left(\frac{y}{\sqrt{2s}}\right) + \frac{1}{2s} = \tilde{D}(y, s). \tag{5.15}$$

6. The free boundary problem

In section 4 we pointed out that there was a continuation set in terms of which the Bayes strategies for all a priori normal distributions could be represented. This set has the property that it simultaneously minimizes $B(x, t)$ *everywhere*. (In relation to the condition posed at the beginning of section 5, it is not difficult to show that for the optimal strategy $B(x, t)$ is continuous and bounded.) In particular the optimal set or associated optimal boundary has the property that the corresponding solution of the diffusion equation is a minimum *everywhere*.

We shall now heuristically indicate that this minimization property is equivalent to the extra boundary condition

$$B_x = D_x \tag{6.1}$$

which determines the "free" boundary.

Suppose that the boundary is specified by two functions $u_1(t)$ and $u_2(t)$ from t_0 to ∞. This determines the Bayes risk $B(x, t)$ for all (x, s) with $s \geqq t_0$. Then it is desired to extend the boundary backward so as to uniformly minimize

$$H(t) = \int_{u_1(t)}^{u_2(t)} B\,dx + \int_{-\infty}^{u_1(t)} D\,dx + \int_{u_2(t)}^{\infty} D\,dx \tag{6.2}$$

for $t < t_0$. Since $B = D$ on the boundary,

$$\frac{dH}{dt} = \int_{u_1}^{u_2} B_t\,dx + \int_{-\infty}^{u_1} D_t\,dx + \int_{u_2}^{\infty} D_t\,dx. \tag{6.3}$$

If

$$B_t(u_1, t) \neq D_t(u_1, t), \tag{6.4}$$

an instantaneous change in u_1 (increase if $B_t < D_t$ and decrease if $B_t > D_t$) would increase dH/dt and lead to smaller values of H for $t < t_0$. Thus if the optimal boundary has finite slope we must have

$$B_t = D_t \tag{6.5}$$

on the boundary. Differentiating $B = D$ along the boundary $u_i(t)$, we have

$$B_x \frac{du_i}{dt} + B_t = D_x \frac{du_i}{dt} + D_t, \tag{6.6}$$

$$B_x = D_x, \tag{6.7}$$

and hence

$$B_{xx} \frac{du_i}{dt} + B_{xt} = D_{xx} \frac{du_i}{dt} + D_{xt} \tag{6.8}$$

and

$$\frac{du_i}{dt} = \frac{B_{xt} - D_{xt}}{D_{xx} - B_{xx}}. \tag{6.9}$$

Furthermore, differentiating the diffusion equation

$$B_{xt} = -B_{xx}\frac{x}{t} - B_x\frac{1}{t} - \frac{B_{xxx}}{2}, \tag{6.10}$$

we can express du_i/dt in terms of the known function D and $B(x, t_0)$. This should permit us to approximate the extension of the optimal boundary backward in time.

To summarize, we have heuristically indicated that the minimizing property of the optimal boundary yields the extra boundary condition $B_t = D_t$ or equiv alently $B_x = D_x$ and that these conditions characterize the optimal boundary. Needless to say there are a number of serious shortcomings of the above derivation. The following remarks may serve to clarify the situation.

First, the optimal boundary from t_0 to ∞ determines $B(x, t_0)$ on the continuation set as the solution of the diffusion equation. Assuming that $B(x, t_0)$ defined on the continuation set and boundary has third order continuous derivatives with respect to x at the boundary point (x_0, t_0) it is possible to express $B(x_0 - r\delta, t_0 - \delta)$ in terms of the slopes of a straight-line boundary from t_0 to $t_0 - \delta$. This expression is

$$B(x_0 - r\delta, t_0 - \delta) = B(x_0, t_0) - \delta[(D_t - rB_x) + s(B_x - D_x)] + o(\delta). \tag{6.11}$$

For the optimal boundary to have finite slope, we must have $B_x = D_x$. Expanding the above expression further, and substituting $B_x = D_x$ and its consequence $B_t = D_t$, the next term is

$$\frac{4(r - s)}{\sqrt{2\pi}}\delta^{3/2}\left[1 + B_t + B_x\frac{x}{t} + \frac{B_{xx}}{2}\right]. \tag{6.12}$$

Clearly, this implies that the diffusion equation is satisfied at the boundary. Incidentally, for the specified regret function $r(\mu) = |\mu|$, it follows that

$$D_t + D_x\frac{x}{t} + \frac{D_{xx}}{2} = 0 \tag{6.13}$$

everywhere except for $x = 0$. Hence

$$D_{xx} = B_{xx} + 2 \tag{6.14}$$

on the boundary. The next term, which is the one requiring the use of B_{xxx}, is of order δ^2 and s appears in it quadratically. This term is

$$\delta^2\left[s^2\left(\frac{D_{xx} - B_{xx}}{2}\right) + s\left(D_{xt} + B_{xx}\frac{x}{t} + \frac{D_x}{t} + \frac{B_{xxx}}{2}\right) + \left(D_{tt} - \frac{D_x r^2 x}{t} - B_{xx}\frac{x}{t}r - r\frac{B_{xxx}}{2} - r^2 D_t - r^2\right)\right]. \tag{6.15}$$

Selecting s to minimize this expression yields

$$s = \frac{D_{xt} + B_{xx}\frac{x}{t} + \frac{D_x}{t} + \frac{B_{xxx}}{2}}{\frac{B_{xx}}{2} - \frac{D_{xx}}{2}}, \tag{6.16}$$

which is equivalent to equation (6.9) if one assumed that the diffusion equation may be differentiated with respect to x at the boundary. Here only derivatives of B with respect to x appear.

This approach serves to show that if the boundary and B are sufficiently well behaved, $B_x = G_x$ on the boundary. Furthermore, it can be used to show that a differentiable boundary for which the solution satisfies the two boundary conditions yields the desired optimal strategy. Such a proof is complex and requires using the fact that replacing a small portion of the curve by a chord increases B everywhere from the minimum attainable by at most $o(\delta)$ and differs from the B corresponding to the curve by at most $o(\delta)$. Dividing a section of the curve into many such small sections yields a function B arbitrarily close to the minimum and to that of the curve. Thus the curve has the desired minimum property.

7. Remarks

7.1. *Truncation rules.* Suppose that a rule is imposed that the sampling procedure be truncated at time t_0 and that if sampling does not terminate before t_0, the statistician should be assigned a positive risk depending upon his position at time t_0. Then the derivation of the diffusion equation and the free boundary conditions apply for $0 < t < t_0$ for the optimal strategy subject to this termination rule. In particular if the assigned risk coincides with B for the nontruncated problem, we get the solution of the nontruncated problem. If the assigned risk at t_0 is changed by less than ϵ, the corresponding B is modified by at most ϵ. Since $D(x, t)$ approaches 0 uniformly in x as $t \to \infty$, it follows that an arbitrary assignment of termination risk for large t has negligible effect on the optimal B and therefore on the optimal boundary for finite values of t.

Incidentally this is an appropriate place to indicate that for the problem which is not arbitrarily truncated, the optimal boundary is not truncated. To prove this all that is required is to compare $D(0, t)$ with the risk $B^*(0, t)$ for the nonoptimal procedure which proceeds for ϵ units of time after t and then terminates. In fact,

$$D(0, t) = (2\pi t)^{-1/2} \tag{7.1}$$

and

$$B^*(0, t) = \frac{1}{(2\pi t)^{1/2}}\left[1 - \left(\frac{\epsilon}{t}\right)^{1/2} + O(\epsilon)\right]. \tag{7.2}$$

On the other hand a similar comparison shows that the optimal procedure for the discrete time sequential problem does terminate. This termination point is of the order of magnitude of c^{-1} which is large compared to $c^{-2/3}$, the scale in which t is measured.

7.2. *Relation between the discrete and continuous solutions.* The discrete problem, when normalized, can be regarded as a solution of the continuous problem where the stopping set is confined to sets corresponding to a sequence of t,

$t_n = \beta + n\epsilon$, where $0 \leqq \beta \leqq \epsilon$ and $\epsilon = O(c^{2/3})$. In the continuous problem consider the procedure which consists of terminating sampling at $\alpha + n\epsilon$ if $X(t)$ crosses the optimal boundary for the first time between $\alpha + (n - 1)\epsilon$ and $\alpha + n\epsilon$. The Bayes risk for this procedure differs from the optimal by at most ϵ, which is more than the cost of extra sampling. On the other hand an optimal procedure which permits termination only at the times $\alpha + n\epsilon$ can be expressed in terms of the values of X at these times alone. It follows that the minimum risk for the discrete problem converges to that of the continuous problem as $\epsilon \to 0$. At the same time the definition of the optimal stopping sets obtained by comparing B and D shows that the optimal boundary of the discrete problem also converges to that of the continuous problem. In fact the convergence of both B and the continuation sets are monotone when $\epsilon \to 0$ through a sequence $\gamma 2^{-i}$. At present J. V. Breakwell and the author are engaged in the numerical computation of the optimal boundary using the solution of the discrete problem as the approximation.

Since Sobel [14] has shown that the Bayes strategies for the discrete problem yield continuation sets which are intervals for each n, the above argument extends this result to the continuous case.

7.3. *Application to other regret functions and costs of sampling.* It is evident that the free boundary approach is applicable when the marginal cost of sampling is a function of time or when the regret due to error is given by some other function than $|\mu|$. The differential equation is replaced by

$$c(t) + B_t + \frac{x}{t} B_x + \frac{B_{xx}}{2} = 0, \tag{7.3}$$

where $c(t)$ is the additional cost of observation per unit time at time t. The boundary condition is changed in that D is replaced by a modified function depending on $r(\mu)$.

7.4. *Applicability of the diffusion equation to bounds.* The behavior of $B(x, t)$ and the optimal boundary as $t \to 0$ is of special interest. The diffusion equation may furnish simple bounds as follows. Let B^* be the solution of the equation for a prescribed boundary. Since the optimal boundary minimizes B everywhere, $B^* < D$ implies $B < D$ and consequently the optimal rule calls for continued sampling at those points where $B^* < D$.

A trivial related argument uses the fact that B is bounded. This is so because one may sample for one unit of time for a cost of one and $B(x, t) < G(x, t) \leqq 1$ for $t \geqq 1$. Hence (x, t) is in the continuation set if $G(x, t) \geqq 2$. But for $x = (-at \log t)^{1/2}$ as $t \to 0$,

$$G(x, t) = \frac{1}{\sqrt{t}} \psi[(-a \log t)^{1/2}] \sim \frac{t^{(a-1)/2}}{\sqrt{2\pi}\,(-a \log t)}. \tag{7.4}$$

Hence the value of x at the upper boundary is larger than $(-at \log t)^{1/2}$ for $a < 1$ and $t \to 0$. The author has reason to conjecture that along the optimal boundary $x \sim (-3t \log t)^{1/2}$.

7.5. *A property of the optimal boundary.* The following argument due to Stein indicates that for the upper portion of the optimal boundary, $x/\sqrt{t}$ is a decreasing function of t. This, together with the continuity of B, implies that the optimal upper boundary is continuous and is differentiable except for a set of t measure 0 on which the slope may be $-\infty$.

Let

$$X^*_{t^*} = \frac{X_t}{\sqrt{t_0}},$$

$$\mu^* = \mu\sqrt{t_0}, \tag{7.5}$$

$$t^* = \frac{t}{t_0}.$$

Then

$$\mathcal{L}(X^*_{t^*}|\mu^*) = \mathfrak{N}(\mu^* t^*, t^*) \tag{7.6}$$

and

$$\mathcal{L}(\mu^*|X^*_t = x^*) = \mathfrak{N}\left(\frac{x^*}{t^*}, \frac{1}{t^*}\right). \tag{7.7}$$

At time t_0 with $t^* = 1$, the problem of minimizing the Bayes risk

$$E[T + \epsilon(\mu)|\mu|] = E\left[t_0 T^* + \frac{\epsilon^*(\mu^*)|\mu^*|}{\sqrt{t_0}}\right] \tag{7.8}$$

is equivalent to that of minimizing

$$E[t_0^{3/2} T + \epsilon^*(\mu^*)|\mu^*|] \tag{7.9}$$

which is the same as our original problem except that the cost of sampling is $t_0^{3/2}$. Clearly the continuation set decreases as the cost of sampling increases. Hence the upper boundary value of x^* at time t_0 decreases with t_0. The desired result follows.

8. Some historical comments and acknowledgments

The sequential probability ratio test, for testing a simple hypothesis versus a simple alternative was introduced by Wald in 1943 (see [15]). He conjectured that the test was optimal in the sense that among tests with error probabilities as small, this minimized the expected sample size under both hypotheses *simultaneously*. This property, together with the Bayes nature of the test, was obtained in 1948 by Wald and Wolfowitz [16] and Arrow, Blackwell, and Girshick [3]. This test has been customarily applied to problems, such as ours, which involve composite hypotheses by replacing the composite hypotheses by simple ones corresponding to somewhat arbitrarily selected points of the related parameter sets.

Sobel [14] characterized an essentially complete class of one-sided sequential tests of composite hypotheses with bounded loss functions where the observations have a Koopman-Darmois type distribution. This essentially complete class, which could be called the class of generalized sequential probability ratio tests, was also discussed for more general distributions by Kiefer and Weiss [11].

Breakwell [4] conjectured and DeGroot [8] proved that the minimax sequential test for the Wiener process problem with linear regret function is the Wald test. The derivation involves demonstrating that the least favorable distribution is a two point distribution and thus the minimax test has a continuation set consisting of two appropriately placed horizontal parallel straight lines which is essentially a special case of the Wald sequential probability ratio test.

Several authors have proposed modifications of the Wald test. In particular Armitage [2] proposed a two-sided test for $\mu = 0$ for which the boundary consists of straight line segments. He gave approximations to the operating characteristics. This term is used loosely to include the expected sample size or its distribution in addition to the error probabilities. Donnelly [9] and Anderson [1] both proposed one-sided tests consisting of pairs of not necessarily parallel straight lines and derived the operating characteristics. All three converted to the Wiener process. Donnelly obtained his results by solving the diffusion equation which has always been closely associated with the Wiener process.

The free boundary approach, which is the substance of this paper, was independently developed by Dennis V. Lindley. He derived the boundary condition $B_x = G_x$ but was apparently unaware at the time we communicated that this extra condition should suffice to determine the optimal boundary.

Free boundary problems for diffusion equations have recently been the subject for considerable research. These problems are somewhat different than ours. Some examples are in the work of Friedman [10] and Kolodner [12].

I wish to thank C. M. Stein and M. M. Schiffer for their comments and discussions, which proved helpful to me.

REFERENCES

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

[2] P. Armitage, "Restricted sequential procedures," *Biometrika*, Vol. 44 (1957), pp. 9–26.

[3] K. J. Arrow, D. Blackwell, and M. A. Girshick, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.

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

[5] R. H. Cameron and W. T. Martin, "Transformation of Wiener integrals under a general class of linear transformations," *Trans. Amer. Math. Soc.*, Vol. 58 (1945), pp. 184–219.

[6] H. Chernoff, "Large-sample theory: parametric case," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 1–22.

[7] ———, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 493–507.

[8] M. H. DeGroot, "Minimax sequential tests of some composite hypotheses" (abstract), *Ann. Math. Statist.*, Vol. 31 (1960), p. 233.

[9] T. G. Donnelly, "A family of truncated sequential tests," unpublished Ph.D. thesis, University of North Carolina, 1957.

[10] A. FRIEDMAN, "Free boundary problems for parabolic equations I: melting of solids," *J. Math. Mech.*, Vol. 8 (1959), pp. 499–518.

[11] J. KIEFER and L. WEISS, "Some properties of generalized sequential probability ratio tests," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 57–75.

[12] I. I. KOLODNER, "Free boundary problems for the heat equation with applications to problems of change of phase," *Comm. Pure Appl. Math.*, Vol. 9 (1956), pp. 1–31.

[13] G. SCHWARZ, "Asymptotic shapes of optimal stopping regions for sequential testing" (abstract), *Ann. Math. Statist.*, Vol. 31 (1960), p. 537.

[14] M. SOBEL, "An essentially complete class of decision functions for certain standard sequential problems," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 319–337.

[15] A. WALD, *Sequential Analysis*, New York, Wiley, 1947.

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