ASYMPTOTIC STOPPING REGIONS FOR SEQUENTIAL TESTING WITH AN INDIFFERENCE REGION

by

GIDEON SCHWARZ

TECHNICAL REPORT NO. 54

January 19, 1960

PREPARED UNDER CONTRACT N6onr-25140

(NR 342-022)

OFFICE OF NAVAL RESEARCH

This work was sponsored by the Army, Navy, and Air Force through the Joint Services Advisory Committee for Research Groups in Applied Mathematics and Statistics by Contract No. N6onr-25140 (NR 342-022).

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government

APPLIED MATHEMATICS AND STATISTICS LABORATORIES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
I. Introduction.

Wald's sequential probability ratio test solves in an optimal way the problem of testing for the mean of a normal random variable with given variance, when the mean is known to be one of two values \( M_1 \) and \( M_2 \). If, however, a third possible value is added, say \( M_0 \), which lies between \( M_1 \) and \( M_2 \), and if the loss for making a wrong decision is zero when \( M_0 \) is the true mean, the Wald test cannot be optimal any more. Its non-optimality becomes especially apparent when the cost \( c \) of taking an observation is small compared to the losses for a wrong decision. It has been shown by Chernoff [1] that in that case the expected loss of any reasonable test is of the same order of magnitude as the part of the loss due to sampling costs, while the part due to wrong decisions becomes relatively negligible. On the other hand, the expected sample size of the Wald test is largest at the intermediate point, and, when \( c \to 0 \), grows faster at that point than at \( M_1 \) and \( M_2 \).

As the determination of the optimal tests for the problem seems hopelessly complicated, the question arises whether a family of tests - one for each value of \( c \) - can be found, such that asymptotically, as \( c \) approaches zero, the tests are as good as the optimal ones.

The question is answered in the affirmative by the results of the present paper, and of a second paper that is in preparation.
In the present paper a family of tests is defined by a property that could be used in the 2-Hypotheses problem to characterize the Wald test: For every test in the family, the a posteriori risk due to stopping and taking the more favorable action is a constant $r$ along the boundaries of the stopping region.

The regions in the $(n, S_n)$-plane are shown to grow, as $r \to 0$ in all directions like $-\log r$. The "asymptotic shape" of the regions as defined below, is found explicitly for the case described, and for a generalization to composite hypotheses. This generalization essentially covers the case where there are two hypotheses separated by a true indifference zone so that if the mean lies in the indifference zone there is no loss attached to either terminal decision.

In the last section, the asymptotic shapes found are shown to be valid for a whole class of $r$-families of tests.

In the second paper, we show that if $r$ is chosen equal to the cost of sampling, the above class of families can be enlarged so as to include the family of optimal tests. The asymptotic shapes computed here are thus the optimal asymptotic shapes for the problem.

We define the asymptotic shape of a stopping region as follows:

**Definition.** For $0 < r < \infty$, let $\Sigma(r)$ be a one parameter family of sets in a plane with a Cartesian coordinate system $[x, y]$. Denote by $x(r)$ the supremum of $x$ in $\Sigma(r)$. If, as $r$ tends to zero, the transformed region $\Sigma(r)/x(r)$ approaches a limit-set, that set is the asymptotic shape of the family.
In the families considered here, \( x(r) \) tends to infinity, and the asymptotic shapes are therefore significant for the "large sample theory" of the problems.

We start by finding the asymptotic shapes for some well defined families of tests, and proceed to show that the results are valid for all families that are in a certain sense close enough to those defined. The class of those families is characterized by certain bounds on the risk along the boundaries of the stopping regions.

We hope to show in a following paper that those classes are essentially complete for testing the mean of a normal distribution with known variance. The cases considered are those of a three point \textit{a priori} distribution over the mean, and that of an \textit{a priori} distribution that gives positive weight to every set of positive Lebesgue measure. Other cases can be easily treated according to the same ideas.

The author wishes to thank the Statistics Department of Stanford University, and in particular Professor Herman Chernoff for their help.

II. The case of three simple hypotheses.

Let \( x_n \) be a sequence of independent, normally distributed random variables with variance one and unknown mean \( \theta \). Three simple hypotheses are given:

\[
H_0 : \theta = m, \quad H_0 : \theta = m, \quad H_+ : \theta = M, \]

where \( M \) is positive, and \(-M < m < M\). The assumptions of unit variance and of symmetry of \( H_- \) and \( H_+ \) around zero can always be fulfilled by a
linear transformation on the \( X_n \). Let \( A_- \) and \( A_+ \) be the available terminal actions, with a loss function defined by

\[
L(H_-, A_+) = L_- , \quad L(H_+, A_-) = L_+ , \quad \text{and} \quad L = 0 \ \text{otherwise}.
\]

Let \( W = (W_-, W_0, W_+) \) be a given a priori distribution. After taking \( n \) observations the pair \( (n, S_n) \), where \( S_n \) denotes the sum of the observations, is a sufficient statistic. The \textit{a posteriori} probabilities can be expressed in terms of this statistic as follows:

\[
\begin{align*}
P_- &= K W_- \exp(-mS_n - \frac{1}{2}m^2 n) \\
P_0 &= K W_0 \exp(ms_n - \frac{1}{2}m^2 n) \\
P_+ &= K W_+ \exp(mS_n - \frac{1}{2}m^2 n)
\end{align*}
\]

where \( K \) is chosen so as to make the sum of all three equal to one. If the \( P_i \) \( (i = -, +, 0) \) are represented by barycentric coordinates in a triangle, the points corresponding to fixed \( n \) and varying \( S_n \) lie on a curve whose equation can be found by eliminating \( K \) and \( S_n \) from (1):

\[
(2) \quad P_-^{M-m} P_+^{M+m} = F(W, n)P_0^{2M}.
\]

The set of all \( (n, S_n) \) is mapped by (1) on the curves (2); however, if we do not restrict \( n \) to positive integer values, (1) describes a one-to-one mapping of the \( (n, S_n) \)-plane onto the interior of the triangle. By inverting the mapping, subsets of the triangle can be transformed into corresponding sets in the \( (n, S_n) \)-plane.
Consider for instance the set of all posterior-probability points for which the risk when taking action \( A_+ \) is greater than or equal to a given constant \( r \). In the triangle this set is described by \( L_+ P_+ \geq r \). After substituting the right expression for \( K \) in (1), this set has as its image in the \((n,S_n)\)-plane the set given by

\[
(3) \quad W_+ (L_- - r) \exp(-MS_n - \frac{1}{2} M^2 n) \geq rW_0 \exp(mS_n - \frac{1}{2} m^2 n) + rW_+ \exp(MS_n - \frac{1}{2} M^2 n).
\]

The derivatives of the two sides of inequality (3) with respect to \( S_n \) fulfill an inequality in the reverse direction; hence, the set described by (3) includes with any point \((n,S_n)\) all the points with the same \( n \) and smaller \( S_n \). The boundary of the set consists of the points \((n,S_n)\) for which the equality sign in (3) holds. For each \( n \) there can be only one such point. We denote its \( S_n \)-value by \( S(n,r,+) \). Generally the equation corresponding to (3) is transcendental in \( \exp(S_n) \), and no explicit form of \( S(n,r,+) \) can be found. In the "symmetrical" case \( m = 0 \), however, the equation can be reduced to quadratic form, and solved, yielding the expression

\[
(4) \quad S(n,r,+) = \frac{1}{M} \log \frac{- W_0 \exp(\frac{1}{2} M^2 n) + \sqrt{W_0^2 \exp(M^2 n) + 4W_+ W_0 (L_- - r)/r}}{2W_+}.
\]
Returning to the general case, we can define a completely analogous
function \( S(n,r,-) \), such that the posterior risk when taking action \( A_- \)
is greater than or equal to \( r \). The set of points for which the posterior
risk is at least \( r \) for all actions is described by \( S(n,r,-) \leq S_n \leq S(n,r,+) \).
Assume now that the following decision procedure is used: "Continue sampling
as long as the posterior risks for both actions are at least \( r \). The first
time one of the risks is smaller than \( r \), stop sampling and take the corre-
sponding action." In the \((n, S_n)\)-plane the set for which \( A_+ \) has a smaller
risk than \( A_- \) is characterized by
\[
L^- W^- \exp(-MS_n) \leq L^+ W^+ \exp(\text{MS}_n) ,
\]
or simply
\[
S_n = \frac{1}{2M} \log \frac{L^- W^-}{L^+ W^+} .
\]

We can now define the stopping regions for the given decision procedure:

\[
\text{If } S_n > \max[S(n,r,+) \text{, } \frac{1}{2M} \log(L^- W^- / L^+ W^+)] \text{, stop and take action } A_+ ;
\]

\[
\text{if } S_n < \min[S(n,r,-) \text{, } \frac{1}{2M} \log(L^- W^- / L^+ W^+)] \text{, stop and take action } A_- .
\]

III. The asymptotic shape in the case of three simple hypotheses.

In the previous section a stopping region, or rather a family of stopping
regions -- one for each value of \( r \) -- has been defined. We shall now proceed
to find the asymptotic shape of the family for \( r \) approaching zero.

When we set \( n = 0 \) in the inequality (3), and replace the greater than
or equal sign by an equal sign, we can easily see that as \( r \) tends to zero,
$S(0,r,+)$ is of the same order of magnitude as $\log l/r$. Therefore, we subject the $(n,S_n)$-plane to the transformation

$$S_n = y \log l/r, \quad n = t \log l/r,$$

and find the limiting intersection of the transformed region with a fixed line $y = kt$. After carrying out the necessary substitutions, equation (3) becomes

$$W_-(L_+ - r) \exp\left[(-M - \frac{1}{2} M^2/k) y \log l/r\right]$$

$$= rW_0 \exp\left[(m - \frac{1}{2} m^2/k) y \log l/r\right] + rW_+ \exp\left[(M - \frac{1}{2} M^2/k) y \log l/r\right].$$

As $r$ approaches zero, the $W_1$ and the number $L_+ - r$ can be ignored. On the right side, the first or second term will dominate, according as to whether $M - \frac{1}{2} M^2/k$ is smaller or larger than $m - \frac{1}{2} m^2/k$. If $0 \leq k < \frac{1}{2} (M+m)$, the first will be the case, and we have, denoting now by $y$ the limit,

$$-M - \frac{1}{2} M^2/k)y \log l/r = \log r + (m - \frac{1}{2} m^2/k)y \log l/r,$$

$$y = \frac{2k}{(M+m)(M-m+2k)}.$$

For values of $k$ larger than $\frac{1}{2} (M+m)$ the second term dominates the right side of the equation, and we have
\[(11) \quad (- M - \frac{1}{2} M^2/k)y \log l/r = \log r + (M - \frac{1}{2} M^2/k)y \log l/r , \]

\[(12) \quad y = 1/2M . \]

By substituting for \( k \) its value \( y/t \) we get the following asymptotic shape for \( S(n,r,+) \):

For \( 0 \leq t = \frac{1}{M(M+m)} \) \( y = \frac{1}{2M} ; \)

\[(13) \quad \text{for} \quad t > \frac{1}{M(M+m)} \quad y = \frac{1}{(M+m)} - \frac{1}{2} (M-m)t . \]

The rest of the boundary of the upper stopping region consists of the horizontal line \( S_n = (1/2M) \log (L W/L W_+), \) independently of \( r \), and the asymptotic shape of the line is obviously the line \( y = 0 \).

After computing the asymptotic shape of the lower region in a similar way, we get the following partition of the \((t,y)\)-plane:
IV. The case of three composite hypotheses.

We again assume $X_n$ to be distributed as in the previous section. The
three hypotheses, however, are composite, and together cover all $\theta$ on the
real line:

$$H_- : \theta \leq -M , \quad H_0 : -M < \theta < M , \quad H_+ : \theta \geq M .$$

The loss function is defined as before, and is therefore independent of $\theta$
within each hypothesis. The a priori distribution on $\theta$ is assumed to
dominate the Lebesgue measure on the $\theta$-line, that is, for a set of $\theta$ of
positive measure, the probability, $W$, is positive.

Under those assumptions, the set in the $(t,y)$-plane for which the risk
when taking action $A_+$ is exactly $r$, is described by an equation analo-
gous to equation (8) in the preceding section:

$$L \int_{-\infty}^{-M} \exp[(\theta - \theta^2/2k)y \log 1/r]dW(\theta) = r \int_{-\infty}^{\infty} \exp[(\theta - \theta^2/2k)y \log 1/r]dW(\theta) .$$

As $r$ tends to zero, we have essentially the same situation as in the pre-
ceding section; the exponentials with the highest exponent will dominate.
This can be seen by the following argument: when the $(y \log 1/r)$-th roots
of both sides of (14) are computed, we obtain

$$L \int_{-\infty}^{-M} \exp(\theta - \theta^2/2k)/y \log 1/r \|\exp(\theta - \theta^2/2k)\|_1 = e^{-1/y} \|\exp(\theta - \theta^2/2k)\|_2 .$$
The terms $\| \cdot \|_1$ and $\| \cdot \|_2$ denote the $L_p$-norms, in the space of functions with $W$-integrable $p$-th powers, of the functions $f_1(\theta)$ and $f_2(\theta)$ respectively, where

$$f_1(\theta) = \exp(\theta - \theta^2/2k) \text{ for } \theta < -M,$$

$$= 0 \text{ otherwise},$$

$$f_2(\theta) = \exp(\theta - \theta^2/2k) \text{ for all } \theta,$$

and

$$p = y \log 1/r = S(n,r,+).$$

Now let $r$ approach zero. $p$ tends to infinity, and $L_p$-norms approach the $L_\infty$-norm, which equals the essential supremum of the functions concerned. As the measure $W$ dominates the Lebesgue measure, the $W$-essential supremum equals at least the Lebesgue essential supremum. The latter is in our case simply the maximum, $\exp(-M-M^2/2k)$ on the left, and $e^{k/2}$ on the right hand side of equation (15). Passing to the limit, and taking logarithms, we get

$$-M - M^2/2k = -1/y + k/2,$$

and finally, after substituting $k = y/t$, we find for the asymptotic shape of $S(n,r,+)$ the following curve:

$$y = \sqrt{2t} - Mt.$$
The asymptotic shape of \( S(n,r,\cdot) \) is given by the mirror image of (17) with respect to the \( t \)-axis. The two curves intersect at the origin and at the point \( t = 2/M^2, y = 0 \). Each of them is a parabola touching the \( y \)-axis at the origin.

V. Relaxation of the "constant risk" assumption.

For \( 0 < r \), let \( \Sigma(r) \) be the boundary of a region, such that the following holds:

The risk when taking action \( A_+ \) along the boundary \( \Sigma(r) \) is bounded below by \( rg \) and above by \( r \); the positive constant \( g < 1 \) is the same for all \( r \).

We proceed to show that the asymptotic shape for this family is the same as for the family with risk exactly equal to \( r \) along the boundary.
Denoting the function whose graph is \( \Sigma(r) \) by \( \Sigma(n,r,+) \), we have the inequality

\[
(18) \quad S(n,r,+)^{-1} \leq \Sigma(n,r,+) \leq S(n,gr,+)^{-1}
\]

with \( S(n,r,+) \) denoting, as before, the function describing the curve of constant risk \( r \). Dividing through by \( \log 1/r \), we have

\[
(19) \quad \frac{S(n,r,+)}{\log 1/r} \leq \frac{\Sigma(n,r,+)}{\log 1/r} \leq \frac{S(n,gr,+)}{\log 1/r} = \frac{S(n,gr,+)}{\log g + \log 1/gr}.
\]

Now, if as \( r \) tends to zero, the first term tends to a limit, the last term will tend to the same limit, and so will the terms in between. The asymptotic shapes found in the previous sections are therefore valid for the more general class of procedures with bounded relative variation of the risk along the boundaries.

We conjecture that this class is essentially complete.

REFERENCE
