ON MINIMIZING THE RISK IN CERTAIN SEQUENTIAL TESTS, FOR KNOWN OR UNKNOWN COST

by

CAMPBELL BURGESS READ
Department of Statistics

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 626

JUNE 1969

Research partially supported by the United States Air Force Office of Scientific Research Grant No. AFOSR-68-1415 and the National Institutes of Health, Institute of General Medical Science Grant No. GM-12865-05.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to acknowledge the guidance and inspiration of Professor Normal L. Johnson throughout the preparation of this dissertation. His invaluable comments and suggestions were a constant encouragement; it has been a privilege to have been taught by him and to have worked under him.

I would also like to thank Professor Walter L. Smith for his guidance and encouragement during my graduate study in Chapel Hill, as well as the faculty of the Department of Statistics for their teaching, and the members of my dissertation committee for their interest and suggestions.

The Department of Statistics provided financial support during my period of study and work at Chapel Hill. Research grants for the computer programs and for typing came from the Air Force Office of Scientific Research and from the National Institute of Health. These are greatly appreciated.

Finally, I wish to thank James O. Kitchen for his assistance in computer programming, and Mrs. Cynthia Grossman for coping so cheerfully with the formidable task of typing.

CHAPTER I

INTRODUCTION

1.1.

Between 1944 and 1946 Abraham Wald published a series of papers ([18], [19], [20], [21], [22], [23]) which effectively brought Sequential Analysis into the main stream of statistical theory. Using the important Fundamental Lemma [18], Wald developed the properties of the Sequential Probability Ratio Test (SPRT) of a simple hypothesis against a simple alternative, and evaluated approximations to the Operating Characteristic (O.C.) Function and to the Average Sample Number (ASN) of the procedure for a large class of sequences of independently and identically distributed (i.i.d.) random variables. He also calculated bounds on the O.C. function and on the ASN ([19]). Since these approximations play a large part in subsequent chapters, a summary of the main results is given in 1.2.

In 1.3., the optimum property of Wald's test ([25], [26]) is related to restrictions which are usually placed on the boundaries of the continuation region of the test and is also related to the "risk" of the procedure when the prior distribution of the unknown parameter θ is given. The background of Chapter II is completed in 1.3.

A modification of Wald's SPRT is introduced and examined in Chapter III, and in 1.4. the background and main results of the rest of this paper are set out. The modified SPRT is used in Chapter V, which together with Chapter IV explores the problem of minimizing the risk when the cost is unknown. A multivariate test discussed in Chapter VII is also introduced in 1.4.

1.2. The SPRT: Notation and Main Results.

$X_1, X_2, \ldots\ldots$ is a sequence of i.i.d. random variables, and for every j, X_j has a probability density function (p.d.f.) or likelihood $p_\theta(x_j)$. The likelihood of

$$\underline{X}_n = (X_1, X_2, \ldots\ldots, X_n)$$

is denoted by

$$p_\theta(\underline{x}_n) = \prod_{j=1}^{n} p_\theta(x_j) .$$

In the standard Wald SPRT, we wish to test the simple hypothesis

$$H_0: \quad \theta = \theta_0$$

against the simple alternative

$$H_1: \quad \theta = \theta_1 \quad .$$

The test $\mathcal{T}$ is then as follows:

given two constants A and B such that $0 < B < A$, continue sampling so long as

$$B < \frac{p_{\theta_1}(\underline{x}_n)}{p_{\theta_0}(\underline{x}_n)} < A \quad .$$

If $\dfrac{p_{\theta_1}(\underline{x}_n)}{p_{\theta_0}(\underline{x}_n)} \geq A$, stop and accept H_1 .

If $\dfrac{p_{\theta_1}(\underline{x}_n)}{p_{\theta_0}(\underline{x}_n)} \le B$, stop and accept H_0 .

If $a = \log A$,

$b = \log B$,

$Z_0 = 0$,

$$Z_j = \log \frac{p_{\theta_1}(\underline{x}_j)}{p_{\theta_0}(\underline{x}_j)} \;, \quad j = 1, 2, \ldots,$$

and

$$z_j = Z_j - Z_{j-1} \;, \qquad j = 1, 2, \ldots;$$

then sampling is continued until

$$b < \sum_{j=1}^{n} z_j < a$$

is violated, and T may be rephrased in terms of the log-likelihood-ratio; in many situations it can be discussed in terms of a particle on a random walk between two absorbing barriers.

The O.C. function is denoted by $L(\theta)$, and

$$L(\theta) = \Pr(\text{accept } H_0 | \theta) .$$

Let N denote the stopping variable. In this paper, the stopping variable is synonymous with the sample size. The ASN of any sequential procedure is denoted by $E(N|\theta)$.

Wald's Fundamental Lemma [18], as applied to z_1, states that under certain conditions $\exists\ h = h(\theta) \neq 0$ such that

$$E\, e^{hz_1} = 1,$$

i.e.

$$\int_G \left(\frac{p_{\theta_1}(x_1)}{p_{\theta_0}(x_1)}\right)^h dF_\theta(x_1) = 1,$$

where the integral is a sum if X_1 is a discrete variable, and F_θ is the cumulative distribution function of X_1 . G is a set, not depending on θ, such that

$$\Pr(X_1 \in G) = 1 .$$

Let $1 - \alpha' = L(\theta_0)$

$\beta' = L(\theta_1)$

and let α, β be the nominal error probabilities when excess over the boundaries is neglected.

Further, we introduce the constraint

$$0 < B < 1 < A$$

or

$$-\infty < b < 0 < a < \infty \quad .$$

This constraint is important in considering the optimum property of the SPRT in the next section. When it holds, Wald showed that

$$a = \log \frac{1-\beta}{\alpha} \leq \log \frac{1-\beta'}{\alpha'}$$

$$b = \log \frac{\beta}{1-\alpha} \qquad \log \frac{\beta'}{1-\alpha'}$$

$$\alpha' + \beta' < \alpha + \beta < 1$$

$$L(\theta) \simeq \frac{\exp[ah(\theta)] - 1}{\exp[ah(\theta)] - \exp[bh(\theta)]} \, , \qquad h(\theta) \neq 0.$$

If $E(z_1|\theta) \neq 0$,

$$E(N|\theta) \simeq \frac{bL(\theta) + a\{1 - L(\theta)\}}{E(z_1|\theta)} \quad .$$

The above approximations are better when α' and β' are small, or when $E(N|\theta)$ is large. Page [15] and Kemp [13] improved Wald's approximations to $L(\theta)$ and to $E(N|\theta)$ for the case when X_j is a normal r.v. with known variance and mean θ. Bhate [4] derived upper and lower bounds for the cumulative distribution of N, and applied them to special cases.

Wald's bounds on $L(\theta)$ will be used in Chapter VI: --

let

$$\underline{\lim_{\zeta}} \left[\zeta \, E(e^{hz} | e^{hz} \leq \frac{1}{\zeta})\right] = \eta$$

$$\overline{\lim_{\rho}} \left[\rho \, E(e^{hz} | e^{hz} \geq \frac{1}{\rho})\right] = \delta ,$$

where it is understood that η and δ depend on θ. Then

$$\frac{e^{ah} - 1}{e^{ah} - \eta e^{bh}} \leq L(\theta) \leq \frac{\delta e^{ah} - 1}{\delta e^{ah} - e^{bh}} \quad \text{if} \quad h > 0$$

and

$$\frac{1 - e^{ah}}{\delta e^{bh} - e^{ah}} \leq L(\theta) \leq \frac{1 - \eta e^{ah}}{e^{bh} - \eta e^{ah}} \quad \text{if} \quad h < 0 \quad .$$

Wald calculated η and δ for the case where X_i is normally distributed r.v. having mean θ and known variance. These values are used in Section 6.2.

1.3. Optimum Property of the SPRT.

Wald and Wolfowitz ([25], [26]) showed that in the class of all sequential tests between two simple hypotheses H_0 and H_1, for which

$$\Pr(\text{accept } H_1 | \theta = \theta_0) \leq \alpha'$$

$$\Pr(\text{accept } H_0 | \theta = \theta_1) \leq \beta'$$

hold, the SPRT is "best" in the sense that whenever $\theta = \theta_0$ or $\theta = \theta_1$, it requires on the average fewest observations. We denote this property by P.

Burkholder and Wijsman [6] pointed out, however, that property $\mathcal{P}$ had been shown valid only when

(i) $B \le 1 \le A$

(ii) $E(N|\theta) < \infty$.

They showed that condition (ii) is not necessary, but that condition (i) is necessary. It is for this reason that (i) will be assumed to hold throughout this paper.

Burkholder and Wijsman also showed that if attention is restricted to the class of tests having at least one observation, then property $\mathcal{P}$ of the SPRT holds without condition (i), other than the obvious condition $B \le A$. Tests having no observations will, however, be considered in Chapter II.

The optimum property may be regarded in another light [14]. Suppose θ to have a prior distribution in which

$$w = Pr(\theta = \theta_0)$$

$$1 - w = Pr(\theta = \theta_1)$$

denoted by

$$\lambda_w = (w, \ 1 - w) \ .$$

Also let c be the cost per observation, and L_0, L_1 the losses for incorrectly rejecting H_0 and H_1 respectively. The risk function of a test procedure δ is

$$R_\delta(\lambda_w) = L_0 w\alpha' + L_1(1-w)\beta' + cE(N|\theta)$$

where α' and β' are the exact error probabilities of δ. Then in the class of tests δ having α' and β' as error probabilities, the SPRT minimizes $R_\delta(\lambda_w)$. This is property $\mathcal{P}$. If the class of tests is not restricted by bounds on the error probabilities, then the optimum test minimizing $R_\delta(\lambda_w)$ is either a SPRT or a decision made without taking any observations. In Chapter II, this will be discussed in detail, using an iterative procedure, in terms of Wald's nominal error probabilities α and β.

Chernoff [7] suggested approximate error probabilities leading to a nominal minimum risk, and although these error probabilities are not always close to those obtained by iteration, we find that they often lead to a risk which is close to that obtained by iteration. The special case of testing a normal mean is discussed, and also the class of symmetric tests, as well as more general testing problems. When $w = \frac{1}{2}$, a symmetric test is frequently as nearly optimal as a nonsymmetric test. Tables give results when X_i has a variety of distributions.

1.4. Unknown Cost and the Partial Sequential Probability Ratio Test.

If the cost per observation c is unknown, and we wish to minimize the risk, we can replace c by an estimator. The question of the most suitable estimator arises, and is touched on in Appendix II. In the main part of the thesis, we use $\bar{c}_n$, the mean observed cost based on n observations. Chapter IV covers the case in which $\bar{c}_n$ is observed in a previous experiment. The risk is minimized as if the true cost were $\bar{c}_n$, and the excess risk over that attainable for the case in which c is known is calculated. If the observed cost has a Gamma distribution with small variance, then the excess risk is averaged over the distribution of $\bar{c}_n$. The analysis uses Chernoff's approximations [7], for SPRT procedures having at least one observation.

For the case in which no previous estimate of c is available, we have adopted a two-stage procedure. In the first, n observations are made; $\underline{x}_n$ and $\bar{c}_n$ are observed. In the second stage, a Wald SPRT procedure is followed, with boundaries A_n and B_n. n is determined in advance, but the observed cost $\bar{c}_n$ determines A_n and B_n if the risk is to be minimized (Chapter V).

This procedure is termed the Partial Sequential Probability Ratio Test (PSPRT), and is of interest in itself (Chapter III). A conditional test $\mathcal{T}(\underline{x}_n)$ is first defined, given $x_1, x_2, \ldots, x_n$. Sampling continues if

$$B'_n(\underline{x}_n) = B_n \frac{p_{\theta_0}(\underline{x}_n)}{p_{\theta_1}(\underline{x}_n)} < \sum_{j=n+1}^{m} \left\{\frac{p_{\theta_1}(x_j)}{p_{\theta_0}(x_j)}\right\} < A_n \frac{p_{\theta_0}(\underline{x}_n)}{p_{\theta_1}(\underline{x}_n)} = A'_n(\underline{x}_n)$$

where $m > n$.

In order that property $\mathcal{P}$ shall hold, i.e., for sampling to proceed beyond n observations,

$$B'_n(\underline{x}_n) \le 1 \le A'_n(\underline{x}_n) \ .$$

If $B'_n(\underline{x}_n) > 1$, sampling stops with the first stage and H_0 is accepted.

If $A'_n(\underline{x}_n) < 1$, sampling stops with the first stage and H_1 is accepted.

The approximate OC function and ASN of $\mathcal{T}(\underline{x}_n)$ are averaged over the joint distribution of $\underline{x}_n$ to give extensions of Wald's approximations to the unconditional PSPRT. For an exponential family, these are obtained from the univariate distribution of a sufficient statistic $\bar{s}_n$, and the cases in which X_j is a Gamma variable and a normal variable are considered, with numerical analysis in the latter case. Wald's bounds on $L(\theta)$ and on $E(N|\theta)$ are extended to the PSPRT in Chapter VI for the same case.

The PSPRT does not appear to have been studied in the literature, although Billard [5] recently developed modifications of procedures of Sobel and Wald [16] and of Armitage [2] for testing three hypotheses jointly. Her modifications are also analogs of the PSPRT.

Tables show that Anderson's modified SPRT [1] is better than the PSPRT, but it is not appropriate to the problem of unknown cost.

If $h(\theta) = 0$, a fixed sample-size test of equivalent power may require fewer observations than a Wald SPRT on the average (see, for example, [12]). The PSPRT may then be better than either (see Tables 3.5.7 and 3.5.8).

If $n = 1$, the PSPRT is a Wald SPRT. The approximations to $L(\theta)$ and $E(N|\theta)$ derived in Chapter III seem to improve very slightly on Wald's but not as substantially as Kemp's [13] (see Tables 3.5.4 and 3.5.5).

Approximations to the minimum risk, and the average excess risk using the estimator $\bar{c}_n$ are considered in Chapter V; X_j is assumed to be normal with unit variance, $L_0 = L_1$, and $w = \frac{1}{2}$. Numerical results by the method of statistical differentials (shown to be reasonably close when compared with quadrature for small values of c) demonstrate an optimum value of n which minimizes the average excess risk; tables and diagrams are included.

Chapter VIII extends work by Baker [3] and Hall [9] for a two-stage test of the mean of a normally distributed variable with unknown variance. Hall suggested a modified SPRT, replacing the variance by an estimator. The error probabilities have known bounds. This extends to a test of the mean of a multinormally distributed population in which the covariance matrix is known except for a scalar multiplier.

NOTATION.

$\simeq$ denotes approximate equality throughout.

$\sim$ means "is distributed as ..." in referring to random variables.

CHAPTER II

MINIMIZING THE AVERAGE RISK IN A S.P.R.T.

2.1. Introduction.

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables, indexed by a parameter θ. In this chapter, we shall be interested in testing a simple hypothesis H_0: $\theta = \theta_0$ against a simple alternative H_1: $\theta = \theta_1$, where the parameter has a prior distribution λ_w: $\Pr(\theta = \theta_0) = w$ and $\Pr(\theta = \theta_1) = 1-w$. L_0 and L_1 are losses relating to wrong rejection of H_0 and H_1 respectively; c is the cost per observation.

In the sequel, we shall be considering SPRT's of H_0 vs H_1, but the error probabilities considered will be the nominal error probabilities when excess over the boundaries is neglected. To this extent, the results will be approximate, since we shall minimize the risk function using these nominal error probabilities rather than the exact ones. When the error probabilities are small, however, the nominal error probabilities come very close to the exact ones. In general, we shall find that if cost per observation is small compared to loss L_0 or L_1, then the error probabilities appropriate to the minimum risk are small, and in many practical applications this will be so.

Let δ be the SPRT for testing H_0 vs H_1, having α and β as (nominal) error probabilities, and constant cost c per observation. Let $R_\delta(\lambda_w)$ be the risk corresponding to δ when the distribution of the prior on δ is λ_w. Then

$$R_\delta(\lambda_w) = L_0 w\alpha' + L_1(1-w)\beta' + cE(N) , \qquad (2.1.1)$$

where N is the stopping variable, possibly equal to zero, and α', β' the error probabilities.

Using Wald's approximation for $E(N)$, and nominal error probabilities α,β,

$$R_\delta(\lambda_w) \simeq L_0 w\alpha + L_1(1-w)\beta + \frac{cw}{E_0(Z_1)}[(1-\alpha)\log\frac{\beta}{1-\alpha} + \alpha\log\frac{1-\beta}{\alpha}]$$

$$+ \frac{c(1-w)}{E_1(Z_1)}[\beta\log\frac{\beta}{1-\alpha} + (1-\beta)\log\frac{1-\beta}{\alpha}] . \qquad (2.1.2)$$

In (2.1.2), we suppose $X_1, X_2, \ldots$ to be a sequence of i.i.d. random variables, each having a p.d.f. $p_\theta(\cdot)$. If the respective observed values are $x_1, x_2, \ldots$,

$$z_1 = \log \left\{\frac{p_{\theta_1}(x_1)}{p_{\theta_0}(x_1)}\right\}$$

and

$$E_i(Z_1) = E(Z_1 | \theta = \theta_i), \quad i = 0, 1 .$$

We shall consistently use the notation

$$k_0 = \frac{cw}{E_0(Z_1)}, \quad k_1 = \frac{c(1-w)}{E_1(Z_1)} . \tag{2.1.3}$$

It is a property of the likelihood ratio that

$$\int_S \log\left(\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}\right) p_{\theta_0}(x)\, dx < 0$$

and

$$\int_S \log\left(\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}\right) p_{\theta_1}(x)\, dx > 0$$

where S is a set such that $\int_S p_{\theta_i}(x)dx = 1;\ i = 0, 1,$ and such that $p_{\theta_i}(x) = 0$ at most on a set of measure zero in S.

Hence we have, if we also assume $c > 0,\ 0 < w < 1$,

$$\left.\begin{aligned} &k_1 > 0, \quad k_0 < 0 \\ \text{and} \quad &k_1 - k_0 > 0 . \end{aligned}\right\} \tag{2.1.4}$$

We now write $R_\delta(\lambda_w)$ in the form

$$R_\delta(\lambda_w) \simeq L_0 w\alpha + L_1(1-w)\beta + \{k_0(1-\alpha) + k_1\beta\}\log\frac{\beta}{1-\alpha} + \{k_0\alpha + k_1(1-\beta)\}\log\frac{1-\beta}{\alpha} \tag{2.1.5}$$

In problems in discriminant analysis, we may wish to minimize the risk for a given prior λ_w. The experimenter is free to choose nominal error probabilities α and β if he uses a Wald SPRT procedure, by appropriate choice of boundaries. We know that among the class C_0 of all sequential procedures (including those for which decisions are made without making any observations at all), Wald's SPRT is a Bayes rule for given λ_w, α and β [14]. In minimizing the risk for procedures in C_0, therefore, it seems reasonable to restrict attention to the class C_1 of Wald SPRT's. Thus the (approximate) minimum risk in C_1 for given λ_w will be the (approximate) absolute minimum risk in C_0 for given λ_w, which will therefore be termed the "nominal a.m. risk."

The problem will be analyzed in two special cases first; those of testing a multinormal mean with known covariance matrix, and of certain symmetric non-normal tests in which $\beta = \alpha$. Then a procedure for simple tests in the general case is discussed. The chapter concludes with some comparisons.

2.2. The Absolute Minimum Risk for a Multinormal Discrimination Problem with Equal Loss.

Let $\underline{X}_1, \underline{X}_2, \ldots$ be a sequence of mutually independent $p \times 1$ random vectors, each distributed multinormally with mean $\underline{\theta}$ and known variance-covariance matrix $\underline{\Sigma}$. We wish to test $H_0\colon \underline{\theta} = \underline{\theta}_0$ against $H_1\colon \underline{\theta} = \underline{\theta}_1$.

Then the log likelihood ratio is

$$Z_1 = (\underline{X}_1 - \tfrac{1}{2}(\underline{\theta}_1 + \underline{\theta}_0))' \, \underline{\Sigma}^{-1} \, (\underline{\theta}_1 - \underline{\theta}_0) .$$

So

$$E_0(Z_1) = - E_1(Z_1) = - \tfrac{1}{2}(\underline{\theta}_1 - \underline{\theta}_0)' \, \underline{\Sigma}^{-1} \, (\underline{\theta}_1 - \underline{\theta}_0). \qquad (2.2.1)$$

To minimize $R_\delta(\lambda_w)$ by a suitable procedure $\delta' \in C_1$, we have

$$\frac{\partial R}{\partial \alpha} = L_0 w - k_0 \log\frac{\beta}{1-\alpha} + \{k_0(1-\alpha) + k_1\beta\} \frac{1}{1-\alpha} + k_0 \log\frac{1-\beta}{\alpha} - \{k_0\alpha + k_1(1-\beta)\} \frac{1}{\alpha}.$$

If we write $\eta = \frac{\alpha}{1-\beta}$, $\nu = \frac{\beta}{1-\alpha}$, then

$$\frac{\partial R}{\partial \alpha} = 0 \implies L_0 w - k_0 \log(\eta\nu) - k_1(\frac{1}{\eta} - \nu) = 0$$

Similarly (2.2.2)

$$\frac{\partial R}{\partial \beta} = 0 \implies L_1(1-w) + k_1 \log(\eta\nu) + k_0(\frac{1}{\nu} - \eta) = 0$$

Eqs. (2.2.2) apply to the general problem to be solved later. In the meantime we note that (2.2.2) would be easier to deal with if the terms in $\log(\eta\nu)$ could be eliminated.

An interesting approach is to solve the problem first with the restriction

$$w\alpha + (1-w)\beta = \pi \tag{2.2.3}$$

so that the expected error rate π is fixed.

Putting $\psi(\alpha, \beta) = w\alpha + (1-w)\beta$, we can use Lagrange's method to solve

$$\frac{\partial R}{\partial \alpha} + h\frac{\partial \psi}{\partial \alpha} = wL_0 + k_0 \log(\eta\nu) - k_1(\eta - \frac{1}{\nu}) + hw = 0$$

$$\frac{\partial R}{\partial \beta} + h\frac{\partial \psi}{\partial \beta} = (1-w)L_1 - k_1 \log(\eta\nu) + k_0(\nu - \frac{1}{\eta}) + h(1-w) = 0 .$$

Since $wk_1 + (1-w)k_0 = 0$, eliminating h gives an equation free of $\log(\eta\nu)$, i.e.,

$$w(1-w)(L_1-L_0) + wk_0(\frac{1-\alpha}{\beta} - \frac{\alpha}{1-\beta}) - (1-w)k_1(\frac{\beta}{1-\alpha} - \frac{1-\beta}{\alpha}) = 0. \tag{2.2.4}$$

Special Case: $L_0 = L_1 = L$

Solving (2.2.4) with (2.2.3) gives concise solutions when the losses are equal. In this case, (2.2.4) reduces to

$$wk_0\alpha(1-\alpha) + (1-w)k_1\beta(1-\beta) = 0,$$

i.e.,

$$w^2\alpha(1-\alpha) - (1-w)^2\beta(1-\beta) = 0 . \qquad (2.2.5)$$

Solving (2.2.5) with (2.2.3), we get the formal solutions

$$\left.\begin{aligned} \alpha &= \frac{\pi(1 - w - \pi)}{w(1-2\pi)} \\ \beta &= \frac{\pi(w-\pi)}{(1-w)(1-2\pi)} \end{aligned}\right\} \qquad (2.2.6)$$

leading to nominal error probabilities minimizing the risk subject to (2.2.3), for tests having at least one observation.

These results can be incorporated into

THEOREM 2.1. $\underline{X}_1, \underline{X}_2 \cdots$ is a sequence of $p \times 1$ random vectors, mutually independent and having the multinormal $N_p(\underline{\theta}, \underline{\Sigma})$ distribution. If in testing $\underline{\theta} = \underline{\theta}_0$ vs $\underline{\theta} = \underline{\theta}_1$ the expected error rate $\pi = w\alpha + (1-w)\beta$ is fixed, and the losses $L_0 = L_1 = L$, then

(i) $\pi < \max(w, 1-w) \quad \forall\ \delta \in C_1$

(ii) $\pi < \min(w, 1-w) \iff$ in the class C_1, the procedure minimizing $R_\delta(\lambda_w)$ subject to (2.2.3) is a SPRT given by (2.2.6), having at least one observation.

(iii) $\min(w, 1-w) < \pi < \max(w, 1-w) \iff$ in the class C_1, the procedure minimizing $R_\delta(\lambda_w)$ subject to (2.2.3) is a decision made without taking any observations.

Proof: (i) In any SPRT having at least one observation, we assume $\alpha + \beta < 1$. So $\pi = w\alpha + (1-w)\beta < \max(w, 1-w)(\alpha + \beta) < \max(w, 1-w)$. Now suppose we have a randomized decision with no observations. If d_i is the decision which accepts H_i, $i = 0, 1$, then

$$p = \Pr(d_0) = \Pr(d_0 | \underline{\theta} = \underline{\theta}_0) = 1 - \alpha$$

$$= \Pr(d_0 | \underline{\theta} = \underline{\theta}_1) = \beta, \text{ so that for such a decision,}$$

$$\alpha + \beta = 1.$$

Then (2.2.3) $\Longrightarrow$ $p = \frac{\pi - w}{1-2\pi}$ if $w \neq \frac{1}{2}$, and $p = \frac{1}{2}$ if $w = \frac{1}{2}$

i.e., $p \geq 0 \Longrightarrow \pi \geq w$ and $w < \frac{1}{2}$

or $\pi \leq w$ and $w > \frac{1}{2}$

and
$$p \leq 1 \Longrightarrow \begin{cases} \pi \leq 1-w \text{ and } w < \frac{1}{2}, & \text{or} \\ \pi \geq 1-w \text{ and } w > \frac{1}{2}. \end{cases}$$

So for such a procedure, π lies between w and $1-w$; the risk is $L\pi$. This proves (i), and sufficiency in (iii).

(ii) We showed above that for SPRTs having one observation, compatable α and β are given by (2.2.6). Then

$$\alpha \geq 0 \text{ in (2.2.6)} \Longrightarrow \begin{cases} 1-w - \pi \geq 0 \text{ and } \pi < \frac{1}{2}, & \text{or} \\ 1-w - \pi \leq 0 \text{ and } \pi > \frac{1}{2} \end{cases}$$

$$\beta \geq 0 \text{ in (2.2.6)} \Longrightarrow \begin{cases} w - \pi \geq 0 \text{ and } \pi < \frac{1}{2}, & \text{or} \\ w - \pi \leq 0 \text{ and } \pi > \frac{1}{2} \end{cases}$$

$\pi > \frac{1}{2}$ is not admissible, since then $\pi > \max(w, 1-w)$. Hence $\pi < \frac{1}{2}$, $\Longrightarrow$ $\pi \leq \min(w, 1-w)$. It can be checked that $\alpha \leq 1$, $\beta \leq 1$, $\alpha + \beta < 1$ if $\pi \leq \min(w, 1-w)$ in (2.2.6).

To prove necessity in (ii) and (iii), we need only show that if the minimizing procedure subject to (2.2.3) is not a SPRT having at least one observation, then $\pi \nless \min(w, 1-w)$. But this must be so, since such a procedure must then be based on no observations, implying what we showed in (iii).

This completes the proof of the theorem, except that we need to establish that the solution (2.2.6) in (ii) does give a minimum.

Expressing the risk in terms of β only,

$$R_\delta(\lambda_w) = L\pi + \frac{c}{E_1(Z_1)}[\{(1-w)\beta - w + (\pi - (1-w)\beta)\beta\} \log \frac{w\beta}{w-\pi+(1-w)\beta}$$

$$+ \{(1-\beta)(1-w) - (\pi - (1-w)\beta)\} \log \frac{w(1-\beta)}{\pi - (1-w)\beta}].$$

So

$$\frac{E_1(Z_1)}{c}\frac{\partial R}{\partial \beta} = -(w-\pi)\left\{\frac{w-\pi}{\beta\{w-\pi+(1-w)\beta\}}\right\} + (1-w-\pi)\left\{\frac{1-w-\pi}{(1-\beta)(\pi-(1-w)\beta)}\right\}.$$

For β small and $\pi < \min(w, 1-w)$, $\frac{\partial R}{\partial \beta} < 0$.

For $1-\beta$ small and $\pi < \min(w, 1-w)$, $\frac{\partial R}{\partial \beta} > 0$.

Since there is only one turning value for R, this gives the required minimum when (2.2.6) holds. Q.E.D.

Situations in sequential analysis could arise in which the experimenter would be more interested in fixing the expected error rate than in fixing both error probabilities. Such an approach would only apply if a prior distribution for H_0 and H_1 were known, but as shown in the theorem for this problem of testing a multinormal mean, it leaves scope for flexibility in choosing α and β to reduce the risk involved.

The nominal minimum risk $R_\delta(\lambda_w, \pi)$, subject to (2.2.3), can be expressed in terms of π. For when (2.2.6) holds,

$$(1-w)\beta - w(1-\alpha) = -(w-\pi),$$

$$(1-w)(1-\beta) - w\alpha = 1 - w - \pi,$$

$$\log \frac{\beta}{1-\alpha} = \log\left\{\frac{w}{(1-w)} \cdot \frac{\pi}{(1-\pi)}\right\},$$

and

$$\log \frac{1-\beta}{\alpha} = \log\left\{\frac{w}{(1-w)} \cdot \frac{(1-\pi)}{\pi}\right\}.$$

Then we substitute in

$$R_\delta(\lambda_w,\pi) = \pi L + \frac{c}{E_1(Z_1)}[\{(1-w)\beta-w(1-\alpha)\}\log\frac{\beta}{1-\alpha} + \{(1-\beta)(1-w) - w\alpha\}\log\frac{1-\beta}{\alpha}]$$

to get the solution

$$R_\delta(\lambda_w,\pi) = \begin{cases} \pi L + \frac{c}{E_1(Z_1)}\{(1-2\pi)\log\frac{1-\pi}{\pi} - (1-2w)\log\frac{1-w}{w}\}, & \pi < \min(w, 1-w) \\ \pi L, & |\pi - \frac{1}{2}| < |w - \frac{1}{2}|. \end{cases} \tag{2.2.7}$$

Now that the problem is solved for fixed expected error rate π, it remains to evaluate the nominal a.m. risk by minimizing $R_\delta(\lambda, \pi)$ with respect to π. For the case $|\pi - \frac{1}{2}| < |w - \frac{1}{2}|$, the risk πL is minimized with a greatest lower bound $L \times \min(w, 1-w)$, at $\pi = \min(w, 1-w)$.

For $\pi < \min(w, 1-w)$, the function in (2.2.7) is to be minimized. $\frac{d}{d\pi} R_\delta(\lambda_w, \pi) = 0$ gives

$$\phi(\pi) = \rho^{-1} - 2\log\frac{1-\pi}{\pi} + (1-2\pi)(-\frac{1}{\pi} - \frac{1}{1-\pi}) = 0$$

or

$$\phi(\pi) = \rho^{-1} - 2\log\frac{1-\pi}{\pi} - \frac{1-2\pi}{\pi(1-\pi)} = 0 \tag{2.2.8}$$

where

$$\rho = \frac{c}{LE_1(Z_1)} = \frac{2c}{L\Delta^2}. \tag{2.2.9}$$

The value of π required is a root of (2.2.8). The Newton-Raphson iterative method can be used to solve for π, noting that ρ is free of w, and that therefore the prior distribution appears only in the boundary values for π in (2.2.7).

$$\frac{d\phi}{d\pi} = \frac{2}{\pi} + \frac{1}{\pi^2} + \frac{2}{1-\pi} + \frac{1}{(1-\pi)^2}$$

$$= [\frac{1}{\pi(1-\pi)}]^2 . \qquad (2.2.10)$$

Thus $\frac{d\phi}{d\pi} > 0$ for $0 < \pi < 1$, and

$$\inf_{0 < \pi < 1} \frac{d\phi}{d\pi} = 16 \quad \text{at} \quad \pi = \frac{1}{2} .$$

Hence ϕ is monotonic increasing from $-\infty$ to $+\infty$ for π in $(0,1)$, and the method will give a rapid convergence. Further, the function ϕ is skew-symmetric about the point $(\frac{1}{2}, \rho^{-1})$. If the cost c is small compared with L or $E_1(Z_1)$, then ρ is small, and the single root of (2.2.8) will be small. Figure 2.2.2 shows a graph of ϕ, and Table 2.2.1 gives values of ρ^{-1} which correspond to roots π of (2.2.8).

For values of ρ other than $\frac{1}{40}$, the graph of ϕ in Figure I 'slides' up or down but is similar in shape. The linearity on log paper suggests that $\rho = H\pi^r$ for some H and r, and certainly when π is small, $\pi \simeq \rho^{-1}$. A good first approximation to the solution of (2.2.8) was given by Chernoff [7], who showed that

$$\alpha \simeq \frac{(1-w)c}{E_1(Z_1)L_0 w} , \qquad \beta \simeq \frac{wc}{E_0(Z_1)L_1(1-w)} \qquad (2.2.11)$$

give an approximate a.m. risk, using Wald's crude approximations, for procedures in C_1 having at least one observation. Here this gives a

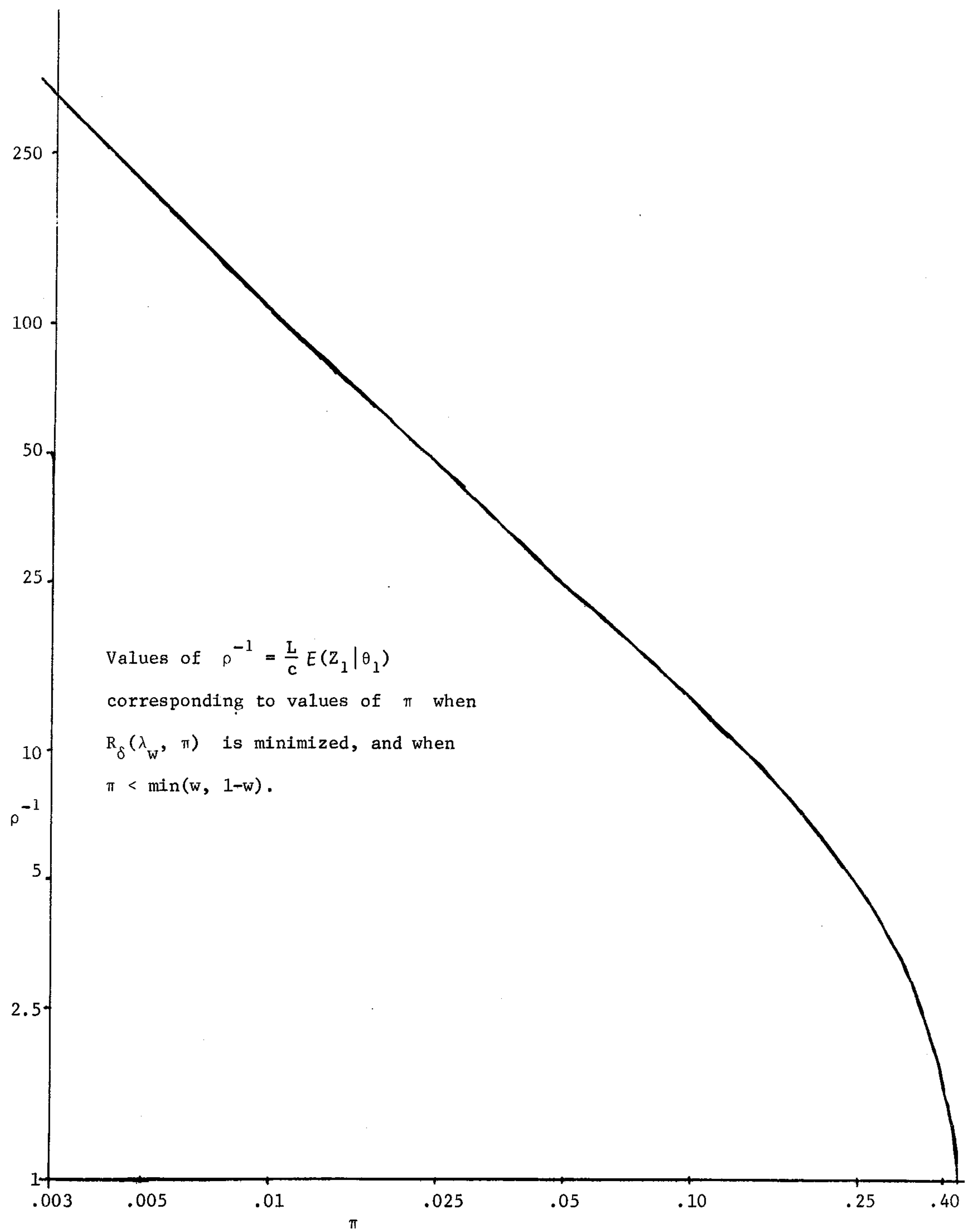
250
100
50
25
10
5
2.5
1
ρ^{-1}
.003
.005
.01
.025
.05
.10
.25
.40
π
Values of $\rho^{-1} = \frac{L}{c} E(Z_1 | \theta_1)$
corresponding to values of π when
$R_\delta(\lambda_w, \pi)$ is minimized, and when
$\pi < \min(w, 1-w)$.

FIGURE 2.2.2. $\phi(\pi)$ vs π, where $\frac{1}{\rho} = 40$

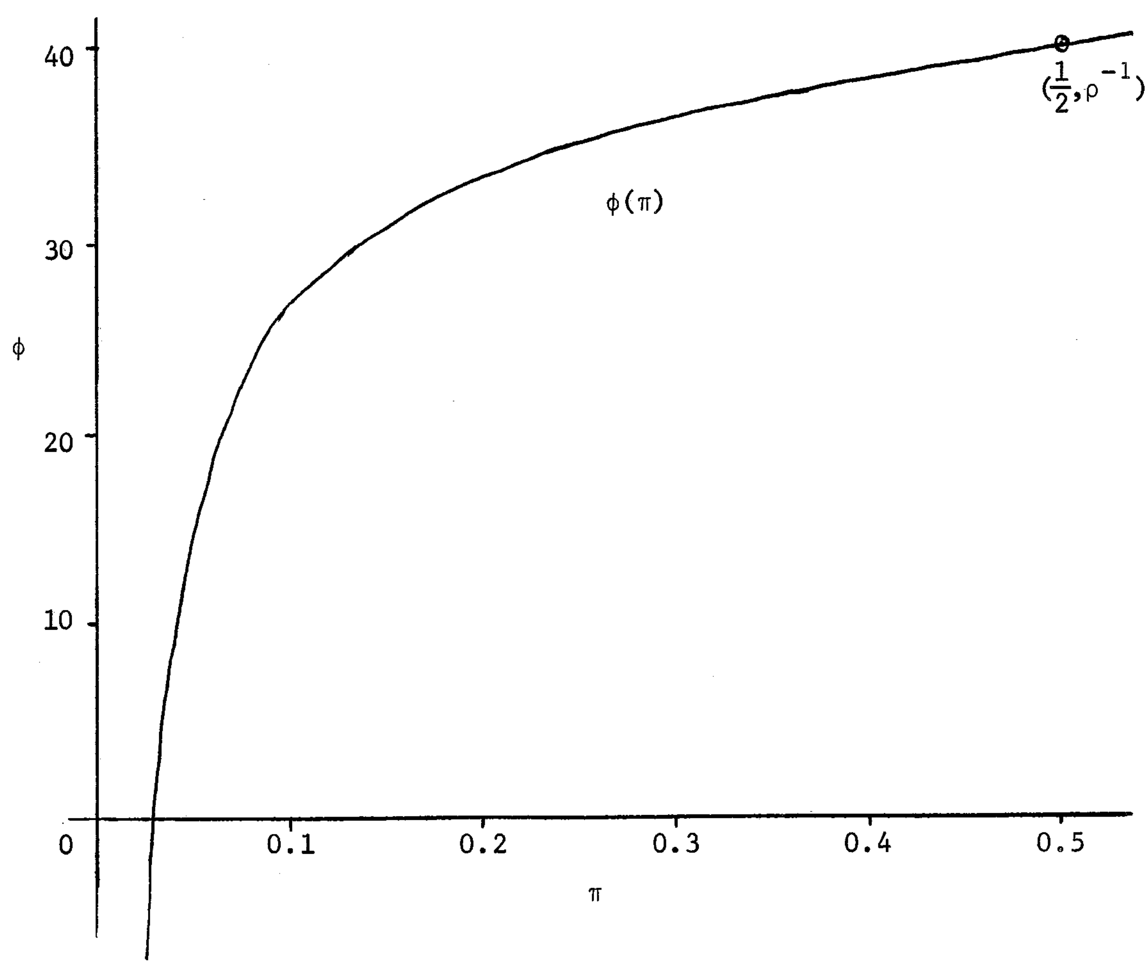

TABLE 2.2.1.

Values of π minimizing $R_\delta(\lambda_w,\pi)$ for given ρ: $\pi < \min(w, 1-w)$

π	.001	.0025	.005	.01	.05	.10	.15	.20	.30	.40	.5
ρ^{-1}	1012.81	409.98	209.58	108.18	24.84	13.28	8.960	6.523	3.599	1.644	0.0

first approximate root

$$\pi_1 = \frac{(1-w)c}{LE_1(Z_1)} + \frac{wc}{E_1(Z_1)L} = \frac{c}{E_1(Z_1)L} = \rho \qquad (2.2.12)$$

The nominal risk arising from $\pi_1 = \rho$ will be called the "Chernoff a.m. risk." The risk arising from the solution π^* to (2.2.8) will be called the "nominal a.m. risk."

A second approximation, by Newton's method, is

$$\pi_2 = \pi_1 - \frac{\phi(\pi_1)}{\phi'(\pi_1)}$$

i.e.,

$$\pi_2 = \rho - \{\rho(1-\rho)\}^2 \{\frac{1}{\rho} - 2\log\frac{1-\rho}{\rho} - \frac{1-2\rho}{\rho(1-\rho)}\}$$

$$= \rho - \rho^2(1-\rho)^2 \{\frac{1}{1-\rho} - 2\log\frac{1-\rho}{\rho}\} . \qquad (2.2.13)$$

If ρ is small, we can write

$$\pi_2 = \rho - \rho^2(1 + 2\log\rho) + O(\rho^3\log\rho). \qquad (2.2.14)$$

If ρ is large, and $\pi \simeq \frac{1}{2}$, it is conceivable that $\pi_1 = \rho$ may be a poor first approximation to the root π^* of $\phi(\pi) = 0$, leading to $\pi_2 < 0$. If $\pi_2 < 0$, we replace π_1 by $p\rho$ for $0 < p < 1$, say $p = 0.25$; then we start the iterative process again. But if $\pi_2 > 0$, it can be shown (see Appendix, I) that $\pi^* = \pi_2 + O(\rho^3\log\rho)$ in the sense that the n-th approximation π_n is of this form. Eventually, and usually quickly, there is rapid convergence to any disired tolerance.

Having obtained π^*, the analysis leading to (2.2.7) is then applied. The nominal a.m. risk is then $R_\delta(\lambda_w,\pi^*)$, with nominal error probabilities given by substituting π^* for π in (2.2.6), provided that $\pi^* < \min(w, 1-w)$. But if π^* lies between w and $1-w$, then the minimum risk is

$$\min[Lw,\ L(1-w)]$$

using a procedure having no observations, where

$$p = \Pr(d_0) = 0 \text{ or } 1.$$

We state the conclusions in

THEOREM 2.2. For the multinormal problem defined in Theorem 2.1., let π^* be the root of $\phi(\pi) = \rho^{-1} - 2\log\frac{1-\pi}{\pi} - \frac{1-2\pi}{\pi(1-\pi)} = 0$. Then the nominal a.m. risk among all procedures $\delta \in C_0$ is defined by

$$R^*_\delta(\lambda_w) = \begin{cases} \pi^* L + \frac{c}{E_1(Z_1)}\{(1-2\pi^*)\log\frac{1-\pi^*}{\pi^*} - (1-2w)\log\frac{1-w}{w}\}, & \pi^* < \min(w,1-w) \\ L \times \min(w,\ 1-w), & |\pi^* - \frac{1}{2}| < |w - \frac{1}{2}|\ . \end{cases} \tag{2.2.15}$$

It can be checked that when $\pi^* = \min(w, 1-w)$, the two forms agree, i.e. for the a.m. procedure, Wald's approximation to $E(N)$ is zero at $\pi^* = \min(w, 1-w)$.

Note: $\inf_{\lambda_w} R_\delta(\lambda_w,\pi)$ in (2.2.7) is attained when $w = \frac{1}{2}$, since

$-(1-2w)\log\frac{1-w}{w}$ decreases as w increases from 0 to $\frac{1}{2}$

$(\frac{d}{dw}[-(1-2w)\log\frac{1-w}{w}] = \phi(w))$.

Then δ' is a SPRT with at least one observation, and from (2.2.6) the minimizing nominal error probabilities are $\alpha^* = \beta^* = \pi^*$.

2.3. The Absolute Minimum Risk for Symmetric Nonnormal Tests.

We consider procedures $\delta' \in C_1$, not restricted to the multinormal case, in which the error probabilities α and β are equal, in tests of a simple $H_0: \theta = \theta_0$ against a simple alternative $H_1: \theta = \theta_1$, where the losses L_0 and L_1 may be unequal, and δ' has at least one observation.

(2.1.5) may be expressed

$$R_{\delta'}(\lambda_w) \simeq [L_0 w + L_1(1-w)]\alpha + (k_0 - k_1)(1-2\alpha)\log\frac{\alpha}{1-\alpha} . \qquad (2.3.1)$$

To minimize, we have

$$\frac{\partial R}{\partial \alpha} = L_0 w + L_1(1-w) - 2(k_1 - k_0)\log\frac{1-\alpha}{\alpha} - \frac{1-2\alpha}{\alpha(1-\alpha)}(k_1 - k_0).$$

So $\frac{\partial R}{\partial \alpha} = 0$ gives

$$\phi(\alpha) = \frac{1}{k^*} - 2\log\frac{1-\alpha}{\alpha} - \frac{1-2\alpha}{\alpha(1-\alpha)} = 0 . \qquad (2.3.2)$$

This equation has the same form as (2.2.8), with

$$k^* = \frac{k_1 - k_0}{L_0 w + L_1(1-w)} \qquad (2.3.3)$$

instead of ρ. We note that $k^* > 0$ by virtue of (2.1.4), and since $\rho > 0$, the formal analysis for solving (2.3.2) is the same as discussed previously.

A first approximation to α is required, and the argument used by Chernoff [7] for general tests is applied. The error probability is supposed small, and Wald's crude approximations are used, so that

$$E(N/\theta_0) \simeq +\frac{1}{E_0(Z_1)}\log\beta, \quad E(N/\theta_1) \simeq -\frac{1}{E_1(Z_1)}\log\alpha. \qquad (2.3.4)$$

Putting $\beta = \alpha$, (2.1.2) gives

$$R_{\delta'}(\lambda_w) \simeq \{L_0 w + L_1(1-w)\}\alpha + \frac{cw}{E_0(Z_1)} \log\alpha - \frac{c(1-w)}{E_1(Z_1)} \log\alpha$$

$$= \{L_0 w + L_1(1-w)\}\alpha - (k_1 - k_0)\log \alpha. \qquad (2.3.5)$$

Then

$$\frac{\partial R}{\partial \alpha} \simeq L_0 w + L_1(1-w) - \frac{k_1 - k_0}{\alpha}$$

$$= 0 \quad \text{at}$$

$$\alpha_1 = \frac{k_1 - k_0}{L_0 w + L_1(1-w)} = k^* . \qquad (2.3.6)$$

Since earlier $\pi_1 = \rho$ gave (2.12.13) and (2.12.14), $\alpha_1 = k^*$ leads to the improved approximation α_2 to the root α of (2.3.2)

$$\alpha_2 = k^* - k^{*2}(1-k^*)^2 \{\frac{1}{1-k^*} - 2\log \frac{1-k^*}{k^*} \}$$

$$= k^* - k^{*2}(1-2\log k^*) + 0(k^{*3}\log k^*) \qquad (2.3.7)$$

$\alpha + \beta < 1 \implies \alpha < \frac{1}{2}$, and numerical results indicate that if $\alpha_1 = k^* > \frac{1}{2}$, it can be replaced by $\alpha_1 = 0.45$ as a suitable first approximation.

Hence the method of solution is straightforward, and covers a wide class of problems, including those with unequal loss. The risk (2.3.1) arising from $\alpha_1 = k^*$ will be called the "Chernoff symmetric a.m. risk," and the risk (2.3.1) arising from the solution of (2.3.2) will be called the "nominal symmetric a.m. risk."

For the problem of testing a multinormal mean with equal loss, as considered in 2.2, the case in which $w = \frac{1}{2}$ yields the above symmetric solution, since (2.2.6) gives $\alpha = \beta$ when $w = \frac{1}{2}$.

Tables 2.3.1 to 2.3.4 show solutions α to (2.3.2), together with the initial $\alpha_1 = k^*$, the Chernoff symmetric a.m. risk, the nominal symmetric a.m. risk, and the nominal a.m. risk for $\beta \neq \alpha$ as discussed in the next section. When c is small, the Chernoff risk is close to the iterated solution; and the iterated solution appears to come closest to the non-symmetric nominal a.m. risk whenever $w = \frac{1}{2}$.

The question may be raised: does a non-symmetric test procedure give a substantial improvement in the a.m. risk over a symmetric procedure? The computed results for Bernoulli and Gamma distributions with equal loss indicate that a symmetric procedure is effectively as good when $w = \frac{1}{2}$, but not necessarily otherwise. This statement appears to hold whether the cost is large or small. Further discussion appears later in 2.5.

If there is no prior information about θ, the experimenter might well take θ_0 to be as equally likely a priori as θ_1, and the numerical results suggest that one would do almost as well with a symmetric test.

In the case of normality, $E_0(Z) = -E_1(Z) \Rightarrow k_1 - k_0 = c/E_1(Z)$. If $L_0 = L_1 = 1$ in addition, then the risk (2.3.1) is free of λ_w. Hence in this case the nominal symmetric a.m. risk is invariant under changes in λ_w, and is effectively the same as the non-symmetric nominal a.m. risk for $w = \frac{1}{2}$.

Tables 2.3. * means the 'no observations' rule is best. $L_0 = L_1 = 1$ in all cases.

TABLE 2.3.1

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{1}{2\theta^2} x^2} ; \quad \theta_0 = 1, \quad \theta_1 = 2$$

w	c	α Chernoff	α nominal	EN Chernoff	EN nominal	Symmetric a.m. Risk Chernoff	Symmetric a.m. Risk nominal	Nominal a.m. risk
.25	.1	.1715	.2190	3.024	1.226	.4739	.3416	.2499
.5	.1	.2191	.2609	3.327	1.092	.5518	.37002	.36947
.75	.1	.2667	.2935	3.525	.968	.6192	.3903	.25*
.25	.01	.0172	.0194	6.974	6.465	.0869	.0841	.0798
.5	.01	.0219	.0254	8.372	7.586	.1056	.10126	.09985
.75	.01	.0267	.0316	9.666	8.555	.1233	.1171	.0958
.25	.001	.00172	.00175	10.924	10.848	.0126	.0126	.0123
.5	.001	.0022	.0023	13.418	13.299	.0156	.01555	.01534
.75	.001	.0027	.0028	15.808	15.636	.0185	.0184	.0160
.25	.0001	.00017	.00017	14.873	14.863	.00166	.00166	.00163
.5	.0001	.00022	.00022	18.463	18.447	.00207	.002065	.002043
.75	.0001	.00027	.00027	21.950	21.93	.00246	.00246	.00222

TABLE 2.3.2

$p_\theta(x) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$; $\theta_0 = 1$, $\theta_1 = 1.2$ {For $c \geq .1$, the 'no observations' procedure is nominally the best.

w	c	α Chernoff	α nominal	EN Chernoff	EN nominal	a.m. Risk Chernoff	a.m. Risk nominal	Nominal a.m. risk
.25	.01		.4019		4.842	*	.4503	
.5	.01		.3991		4.977	*	.4488	
.75	.01		.3661		5.119	*	.4473	
.25	.001	.0621	.0810	172.49	126.30	.2345	.2073	.1719
.5	.001	.0602	.0784	169.21	125.17	.2294	.20353	.20351
.75	.001	.0584	.0757	165.87	123.96	.2242	.1997	.1691
.25	.0001	.0310	.0374	215.50	186.56	.1388	.1306	.1125
.5	.0001	.0301	.0361	210.95	183.47	.1356	.12786	.12784
.75	.0001	.0292	.0349	206.35	180.31	.1324	.1251	.1101

TABLE 2.3.3

$$p_\theta(x) = \theta^2 x e^{-\theta x}, \quad x > 0, \quad \theta > 0; \quad \theta_0 = 1, \quad \theta_1 = 2$$

w	c	α		EN		Symmetric a.m. Risk		Nominal
		Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	a.m. Risk
.25	.1	.2349	.2726	3.403	1.049	.5752	.3775	*
.5	.1	.2109	.2544	3.282	1.114	.5392	.36580	.36564
.75	.1	.1869	.2337	3.135	1.183	.5004	.3519	*
.25	.01	.0235	.0274	8.811	7.922	.1116	.1067	.0905
.5	.01	.0211	.0244	8.139	7.404	.1025	.09840	.09803
.75	.01	.0187	.0213	7.439	6.846	.0931	.0898	.0824
.25	.001	.0024	.0024	14.220	14.084	.0166	.0165	.0148
.5	.001	.0021	.0022	12.995	12.884	.0151	.01504	.01499
.75	.001	.00187	.00191	11.743	11.654	.0136	.0136	.0129
.25	.0001	.00024	.00024	19.628	19.610	.0022	.0022	.0020
.5	.0001	.00021	.00021	17.851	17.836	.0020	.001995	.001990
.75	.0001	.000187	.000187	16.047	16.035	.00179	.00179	.00172

TABLE 2.3.4

$$Pr(x = {1 \atop 0}) = \begin{cases} \theta \\ 1 - \theta \end{cases}, \quad 0 < \theta < 1 \qquad \theta_0 = .5, \quad \theta_1 = .7$$

w	c	α Chernoff	α nominal	EN Chernoff	EN nominal	Symmetric a.m. Risk Chernoff	Symmetric a.m. Risk nominal	Nominal a.m. Risk
.25	.1					*	*	
.5	.1		.4475		0.262	*	.47364	.47364
.75	.1					*	*	
.25	.01	.1198	.1606	25.42	13.456	.3740	.2951	.2287
.5	.01	.1181	.1584	25.23	13.480	.3704	.29318	.29318
.75	.01	.1164	.1562	25.04	13.501	.3668	.2912	.2279
.25	.001	.0120	.0132	53.02	50.35	.0650	.0635	.0567
.5	.001	.0118	.0130	52.43	49.83	.0642	.06281	.06281
.75	.001	.0116	.0128	51.84	49.31	.0635	.06209	.0559
.25	.0001	.00120	.00122	80.61	80.22	.0093	.0092	.0086
.5	.0001	.00118	.00120	79.63	79.25	.0091	.00912	.00912
.75	.0001	.00116	.00118	78.65	78.28	.0090	.0090	.0084

2.4. The General Case.

We now return to the problem of minimizing (2.1.5), viz.

$$R_\delta(\lambda_w) = L_0 w\alpha + L_1(1-w)\beta + \{k_0(1-\alpha) + k_1\beta\}\log\frac{\beta}{1-\alpha} + \{k_0\alpha + k_1(1-\beta)\}\log\frac{1-\beta}{\alpha} .$$

We seek solutions of equations (2.2.2); if $\eta = \frac{\alpha}{1-\beta}$, and $\nu = \frac{\beta}{1-\alpha}$, these are $\phi(\eta, \nu) = \frac{\partial R}{\partial \alpha} = L_0 w - k_0 \log(\eta\nu) - k_1(\frac{1}{\eta} - \nu) = 0$,
$\psi(\eta, \nu) = \frac{\partial R}{\partial \beta} = L_1(1-w) + k_1\log(\eta\nu) + k_0(\frac{1}{\nu} - \eta) = 0.$

We note that if $\alpha + \beta < 1$ for Wald's SPRT, then

$$0 < \eta < 1, \quad 0 < \nu < 1. \tag{2.4.1}$$

Also

$$\left.\begin{aligned} \alpha &= \frac{\eta(1-\nu)}{1-\eta\nu} < \eta \\ \beta &= \frac{\nu(1-\eta)}{1-\eta\nu} < \nu. \end{aligned}\right\} \tag{2.4.2}$$

We apply the Newton-Raphson method for two unknowns to solve for η and ν. If (η_0, ν_0) is a first approximation to the roots of (2.2.2), then we take as our second approximation in (η, ν, Z)-space the intersection with $Z = 0$ of the tangent planes to the surfaces $Z = \phi(\eta, \nu)$ at $(\eta_0, \nu_0, \phi(\eta_0, \nu_0))$ and $Z = \psi(\eta, \nu)$ at $(\eta_0, \nu_0, \psi(\eta_0, \nu_0))$.

So we solve

$$\left.\begin{aligned} \phi(\eta_0, \nu_0) + \varepsilon\left(\frac{\partial\phi}{\partial\eta}\right)_0 + \varepsilon'\left(\frac{\partial\phi}{\partial\nu}\right)_0 &= 0 \\ \psi(\eta_0, \nu_0) + \varepsilon\left(\frac{\partial\psi}{\partial\eta}\right)_0 + \varepsilon'\left(\frac{\partial\psi}{\partial\nu}\right)_0 &= 0 \end{aligned}\right\} \tag{2.4.3}$$

for ε and ε', where $\left(\frac{\partial\phi}{\partial\eta}\right)_0$ is the value of $\frac{\partial\phi}{\partial\eta}$ at (η_0, ν_0), etc.

Then, $\left.\begin{aligned} \eta_1 &= \eta_0 + \varepsilon \\ \nu_1 &= \nu_0 + \varepsilon' \end{aligned}\right\}$ gives a second approximation, provided $0 < \eta_1 < 1$, and $0 < \nu_1 < 1$. In solving (2.4.3), we need that the determinant $D \neq 0$, where

$$D = \begin{vmatrix} \frac{\partial\phi}{\partial\eta}\Big)_0 & \frac{\partial\phi}{\partial\nu}\Big)_0 \\ \frac{\partial\psi}{\partial\eta}\Big)_0 & \frac{\partial\psi}{\partial\nu}\Big)_0 \end{vmatrix},$$

$$= \begin{vmatrix} -\frac{k_0}{\eta_0} + \frac{k_1}{\eta_0^2} & -\frac{k_0}{\nu_0} + k_1 \\ -k_0 + \frac{k_1}{\eta_0} & -\frac{k_0}{\nu_0^2} + \frac{k_1}{\nu_0} \end{vmatrix}$$

i.e.,

$$D = \left(\frac{k_1}{\eta_0} - k_0\right)\left(k_1 - \frac{k_0}{\nu_0}\right)\left(\frac{1}{\eta_0\nu_0} - 1\right) . \qquad (2.4.4)$$

Hence $D > 0$, since $0 < \eta_0 < 1$, $0 < \nu_0 < 1$ from (2.4.1), and $k_1 > 0$, $k_0 < 0$ by (2.1.4); thus each factor in (2.4.4) is strictly positive.

The second approximation to the set of minimizing error probabilities is then derived from

$$\left.\begin{aligned} \eta_1 &= \eta_0 + \frac{\begin{vmatrix} -\phi(\eta_0,\nu_0) & \frac{\partial\phi}{\partial\nu}\Big)_0 \\ -\psi(\eta_0,\nu_0) & \frac{\partial\psi}{\partial\nu}\Big)_0 \end{vmatrix}}{D} \\ \nu_1 &= \nu_0 + \frac{\begin{vmatrix} \frac{\partial\phi}{\partial\eta}\Big)_0 & -\phi(\eta_0,\nu_0) \\ \frac{\partial\psi}{\partial\eta}\Big)_0 & -\psi(\eta_0,\nu_0) \end{vmatrix}}{D} \end{aligned}\right\} . \qquad (2.4.5)$$

The Chernoff Approximation, [7].

Using Wald's cruder approximations to the ASN

$$E(N|\theta_0) \simeq \frac{\log \beta}{E_0(Z_1)}, \qquad E(N|\theta_1) = -\frac{\log \alpha}{E_1(Z_1)}, \tag{2.4.6}$$

we have

$$R_\delta(\lambda_w) \simeq L_0 w\alpha + L_1(1-w)\beta + k_0 \log \beta - k_1 \log \alpha .$$

Then

$$\left.\begin{aligned} \frac{\partial R}{\partial \alpha} &\simeq L_0 w - \frac{k_1}{\alpha} \\ \frac{\partial R}{\partial \beta} &\simeq L_1(1-w) + \frac{k_0}{\beta} \end{aligned}\right\}, \text{ giving } \left\{\begin{aligned} \alpha_0 &= \frac{k_1}{L_0 w} \\ \beta_0 &= -\frac{k_0}{L_1(1-w)} \end{aligned}\right. \tag{2.4.7}$$

as first approximations, provided that $\alpha_0 + \beta_0 < 1$. If $\alpha_0 + \beta_0 \geq 1$, we start again with α_0 and β_0 replaced by $p\alpha_0$ and $p\beta_0$, where $p < (\alpha_0 + \beta_0)^{-1}$. Unless w or $1-w$ is small, however, in most practical situations the cost will be small compared to L_0 or to L_1, and will be of the same or smaller order than $|E_0(Z_1)|$ or $|E_1(Z_1)|$; then α_0 and β_0 will be small. Cases where w or $1-w$ is small are of less interest, and often lead to solutions for α^* and β^* with $\alpha^* + \beta^* > 1$, and the minimum risk would then be taken by making a decision without taking any observations at all.

α_0 and β_0 give η_0 and ν_0; in turn (2.4.5) and (2.4.2) give α_1 and β_1 as second approximations. As before, if $\alpha_1 < 0$, it is necessary to start again with α_0 replaced by $p\alpha_0$; if $\beta_1 < 0$, replace β_0 by $p\beta_0$. Eventually the convergence process should commence within the unit square $0 < \eta < 1$, $0 < \nu < 1$.

Before continuing further, we refer to Tables 2.4, which give the minimizing error probabilities and nominal a.m. risks for a number of hypothesis-testing problems with $L_0 = L_1 = L$ and for differing values of c and w. The Chernoff approximations α_0 and β_0 in (2.4.7) are entered, as are the risks (2.1.5) based upon them; these risks turn out to be close to the nominal a.m. risks evaluated by computer, and a discussion of this follows later. Originally, Chernoff approximated the minimum risk by

$$\begin{aligned} R_0(\lambda_w) &= cwE_0(N|\alpha_0, \beta_0) + c(1-w)\ E_1(N|\alpha_0, \beta_0) \\ &= (k_1 - k_0) \log \frac{1}{c} \end{aligned}$$

using (2.4.6). Computer results show this not to be a better approximation to the a.m. risk. What we wish to make clear is that Chernoff's approximations (2.4.7) to the minimizing error probabilities do give a close value to the a.m. risk when entered into $R_\delta(\lambda_w)$ in (2.1.5). A discussion on why this is so follows later in this section. For the moment, we notice from the tables that although Chernoff's a.m. risk is close to the computed value, the appropriate error probabilities are not necessarily close.

To show that the stationary point (α^*, β^*) does give a minimum for $R_\delta(\lambda_w)$, we examine $\frac{\partial^2 R}{\partial \alpha^2}$ and $[\frac{\partial^2 R}{\partial\alpha\partial\beta}]^2 - \frac{\partial^2 R}{\partial\alpha^2} \cdot \frac{\partial^2 R}{\partial\beta^2}$ at (α^*, β^*); we show that (α^*, β^*) gives a minimum when α^* and β^* are small.

$$\frac{\partial R}{\partial \alpha} = \phi(\eta, \nu); \qquad \frac{\partial R}{\partial \beta} = \psi(\eta, \nu).$$

Under the usual conditions of continuity, and existence of partial derivatives,

$$\frac{\partial^2 R}{\partial \alpha^2} = \frac{\partial \phi}{\partial \eta}\frac{\partial \eta}{\partial \alpha} + \frac{\partial \phi}{\partial \nu}\frac{\partial \nu}{\partial \alpha}$$

$$= \left(\frac{k_1}{\eta^2} - \frac{k_0}{\eta}\right)\frac{1}{1-\beta} + \left(k_1 - \frac{k_0}{\nu}\right)\left(-\frac{\beta}{(1-\alpha)^2}\right)$$

> 0 for small α and β (and so for small η and ν) since the first term dominates.

In a similar manner,

$$\frac{\partial^2 R}{\partial \alpha \partial \beta} = -\frac{\alpha}{(1-\beta)^2}\left(\frac{k_1}{\eta^2} - \frac{k_0}{\eta}\right) + \frac{1}{1-\alpha}\left(k_1 - \frac{k_0}{\nu}\right)$$

$$\frac{\partial^2 R}{\partial \beta^2} = -\frac{\alpha}{(1-\beta)^2}\left(\frac{k_1}{\eta} - k_0\right) + \frac{1}{1-\alpha}\left(\frac{k_1}{\nu} - \frac{k_0}{\nu^2}\right) .$$

Hence

$$\left[\frac{\partial^2 R}{\partial \alpha \partial \beta}\right]^2 - \frac{\partial^2 R}{\partial \alpha^2}\cdot\frac{\partial^2 R}{\partial \beta^2} = \left[-\frac{k_1 - k_0\eta}{\eta}\cdot\frac{1}{1-\beta} + \frac{k_1\nu - k_0}{\nu(1-\alpha)}\right]^2$$

$$- \left[\frac{k_1 - k_0\eta}{\eta^2(1-\beta)} - \frac{k_1\nu - k_0}{1-\alpha}\right]\left[-\frac{k_1 - k_0\eta}{1-\beta} + \frac{k_1\nu - k_0}{\nu^2(1-\alpha)}\right]$$

$$= \left[-\frac{u}{\eta} + \frac{v}{\nu}\right]^2 - \left[\frac{u}{\eta^2} - v\right]\left[-u + \frac{v}{\nu^2}\right] ,$$

say, where

$$u = \frac{k_1 - k_0\eta}{1-\beta}, \quad v = \frac{k_1\nu - k_0}{1-\alpha}; \quad u > 0, \quad v > 0.$$

When η and ν (hence α and β) are small, this quantity is dominated by terms in $(\eta^2\nu^2)^{-1}$, $(\eta^2\nu)^{-1}$ and $(\eta\nu^2)^{-1}$, i.e., by

$$\frac{k_1k_0}{\eta^2\nu^2}\cdot\frac{1}{1-\alpha}\cdot\frac{1}{1-\beta} - \frac{k_0{}^2}{\eta\nu^2}\cdot\frac{1}{1-\alpha}\cdot\frac{1}{1-\beta} - \frac{k_1{}^2}{\eta^2\nu}\cdot\frac{1}{1-\beta}\cdot\frac{1}{1-\alpha}, \quad < 0$$

since each term < 0.

So, if $\alpha^* + \beta^* < 1$, $R_\delta(\lambda_w)$ is approximately minimized at (α^*, β^*). But, as mentioned earlier, $\alpha^* + \beta^* > 1$ implies that a decision should be made without taking any observations. It should be noted that the Wald approximations are only good when α and β are small, and when $E(N)$ is large, so that an investigation of whether (α^*, β^*) is a minimizing point when α^* and β^* are not small is not of very much interest.

The tables indicate that for $L_0 = L_1$, the nominal a.m. risk is greatest around $w = \frac{1}{2}$, which is reasonable since we have the least prior information about θ when $w = \frac{1}{2}$. We remark that if the iterated solution is greater than $\min(L_0w, L_1(1-w))$, then the best rule is a decision with no observations.

<u>Closeness of Chernoff a.m. Risk to Nominal a.m. Risk</u> $R_{\delta^*}(\lambda_w)$:

The following argument is intended to account for the closeness of $R_{\delta_0}(\lambda_w)$ to $R_{\delta^*}(\lambda_w)$, when δ_0 is the SPRT defined by α_0 and β_0 in (2.4.7). Since (α^*, β^*) and (α_0, β_0) need not be close, it appears that the surface, $R = g(\alpha,\beta)$ say, is somewhat flat near the minimum. We examine the behavior of $\frac{\partial R}{\partial \alpha}$ when β is fixed.

$$\frac{\partial R}{\partial \alpha} = L_0 w - k_0 (\log \frac{\beta}{1-\alpha} - \log \frac{1-\beta}{\alpha}) + \{k_0(1-\alpha) + k_1 \beta\} \frac{1}{1-\alpha} + \{k_0 \alpha + k_1 (1-\beta)\}(-\frac{1}{\alpha})$$

$$= f(\beta) - k_0 \log \frac{\alpha}{1-\alpha} + k_1 (\frac{\beta}{1-\alpha} - \frac{1-\beta}{\alpha})$$

where

$$f(\beta) = L_0 w - k_0 \log \frac{\beta}{1-\beta} .$$

Let $\alpha_0 = \alpha^* + \delta$, where α_0 is the Chernoff minimizing error probability. Then

$$\left.\frac{\partial R}{\partial \alpha}\right|_{(\alpha_0, \beta^*)} = f(\beta^*) - k_0 \log \frac{\alpha^* + \delta}{1-\alpha^*-\delta} + k_1 (\frac{\beta^*}{1-\alpha^*-\delta} - \frac{1-\beta^*}{\alpha^*+\delta})$$

$$= f(\beta^*) - k_0 [\log(1 + \frac{\delta}{\alpha^*}) - \log(1 - \frac{\delta}{1-\alpha^*}) - \log \frac{1-\alpha^*}{\alpha^*}]$$

$$+ k_1 \frac{\beta^*}{1-\alpha^*} \cdot \frac{1}{1 - \frac{\delta}{1-\alpha^*}} - k_1 \frac{1-\beta^*}{\alpha^*} \cdot \frac{1}{1 + \frac{\delta}{\alpha^*}}$$

$$= - k_0 [\log(1 + \frac{\delta}{\alpha^*}) - \log(1 - \frac{\delta}{1-\alpha^*})] + k_1 \frac{\beta^*}{1-\alpha^*}(\frac{1}{1 - \frac{\delta}{1-\alpha^*}} - 1)$$

$$- k_1 \frac{1-\beta^*}{\alpha^*}(\frac{1}{1 + \frac{\delta}{\alpha^*}} - 1)$$

since

$$\left.\frac{\partial R}{\partial \alpha}\right|_{(\alpha^*, \beta^*)} = 0.$$

For small error probabilities, the dominating term is the last. We assume k_0 and k_1 to be small, which is so if $c \leq .001$ in most cases, or if $c < .01$ in many cases. If $|\frac{\delta}{\alpha^*}|$ is small (i.e., if the Chernoff α_0 is close to α^*), then $\left.\frac{\partial R}{\partial \alpha}\right|_{(\alpha_0, \beta^*)}$ is going to be small;

or if $\frac{k_1}{\alpha^*}$ is small (and this is the important criterion in cases where $|\frac{\delta}{\alpha^*}|$ may not be small), then $\left.\frac{\partial R}{\partial \alpha}\right|_{(\alpha_0, \beta^*)}$ will be small.

The same kind of argument holds about any point (α', β') for which $\frac{\partial R}{\partial \alpha}$ is small; i.e., it can be applied to $\alpha_0' = \alpha' + \delta'$.

Similarly, if α is held fixed, $\left.\frac{\partial R}{\partial \beta}\right|_{(\alpha^*, \beta_0)}$ is small if $|\frac{\varepsilon}{\beta^*}|$ is small, where $\beta_0 = \beta^* + \varepsilon$, or if $\frac{k_1}{\beta^*}$ is small.

These results do not prove the flatness of $R_\delta(\lambda_w)$ near (α^*, β^*), even if all the quantities involved are small, but they are indicative of reasons why this flatness exists. The tables show that as c decreases (and α^*, β^* decrease), the percentage saving in using the iterative method, rather than the Chernoff approximation, becomes negligible.

Tables 2.4.1 to 2.4.7 show values of the Chernoff a.m. risk and nominal a.m. risk for a number of cases, including the univariate normal of Section 2.2. They are given on pp 43-49.

2.5. Comparison of a.m. risks: $w = \frac{1}{2}$.

The tables also indicate that for $L_0 = L_1 = 1$ and $w = \frac{1}{2}$, the nominal a.n. risk in the class of symmetric tests is very close to the nominal a.m. risk in the class of general tests. We write R_{min} for the a.m. risk in the latter class, and R_{sym} for the former. We also drop the star (*) notation. Thus, let (α, β) minimize $R_\delta(\lambda_w)$ for general tests, and α' for symmetric tests. In all cases in the tables, $\min(\alpha,\beta) \le \alpha' \le \max(\alpha,\beta)$, and indeed α' seems to be very close to

$$\gamma = \frac{\alpha + \beta}{2} .$$

Let R'' be the risk of the symmetric test in which the error probability is γ. Consider $R'' - R_{min}$, and $R'' - R_{sym}$. The first of these appears from computed results to be much larger numerically than the second.

Consider first $R'' - R_{min}$. The following applies to any (α, β) satisfying the conditions below, and not only to the minimizing values.

From (2.1.5) and (2.3.1),

$$R'' - R_{min} = L_0(\tfrac{1}{2} - w)\alpha + L_1(\tfrac{1}{2} - (1-w))\beta + J_0k_0 + J_1k_1$$

where

$$\left.\begin{aligned} J_0 &= -[1-(\alpha+\beta)]\log\frac{2-(\alpha+\beta)}{(\alpha+\beta)} - (1-\alpha)\log\frac{\beta}{1-\alpha} - \alpha\log\frac{1-\beta}{\alpha} \\ J_1 &= [1-(\alpha+\beta)]\log\frac{2-(\alpha+\beta)}{\alpha+\beta} - \beta\log\frac{\beta}{1-\alpha} - (1-\beta)\log\frac{1-\beta}{\alpha}\ . \end{aligned}\right\} \qquad (2.5.1)$$

(i) Hence, if $w = \frac{1}{2}$,

$$R'' - R_{min} = J_0k_0 + J_1k_1 \qquad (2.5.2)$$

and there is no contribution from the loss term. This factor could make a substantial difference; for example, if $w = .25$ and $L_0 = L_1 = L$,

$$R'' - R_{min} = \frac{1}{2}\gamma L + J_0k_0 + J_1k_1.$$

The first term $\frac{1}{2}\gamma L$ could be large enough to make the ratio $(R'' - R_{min})/R_{min}$ substantially greater than zero. The disappearance of this term when $w = \frac{1}{2}$ may explain why R'' and R_{min} are close (when $w = \frac{1}{2}$), but not necessarily close when $w \neq \frac{1}{2}$. A further reason arises from ii). Examining the terms in k_0 and k_1, let $\alpha = \gamma - \varepsilon$, $\beta = \gamma + \varepsilon$, so that $\varepsilon = \frac{1}{2}(\beta - \alpha)$. Suppose further that γ and $\frac{\varepsilon}{\gamma}$ are small, and consider J_1.

$$J_1 = (1-2\gamma)\log\frac{1-\gamma}{\gamma} - (\gamma+\varepsilon)\left[\log\frac{\gamma}{1-\gamma} + \log\frac{1+\frac{\varepsilon}{\gamma}}{1+\frac{\varepsilon}{1-\gamma}}\right]$$

$$- (1-\gamma-\varepsilon)\left[\log\frac{1-\gamma}{\gamma} + \log\left(\frac{1-\frac{\varepsilon}{1-\gamma}}{1-\frac{\varepsilon}{\gamma}}\right)\right]$$

$$= 2\varepsilon\log\frac{1-\gamma}{\gamma} - (1-2\gamma)\frac{\varepsilon}{\gamma(1-\gamma)} - \frac{1}{2}\frac{\varepsilon^2}{\gamma^2}\left(1-\frac{\gamma}{1-\gamma}\right)^2 + 0\left(\frac{\varepsilon^3}{\gamma^3}\right). \qquad (2.5.3)$$

Since $\varepsilon\log\gamma \to 0$ as $\frac{\varepsilon}{\gamma} \to 0$, it follows that whenever

$$\left|\frac{\varepsilon}{\gamma}\right| = \left|\frac{\alpha - \beta}{\alpha + \beta}\right|$$

is small, then J_1, and similarly J_0, are small; and hence for $w = \frac{1}{2}$ that $R'' - R_{min}$ is small.

(iii) An examination of the numerical results showed that the conditions for (ii) to hold were often, but not always, present. The quantity $\left|\frac{\alpha-\beta}{\alpha+\beta}\right|$ rises to as much as .431. In such cases J_0 and J_1 are not small, but neither will they be large, and hence whenever k_0 and k_1 are small, (2.5.2) shows that $R'' - R_{min}$ will be small. This is so whenever the cost c is small enough.

(iv) Next, consider $R'' - R_{sym}$.

Writing $R'' = f(\gamma)$, and $\gamma = \alpha' + \varepsilon'$, we get by Taylor expansion

$$R'' = R_{sym} + \varepsilon' f'(\alpha') + \frac{1}{2}\varepsilon'^2 f''(\alpha) + \cdots .$$

But from (2.3.2), $f'(\alpha') = \phi(\alpha') = 0$, and with (2.2.10),

$$R'' - R_{sym} = \frac{1}{2}\varepsilon'^2(k_1 - k_0)\frac{1}{\alpha'^2(1-\alpha')^2} + 0\left(\frac{\varepsilon'^3}{\alpha'^3}\log\frac{\varepsilon'}{\alpha'}\right) . \qquad (2.5.4)$$

This is small if k_1 and k_0 are small, or if $\frac{\varepsilon'}{\alpha'}$ is small. In most of the numerical cases investigated, ε' is very small, since α' is so close to $\frac{1}{2}(\alpha+\beta)$, and (2.5.4) is negligible.

The above discussion throws some light on the remarkable agreement in computed results between R_{min} and R_{sym} when $w = \frac{1}{2}$. The problem of whether such agreement holds under wider conditions might repay further investigation.

Tables 2.4.1 - 2.4.7 follow. * means that the 'no observations' procedure is best. $L_0 = L_1 = 1$ in all cases. The Chernoff and nominal values are both shown; the % Saving is also shown. This is given by

$$\text{Percentage Saving} = \frac{\text{Chernoff a.m. risk} - \text{nominal a.m. risk}}{\text{nominal a.m. risk}} \times 100 .$$

TABLE 2.4.1

Nominal a.m. Risk

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}; \quad \theta_1 - \theta_0 = 0.1 \quad E_1(Z) = -E_0(Z) = .005$$

For $c \geq 0.005$, the 'no observations' procedure is best.

w	c	α Cher.	α Nom.	β Cher.	β Nom.	EN Cher.	EN Nom.	a.m. Risk Cher.	a.m. Risk Nom.	% Sav.
.1	.001							*	*	
.2	.001							*	*	
.3	.001	.4667	.7298	.0857	.0376	103.82	46.737	*	.2920	2.74
.4	.001	.3000	.4270	.1333	.1242	151.20	98.302	.3512	.3436	2.22
.5	.001	.2000	.2453	.2000	.2453	166.36	114.52	.3664	.3598	1.82
.1	.0005	.9000		.0111		15.070		*	*	
.2	.0005	.4000	.6125	.0250	.0151	188.30	105.75	.1942	.1874	3.59
.3	.0005	.2333	.3470	.0429	.0435	284.75	204.32	.2424	.2367	2.40
.4	.0005	.1500	.2142	.0667	.0814	335.54	255.89	.2678	.2625	2.02
.5	.0005	.1000	.1345	.1000	.1345	351.56	272.10	.2758	.2706	1.84
.1	.0001	.1800	.2112	.0022	.0021	395.99	363.99	.0596	.0594	.37
.2	.0001	.0800	.0936	.0050	.0053	580.97	549.19	.0781	.0779	.26
.3	.0001	.0467	.0543	.0086	.0095	679.5	647.8	.0880	.0878	.23
.4	.0001	.0300	.0347	.0133	.0151	731.0	699.3	.0931	.0929	.20
.5	.0001	.0200	.0230	.0200	.0230	747.2	715.5	.0947	.0945	.20

TABLE 2.4.2

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}; \qquad \theta_1 - \theta_0 = 0.25 \qquad E_1(Z) = -E_0(Z) = .03125$$

w	c	α		β		EN		a.m. Risk		% Sav.
		Cher.	Nom.	Cher.	Nom.	Cher.	Nom.	Cher.	Nom.	
.1	.01							*	*	
.2	.01							*	*	
.3	.01	.7467		.1371		1.366		*	*	
.4	.01	.4800	.6278	.2133	.1177	6.707	5.912	.3871	.3809	1.63
.5	.01	.3200	.3218	.3200	.3218	8.684	8.507	.4068	.4068	.005
.1	.001	.2880	.3609	.0036	.0029	46.01	38.63	.0780	.0773	.93
.2	.001	.1280	.1595	.0080	.0085	75.55	68.26	.1075	.1069	.57
.3	.001	.0747	.0924	.0137	.0156	91.29	84.03	.1233	.1227	.48
.4	.001	.0480	.0588	.0213	.0252	99.53	92.28	.1315	.1310	.44
.5	.001	.0320	.0387	.0320	.0387	102.12	94.88	.1341	.1335	.3
.1	.0001	.0288	.0299	.00036	.00036	126.30	125.18	.0158	.0158	0
.2	.0001	.0128	.0133	.00080	.00082	155.93	154.81	.0188	.0188	0
.3	.0001	.0075	.0077	.00137	.00141	171.70	170.58	.0204	.0204	0
.4	.0001	.0048	.0050	.0021	.0022	179.95	178.83	.0212	.0212	0
.5	.0001	.0032	.0033	.0032	.0033	182.55	181.43	.0215	.0215	0

TABLE 2.4.3

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}; \quad \theta_1 - \theta_0 = 0.5 \quad E_1(Z) = -E_0(Z) = .125$$

w	c	α		β		EN		a.m. Risk		
		Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	% saving
.25	.1							*	*	
.5	.1		.4231		.4231		.381	*	.4613	
.25	.01	.2400	.3491	.0267	.0259	12.060	8.975	.2006	.1965	2.10
.5	.01	.0800	.1067	.0800	.1067	16.413	13.369	.2441	.2404	1.54
.25	.001	.0240	.0259	.0027	.0028	33.55	32.95	.0416	.0415	.05
.5	.001	.0080	.0086	.0080	.0086	37.95	37.34	.0460	.0459	.04
.25	.0001	.0024	.0024	.00027	.00027	52.56	52.47	.00606	.00606	0
.5	.0001	.00080	.00081	.00080	.00081	56.95	56.86	.00650	.00650	0

TABLE 2.4.4

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}\theta} \exp(-\frac{1}{2\theta^2} x^2); \quad \theta_0 = 1, \quad \theta_1 = 2; \quad E_0(Z) = -.31815, \quad E_1(Z) = .80685$$

w	c	α		β		EN		a.m. Risk		
		Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	% saving
.25	.1	.3718	.9672	.1048	.0041	1.184		*	.2499	
.5	.1	.1239	.2305	.3143	.2879	1.605	1.103	.3797	.3695	2.76
.75	.1	.0413		.9430				*	*	
.25	.01	.0372	.0481	.0105	.0110	6.299	5.958	.0801	.0798	.38
.5	.01	.0124	.0156	.0314	.0341	7.815	7.498	.1001	.0999	.21
.75	.01	.0041	.0048	.0943	.1034	6.925	6.631	.0959	.0958	.14
.25	.001	.00372	.0039	.00105	.00106	10.539	10.493	.0123	.0123	.08
.5	.001	.00124	.00129	.0031	.0032	13.152	13.106	.0153	.0153	0
.75	.001	.00041	.00043	.0094	.0096	13.357	13.311	.0160	.0160	0
.25	.0001	.00037	.00037	.000105	.000105	14.534	14.536	.0016	.0016	0
.5	.0001	.000124	.000125	.00031	.00031	18.243	18.240	.0020	.0020	0
.75	.0001	.000041	.000041	.00094	.00094	19.543	19.539	.0022	.0022	0

TABLE 2.4.5

$$p_\theta(x) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0; \quad \frac{\theta_1}{\theta_0} = 1.2 \qquad E_0(Z) = -.01768, \quad E_1(Z) = .01565$$

For $c \geq .01$, the 'no observations' procedure is best.

w	c	α		β		EN		a.m. Risk		
		Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	% saving
.25	.005							*	*	
.5	.005	.3195	.3159	.2828	.3092	20.187	17.795	.4021	.4015	
.75	.005							*	*	
.25	.001	.1917	.2586	.0189	.0203	111.91	91.92	.1740	.1718	.13
.5	.001	.0639	.0814	.0566	.0753	145.51	125.18	.2057	.2035	1.10
.75	.001	.0213	.0223	.1697	.2405	113.28	92.27	.1717	.1691	1.52
.25	.0005	.0958	.1163	.0094	.0105	163.99	151.11	.1130	.1125	.44
.5	.0005	.0319	.0378	.0283	.0344	196.47	183.46	.1284	.1278	
.75	.0005	.0107	.0117	.0848	.1060	162.85	149.63	.1106	.1101	

TABLE 2.4.6

$p_\theta(x) = \theta^2 x e^{-\theta x}, \quad x > 0, \quad \theta > 0; \quad \theta_1 = 1, \quad \theta_2 = 2 \qquad E_0(Z) = -.6137, \quad E_1(Z) = .3863$

w	c	α		β		EN		a.m. Risk		% saving
		Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	Chernoff	nominal	
.25	.1							*	*	
.5	.1	.2589	.2689	.1629	.2389	1.619	1.118	.3728	.3656	1.96
.75	.1							*	*	
.25	.01	.0777	.0878	.0054	.0061	6.718	6.396	.0907	.0905	.18
.5	.01	.0259	.0289	.0163	.0196	7.715	7.381	.0982	.0980	.21
.75	.01	.0086	.0092	.0489	.0599	6.396	6.049	.0827	.0824	.32
.25	.001	.0078	.0079	.00054	.00056	12.438	12.389	.0148	.0148	0
.5	.001	.0026	.0026	.00163	.00168	12.884	12.834	.0150	.0150	0
.75	.001	.00086	.00088	.0049	.0050	11.012	10.963	.0129	.0129	0
.25	.0001	.00078	.00078	.000054	.000054	17.895	17.890	.00202	.00202	0
.5	.0001	.00026	.00024	.000163	.00016	17.789	17.870	.00199	.00199	0
.75	.0001	.000086	.000087	.00049	.00049	15.365	15.359	.00172	.00172	0

TABLE 2.4.7

$Pr(x = \{^1_0) = \{^{\theta}_{1-\theta}; \quad \theta_0 = .5, \quad \theta_1 = .7; \quad E_0(Z) = -.08718, \quad E_1(Z) = .08228$

w	c	α Chernoff	α nominal	β Chernoff	β nominal	EN Chernoff	EN nominal	a.m. Risk Chernoff	a.m. Risk nominal	% saving
.25	.1							*	*	
.5	.1		.4480		.4470		.262	*	.4736	
.75	.1							*	*	
.25	.01	.3646	.5544	.0382	.0277	11.631	6.936	.2361	.2287	3.24
.5	.01	.1215	.1603	.1147	.1564	18.136	13.480	.2995	.2932	2.15
.75	.01	.0405	.0290	.3441	.5426	11.971	7.048	.2361	.2279	3.62
.25	.001	.0365	.0403	.0038	.0041	44.77	43.55	.0568	.0567	.09
.5	.001	.0122	.0133	.0115	.0126	51.05	49.83	.0629	.0628	.08
.75	.001	.00405	.0043	.0344	.0383	44.35	43.12	.0560	.0559	.09
.25	.0001	.00365	.0037	.00038	.00039	73.54	73.36	.00855	.00855	0
.5	.0001	.00122	.00123	.00115	.00116	79.42	79.24	.00912	.00912	0
.75	.0001	.00042	.00041	.0034	.0035	72.33	72.15	.00840	.00840	0

CHAPTER III

THE PARTIAL SPRT PROCEDURE

3.1. Definition, and Application to the Exponential Family.

In Chapter II, the problem of minimizing the risk for Wald SPRTs was considered. We shall be concerned with the problem of minimum risk when the cost is unknown in Chapters IV and V. The procedure to be discussed in Chapter V, when no estimate of the cost is available, is a variant of the SPRT, in which n observations are taken initially, and a SPRT procedure is introduced at stage n. This will be defined presently, and will be called the Partial Sequential Probability Ratio Test, or PSPRT.

Although no discussion of such a procedure has appeared in the literature, Billard [5] has recently derived properties of an analogous type of test procedure for two-sided alternative hypotheses, deriving the OC function and ASN for testing a normal mean with known variance, and a binomial variable.

In this chapter, the properties of the PSPRT will be derived, and compared with other procedures. In this section, general formulae for the O.C. function and the ASN will be derived when the r.v. under consideration belongs to the exponential family. In Section 3.2., these formulae will be further developed when the r.v. X has a Gamma distribution; in Sections 3.3. and 3.4., X will be a $N(\theta, 1)$ variable.

It can be noted at this stage that when $n = 1$, the PSPRT is a Wald SPRT. The results of Sections 3.3. and 3.4. will be compared in this case numerically with exact results and approximations of Wald and Kemp [13], in Section 3.5. A comparison with Anderson's modified SPRT [1] is also made.

Definition of PSPRT:

$X_1, X_2, \ldots,$ is a sequence of independently and identically distributed random variables indexed by a parameter θ. We want to test a hypothesis H_0: $\theta = \theta_0$ vs an alternative H_1: $\theta = \theta_1$. A fixed numbers n of observations is taken, and further observations are taken according to whether

$$B < \frac{p_{1n'}}{p_{0n'}} < A, \quad n' \geq n$$

or not, where $p_{in'}$ = joint likelihood of the first n observations $\underline{x}_{n'} = (x_1, x_2, \ldots, x_{n'})$ under H_i, $i = 0, 1$.

This is a GSPRT, and may be relevant to the problem of unknown cost considered earlier, since the information from the first n observations may be used to estimate the cost (or mean cost), and hence the values of A and B chosen to minimize the risk.

Let $T(\underline{x}_n)$ be the conditional test, given $\underline{x}_n$, and starting at the close of stage n, in which the continuation region is

$$B'_n(\underline{x}_n) = B\frac{p_{0n}}{p_{1n}} < \frac{p_1(x_{n+1}, \ldots, x_{n'})}{p_0(x_{n+1}, \ldots, x_{n'})} < A\frac{p_{0n}}{p_{1n}} = A'_n(\underline{x}_n); \; n' \geq n. \qquad (3.1.1)$$

If the inequality

$$0 < B'_n < 1 < A'_n \tag{3.1.2}$$

does not hold, then we decide at stage n. If $A'_n < 1$, H_1 is accepted; if $B'_n > 1$, H_0 is accepted.

Let $b = \log B$, $a = \log A$, $b'_n = \log B'_n$, $a'_n = \log A'_n$, etc.

Let $\mathcal{T}_n$ be the unconditional test procedure with n fixed in advance, having OC function $L_n(\theta)$, ASN $\mathcal{E}_n(N|\theta)$, and error probabilities α_n, β_n.

If $L_{\underline{x}_n}(\theta)$, $\mathcal{E}_{\underline{x}_n}(N|\theta)$, $\alpha(\underline{x}_n)$ and $\beta(\underline{x}_n)$ are the equivalent quantities for test $\mathcal{T}(\underline{x}_n)$, then

$$\left.\begin{aligned} L_n(\theta) &= \mathcal{E}L_{\underline{x}_n}(\theta) \\ \mathcal{E}_n(N|\theta) &= \mathcal{E}\mathcal{E}_{\underline{x}_n}(N|\theta), \end{aligned}\right\} \tag{3.1.3}$$

taking expectations over the joint distribution of $\underline{x}_n$.

<u>Exponential Family</u>.

We confine attention to the case where

$$p_\theta(x) = A(x)\exp[\theta\, s(x) - \tau(\theta)]. \tag{3.1.4}$$

If $$\bar{s}_n = \frac{1}{n}\sum_{i=1}^{n} s(x_i), \tag{3.1.5}$$

then $\bar{s}_n$ is a sufficient statistic for θ up to n observations, and hence expectations in (3.1.4) may be taken over the distribution of $\bar{s}_n$.

We have

$$\begin{aligned} \log\frac{p_{1n}}{p_{0n}} &= n(\theta_1 - \theta_0)\bar{s}_n - n(\tau(\theta_1) - \tau(\theta_0)) \\ &= n(\Delta\bar{s}_n - T), \text{ say.} \end{aligned} \tag{3.1.6}$$

Dropping the subscript n on $\bar{s}_n$, we have from (3.1.2) that observations are continued beyond stage n if and only if

$$b < n(\Delta\bar{s} - T) < a$$

or

$$b''_n = \frac{\frac{1}{n}b + T}{\Delta} < \bar{s} < \frac{\frac{1}{n}a + T}{\Delta} = a''_n \qquad (3.1.7)$$

where we assume $\theta_1 > \theta_0$ (i.e., $\Delta > 0$), and will retain this assumption for the rest of the chapter.

O. C. Function.

Using Wald's approximation, the O.C. function in test $T(\underline{x}_n)$ is

$$L_{\underline{x}_n}(\theta) \simeq \begin{cases} 1 & \text{if} \quad \bar{s} < b''_n \\ \dfrac{1 - \exp\{h(\theta)a'_n\}}{\exp\{h(\theta)b'_n\} - \exp\{h(\theta)a'_n\}} & \text{if} \quad b''_n < \bar{s} < a''_n \\ 0 & \text{if} \quad \bar{s} > a''_n \end{cases} \qquad (3.1.8)$$

where

$$E[\{\frac{p_1(x)}{p_0(x)}\}^{h(\theta)}|\theta] = 1, \quad h \neq 0.$$

So

$$L_{\underline{x}_n}(\theta) = \begin{cases} 1 & \text{if} \quad \bar{s} < b''_n \\ \dfrac{e^{hn(\Delta\bar{s} - T)} - e^{ha}}{e^{hb} - e^{ha}} & \text{if} \quad b''_n < \bar{s} < a''_n . \\ 0 & \text{if} \quad \bar{s} > a''_n \end{cases} \qquad (3.1.8a)$$

Hence

$$L_n(\theta) \quad = \quad E(L_{\underline{x}_n}(\theta)),$$

and if $\bar{s}$ has c.d.f. F,

$$= \quad F_\theta(b''_n) + \frac{1}{e^{hb} - e^{ha}} \int_{b''_n}^{a''_n} \{e^{hn(\Delta\bar{s} - T)} - e^{ha}\} dF_\theta(\bar{s})$$

$$= \quad \frac{e^{hb}F_\theta(b''_n) - e^{ha}F_\theta(a''_n)}{e^{hb} - e^{ha}} + \frac{1}{e^{hb} - e^{ha}} \int_{b''_n}^{a''_n} e^{hn(\Delta\bar{s} - T)} dF_\theta(\bar{s}). \qquad (3.1.9)$$

<u>A.S.N.</u>

Denote by g_θ the quantity $E_\theta(\log \frac{p_1(x)}{p_0(x)})$

i.e.,

$$g_\theta \quad = \quad \Delta E(s(x)|\theta) - T. \qquad (3.1.10)$$

Then

$$E_{\underline{x}_n}(N|\theta) \quad \simeq \quad \begin{cases} n \quad \text{if} \quad \bar{s} < b''_n \quad \text{or} \quad > a''_n \\ n + \frac{1}{g_\theta}\{L_{\underline{x}_n}(\theta)b'_n + (1 - L_{\underline{x}_n}(\theta))a'_n\} \qquad (3.1.11) \\ \qquad \text{otherwise.} \end{cases}$$

Hence

$$E_n(N|\theta) = n + \frac{1}{g_\theta}\int_{b_n''}^{a_n''}\left[a_n' - (a_n' - b_n')\frac{e^{hn(\Delta\bar{s} - T)} - e^{ha}}{e^{hb} - e^{ha}}\right]dF_\theta(\bar{s})$$

$$= n + \frac{1}{g_\theta}\int_{b_n''}^{a_n''}\left[a - n(\Delta\bar{s} - T) - (a - b)\frac{e^{hn(\Delta\bar{s}-T)} - e^{ha}}{e^{hb} - e^{ha}}\right]dF_\theta(\bar{s}) \tag{3.1.12}$$

$$= n + \frac{1}{g_\theta}\left[nT + \frac{ae^{hb} - be^{ha}}{e^{hb} - e^{ha}}\right][F_\theta(a_n'') - F_\theta(b_n'')]$$

$$- \frac{n\Delta}{g_\theta}\int_{b_n''}^{a_n''} s\ dF_\theta(\bar{s}) - \frac{a - b}{g_\theta(e^{hb} - e^{ha})}\int_{b_n''}^{a_n''} e^{hn(\Delta\bar{s}-T)}dF_\theta(\bar{s}). \tag{3.1.12a}$$

We shall consider two particular cases for study; those in which $X \sim N(\theta, 1)$ and X is a Gamma variable with m degrees of freedom.

3.2. Testing a Gamma Parameter.

Here

$$p_\theta(x) = \frac{1}{\Gamma(m)}\theta^m x^{m-1} e^{-\theta x}, \qquad x > 0; \quad \theta > 0. \tag{3.2.1}$$

And

$$p_\theta(\underline{x}_n) = K(\underline{x}_n)e^{-n(\theta\bar{x} - m\log\theta)}. \tag{3.2.2}$$

So

$$\bar{s} = -\bar{x}$$

and

$$T = -m\log(\theta_1/\theta_0)$$

$\bar{x}$ has a Gamma distribution where

$$p_\theta^{\bar{x}}(x) = \frac{1}{\Gamma(mn)} (n\theta)^{mn} x^{mn-1} e^{-n\theta x}, \quad x > 0. \tag{3.2.3}$$

Representing the distribution of $\bar{x}$ by G,

$$F_\theta(t) = 1 - G_\theta(-t) . \tag{3.2.4}$$

Hence (3.1.9) gives

$$L_n(\theta) = 1 - \frac{e^{hb} G_\theta(-b_n'') - e^{ha} G_\theta(-a_n'')}{e^{hb} - e^{ha}}$$

$$+ \frac{1}{e^{hb} - e^{ha}} \int_{-a_n''}^{-b_n''} \frac{1}{\Gamma(mn)} (n\theta)^{nm} y^{nm-1} e^{hnm \log(\frac{\theta_1}{\theta_0})}$$

$$\cdot e^{-nh(\theta_1-\theta_0)y} e^{-n\theta y} dy$$

$$= 1 - \frac{e^{hb}(G_\theta(-b_n'') - e^{ha} G_\theta(-a_n'')}{e^{hb} - e^{ha}} + \frac{1}{e^{hb} - e^{ha}} (\frac{\theta_1}{\theta_0})^{hnm} (\frac{\theta}{\theta + h(\theta_1 - \theta_0)})^{nm}$$

$$\cdot \int_{-a_n''}^{-b_n''} \frac{1}{\Gamma(nm)} [n(\theta + (\theta_1 - \theta_0)h]^{nm} y^{nm-1} e^{-n(\theta + h(\theta_1-\theta_0))} dy.$$

If we use the notation for the Incomplete Gamma Function ratio

$$P(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt, \tag{3.2.5}$$

then this reduces to

$$L_n(\theta) \simeq 1 - \frac{1}{e^{hb}-e^{ha}} [e^{hb}P(mn,\frac{\theta}{\Delta}(-b+mn\log\frac{\theta_1}{\theta_0})) - e^{ha}P(mn,\frac{\theta}{\Delta}(-a+mn\log\frac{\theta_1}{\theta_0}))]$$

$$+ \frac{1}{e^{hb}-e^{ha}} [(\frac{\theta_1}{\theta_0})^{hnm} (\frac{\theta}{\theta+h\Delta})^{nm} \{P(mn,\frac{\theta+h\Delta}{\Delta} (-b + mn \log \frac{\theta_1}{\theta_0}))$$

$$- P(mn, \frac{\theta+h\Delta}{\Delta} (-a + mn \log \frac{\theta_1}{\theta_0}))\}] \tag{3.2.6}$$

or

$$L_n(\theta) \simeq 1 - \frac{1}{e^{hb} - e^{ha}} [e^{hb} P(mn, - n\theta b_n'') - e^{ha} P(mn, - n\theta a_n'')]$$

$$+ \frac{1}{e^{hb}-e^{ha}} (\frac{\theta_1}{\theta_0})^{hnm}(\frac{\theta}{\theta+h\Delta})^{nm}\{P(mn, - n(\theta+h\Delta)b_n'')- P(mn, - n(\theta+h\Delta)a_n'')\} \tag{3.2.7}$$

where

$$a_n'' = \frac{1}{(\theta_1 - \theta_0)} \{\frac{1}{n} a - m \log \frac{\theta_1}{\theta_0}\}$$

$$b_n'' = \frac{1}{(\theta_1 - \theta_0)} \{\frac{1}{n} b - m \log \frac{\theta_1}{\theta_0}\}.$$

For any value of θ, if $h(\theta)$ is known, $L_n(\theta)$ can be approximately evaluated. To determine h, since $\log \frac{p_1(x)}{p_0(x)} = -\Delta x + m\log \frac{\theta_1}{\theta_0}$,

$$\frac{1}{\Gamma(m)} \int_0^\infty \theta^m x^{m-1} (\frac{\theta_1}{\theta_0})^{mn} e^{-(h\Delta + \theta)x} dx = 1$$

i.e.,

$$[(\frac{\theta_1}{\theta_0})^h]^m = (\frac{h\Delta}{\theta} + 1)^m . \tag{3.2.8}$$

For the case $m = 1$, this gives h as the non-zero solution of

$$\left(\frac{\theta_1}{\theta_0}\right)^h = \frac{h\Delta}{\theta} + 1 . \tag{3.2.9}$$

The solution of (3.2.9) will, by Wald's Lemma [18], satisfy (3.2.8), so that for distributions of the form (3.2.1), $h(\theta)$ is independent of m, and given by (3.2.9).

The A.S.N.

Using the notation of (3.1.10), with one obvious modification,

$$\begin{aligned} g_\theta &= E\left(\log \frac{p_1(x)}{p_0(x)} \middle| \theta\right) \\ &= E(-\Delta \bar{x}) + m \log \left(\frac{\theta_1}{\theta_0}\right) \\ &= -\frac{\Delta}{\theta} m + m \log \left(\frac{\theta_1}{\theta_0}\right) . \end{aligned} \tag{3.2.10}$$

From (3.1.12), (3.2.3) and (3.2.4), we get

$$\begin{aligned} E_n(N|\theta) \approx{} & n + \frac{1}{g_\theta} \left[-nm \log \frac{\theta_1}{\theta_0} + \frac{ae^{hb} - be^{ha}}{e^{hb} - e^{ha}}\right] \{G_{\theta,mn}(-b_n'') - G_{\theta,mn}(-a_n'')\} \\ & + \frac{n\Delta}{g_\theta} \int_{-a_n''}^{-b_n''} \frac{1}{\Gamma(mn)} \cdot (n\theta)^{mn} x^{mn} e^{-n\theta x} \, dx \\ & - \frac{(a - b)}{g_\theta(e^{hb} - e^{ha})} \int_{-a_n''}^{-b_n''} e^{hnm\log\left(\frac{\theta_1}{\theta_0}\right)} \frac{1}{\Gamma(mn)} (n\theta)^{nm} x^{mn-1} e^{-n(h\Delta+\theta)x} dx \end{aligned}$$

i.e.,

$$E_n(N|\theta) \simeq n + \frac{1}{m(\log\frac{\theta_1}{\theta_0} - \frac{\Delta}{\theta})} (\{-nm \log \frac{\theta_1}{\theta_0} + \frac{ae^{hb}-be^{ha}}{e^{hb}-e^{ha}}\}\{G_{\theta,mn}(-b''_n)$$

$$- G_{\theta,mn}(-a''_n)\} + n\Delta \cdot \frac{m}{\theta} \{G_{\theta,mn+1}(-b''_n) - G_{\theta,mn+1}(-a''_n)\}$$

$$- \frac{a - b}{e^{hb}-e^{ha}} \cdot (\frac{\theta_1}{\theta_0})^{hnm} (\frac{\theta}{\theta+\Delta h})^{nm} \{G_{\theta+h\Delta,mn}(-b''_n) - G_{\theta+h\Delta,mn}(-a''_n)\}).$$

We can relate $G_{\theta,mn+1}$ to $G_{\theta,mn}$. For

$$\int y^{mn} e^{-n\theta y} dy = -\frac{1}{n\theta} y^{mn} e^{-n\theta y} + \frac{1}{n\theta} \int mn\, y^{mn-1} e^{-n\theta y} dy.$$

Hence

$$G_{\theta,mn+1}(-b''_n) - G_{\theta,mn+1}(-a''_n) = \frac{1}{\Gamma(mn+1)} \frac{(n\theta)^{mn+1}}{\theta} [(-a''_n)^{mn} e^{n\theta a''_n}$$

$$- (-b''_n)^{mn} e^{n\theta b''_n}] + \{G_{\theta,mn}(-b''_n)$$

$$- G_{\theta,mn}(-a''_n)\}.$$

So

$$E_n(N|\theta) \simeq n + \frac{\Delta(n\theta)^{mn+1}}{\Gamma(mn)\theta^2} [(-a''_n)^{mn} e^{n\theta a''_n} - (-b''_n)^{mn} e^{n\theta b''_n}] \cdot \frac{1}{m(\log\frac{\theta_1}{\theta_0} - \frac{\Delta}{\theta})}$$

$$+ \frac{1}{m(\log\frac{\theta_1}{\theta_0} - \frac{\Delta}{\theta})} (\{-mn\log\frac{\theta_1}{\theta_0} + \frac{n\Delta m}{\theta} + \frac{ae^{hb}-be^{ha}}{e^{hb}-e^{ha}}\}\{G_{\theta,mn}(-b''_n)$$

$$- G_{\theta,mn}(-a''_n)\} - \frac{a - b}{e^{hb}-e^{ha}} \cdot (\frac{\theta_1}{\theta_0})^{hmn} (\frac{\theta}{\theta+\Delta h})^{mn} \{G_{\theta+\Delta h,mn}(-b''_n)$$

$$- G_{\theta+\Delta h,mn}(-a''_n)\}) \qquad (3.2.11)$$

or

$$E_n(N|\theta) \simeq n\{1 - G_{\theta,mn}(-b_n'') + G_{\theta,mn}(-a_n'')\}$$

$$+ \frac{1}{m(\log\frac{\theta_1}{\theta_0} - \frac{\Delta}{\theta})} (\frac{\Delta n}{\theta} \cdot \frac{(n\theta)^{mn}}{\Gamma(mn)} \{(-a_n'')^{mn} e^{n\theta a_n''} - (-b_n'')^{mn} e^{n\theta b_n''}\}$$

$$+ \frac{ae^{hb}-be^{ha}}{e^{hb}-e^{ha}} \{G_{\theta,mn}(-b_n'') - G_{\theta,mn}(-a_n'')\}$$

$$- \frac{(a-b)}{e^{hb}-e^{ha}} (\frac{\theta_1}{\theta_0})^{hmn} (\frac{\theta}{\theta+\Delta h})^{mn} \{G_{\theta+\Delta h,mn}(-b_n'') - G_{\theta+\Delta h,mn}(-a_n'')\})$$

i.e.,

$$E_n(N|\theta) \simeq n\{1 - P(mn, -n\theta b_n'') + P(mn, -n\theta a_n'')\}$$

$$+ \frac{1}{m(\log\frac{\theta_1}{\theta_0} - \frac{\Delta}{\theta})} (\frac{\Delta n}{\theta} \cdot \frac{(n\theta)^{mn}}{\Gamma(mn)} \{(-a_n'')^{mn} e^{n\theta a_n''} - (-b_n'')^{mn} e^{n\theta b_n''}\}$$

$$+ \frac{ae^{hb}-be^{ha}}{e^{hb}-e^{ha}} \{P(mn, -n\theta b_n'') - P(mn, -n\theta a_n'')\}$$

$$- \frac{a-b}{e^{hb}-e^{ha}} (\frac{\theta_1}{\theta_0})^{hmn} (\frac{\theta}{\theta+\Delta h})^{mn} \{P(mn, -n(\theta+\Delta h)b_n'')$$

$$- P(mn, -n(\theta+\Delta h)a_n'')\}). \qquad (3.2.12)$$

3.3. Testing a Normal Mean with Known Variance: the O.C. Function.

Suppose without loss of generality that

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}. \tag{3.3.1}$$

Then in (3.1.4), $s(x) = x$ and $\tau(\theta) = \frac{1}{2}\theta^2$; $T = \frac{1}{2}(\theta_1^2 - \theta_0^2)$.

Hence

$$\bar{s} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{3.3.2}$$

and

$$\left.\begin{aligned} a_n'' &= \frac{a}{n\Delta} + \tfrac{1}{2}(\theta_1 + \theta_0) \\ b_n'' &= \frac{b}{n\Delta} + \tfrac{1}{2}(\theta_1 + \theta_0) \ . \end{aligned}\right\} \tag{3.3.3}$$

In the case of the present problem,

$$h(\theta) = \frac{\theta_1 + \theta_0 - 2\theta}{\Delta} . \tag{3.3.4}$$

So for the O.C. function, (3.1.8a) gives for the conditional test $T(\underline{x}_n)$,

$$\begin{aligned} \Delta\bar{s} - T &= \Delta\bar{x} - \tfrac{1}{2}\Delta(\theta_1 + \theta_0) \\ &= \Delta(\bar{x} - \theta) - \tfrac{1}{2}h\Delta^2 \ . \end{aligned}$$

Hence

$$L_{\underline{x}_n}(\theta) \approx \begin{cases} 1 & \text{if} \quad \bar{x} < b_n'' \\ \dfrac{\exp[hn\Delta\{(\bar{x}-\theta) - \frac{1}{2}h\Delta\}] - e^{ha}}{e^{hb} - e^{ha}} & \text{if} \quad b_n'' < \bar{x} < a_n''. \\ 0 & \text{if} \quad \bar{x} > a_n'' \end{cases} \tag{3.3.5}$$

So for the PSPRT T_n, if Φ is the c.d.f. of a N(0,1) variable, (3.1.9) gives

$$L_n(\theta) \simeq \frac{e^{hb}}{e^{hb}-e^{ha}} \Phi(\sqrt{n}(b_n''-\theta)) - \frac{e^{ha}}{e^{hb}-e^{ha}} \Phi(\sqrt{n}(a_n''-\theta))$$

$$+ \frac{1}{e^{hb}-e^{ha}} \int_{b_n''}^{a_n''} \sqrt{\frac{n}{2\pi}} \exp[hn\Delta(\bar{x} - \theta - \tfrac{1}{2}h\Delta) - \frac{n}{2}(\bar{x}-\theta)^2]d\bar{x} \qquad (3.3.6)$$

and since $hn\Delta(\bar{x}-\theta-\tfrac{1}{2}n\Delta h) - \tfrac{1}{2}n(\bar{x}-\theta)^2 = -\tfrac{1}{2}n(\bar{x}-\theta+h\Delta)^2$, the last integral

$$= \int_{b_n''-\theta+h\Delta}^{a_n''-\theta+h\Delta} \sqrt{\frac{n}{2\pi}}\, e^{-\frac{1}{2}ny^2}\, dy$$

where

$$y = \bar{x} - \theta + h\Delta \ .$$

Hence, since

$$\sqrt{n}(b_n'' - \theta) = (\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta)$$

$$\sqrt{n}(b_n'' - \theta + h\Delta) = (\frac{b}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}h\Delta),$$

and similarly for terms in a_n'', (3.3.6) gives

$$L_n(\theta) \simeq \frac{e^{hb}}{e^{hb}-e^{ha}} \Phi(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta) - \frac{e^{ha}}{e^{hb}-e^{ha}} \Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta)$$

$$+ \frac{1}{e^{hb}-e^{ha}} \{\Phi(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}h\Delta) - \Phi(\frac{b}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}h\Delta)\}. \qquad (3.3.7)$$

An alternative form is

$$L_n(\theta) \simeq \frac{e^{hb}}{e^{hb}-e^{ha}} \Phi(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}(\theta_1+\theta_0-2\theta)) - \frac{e^{ha}}{e^{hb}-e^{ha}} \Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}(\theta_1+\theta_0-2\theta))$$

$$+ \frac{1}{e^{hb}-e^{ha}} \{\Phi(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}(\theta_1+\theta_0-2\theta)) - \Phi(\frac{b}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}(\theta_1+\theta_0-2\theta))\}. \qquad (3.3.8)$$

We introduce the following notation, to be used here and in Chapter V:

$$\left.\begin{aligned} \Phi_+(a,\theta) &= \Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta) \\ \Phi_-(a,\theta) &= \Phi(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}h\Delta) . \end{aligned}\right\} \qquad (3.3.9)$$

The derivatives will be required later; if $\Phi'(x) = \phi(x)$, write

$$\begin{aligned} \phi_+(a,\theta) &= \Phi_+'(a,\theta) \\ &= \frac{1}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta)^2] \end{aligned} \qquad (3.3.10)$$

and similarly for $\phi_-(a,\theta)$; also expressions in b are similarly defined. The dependence of Φ and ϕ on n and Δ is to be understood.

Then concisely, (3.3.7) gives

$$L_n(\theta) \simeq \frac{e^{hb}}{e^{hb}-e^{ha}} \Phi_+(b,\theta) - \frac{e^{ha}}{e^{hb}-e^{ha}} \Phi_+(a,\theta) + \frac{1}{e^{hb}-e^{ha}} \{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}. \qquad (3.3.11)$$

We note that as $n \downarrow 0$, stochastically; i.e., as a continuous real number,

$$L_n(\theta) \to \frac{1 - e^{ha}}{e^{hb} - e^{ha}},$$

the approximate O.C. function of a Wald SPRT. Further, as $n \to \infty$, $L_n(\theta) \to 1$ or 0 according as $\theta <$ or $> \frac{1}{2}(\theta_1+\theta_0)$.

Error Probabilities.

Let α_n and β_n be the nominal error probabilities $1 - L(\theta_0)$ and $L(\theta_1)$. Then, writing in terms of the boundaries A and B of the PSPRT, at $\theta = \theta_0$, $h(\theta) = 1$, and at $\theta = \theta_1$, $h = -1$. So

$$\alpha_n = 1 - L(\theta_0)$$

$$= 1 + \frac{B}{A-B}\,\Phi\{\frac{b}{\sqrt{n}\Delta} + \frac{\sqrt{n}\Delta}{2}\} - \frac{A}{A-B}\,\Phi\{\frac{a}{\sqrt{n}\Delta} + \frac{\sqrt{n}\Delta}{2}\}$$

$$+ \frac{1}{A-B}\,[\Phi\{\frac{a}{\sqrt{n}\Delta} - \frac{\sqrt{n}\Delta}{2}\} - \Phi\{\frac{b}{\sqrt{n}\Delta} - \frac{\sqrt{n}\Delta}{2}\}]$$

$$\beta_n = L(\theta_1)$$

$$= \frac{A}{A-B}\,\Phi\{\frac{b}{\sqrt{n}\Delta} - \frac{\sqrt{n}\Delta}{2}\} - \frac{B}{A-B}\,\Phi\{\frac{a}{\sqrt{n}\Delta} - \frac{\sqrt{n}\Delta}{2}\}$$

$$+ \frac{AB}{A-B}\,[\Phi\{\frac{a}{\sqrt{n}\Delta} + \frac{\sqrt{n}\Delta}{2}\} - \Phi\{\frac{b}{\sqrt{n}\Delta} + \frac{\sqrt{n}\Delta}{2}\}]. \qquad (3.3.12)$$

If we use the notation

$$\left.\begin{aligned} \Phi_+(a) &= \Phi\left(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}\Delta\right) \\ \text{and} \qquad & \\ \Phi_-(a) &= \Phi\left(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta\right) \end{aligned}\right\} \tag{3.3.13}$$

with similar notation for expressions in b, and for $\phi_+(a)$, $\phi_-(a)$, etc., then

$$\alpha_n = 1 + \frac{B}{A-B}\Phi_+(b) - \frac{A}{A-B}\Phi_+(a) + \frac{1}{A-B}\{\Phi_-(a) - \Phi_-(b)\}$$

$$\beta_n = \frac{A}{A-B}\Phi_-(b) - \frac{B}{A-B}\Phi_-(a) + \frac{AB}{A-B}\{\Phi_+(a) - \Phi_+(b)\} . \tag{3.3.14}$$

We next investigate some properties of the error probabilities. For a Wald SPRT, $\alpha + \beta < 1$ whenever $0 < B < 1 < A$. The same holds for the PSPRT considered in this section.

LEMMA 3.1. $\alpha_n + \beta_n < 1$ whenever $0 < B < 1 < A$.

Proof:

$$\alpha_n + \beta_n < 1 + \frac{AB - A + 1 - B}{A - B}\Phi_+(a) + \frac{B - AB + A - 1}{A - B}\Phi_+(b)$$

since

$$\Phi_-(a) < \Phi_+(a).$$

i.e.,

$$\alpha_n + \beta_n < 1 - \frac{(A-1)(1-B)}{A - B}\{\Phi_+(a) - \Phi_+(b)\}$$

$$< 1,$$

since $A - 1 < A - B$ and $B < 1$; and $\Phi_+(a) - \Phi_+(b) < 1$. Q.E.D.

It is easy to show that the conditional and nominal conditional error probabilities $\alpha'_{\underline{x}_n}$, $\beta'_{\underline{x}_n}$; and $\alpha_{\underline{x}_n}$, $\beta_{\underline{x}_n}$ satisfy

$$\alpha'_{\underline{x}_n} + \beta'_{\underline{x}_n} < \alpha_{\underline{x}_n} + \beta_{\underline{x}_n} . \tag{3.3.15}$$

But one cannot deduce that $\alpha'_n + \beta'_n < \alpha_n + \beta_n$, since integration is performed w.r.t. two different measures.

Behavior of Error Probabilities.

THEOREM 3.2. For given boundaries A and B, α_n decreases as n increases, whenever $a < \frac{1}{2}n\Delta^2$, or whenever $a > \frac{1}{2}n\Delta^2$ and $a - |b| < n\Delta^2$. Otherwise α_n increases as n increases.

β_n decreases as n increases, whenever $|b| < \frac{1}{2}n\Delta^2$, or whenever $|b| > \frac{1}{2}n\Delta^2$ and $|b| - a < n\Delta^2$. Otherwise β_n increases.

Proof: We consider the procedure as a continuous time process, and write t instead of n, with $t > 0$.

From Eqs. (3.3.12),

$$\begin{aligned}
\frac{\partial \alpha t}{\partial t} = {} & \frac{B}{A-B}\left[\frac{-b}{2t^{3/2}\Delta} + \frac{\Delta}{4t^{\frac{1}{2}}}\right]\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{b^2}{t\Delta^2} + \frac{t\Delta^2}{4} + b\right)\right] \\
& - \frac{1}{A-B}\left[\frac{b}{2t^{3/2}\Delta} - \frac{\Delta}{4t^{\frac{1}{2}}}\right]\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{b^2}{t\Delta^2} + \frac{t\Delta^2}{4} - b\right)\right] \\
& - \frac{A}{A-B}\left[\frac{-a}{2t^{3/2}\Delta} + \frac{\Delta}{4t^{\frac{1}{2}}}\right]\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{a^2}{t\Delta^2} + \frac{t\Delta^2}{4} + a\right)\right] \\
& + \frac{1}{A-B}\left[\frac{-a}{2t^{3/2}\Delta} - \frac{\Delta}{4t^{\frac{1}{2}}}\right]\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{a^2}{t\Delta^2} + \frac{t\Delta^2}{4} - a\right)\right]
\end{aligned}$$

and since $e^{-\frac{1}{2}b} = \frac{1}{\sqrt{B}}$, etc., we get

$$\frac{\partial \alpha_t}{\partial t} = \frac{\sqrt{B}}{A-B} \cdot \frac{\Delta}{2\sqrt{t}} \cdot \frac{1}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}(\frac{b^2}{t\Delta^2} + \frac{t\Delta^2}{4})]$$

$$- \frac{\sqrt{A}}{A-B} \cdot \frac{\Delta}{2\sqrt{t}} \cdot \frac{1}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}(\frac{a^2}{t\Delta^2} + \frac{t\Delta^2}{4})]$$

$$= \frac{1}{A-B} \cdot \frac{\Delta}{2\sqrt{t}} \cdot \frac{1}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}(\frac{b}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2})^2] - \exp[-\tfrac{1}{2}(\frac{a}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2})^2]\}$$

< 0 if

$$\exp[-\tfrac{1}{2}(\frac{b}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2})^2] < \exp[-\tfrac{1}{2}(\frac{a}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2})^2]$$

i.e.
$$(\frac{b}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2})^2 > (\frac{a}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2})^2$$

i.e.
$$-\frac{b}{\sqrt{t}\Delta} + \frac{\sqrt{t}\Delta}{2} > \left|\frac{a}{\sqrt{t}\Delta} - \frac{\sqrt{t}\Delta}{2}\right|,$$

since $b < 0$, $t > 0$,

i.e. $\frac{\partial \alpha_t}{\partial t} < 0$ if $-b + \frac{t\Delta^2}{2} > \left|a - \frac{t\Delta^2}{2}\right|$.

If $a < \frac{1}{2} t\Delta^2$, this gives $a > b$, which is so.

If $a > \frac{1}{2} t\Delta^2$, it gives

$$a + b = a - |b| < t\Delta^2.$$

Similarly

$$\frac{\partial\beta_t}{\partial t} = \frac{A}{A-B}\left\{\frac{-b}{2t^{3/2}\Delta} - \frac{\Delta}{4\sqrt{t}}\right\}\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{b^2}{t\Delta^2} + \frac{t\Delta^2}{4} - b\right)\right]$$

$$- \frac{B}{A-B}\left\{\frac{-a}{2t^{3/2}\Delta} - \frac{\Delta}{4\sqrt{t}}\right\}\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{a^2}{t\Delta^2} + \frac{t\Delta^2}{4} - a\right)\right]$$

$$+ \frac{AB}{A-B}\left\{\frac{-a}{2t^{3/2}\Delta} + \frac{\Delta}{4\sqrt{t}}\right\}\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{a^2}{t\Delta^2} + \frac{t\Delta^2}{4} + a\right)\right]$$

$$- \frac{AB}{A-B}\left\{\frac{-b}{2t^{3/2}\Delta} + \frac{\Delta}{4\sqrt{t}}\right\}\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{b^2}{t\Delta^2} + \frac{t\Delta^2}{4} + b\right)\right]$$

$$= \frac{\sqrt{A}\cdot B}{A-B}\,\frac{\Delta}{2\sqrt{t}}\cdot\frac{1}{\sqrt{2\pi}}\left\{\exp\left[-\tfrac{1}{2}\left(\frac{a^2}{t\Delta^2} + \frac{t\Delta^2}{4}\right)\right]\right\}$$

$$- \frac{A\cdot\sqrt{B}}{A-B}\,\frac{\Delta}{2\sqrt{t}}\,\frac{1}{\sqrt{2\pi}}\left\{\exp\left[-\tfrac{1}{2}\left(\frac{b^2}{t\Delta^2} + \frac{t\Delta^2}{4}\right)\right]\right\}$$

$$= \frac{AB}{A-B}\,\frac{\Delta}{2\sqrt{t}}\,\frac{1}{\sqrt{2\pi}}\left\{\exp\left[-\tfrac{1}{2}\left(\frac{a}{\sqrt{t}\Delta} + \frac{\sqrt{t}\Delta}{2}\right)^2\right] - \exp\left[-\tfrac{1}{2}\left(\frac{b}{\sqrt{t}\Delta} + \frac{\sqrt{t}\Delta}{2}\right)^2\right]\right\}$$

< 0 if

$$\left|\frac{a}{\sqrt{t}\Delta} + \frac{\sqrt{t}\Delta}{2}\right| > \left|\frac{b}{\sqrt{t}\Delta} + \frac{\sqrt{t}\Delta}{2}\right|$$

i.e. $\quad a + \frac{t\Delta^2}{2} > \left|b + \frac{t\Delta^2}{2}\right|, \quad t > 0, \quad a > 0.$

If $|b| > \tfrac{1}{2}t\Delta^2$, this gives

$$a + \tfrac{1}{2}t\Delta^2 > -b - \tfrac{1}{2}t\Delta^2$$

i.e. $\quad a + b > -t\Delta^2$

or $\quad (-b) - a < t\Delta^2$.

If $|b| < \tfrac{1}{2}t\Delta^2$, we get $a > b$ (true).

Q.E.D.

COROLLARY 3.3. For all symmetric procedures, with $b = -a$, α_n and β_n decrease as n increases, for all n.

The case $h(\theta) = 0$: i.e. $\theta = \frac{1}{2}(\theta_0 + \theta_1)$.

(3.3.11) gives $\frac{0}{0}$ if $h(\theta) = 0$, and hence we use d'Hôpital's Rule;

$$L_n\left(\frac{\theta_0+\theta_1}{2}\right) = \frac{1}{a-b}\left[a\Phi\left(\frac{a}{\sqrt{n}\Delta}\right) - b\Phi\left(\frac{b}{\sqrt{n}\Delta}\right) + \sqrt{n}\Delta\left\{\phi\left(\frac{a}{\sqrt{n}\Delta}\right) - \phi\left(\frac{b}{\sqrt{n}\Delta}\right)\right\}\right]. \quad (3.3.16)$$

Monotonicity of the O.C. Function.

This can be easily established, using the sufficient conditions for the exponential family. We know that the procedure terminates with probability 1;

$$\Pr\{\text{accept } H_1 | N = n'\} \neq 1 \implies \bar{x}_{n'} < \frac{a}{n'(\theta_1-\theta_0)} + \tfrac{1}{2}(\theta_1+\theta_0)$$

i.e. $$\Sigma_1^{n'} x_i < \frac{a}{\theta_1-\theta_0} + n \cdot \tfrac{1}{2}(\theta_1+\theta_0)$$

$$< nK \quad \text{for} \quad K = \tfrac{1}{2}(\theta_1+\theta_0), \quad \text{where} \quad \theta_1 > \theta_0 .$$

Similarly $\Pr\{\text{accept } H_1 | N = n'\} \neq 0 \implies \Sigma_1^{n'} x_i > nK$.

Hence $L_n(\theta)$ is non-increasing, in θ.

If $\theta_1 < \theta_0$, $L_n(\theta)$ is non-decreasing in θ.

3.4. Testing a Normal Mean with Known Variance: the Average Sample Number.

This again is constructed using Wald's approximation to the ASN of the conditional test $T(\underline{x}_n)$, and integrating over the distribution of $\bar{x}$ for test T_n.

From (3.1.10), and (3.3.4),

$$g_\theta = \Delta\theta - \tfrac{1}{2}\Delta(\theta_1+\theta_0)$$
$$= -\tfrac{1}{2}h\Delta^2.$$

So (3.1.12) gives

$$E_n(N|\theta) \simeq n - \frac{2}{h\Delta^2}\int_{b''_n}^{a''_n} [a - n(\Delta(\bar{x}-\theta) - \tfrac{1}{2}h\Delta^2)$$

$$- (a-b)\cdot\frac{e^{hn(\Delta(\bar{x}-\theta) - \frac{1}{2}h\Delta^2)} - e^{ha}}{e^{hb} - e^{ha}}]\, dF_\theta(\bar{x})$$

$$= n - \frac{2}{h\Delta^2}\{\tfrac{1}{2}h\Delta^2 + \frac{ae^{hb} - be^{ha}}{e^{hb} - e^{ha}}\}\{\Phi(\sqrt{n}(a''_n-\theta)) - \Phi(\sqrt{n}(b''_n-\theta))\}$$

$$+ \frac{2}{h\Delta^2}\cdot\frac{(a-b)}{e^{hb}-e^{ha}}\int_{b''_n}^{a''_n}\sqrt{\frac{n}{2\pi}}\exp[hn\Delta(\bar{x}-\theta-\tfrac{1}{2}h\Delta) - \frac{n}{2}(\bar{x}-\theta)^2]\, d\bar{x}$$

$$+ \frac{2n\Delta}{h\Delta^2}\int_{b''_n}^{a''_n}\sqrt{\frac{n}{2\pi}}\,(\bar{x}-\theta)\, e^{-\frac{n}{2}(\bar{x}-\theta)^2}\, d\bar{x}\ .$$

Since

$$\int_{b''_n}^{a''_n}(\bar{x}-\theta)\,dF_\theta(\bar{x}) = \left[-\frac{1}{\sqrt{2\pi n}}\, e^{-\frac{1}{2}ny^2}\right]_{y=b''_n-\theta}^{y=a''_n-\theta}$$

$$= \frac{1}{\sqrt{2\pi n}}\,[\exp\{-\tfrac{1}{2}(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}(\theta_1 + \theta_0 - 2\theta)^2\}$$

$$- \exp\{-\tfrac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}(\theta_1 + \theta_0 - 2\theta)^2\}],$$

and using the result of (3.3.6), in which

$$\sqrt{n}(a''_n - \theta) = \frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta, \quad \text{etc.},$$

we get

$$\begin{aligned} E_n(N|\theta) \simeq\ & n[1 - \Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta) + \Phi(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta)] \\ & - \frac{2}{h\Delta^2} \cdot \frac{ae^{hb}-be^{ha}}{e^{hb}-e^{ha}} \{\Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta) - \Phi(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta)\} \\ & + \frac{2}{h\Delta^2} \cdot \frac{a-b}{e^{hb}-e^{ha}} \{\Phi(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}h\Delta) - \Phi(\frac{b}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}h\Delta)\} \\ & + \frac{2}{h\Delta}\sqrt{\frac{n}{2\pi}} [\exp - \tfrac{1}{2}(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta)^2\} - \exp\{- \tfrac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}h\Delta^2\}] \end{aligned} \tag{3.4.1}$$

$$\begin{aligned} =\ & n[1 - \Phi_+(a,\theta) + \Phi_+(b,\theta)] \\ & - \frac{2}{h\Delta^2} \cdot \frac{ae^{hb}-be^{ha}}{e^{hb}-e^{ha}} \{\Phi_+(a,\theta) - \Phi_+(b,\theta)\} \\ & + \frac{2}{h\Delta^2} \cdot \frac{a-b}{e^{hb}-e^{ha}} \{\Phi_-(a,\theta) - \Phi_-(b,\theta)\} \\ & + \frac{2\sqrt{n}}{h\Delta} \{\phi_+(b,\theta) - \phi_+(a,\theta)\}\ . \end{aligned} \tag{3.4.2}$$

An alternative form can be obtained using the concise form (3.3.11) for the O.C. function.

$$E_n(N|\theta) \approx \frac{1}{E_\theta(Z)} \{bL_n(\theta) + a(1 - L_n(\theta))\}$$

$$+ n[1 - \Phi_+(a,\theta) + \Phi_+(b,\theta)]$$

$$- \frac{1}{E_\theta(Z)} [b\Phi_+(b,\theta) + a(1 - \Phi_+(a,\theta)) + \sqrt{n}\Delta\{\phi_+(b,\theta) - \phi_+(a,\theta)\}] \quad (3.4.3)$$

where the first term is the Wald form of the approximation to $E(N|\theta)$ for a SPRT procedure. It can be seen from (3.4.3) that as n increases, the terms in n and $\sqrt{n}$ begin to dominate the right-hand side.

The A.S.N. when $h = 0$ (i.e. $\theta = \frac{1}{2}(\theta_0+\theta_1)$)

Since we cannot apply Wald's equation

$$E(\sum_{i=1}^{N} Z_i) = E(N)\, E(Z_1)$$

when $E(Z_1) = 0$, the case $h = 0$ leads to a different approximation to the ASN in a SPRT. We use the result [10]

$$E\{(\sum_{i=1}^{N} Z_i')^2\} = E(N)\, \mathrm{Var}(Z_1'), \quad \text{where} \quad E(Z_1') = 0. \quad (3.4.4)$$

Here
$$Z_i' = (\theta_1-\theta_0)X_i - \tfrac{1}{2}(\theta_1^2-\theta_0^2)$$

$$= \log \frac{p_1(X_i)}{p_0(X_i)} .$$

So
$$\mathrm{Var}(Z_i') = (\theta_1-\theta_0)^2 = \Delta^2 .$$

For the conditional test $T(\underline{x}_n)$, suppose (3.1.2) holds, so that a SPRT procedure begins with the $(n+1)$th observation.

Then

$$E\{(\sum_{i=n+1}^{N} Z_i')^2\} = b_n'^2 \Pr(\text{accept } H_0) + a_n'^2 \Pr(\text{accept } H_1) ,$$

neglecting excess over the conditional boundaries a_n' and b_n',

$$= (b - n(\theta_1-\theta_0)\bar{x}_n + \tfrac{n}{2}(\theta_1^2-\theta_0^2))^2 \, L_{\underline{x}_n}(\tfrac{\theta_0+\theta_1}{2})$$

$$+ (a - n(\theta_1-\theta_0)\bar{x}_n + \tfrac{n}{2}(\theta_1^2-\theta_0^2))^2 \, \{1 - L_{\underline{x}_n}(\tfrac{\theta_1+\theta_0}{2})\}.$$

When (3.1.2) holds and $h = 0$,

$$L_{\underline{x}_n}(\theta) = \frac{a_n'}{a_n'-b_n'}, \quad \text{so} \quad 1 - L_{\underline{x}_n}(\theta) = -\frac{b_n'}{a_n'-b_n'}; \quad \text{and} \quad a_n' - b_n' = a - b;$$

therefore

$$E\{\sum_{i=n+1}^{N} Z_i'\}^2 = b_n'^2 \frac{a_n'}{a-b} - a_n'^2 \frac{b_n'}{a-b}$$

$$= - a_n'b_n' \qquad (3.4.5)$$

and so if (3.1.2) holds,

$$E_{\underline{x}_n}(N|\theta = \frac{\theta_0+\theta_1}{2}) = n - \frac{1}{\Delta^2} a_n'b_n'$$

i.e.

$$E_{\underline{x}_n}(N|\theta = \frac{\theta_1+\theta_0}{2}) \simeq \begin{cases} n & \text{if } \frac{p_{1n}}{p_{0n}} \le B \text{ or } \ge A \\ n - \{b - n\Delta(\bar{x}-\theta)\}\{a - n\Delta(\bar{x}-\theta)\} \cdot \frac{1}{\Delta^2} & \\ \qquad \text{if } B < \frac{p_{1n}}{p_{0n}} < A \end{cases} .$$

Hence unconditionally,

$$E_n(N|\theta = \frac{\theta_1+\theta_0}{2}) \simeq n - \frac{1}{\Delta^2}\int_{b''_n < \bar{x} < a''_n} [ab - n\Delta(a+b)(\bar{x}-\theta)$$

$$+ n^2\Delta^2(\bar{x}-\theta)^2]dF(\bar{x})$$

$$= n - \frac{ab}{\Delta^2}\{\Phi_+(a,\theta) - \Phi_+(b,\theta)\}$$

$$+ \frac{(a+b)\sqrt{n}}{\Delta}\{\phi_+(b,\theta) - \phi_+(a,\theta)\}$$

$$- n^2\int_{b''_n}^{a''_n}\sqrt{\frac{n}{2\pi}}(\bar{x}-\theta)^2 \exp(-\frac{n}{2}(\bar{x}-\theta)^2)d\bar{x} \ . \qquad (3.4.6)$$

The last term

$$= -\frac{n}{\sqrt{2\pi}}\{\left[-ye^{-\frac{1}{2}y^2}\right]_{\frac{b}{\sqrt{n}\Delta}}^{\frac{a}{\sqrt{n}\Delta}} + \int_{\frac{b}{\sqrt{n}\Delta}}^{\frac{a}{\sqrt{n}\Delta}} e^{-\frac{1}{2}y^2}\,dy\}$$

where $y = \sqrt{n}(\bar{x}-\theta)$

at $\theta = \frac{1}{2}(\theta_1+\theta_0)$.

So the last term

$$= \frac{n}{\sqrt{2\pi}}\left\{\frac{a}{\sqrt{n}\Delta}\, e^{-\frac{1}{2}\frac{a^2}{n\Delta^2}} - \frac{b}{\sqrt{n}\Delta}\, e^{-\frac{1}{2}\frac{b^2}{n\Delta^2}}\right\}$$

$$- n\{\Phi_+(a,\theta) - \Phi_+(b,\theta)\}$$

$$= \frac{a\sqrt{n}}{\Delta}\phi_+(a,\theta) - \frac{b\sqrt{n}}{\Delta}\phi_+(b,\theta) - n\{\Phi_+(a,\theta) - \Phi_+(b,\theta)\} \ .$$

Hence (3.4.6) gives concisely

$$E_n(N|\frac{\theta_1+\theta_0}{2}) \simeq n - (n + \frac{ab}{\Delta^2})\{\Phi(\frac{a}{\sqrt{n}\Delta}) - \Phi(\frac{b}{\sqrt{n}\Delta})\}$$

$$+ \frac{a\sqrt{n}}{\Delta}\,\phi(\frac{b}{\sqrt{n}\Delta}) - \frac{b\sqrt{n}}{\Delta}\,\phi(\frac{a}{\sqrt{n}\Delta}) \; . \qquad (3.4.7)$$

An alternative form can be obtained in terms of $L_n(\theta)$. In the case of a SPRT, with OC function $L(\theta)$, the approximation to the ASN is given by

$$E(N|\frac{\theta_1+\theta_0}{2}) \simeq \frac{1}{\Delta^2}\,[b^2\,L(\theta) + a^2(1 - L(\theta))] \; .$$

Here, (3.3.16) and (3.4.7) lead to the form

$$E_n(N|\frac{\theta_1+\theta_0}{2}) \simeq n(1 - \Phi(\frac{a}{\sqrt{n}\Delta}) + \Phi(\frac{b}{\sqrt{n}\Delta}))$$

$$+ \frac{1}{\Delta^2}\,[b^2\,L_n(\frac{\theta_1+\theta_0}{2}) + a^2(1 - L_n(\frac{\theta_1+\theta_0}{2}))]$$

$$- \frac{b^2}{\Delta^2}\,\Phi(\frac{b}{\sqrt{n}\Delta}) - \frac{a^2}{\Delta^2}(1 - \Phi(\frac{a}{\sqrt{n}\Delta})) + \frac{a\sqrt{n}}{\Delta}\,\phi(\frac{a}{\sqrt{n}\Delta}) - \frac{b\sqrt{n}}{\Delta}\,\phi(\frac{b}{\sqrt{n}\Delta}).$$

(3.4.8)

3.5. Numerical Results and Comparisons.

In this section, the results obtained for testing a normal mean will be used to present some interesting and useful numerical results and comparisons.

Tables 3.5.1 and 3.5.2 give the OC function and ASN of the PSPRT when $\Delta = .50$, $n = 1, 5, 15$ and 30, for the two cases $a = -b = 3$

and $a = -b = 5$. In this last case, when the boundaries are farther apart, the ASN shows virtually no increase from $n = 1$ to $n = 5$, and only begins to increase substantially with n for $n > 15$.

TABLES 3.5.

OC function and ASN for PSPRT; test of normal mean.

$\Delta = \theta_1 - \theta_0$; $n = 1, 5, 15, 30$.

Table 3.5.1. $\Delta = .50$ $a = -b = 3$

OC Function

$\theta - \theta_0 \backslash n$	1	5	15	30
.50	.0474	.0472	.0398	.0251
.375	.1824	.1822	.1734	.1510
.25	.5000	.5000	.5000	.5000

ASN

$\theta - \theta_0 \backslash n$	1	5	15	30
.50	21.724	21.787	24.758	34.811
.375	30.487	30.550	33.386	42.784
.25	36.000	36.066	39.036	48.751

Table 3.5.2. $\Delta = .50$ $a = -b = 5$

OC Function

$\theta - \theta_0 \backslash n$	1	5	15	30
.50	.00669	.00669	.00653	.00522
.375	.0759	.0759	.0755	.0718
.25	.5000	.5000	.5000	.5000

ASN

$\theta - \theta_0 \backslash n$	1	5	15	30
.50	39.465	39.465	39.826	44.214
.375	67.863	67.863	68.112	71.197
.25	100.00	100.00	100.27	103.50

Comparison with Wald SPRT.

The following table compares the ASN of a PSPRT with a Wald SPRT having the same error probabilities α, and a fixed sample-size test:

Pr{wrong decision$|H_i$} = α; $i = 0, 1$. α, β and $E(N)$ are nominal.

Table 3.5.3. $\Delta = \theta_1 - \theta_0 = 1$, $n = 15$, $a = -b = 5$, $\alpha = \beta = .002356$

$\theta - \theta_0$	$E(N)$ (Wald)	$E(N\|n = 15)$ (PSPRT)	Fixed sample size n'
0.5	36.554	29.748	31.945
0.0	12.035	16.164	31.945

The PSPRT procedure does better than the fixed sample size procedure, and improves the SPRT at $h = 0$ $(\theta - \theta_0 = \frac{1}{2})$. The difference in observations required at θ_1 or θ_0 is 4.

However, for $a = -b = 5.0$, the nominal SPRT error probabilities are .0067, so that fixing $n = 15$ with these boundaries reduces the error probabilities here by about 3/5.

Tables 3.5.1 and 3.5.2 demonstrate what intuition would suggest, namely that separating the boundaries a and b means that a larger value of n may be allowed before $E(N)$ begins to increase substantially, or $L_n(\theta)$ to change substantially. Separating the boundaries results of course in decreased error probabilities.

Comparison with Other Methods and Exact Results.

As noted earlier, the approximations given above for $L_n(\theta)$ and $E_n(N|\theta)$ apply to a Wald SPRT when $n = 1$. Effectively they give exact probabilities of stopping at the first observation, and are approximate thereafter. We give some numerical results below. It will be seen that the differences from Wald's approximations are small, but that they are nevertheless an improvement.

Using a method of Page [15], Kemp [13] developed in 1958 formulae for calculating other approximations to $L(\theta)$ and $E(N|\theta)$ in the SPRT procedure. The following tables are partially based upon his paper.

Table 3.5.4. ASN: $\theta_0 = -\frac{1}{2}$, $\theta_1 = \frac{1}{2}$, $h = -2\theta$. $a = 7.5$, $b = -2.5$; $n = 1$ in PSPRT

θ	Wald	PSPRT	Kemp	Exact
-.50	4.990	5.007	7.0	6.4
-.25	9.324	9.341	13.0	12.0
0	18.750	18.771	27.7	25.2
.25	18.733	18.751	24.8	23.2
.50	13.359	13.369	16.2	15.4

The improvement in approximating to the ASN using the PSPRT approach is less than 1.2% over Wald's formulae. Kemp appears to have some errors in computing the Wald ASN: for $\theta = -.25$, .25 and .50, he gets ASN values of 8.7, 15.8 and 12.3 respectively.

The next table gives a similar comparison for the O.C. function.

Table 3.5.5. OC Function: $\theta_0 = -\frac{1}{2}$, $\theta_1 = \frac{1}{2}$, $h = -2\theta$. $a = 7.5$, $b = -2.5$. $n = 1$ in PSPRT.

θ	Wald	PSPRT	Kemp	Exact
-0.25	.9831	.9831	.9872	.9861
-0.125	.9223	.9223	.9254	.9241
0	.7500	.7498	.7155	.7239
0.125	.4942	.4939	.4058	.4276
0.25	.2817	.2811	.1896	.2113

Kemp also makes a comparison for $a = 5.0 = -b$. (3.3.11) gives the same results as Wald's approximation, since the c.d.f. terms are $\Phi(5\pm\theta)$ or $\Phi(-5\pm\theta)$, which are nearly 1 or 0 in the range of θ compared. In the above table, however, terms in $\Phi(-2.5\pm\theta)$ and $\Phi(7.5\pm\theta)$ appear, allowing small improvements over Wald's results.

It appears therefore that the PSPRT, which is a SPRT when $n = 1$, only shows small improvements in the approximations to $L(\theta)$ and $E(N|\theta)$, and these become negligible when the boundaries of the test procedure are further apart.

Although exact values are not available, we give a second example, with closer boundaries, giving approximate error probabilities .022 and .097.

Table 3.5.6. $\theta_1 - \theta_0 = 1$; $a = 3.2$, $b = -2.3$. $n = 1$ for PSPRT data.

$\theta - \theta_0$	Wald OC	PSPRT OC	Wald ASN	PSPRT ASN
0	.96317	.96316	4.1949	4.2233
.25	.85261	.85244	5.9574	5.9830
.50	.58182	.58119	7.3600	7.3836
.75	.26997	.26891	6.8607	6.8790
1.00	.09657	.09552	5.3378	5.3499

Comparison with Anderson's Modified SPRT [1].

It was noted earlier that at $h = 0$, the procedure we are considering can do better than either the SPRT or fixed sample-size procedures (the latter may indeed do better than the SPRT at $h = 0$). Anderson [1] developed approximations to the OC function and ASN for a sequential test having boundaries which are linear functions of n', the number of observations. He obtained ASN values which are close to the lower bound at $h = 0$, but which are only slightly greater than the optimum values of the SPRT at $h = \pm 1$. He obtained his results for symmetric tests having $\alpha = \beta = 0.05$ and 0.01, with $\theta_1 - \theta_0 = .25$.

In the PSPRT procedure, it was found that by varying both n and $a(= -b)$, and keeping the error probabilities constant, the ASNs at all values of θ would vary, increasing with n as shown below.

Tables 3.5.7./8. Comparison of PSPRT with Anderson's Modified SPRT for given Error Probabilities.

Table 3.5.7. $\alpha = \beta = .05$ $\Delta = \theta_1 - \theta_0 = 0.25$ ($n = 0$ gives SPRT data)

$a = -b$	Initial fixed n	ASN (approx: $h = \pm 1$)	ASN (approx: $h = 0$)
2.944	0	132.50	216.70
2.7	96	143.88	208.59
2.6	111	149.86	207.44
2.5	124	155.96	206.96
2.4	136	162.40	207.33
2.2	158	176.04	210.34
Lower Bound		132.5	187.0
Anderson Modified SPRT		139.2	192.2
Fixed Sample-size		270.6	270.6

The ASN of the PSPRT reaches a minimum for $h = 0$ at $n \simeq 125$, but the ASN for $h = \pm 1$ is increasing with n.

Table 3.5.8. $\alpha = \beta = .01$ $\Delta = \theta_1 - \theta_0 = 0.25$ ($n = 0$ gives SPRT data)

a = -b	Initial fixed n	ASN (approx: h = ±1)	ASN (approx: h = 0)
4.59	0	225.20	527.90
4.3	173	244.79	492.03
4.0	228	271.20	468.35
3.7	272	299.43	453.61
3.5	298	318.50	447.87
3.4	310	327.78	445.86
3.3	323	338.29	445.69
3.0	357	366.88	446.41
Lower Bound		225.2	388.3
Anderson Modified SPRT		249.4	402.2
Fixed Sample-size		541.2	541.2

It can be seen that Anderson's procedure is better than the PSPRT in general. The ASN of the PSPRT does reach a minimum for $h = 0$ at $n \simeq 320$, but the ASN for $h = \pm 1$ shows consistent changes as this occurs. The optimum combination for a PSPRT thus depends on the circumstances, but we might take $a = -b = 3.5$, $n = 298$, since the extra 20 observations (on average) at $h = \pm 1$ might not make the 2.18 improvement at $h = 0$ worthwhile.

In Table 3.5.7, the partial procedure does produce an ASN at $h = 0$ which is a substantial gain; an optimum procedure might be for $a = -b = 2.7$, $n = 96$. The results indicate that in order to improve on the SPRT at $h = 0$, the error probabilities should be fairly small.

Anderson's results are much better. The PSPRT procedure could be used, however, when some results are already available, or if the cost is unknown, a situation which will be considered in Chapter V. Computationally, the PSPRT may be easier to work with than Anderson's procedure, although there seems no analytical method of finding optimum values of a, b and n corresponding to given error probabilities in the case of the PSPRT. The combinations of a, b and n in the above examples were obtained by trial and error.

CHAPTER IV

ON MINIMIZING THE EXCESS RISK WHEN THE COST PER OBSERVATION IS UNKNOWN.

4.1. Introduction.

In this chapter, we shall consider the problem of obtaining optimum procedures when the cost per observation is unknown. The criterion for an 'optimum' procedure in a class C of tests will be that such a procedure minimizes the excess risk over the Chernoff a.m. risk, or that it minimizes the average excess risk when the cost of an observation is a random variable having a known distribution. Throughout the discussion, the variability in cost will be assumed to be independent of the observations.

In Section 4.2, expressions will be derived for the nominal excess risk when the true cost c is erroneously assumed to be some quantity $c' \neq c$. These will be derived both for symmetric SPRTs and general SPRTs, using results in Chapter 2. In Section 4.3, the approximate average excess risk will be derived when a previous estimate $\bar{c}_n$ of the cost is available, for symmetric SPRTs and general SPRTs, and when the cost c_i per observation is assumed to have a certain Gamma distribution.

Throughout the chapter, it will be assumed that the appropriate a.m. risks are based on SPRTs having at least one observation, and that the nominal error probabilities which minimize the risk also satisfy $\alpha + \beta < 1$.

4.2. Excess Risk due to Use of Incorrect Cost.

Symmetric SPRTS.

We use the notation introduced in Chapter II. It will be recalled that for a minimizing error probability α, the corresponding nominal risk is

$$R_{\delta'}(\lambda_w) \simeq [L_0 w + L_1(1-w)]\alpha + (k_1-k_0)\ (1-2\alpha)\ \log \frac{1-\alpha}{\alpha}\ . \tag{4.2.1}$$

We shall express (4.2.1) in terms of Chernoff's approximation to the minimizing error probability, i.e., from (2.3.6)

$$\alpha = \frac{k_1 - k_0}{L_0 w + L_1(1-w)}, = \gamma c, = \frac{k_1 - k_0}{V}, \tag{4.2.2}$$

where the notation γ and V is introduced for convenience. Then

$$R_{\delta'}(\lambda_w) \simeq (k_1-k_0)[1 + (1 - \frac{2(k_1-k_0)}{V})\ \log \{\frac{V - (k_1-k_0)}{k - k}\}]$$

$$= V\gamma c[1 + (1-2\gamma c)\ \log\ (\frac{1 - \gamma c}{\gamma c})]\ . \tag{4.2.3}$$

If $\gamma c = \alpha$ is supposed small,

$$R_{\delta'}(\lambda_w) \simeq V\gamma c[1 + (1-2\gamma c)\ \{-\gamma c - \log\gamma c\}] + 0(\gamma^3 c^3)$$

$$= V\gamma c(1 - \log\gamma c + 2\gamma c\ \log\gamma c - \gamma c) + 0(\gamma^3 c^3). \tag{4.2.4}$$

Now suppose that the value c' is assumed for the cost per observation, while the true cost is c. Then the risk will be "minimized" by the error probability $\alpha' = \gamma c'$, and the actual risk is approximately

$$V\alpha' + (k_1-k_0)(1-2\alpha') \log \frac{1-\alpha'}{\alpha'}$$

$$= V\gamma[c' + c(1-2\gamma c') \log (\frac{1-\gamma c'}{\gamma c'})]. \qquad (4.2.5)$$

Since the wrong value has been assigned to cost, this risk will be larger than the desired minimum. Hence if we define excess risk to be that amount by which it exceeds the a.m. risk, and using Chernoff's approximation,

$$\text{Excess risk} \simeq V\gamma[(c'-c) + c\log \frac{1-\gamma c'}{\gamma c'} - c\log \frac{1-\gamma c}{\gamma c}$$

$$- 2\gamma c\{c'\log \frac{1-\gamma c'}{\gamma c'} - c\log \frac{1-\gamma c}{\gamma c}\}]$$

$$\simeq V\gamma[(c' - c + c\log \frac{c}{c'}) + c\gamma(c-c')$$

$$- 2\gamma c(c\log\gamma c - c'\log\gamma c')] + O(\gamma^3c^3),$$

where c' is assumed to be of the same order of smallness as c. So

$$\text{Excess risk} \simeq V\gamma(c' - c + c\log \frac{c}{c'}) + V\gamma^2c[-(1-2\log\gamma)\ (c'-c)$$

$$- 2(c\log c - c'\log c')] + O(\gamma^3c^3). \qquad (4.2.6)$$

In (4.2.6), the expression $V\gamma(c' - c + c\log \frac{c}{c'})$ plays the dominant part, under the assumption that c' and c are both small. It will be seen presently that a similar expression dominates the Chernoff excess risk for more general SPRT's; later it will be studied more closely.

General SPRTs.

We return to Chernoff's approximating error probabilities (2.4.7), i.e.,

$$\alpha_0 = \frac{k_1}{L_0 w}, \quad \beta_0 = -\frac{k_0}{L_1(1-w)}, \tag{4.2.7}$$

where, if c is the true cost,

$$k_0 = -\frac{2cw}{\Delta^2}, \quad k_1 = \frac{2c(1-w)}{\Delta^2},$$

from (2.1.3). The Chernoff a.m. risk is

$$R_\delta(\lambda_w) \simeq k_1 - k_0 + \{k_0(1 - \frac{k_1}{L_0 w}) + k_1(-\frac{k_0}{L_1(1-w)})\} \log\left(\frac{-\frac{k_0}{L(1-w)}}{1 - \frac{k_1}{L_0 w}}\right)$$

$$+ \{\frac{k_0 k_1}{L_0 w} + k_1(1 + \frac{k_0}{L_1(1-w)})\} \log\left(\frac{1 + \frac{k_0}{L_1(1-w)}}{\frac{k_1}{L_0 w}}\right).$$

Let $k_0 = \gamma_0 c$ and $k_1 = \gamma_1 c$; write

$$\Lambda = \frac{1}{L_0 w} + \frac{1}{L_1(1-w)}. \tag{4.2.8}$$

We temporarily introduce the notation

$$\left.\begin{aligned} U(c) &= \frac{-\dfrac{k_0}{L_1(1-w)}}{1-\dfrac{k_1}{L_0 w}} = \frac{-\dfrac{\gamma_0}{L_1(1-w)}}{\dfrac{1}{c}-\dfrac{\gamma_1}{L_0 w}} . \\ W(c) &= \frac{1+\dfrac{k_0}{L_1(1-w)}}{\dfrac{k_1}{L_0 w}} = \frac{\dfrac{1}{c}+\dfrac{\gamma_0}{L_1(1-w)}}{\dfrac{\gamma_1}{L_0 w}} \end{aligned}\right\} . \qquad (4.2.9)$$

Then

$$R_\delta(\lambda_w) \simeq (\gamma_1-\gamma_0)c + c\{\gamma_0-\gamma_0\gamma_1 c\Lambda\}\log(U(c)) + c\{\gamma_1+\gamma_1\gamma_0 c\Lambda\}\log(W(c)). \qquad (4.2.10)$$

If c' is supposed - wrongly - to be the value of the cost, the 'supposed' a.m. risk will be based on c'. The actual nominal risk is then

$$(\gamma_1-\gamma_0)c' + c\{\gamma_0-\gamma_0\gamma_1 c'\Lambda\}\log(U(c')) + c\{\gamma_1+\gamma_0\gamma_1 c'\Lambda\}\log(W(c')). \qquad (4.2.11)$$

The excess risk is defined:

$$\text{Excess risk} \simeq (\gamma_1-\gamma_0)(c'-c) + c\gamma_0 \log\left(\frac{U(c')}{U(c)}\right) + c\gamma_1 \log\left(\frac{W(c')}{W(c)}\right)$$

$$+ \gamma_0\gamma_1\Lambda c[c'\log\left(\frac{W(c')}{U(c')}\right) - c\log\left(\frac{W(c)}{U(c)}\right)]$$

$$= (\gamma_1-\gamma_0)(c'-c) + c\gamma_0[\log(\frac{1}{c} - \frac{\gamma_1}{L_0w}) - \log(\frac{1}{c'} - \frac{\gamma_1}{L_0w})]$$

$$+ c\gamma_1[\log(\frac{1}{c'} + \frac{\gamma_0}{L_1(1-w)}) - \log(\frac{1}{c} + \frac{\gamma_0}{L_1(1-w)})]$$

$$- \gamma_0\gamma_1 c\Lambda(c'-c) \log(- \frac{\gamma_0\gamma_1}{L_0L_1w(1-w)})$$

$$+ \gamma_0\gamma_1 c\Lambda[c'\log\{(\frac{1}{c'} + \frac{\gamma_0}{L_1(1-w)})(\frac{1}{c'} - \frac{\gamma_1}{L_0w})\}$$

$$- c\log\{(\frac{1}{c} + \frac{\gamma_0}{L_1(1-w)})(\frac{1}{c} - \frac{\gamma_1}{L_0w})\}]$$

$$\approx (\gamma_1-\gamma_0)(c' - c + c\log \frac{c}{c'}) + c\gamma_0 \frac{\gamma_1}{L_0w} (c'-c)$$

$$+ c\gamma_1 \frac{\gamma_0}{L_1(1-w)} (c'-c) - \gamma_0\gamma_1 c\Lambda(c'-c) \log(- \frac{\gamma_0\gamma_1}{L_0L_1w(1-w)})$$

$$+ \gamma_1\gamma_0 c\Lambda(2c'\log \frac{1}{c'} - 2c\log \frac{1}{c}) + O(c^3)$$

$$= (\gamma_1-\gamma_0) (c' - c + c\log \frac{c}{c'})$$

$$+ c(c'-c)\gamma_0\gamma_1\Lambda\{1 - \log(- \frac{\gamma_0\gamma_1}{L_0L_1w(1-w)})\}$$

$$+ 2c\gamma_0\gamma_1\Lambda(c\log c - c'\log c') + O(c^3). \qquad (4.2.12)$$

In (4.2.10), as in (4.2.6), the leading term dominates, i.e.,

$$(\gamma_1 - \gamma_0) (c' - c + c \log \frac{c}{c'}) .$$

The coefficients $\gamma_1 - \gamma_0$ and $V\gamma$ represent the same quantity, i.e. $\frac{k_1 - k_0}{c}$

$$c' - c + c \log \frac{c}{c'} = c(\frac{c'}{c} - 1 - \log \frac{c'}{c})$$

$$\geq 0,$$

since $x - 1 - \log x \geq 0$ for every $x > 0$, and we assume $c > 0$, $c' > 0$.

So the dominant term is non-negative. This is of course desirable, since here we are dealing with approximations to the a.m. risk and to the excess risk, and it is important that non-negativeness of excess in the true risks should be reflected in the approximations.

4.3. Average Excess with Variable Cost.

Suppose that the cost per observation, c_i, is a random variable, independent of X. In this section, we shall derive the Chernoff excess risk, averaged over the distribution of c_i, when the mean observed cost

$$\bar{c}_n = \frac{1}{n} \sum_{j=1}^{n} c_j$$

from a previous experiment is used to estimate the true mean cost c. Thus $E(\bar{c}_n) = c$, and if c were known, the a.m. risk could be attained as shown in Chapter II.

We suppose that c_i has a Gamma distribution, i.e.,

$$p(c_i) = \frac{1}{\Gamma(m)} (\frac{m}{c})^m c_i^{m-1} e^{-\frac{m}{c} c_i}, \quad c_i > 0. \qquad (4.3.1)$$

Then $E(c_i) = c$, and $Var(c_i) = \frac{c^2}{m}$

Notationally, write

$$c_i \sim \Gamma(m, \frac{m}{c}) .$$

Then

$$c_n \sim \Gamma(mn, \frac{mn}{c}) .$$

Symmetric SPRTs.

In (4.2.6), if c' is replaced by $\bar{c}_n$, we get the Chernoff excess risk when the estimate $\bar{c}_n$ is used for cost instead of c. Hence if the average excess is defined as the expectation of the excess risk,

$$\begin{aligned} \text{Average excess} \simeq\ & V\gamma E(\bar{c}_n - c + c\log c - c\log\bar{c}_n) \\ & + V\gamma^2 cE[-(1-2\log\gamma)(\bar{c}_n - c) - 2(c\log c - \bar{c}_n\log\bar{c}_n)] \\ & + O(\gamma^3 c^3) \\ =\ & V\gamma c(\log c - E\log\bar{c}_n) - 2V\gamma^2 c(c\log c - E(c_n\log\bar{c}_n)) \\ & + O(\gamma^3 c^3). \end{aligned}$$

Since $\bar{c}_n$ is a Gamma variable,

$$E \log \bar{c}_n = - \log \frac{mn}{c} + \psi(mn), \tag{4.3.2}$$

where $\psi(\cdot)$ is the Digamma Function $\frac{d}{dZ}\{\log \Gamma(Z)\}$; and if $X \sim \Gamma(m', \theta)$,

$$E(X\log X) = \int_0^\infty \frac{\theta^{m'}}{\Gamma(m')} (\log x) x^{(m'+1)-1} e^{-\theta x} dx$$

$$= \frac{m'}{\theta} E(\log Y) \quad \text{where} \quad Y \sim \Gamma(m'+1, \theta)$$

$$= \frac{m'}{\theta} (-\log\theta + \psi(m'+1))$$

$$= \frac{m'}{\theta} (-\log\theta + \psi(m') + \frac{1}{m'}) .$$

Hence

$$E(\bar{c}_n \log\bar{c}_n) = -c\log\frac{mn}{c} + c(\psi(mn) + \frac{1}{mn}). \qquad (4.3.3)$$

(4.3.2) and (4.3.3) give above,

$$\text{Average excess} \simeq V\gamma c(\log(mn) - \psi(mn))$$

$$- 2V\gamma^2 c^2\{\log(mn) - \psi(mn) - \frac{1}{mn}\} + 0(\gamma^3 c^3)$$

$$= V\gamma c(1-2\gamma c)\{\log(mn) - \psi(mn)\} + 2V\gamma^2 \frac{c^2}{mn} + 0(\gamma^3 c^3) . \qquad (4.3.4)$$

In many problems, the variance of c_i will be small, even when $E(c_i)$ is not necessarily small. Hence m may be large. Alternatively, n may be large, and the Gamma-type distribution of $\bar{c}_n$ will be close to normality. The asymptotic form of the Diagamma Function can then be used, i.e.,

$$\psi(Z) \sim \log Z - \frac{1}{2Z} - \sum_{j=1}^{\infty} \frac{B_{2j}}{2jZ^{2j}} , \qquad (4.3.5)$$

where B_{2j} are Bernoulli numbers. Then

$$\psi(Z) \sim \log Z - \frac{1}{2Z} - \frac{1}{12Z^2} + \frac{1}{120Z^4} - \frac{1}{252Z^6} + \frac{1}{240Z^8} - \cdots\cdots \quad . \tag{4.3.6}$$

Thus, if mn is reasonably large, (4.3.4) and (4.3.6) give

$$\text{Average excess} \sim \frac{V\gamma c(1+2\gamma c)}{2mn} + V\gamma c(1-2\gamma c)\left(\frac{1}{12m^2n^2} - \frac{1}{120m^4n^4} + \cdots\right) . \tag{4.3.7}$$

In (4.3.7), it can be seen that if $mn \geq 10$, and if c is small, the average excess, given that the true cost is c, is of the order of

$$\frac{k_1-k_0}{2mn} = \frac{c}{2mn}\left\{\frac{1-w}{E(Z_1|\theta_1)} - \frac{w}{E(Z_1|\theta_0)}\right\} . \tag{4.3.8}$$

General SPRTs.

We now replace c' by $\bar{c}_n$ in (4.2.10) and take expectations, to obtain

$$\begin{aligned}
\text{Average excess} &\simeq (\gamma_1-\gamma_0)E(\bar{c}_n - c + c\log c - c\log\bar{c}_n) \\
&\quad + c\gamma_0\gamma_1\Lambda\{1 - \log(-\frac{\gamma_0\gamma_1}{L_0L_1w(1-w)})\}\, E(\bar{c}_n-c) \\
&\quad + 2c\gamma_0\gamma_1\Lambda\{c\log c - E(\bar{c}_n\log\bar{c}_n)\} + O(c^3) \\
&= (\gamma_1-\gamma_0)c(\log(mn) - \psi(mn)) \\
&\quad + 2c^2\gamma_0\gamma_1\Lambda(\log(mn) - \psi(mn) - \frac{1}{mn}) + O(c^3) \\
&= c\{(\gamma_1-\gamma_0) + 2\gamma_0\gamma_1\Lambda c\}(\log(mn) - \psi(mn)) - \frac{2c^2\gamma_0\gamma_1\Lambda}{mn} \\
&\quad + O(c^3).
\end{aligned} \tag{4.3.9}$$

Asymptotically, if mn is large, we may use (4.3.6), obtaining

$$\text{Average excess} \sim \frac{c\{(\gamma_1-\gamma_0) - 2\gamma_0\gamma_1\Lambda c\}}{2mn} + c\{(\gamma_1-\gamma_0)$$

$$+ 2\gamma_0\gamma_1\Lambda c\} \left(\frac{1}{12m^2n^2} - \frac{1}{120m^4n^4} + \cdots\right) . \tag{4.3.10}$$

If $mn \geq 10$, say, and if c is small, then as in (4.3.7), the average excess is of the order of $(k_1 - k_0)/2mn$.

Since $\gamma_0 < 0$ and $\gamma_1 > 0$, there is no inconsistency in the opposite signs which appear in the first terms of (4.3.7) and (4.3.10), i.e. in the quantity $\gamma(1+2\gamma c)$ of (4.3.7) and in the quantity $(\gamma_1-\gamma_0) - 2\gamma_0\gamma_1\Lambda c$ of (4.3.10).

The smallness of the approximate average excess is derived from (4.3.7) and (4.3.10), i.e. $(k_1 - k_0)/2mn$, may more appropriately be studied in its ratio to the Chernoff a.m. risks in (4.2.3) and (4.2.10).

To terms in $c \log c$, the leading quantity in (4.2.3) is

$$V\gamma c(1 - \log\gamma c)$$

which dominates the risk $R_{\delta'}(\lambda_w)$ for symmetric SPRTs. The ratio of the approximate average excess to this risk is of the order

$$\frac{V\gamma c}{2mn} \cdot \frac{1}{V\gamma c(1-\log\gamma c)} , = \frac{1}{2mn(1-\log\gamma c)} . \tag{4.3.11}$$

For general SPRTs, the equivalent ratio, from (4.2.10), is

$$\frac{c(\gamma_1-\gamma_0)}{2mn} \cdot \frac{1}{k_1 - k_0 + k_0\log(-\frac{k_0}{L_1(1-w)}) - k_1\log(\frac{k_1}{L_0 w})}$$

$$= \frac{(\gamma_1-\gamma_0)}{2mn[\gamma_1 - \gamma_0 + \gamma_0\log(-\frac{\gamma_0 c}{L_1(1-w)}) - \gamma_1\log(\frac{\gamma_1 c}{L_0 w})]}$$

$$= \frac{1}{2mn[1 - \log c + \frac{\gamma_0}{\gamma_1-\gamma_0}\log(-\frac{\gamma_0}{L_1(1-w)}) - \frac{\gamma_1}{\gamma_1-\gamma_0}\log(\frac{\gamma_1}{L_0 w})]} . \tag{4.3.12}$$

Hence under circumstances in which the expansions above could be used, i.e., when

i) c is small

ii) mn is large,

the Chernoff average excess risk is small in proportion to the Chernoff a.m. risk, as evidenced by (4.3.11) and (4.3.12).

The situation considered in this chapter, when some estimate of the cost is available, is likely to be more frequent than one in which no estimate is available at all. However, the latter situation, to be considered in the next chapter, is one in which the Partial Sequential Probability Ratio Test procedure, whose properties were studied in Chapter III, is a suitable one to use.

CHAPTER V

ON MINIMIZING THE RISK WHEN THE COST IS UNKNOWN: A SPECIAL CASE.

5.1. Introduction.

When no estimate of the mean cost per observation in a sequential testing procedure is available, the analysis of Chapter IV does not apply.

A procedure which has some appeal in this situation is the PSPRT of Chapter III. The first n observations are used to estimate the cost, and we shall use the observed mean cost

$$\bar{c}_n = \frac{1}{n} \sum_{i=1}^{n} c_i, \qquad (5.1.1)$$

where c_i is the observed cost of the i-th observation. Using the notation of Chapter III, the conditional test $T(\underline{x}_n)$ is constructed as an SPRT on the basis of the observed $\underline{x}_n = (x_1, x_2, \ldots, x_n)$ and $\bar{c}_n$.

To minimize the conditional risk, given $\underline{x}_n$ and $\bar{c}_n$, the test boundaries A_n and B_n of $T(\underline{x}_n)$ will generally be functions of $w(\underline{x}_n)$ and of $\bar{c}_n$. In one special case, however, if Chernoff's a.m. risk is used, A_n and B_n are free of $w(\underline{x}_n)$ and depend on $\bar{c}_n$, on $\theta_1 - \theta_0$, and on the expected error rate. This is the case in which $L_0 = L_1 = L$, X is a normal r.v. with known variance, and $E(X) = \theta$.

This leads to a PSPRT procedure, in the sense that boundaries A and B are to be chosen which depend on the first n observations only through $\bar{c}_n$, but which are free of the posterior probabilities $w(\underline{x}_n)$ and $1-w(\underline{x}_n)$ of θ_0 and θ_1 after $\underline{x}_n$ has been observed. If $\bar{c}_n$ were the true cost, these boundaries would lead to an approximately minimum conditional risk for $T(\underline{x}_n)$, and since they would do so for all $\underline{x}_n$, the procedure would approximately minimize the unconditional risk for T_n.

In Section 5.2, the boundaries A and B are obtained, and the procedure is developed. In Section 5.3, the average excess is investigated when the cost c_i has the Gamma distribution of Chapter IV, and for the special case $w = \frac{1}{2}$. In Sections 5.4 and 5.5, the average excess is espressed in explicit but approximate forms using the methods of statistical differentials and of quadrature respectively. Section 5.5 gives numerical results and graphs which show

i) the existence of an optimum value of n minimizing the average excess.

ii) that the method of statistical differentials geves reasonably accurate results when compared with quadrature.

5.2. Extension of the Chernoff a.m. Risk to the PSPRT of a Normal Mean.

In this section, the boundaries A and B are explicitly obtained and the procedure is justified. Later, the lack of information on cost is considered, but in developing the procedure we shall denote by $\bar{c}_n$ the estimator of the true cost.

The posterior probabilities of θ_0 and θ_1 are $w(\underline{x}_n)$ and $1-w(\underline{x}_n)$ where

$$w(\underline{x}_n) = \frac{wp(\underline{x}_n|\theta_0)}{wp(\underline{x}_n|\theta_0) + (1-w)p(\underline{x}_n|\theta_1)}$$

$$= \frac{w}{w + (1-w)\exp\{n\Delta\bar{x}_n - \frac{1}{2}n(\theta_1^2-\theta_0^2)\}} \; . \tag{5.2.1}$$

<u>The a.m. Conditional Risk at Stage</u> n.

Let $r(\underline{x}_n)$ be the nominal a.m. risk, given that at least one more observation is taken, and that beyond stage n the procedure is that of a SPRT. In Theorem 2.2, it was shown that the nominal a.m. risk is

$$\pi^*L + \frac{\bar{c}_n}{E_1(Z)} \{(1-2\pi^*)\log\frac{1-\pi^*}{\pi^*} - (1-2w(\underline{x}_n))\log(\frac{1-w(\underline{x}_n)}{w(\underline{x}_n)})\} \; ,$$

where π^* is the root of (2.2.8). π^* is a function of $\underline{x}_n$, but Chernoff's approximation (2.2.12), viz.

$$\pi_1^* = \frac{\bar{c}_n}{E_1(Z)L} = \frac{2\bar{c}_n}{\frac{1}{2}\Delta^2 L}$$

gives a good approximation to the a.m. risk, and is free of $\underline{x}_n$. Thus π^* does not depend on $\underline{x}_n$ very much, and approximately

$$r(\underline{x}_n) = \pi^*L[1 + (1-2\pi^*)\log\frac{1-\pi^*}{\pi^*} - (1-2w(\underline{x}_n))\log\frac{1-w(\underline{x}_n)}{w(\underline{x}_n)}] \; . \tag{5.2.2}$$

If

$$\min[Lw(\underline{x}_n),\ L(1 - w(\underline{x}_n))] < r(\underline{x}_n) , \qquad (5.2.3)$$

we stop and decide at stage n.

In this case,

$$w(\underline{x}_n) < \tfrac{1}{2} \Rightarrow \text{accept } H_1:\ \theta = \theta_1$$

$$w(\underline{x}_n) > \tfrac{1}{2} \Rightarrow \text{accept } H_0:\ \theta = \theta_0 .$$

From Theorem 2.2, it can be seen that we accept H_1 at stage n if

$$w(\underline{x}_n) < \tfrac{1}{2} \quad \text{and} \quad \pi^* > w(\underline{x}_n) .$$

The second inequality gives

$$w < \pi^*[w + (1-w)\exp\{n\Delta\bar{x}_n - n(\theta_1^2-\theta_0^2)\}] .$$

So

$$\frac{w(1-\pi^*)}{(1-w)\pi^*} < \exp\{n\Delta\bar{x}_n - \tfrac{1}{2}n(\theta_1^2-\theta_0^2)\},$$

i.e., on taking logs and rearranging,

$$\bar{x}_n > \tfrac{1}{2}(\theta_1+\theta_0) + \frac{1}{n\Delta}\log\left(\frac{w}{1-w}\cdot\frac{1-\pi^*}{\pi^*}\right) . \qquad (5.2.4)$$

This is the condition for stopping at stage n and accepting H_1. If we assume that the expected error rate π^* is less than $\tfrac{1}{2}$, then the condition $w(\underline{x}_n) < \tfrac{1}{2}$ gives

$$\bar{x}_n > \tfrac{1}{2}(\theta_1+\theta_0) + \frac{1}{n\Delta}\log\frac{w}{1-w}$$

and follows from (5.2.4), since $\frac{1}{n\Delta} \log \frac{1-\pi^*}{\pi^*} > 0$ if $\pi^* < \frac{1}{2}$.

Similarly, from Theorem 2.2, we stop and accept H_0 at stage n if

$$w(\underline{x}_n) > \tfrac{1}{2} \quad \text{and} \quad \pi^* > 1 - w(\underline{x}_n) .$$

The second inequality leads, analogously to (4.1.4), to

$$\bar{x}_n < \tfrac{1}{2}(\theta_1+\theta_0) + \frac{1}{n\Delta} \log\left(\frac{w}{1-w} \cdot \frac{\pi^*}{1-\pi^*}\right) . \tag{5.2.5}$$

Then $\pi^* < \frac{1}{2} \Rightarrow w(\underline{x}_n) > \frac{1}{2}$.

Consider now the PSPRT procedure of Chapter III. (5.2.4) and (5.2.5) can be compared with Eqs. (3.3.3) of that chapter, and it is clear that a PSPRT procedure which approximately minimizes the unconditional risk (since it does so for every $\bar{x}_n$) is given by boundaries

$$\left.\begin{aligned} a &= \log\left(\frac{w}{1-w} \cdot \frac{1-\pi^*}{\pi^*}\right) \\ b &= \log\left(\frac{w}{1-w} \cdot \frac{\pi^*}{1-\pi^*}\right) \end{aligned}\right\} , \tag{5.2.6}$$

or

$$\left.\begin{aligned} A &= \frac{w}{1-w} \cdot \frac{1-\pi^*}{\pi^*} \\ B &= \frac{w}{1-w} \cdot \frac{\pi^*}{1-\pi^*} \end{aligned}\right\} . \tag{5.2.7}$$

Further, $0 < B < 1 < A$ if

$$\pi^* < \min(w, 1-w) \tag{5.2.8}$$

as for the SPRT in Chapter II.

Inequality (5.2.8) ensures that $\pi^* = \frac{2\bar{c}_n}{\Delta^2 L} < \frac{1}{2}$.

It is understood that A, B, a and b are functionally dependent upon $\bar{c}_n$, through π^*.

With these parameter values, the approximate a.m. risk is given by

$$R(w,n,\bar{c}_n|c) = L(w\alpha_n + (1-w)\beta_n) + c\{wE_n(N|\theta_0) + (1-w)E_n(N|\theta_1)\}, \quad (5.2.9)$$

where α_n and β_n are given by (3.3.12) of Chapter III and $E_n(N|\theta)$ by (3.4.1).

We note from (5.2.7) that if $w = \frac{1}{2}$, the best procedure is a symmetric one, with $B = 1/A$, $b = -a$.

<u>Numerical Results</u> = <u>Chernoff a.m. Risks for PSPRT Procedures</u>.

Tables 5.2 following show the approximate a.m. risks when $w = \frac{1}{2}$ and $\theta_1 - \theta_0 = \frac{1}{2}$ and $\frac{1}{4}$, for costs in the neighborhood of .01 and .001, and for unit loss L. The a.m. risks are calculated for procedures having n initial observations. If the cost c is known, then of course the optimum rule will be for $n = 1$, or not to take any observations at all. H_0 or H_1 is accepted in the latter case, according as $w >$ or $< \frac{1}{2}$.

The "no observations" rule would apply for the case $\Delta = .25$, $w = .25$, $c = .01$, since risk $> Lw$ for $n = 1$ (and $a = -.345$, $b = -1.852$, so that $0 < B < 1 < A$ does not hold).

For $\Delta = 0.5$, the a.m. risk more than doubles from $n = 1$ to $n = 100$ (and $n = 1$ to $n = 50$ for $c = .01$), while for $\Delta = 0.25$ the increase is less, particularly for smaller cost.

Tables 5.2. Chernoff a.m. Risks for PSPRT Procedures with Known Cost. Test of H_0: $\theta = \theta_0$ vs H_1: $\theta = \theta_1$, where θ is mean of $N(\theta, 1)$ distribution. $\Delta = \theta_1 - \theta_0$.

TABLE 5.2.1

$\Delta = 0.5$, $w = .5$, $L = 1$ $\qquad$ $w = .25$

n\c	.0075	.009	.01	.011	.0125	.01
1	.2053	.2296	.2441	.2576	.2758	.2229
5	.2056	.2303	.2452	.2590	.2781	.2233
10	.2105	.2377	.2547	.2708	.2937	.2368
15	.2219	.2535	.2736	.2931	.3214	.2597
25	.2599	.3029	.3309	.3586	.3995	.3222
50	.4026	.4794	.5304	.5813	.6574	.5274

TABLE 5.2.2

$\Delta = 0.5$, $w = .5$, $L = 1$ $\qquad$ $w = .25$

n\c	.00075	.001	.00125	.001
1	.03629	.04595	.05503	.04197
5	.03629	.04595	.05503	.04197
25	.03737	.04775	.05770	.04481
50	.04569	.05980	.07371	.05846
100	.07652	.10178	.12701	.10154

TABLE 5.2.3

$\Delta = 0.5$, $w = .5$, $L = 1$

n\c	.0001	.00001	.000001
1	.006495	.000835	.0001019
50	.007064	.000853	.0001023
100	.010443	.001093	.0001172
200	.020018	.002004	.0002010
500	.050000		

TABLE 5.2.4

Δ = 0.25, w = .5, L = 1 w = .25

n\c	.0075	.01	.0125	.01
1	.3839	.4068	.4319	.2726
5	.3847	.4106	.4373	.3288
10	.3909	.4256	.4608	.3693
15	.4024	.4476	.4934	.4047
25	.4359	.5040	.5725	.4750
50	.5531	.6814	.8099	.6658

TABLE 5.2.5

Δ = 0.25, w = .5, L = 1 w = .25

n\c	.00075	.001	.00125	.001
1	.10866	.13412	.15695	.12136
5	.10866	.13412	.15695	.12136
25	.10870	.13423	.15720	.12160
50	.10990	.13656	.16104	.12564
100	.12069	.15359	.18513	.14667

TABLE 5.2.6

Δ = 0.25, w = .5, L = 1

n\c	.0001	.00001	.000001
1	.02146	.002893	.0003632
100	.02180	.002897	.0003632
200	.02544	.003043	.0003675
300		.003515	.0003937

Tables 5.2.3 and 5.2.6 show how the Chernoff a.m. risk for the SPRT procedure behaves when the cost becomes small. As the cost $\to 0$, the value of n, at which the increase of the risk ceases to be slow, becomes larger.

It can be noted that the steepest increase in the risk for larger n results from the growing domination of the term cn, due to the cost of the first n observations. For very large n, the other terms appear, in fact, to become negligible, as evidenced in Table 5.2.3 for $n = 500$.

When Δ is smaller, the risks are larger, but the proportional increases in risk are smaller over the same change in n.

5.3. Average Excess.

For the special case in which $w = \frac{1}{2}$ and loss $L = 1$, we next investigate the effect of lack of information on cost. In the previous section, the apparent a.m. risk is based upon a supposed cost $\bar{c}_n$, although $\bar{c}_n$ is an estimator. If the true cost is c, boundaries for the same test procedure would be determined by

$$a = \log\left(\frac{w}{1-w} \cdot \frac{1-\pi}{\pi}\right), \quad \text{with} \quad w = \tfrac{1}{2}; \quad b = -a$$

and

$$\pi = \frac{2c}{\Delta^2}, \quad \Delta = \theta_1 - \theta_0 .$$

Using the notation of (5.2.9), we are interested in

$$ER(\tfrac{1}{2}, n, \bar{c}_n \mid c)$$

over the distribution of $\bar{c}_n$. As in Section 4.3, suppose that

$c_i \sim \Gamma(m, \frac{m}{c})$, so that

$$p(c_i) = \frac{1}{\Gamma(m)} (\frac{m}{c})^m x^{m-1} e^{-\frac{m}{c}x}$$

$$E(c_i) = c, \quad Var(c_i) = \frac{c^2}{m} .$$

Then

$$\bar{c}_n \sim (mn, \frac{mn}{c}), \quad E\bar{c}_n = c, \ Var(\bar{c}_n) = \frac{c^2}{mn} .$$

Evaluating (5.2.9), we get from (3.3.12)

$$\alpha_n + \beta_n = 1 - \frac{B(A-1)}{A-B} \Phi(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}\Delta) - \frac{A(1-B)}{A-B} \Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}\Delta)$$

$$+ \frac{1-B}{A-B} \Phi(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta) + \frac{A-1}{A-B} \Phi(\frac{b}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta) .$$

Now put

$$g_+ = \frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}\Delta, \quad \text{and} \quad g_- = \frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta . \qquad (5.3.1)$$

If we also write

$$\left. \begin{aligned} \Phi_+ &= \Phi(\frac{a}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}\Delta) = \Phi(g_+) \\ \Phi_- &= \Phi(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta) = \Phi(g_-) \end{aligned} \right\} \qquad (5.3.2)$$

and since

$$\left. \begin{aligned} b = -a \implies \Phi(\frac{b}{\sqrt{n}\Delta} + \tfrac{1}{2}\sqrt{n}\Delta) &= 1 - \Phi_- \\ \Phi(\frac{b}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta) &= 1 - \Phi_+ , \end{aligned} \right\} \qquad (5.3.3)$$

and since, further,

$$A - B = \frac{1-2\pi^*}{\pi^*(1-\pi^*)}\ , \quad A-1 = \frac{1-2\pi^*}{\pi^*}\ , \quad 1-B = \frac{1-2\pi^*}{1-\pi^*}\ , \quad \text{and,} \quad \pi^* = \frac{2\bar{c}_n}{\Delta^2}\ ,$$

we have

$$\begin{aligned} \alpha_n + \beta_n &= 1 - \pi^*(1-\Phi_-) - (1-\pi^*)\Phi_+ + \pi^*\Phi_- + (1-\pi^*)(1-\Phi_+) \\ &= 2[\pi^*\Phi_- + (1-\pi^*)(1-\Phi_+)]\ . \end{aligned} \tag{5.3.4}$$

Next, we require $E_n(N|\theta_0)$ and $E_n(N|\theta_1)$. From (3.4.1), by putting $h = 1$ and -1 we get

$$\begin{aligned} E_n(N|\theta_0) + E_n(N|\theta_1) &= n + \frac{2}{\Delta^2}\ [\{-\tfrac{1}{2}n\Delta^2 + \frac{bA-aB}{B-A}\}\{\Phi_+ + \Phi_- - 1\} \\ &\quad + \frac{a-b}{B-A}\ \{\Phi_- + \Phi_+ - 1\} \\ &\quad + \sqrt{\frac{n}{2\pi}}\ \Delta\ \{e^{-\frac{1}{2}(\frac{b}{\sqrt{n}\Delta} + \frac{1}{2}\sqrt{n}\Delta)^2} - e^{-\frac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \frac{1}{2}\sqrt{n}\Delta)^2}\}] \\ &\quad + n - \frac{2}{\Delta^2}\ [\{\tfrac{1}{2}n\Delta^2 + \frac{bB-aA}{A-B}\}\{\Phi_- + \Phi_+ - 1\} \\ &\quad + \frac{a-b}{A-B}\ \{\Phi_+ + \Phi_- - 1\} \\ &\quad + \sqrt{\frac{n}{2\pi}}\ \Delta\ \{e^{-\frac{1}{2}(\frac{b}{\sqrt{n}\Delta} - \frac{1}{2}\sqrt{n}\Delta)^2} - e^{-\frac{1}{2}(\frac{a}{\sqrt{n}\Delta} - \frac{1}{2}\sqrt{n}\Delta)^2}\}]\ . \end{aligned} \tag{5.3.5}$$

Now

$$\frac{bA-aB}{B-A} = \frac{a(A^2+1)}{A^2-1}\ , \quad \frac{bB-aA}{A-B} = -\frac{a(A^2+1)}{A^2-1}\ , \quad \frac{a-b}{A-B} = \frac{2aA}{A^2-1}\ .$$

Hence

$$E_n(N|\theta_0) + E_n(N|\theta_1) = -2n(\Phi_+(\bar{c}_n) + \Phi_-(\bar{c}_n) - 2)$$

$$+ \frac{4a(A-1)}{\Delta^2(A+1)} (\Phi_+(\bar{c}_n) + \Phi_-(\bar{c}_n) - 1)$$

$$+ \frac{4}{\Delta}\sqrt{\frac{n}{2\pi}} \{e^{-\frac{1}{2}g_-^2} - e^{-\frac{1}{2}g_+^2}\}. \qquad (5.3.6)$$

Hence (5.2.9) gives

$$R(\tfrac{1}{2},n,\bar{c}_n|c) = \tfrac{1}{2}(\alpha_n+\beta_n) + c[E(N|\theta_0) + E(N|\theta_1)]$$

$$= \pi^*\Phi_- + (1-\pi^*)(1-\Phi_+) + nc$$

$$+ c[(n - \frac{2a}{\Delta^2}(1-2\pi^*)(1-\Phi_+-\Phi_-) + \frac{2\sqrt{n}}{\Delta}\{\phi(g_-) - \phi(g_+)\}] , \qquad (5.3.7)$$

where $\phi(t) = \Phi'(t)$ and $\pi^* = \frac{2\bar{c}_n}{\Delta^2}$.

If, for given n, the apparent risk is (5.3.7), then the excess over the (approx.) a.m. risk when c is known, is

$$\text{Excess} = R(\tfrac{1}{2}, n, \bar{c}_n|c) - R(\tfrac{1}{2}, n, c|c). \qquad (5.3.8)$$

Over the distribution of $\bar{c}_n$, the average excess is

$$ER(\tfrac{1}{2}, n, \bar{c}_n|c) - R(\tfrac{1}{2}, n, c|c).$$

The exact evaluation of this quantity presents analytical difficulties, but when c is small, and m is large, the coefficient of variation of $\bar{c}_n$, i.e., $\frac{1}{\sqrt{mn}}$, is small, and we can apply the statistical differential method to approximate the average excess. This method has the advantage that if we only take the expansion to the variance term, then the distributional form of $\bar{c}_n$ doesn't matter beyond knowing the first two moments. Thus

$$\mathcal{E}R(\tfrac{1}{2},n,\bar{c}_n|c) \simeq R(\tfrac{1}{2},n,c|c) + \frac{1}{2!}\,\mathrm{Var}(\pi^*)\,\frac{\partial^2}{\partial\pi^{*2}}\,[R(\tfrac{1}{2},n,\bar{c}_n|c)]\Big|_{\bar{c}_n=c} \,. \tag{5.3.9}$$

Evaluating the required terms separately,

$$\left.\begin{aligned}
\mathrm{Var}(\pi^*) &= \frac{4}{\Delta^4}\cdot\frac{c^2}{mn} \\
g'_+ &= \frac{\partial}{\partial\pi^*}\left[\frac{1}{\sqrt{n}\Delta}\log\frac{1-\pi^*}{\pi^*} + \tfrac{1}{2}\sqrt{n}\Delta\right] \\
&= -\frac{1}{\sqrt{n}\Delta}\cdot\frac{1}{\pi^*(1-\pi^*)} \\
&= g'_-(\pi^*) \\
g''_+(\pi^*) &= \frac{\partial}{\partial\pi^*}\,[g'_+(\pi^*)] \\
&= \frac{1}{\sqrt{n}\Delta}\cdot\frac{(1-2\pi^*)}{\pi^{*2}(1-\pi^*)^2} \\
&= g''_-(\pi^*)
\end{aligned}\right\} \tag{5.3.10}$$

$$\frac{d}{d\pi^*}\,[\Phi_+(\bar{c}_n)] = g'_+(\pi^*)\;\phi(g_+(\pi^*))$$

$$\frac{d^2}{d\pi^{*2}}\,[\Phi_+(\bar{c}_n)] = \{g''_+(\pi^*) - g_+(\pi^*)\;g'^{\,2}_+(\pi^*)\}\;\phi(g_+(\pi^*)) \,.$$

Abbreviating the notation, write

$$g_+ = g_+(\pi^*), \quad g_+' = g_+'(\pi^*), \quad g_+'' = g_+''(\pi^*),$$

$$\phi_+ = \phi(g_+(\pi^*)), \quad \text{and} \quad \Phi_+ = \Phi(g_+(\pi^*)) \quad \text{as before.}$$

g_-, ϕ_-, etc., are similarly defined.

Then

$$\frac{d^2}{d\pi^{*2}}[\Phi_+] = (g_+'' - g_+ g_+'^2)\phi_+$$

$$\frac{d}{d\pi^*}(\phi_+) = -g_+ g_+' \phi_+$$

$$\frac{d^2}{d\pi^{*2}}(\phi_+) = -(g_+'^2 + g_+ g_+'' - g_+^2 g_+'^2)\phi_+$$

$$\frac{d}{d\pi^*}(\Phi_+ \cdot \pi^*) = \Phi_+ + \pi^* g_+' \phi_+$$

$$\frac{d^2}{d\pi^{*2}}(\pi^*\Phi_+) = 2g_+'\phi_+ + \pi^*(g_+'' - g_+ g_+'^2)\phi_+$$

$$\frac{d}{d\pi^*}\{(1-2\pi^*)\log\frac{1-\pi^*}{\pi^*}\} = -\frac{(1-2\pi^*)}{\pi^*(1-\pi^*)} - 2\log\left(\frac{1-\pi^*}{\pi^*}\right)$$

$$\frac{d^2}{d\pi^{*2}}\{(1-2\pi^*)\log\frac{1-\pi^*}{\pi^*}\} = \frac{1-2\pi^*}{\pi^{*2}(1-\pi^*)^2} - \frac{2}{(1-\pi^*)^2} + \frac{2}{\pi^*(1-\pi^*)}$$

$$= \frac{1-4\pi^{*2}}{\pi^{*2}(1-\pi^*)^2} .$$

Also, for an arbitrary twice-differentiable function $f(\pi^*)$,

$$\frac{d}{d\pi^*}\{f(\pi^*)\Phi_+\} = f'(\pi^*)\Phi_+ + f(\pi^*)g_+'\phi_+, \quad \text{where} \quad \Phi_+ = \Phi(g_+(\pi^*))$$

$$\frac{d^2}{d\pi^{*2}}\{f(\pi^*)\Phi_+\} = f''(\pi^*)\Phi_+ + [f'(\pi^*)g_+' + f(\pi^*)\{g_+'' - g_+ g_+'^2\}]\phi_+ .$$

Hence

$$\frac{d^2}{d\pi^{*2}} \{(1-2\pi^*) \log (\frac{1-\pi^*}{\pi^*})\Phi_+\} = \frac{1-4\pi^{*2}}{\pi^{*2}(1-\pi^*)^2} \Phi_+ + [-\{\frac{1-2\pi^*}{\pi^*(1-\pi^*)}$$

$$+ 2\log \frac{1-\pi^*}{\pi^*}\} g'_+$$

$$+ (1-2\pi^*) \log (\frac{1-\pi^*}{\pi^*})(g''_+ - g_+ g'^2_+)]\phi_+ .$$

Finally, in order to evaluate (5.3.9), we substitute the above quantities, and evaluate them at $\pi^* = \pi = \frac{2c}{\Delta^2}$, i.e., at $\bar{c}_n = E\bar{c}_n = c$. This gives

$$ER(\tfrac{1}{2},n,\bar{c}_n|c) - R(\tfrac{1}{2},n,c|c) \simeq \frac{2c^2}{\Delta^4 mn} [2g'_-\phi_- + \pi(g''_- - g_- g'^2_-)\phi_- + 2g'_+\phi_+$$

$$+ \pi(g''_+ - g_+ g'^2_+)\phi_+ + \frac{2c}{\Delta^2} \cdot \frac{1-4\pi^2}{\pi^2(1-\pi)^2} (\Phi_+ + \Phi_- - 1) - \frac{2c}{\Delta^2} \{\frac{1-2\pi}{\pi(1-\pi)}$$

$$+ 2\log \frac{1-\pi}{\pi}\}(g'_+\phi_+ + g'_-\phi_-) + \frac{2c}{\Delta^2} (1-2\pi) \log \frac{1-\pi}{\pi} \cdot \{(g''_+ - g_+ g'^2_+)\phi_+$$

$$+ (g''_- - g_- g'^2_-)\phi_-\} - nc\{(g''_+ - g_+ g'^2_+)\phi_+ + (g''_- - g_- g'^2_-)\phi_-\}$$

$$+ \frac{2\sqrt{n}c}{\Delta} \{(g'^2_+ + g_+ g''_+ - g^2_+ g'^2_+)\phi_+ - (g'^2_- + g_- g''_- - g^2_- g'^2_-)\phi_-\}]$$

$$= \frac{2c^2}{\Delta^4 mn} [\frac{1-4\pi^2}{\pi(1-\pi)^2} (\Phi_+ + \Phi_- - 1) + (g'_+\phi_+ + g'_-\phi_-) \cdot$$

$$\cdot (\frac{1}{1-\pi} - 2\pi \log \frac{1-\pi}{\pi}) + \{(g''_+ - g_+ g'^2_+)\phi_+ + (g''_- - g_- g'^2_-)\phi_-\} \cdot$$

$$\cdot \{\pi - nc + \pi(1-2\pi) \log \frac{1-\pi}{\pi}\} + \frac{2\sqrt{n}c}{\Delta} \{(g'^2_+ + g_+ g''_+ - g^2_+ g'^2_+)\phi_+$$

$$- (g'^2_- + g_- g''_- - g^2_- g'^2_-)\phi_-\}]$$

evaluated at $\pi^* = \pi = \frac{2c}{\Delta^2}$. (5.3.11)

For the sake of a better comparison, it might be desirable to measure the average excess over the a.m. risk in the class $\mathcal{C}$ of SPRT procedures, including those in which no observations are made. The boundaries of such an SPRT, leading through Chernoff's approximations to a risk very close to the a.m. risk, are the same as above, with error probability $\alpha = \pi = \frac{2c}{\Delta^2}$ under both hypotheses.

Using the notation of Chapter 2, when at least one observation is taken,

$$\begin{aligned} R(\tfrac{1}{2},n,c|c) - R_\delta(\lambda_w|c) &= \pi\Phi_- + (1-\pi)(1-\Phi_+) + nc \\ &\quad + \{nc - \pi\log(\tfrac{1-\pi}{\pi})(1-2\pi)\}(1 - \Phi_+ - \Phi_-) \\ &\quad + \frac{2\sqrt{nc}}{\Delta}(\phi_- - \phi_+) - \pi[1 + (1-2\pi)\log\frac{1-\pi}{\pi}] \\ &= [nc - \pi\{1 + (1-2\pi)\log\frac{1-\pi}{\pi}\}](2 - \Phi_+ - \Phi_-) \\ &\quad + 1 - \Phi_+ + \sqrt{n}\Delta\pi(\phi_- - \phi_+). \end{aligned} \tag{5.3.12}$$

So $\mathcal{E}R(\tfrac{1}{2}, n, \bar{c}_n|c) - R_\delta(\lambda_w|c)$ is approximately given by summing (5.3.11) and (5.3.12).

5.4. Method of Quadrature.

A more accurate measurement of the average excess is given by evaluating

$$\int R(\tfrac{1}{2}, n, \bar{c}_n|c)\,dF(\bar{c}_n) \tag{5.4.1}$$

in the form

$$\frac{h}{\Gamma(mn)}\left(\frac{mn}{c}\right)^{mn} \Sigma\, R(\tfrac{1}{2}, n, \bar{c}_n|c)\,\bar{c}_n^{\,mn-1}\exp(-\frac{mn}{c}\bar{c}_n) \tag{5.4.2}$$

summed at intervals of length h and truncated at suitable points to be set out presently.

The s.d. of the distribution of $\bar{c}_n$ is $\frac{c}{\sqrt{mn}}$, and a suitable degree of accuracy to the above integral seems to be given if

$$h = \frac{c}{10\sqrt{mn}} , \qquad (5.4.3)$$

summed from $\bar{c}_n = \max(0, c - \frac{6c}{\sqrt{mn}})$ to $c + \frac{6c}{\sqrt{mn}}$.

The choice of h, as one-tenth of the standard deviation of $\bar{c}_n$, is chosen because, for bell-shaped distributions whose density functions are not too far removed from normality, it can give a very close approximation to the area under the curve, i.e., for this value of h

$$\frac{h}{\Gamma(mn)} \left(\frac{mn}{c}\right)^{mn} \Sigma\, \bar{c}_n^{\,mn-1} \exp\left(- \frac{mn}{c} \bar{c}_n\right)$$

is very close to 1 when mn is not small. Since $R(\frac{1}{2}, n, \bar{c}_n | c)$ is generally smaller than 1, and decreases overall as c decreases, it would seem reasonable that (5.4.2) and (5.4.3) give a close fit to (5.4.1). This is a heuristic argument, but the probability integral was computed by the method, in a few cases, and found to be extremely close to 1.

We use the refined form of Stirling's approximation [8]

$$(mn)! \sim (2\pi)^{\frac{1}{2}} (mn)^{mn+\frac{1}{2}} \exp\left(- mn + \frac{1}{12mn} - \frac{1}{360m^3n^3}\right) . \qquad (5.4.4)$$

The supplementary factors in (5.4.4) are required in order that the average excess computed by quadrature gives a sufficiently accurate result. It turns out that without these factors, the approximation to average excess using statistical differentials gives a more accurate result in many cases; but using (5.4.4), the statistical differential approach still gives results agreeing with those obtained by quadrature to within 3.5%.

(5.4.2), (5.4.3) and (5.4.4) now give (replacing $\frac{1}{\Gamma(mn)}$ by $\frac{mn}{(mn)!}$)

$$\int R(\tfrac{1}{2},n,\bar{c}_n|c)\,dF(\bar{c}_n) \approx \frac{(mn)^{mn}}{\sqrt{2\pi}(mn)^{mn+}\,e^{-mn}} \cdot \frac{hmn}{c} \cdot \Sigma R(\bar{c}_n)\left(\frac{\bar{c}_n}{c}\right)^{mn-1} e^{-mn} \cdot$$

$$\cdot\ e^{-\frac{mn}{c}(\bar{c}_n-c)} \cdot \exp\left[-\frac{1}{12mn} + \frac{1}{360m^3n^3}\right.$$

$$= \frac{1}{10\sqrt{2\pi}} \exp\left(-\frac{1}{12mn} + \frac{1}{360m^3n^3}\right) \Sigma\ R(\bar{c}_n)\left(\frac{\bar{c}_n}{c}\right)^{mn-1} e^{-\frac{mn}{c}(\bar{c}_n-c)} \tag{5.4.5}$$

The summation is as described above.

Although the analytical presentation here of the quadrature method is considerably briefer and more concise than that of 5.3 by statistical differentials, the computation of (5.4.5) involves a substantially greater amount of time on the computer than do (5.3.11) and (5.3.12), and this is a disadvantage. The only reason, therefore, to use quadrature in preference to statistical differentials would be a gain in accuracy. Numerical comparisons, to be presented in 5.5, showed, however, that for smaller values of c, the method of 5.3 is adequate, both in evaluating the average excess based on Chernoff's a.m. risk, and in evaluating the optimum value of n in PSPRT procedures leading to an overall minimum average excess.

5.5. Numerical Results.

Tables 5.5.1 to 5.5.4 show the approximate average excess as evaluated using the method of statistical differentials for cases in which Δ and c are fixed. It will be seen from these, and from the diagrams in Figs. 5.5.1 to 5.5.4 that although m varies, the optimum value of n, minimizing the average excess, does not change very much under these conditions.

The value of the Chernoff a.m. risk for a Wald SPRT, given that the corresponding cost c is known, is stated with each table, and it should be noted that throughout the entire range of n displayed in each table, the ratio (average excess risk)/(Chernoff a.m. risk) is small. Examples might arise where this would not necessarily be so -- for instance, if the variance of c_i were larger.

The curves in each figure lie inside one another as m decreases, and the (approximate) minimum average excess is greater when m is smaller. This last feature is what one would expect, since $\mathrm{Var}(c_i) = c^2/m$ is greater when m is smaller.

Tables 5.5. Approximate Average Excess of PSPRT procedures for tests of normal mean; $\sigma^2 = 1$.

TABLE 5.5.1

$\Delta = \theta_1 - \theta_0 = .25$, $c = .001$, Chernoff a.m. risk = .13412

n\m	3	10	17	25
1	.005668	.001701	.0010003	.0006802
3	.001889	.000567	.0003334	.0002267
5	.001134	.000340	.0002001	.0001360
8	.000708	.000213	.0001250	.0000850
10	.000566	.000170	.0001000	.0000680
12	.000470	.000141	.0000834	.0000569
13	.000433	.000131	.0000774	.0000530
14	.000402	.000122	.0000726	.0000500
15	.000374	.000115	.0000691	.0000482
16	.000351	.000110	.0000671	.0000476
17	.000332	.000106	.0000666	.0000484
18	.000316	.000105	.0000678	.0000507
19	.000304	.000106	.0000710	.0000550
20	.000296	.000109	.0000764	.0000614
22	.000290	.000124	.0000951	.0000818
25	.000312	.000172	.0001470	.0001357
30	.000443	.000334	.0003152	.0003064

TABLE 5.5.2

$\Delta = .50$, $c = .001$, Chernoff a.m. risk = .04595

n\m	3	10	17	25
1	.001355	.0004064	.0002390	.0001625
3	.000452	.0001355	.0000797	.0000542
5	.000271	.0000812	.0000478	.0000325
6	.000225	.0000677	.0000400	.0000273
7	.000192	.0000587	.0000351	.0000243
8	.000169	.0000536	.0000332	.0000239
10	.000148	.0000585	.0000428	.0000356
11	.000150	.0000712	.0000572	.0000508
12	.000164	.0000930	.0000805	.0000748
15	.000282	.0002300	.0002209	.0002167
20	.000812	.0007787	.0007729	.0007702

TABLE 5.5.3

Δ = .25, c = .00001, Chernoff a.m. risk = .002893

n\m	3	10	17	25
35	.000001524	.000000457	.000000268	.000000183
40	.000001334	.000000401	.000000235	.000000160
45	.000001187	.000000356	.000000211	.000000144
48	.000001113	.000000337	.000000199	.000000136
50	.000001071	.000000324	.000000193	.000000133
52	.000001032	.000000316	.000000189	.000000131
53	.000001015	.000000312	.000000188	.000000131
54	.000000999	.000000309	.000000187	.000000131
55	.000000983	.000000306	.000000187	.000000132
56	.000000969	.000000304	.000000186	.000000133
57	.000000957	.000000303	.000000189	.000000136
58	.000000946	.000000304	.000000190	.000000138
60	.000000927	.000000307	.000000198	.000000147
65	.000000909	.000000338	.000000238	.000000192
70	.000000945	.000000418	.000000325	.000000282
75	.000001059	.000000570	.000000483	.000000444

TABLE 5.5.4

Δ = .50, c = .00001, Chernoff a.m. risk = .0008346

n m	3	10	17	25
1	.000013335	.000003999	.000002353	.000001600
3	.000004444	.000001332	.000000784	.000000533
5	.000002665	.000000799	.000000469	.000000319
8	.000001666	.000000500	.000000293	.000000200
10	.000001333	.000000400	.000000236	.000000159
12	.000001111	.000000334	.000000196	.000000133
15	.000000888	.000000266	.000000156	.000000106
17	.000000786	.000000238	.000000141	.000000097
18	.000000744	.000000227	.000000136	.000000094
19	.000000709	.000000220	.000000133	.000000094
20	.000000681	.000000217	.000000135	.000000097
21	.000000661	.000000220	.000000141	.000000106
23	.000000648	.000000247	.000000176	.000000144
25	.000000686	.000000320	.000000256	.000000226
28	.000000888	.000000567	.000000510	.000000484
30	.000001163	.000000867	.000000815	.000000792
33	.000001875	.000001613	.000001566	.000001546
37	.000003579	.000003356	.000003317	.000003299

From computed results, the method of statistical differentials is undesirable for larger values of c, yielding as it does for $c = .01$, $\Delta = .25$, a negative excess. (However, the a.m. risk for an SPRT procedure with $c = .01$ and $\Delta = .25$ comes close to that obtained if a decision is made with no observations; see Table 2.2.2.) A comparison with the results obtained by quadrature follows in 5.6, but here it can be noted that for cases such as $c = .01$, $\Delta = .25$, quadrature does give valid results.

If c and m are fixed, the optimum value of n appears to increase as Δ decreases (see Figs. 5.5.5 and 5.5.6), as does the approximate average excess. This is what intuition would suggest.

It is of interest to give an example illustrating how the average excess splits into two non-negative components. If we denote the approximate average excess by E, then

$$E = E_1 + E_2,$$

where

E_1 = average risk of PSPRT using estimator $\bar{c}_n$ for cost
- (risk of PSPRT when true cost c is known)

E_2 = Risk of PSPRT when true cost c is known
- (risk of Wald SPRT when c is known).

The risks in E_1 and E_2 are Chernoff a.m. risks for given n. As one would expect, E_1 dominates E for smaller values of n, but as n increases and $\bar{c}_n \to c$ in probability, E_2 becomes dominant.

TABLE 5.5.5

Components of approximate average excess.

m = 25, $\theta_1 - \theta_0$ = .25, c = .001, Chernoff a.m. risk = .13412

n	E_1	E_2	E = approx. Av. Excess
1	.00068022	0	.00068022
10	.00006792	.00000008	.00006800
15	.00004449	.00000372	.00004821
18	.00003622	.00001453	.00005075
25	.00002403	.00011163	.00013566

The next comparison shows the effect of changes in cost when m and Δ are fixed. Table 5.5.6 and Fig. 5.5.7 are compiled to show the correspondence between the optimum value of n and the cost. On a log-scale for cost, the graph indicates that the increase in optimum n is linear with log (decrease in cost). For each combination of m and Δ, the ratio

$$\frac{\text{Approx. Average Excess}}{\text{Cost}}$$

appears to remain in the same order of magnitude as the cost varies.

Figs. 5.5. Graphs of Average Excess Risk E against n for tests of normal means by PSPRT procedures.

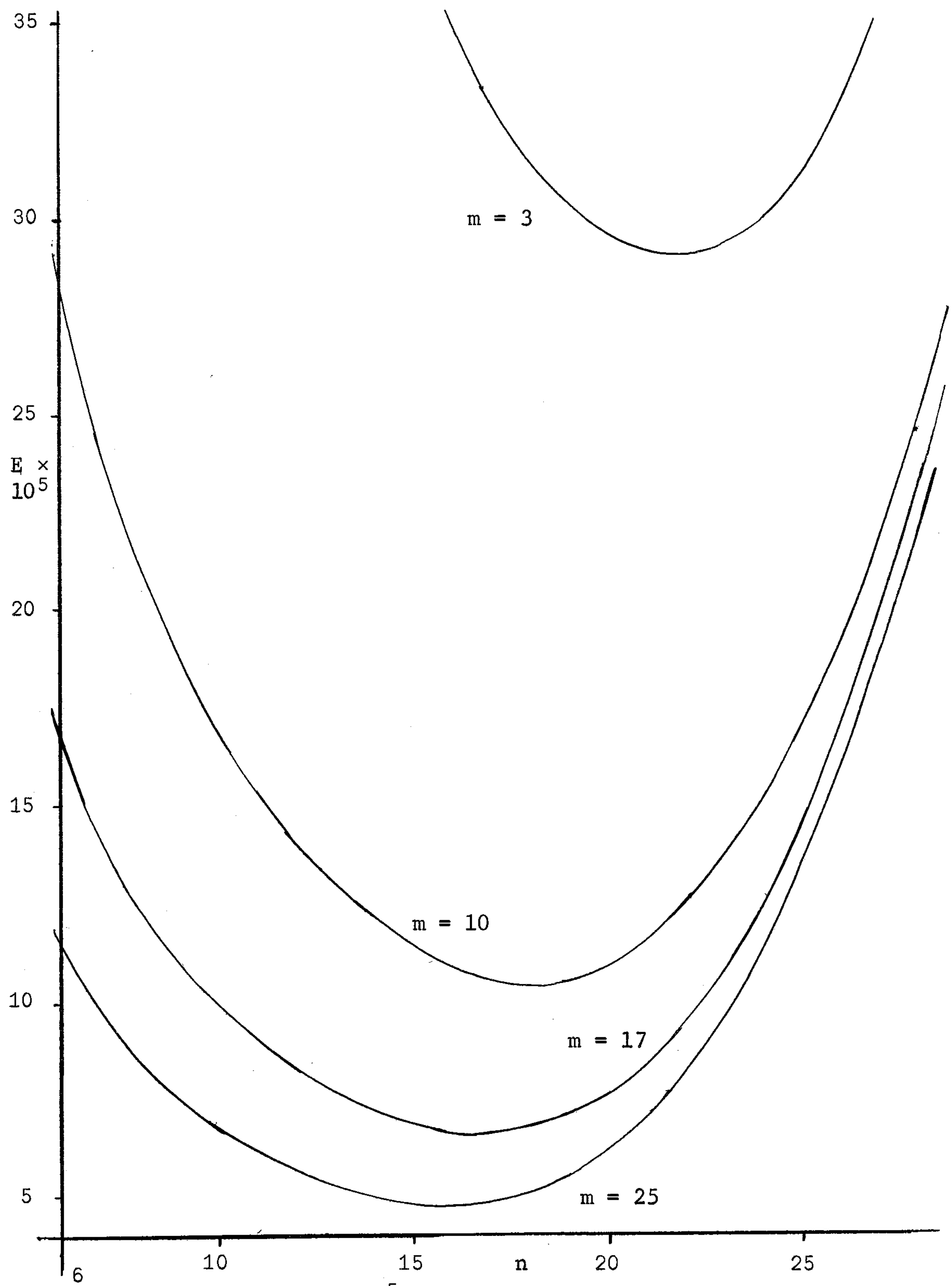

Fig. 5.5.1. $E \times 10^5$ against n. c = .001, Δ = .25.

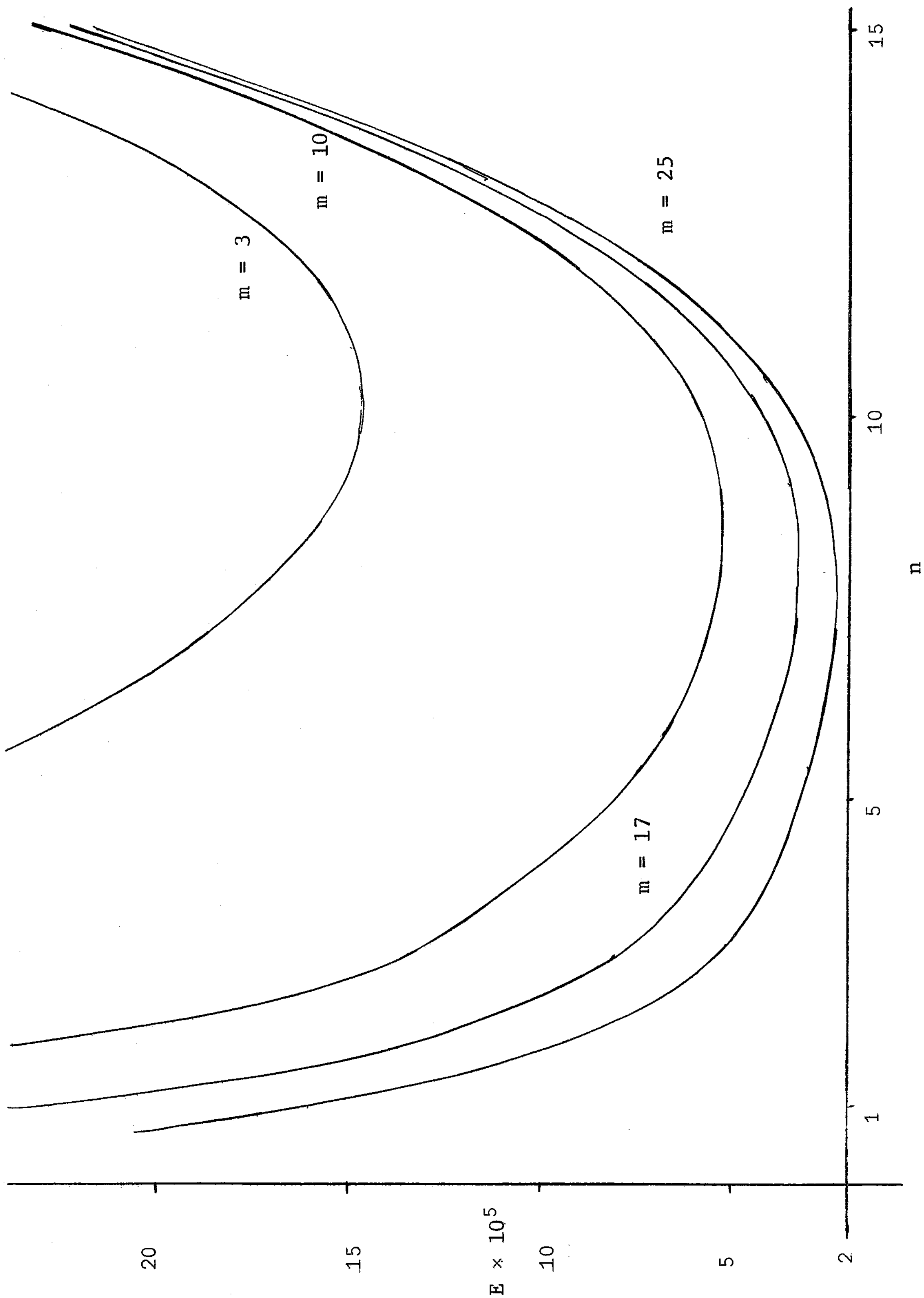

Fig. 5.5.2. $E \times 10^5$ against n. $c = .001$, $\Delta = .50$.

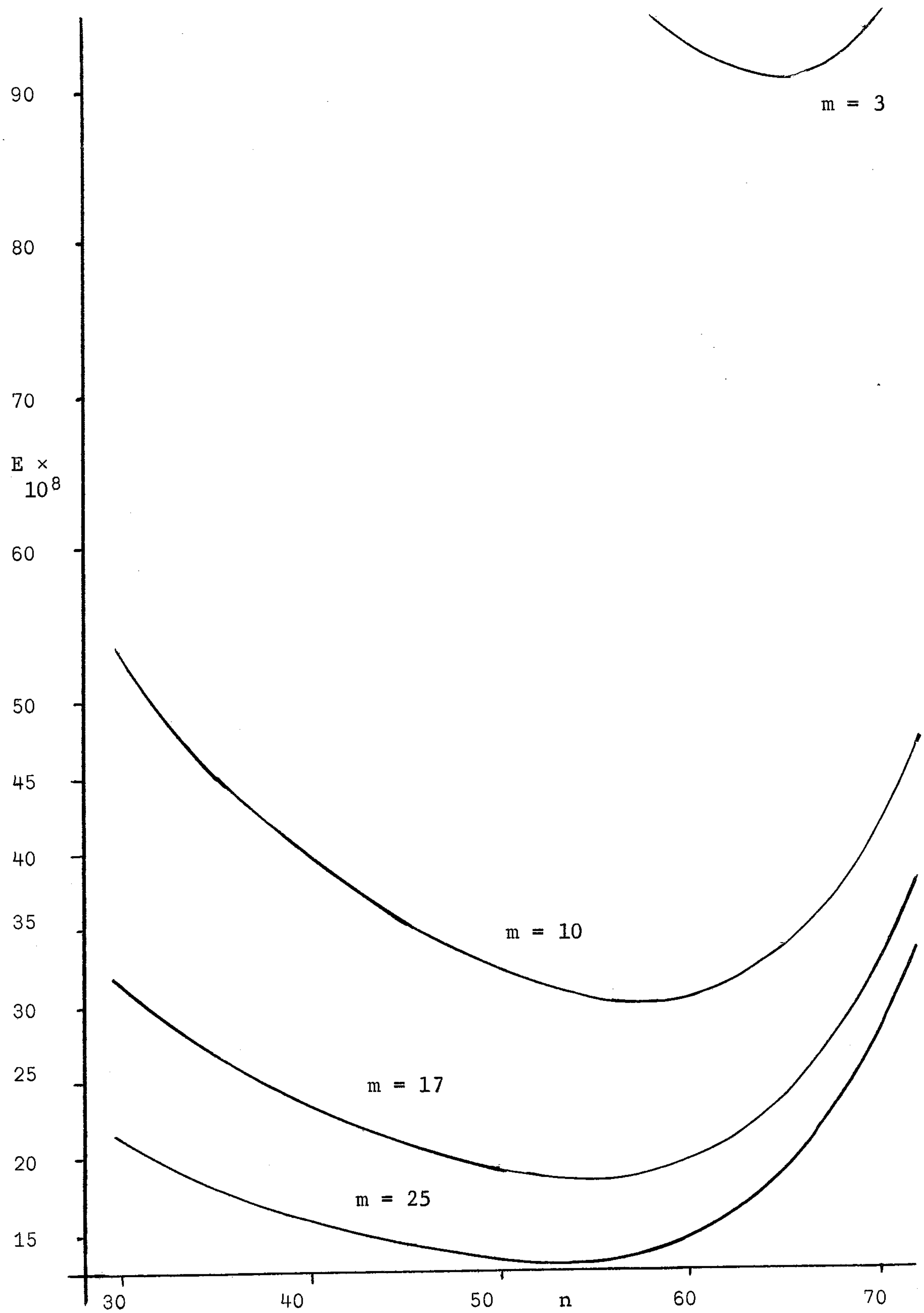

Fig. 5.5.3. $E \times 10^8$ against n. c = .00001, Δ = .25.

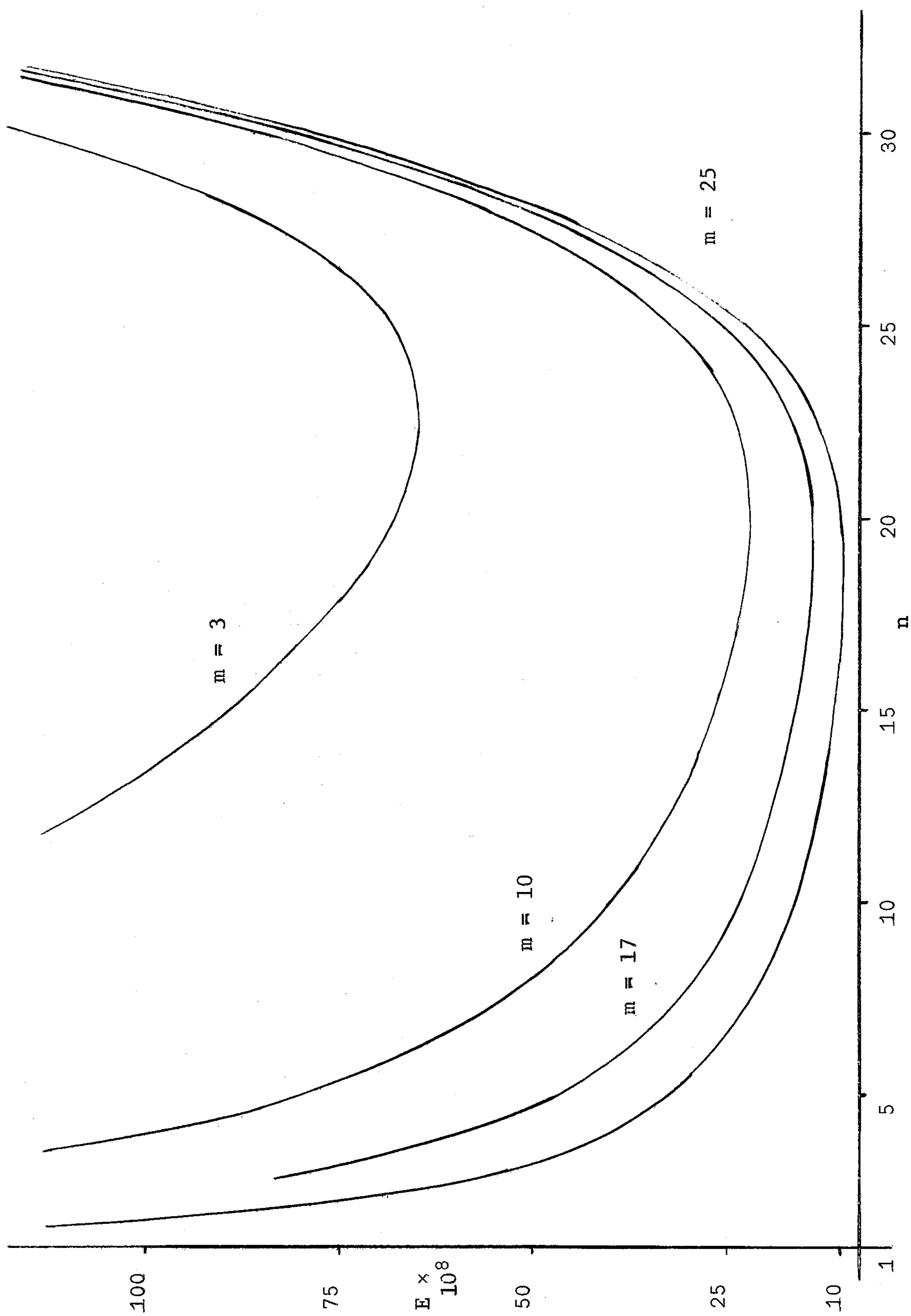

Fig. 5.5.4. $E \times 10^8$ against n. c = .00001, Δ = .50.

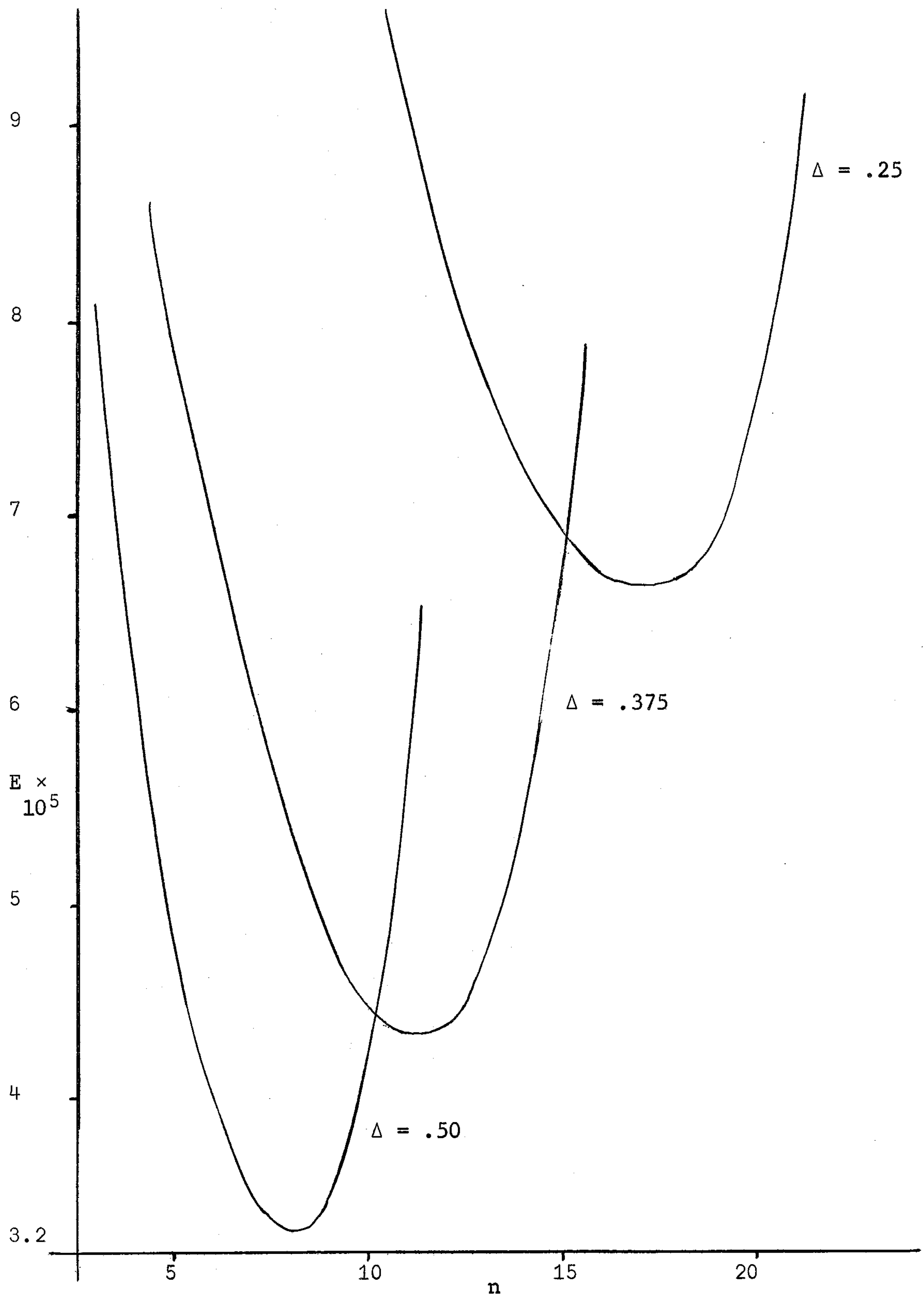

Fig. 5.5.5. E × 10^5 against n. c = .001, m = 17.

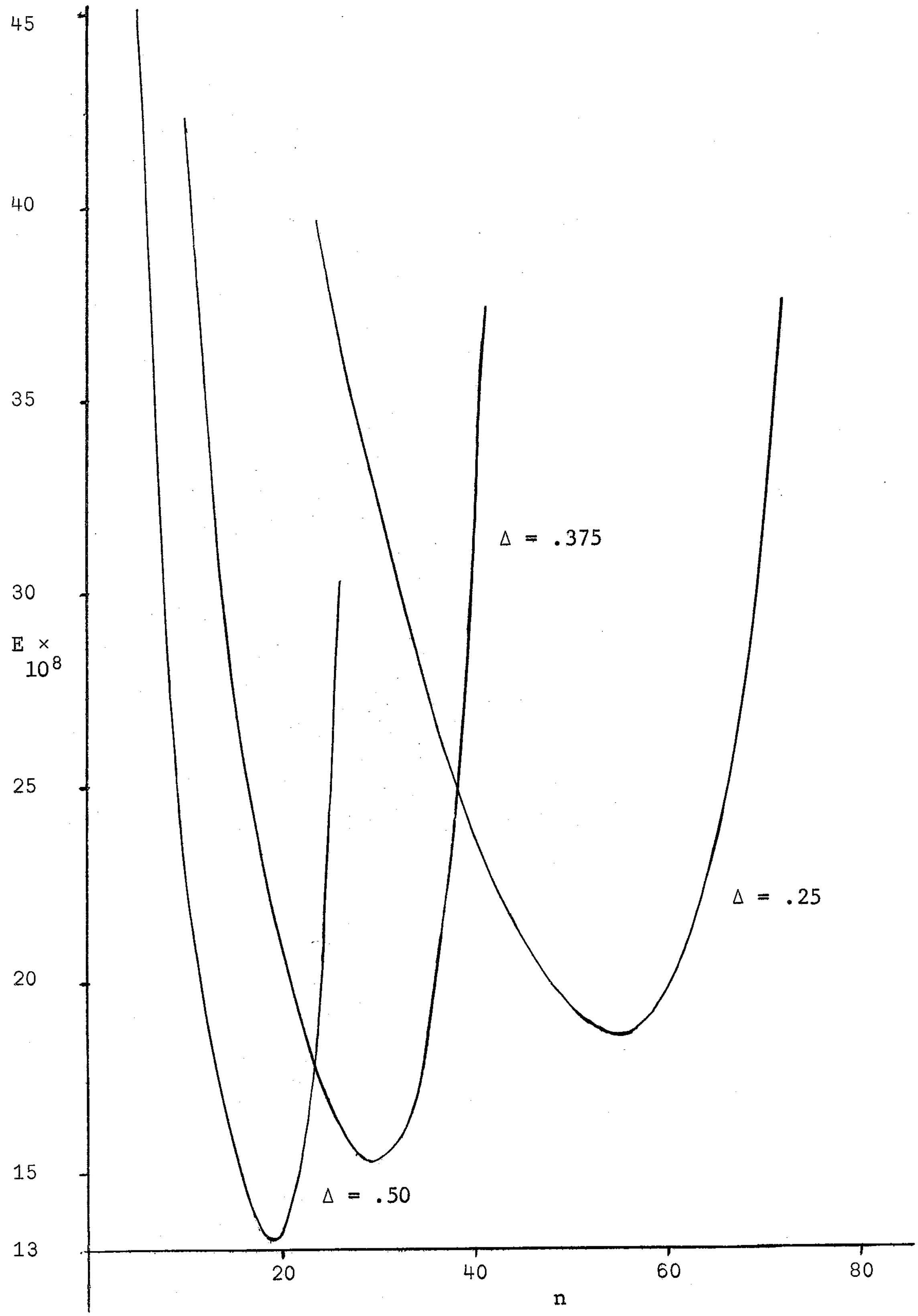

Fig. 5.5.6. $E \times 10^8$ against n. c = .00001, m = 17.

Fig. 5.5.7. Correspondence between optimum values of n and unknown true cost when Average Excess Risk is approximately minimized (logarithmic scale for cost).

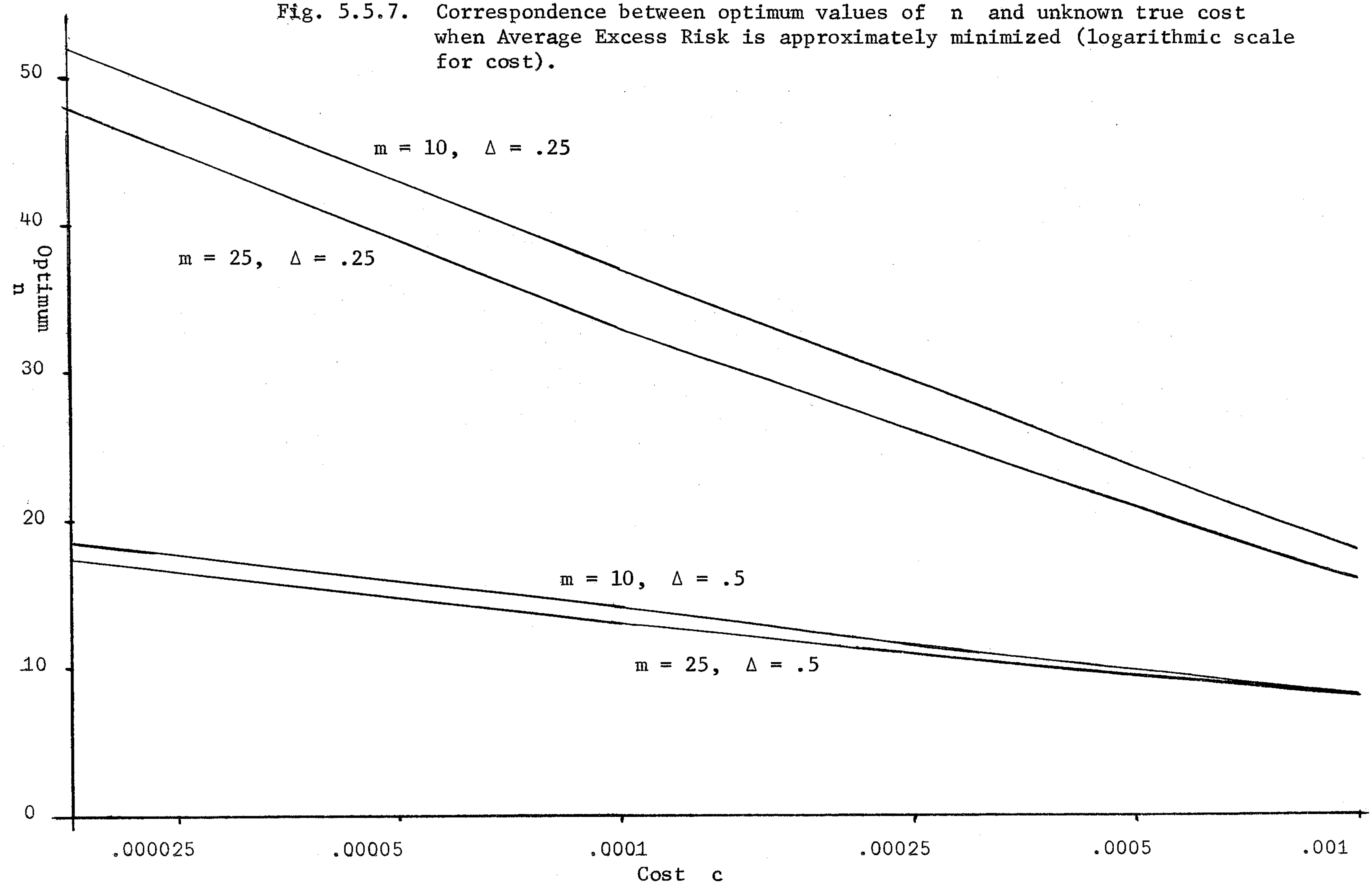

TABLE 5.5.6

Minimum (approx.) average excess and Optimum values of n in PSPRT procedures for given values of m and Δ.

	m = 10, Δ = .25		m = 25, Δ = .25		m = 10, Δ = .5		m = 25, Δ = .5	
True Cost	Opt. n	Min. Av. Excess	Opt. n	Min. Av. Excess	Opt. n	Min. Av. Excess	Opt. n	Min. Av. Excess
.01					4	.001282	3	.000630
.001	18	.00010507	16	$10^{-5} \times 4.76$	8	$10^{-5} \times 5.36$	8	$10^{-5} \times 2.39$
.0001	37	$10^{-6} \times 4.87$	33	$10^{-6} \times 2.13$	14	$10^{-6} \times 3.14$	13	$10^{-6} \times 1.38$
.00001	57	$10^{-7} \times 3.04$	53	$10^{-7} \times 1.31$	20	$10^{-7} \times 2.17$	19	$10^{-7} \times 0.94$

Comparison with Quadrature.

Returning to the discussion at the end of 5.4, a comparison was made for similar values of m, Δ and c of the average excess as approximated by statistical differentials and by quadrature. Table 5.5.7 shows the results.

TABLE 5.5.7

Comparison of Approximations to Average Excess in PSPRT Procedures: Test of Normal Mean

R = Chernoff a.m. Risk in Wald SPRT	n	Av. Excess by Quadrature $= E_Q$	Av. Excess by Stat. Diffls. $= E_D$	Ratio E_Q/E_D
m = 25	1	$10^{-6} \times 6.443$	$10^{-6} \times 6.404$	1.006
Δ = .25	35	$10^{-7} \times 1.872$	$10^{-7} \times 1.827$	1.025
c = .00001	50	$10^{-7} \times 1.385$	$10^{-7} \times 1.333$	1.039
	52	$10^{-7} \times 1.367$	$10^{-7} \times 1.311$	1.043
R = .002893	53	$10^{-7} \times 1.351$	$10^{-7} \times 1.308$	1.033
	54	$10^{-7} \times 1.365$	$10^{-7} \times 1.308$	1.044
	60	$10^{-7} \times 1.528$	$10^{-7} \times 1.474$	1.037
	1	$10^{-5} \times 1.624$	$10^{-5} \times 1.601$	1.014
m = 10	30	$10^{-7} \times 5.362$	$10^{-7} \times 5.339$	1.004
Δ = .25	40	$10^{-7} \times 4.027$	$10^{-7} \times 4.005$	1.006
c = .00001	50	$10^{-7} \times 3.285$	$10^{-7} \times 3.242$	1.013
	56	$10^{-7} \times 3.076$	$10^{-7} \times 3.035$	1.014
R = .002893	57	$10^{-7} \times 3.068$	$10^{-7} \times 3.032$	1.012
	58	$10^{-7} \times 3.070$	$10^{-7} \times 3.038$	1.011
	60	$10^{-7} \times 3.111$	$10^{-7} \times 3.067$	1.014
	70	$10^{-7} \times 4.232$	$10^{-7} \times 4.182$	1.012
	80	$10^{-7} \times 8.312$	$10^{-7} \times 8.232$	1.010
	1	$10^{-6} \times 1.610$	$10^{-6} \times 1.600$	1.006
m = 25	15	$10^{-7} \times 1.082$	$10^{-7} \times 1.058$	1.023
Δ = .50	18	$10^{-7} \times 0.954$	$10^{-7} \times 0.943$	1.012
c = .00001	19	$10^{-7} \times 0.955$	$10^{-7} \times 0.942$	1.014
	20	$10^{-7} \times 0.991$	$10^{-7} \times 0.969$	1.023
R = .0008346	25	$10^{-7} \times 2.274$	$10^{-7} \times 2.257$	1.008
	30	$10^{-7} \times 7.943$	$10^{-7} \times 7.915$	1.004

Table 5.5.7 (continued)

m = 17	1	$10^{-5} \times 43.417$	$10^{-5} \times 43.011$	1.009
Δ = .375	7	$10^{-5} \times 6.166$	$10^{-5} \times 6.141$	1.004
c = .001	10	$10^{-5} \times 4.531$	$10^{-5} \times 4.459$	1.016
R = .07279	11	$10^{-5} \times 4.395$	$10^{-5} \times 4.298$	1.023
	12	$10^{-5} \times 4.508$	$10^{-5} \times 4.385$	1.028
	13	$10^{-5} \times 4.928$	$10^{-5} \times 4.779$	1.031
	14	$10^{-5} \times 5.723$	$10^{-5} \times 5.549$	1.031
	15	$10^{-5} \times 6.968$	$10^{-5} \times 6.771$	1.029
m = 25	1	$10^{-4} \times 18.991$	$10^{-4} \times 18.416$	1.031
Δ = .5	2	$10^{-4} \times 9.540$	$10^{-4} \times 9.005$	1.059
c = .01	3	$10^{-4} \times 7.217$	$10^{-4} \times 6.298$	1.146
R = .244126	4	$10^{-4} \times 8.662$	$10^{-4} \times 7.515$	1.153
	5	$10^{-4} \times 14.110$	$10^{-4} \times 12.916$	1.093
	10	$10^{-4} \times 107.003$	$10^{-4} \times 106.248$	1.007

The results shown by quadrature can be considered more accurate, and they give an approximate average excess which is consistently a little greater than that given by statistical differentials (and if $c \leq .001$ only very slightly so), as shown in the last column of Table 5.5.7. This indicates that the statistical differential method gives reasonably accurate results for the average excess based on the Chernoff a.m. risk, whenever the true value c of the cost is small enough. Since the results by the latter method seem to lag behind the corresponding results by quadrature, it may be that the contribution from the third-order term in the differential expansion is sizeable enough to correct the lag, viz.,

$$\frac{1}{3!} E[\pi^* - \pi)^3] \frac{\partial^3}{\partial \pi^{*3}} [R(\tfrac{1}{2}, n, \bar{c}_n | c)] \Big|_{\bar{c}_n = c} .$$

What is important, however, is that the optimum value of n in all examples studied never varies by more than 1, even when $c = .01$ and the average excess approximations differ by as much as 15%. This perhaps is a more important point to be noted than the comparison of average excess values.

CHAPTER VI

ACCURACY OF APPROXIMATIONS TO THE A.M. RISK OF A P.S.P.R.T.

6.1. Introduction.

In this chapter, some attempt is made to study how close the approximations used in the last three chapters are to their true values. Two different kinds of approximations have been used. The first is the set of Wald approximations to the O.C. function and ASN of a SPRT procedure, based on nominal error probabilities [24]. These were used for the conditional test $T(\underline{x}_n)$ of Chapter III to evaluate approximations to the O.C. function and ASN of a PSPRT procedure. Wald's bounds on $L(\theta)$ and on $E(N|\theta)$ for a SPRT are generally closer if $\theta \leq \theta_0$ or $\theta \geq \theta_1$ than if $\theta_0 < \theta < \theta_1$, and since the main interest in these chapters is the risk function for the two-point parameter space (θ_0, θ_1), the analysis should not lead to bounds which are too far apart to be of interest in the case of the PSPRT. In Sections 6.2 and 6.3, these bounds are obtained for the OC function and ASN respectively, and in Section 6.4 some numerical results are given for the problem of testing a normal mean when the variance is known. In Section 6.5, bounds for the Chernoff a.m. risk are given, and further numerical results are tabled.

The second kind of approximation which has been used is Chernoff's approximation to the nominal a.m. risk, based upon nominal error probabilities [7]. We noted at the end of 2.4 that when the cost is small the Chernoff a.m. risk in the computed results is close to the nominal a.m. risk. Analytically, the determination of general bounds on the true a.m. risk would repay consideration, but has not been successfully considered here.

6.2. Bounds on the O.C. function. (based on Wald)

If

$$\delta_\theta = \frac{1 - \Phi(-|\theta - \frac{\theta_1+\theta_0}{2}|)}{1 - \Phi(|\theta - \frac{\theta_1+\theta_0}{2}|)}, = \frac{\Phi(|\theta - \frac{\theta_0+\theta_1}{2}|)}{\Phi(-|\theta - \frac{\theta_0+\theta_1}{2}|)} \tag{6.2.1}$$

and

$$\eta_\theta = \frac{1}{\delta_\theta} \tag{6.2.2}$$

then for $h(\theta) = h > 0$, and when (3.1.2) holds,

$$\frac{e^{ha_n'} - 1}{e^{ha_n'} - \eta_\theta e^{hb_n'}} \le L_{\underline{x}_n}(\theta) \le \frac{\delta_\theta e^{ha_n'} - 1}{\delta_\theta e^{ha_n'} - e^{hb_n'}}$$, using the notation of Wald [22] and of 3.1, i.e.,

$$\frac{e^{ha} - e^{hn(\Delta\bar{x}-\frac{1}{2}(\theta_1^2-\theta_0^2))}}{e^{ha} - \eta_\theta e^{hb}} \le L_{\underline{x}_n}(\theta) \le \frac{\delta_\theta e^{ha} - e^{hn(\Delta\bar{x}-\frac{1}{2}(\theta_1^2-\theta_0^2))}}{\delta_\theta e^{ha} - e^{hb}} . \tag{6.2.3}$$

$L_n(\theta)$: -- For the lower bound, we replace e^{hb} by $\eta_\theta e^{hb}$ in (3.3.11).

For the upper bound, we replace e^{ha} by $\delta_\theta e^{ha}$ in (3.3.11).

Then, as in (3.3.9), let

$$\left.\begin{aligned} \Phi_+(a,\theta) &= \Phi\left\{\frac{a}{\sqrt{n}(\theta_1-\theta_0)} + \tfrac{1}{2}\sqrt{n}\,(\theta_1+\theta_0-2\theta)\right\} \\ \Phi_-(a,\theta) &= \Phi\left\{\frac{a}{\sqrt{n}(\theta_1-\theta_0)} - \tfrac{1}{2}\sqrt{n}\,(\theta_1+\theta_0-2\theta)\right\} \end{aligned}\right\} . \tag{6.2.4}$$

Similarly, defining $\Phi_+(b,\theta)$, $\Phi_-(b,\theta)$, we get

$$\frac{\eta_\theta e^{hb}}{\eta_\theta e^{hb}-e^{ha}}\Phi_+(b,\theta) - \frac{e^{ha}}{\eta_\theta e^{hb}-e^{ha}}\Phi_+(a,\theta) + \frac{1}{\eta_\theta e^{hb}-e^{ha}}\{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$\le L_n(\theta) \le \frac{e^{hb}}{e^{hb}-\delta_\theta e^{ha}}\Phi_+(b,\theta) - \frac{\delta_\theta e^{ha}}{e^{hb}-\delta_\theta e^{ha}}\Phi_+(a,\theta) + \frac{1}{e^{hb}-\delta_\theta e^{ha}}\{\Phi_-(a,\theta)$$

$$- \Phi_-(b,\theta)\} ,$$

i.e.,

$$\frac{e^{hb}}{e^{hb}-\delta_\theta e^{ha}}\Phi_+(b,\theta) - \frac{\delta_\theta e^{ha}}{e^{hb}-\delta_\theta e^{ha}}\Phi_+(a,\theta) + \frac{\delta_\theta}{e^{hb}-\delta_\theta e^{ha}}\{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$\le L_n(\theta) \le \frac{e^{hb}}{e^{hb}-\delta_\theta e^{ha}}\Phi_+(b,\theta) - \frac{\delta_\theta e^{ha}}{e^{hb}-\delta_\theta e^{ha}}\Phi_+(a,\theta) + \frac{1}{e^{hb}-\delta_\theta e^{ha}}\{\Phi_-(a,\theta)$$

$$- \Phi_-(b,\theta)\} . \tag{6.2.5}$$

The accuracy of these bounds is measured by their difference, viz.,

$$\frac{\delta_\theta - 1}{|e^{hb} - \delta_\theta e^{ha}|} \{\Phi_-(a,\theta) - \Phi_-(b,\theta)\} \tag{6.2.6}$$

as compared with $\dfrac{\delta_\theta - 1}{|e^{hb} - \delta_\theta e^{ha}|}$ for the SPRT procedure.

Clearly, closer bounds are given in the PSPRT procedure, since

$$0 < \Phi_-(a,\theta) - \Phi_-(b,\theta) < 1.$$

Further, $\Phi_-(a,\theta) - \Phi_-(b,\theta)$ is decreasing in n for all a, b, θ and $\theta_1 - \theta_0 > 0$, where $h > 0 \Rightarrow \theta < \dfrac{\theta_1 + \theta_0}{2}$. Hence the bounds on $L_n(\theta)$ become closer as n increases.

$$h < 0: -\frac{1 - e^{ha_n'}}{\delta_\theta e^{hb_n'} - e^{ha_n'}} \le L_{\underline{x}_n}(\theta) \le \frac{1 - \eta_\theta e^{ha_n'}}{e^{hb_n'} - \eta_\theta e^{ha_n'}} . \tag{6.2.7}$$

So, for the lower bound of $L_n(\theta)$, replace e^{hb} by $\delta_\theta e^{hb}$ in (3.3.11), and for the upper bound of $L_n(\theta)$, replace e^{ha} by $\eta_\theta e^{ha}$ in (3.3.11). This leads to

$$\frac{\delta_\theta e^{hb}}{\delta_\theta e^{hb} - e^{ha}} \Phi_+(b,\theta) - \frac{e^{ha}}{\delta_\theta e^{hb} - e^{ha}} \Phi_+(a,\theta) + \frac{1}{\delta_\theta e^{hb} - e^{ha}} \{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$\le L_n(\theta) \le \frac{\delta_\theta e^{hb}}{\delta_\theta e^{hb} - e^{ha}} \Phi_+(b,\theta) - \frac{e^{ha}}{\delta_\theta e^{hb} - e^{ha}} \Phi_+(a,\theta) + \frac{\delta_\theta}{\delta_\theta e^{hb} - e^{ha}} \{\Phi_-(a,\theta)$$

$$- \Phi_-(b,\theta)\} . \tag{6.2.8}$$

The accuracy is again measured by the difference

$$\frac{\delta_\theta - 1}{|\delta_\theta e^{hb} - e^{ha}|}\{\Phi_-(a,\theta) - \Phi_-(b,\theta)\},$$

again decreasing as n increases.

Since we shall be largely concerned with the risk when $(w, 1-w)$ is the prior distribution on $\Theta = (\theta_0, \theta_1)$, the above bounds are of interest when $\theta = \theta_0$ or θ_1, when

$$\delta_{\theta_0} = \delta_{\theta_1} = \frac{\Phi(\frac{1}{2}\Delta)}{\Phi(-\frac{1}{2}\Delta)} . \qquad (6.2.9)$$

6.3. Bounds on the Average Sample Number.

Again, using Wald's bounds, and writing $\bar{L}$, $\underline{L}$ for upper and lower bounds of the O.C. function, let

$$\left.\begin{aligned} \xi_\theta &= \Delta[\bar{\theta} + \frac{\phi(\bar{\theta})}{\Phi(\bar{\theta})}] \\ \xi'_\theta &= -\Delta[-\theta + \frac{\phi(-\bar{\theta})}{\Phi(-\bar{\theta})}] \end{aligned}\right\} \qquad (6.3.1)$$

where $\bar{\theta} = \theta - \frac{1}{2}(\theta_1+\theta_0)$.

Then if $h < 0$, and $Z = \log \frac{p_{\theta_1}(X)}{p_{\theta_0}(X)}$, with the notation of 3.1,

$$\frac{\bar{L}_{\underline{x}_n}(\theta)(b'_n+\xi'_\theta) + (1-\bar{L}_{\underline{x}_n}(\theta))a}{E_\theta(Z)} \le E(N|\underline{x}_n,\theta) - n \le \frac{\underline{L}_{\underline{x}_n}(\theta)b + (1-\underline{L}_{\underline{x}_n}(\theta))(a+\xi_\theta)}{E_\theta(Z)}$$

(6.3.2)

whenever sampling is continued beyond stage n. $(h < 0 \Rightarrow E_\theta(Z) > 0)$.

So for the conditional test $T(\underline{x}_n)$, if $N > n$ and $h < 0$,

$$\frac{a_n' - \bar{L}_{\underline{x}_n}(\theta)(a_n'-b_n'-\xi_\theta')}{E_\theta(Z)} \le E_{\underline{x}_n}(N|\theta) \le \frac{a_n' + \xi_\theta - \underline{L}_{\underline{x}_n}(\theta)(a_n'+\xi_\theta-b_n')}{E_\theta(Z)} .$$

$E_\theta(Z) = -\frac{1}{2}h\Delta^2$, so if $N > n$,

lower bound for

$$E_{\underline{x}_n}(N|\theta) \times E_\theta(Z) \text{ is } a_n' - \frac{\delta_\theta - e^{ha_n'}}{\delta_\theta e^{hb_n'} - e^{ha_n'}}(a_n' - b_n' - \xi_\theta') + n,$$

$$= a - n\Delta(\bar{x}_n-\theta) + \tfrac{1}{2}nh\Delta^2 - \frac{\delta_\theta \exp\{n\Delta(\bar{x}_n-\theta)+\frac{1}{2}nh\Delta^2\} - e^{ha}}{\delta_\theta e^{hb} - e^{ha}} \cdot$$

$$\cdot\ (a - b - \xi_\theta') + n \qquad (6.3.3)$$

from (6.2.7).

Hence, integrating, the lower bound for $E_n(N|\theta)$ is given by

$$n - \frac{1}{\frac{1}{2}h\Delta^2}\left[\int_{b_n''}^{a_n''}\{a + \tfrac{1}{2}nh\Delta^2 + \frac{e^{ha}}{\delta_\theta e^{hb} - e^{ha}}(a-b-\xi_\theta')\}dF(\bar{x}_n)\right.$$

$$- n\Delta\int_{b_n''}^{a_n''}(\bar{x}_n - \theta)dF(\bar{x}_n)$$

$$\left. - \frac{\delta_\theta(a-b-\xi_\theta')}{\delta_\theta e^{hb} - e^{ha}} e^{\frac{1}{2}nh\Delta^2}\int_{b_n''}^{a_n''}\exp\{n\Delta(\bar{x}_n - \theta)\}dF(\bar{x}_n)\right], \qquad (6.3.4)$$

where a_n'', b_n'' are defined in (3.3.3).

The integrals in (6.3.4) have been already evaluated in 3.3 and 3.4. We get for $h < 0$

$$\underline{E}_n(N|\theta) = n + \frac{2}{h\Delta^2}[-\{a + \tfrac{1}{2}nh\Delta^2 + \frac{e^{ha}}{\delta_\theta e^{hb}-e^{ha}}(a-b-\xi'_\theta)\}\{\Phi_+(a,\theta) - \Phi_+(b,\theta)\}$$

$$+ \frac{\delta_\theta(a-b-\xi'_\theta)}{\delta_\theta e^{hb}-e^{ha}} \cdot \{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$+ \sqrt{\frac{n}{2\pi}} \cdot \Delta \{\exp(-\tfrac{1}{2}(\frac{b}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta)^2) - \exp(-\tfrac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta)^2)\}] . \tag{6.3.5}$$

Using a similar procedure, we get for $h < 0$, when $N > n$, $E_\theta(Z) \times$ (upper bound for $E\underline{x}_n(N|\theta)$) is given by

$$a'_n + \xi_\theta - \frac{1-e^{ha'_n}}{\delta_\theta e^{hb'_n}-e^{ha'_n}}(a - b + \xi_\theta) \tag{6.3.6}$$

and

$$\bar{E}_n(N|\theta) = n + \frac{2}{h\Delta^2}[-\{a + \xi_\theta + \tfrac{1}{2}nh\Delta^2 + \frac{e^{ha}}{\delta_\theta e^{hb}-e^{ha}}(a-b+\xi_\theta)\}\{\Phi_+(a,\theta)$$

$$- \Phi_+(b,\theta)\} + \frac{a-b+\xi_\theta}{\delta_\theta e^{hb}-e^{ha}}\{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$+ \sqrt{\frac{n}{2\pi}} \cdot \Delta \{\exp(-\tfrac{1}{2}(\frac{b}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta)^2) - \exp(-\tfrac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta)^2)\}] . \tag{6.3.7}$$

The accuracy of the approximation to $E_n(N|\theta)$ in 3.4 when $h(\theta) < 0$ can be measured by the difference

$$\bar{E}_n(N|\theta) - \underline{E}_n(N|\theta) = \frac{2}{h\Delta^2}\left[-\left\{\xi_\theta + \frac{e^{ha}}{\delta_\theta e^{hb}-e^{ha}}(\xi_\theta+\xi'_\theta)\right\}\{\Phi_+(a,\theta) - \Phi_+(b,\theta)\}\right.$$

$$\left. + \frac{(a-b)(1-\delta_\theta) + \xi_\theta + \delta_\theta\xi'_\theta}{\delta_\theta e^{hb}-e^{ha}}\{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}\right] . \tag{6.3.8}$$

However, for comparison, a better measure of accuracy might be the ratio

$$\frac{2\{\bar{E}_n(N|\theta) - \underline{E}_n(N|\theta)\}}{\bar{E}_n(N|\theta) + \underline{E}_n(N|\theta)} .$$

If $h > 0$, we get for the conditional test $T(\underline{x}_n)$

$$\frac{\underline{L}_{\underline{x}_n}(\theta)b'_n + (1-\underline{L}_{\underline{x}_n}(\theta))(a'_n+\xi_\theta)}{E_\theta(Z)} \le E_{\underline{x}_n}(N|\theta) \le \frac{\bar{L}_{\underline{x}_n}(\theta)(b'_n+\xi'_\theta) + (1-\bar{L}_{\underline{x}_n}(\theta))a'_n}{E_\theta(Z)} \tag{6.3.9}$$

therefore

$$E_\theta(Z) \times \text{lower bound} = a'_n + \xi_\theta - \frac{1-e^{ha'_n}}{\eta_\theta e^{hb'_n}-e^{ha'_n}}(a - b + \xi_\theta) . \tag{6.3.10}$$

So, for test T_n, we get when $h > 0$

$$\underline{E}_n(N|\theta) = n + \frac{2}{h\Delta^2}\left[-\left\{a + \xi_\theta + \tfrac{1}{2}nh\Delta^2 + \frac{e^{ha}}{\eta_\theta e^{hb}-e^{ha}}(a-b+\xi_\theta)\right\}\{\Phi_+(a,\theta)\right.$$

$$- \Phi_+(b,\theta)\} + \frac{(a-b+\xi_\theta)}{\eta_\theta e^{hb}-e^{ha}}\{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$\left. + \sqrt{\frac{n}{2\pi}} \cdot \Delta\left\{\exp\left(-\tfrac{1}{2}\left(\frac{b}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta\right)^2\right) - \exp\left(-\tfrac{1}{2}\left(\frac{a}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta\right)^2\right)\right\}\right] . \tag{6.3.11}$$

Similarly, for $T(\underline{x}_n)$

$$E_\theta(Z) \times \text{upper bound} = a_n' - \frac{\eta_\theta - e^{ha_n'}}{\eta_\theta e^{hb_n'} - e^{ha_n'}} (a - b - \xi_\theta') \tag{6.3.12}$$

and for test T_n, when $h > 0$,

$$\bar{E}_n(N|\theta) = n + \frac{2}{h\Delta^2} [-\{a + \tfrac{1}{2}nh\Delta^2 + \frac{e^{ha}}{\eta_\theta e^{hb} - e^{ha}} (a-b-\xi_\theta')\}\{\Phi_+(a,\theta) - \Phi_+(b,\theta)\}$$

$$+ \frac{\eta_\theta(a-b-\xi_\theta')}{\eta_\theta e^{hb} - e^{ha}} \{\Phi_-(a,\theta) - \Phi_-(b,\theta)\}$$

$$+ \sqrt{\frac{n}{2\pi}} \cdot \Delta \{\exp(-\tfrac{1}{2}(\frac{b}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta)^2) - \exp(-\tfrac{1}{2}(\frac{a}{\sqrt{n}\Delta} + \frac{h}{2}\sqrt{n}\Delta)^2)\}] \ . \tag{6.3.13}$$

The accuracy for $h > 0$ is again measured by $\bar{E}_n(N|\theta) - \underline{E}_n(N|\theta)$ in a result similar to (6.3.8)

6.4. Numerical Results.

As one would expect, the bounds on $L_n(\theta)$ and on $E_n(N|\theta)$ narrow as n increases, a reflection on the increasing probability of stopping observations at stage n, as n increases, and since this event has a probability whose value is known exactly, the information we have on the O.C. function and on the ASN is correspondingly more complete.

The following tables show some of the results of numerical analysis. One interesting case, when $n = 1$, gives bounds for a Wald SPRT procedure, and (6.2.6) shows that they are closer than those given by Wald. However, the factor $\Phi_-(a,\theta) - \Phi_-(b,\theta)$ is generally too close to 1 to be at all substantial. Only if a is not too large, say $a \le 2$, and Δ not too small, say $\Delta \ge .5$, is any noticeable improvement observed, and Table 6.4.1 shows this for $a = -b = 2$, $\Delta = 1$. (The factor here, for $n = 1$, is

$$\Phi_{-}(a,\theta) - \Phi_{-}(b,\theta) = \Phi(\frac{a}{\Delta} - \tfrac{1}{2}\Delta) - \Phi(\frac{b}{\Delta} - \tfrac{1}{2}\Delta) \quad .)$$

Comparing the difference in bounds in Table 6.4.1, the following is of interest.

When $\theta = \theta_0$ or $\theta = \theta_1$, $\bar{L}(\theta) - \underline{L}(\theta)$ improves by .0055 in .07
and $\bar{E}(N|\theta) - \underline{E}(N|\theta)$ improves by .196 in 2.485.

When $\theta = \theta_i + (-1)^i\Delta$; $i = 0, 1$, $\bar{L}(\theta) - \underline{L}(\theta)$ improves by .0073 in .143
and $\bar{E}(N|\theta) - \underline{E}(N|\theta)$ improves by .450 in 8.823.

TABLE 6.4.1.

Bounds for O.C. function and A.S.N. in PSPRT and SPRT

$a = -b = 2, \quad \theta_1 - \theta_0 = 1$

n	$\theta-\theta_0$	$\underline{L}_n(\theta)$	$\bar{L}_n(\theta)$	Difference	$\underline{E}_n(N\|\theta)$	$\bar{E}_n(N\|\theta)$	Difference
SPRT: Wald's Approximations	1	.0526	.1282	.0756	2.810	5.491	2.681
1	1	.0550	.1251	.0701	2.902	5.387	2.485
5	1	.0426	.0720	.0294	5.652	6.696	1.044
10	1	.0217	.0337	.0120	10.255	10.679	.424
SPRT: Wald's Approximations	.667	.2399	.3903	.1504	.900	10.173	9.273
1	.667	.2434	.3865	.1431	1.122	9.945	8.823
5	.667	.2456	.3354	.0898	4.827	10.359	5.532
10	.667	.2236	.2866	.0630	9.851	13.735	3.884

The difference in bounds drops noticeably as n increases.

TABLE 6.4.2.

Bounds in PSPRT; $a = 4, \quad b = -2.5, \quad \theta_1 - \theta_0 = .25$

n	$\theta-\theta_0$	$\underline{L}_n(\theta)$	$\bar{L}_n(\theta)$	Difference	$\underline{E}_n(N\|\theta)$	$\bar{E}_n(N\|\theta)$	Difference
1	0	.9829	.9862	.0033	76.34	83.80	7.46
50	0	.9834	.9864	.0030	87.12	92.49	5.37
1	.083	.8248	.8438	.0190	127.39	159.00	31.61
50	.083	.8233	.8410	.0177	137.42	165.54	28.12
1	.167	.3354	.3667	.0313	148.28	187.68	39.40
50	.167	.3251	.3521	.0270	157.03	192.38	35.35
1	.25	.0661	.0810	.0149	110.67	120.57	9.90
50	.25	.0569	.0673	.0104	115.63	123.94	8.31

TABLE 6.4.3.

Bounds in PSPRT; $a = 4$, $b = -2.5$, $\theta_1-\theta_0 = .50$

n	$\theta-\theta_0$	$\underline{L}_n(\theta)$	$\bar{L}_n(\theta)$	Diff.	$\underline{E}_n(N\|\theta)$	$\bar{E}_n(N\|\theta)$	Diff.
1	0	.9827	.9887	.0060	19.05	22.96	3.91
15	0	.9838	.9890	.0052	23.11	25.63	2.52
30	0	.9884	.9915	.0031	33.99	35.31	1.32
1	.167	.8185	.8550	.0365	30.01	45.88	15.87
15	.167	.8181	.8511	.0330	34.00	47.57	13.57
30	.167	.8229	.8505	.0276	44.21	55.13	10.92
1	.333	.3114	.3716	.0602	34.61	54.27	19.66
15	.333	.3012	.3508	.0496	38.24	55.28	17.04
30	.333	.2753	.3143	.0390	47.72	61.56	13.84
1	.500	.0541	.0812	.0271	27.55	32.58	5.03
15	.500	.0455	.0625	.0170	29.61	33.57	3.96
30	.500	.0284	.0371	.0107	37.69	40.03	2.34

With the same test boundaries a and b, but Δ double its value in Table 6.4.2, Table 6.4.3 shows a more spectacular decrease in the difference in bounds (proportionally) as n increases. For the O.C. function, the factor

$$\Phi_{-}(a,\theta) - \Phi_{-}(b,\theta)$$

would suggest this to be so; for given a and n,

$$\Phi\left(\frac{a}{\sqrt{n}\Delta} - \tfrac{1}{2}\sqrt{n}\Delta\right)$$

decreases as Δ increases.

A final example illustrates how the upper bound $\bar{\bar{E}}_n(N|\theta)$ may decrease initially as n increases, before beginning to increase with n. Also, the lower bound for the error probabilities may increase initially.

TABLE 6.4.4a.

Behavior of $\bar{\bar{E}}_n(N|\theta)$ as n increases. $a = -b = 3$, $\theta_1 - \theta_0 = .50$

n	1	5	15	30
$\theta-\theta_0 = 0$	25.945	25.935	27.77	36.44
$\theta-\theta_0 = .333$	46.395	46.350	47.50	54.64

TABLE 6.4.4b.

Upper bounds $\bar{\bar{\alpha}}_n$ of error probabilities as n increases;
$a = -b = 3$, $\theta_1 - \theta_0 = .50$

n	1	5	15	30
$\bar{\bar{\alpha}}_n = \bar{\bar{\beta}}_n$	.03176	.03180	.02857	.01909

6.5. Bounds on the Risk of a P.S.P.R.T.

The bounds calculated earlier for the O.C. function and ASN of a PSPRT can be used to calculate bounds for the risk. Suppose θ_0 and θ_1 have prior probabilities w and $1-w$; then for cost c and unit loss, the true risk lies between

$$w(1 - \bar{L}_n(\theta_0)) + (1-w)\underline{L}_n(\theta_1) + cw\, \underline{E}_n(N|\theta_0) + c(1-w)\, \underline{E}_n(N|\theta_1)$$

and

$$w(1 - \underline{L}_n(\theta_0)) + (1-w)\bar{L}_n(\theta_1) + cw\, \bar{E}_n(N|\theta_0) + c(1-w)\, \bar{E}_n(N|\theta_1).$$

Of particular interest are bounds for the a.m. risk. The procedure used earlier to yield the Chernoff nominal a.m. risk and the bounds developed here will not give bounds for the true a.m. risk. All that is known is that the Chernoff nominal a.m. risk is usually close to the nominal a.m. risk produced by the iterative methods of Chapter II. The tables below, therefore, give bounds for the risk when the boundaries of the test procedure are those of (5.2.6), viz.,

$$a = \log \left(\frac{w}{1-w} \cdot \frac{1-\frac{2c}{\Delta^2}}{\frac{2c}{\Delta^2}} \right)$$

$$b = \log \left(\frac{w}{1-w} \cdot \frac{\frac{2c}{\Delta^2}}{1-\frac{2c}{\Delta^2}} \right)$$

so that a and b are given by

$$\log\left(\frac{w}{1-w}\right) \pm \log\left\{\frac{1-\frac{2c}{\Delta^2}}{\frac{2c}{\Delta^2}}\right\} . \qquad (6.5.1)$$

These boundaries arise from the Chernoff nominal error probabilities, and yield the Chernoff a.m. risk. The optimum value of n, minimizing the approximate average excess is also shown for the case in which (using the notation of Chapter V),

$\bar{c}_n \sim \Gamma(mn, \frac{mn}{c})$ and $m = 25$.

TABLE 6.5.1.

Bounds on the Chernoff a.m. risk; $w = \frac{1}{2}$, $c_n \sim \Gamma(mn, \frac{mn}{c})$, $m = 25$

	n	Lower bound	Chernoff a.m. risk	Upper bound
Δ = .5	1	.006231	.006495	.006856
c = .0001	25	.006269	.006518	.006860
a = -b = 7.1301	50	.006906	.007064	.007281
Optimum n = 13	100	.010406	.010443	.010494
Δ = .5	1	.04328	.04595	.04973
c = .001	15	.04366	.04615	.04970
ε = -b = 4.8203	25	.04575	.04775	.05059
Optimum n = 8	50	.05889	.05980	.06109
	100	.10162	.10178	.10202
Δ = .5	1	.2144	.2441	.2907
c = .01	5	.2170	.2452	.2892
a = -b = 2.4423	10	.2319	.2547	.2903
Optimum n = 3	15	.2557	.2736	.3017
	25	.3196	.3309	.3486
Δ = .25	1	.02087	.02146	.02215
c = .0001	100	.02130	.02180	.02239
a = -b = 5.7414	200	.02519	.02544	.02575
Optimum n = 33	300	.03651	.03659	.03668
Δ = .25	1	.1281	.1341	.1422
c = .001	25	.1283	.1342	.1421
a = -b = 3.4095	50	.1314	.1366	.1434
Optimum n = 16	150	.1814	.1834	.1861
	250	.2626	.2634	.2644
Δ = .25	1	.2990	.4068	.5247
c = .01	10	.3582	.4256	.4992
a = -b = .7538	15	.3922	.4476	.5082
Optimum n = 3	25	.4628	.5040	.5491
	50	.6566	.6814	.7084

The above data show:

i) The bounds are proportionally closer when n is larger, for any given cost.

ii) The difference narrows considerably as n increases, and becomes very narrow as the risk $\to$ cn. This is to be expected, because as the contribution from the term cn begins to dominate, the information on the ASN becomes correspondingly more complete.

iii) The upper bound for the risk again decreases initially as n increases, in many instances; this is reflected in similar behavior noted earlier in upper bounds for the ASN.

iv) The (approximate) optimal values of n minimizing the average excess risk for unknown cost generally are such that the bounds above have not altered substantially beyond their values for $n = 1$; a slight sharpening of bounds is all that seems possible to claim.

This last point is perhaps important. The approximate optimal values of n for minimizing average excess are those found earlier in Chapter V using approximate methods, for $m = 25$, where the cost has a Gamma distribution. Since the optimal n values did not appear to vary much with m, the choice of m is not too crucial. It seems intuitive, however, that the optimal n would lie in the range of n for which the risk function does not begin to increase markedly. Consequently, the PSPRT bounds for the true risk when the cost is unknown are not going to be much narrower than for a Wald SPRT with the same test boundaries, except in cases noted earlier (i.e., $a = -b \leq 2$, $\Delta \geq .5$, say).

CHAPTER VII

A MULTIVARIATE EXTENSION OF A SEQUENTIAL DISCRIMINATION PROCEDURE

7.1. Introduction.

In this chapter, we shall extend a procedure developed by Baker [3] and Hall [9] as a sequential analog of Stein's two-stage test [17] of hypotheses about the mean of a normal population with unknown variance and given bounds on the error probabilities. The notation is based on Hall's, and hence differs from that used in earlier chapters.

Let μ and σ^2 be the mean and variance of the population respectively, both unknown. We wish to test the composite hypotheses

$$H_0: \ \mu \le 0, \ \ \sigma > 0$$

vs

$$H_1: \ \mu \ge \Delta > 0, \ \ \sigma > 0$$

with error probabilities bounded above by α and β. When no bound on σ is known, a fixed sample-size test does not exist, and the sequential t-test may be unsatisfactory because H_1 has to be stated in s.d. units, $\mu \ge \delta\sigma$.

In Hall's test procedure T, a first stage estimates σ^2 as in Stein's two-stage test, but sampling then proceeds one observation at a time. The procedure relates closely to that of Baker [3]; compared with results by Stein's procedure, savings are possible in the ASN of T. Hall obtained a sequence of upper and lower bounds on both the O.C. function and the ASN of T.

The extension discussed in this chapter is to the case of a multinormal population in which the covariance matrix $\underline{V}$ is known except for a scalar multiplier. When $\underline{V}$ is generally unknown, the sequential T^2-test can be used, but no approximations to the OC function or to the ASN exist. (See Jackson and Bradley, 1961 [10].)

Suppose $\underline{X}_1, \underline{X}_2, \ldots$ are independent $p \times 1$ random vectors, multinormally distributed with common mean $\underline{\mu}$ and variance-covariance matrix $\underline{V}$, and it is known that either $\underline{\mu} = \underline{\mu}_0$ or $\underline{\mu} = \underline{\mu}_1$. If $\underline{V}$ is known, then Wald's SPRT can be applied as follows:

Let $\underline{z}_i = \underline{x}_i - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_0)$, then

if $\log B = b < (\sum_{i=1}^{N} \underline{z}_i)'\underline{V}^{-1}(\underline{\mu}_1 - \underline{\mu}_0) < a = \log A$,
take a further observation;

if $(\sum_{i=1}^{N} \underline{z}_i)'\underline{V}^{-1}(\underline{\mu}_1 - \underline{\mu}_0) < b$,
assign to population with mean $\underline{\mu}_0$;

if $(\sum_{i=1}^{N} \underline{z}_i)'\underline{V}^{-1}(\underline{\mu}_1 - \underline{\mu}_0) > a$,
assign to population with mean $\underline{\mu}_1$.

Frequently the elements of $\underline{V}$ are not known in full, and here we consider the case in which it is known that $\underline{V} = \lambda \underline{V}_0$, where $\underline{V}_0$ is known, but the positive scalar λ is replaced by an estimator $\hat{\lambda}$ based on m observations. These m observations come either from the first stage of the procedure or from a previous experiment.

The above SPRT procedure is then carried out with $\underline{V}$ replaced by $\hat{\lambda}\underline{V}_0$. a and b are replaced by quantaties such that the error probabilities are bounded by required values.

The properties of this test, as well as the procedure itself, are almost identical with those of Hall's in the univariate case, in which the variance σ^2 is replaced by an estimator s^2. The O.C. function has the same form, provided $\underline{\mu} = \theta\underline{\mu}_1 + (1-\theta)\underline{\mu}_0$, and for the multinormal problem, $h(\underline{\mu}) = -2(\theta-\frac{1}{2})$. The approximations to the O.C. function and to the A.S.N. function are then notationally the same as those demonstrated by Hall and Baker, except that we have $p\nu$ (or pf) degrees of freedom instead of ν (or f), $\hat{\lambda}$ being based on ν (or f) degrees of freedom. Baker's and Hall's tables can be used with this modification.

In 7.2, the test procedure is presented, and in 7.3 and 7.4 the O.C. function and ASN approximations are discussed.

It should be stated that Hall's and Baker's procedures apply whether the estimate s^2 of the variance σ^2 is based upon a previous experiment or upon the first m observations of a two-stage test. In the latter case, however, it should be pointed out that not all the information is used, since Hall's approximations assume that the test dies not terminate with the first stage. If it does so terminate after m observations, it would seem appropriate to study an analog of the PSPRT procedure of Chapter III, replacing σ^2 by s^2. A discussion of this possibility is presented in 7.5.

In the sequel, some familiarity with the notation and content of Sections 1-5 of Hall's paper will be assumed.

7.2. The Test Procedure.

With the above notation, the maximum likelihood estimator of λ based on m observations is

$$\hat{\hat{\lambda}} = \frac{1}{mp} \sum_{i=1}^{m} (\underline{X}_i-\bar{\underline{X}})'\underline{V}_0^{-1}(\underline{X}_i-\bar{\underline{X}}), \tag{7.2.1}$$

where $\bar{\underline{X}} = \frac{1}{m} \sum_{i=1}^{m} \underline{X}_i$.

Now

$$\sum_{i=1}^{m} (\underline{X}_i-\underline{\mu})'\underline{V}^{-1}(\underline{X}_i-\underline{\mu}) = \sum_{i=1}^{m} (\underline{X}_i-\bar{\underline{X}})'\underline{V}^{-1}(\underline{X}_i-\bar{\underline{X}}) + m(\bar{\underline{X}}-\underline{\mu})'\underline{V}^{-1}(\bar{\underline{X}}-\underline{\mu})$$

i.e., in terms of distributions, $\chi^2_{mp} = \chi^2_{p(m-1)} + \chi^2_p$ (7.2.2)

with the components on the right independent, so that $\frac{mp}{\lambda}\hat{\hat{\lambda}} \sim \chi^2_{p(m-1)}$.

Hence $E\hat{\hat{\lambda}} = \frac{m-1}{m}\lambda$.

We shall use the unbiased estimator

$$\hat{\lambda} = \frac{m}{m-1}\hat{\hat{\lambda}} \tag{7.2.3}$$

$$= \frac{1}{p(m-1)} \sum_{i=1}^{m} (\underline{X}_i-\bar{\underline{X}})'\underline{V}_0^{-1}(\underline{X}_i-\bar{\underline{X}}) \tag{7.2.4}$$

which is distributed as a $\frac{\lambda}{\nu}\chi^2_\nu$ variable, with

$$\nu = p(m-1) . \tag{7.2.5}$$

(In the univariate case, $\nu = m-1$, as used by Hall.)

Using Hall's notation, let T be the test based on the SPRT for known $\underline{V}$ described in 7.1, but with $\underline{V} = \lambda\underline{V}_0$ replaced by $\hat{\lambda}\underline{V}_0$.

So, if

$$\underline{Z}_n = \sum_{i=1}^{n} \underline{z}_i = \sum_{i=1}^{n} \underline{x}_i - \frac{n}{2}(\underline{\mu}_1+\underline{\mu}_0), \tag{7.2.6}$$

let

$$r_n(\hat{\lambda}) = \frac{1}{\hat{\lambda}} \underline{Z}_n' \underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0). \tag{7.2.7}$$

Then T is: stop and accept H_0 (decision d_0) if $r_n(\hat{\lambda}) \le b_m = \log B_m$

stop and accept H_1 (decision d_1) if $r_n(\hat{\lambda}) \ge a_m = \log A_m$

otherwise continue sampling.

Analogous to Hall's $T(s, \sigma)$, define $T(\hat{\lambda}, \lambda)$ as the conditional SPRT of $\underline{\mu}_0$ vs $\underline{\mu}_1$, given $\hat{\lambda}$ and λ, with boundaries

$$\bar{a}_m = \log \bar{A}_m = a_m \frac{\hat{\lambda}}{\lambda}, \quad \bar{b}_m = \log \bar{B}_m = b_m \frac{\hat{\lambda}}{\lambda}; \tag{7.2.8}$$

then it is seen that we decide according as

$$r_n(\lambda) \le \bar{b}_m \quad \text{or} \quad r_n(\lambda) \ge \bar{a}_m; \quad n \ge m,$$

and since $r_n(\lambda) = \frac{\hat{\lambda}}{\lambda} r_n(\hat{\lambda})$, decisions based on $T(\hat{\lambda}, \lambda)$ and on T are the same if $\hat{\lambda}$ is computed from the first m observations and $n \ge m$. Hence

$$E \Pr\{d_i \text{ from test } T(\hat{\lambda},\lambda) | \hat{\lambda}, \lambda, \underline{\mu}\} = \Pr\{d_i \text{ from test } T | \lambda, \underline{\mu}\} . \tag{7.2.9}$$

But using Wald's bounds on the error probabilities,

$$\left.\begin{aligned} \Pr\{d_1 \text{ from test } T(\hat{\lambda},\lambda) | \hat{\lambda}, \lambda, \underline{\mu} = \underline{\mu}_0\} &< \frac{1}{\bar{A}_m} = \exp(- a_m \frac{\hat{\lambda}}{\lambda}) \\ \Pr\{d_0 \text{ from test } T(\hat{\lambda},\lambda) | \hat{\lambda}, \lambda, \underline{\mu} = \underline{\mu}_1\} &< \bar{B}_m = \exp(b_m \frac{\hat{\lambda}}{\lambda}). \end{aligned}\right\} \tag{7.2.10}$$

Hence, since $E(\exp(t\chi_\nu^2)) = (1-2t)^{-\frac{1}{2}\nu}$,

we get

$$E \Pr\{d_1 \text{ from test } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda, \underline{\mu} = \underline{\mu}_0\} < E \exp(- a_m \hat{\lambda}/\lambda)$$

$$\text{i.e.} \quad < (1 + \frac{2a_m}{\nu})^{-\frac{1}{2}\nu} ,$$

$$< \alpha$$

$$\left.\begin{aligned} \text{if} \qquad a_m &= \tfrac{1}{2}\nu(\alpha^{-2/\nu} - 1) \\ &= (- \log \alpha)[1 + \frac{-\log\alpha}{\nu} + o(\frac{1}{\nu})] . \end{aligned}\right\} \qquad (7.2.11)$$

Similarly

$$E \Pr\{d_0 \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda, \underline{\mu} = \underline{\mu}_1\} < \beta$$

$$\left.\begin{aligned} \text{if} \qquad b_m &= \tfrac{1}{2}\nu(1 - \beta^{-2/\nu}) \\ &= - (- \log \beta)[1 + \frac{-\log\beta}{\nu} + o(\frac{1}{\nu})] . \end{aligned}\right\} \qquad (7.2.12)$$

We shall demonstrate shortly that the bounds α and β above also hold for the composite hypotheses

$$H_0': \underline{\mu} = \theta\underline{\mu}_1 + (1-\theta)\underline{\mu}_0, \quad \theta \le 0$$

$$H_1': \underline{\mu} = \theta\underline{\mu}_1 + (1-\theta)\underline{\mu}_0, \quad \theta \ge 1 .$$

7.3. The O.C. Function.

We first point out that, in this multivariate problem, the form of the O.C. Function, neglecting excess over the boundaries, is given by

$$L(\theta) = \Pr\{d_0 \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda, \underline{\mu}\}$$

$$\cong \frac{\bar{A}_m^h - 1}{\bar{A}_m^h - \bar{B}_m^h}, \ (h \neq 0) \qquad (7.3.1)$$

for the conditional test, given $\hat{\lambda}$ and λ; and for $\underline{\mu} = \theta\underline{\mu}_0 + (1-\theta)\underline{\mu}_1$, there exists $h = h(\underline{\mu}) = -2(\theta-\frac{1}{2})$ such that (7.3.1) is satisfied. We also remark that $L(\theta)$ is monotonic decreasing in θ, so that bounds α and β hold on error probabilities for H_0' and H_1' respectively, as stated earlier. So

$$\left.\begin{array}{l} \Pr\{d_1 \text{ from } T|\lambda, \underline{\mu}\} < \alpha \text{ for all } \theta \leq 0, \quad \underline{\mu} = \theta\underline{\mu}_0 + (1-\theta)\underline{\mu}_1 \\ \Pr\{d_0 \text{ from } T|\lambda, \underline{\mu}\} < \beta \text{ for all } \theta \quad 1, \quad \underline{\mu} = \theta\underline{\mu}_0 + (1-\theta)\underline{\mu}_1 \end{array}\right\} (7.3.2)$$

Successive upper and lower bounds on the O.C. function $\Pr\{d_0|\lambda,\underline{\mu}\}$ of test T follow as in Hall's paper. Taking expectations in (7.3.1),

$$\Pr\{d_0|\lambda, \mu\} \cong E\left[\frac{1 - \exp(-a_m h \hat{\lambda}/\lambda)}{1 - \exp\{-(a_m-b_m)h \hat{\lambda}/\lambda\}}\right] \text{ if } h > 0, \text{ i.e., if } \theta < \tfrac{1}{2}$$

$$= E[(1 - \exp(-a_m \hat{\lambda}/\lambda\, h)) \{\sum_{r=0}^{\infty} \exp(-r(a_m-b_m)\hat{\lambda}/\lambda\, h)\}]$$

$$= 1 - (1 + \frac{2a_m}{\nu} h)^{-\frac{1}{2}\nu} + (1 + \frac{2(a_m-b_m)}{\nu} h)^{-\frac{1}{2}\nu}$$

$$- (1 + \frac{2(2a_m-b_m)}{\nu} h)^{-\frac{1}{2}\nu} + (1 + \frac{2(2a_m-2b_m)}{\nu} h)^{-\frac{1}{2}\nu} - \ldots \qquad (7.3.3)$$

$$= 1 - (1-h+h\alpha^{-2/\nu})^{-\frac{1}{2}\nu} + (1-2h+h\alpha^{-2/\nu} + h\beta^{-2/\nu})^{-\frac{1}{2}\nu} - \ldots, \qquad (7.3.4)$$

using (7.2.11) and (7.2.12).

These results are notationally Hall's, with h and ν appropriate to the multivariate problem being considered (his equations (9)). If $h < 0$, Hall's equations (10) apply, and if $\beta = \alpha$, partial sums again give successive upper and lower bounds in the result (Hall's (11))

$$\Pr\{d_j|\lambda_1, \underline{\mu}\} = \sum_{r=0}^{\infty} (-1)^r \, [1 + r|h|\alpha^{-2/\nu} - 1)]^{-\frac{1}{2}\nu}, \qquad (7.3.5)$$

where $j = 0$ if $\theta < \frac{1}{2}$, and $j = 1$ if $\theta > \frac{1}{2}$.

Case $h = 0$. (i.e., $\theta = \frac{1}{2}$). For this case, $\underline{\mu} = \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_0)$. We evaluate in (7.3.1)

$$\lim_{h \to 0} \Pr\{d_0|\hat{\lambda}, \lambda, \underline{\mu}\} = \lim_{h \to 0} \frac{\exp(a_m h\, \hat{\lambda}/\lambda) - 1}{\exp(a_m h\, \hat{\lambda}/\lambda) - \exp(b_m h\, \hat{\lambda}/\lambda)}$$

$$= \lim_{h \to 0} \frac{a_m \exp(a_m h\, \hat{\lambda}/\lambda)}{a_m \exp(a_m h\, \hat{\lambda}/\lambda) - b_m \exp(b_m h\, \hat{\lambda}/\lambda)},$$

using L'Hôpital's rule,

$$= \frac{a_m}{a_m - b_m}, \qquad (7.3.6)$$

$$= \tfrac{1}{2} \quad \text{if} \quad \beta = \alpha.$$

These results are independent of $\hat{\lambda}$, and hold therefore for the O.C. function of T., i.e., for $\Pr\{d_0|\lambda, \underline{\mu}\}$.

Hall illustrates the rapid convergence of the series (7.3.5) when $\alpha = \beta = .05$, and $m = 16, 31$. His remarks apply here, as does the table, modified in terms of θ.

7.4. The A.S.N.

Again, we follow Hall's method. Under certain conditions, Wald's Equation

$$E\left(\sum_{i=1}^{N} Y_i\right) = E(Y_1)\, E(N)$$

holds for a sequence $Y_1, Y_2, \ldots$ of i.i.d. random variables. A simple extension gives the same result for multivariate variables, where each "observation" covers p characters.

LEMMA 7.1. a) $\underline{Z}_1, \underline{Z}_2, \ldots$ is a sequence of independent $p \times 1$ random vectors.

b) If $\underline{Z}_1' = [Z_{11} Z_{21} \ldots Z_{p1}]$, then $E|Z_{i1}| < \infty$, $i = 1, 2, \ldots, p$.

c) N is a positive integer-valued r.v. and

i) the event $\{N \le j\}$ and $\underline{Z}_k$ are independent if $j < k$,

ii) $E(N) < \infty$.

Then $E\left(\sum_{k=1}^{N} \underline{Z}_k\right) = E(N)\, E(\underline{Z}_1)$.

Following Johnson's method [11], the proof consists in applying the univariate result to each component of $\underline{Z}_1$.

Then from (7.2.6), where $\underline{Z}_n = \sum_{i=1}^{n} \underline{z}_i$,

$$\begin{aligned} E\, r_n(\lambda) &= (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}_0^{-1}\, E\underline{Z}_N \\ &\quad\ \ \overline{\lambda} \\ &= (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}^{-1}\, E(\underline{z}_1)\, E(N) \\ &= (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}^{-1}(\underline{\mu} - \tfrac{1}{2}(\underline{\mu}_1+\underline{\mu}_0))\, E(N) \\ &= -\tfrac{1}{2}h\, E(N) \cdot (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}^{-1}(\underline{\mu}_1-\underline{\mu}_0) \ . \end{aligned} \qquad (7.4.1)$$

Again, using conditional expectations, and following Hall's approach,

$$E\, r_N(\lambda) = E[E(r_N(\lambda)|\hat{\lambda})]$$

$$= E[\sum_{i=0}^{1} E\{r_N(\lambda)|\hat{\lambda},\, d_i\}\, Pr(d_i|\hat{\lambda})]$$

$$\simeq E[a_m\, \hat{\lambda}/\lambda\, Pr(d_1|\hat{\lambda}) + b_m\, \hat{\lambda}/\lambda\, Pr(d_0|\hat{\lambda})]\ ,$$

since $E[r_N(\lambda)|\hat{\lambda},d_1] \simeq \bar{a}_m = a_m\, \hat{\lambda}/\lambda$

and $E[r_N(\lambda)|\hat{\lambda},d_0] \simeq \bar{b}_m = b_m\, \hat{\lambda}/\lambda$,

neglecting excess over the boundaries. Hence if $h > 0$,

$$E(r_N|\lambda) \simeq a_m - (a_m - b_m)E[(\hat{\lambda}/\lambda)Pr(d_0|\hat{\lambda})]\ ,$$

$$\simeq a_m - (a_m - b_m)E[\hat{\lambda}/\lambda\, \{1 - \exp(-\, a_m\, \hat{\lambda}/\lambda\, h)\} \cdot$$

$$\cdot \{\sum_{r=0}^{\infty} \exp(-\, r(a_m - b_m)\, \hat{\lambda}/\lambda\, h)\}], \qquad (7.4.2)$$

using Wald's approximation

$$Pr(d_0|\hat{\lambda},\, \lambda,\, \underline{\mu}) \simeq \frac{\exp(a_m h\, \hat{\lambda}/\lambda) - 1}{\exp(a_m h\, \hat{\lambda}/\lambda) - \exp(b_m h\, \hat{\lambda}/\lambda)}\ .$$

Hence

$$E(r_N|\lambda) \simeq a_m - (a_m-b_m)[E\,\hat{\lambda}/\lambda - E\{\hat{\lambda}/\lambda \exp(-\,a_m\,\hat{\lambda}/\lambda\;h)\}$$

$$+ E\{\hat{\lambda}/\lambda \exp(-(a_m-b_m)\,\hat{\lambda}/\lambda\;h)\} - E\{\hat{\lambda}/\lambda \exp(-(2a_m-b_m)\,\hat{\lambda}/\lambda\;h)\}$$

$$+ E\{\hat{\lambda}/\lambda \exp(-\,2(a_m-b_m)\,\hat{\lambda}/\lambda\;h)\} - \ldots] \;,$$

and since $E[(\chi^2_\nu/\nu)\exp(-\,t\chi^2_\nu/\nu)] = (1+2t)^{-1-\frac{1}{2}\nu}$,

$$E(r_N|\lambda) \simeq a_m - (a_m-b_m)[\nu/\nu - (1 + 2\,\frac{a_m h}{\nu})^{-1-\frac{1}{2}\nu} + (1 + 2\,\frac{(a_m-b_m)h}{\nu})^{-1-\frac{1}{2}\nu}$$

$$- (1 + 2\,\frac{(2a_m-b_m)h}{\nu})^{-1-\frac{1}{2}\nu} + \ldots]$$

$$\simeq b_m + (a_m-b_m)[(1 + 2\,\frac{a_m h}{\nu})^{-1-\frac{1}{2}\nu} - (1 + 2\,\frac{(a_m-b_m)h}{\nu})^{-1-\frac{1}{2}\nu}$$

$$+ (1 + 2\,\frac{(2a_m-b_m)h}{\nu})^{-1-\frac{1}{2}\nu} - (1 + 4\,\frac{(a_m-b_m)h}{\nu})^{-1-\frac{1}{2}\nu} + \ldots] \,. \tag{7.4.3}$$

Combining with (7.3.1), we obtain for $h > 0$,

$$E(N)\cdot(\underline{\mu}_1-\underline{\mu}_0)'\underline{V}^{-1}(\underline{\mu}_1-\underline{\mu}_0) \simeq \frac{\nu}{h}(\beta^{-2/\nu} - 1) - \frac{\nu}{h}(\alpha^{-2/\nu} + \beta^{-2/\nu} - 2)\cdot$$

$$\cdot[(1 + h(\alpha^{-2/\nu} - 1))^{-1-\frac{1}{2}\nu}$$

$$- (1 + h(\alpha^{-2/\nu} + \beta^{-2/\nu} - 2))^{-1-\frac{1}{2}\nu}$$

$$+ (1 + h(2\alpha^{-2/\nu} + \beta^{-2/\nu} - 3))^{-1-\frac{1}{2}\nu} - \ldots] \tag{7.4.4}$$

which on the right is Hall's (16), and gives successive upper and lower bounds on the ASN of T. For $h < 0$, interchange a_m and b_m in (7.4.3) and put $-h$ for h; in (7.4.4) interchange α and β. With the same notation as in (7.3.5), $\alpha = \beta$ in (7.4.4) leads to

$$E(N) \cdot (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}^{-1}(\underline{\mu}_1-\underline{\mu}_0) \simeq \frac{\nu}{|h|}(\alpha^{-2/\nu} - 1)[1 - 2\sum_{r=1}^{\infty}\{1 + r|h|(\alpha^{-2/\nu} - 1)\}^{-1-\frac{1}{2}\nu}], \quad h \neq 0. \qquad (7.4.5)$$

These approximations involve the unknown λ. At $\underline{\mu} = \underline{\mu}_0$ and $\underline{\mu} = \underline{\mu}_1$, (i.e., $\theta = 0$ and $\theta = 1$), we can use bounds

$$\Pr\{d_1 \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda, \underline{\mu}_0\} < \frac{1}{\bar{A}_m} = \exp(-a_m\hat{\lambda}/\lambda)$$

$$\Pr\{d_0 \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda, \underline{\mu}_1\} < \bar{B}_m = \exp(b_m\hat{\lambda}/\lambda) .$$

Since

$$E\, r_N(\lambda) = b_m + (a_m-b_m)E[(\hat{\lambda}/\lambda)\Pr\{d_1 \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda, \underline{\mu}_0\}] ,$$

(7.4.1) gives for $h = 1$ at $\underline{\mu}_0$,

$$E(N) \cdot (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}^{-1}(\underline{\mu}_1-\underline{\mu}_0) > -2b_m - 2(a_m-b_m)E[(\hat{\lambda}/\lambda)\exp(-a_m\hat{\lambda}/\lambda)]$$

$$\text{i.e.,} \qquad > \nu[\beta^{-2/\nu} - 1 - \alpha(1 + \beta^{-2/\nu}\alpha^{2/\nu} - 2\alpha^{2/\nu})] \qquad (7.4.6)$$

$$\text{and} \qquad > \nu[\alpha^{-2/\nu} - 1 - 2\alpha(1 - \alpha^{2/\nu})] \quad \text{if} \quad \beta = \alpha. \qquad (7.4.7)$$

If $\underline{\mu} = \underline{\mu}_1$, then $\theta = 1$ and $h = -1$; then we get (7.4.6) with α and β interchanged.

There remains the special case $\underline{\mu} = \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_0)$, where $h = 0$. We consider

$$E\{[r_N(\lambda)]^2\} = E\,E\{[r_N(\lambda)]^2|N = n\}$$

$$= (1/\lambda^2)E\,E(\underline{Z}_n'\,\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)\,|N = n),$$

where $E\,\underline{Z}_n = \underline{0}$ if $\underline{\mu} = \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_0)$. Now

$$\{\underline{Z}_n'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)\}^2 = \sum_{i=1}^{n}\{\underline{z}_i'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)\}^2 + \sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\underline{z}_i'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)\cdot$$

$$\cdot\,\underline{z}_j'\,\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0).$$

Given $N = n$, the second sum has zero expectation, by the independence of $\underline{z}_i$ and $\underline{z}_j$, and since $h = 0$ implies $E\underline{z}_i = \underline{0}$. Also, for given n, the distribution of $\underline{z}_i'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)$ is $N(0,\ \lambda(\underline{\mu}_1-\underline{\mu}_0)'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0))$. Hence

$$\sum_{i=1}^{n}\{\underline{z}_i'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)\}^2 \sim \lambda(\underline{\mu}_1-\underline{\mu}_0)'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)\chi_n^2$$

with expected value $n\lambda(\underline{\mu}_1-\underline{\mu}_0)'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0)$. Hence

$$E[\{r_N(\lambda)\}^2|h = 0] = E(N|h = 0)(1/\lambda)(\underline{\mu}_1-\underline{\mu}_0)'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0). \qquad (7.4.8)$$

Again,

$$E[\{r_N(\lambda)\}^2|h = 0] = E[E\{r_N(\lambda)\}^2|h = 0, \hat{\lambda}]$$

$$= E[\sum_{j=0}^{1} E[\{r_N(\lambda)\}^2|\hat{\lambda}, d_j]\Pr(d_j \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda}, \lambda)]$$

if $h = 0$; and since

$$E[r_N^2(\lambda)|\hat{\lambda}, d_0] \simeq \bar{b}_m^2 = (b_m \hat{\lambda}/\lambda)^2,$$

$$E[r_N^2(\lambda)|\hat{\lambda}, d_1] \simeq \bar{a}_m^2 = (a_m \hat{\lambda}/\lambda)^2,$$

$$\simeq a_m^2 E(\hat{\lambda}^2/\lambda^2) - (a_m^2 - b_m^2)E[\hat{\lambda}^2/\lambda^2 \Pr(d_0 \text{ from } T(\hat{\lambda},\lambda)|\hat{\lambda},\lambda)]$$

$$= a_m^2 E(\chi_\nu^4/\nu^2) - (a_m^2 - b_m^2)\frac{a_m}{a_m - b_m} E(\chi_\nu^4/\nu^2),$$

using (7.3.6) at $h = 0$; so that

$$E[\{r_N(\lambda)\}^2|h = 0] \simeq -a_m b_m(1 + 2/\nu). \tag{7.4.9}$$

Hence

$$E(N|h = 0) \cdot (\underline{\mu}_1-\underline{\mu}_0)'\underline{V}_0^{-1}(\underline{\mu}_1-\underline{\mu}_0) \simeq -\tfrac{1}{4}\lambda\nu^2(\alpha^{-2/\nu}- 1)(\beta^{-2/\nu}- 1)(1 + 2/\nu), \tag{7.4.10}$$

and if $\beta = \alpha$,

$$\simeq -\tfrac{1}{4}\lambda\nu^2(\alpha^{-2/\nu} - 1)^2 (1 + 2/\nu). \tag{7.4.11}$$

7.5. Suggestions for Further Research.

We return to the univariate problem of Hall's paper, where the information arising from the statistic $\bar{x}_m = \frac{1}{m}\sum_{i=1}^{m} x_i$ at the m-th observation is not fully taken into consideration.

Suppose that the PSPRT procedure of Chapter III (in which n is here replaced by m) is modified when the variance is unknown, so that σ^2 is replaced by s^2. The test procedure T_m is now as follows:

$$\text{if} \quad r_n(s) = \Delta \sum_{i=1}^{n} (x_i - \tfrac{1}{2}\Delta)/s^2 , \quad n \geq m , \tag{7.5.1}$$

stop and accept H_0 if $r_n(s) \leq b_m$

stop and accept H_1 if $r_n(s) \geq a_m$

continue sampling if $b_m < r_n(s) < a_m$.

Let $T_m(s^2, \bar{x}_m)$ be the conditional test given the first m observations. This will depend on $(x_1, x_2, \ldots, x_m)$ only through the sufficient statistics $\bar{x}_m$ and s^2. Then (cf. (3.1.1), (3.1.7), (3.3.3)) $T_m(s^2, \bar{x}_m)$ is as follows:

$$\text{if} \quad \left.\begin{aligned} a'_m &= a_m - \frac{m\Delta\bar{x}_m}{s^2} + \frac{m\Delta^2}{s^2} \\ b'_m &= b_m - \frac{m\Delta\bar{x}_m}{s^2} + \frac{m\Delta^2}{s^2} \end{aligned}\right\} \tag{7.5.2}$$

$$r^*_n(s) = r_n(s) - r_m(s), \tag{7.5.3}$$

stop and accept H_0 if $r^*_n(s) \leq b'_m$, stop and accept H_1 if $r_n^*(s) \geq a'_m$,

continue sampling if $b'_m < r_n^*(s) < a'_m$.

If the inequality

$$b'_m < 0 < a'_m$$

does not hold then we decide on the m observations of the first stage. If $a'_m < 0$, H_1 is accepted; if $b'_m > 0$, H_0 is accepted.

This analog of the PSPRT seems appropriate, since the probability, given σ^2, that observations terminate with the first stage, can be exactly determined, and we have thus more information on the O.C. function and ASN.

Let $T_m(s^2, \bar{x}_m, \sigma^2)$ be the SPRT based upon the first stage, with known σ, defined by:

$$\text{if} \quad \left.\begin{aligned} \bar{a}'_m &= a'_m\, s^2/\sigma^2 \\ \bar{b}'_m &= b'_m\, s^2/\sigma^2 \end{aligned}\right\} \qquad (7.5.4)$$

stop and accept H_0 if $r_n^*(\sigma) \leq \bar{b}'_m$

stop and accept H_1 if $r_n^*(\sigma) \geq \bar{a}'_m$.

If $\bar{b}'_m < 0 < \bar{a}'_m$ does not hold, stop and decide at the end of the first stage as in $T_m(s^2, \bar{x}_m)$; otherwise continue sampling.

It can be noted that $T_m(s^2, \bar{x}_m, \sigma^2)$ is the same as $T(\underline{x}_m)$ of 3.3., with boundaries $a_m s^2/\sigma^2$ and $b_m s^2/\sigma^2$ for the unconditional test.

Further, since $r_n^*(\sigma) = r_n^*(s) \cdot s^2/\sigma^2$, $T_m(s^2, \bar{x}_m, \sigma^2)$ has the same decision rule at each stage as does $T_m(s^2, \bar{x}_m)$.

The O.C. function of T_m is given by

$$L(\mu|\sigma) = E\, E \Pr\{\text{accept } H_0 \text{ from } T_m(s^2, \bar{x}_m, \sigma^2)|s^2, \bar{x}_m, \sigma^2, \mu\}\,, \tag{7.5.5}$$

where expectations are taken w.r.t. the distributions of the independent statistics $\bar{x}_m$ and s^2.

I.e., $L(\mu|\sigma) = E\, E\, L(\mu|s^2, \bar{x}_m, \sigma)$, say, where

$$L(\mu|s^2, \bar{x}_m, \sigma) \simeq \begin{cases} 1 & \text{if } r_m^*(\sigma) \le \bar{b}'_m \\ \dfrac{1 - \exp(h\bar{a}'_m)}{\exp(h\bar{b}'_m) - \exp(h\bar{a}'_m)} & \text{if } \bar{b}_m < r_m^*(\sigma) < \bar{a}'_m\,. \\ 0 & \text{otherwise} \end{cases} \tag{7.5.6}$$

Then in (3.3.7), we replace a by as^2/σ^2, b by bs^2/σ^2, and n by m/σ^2 to get

$$L(\mu|\sigma) = E\Big[\frac{1}{\exp(hbs^2/\sigma^2) - \exp(has^2/\sigma^2)}\Big\{\exp(hbs^2/\sigma^2)\Phi\Big(\frac{bs^2}{\sqrt{m}\sigma\Delta} + \frac{\sqrt{m}h\Delta}{2\sigma}\Big)$$
$$- \exp(has^2/\sigma^2)\Phi\Big(\frac{as^2}{\sqrt{m}\sigma\Delta} + \frac{\sqrt{m}h\Delta}{2\sigma}\Big) + \Phi\Big(\frac{as^2}{\sqrt{m}\sigma\Delta} - \frac{\sqrt{m}h\Delta}{2\sigma}\Big)$$
$$- \Phi\Big(\frac{bs^2}{\sqrt{m}\sigma\Delta} - \frac{\sqrt{m}h\Delta}{2\sigma}\Big)\Big\}\Big]\,, \tag{7.5.7}$$

where expectations are taken over the distribution of s^2.

In this form (7.5.7) is unlikely to lead to an infinite series of the form of Hall's equations (9), but further research on this approach using bounds on the normal c.d.f. might yield results extending Hall's procedure to the case in which the test terminates with the first stage. The same approach may well be applied to the ASN.

APPENDIX I

LEMMA A.1. Let $\pi_1, \pi_2, \ldots$ be a sequence of approximations to the root π^* of the equation

$$\phi(\pi) = \rho^{-1} - 2 \log \frac{1-\pi}{\pi} - \frac{1-2\pi}{\pi(1-\pi)} = 0,$$

such that

$$\pi_{n+1} = \pi_n - \frac{\phi(\pi_n)}{\phi'(\pi_n)} .$$

If, for some $k \geq 1$,

$$\pi_k = \rho - \rho^2(1 + 2 \log \rho) + O(\rho^3 \log \rho) > 0$$

where ρ is small, then for every $n > k$, π_n has the same form as π_k.

PROOF: [$\pi_1, \pi_2, \ldots$ are obtained by the Newton-Raphson process of iteration. The condition $\pi_k > 0$ follows from monotonicity of ϕ from $-\infty$ to $+\infty$ for π in $(0,1)$, and as shown in 2.2, if $\pi_2 < 0$, a new first approximation π_1 would be chosen. If $\pi_2 > 0$, however, rapid convergence follows (see Fig. 2.2.2).]

By induction, assume true for π_n.

Then

$$\pi_{n+1} = \pi_n - \{\pi_n^2(1-\pi_n)^2\}[\frac{1}{\rho} - 2\log\frac{1-\pi_n}{\pi_n} - \frac{1-2\pi_n}{\pi_n(1-\pi_n)}]$$

$$= \rho - \rho^2(1+2\log\rho) - \{\rho - \rho^2(1+2\log\rho)\}^2\{1 - \rho + \rho^2(1+2\log\rho)\}^2$$

$$\cdot [\frac{1}{\rho} - 2\log(\frac{1-\rho+\rho^2(1+2\log\rho)}{\rho-\rho^2(1+2\log\rho)}) - \frac{1 - 2\rho + 2\rho^2(1+2\log\rho)}{\{\rho-\rho^2(1+2\log\rho)\}\{1-\rho+\rho^2(1+2\log\rho)\}}]$$

$$+ 0(\rho^3 \log \rho)$$

$$= \rho - \rho^2(1 + 2\log\rho) - \{\rho^2 + 0(\rho^3)\}\{1 - 2\rho + 0(\rho^2)\}$$

$$\cdot [\frac{\rho-\rho^2(1+2\log\rho)-\rho^2-\rho+2\rho^2+0(\rho^3)}{\rho^2 + 0(\rho^3)} - 2\log\{(\frac{1-\rho}{\rho}) \cdot \frac{1+\frac{\rho^2}{1-\rho}(1+2\log\rho)}{1-\rho(1+2\log\rho)}\}]$$

$$+ 0(\rho^3 \log \rho)$$

$$= \rho - \rho^2(1+2\log\rho) - \rho^2(-2\log\rho+2\log\rho+2\rho-2\rho+0(\rho)) + 0(\rho^3\log\rho)$$

$$= \rho - \rho^2(1 + 2\log\rho) + 0(\rho^3 \log \rho) ,$$

which is of the same form.

Q.E.D.

APPENDIX II

ON CHOOSING A SUITABLE ESTIMATOR OF COST

The choice of a PSPRT procedure in Chapter V is based largely on mathematical convenience, as is the choice of $\bar{c}_n$ as a suitable estimator. Research into the choice of a "best" estimator, in some sense, has not been done, to the writer's knowledge. The following approach may be useful.

We consider a class of test procedures between two simple hypotheses H_0 and H_1 about the parameter θ, where the i.i.d. sequence of r.v's $X_1, X_2, \ldots$ is indexed by θ. To distinguish the present problem from that of the tests based on $x_1, x_2, \ldots,$ we refer to "cost - loss" and "cost-risk'.

The cost-loss refers to incorrect use of an estimator $\tilde{c}_n$ of the cost per observation c, where $\tilde{c}_n$ is a function of the observed cost on the first n individuals, i.e., $\tilde{c}_n$ is a function of

$c_1, c_2, \ldots, c_n$.

Using the notation of 5.2, let $R(w, n, c|c)$ be the a.m. risk in a class of test procedures, each procedure being well-defined when w, n and c are given. Define the cost loss to be

$$L(\tilde{c}_n|c) = R(w, n, \tilde{c}_n|c) - R(w, n, c|c) . \qquad (A.2.1)$$

Thus the cost-loss is the excess, as for example, in (5.3.8). Ideally, the estimator $\tilde{c}_n$ should be chosen to minimize $EL(\tilde{c}_n|c)$, averaged over the distribution of $\tilde{c}_n$; but c is unknown.

However, if we can postulate a prior distribution ξ for c, then we try to choose $\tilde{c}_n$ to minimize the cost-risk

$$r^{(\xi)}(\tilde{c}_n) = \int \int L(\tilde{c}_n|c) \, dF_\theta(\tilde{c}_n) \, d\xi(c)$$

where the integrals are assumed to exist, and the conditions for Fubini's theorem are assumed to apply. Then, effectively, we try to minimize the posterior cost-risk of c over the posterior distribution of c, given $\tilde{c}_n$. This may lead to a Bayes estimator $\tilde{c}_n$ of the true cost.

BIBLIOGRAPHY

[1] Anderson, T.W., "A modification of the sequential probability ratio test to reduce the sample size", Annals of Mathematical Statistics, Volume 31 (1960), pp 165-197.

[2] Armitage, P., "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis", Journal of the Royal Statistical Society, Series B, Volume 12 (1950), pp 137-144.

[3] Baker, A.G., "Properties of some tests in sequential analysis", Biometrika, Volume 37 (1950), pp 334-346.

[4] Bhate, D.H., "Approximations to the distribution of sample size for sequential tests. I. Tests of Simple hypotheses", Biometrika, Volume 46 (1959), pp 130-138.

[5] Billard, L., "Sequential tests for two-sided alternative hypothesis", University of New South Wales, Ph.D. Thesis, 1968.

[6] Burkholder, D.L. and Wijsman, R.A., "Optimum Properties and Admissibility of Sequential Tests", Annals of Mathematical Statistics, Volume 34 (1963), pp 1-17.

[7] Chernoff, H. "Sequential design of experiments", Annals of Mathematical Statistics, Volume 30 (1959), pp 755-770.

[8] Feller, W., An Introduction to Probability Theory and Its Applications, Volume I, Second Edition, John Wiley and Sons, Inc., New York, 1958.

[9] Hall, W.J., "Some sequential analogs of Stein's two-stage test", Biometrika, Volume 49 (1962), pp 367-378.

[10] Jackson, J.E. and Bradley, R.A., "Sequential χ^2- and T^2 tests", Annals of Mathematical Statistics, Volume 32 (1961), pp 1063-1077.

[11] Johnson, N.L., "A proof of Wald's theorem on cumulative sums", Annals of Mathematical Statistics, Volume 30 (1959), pp 1245-1247.

[12] Johnson, N.L., "Sequential analysis: a survey", Journal of the Royal Statistical Society, Series A, Volume 124 (1961), pp 372-411.

[13] Kemp, K.W., "Formulae for calculating the operating characteristic and the average sample number of some sequential tests", Journal of the Royal Statistical Society, Series B, Volume 20 (1958), pp 379-386.

[14] Lehmann, E.L., Testing Statistical Hypotheses, John Wiley and Sons, Inc., New York, 1959.

[15] Page, E.S., "An improvement to Wald's approximation for some properties of sequential tests", Journal of the Royal Statistical Society, Series B, Volume 16 (1954), pp 136-139.

[16] Sobel, M. and Wald, A., "A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution",

Annals of Mathematical Statistics, Volume 20 (1949), pp 502-522.

[17] Stein, C., "A two-sample test for a linear hypothesis whose power is independent of the variance", Annals of Mathematical Statistics, Volume 16 (1945), pp 243-258.

[18] Wald, A., "On cumulative sums of random variables", Annals of Mathematical Statistics, Volume 15 (1944), pp 283-296.

[19] Wald, A., "Sequential tests of statistical hypotheses", Annals of Mathematical Statistics, Volume 16 (1945), pp 117-186.

[20] Wald, A., "Sequential method of sampling for deciding between two courses of action", Journal of the American Statistical Association, Volume 40 (1945), pp 277-306.

[21] Wald, A., "Some generalizations of the theory of cumulative sums of random variables", Annals of Mathematical Statistics, Volume 16 (1945), pp 287-293.

[22] Wald, A., "Some improvements in setting limits for the expected number of observations required by a sequential probability ratio test", Annals of Mathematical Statistics, Volume 17 (1946), pp 466-474.

[23] Wald, A., "Differentiation under the expectation sign in the fundamental identity of sequential analysis", Annals of Mathematical Statistics, Volume 17 (1946), pp 493-497.

[24] Wald, A., Sequential Analysis, John Wiley and Sons, Inc., New York, 1947.

[25] Wald, A. and Wolfowitz, J., "Optimum character of the sequential probability ratio test", Annals of Mathematical Statistics, Volume 19 (1948), pp 326-339.

[26] Wald, A. and Wolfowitz, J., "Bayes solutions of sequential decision problems", Annals of Mathematical Statistics, Volume 21 (1950), pp 89-99.