

COMPARISON OF TWO DRUGS BY MULTIPLE STAGE SAMPLING  
USING BAYESIAN DECISION THEORY

by

Armand V. Smith, Jr., B.S. Ch.E., M.S.

Thesis submitted to the Graduate Faculty of the  
Virginia Polytechnic Institute  
in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

May 1963

Blacksburg, Virginia

LD  
5655  
V856  
1963  
5647  
C.2

TABLE OF CONTENTS

Chapter		Page
I	INTRODUCTION AND LITERATURE SURVEY	5
II	EXTENSIVE-FORM ANALYSIS	12
	2.1 General Procedure .....	12
	2.2 Some Distribution Theory .....	14
	2.3 Details of the Analysis for a Constant Loss Function .....	16
	2.4 Analysis for a Linear Loss Function ....	19
	2.5 An Example of the Computations .....	22
	2.6 Numerical Results .....	27
III	THE STAGE-BY-STAGE PROCEDURE	34
	3.1 General Procedure .....	34
	3.2 An Example of the Calculations .....	36
	3.3 Numerical Results .....	38
IV	AN APPROXIMATE PROCEDURE	41
	4.1 General Procedure .....	41
	4.2 Some Numerical Results .....	42
V	STAGE-BY-STAGE PROCEDURE ASSUMING NORMAL DISTRIBUTIONS	44
	5.1 Distribution Theory .....	44
	5.2 Procedure for a Linear Loss Function ...	46

TABLE OF CONTENTS (Cont'd)

Chapter		Page
	5.3 Procedure for a Constant Loss Function ..	51
VI	THEORETICAL COMPARISONS OF EXTENSIVE-FORM ANALYSIS, THE STAGE-BY-STAGE PROCEDURE AND THE APPROXIMATE PROCEDURE	62
	6.1 Asymptotic Comparison of the Stage-by-Stage Procedure and Extensive-Form Analysis .....	62
	6.2 Derivation of the Approximate Procedure as an Approximation to the Stage-by-Stage Procedure Assuming Normal Distribution .....	65
VII	NUMERICAL STUDIES OF EXTENSIVE-FORM ANALYSIS, THE STAGE-BY-STAGE PROCEDURE, AND THE APPROXIMATE PROCEDURE	68
VIII	STOPPING RULES FOR THE STAGE-BY-STAGE AND APPROXIMATE PROCEDURES	79
	8.1 Stopping Rules for the Stage-by-Stage Procedure .....	79
	8.2 A Stopping Rule for the Approximate Procedure .....	82

TABLE OF CONTENTS (Cont'd)

Chapter	Page
8.3 Results of Simulation Studies .....	91
IX GENERAL DISCUSSION OF THE RESULTS AND RECOMMENDATIONS	95
X SUGGESTIONS FOR FUTURE RESEARCH	101
XI SUMMARY	104
ACKNOWLEDGMENTS	112
BIBLIOGRAPHY	113
VITA	115

## I. INTRODUCTION AND LITERATURE SURVEY

It is common practice in comparing the parameters of two binomial populations (e.g., the probabilities of success for two drugs) to choose a fixed number of observations and allocate half of these to samples from each of the two populations.

Beginning during World War II Wald (1947) developed the general technique of sequential analysis which makes the total number of observations a random variable depending on the progress of the experiment. This method often reduces the amount of data required to reach statistically valid conclusions. The problem of comparing two treatments of drugs in medical experiments was put into a most convenient form from the user's point of view in a paper by Bross (1952).

More recently Taylor and David (1962) studied a problem which is similar in spirit although quite different in approach to the one considered in this thesis. Their aim is to allocate observations at each stage of a multi-stage experiment involving several drugs on the basis of results in previous stages. In order to do this they use weighting functions to divide up the fixed number of observations at each stage in such a way that the drug showing the highest number of favorable responses is allocated the highest proportion of observations (i.e., patients). The probability of correctly

selecting the best drug by this procedure was found, mainly by simulation, to be generally greater than that for equal allocation of all drugs. It will be noted that the total number of observations in the experiment is pre-determined but that this number is divided up between the various drugs in a manner which tries to take advantage of knowledge gained during the experiment.

The general problem considered in this thesis is that of determining an optimum strategy for deciding how to allocate the observations in each stage of a multi-stage experimental procedure between two binomial populations on the basis of the results of the previous stages. After all of the stages of this experiment have been performed, one must decide which of the two populations has the higher probability of success. The optimum strategy is to be optimum relative to a given loss function and a given prior distribution for the probabilities of success of the two populations. At first it is assumed that there are a fixed number of stages in the experiment, but later in the thesis this restriction is weakened to the assumption that only the maximum number of stages possible in the experiment is fixed and the experiment can be stopped at any stage before the last possible stage is reached. In any case the total number of observations in each of the stages is fixed before the experiment.

In addition, the same general problem is considered

when the two populations are normal and the terminal decision is that one or the other of the two populations has the higher mean.

Since it turns out that the exact procedure for finding the optimum strategy is impractical when the number of stages or the number of observations in each of the stages is at all large, two alternative approximate procedures are presented and compared with the exact procedure.

Throughout this thesis it will be assumed that the two binomial populations are the respective outcomes when two drugs are administered to patients in a multi-stage testing procedure and that the observations in each stage are the numbers of successes of given drugs administered to given patients. However, it should be emphasized that the procedures developed and studied in this thesis have much wider applicability than the comparison of two drugs. For example, one might want to compare the output of two machines or processes in a multi-stage testing procedure.

At this point we may mention a somewhat similar problem, the "Two-Armed Bandit Problem." Here the aim is to allocate  $n$  observations, one at a time, between two binomial populations so as to maximize the expected number of successes. Contributors to this subject include Vogel and Robbins as well as Bradt, Johnson, and Karlin. To obtain results they impose various additional conditions which we now outline.



Vogel (1960a) considers the following problem:

Let the random variables  $X$  and  $Y$ , where

$$\Pr(X = 1) = 1 - \Pr(X = 0) = p$$

and

$$\Pr(Y = 1) = 1 - \Pr(Y = 0) = q \quad ,$$

describe the outcomes of two experiments,  $E_1$  and  $E_2$ . A class of strategies is studied which results from a sequential procedure of  $n$  steps, in which either  $E_1$  or  $E_2$  (but not both) is performed in each stage. However, this paper has the additional restriction that in the first  $2k$  steps ( $k$  is a random variable) each of  $E_1$  and  $E_2$  is performed  $k$  times. Then the rest of the  $n - 2k$  steps are made either with  $E_1$  or  $E_2$ . A loss function based on the expected sum for all  $n$  steps is used.

In another paper (1960b) Vogel derives some asymptotic results for the Two-Armed Bandit Problem both with and without the additional restriction of his first paper (1960a).

Robbins (1956) considers the problem of successively choosing one of two ways of action, each of which may lead to success or failure, in such a way as to maximize the long-run proportion of successes obtained, the choice each time being based on a fixed number of the previous trials.

In another paper Robbins (1952) proposes and discusses in general several problems in the sequential design of experiments. Among these is the problem of how a sample

should be drawn from two populations in order to achieve the greatest possible expected value of the sum of the sample results.

Bradt, Johnson, and Karlin (1956) derive some properties of the Two-Armed Bandit Problem. In addition, they consider the generalized Two-Armed Bandit Problem of maximizing the expected number of successes in  $n$  trials when at each trial one is free to choose between two binomial random variables, whose probabilities of success,  $p$  and  $q$ , are unknown but have a prior distribution  $F(p, q)$ . Finally, they consider in detail the situation in which one of the probabilities of success is unknown but a prior distribution for it is specified and the other probability of success is known.

Dunnett, Samuel, and Chernoff have also worked on problems similar to the one in this thesis.

Dunnett (1960) uses decision theory methods to decide if a potential drug, which is assumed to have two levels of activity, active and inactive, is to be accepted as being worthy of further experimentation or rejected. A sequential procedure is derived in which rejection can occur at any stage. A method for computing critical rejection levels in each stage when the testing errors are normally distributed is given in detail for one, two, and three stage problems.

Samuel (1961) describes a minimal complete class of decision rules for sequentially classifying individuals of a

group which is known to have come from one of two completely specified populations.

Chernoff (1959) presents a procedure for the sequential design of experiments where the problem is one of testing a hypothesis. It is assumed that there are two possible terminal decisions and a class of available experiments. After each observation the statistician decides whether to continue experimentation or not. If he decides to stop, he must select one of the two terminal decisions. The choice of an experiment at any stage is based on Kullback-Leibler information numbers. This procedure is worked out in detail for the case in which each of the experiments yield Bernoulli random variables.

The two most important references for the body of this thesis are Raiffa and Schlaifer's Applied Decision Theory and a paper by Amster.

Raiffa and Schlaifer (1961) have a general discussion of Bayesian decision theory in the first part of their book. Then they give detailed specific analytical solutions for some problems in which there are one stage, two (or more) possible terminal decisions, normal or binomial distributions, and utility functions (They use utility functions instead of loss functions.) which are linear in the population means. In this thesis many of the concepts and results and much of the terminology and notation developed in their book will be used.

Amster (1962) proposes and derives some properties of a stopping rule for sequential sampling which weighs the cost of additional observations against the expected gain to be derived from additional sampling. This rule requires one more observation to be taken as long as the posterior risk is larger than the expected posterior risk for any additional fixed size sample. He shows how this stopping rule can be applied to both estimation and the testing of two simple hypotheses.

## II. EXTENSIVE-FORM ANALYSIS

### 2.1 General Procedure

Suppose one desires to compare two drugs, Drug 1 and Drug 2, in an  $m$ -stage test procedure. Let the probability that Drug 1 results in an improvement in a patient in any of the stages be  $\theta_1$ , and let the corresponding probability for Drug 2 be  $\theta_2$ . Suppose the experimenter has decided prior to the experiment that he will take  $n_k$  observations in the  $k$ 'th stage ( $k = 1, 2, \dots, m$ ). However, he wants to partition  $n_k$  into  $n_{1k}$  and  $n_{2k}$ , the number of observations for Drug 1 and for Drug 2, respectively, in the  $k$ 'th stage on the basis of the results of the observations in the previous  $k-1$  stages. Let the numbers of successes (improvement of patients) of Drug 1 and of Drug 2 in the  $k$ 'th stage be  $r_{1k}$  and  $r_{2k}$ , respectively. After the  $m$ 'th stage, on the basis of his results for all  $m$  stages, he desires to state that either  $\theta_1 > \theta_2$  or that  $\theta_1 < \theta_2$ .

Let  $\theta = (\theta_1, \theta_2)$  and let  $x_k = (r_{1k}, r_{2k})$ . Let  $a_k$  be a decision of what size  $n_{1k}$  (and thus  $n_{2k} = n_k - n_{1k}$ ) should be. (" $a_k$ " will be  $0, 1, 2, \dots, \text{or } n_k$ .) Let  $b$  be a decision (the terminal decision) whether to state that  $\theta_1 > \theta_2$  or  $\theta_1 < \theta_2$ , where  $b = 1$  will represent the decision that  $\theta_1 > \theta_2$ , and  $b = 2$  will represent the decision that  $\theta_1 < \theta_2$ .

Define the following "loss" function on  $x_1, \dots, x_m, a_1, \dots, a_m, b, \theta$ :

$$(2.1) \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) = \begin{cases} 0 & \text{if } b = 1 \text{ and } \theta_1 > \theta_2 \\ 0 & \text{if } b = 2 \text{ and } \theta_1 < \theta_2 \\ q_1 & \text{if } b = 1 \text{ and } \theta_1 < \theta_2 \\ q_2 & \text{if } b = 2 \text{ and } \theta_1 > \theta_2 \end{cases} ,$$

where  $q_1$  and  $q_2$  are positive constants. In what follows (2.1) will be called the "constant loss function."

Notice that  $L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta)$  does not directly depend on  $a_1, \dots, a_m$  since it will be assumed that the costs of making observations with each of the two drugs are equal.

The problem now is first to select  $a_1$ , then  $a_2$  on the basis of  $x_1$  and  $a_1$ , then  $a_3$  on the basis of  $a_1, a_2, x_1$ , and  $x_2$ , etc., and finally  $b$  on the basis of  $a_1, \dots, a_m, x_1, \dots, x_m$ . In order for this to be done, it is proposed that an extended and modified form of the "analysis in extensive form" as presented by Raiffa and Schlaifer (1961) be used.

In general, the analysis will proceed as follows:

- (1) For fixed values of  $x_1, \dots, x_m, a_1, \dots, a_m$ , and  $b$  compute  $E_{\theta|x_1, \dots, x_m} L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta)$ .
- (2) For fixed  $x_1, \dots, x_m, a_1, \dots, a_m$  compute

$$L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \\ = \min_b E_{\theta | x_1, \dots, x_m} L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta)$$

and choose the "b" which gives the minimum.

(3) For fixed values of  $x_1, \dots, x_{m-1}$  and  $a_1, \dots, a_m$  compute  $E_{x_m | x_1, \dots, x_{m-1}} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m)$ .

(4) Then for the fixed values of  $x_1, \dots, x_{m-1}$ , and  $a_1, \dots, a_{m-1}$  compute

$$L^{(2)}(x_1, \dots, x_{m-1}, a_1, \dots, a_{m-1}) \\ = \min_{a_m} E_{x_m | x_1, \dots, x_{m-1}} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m)$$

and choose the " $a_m$ " which gives the minimum.

(5) Continue in this manner with each stage until the first stage.

(6) On the first stage choose  $a_1$  such that

$$E_{x_1} [L^{(m)}(x_1, a_1)] \text{ is a minimum.}$$

## 2.2 Some Distribution Theory

In order to evaluate the expressions in Section 2.1 we need a weighting function, or prior distribution, of  $\theta$ . Let

$$(2.2) \quad \rho(\theta) = \frac{\theta_1^{\alpha_0} (1 - \theta_1)^{\beta_0} \theta_2^{\gamma_0} (1 - \theta_2)^{\delta_0}}{B(\alpha_0 + 1, \beta_0 + 1) B(\gamma_0 + 1, \delta_0 + 1)},$$

where  $\alpha_0, \beta_0, \gamma_0,$  and  $\delta_0$  are non-negative integers,  $B(\alpha_0+1, \beta_0+1)$  and  $B(\gamma_0+1, \delta_0+1)$  are beta functions,  $0 \leq \theta_1 \leq 1,$  and  $0 \leq \theta_2 \leq 1.$  Since each  $r_{1k}$  and  $r_{2k}$  has a binomial distribution,

$$(2.3) \quad P_{\theta}(x_1, \dots, x_k) = \prod_{i=1}^k \binom{n_{1k}}{r_{1k}} \binom{n_{2i}}{r_{2i}} \\ \cdot \theta_1^{\alpha_k - \alpha_0} (1 - \theta_1)^{\beta_k - \beta_0} \theta_2^{\gamma_k - \gamma_0} (1 - \theta_2)^{\delta_k - \delta_0},$$

where  $\alpha_k = \sum_{i=1}^k r_{1i} + \alpha_0,$

$$\beta_k = \sum_{i=1}^k (n_{1i} - r_{1i}) + \beta_0,$$

$$\gamma_k = \sum_{i=1}^k r_{2i} + \gamma_0,$$

and  $\delta_k = \sum_{i=1}^k (n_{2i} - r_{2i}) + \delta_0,$

for  $k = 1, 2, \dots, m.$  Then

$$(2.4) \quad P(\theta | x_1, \dots, x_m) = \frac{\rho(\theta) P_{\theta}(x_1, \dots, x_m)}{\int_0^1 \int_0^1 \rho(\theta) P_{\theta}(x_1, \dots, x_m) d\theta_1 d\theta_2} \\ = \frac{\theta_1^{\alpha_m} (1 - \theta_1)^{\beta_m} \theta_2^{\gamma_m} (1 - \theta_2)^{\delta_m}}{B(\alpha_m + 1, \beta_m + 1) B(\gamma_m + 1, \delta_m + 1)}$$

Also, let



$$P(x_1, \dots, x_k) = \int_0^1 \int_0^1 \rho(\theta) P_\theta(x_1, \dots, x_k) d\theta_1 d\theta_2 .$$

Then

$$(2.5) \quad P(x_k | x_1, \dots, x_{k-1}) = \frac{P(x_1, \dots, x_{k-1}, x_k)}{P(x_1, \dots, x_{k-1})}$$

$$= \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}} \frac{B(\alpha_k+1, \beta_k+1) B(\gamma_k+1, \delta_k+1)}{B(\alpha_{k-1}+1, \beta_{k-1}+1) B(\gamma_{k-1}+1, \delta_{k-1}+1)}$$

for  $k = 2, 3, \dots, m$  and

$$P(x_1) = \binom{n_{11}}{r_{11}} \binom{n_{21}}{r_{21}} \frac{B(\alpha_1+1, \beta_1+1) B(\gamma_1+1, \delta_1+1)}{B(\alpha_0+1, \beta_0+1) B(\gamma_0+1, \delta_0+1)} .$$

### 2.3 Details of the Analysis for a Constant Loss Function

Following the outline of the analysis in Section 2.1 and using (2.4) and (2.5), we obtain:

(1) If  $b = 2$ ,

$$E_\theta | x_1, \dots, x_m L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta)$$

$$= \int_0^1 \int_0^{\theta_1} q_2 P(\theta | x_1, \dots, x_m) d\theta_2 d\theta_1$$

$$= q_2 \int_0^1 \int_0^{\theta_1} \frac{\theta_1^{\alpha_m} (1-\theta_1)^{\beta_m} \theta_2^{\gamma_m} (1-\theta_2)^{\delta_m}}{B(\alpha_m+1, \beta_m+1) B(\gamma_m+1, \delta_m+1)} d\theta_2 d\theta_1$$

$$= \frac{q_2 \int_0^1 \theta_1^{\alpha_m} (1-\theta_1)^{\beta_m} \int_0^{\theta_1} \theta_2^{\gamma_m} \sum_{j=0}^{\delta_m} (-1)^j \binom{\delta_m}{j} \theta_2^j d\theta_2 d\theta_1}{B(\alpha_m+1, \beta_m+1) B(\gamma_m+1, \delta_m+1)}$$

$$= \frac{q_2 \sum_{j=0}^{\delta_m} (-1)^j \binom{\delta_m}{j} \frac{B(\alpha_m + \gamma_m + j + 2, \beta_m + 1)}{\gamma_m + j + 1}}{B(\alpha_m + 1, \beta_m + 1) B(\gamma_m + 1, \delta_m + 1)}$$

Similarly, if  $b = 1$ ,

$$\mathbb{E}_{\theta} | x_1, \dots, x_m L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta)$$

$$= \frac{\int_0^1 \int_0^{\theta_2} q_1 \theta_1^{\alpha_m} (1-\theta_1)^{\beta_m} \theta_2^{\gamma_m} (1-\theta_2)^{\delta_m} d\theta_1 d\theta_2}{B(\alpha_m + 1, \beta_m + 1) B(\gamma_m + 1, \delta_m + 1)}$$

$$= \frac{q_1 \sum_{j=0}^{\beta_m} (-1)^j \binom{\beta_m}{j} \frac{B(\gamma_m + \alpha_m + j + 2, \delta_m + 1)}{\alpha_m + j + 1}}{B(\alpha_m + 1, \beta_m + 1) B(\gamma_m + 1, \delta_m + 1)} \cdot$$

- (2) Thus for fixed  $x_1, \dots, x_m$  and  $a_1, \dots, a_m$  choose  $b = 1$  if

$$q_1 \sum_{j=0}^{\beta_m} (-1)^j \binom{\beta_m}{j} \frac{B(\alpha_m + \gamma_m + j + 2, \delta_m + 1)}{\alpha_m + j + 1}$$

is less than

$$q_2 \sum_{j=0}^{\delta_m} (-1)^j \binom{\delta_m}{j} \frac{B(\alpha_m + \gamma_m + j + 2, \beta_m + 1)}{\gamma_m + j + 1} ,$$

choose either  $b = 1$  or  $b = 2$  if the two expressions are equal, and choose  $b = 2$  otherwise. Call the minimum of the two above expressions  $G(x_1, \dots, x_m, a_1, \dots, a_m)$ .

- (3) In the rest of this section we shall let

$$\begin{aligned}
 (2.6) \quad S_k(x_1, \dots, x_{k-1}, a_1, \dots, a_k) &= \sum_{r_{1k}=0}^{n_{1k}} \sum_{r_{2k}=0}^{n_{2k}} \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}} \\
 &\cdot \left\{ \min_{a_{k+1}} \sum_{r_{1,k+1}=0}^{n_{1,k+1}} \sum_{r_{2,k+1}=0}^{n_{2,k+1}} \binom{n_{1,k+1}}{r_{1,k+1}} \binom{n_{2,k+1}}{r_{2,k+1}} \left\{ \dots \right. \right. \\
 &\cdot \left\{ \min_{a_m} \sum_{r_{1m}=0}^{n_{1m}} \sum_{r_{2m}=0}^{n_{2m}} \binom{n_{1m}}{r_{1m}} \binom{n_{2m}}{r_{2m}} \right. \\
 &\cdot \left. \left. G(x_1, \dots, x_m, a_1, \dots, a_m) \right\} \dots \right\}
 \end{aligned}$$

and

$$\begin{aligned}
 H_k(x_1, \dots, x_{k-1}, a_1, \dots, a_{k-1}) \\
 = \min_{a_k} S_k(x_1, \dots, x_{k-1}, a_1, \dots, a_k) \quad ,
 \end{aligned}$$

where  $k = 1, 2, \dots, m$ .

Also, we shall let

$$\begin{aligned}
 H_{m+1}(x_1, \dots, x_m, a_1, \dots, a_m) \\
 = G(x_1, \dots, x_m, a_1, \dots, a_m) \quad .
 \end{aligned}$$

Note that for  $k = 1, 2, \dots, m$

$$\begin{aligned}
 S_k(x_1, \dots, x_{k-1}, a_1, \dots, a_k) &= \sum_{r_{1k}=0}^{n_{1k}} \sum_{r_{2k}=0}^{n_{2k}} \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}} \\
 &\cdot H_{k+1}(x_1, \dots, x_k, a_1, \dots, a_k) \quad .
 \end{aligned}$$

Then

$$\begin{aligned}
 & E_{x_m | x_1, \dots, x_{m-1}} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \\
 &= \sum_{\text{all } x_m} P(x_m | x_1, \dots, x_{m-1}) L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \\
 &= \frac{S_m(x_1, \dots, x_{m-1}, a_1, \dots, a_m)}{B(\alpha_{m-1}+1, \beta_{m-1}+1) B(\gamma_{m-1}+1, \delta_{m-1}+1)} \cdot
 \end{aligned}$$

(4) Choose  $a_m$  such that  $S_m(x_1, \dots, x_{m-1}, a_1, \dots, a_m)$  is a minimum for fixed  $x_1, \dots, x_{m-1}$  and  $a_1, \dots, a_{m-1}$  thus computing  $H_m(x_1, \dots, x_{m-1}, a_1, \dots, a_{m-1})$ .

(5) At the  $k^{\text{th}}$  stage

$$\begin{aligned}
 & E_{x_k | x_1, \dots, x_{k-1}} L^{(m-k+1)}(x_1, \dots, x_{k-1}, a_1, \dots, a_k) \\
 &= \frac{S_k(x_1, \dots, x_{k-1}, a_1, \dots, a_k)}{B(\alpha_{k-1}+1, \beta_{k-1}+1) B(\gamma_{k-1}+1, \delta_{k-1}+1)} \cdot
 \end{aligned}$$

Choose  $a_k$  such that  $S_k(x_1, \dots, x_{k-1}, a_1, \dots, a_k)$  is a minimum, thus computing

$$H_k(x_1, \dots, x_{k-1}, a_1, \dots, a_{k-1}) \cdot$$

(6) Finally at the first stage choose  $a_1$  such that  $S_1(a_1)$  is a minimum.

#### 2.4 Analysis for a Linear Loss Function

In this section we shall assume the same situation as there was in Sections 2.1, 2.2, and 2.3, except for the loss

function. Let

$$(2.7) \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) = \left\{ \begin{array}{ll} k_{10} + k_{11}\theta_1 + k_{12}\theta_2 & \text{for } b = 1 \\ k_{20} + k_{21}\theta_1 + k_{22}\theta_2 & \text{for } b = 2 \end{array} \right\},$$

where  $k_{10}$ ,  $k_{11}$ ,  $k_{12}$ ,  $k_{20}$ ,  $k_{21}$ , and  $k_{22}$  are any real numbers. Then, following the general outline of extensive form of analysis in Section 2.1 and using (2.4) and (2.5), we obtain:

(1) If  $b = 1$ ,

$$\begin{aligned} & E_{\theta|x_1, \dots, x_m} L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) \\ &= \int_0^1 \int_0^1 \frac{\theta_1^{\alpha_m} (1-\theta_1)^{\beta_m} \theta_2^{\gamma_m} (1-\theta_2)^{\delta_m}}{B(\alpha_m+1, \beta_m+1) B(\gamma_m+1, \delta_m+1)} \\ & \quad \cdot (k_{10} + k_{11}\theta_1 + k_{12}\theta_2) d\theta_1 d\theta_2 \\ &= k_{10} + k_{11} \left( \frac{\alpha_m + 1}{\alpha_m + \beta_m + 2} \right) + k_{12} \left( \frac{\gamma_m + 1}{\gamma_m + \delta_m + 2} \right). \end{aligned}$$

Similarly, if  $b = 2$ ,

$$\begin{aligned} & E_{\theta|x_1, \dots, x_m} L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) \\ &= k_{20} + k_{21} \left( \frac{\alpha_m + 1}{\alpha_m + \beta_m + 2} \right) + k_{22} \left( \frac{\gamma_m + 1}{\gamma_m + \delta_m + 2} \right). \end{aligned}$$

(2) Thus we choose  $b = 1$  if

$$(k_{20}-k_{10})+(k_{21}-k_{11}) \left( \frac{\alpha_m+1}{\alpha_m+\beta_m+2} \right) +(k_{22}-k_{12}) \left( \frac{\gamma_m+1}{\gamma_m+\delta_m+2} \right)$$

is greater than zero, choose either  $b = 1$  or  $b = 2$  if it is equal to zero, and choose  $b = 2$  if it is less than zero. Then

$$(2.8) \quad L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \\ = \min \left\{ k_{10}+k_{11} \left( \frac{\alpha_m+1}{\alpha_m+\beta_m+2} \right) +k_{12} \left( \frac{\gamma_m+1}{\gamma_m+\delta_m+2} \right) , \right. \\ \left. k_{20}+k_{21} \left( \frac{\alpha_m+1}{\alpha_m+\beta_m+2} \right) +k_{22} \left( \frac{\gamma_m+1}{\gamma_m+\delta_m+2} \right) \right\}$$

$$(3) \quad E_{x_m | x_1, \dots, x_{m-1}} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m)$$

$$= \sum_{r_{1m}=0}^{n_{1m}} \sum_{r_{2m}=0}^{n_{2m}} \binom{n_{1m}}{r_{1m}} \binom{n_{2m}}{r_{2m}}$$

$$\cdot \frac{B(\alpha_m+1, \beta_m+1) B(\gamma_m+1, \delta_m+1)}{B(\alpha_{m-1}+1, \beta_{m-1}+1) B(\gamma_{m-1}+1, \delta_{m-1}+1)}$$

$$\cdot L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m)$$

$$= \sum_{r_{1m}=0}^{n_{1m}} \sum_{r_{2m}=0}^{n_{2m}} \frac{\binom{n_{1m}}{r_{1m}} \binom{n_{2m}}{r_{2m}} G^*(x_1, \dots, x_m, a_1, \dots, a_m)}{B(\alpha_{m-1}+1, \beta_{m-1}+1) B(\gamma_{m-1}+1, \delta_{m-1}+1)}$$

$$= \frac{S'_m(x_1, \dots, x_{m-1}, a_1, \dots, a_m)}{B(\alpha_{m-1}+1, \beta_{m-1}+1) B(\gamma_{m-1}+1, \delta_{m-1}+1)},$$

where

$$(2.9) \quad \begin{aligned} G'(x_1, \dots, x_m, a_1, \dots, a_m) \\ = B(\alpha_m+1, \beta_m+1) B(\gamma_m+1, \delta_m+1) \\ \cdot L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \end{aligned}$$

The rest of the analysis proceeds exactly as it does with the constant loss function except that the functions  $G$ ,  $S$ , and  $H$  are replaced by  $G'$ ,  $S'$ , and  $H'$ .

### 2.5 An Example of the Computations

In this section the "decision tree," the diagram showing all of the possible decisions and outcomes which can occur at each stage, will be constructed for the following example:

Let  $m = 3$  with  $n_1 = n_2 = n_3 = 1$ . Let  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  so that

$$\rho(\theta) = \left\{ \begin{array}{l} 1 \text{ for } 0 \leq \theta_1 \leq 1 \text{ and } 0 \leq \theta_2 \leq 1 \\ 0 \text{ otherwise} \end{array} \right\}.$$

Let

$$\begin{aligned} L(x_1, x_2, x_3, a_1, a_2, a_3, b, \theta) \\ = \left\{ \begin{array}{l} \theta_2 - \theta_1 \text{ for } b = 1 \\ \theta_1 - \theta_2 \text{ for } b = 2 \end{array} \right\} \end{aligned}$$

We shall compute  $S_k^*$  ( $x_1, \dots, x_{k-1}, a_1, \dots, a_k$ ) and  $H_k^*$  ( $x_1, \dots, x_{k-1}, a_1, \dots, a_{k-1}$ ) for  $k = 1, 2$ , and  $3$  ( and for  $k = 4$  for the function  $H_k^*$ ), where  $H_k^*$  and  $S_k^*$  are the same as defined by (2.6) except that the function  $G$  is replaced by  $G^*$ , which is defined by (2.9). Then we shall place  $H_k^*$  and  $S_k^*$  on the diagram showing the decision tree without the primes for the sake of clarity.

For the loss function in our example we find from (2.8) that

$$(2.10) \quad L^{(1)}(x_1, x_2, x_3, a_1, a_2, a_3) \\ = \min \left\{ -\frac{(\alpha_3+1)}{\alpha_3+\beta_3+2} + \frac{(\gamma_3+1)}{\gamma_3+\delta_3+2}, \right. \\ \left. \frac{(\alpha_3+1)}{\alpha_3+\beta_3+2} - \frac{(\gamma_3+1)}{\gamma_3+\delta_3+2} \right\},$$

where  $b = 1$  or  $2$  is chosen according as the first expression is less than or greater than the second. Then from (2.6), (2.9), and (2.10) we find that (dropping the arguments of  $S_k^*$  and  $H_k^*$ )

$$H_4^* = B(\alpha_3+1, \beta_3+1) \cdot B(\gamma_3+1, \delta_3+1) \\ \cdot \min \left\{ -\frac{(\alpha_3+1)}{\alpha_3+\beta_3+2} + \frac{(\gamma_3+1)}{\gamma_3+\delta_3+2}, \right. \\ \left. \frac{(\alpha_3+1)}{\alpha_3+\beta_3+2} - \frac{(\gamma_3+1)}{\gamma_3+\delta_3+2} \right\}$$



$$= |B(\alpha_3+2, \beta_3+1) B(\gamma_3+1, \delta_3+1) - B(\alpha_3+1, \beta_3+1) B(\gamma_3+2, \delta_3+1)| .$$

(Note that in this example each  $\binom{n_{1k}}{r_{1k}}$  and  $\binom{n_{2k}}{r_{2k}}$  is 1.)

The possible values of  $H_4^*$  and the terminal decisions corresponding to the possible values of  $\alpha_3, \beta_3, \gamma_3,$  and  $\delta_3$  if we start with  $n_{11} = 1$  and  $n_{21} = 0$  are given in Table 1.

We are now ready to construct the decision tree. We shall construct only half of the entire decision tree for this example since the other half can be obtained by symmetry. The diagram starting with  $n_{11} = 1$ , is shown in Figure 1. The terminal decisions one should take are also shown in Figure 1.

The decision tree for  $m = 3$  with  $n_1 = n_2 = n_3 = 1$  and  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  but with

$$L(x_1, x_2, x_3, a_1, a_2, a_3, b, \theta) = \left\{ \begin{array}{l} 1 \text{ if } b = 1 \text{ and } \theta_1 < \theta_2 \text{ or if } b = 2 \text{ and } \theta_1 > \theta_2 \\ 0 \text{ otherwise} \end{array} \right\}$$

has also been computed. It turns out that for these quite small examples the decision trees have exactly the same decisions at each stage for the same outcomes in previous stages.

Table 1

Results of Computations for Example 1

<u><math>\alpha_3</math></u>	<u><math>\beta_3</math></u>	<u><math>\gamma_3</math></u>	<u><math>\delta_3</math></u>	<u><math>360 H_4^*</math></u>	<u>b</u>
0	1	0	2	-5	1
0	1	1	1	-5	2
0	1	2	0	-25	2
1	0	0	2	-25	1
1	0	1	1	-5	1
1	0	2	0	-5	2
0	2	0	1	-5	2
0	2	1	0	-25	2
1	1	0	1	-5	1
1	1	1	0	-5	2
2	0	0	1	-25	1
2	0	1	0	-5	1
0	3	0	0	-27	2
1	2	0	0	-3	2
2	1	0	0	-3	1
3	0	0	0	-27	1

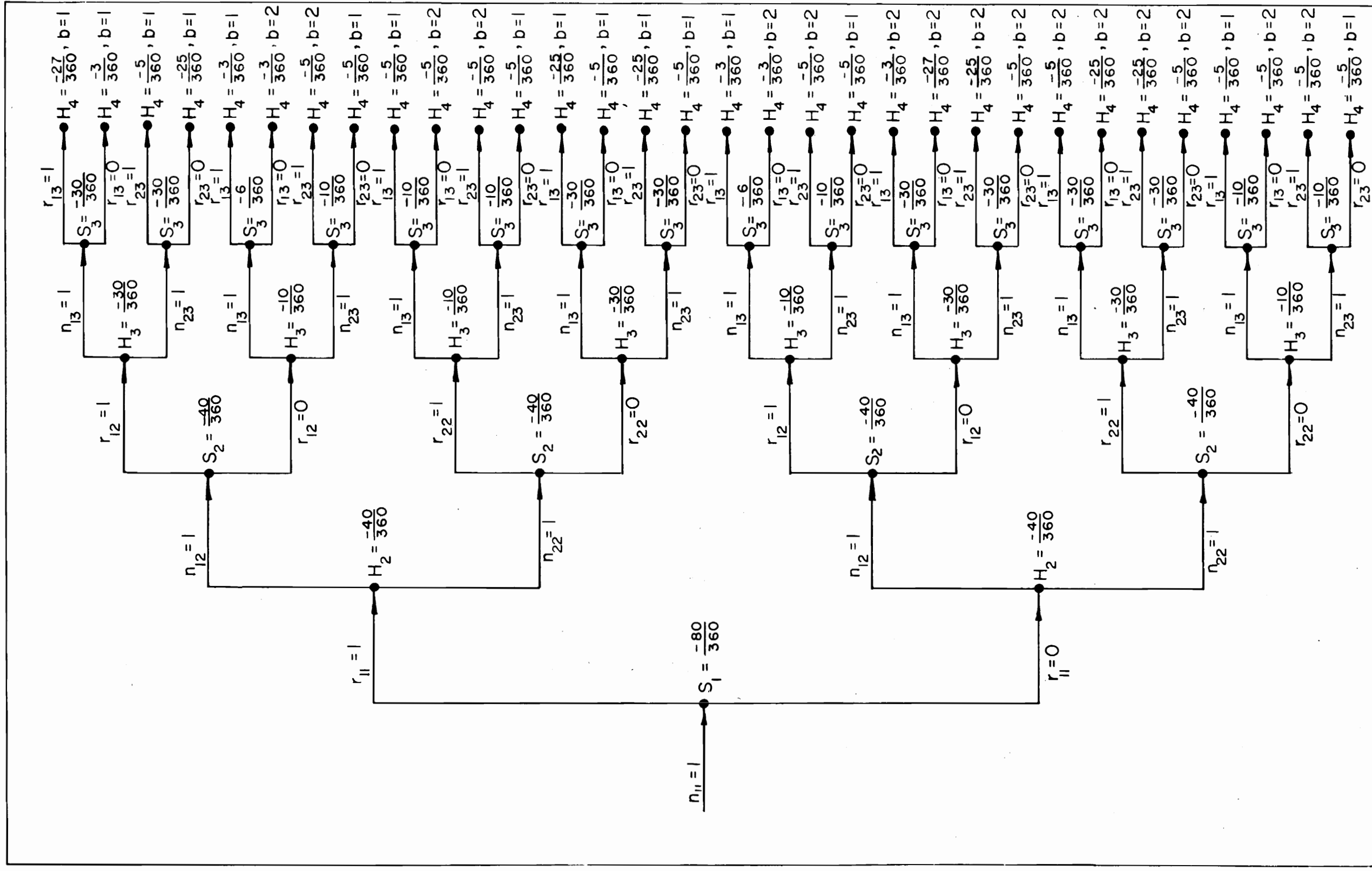


FIGURE 1  
DECISION TREE FOR LINEAR LOSS FUNCTION

## 2.6 Numerical Results

As one can see from the example of the previous section, the computations for finding optimum strategies are long even for very small examples. For larger examples, the computations rapidly "get out of hand." With the above example there are  $2(4)^3$  or 128 "terminal branches" on the decision tree; but if there were five stages with one observation per stage, one would have to construct a "decision tree" with  $2(4)^5$  or 2048 terminal branches. Thus two programs for the IBM 1620 computer have been written for computing decision trees. The first program is for the linear loss function defined by (2.7), and the second is for the constant loss function defined by (2.1). Each of these programs has two main options. One can enter the decision tree at the  $k$ 'th stage ( $k = 1, 2, \dots, m$ ) with the results of the previous  $k - 1$  stages and have the entire decision tree above the point of entry punched out on cards; or one can have only  $S_k(x_1, \dots, x_{k-1}, a_1, \dots, a_k)$  punched out for each of the possible values of  $n_{1k}$ . Thus, using the second option, one can find out how many observations should be taken on each drug in the first stage; perform the experiment; using the data from the first stage, find out how many observations should be taken on each drug in the second stage; etc. It should be emphasized that at each stage the computer has to go through

the entire decision tree above the given stage in the computations even though it may not punch out the entire decision tree.

A number of examples with various prior distributions (all basically joint beta-type distributions as defined by (2.2)) and three different loss functions were considered for computer calculations. The three loss functions were

$$(2.11) \quad \begin{aligned} 1. \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) &= \begin{cases} \theta_2 - \theta_1 & \text{for } b = 1 \\ \theta_1 - \theta_2 & \text{for } b = 2 \end{cases} \\ 2. \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) &= \begin{cases} \theta_2 - \theta_1 & \text{for } b = 1 \\ \theta_1 - 2\theta_2 & \text{for } b = 2 \end{cases} \\ 3. \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) &= \begin{cases} 1 & \text{if } \theta_1 < \theta_2 \text{ and } b = 1 \\ 1 & \text{if } \theta_1 > \theta_2 \text{ and } b = 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

The examples were as follows:

- (1)  $m = 3$  with  $n_1 = 1$ ,  $n_2 = 1$ , and  $n_3 = 1$ , which is the example which was computed by hand.
- (2)  $m = 2$  with  $n_1 = 4$  and  $n_2 = 2$ . Part of the decision tree was calculated by hand for  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  and Loss Function 1, and then the entire decision tree was calculated and punched out on cards by the computer. Then the best values of  $n_{11}$  for various

prior distributions and Loss Functions 1 and 2 were found. Finally the entire decision tree was punched out for  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  and Loss Function 3, and the decisions at each stage were the same as they were for Loss Function 1.

- (3)  $m = 3$  with  $n_1 = 3, n_2 = 2, n_3 = 3$ . The entire decision tree for this example is much too large to be punched out. There are  $(2)(20)(10)(20)$  or 8000 terminal branches on the tree. Not only the results for the first stage were found, but also results for assumed values of  $n_{11}, n_{21}, r_{11}$ , and  $r_{21}$  were found for the second stage; and results for assumed values of  $n_{11}, n_{21}, r_{11}, r_{21}, n_{12}, n_{22}, r_{12}$ , and  $r_{22}$  were found for the third stage.
- (4)  $m = 5$  with  $n_1 = n_2 = n_3 = n_4 = n_5 = 1$ . The entire decision tree assuming  $n_{11} = 1$  was punched out for Loss Functions 1 and 3 with  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ . The decisions in the decision trees for the two loss functions were the same except for a few cases in the fourth stage in which the linear loss function gave decisions that  $n_{14} = 0$  and  $n_{14} = 1$  were equivalent and the constant loss function gave decisions that one was better than the other.
- (5)  $m = 2$  with  $n_1 = 5$  and  $n_2 = 4$ .
- (6)  $m = 1$  with  $n_1 = 10$ .

(7)  $m = 1$  with  $n_1 = 21$ .

The results for these examples are shown in Table 2.

The following prior distributions are used in Table 2:

- (2.12)    (1)  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$   
          (2)  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 2$   
          (3)  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 9$   
          (4)  $\alpha_0 = \beta_0 = \gamma_0 = 2, \delta_0 = 8$   
          (5)  $\alpha_0 = \beta_0 = 2, \gamma_0 = \delta_0 = 8$   
          (6)  $\alpha_0 = \beta_0 = 4, \gamma_0 = \delta_0 = 8$  .

In Table 2 under the heading "Example" the number of observations in each stage is listed first, and then the number of the example as given in this section is listed in parentheses. For example, 3-2-3 (3) means  $m = 3$  with  $n_1 = 3$ ,  $n_2 = 2$ , and  $n_3 = 3$ , which is Example 3.

The results for two other procedures, the stage-by-stage and the approximate procedures, which are defined in Chapter III and IV respectively, are also given in Table 2.

As one can see from Table 2, there seems to be good agreement between the results for Loss Functions 1 and 3 in these small examples. However, there does not seem to be any definite pattern in the results. It should be noted that computing an entire decision tree for even a small problem can become quite time consuming even though one is using a

Table 2

Some Computer Results with Loss Functions and Prior Distributions as Defined by (2.11) and (2.12)

First Stage Results

Example	Loss Function	Prior Distribution	Extensive-Form Analysis		Stage-by-Stage Procedure		Approximate Best Values of $n_{11}$
			Best Values of $n_{11}$	Time to Obtain Results	Best Values of $n_{11}$	Time to Obtain Results	
4-2(2)	1	1	0, 1, 2, 3, 4	6 min.	1, 3		2
4-2(2)	1	2	0, 1, 2, 3, 4				
4-2(2)	1	3	0, 1, 2, 3, 4				
4-2(2)	1	4	3, 4				
4-2(2)	1	5	3, 4				
4-2(2)	1	1	1				
4-2(2)	2	1	0, 1, 2, 3, 4	11 min.	1, 3		
3-2-3(3)	3	1	0, 1, 2, 3	58 min.	1, 2		
3-2-3(3)	1	1	0, 0, 3	99 min.	1, 2		
1-1-1-1(4)	3	1	0, 1	16 min.			
1-1-1-1(4)	1	1	0, 1				
5-4(5)	3	1	0, 1	27 min.	2, 3		
5-4(5)	1	1	0, 5	50 min.	2, 3		
10(6)	1	6	0, 5		4, 6	4 min.	9
10(6)	1	6			8, 9, 10	10 min.	
10(6)	3	1			4, 6		
10(6)	3	1			8, 9, 10	70 min.	9
21(7)	1	3			10, 11	27 min.	
21(7)	3	1			10, 11	122 min.	



Table 2 (continued)

Second Stage Results

Best Values of  $n_{12}$

<u>Example</u>	<u>Loss Function</u>	<u>Prior Distribution</u>	<u>Assumed Results of Stage 1</u>	<u>Extensive-Form Analysis</u>	<u>Stage-by-Stage Procedure</u>	<u>Approximate Procedure</u>
3-2-3(3)	1	1	$n_{11}=1, r_{11}=1$ $n_{21}=1, r_{21}=1$	1,2	1,2	1
3-2-3(3)	3	1	$n_{11}=3, r_{11}=1$ $n_{21}=0, r_{21}=0$	0		

Third Stage Results

Best Values of  $n_{13}$

<u>Example</u>	<u>Loss Function</u>	<u>Prior Distribution</u>	<u>Assumed Results</u>		<u>Extensive-Form Analysis</u>	<u>Approximate Procedure</u>
			<u>Stage 1</u>	<u>Stage 2</u>		
3-2-3(3)	1	1	$n_{11}=1, r_{11}=1$ $n_{21}=1, r_{21}=1$	$n_{12}=1, r_{12}=0$ $n_{22}=1, r_{22}=1$	1,3	2
3-2-3(3)	3	1	$n_{11}=3, r_{11}=1$ $n_{21}=0, r_{21}=0$	$n_{12}=0, r_{12}=0$ $n_{22}=2, r_{22}=1$	0,2	

computer. For very large problems the decision tree quickly becomes too large for a computer to handle. For example, if one added a fourth stage with  $n_4 = 3$  to Example 3, the resulting decision tree would have twenty times the number of terminal branches; and if one assumed that the time necessary for obtaining the results for the first stage was twenty times longer (which is undoubtedly being conservative), one would expect the computer to take almost twenty hours to find the results for the first stage with Loss Function 1 and Prior Distribution 1. Thus it is apparent that, when the number of observations per stage or the number of stages is at all large, using extensive-form analysis becomes quite impractical even when a computer is used.

### III. THE STAGE-BY-STAGE PROCEDURE

#### 3.1 General Procedures

In view of the practical difficulties associated with extensive-form analysis, we turn to the following simpler procedure: Suppose that at each stage the experimenter acts as if he were at the terminal stage; i.e., if he is at the  $k$ 'th stage, he assumes that there are only  $k$  stages in the entire procedure and chooses  $n_{1k}$  and  $n_{2k}$  accordingly. Thus at the  $k$ 'th stage ( $k = 2, 3, \dots, m$ ) he evaluates

$$E_{x_k | x_1, \dots, x_{k-1}} \min_b E_{\theta | x_1, \dots, x_k} L(x_1, \dots, x_k, a_1, \dots, a_k, b, \theta)$$

and chooses the " $a_k$ " which minimizes the expression. At the first stage he evaluates

$$E_{x_1} \min_b E_{\theta | x_1} L(x_1, a_1, b, \theta)$$

and chooses the " $a_1$ " which gives the minimum.

For the linear loss function defined by (2.7) (but dependent only on  $x_1, \dots, x_k, a_1, \dots, a_k, b, \theta$ ) one would evaluate at the  $k$ 'th stage

$$\sum_{r_{1k}=0}^{n_{1k}} \sum_{r_{2k}=0}^{n_{2k}} \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}}$$

$$\cdot \frac{B(\alpha_k+1, \beta_k+1) B(\gamma_k+1, \delta_k+1)}{B(\alpha_{k-1}+1, \beta_{k-1}+1) B(\gamma_{k-1}+1, \delta_{k-1}+1)}$$

$$\cdot \min \left\{ \begin{aligned} &k_{10} + k_{11} \left( \frac{\alpha_k+1}{\alpha_k+\beta_k+2} \right) + k_{12} \left( \frac{\gamma_k+1}{\gamma_k+\delta_k+2} \right), \\ &k_{20} + k_{21} \left( \frac{\alpha_k+1}{\alpha_k+\beta_k+2} \right) + k_{22} \left( \frac{\gamma_k+1}{\gamma_k+\delta_k+2} \right) \end{aligned} \right\}$$

and choose the  $n_{1k}$  which minimized the expression subject to  $n_{1k} + n_{2k} = n_k$ .

For the constant loss function defined by (2.1) (but dependent only on  $x_1, \dots, x_k, a_1, \dots, a_k, b, \theta$ ) one would evaluate at the  $k$ 'th stage

$$\sum_{r_{1k}=0}^{n_{1k}} \sum_{r_{2k}=0}^{n_{2k}} \frac{\binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}}}{B(\alpha_{k-1}+1, \beta_{k-1}+1) B(\gamma_{k-1}+1, \delta_{k-1}+1)}$$

$$\cdot \min \left\{ \begin{aligned} &q_1 \sum_{j=0}^{\beta_k} (-1)^j \binom{\beta_k}{j} \frac{B(\alpha_k+\gamma_k+j+2, \delta_k+1)}{\alpha_k+j+1}, \\ &q_2 \sum_{j=0}^{\delta_k} (-1)^j \binom{\delta_k}{j} \frac{B(\alpha_k+\gamma_k+j+2, \beta_k+1)}{\gamma_k+j+1} \end{aligned} \right\}$$

and choose the  $n_{1k}$  which minimized the expression subject to  $n_{1k} + n_{2k} = n_k$ . Notice that this procedure is equivalent at each stage to a one-stage procedure in which  $\alpha_0 = \alpha_{k-1}$ ,

$\beta_0 = \beta_{k-1}$ ,  $\gamma_0 = \gamma_{k-1}$ , and  $\delta_0 = \delta_{k-1}$ . Thus at each stage one has merely a one-stage procedure with an altered prior distribution. It should also be observed that, for the same results at the end of the  $(m-1)$ th stage, extensive-form analysis and the stage-by-stage procedure give identical results for the  $m$ 'th stage.

### 3.2 An Example of the Calculations

Consider the example of the computations for extensive-form analysis in which  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  and  $m = 3$  with  $n_1 = n_2 = n_3$ . Suppose that on the first stage one took  $n_{11} = 1$  and  $n_{21} = 0$  and that  $r_{11} = 0$ . Then  $\alpha_1 = 0$ ,  $\beta_1 = 1$ ,  $\gamma_1 = 0$ , and  $\delta_1 = 0$ . It is now desired to determine whether Drug 1 or Drug 2 should be used on the second stage. The stage-by-stage procedure will be used, and Loss Function 1 as defined by (2.11) will be assumed. In Table 3, which shows the computations, we let

$$\begin{aligned}
 H &= B(\alpha_2+1, \beta_2+1) B(\gamma_2+1, \delta_2+1) \\
 &\cdot \min \left\{ \frac{\alpha_2+1}{\alpha_2+\beta_2+2} - \frac{\gamma_2+1}{\gamma_2+\delta_2+2}, \frac{\gamma_2+1}{\gamma_2+\delta_2+2} - \frac{\alpha_2+1}{\alpha_2+\beta_2+2} \right\} \\
 &= -|B(\alpha_2+2, \beta_2+1) B(\gamma_2+1, \delta_2+1) \\
 &\quad - B(\alpha_2+1, \beta_2+1) B(\gamma_2+2, \delta_2+1)| .
 \end{aligned}$$

Table 3

Computations for Stage-by-Stage Procedure Example

$\underline{\alpha}_2$	$\underline{\beta}_2$	$\underline{\gamma}_2$	$\underline{\delta}_2$	$\underline{H}$
1	1	0	0	0
0	2	0	0	$-\frac{1}{12}$
0	1	1	0	$-\frac{1}{12}$
0	1	0	1	0

Thus for this example

$$\begin{aligned}
 & B(\alpha_1+1, \beta_1+1) B(\gamma_1+1, \delta_1+1) \\
 & \cdot E_{x_2|x_1} \min_b E_{\theta|x_1, x_2} L(x_1, x_2, a_1, a_2, b, \theta) \\
 & = \left\{ \begin{array}{l} 0 + (-1/12) \text{ or } -1/12 \text{ for } n_{12} = 1 \\ -1/12 + 0 \text{ or } -1/12 \text{ for } n_{22} = 1 \end{array} \right\} .
 \end{aligned}$$

Thus one could use either Drug 1 or Drug 2 on Stage 2, which is the same conclusion reached from Figure 1 when extensive-form analysis was used.

The same example was computed with Loss Function 3 as defined by (2.11) instead of Loss Function 1. The results were the same as those which resulted from using Loss Function 1. Thus in this small example the results were exactly the same as they were when extensive-form analysis was used

for both Loss Functions 1 and 3.

As one can see from the example, the stage-by-stage procedure requires much less computation than extensive-form analysis does. The computations for a linear loss function are especially easy to perform. In fact, if one uses a table of beta functions, and has a small number of observations in each stage, using a computer is not necessary for the calculations.

### 3.3 Numerical Results

If there is a large number of observations in any of the stages for either type of loss function (linear or constant), or if one has a constant loss function and  $\beta_k$  and  $\delta_k$  are very large, one will want to use a computer in the stage-by-stage procedure. However, it is not necessary to write new programs for this procedure since at the  $k$ 'th stage one can use the programs for extensive-form analysis by letting  $\alpha_0 = \alpha_{k-1}$ ,  $\beta_0 = \beta_{k-1}$ ,  $\gamma_0 = \gamma_{k-1}$ ,  $\delta_0 = \delta_{k-1}$ ,  $\alpha_1 = \alpha_k$ ,  $\beta_1 = \beta_k$ ,  $\gamma_1 = \gamma_k$ , and  $\delta_1 = \delta_k$  and assuming that one has only a one stage procedure.

The same examples computed for extensive-form analysis and two other examples -- Example 6 with  $m = 1$  and  $n_1 = 10$  and Example 7 with  $m = 1$  and  $n_1 = 21$  -- were considered using the stage-by-stage procedure. For Example 4 decision trees for Loss Functions 1 and 3 were constructed using the

stage-by-stage procedure and compared with the decision trees resulting from extensive-form analysis. The only difference between the results for the two types of procedures was that the stage-by-stage procedure in some cases stated that using one drug was better than using the other drug while the extensive-form analysis stated that using one of the drugs was equivalent to using the other. This occurred twice in the third stage when the constant loss function was used and eight times in both the third and fourth stages when the linear loss function was used. It was also noted that the stage-by-stage decision trees gave the same decisions at each stage for the two loss functions except for the terminal decisions.

The other results are summarized in Table 2.

These examples indicate that for the symmetric loss functions and for  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ ,  $n_{11} = \frac{1}{2}(n_1-1)$  and  $n_{11} = \frac{1}{2}(n_1+1)$  are best if  $n_1$  is odd and  $n_{11} = \frac{1}{2}n_1 - 1$  and  $n_{11} = \frac{1}{2}n_1 + 1$  (but not  $n_{11} = \frac{1}{2}n_1$ ) are best if  $n_1$  is even. Also, the examples indicate good agreement between the stage-by-stage procedure decisions and the extensive-form analysis decisions except on the first stage. However, in these examples, when extensive-form analysis was used, the values of  $E_{x_1} [L^{(m)}(x_1, a_1)]$  for different  $n_{11}$  were almost equal. For example, when  $n_1 = 5$  and  $n_2 = 4$  with  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$



and Loss Function l,

$$E_{x_1} [L^{(m)}(x_1, a_1)] = \left\{ \begin{array}{l} -.27896822 \text{ for } n_{11} = 0 \\ -.27865073 \text{ for } n_{11} = 1 \\ -.27825392 \text{ for } n_{11} = 2 \\ -.27825390 \text{ for } n_{11} = 3 \\ -.27865074 \text{ for } n_{11} = 4 \\ -.27896822 \text{ for } n_{11} = 5 \end{array} \right\} .$$

#### IV. AN APPROXIMATE PROCEDURE

##### 4.1 General Procedure

Suppose again the prior distribution of  $\theta$  is that given by (2.2). Then from (2.4) at the  $k$ 'th stage ( $k = 1, 2, \dots, m$ )

$$\begin{aligned} P(\theta | x_1, \dots, x_k) &= \left[ \frac{\theta_1^{k(1-\theta_1)} \beta_k}{B(\alpha_k+1, \beta_k+1)} \right] \left[ \frac{\theta_2^{\gamma_k(1-\theta_2)} \delta_k}{B(\gamma_k+1, \delta_k+1)} \right] \\ &= P(\theta_1 | x_1, \dots, x_k) P(\theta_2 | x_1, \dots, x_k) . \end{aligned}$$

Thus

$$\begin{aligned} \text{Var} (\theta_1 - \theta_2 | x_1, \dots, x_k) &= \text{Var} (\theta_1 | x_1, \dots, x_k) + \text{Var} (\theta_2 | x_1, \dots, x_k) \\ &= \frac{(\alpha_k+1)(\beta_k+1)}{(\alpha_k+\beta_k+2)^2(\alpha_k+\beta_k+3)} + \frac{(\gamma_k+1)(\delta_k+1)}{(\gamma_k+\delta_k+2)^2(\gamma_k+\delta_k+3)} \\ &= \frac{\bar{p}_{1k} (1 - \bar{p}_{1k})}{\alpha_k+\beta_k+3} + \frac{\bar{p}_{2k} (1 - \bar{p}_{2k})}{\gamma_k+\delta_k+3} , \end{aligned}$$

where

$$\bar{p}_{1k} = \frac{\alpha_k+1}{\alpha_k+\beta_k+2} \quad \text{and} \quad \bar{p}_{2k} = \frac{\gamma_k+1}{\gamma_k+\delta_k+2} .$$

We shall now use  $\bar{p}_{1,k-1}(1 - \bar{p}_{1,k-1})$  and  $\bar{p}_{2,k-1}(1 - \bar{p}_{2,k-1})$  as approximations for  $\bar{p}_{1k}(1 - \bar{p}_{1k})$  and  $\bar{p}_{2k}(1 - \bar{p}_{2k})$ , respectively. Then

$$\text{Var} (\theta_1 - \theta_2 | x_1, \dots, x_k) = \frac{\bar{p}_{1,k-1}(1-\bar{p}_{1,k-1})}{\alpha_{k-1} + \beta_{k-1} + n_{1k} + 3} + \frac{\bar{p}_{2,k-1}(1-\bar{p}_{2,k-1})}{\gamma_{k-1} + \delta_{k-1} + n_{2k} + 3} .$$

The approximate procedure will be to choose  $n_{1k}$  and  $n_{2k}$  at the  $k$ 'th stage such that  $\text{Var} (\theta_1 - \theta_2 | x_1, \dots, x_k)$  is minimized (approximately). Thus we shall take

$$\frac{\alpha_{k-1} + \beta_{k-1} + 3 + n_{1k}}{\gamma_{k-1} + \delta_{k-1} + 3 + (n_k - n_{1k})} = \sqrt{\frac{\bar{p}_{1,k-1}(1-\bar{p}_{1,k-1})}{\bar{p}_{2,k-1}(1-\bar{p}_{2,k-1})}} = R_{k-1} ,$$

or

$$(4.1) \quad n_{1k} \approx \frac{(\gamma_{k-1} + \delta_{k-1} + n_k + 3) R_{k-1} - \alpha_{k-1} - \beta_{k-1} - 3}{R_{k-1} + 1}$$

If the value of  $n_{1k}$  which satisfies (4.1) to the nearest integer is negative, we shall take  $n_{1k} = 0$  and  $n_{2k} = n_k$ . If it is greater than  $n_k$ , we shall take  $n_{1k} = n_k$  and  $n_{2k} = 0$ .

After  $m$  stages we shall choose the decision  $\theta_1 < \theta_2$  if  $\bar{p}_{1m} < \bar{p}_{2m}$ , the decision  $\theta_1 > \theta_2$  if  $\bar{p}_{1m} > \bar{p}_{2m}$ , and either decision if  $\bar{p}_{1m} = \bar{p}_{2m}$ .

#### 4.2 Some Numerical Results

The approximate procedure outlined above was used in some of the examples previously considered with the stage-by-stage procedure and extensive-form analysis. It should be

noted that the approximate procedure is good only for symmetric loss functions and is independent of the total number of stages as is the stage-by-stage procedure.

Let us assume we are at the first stage and that  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ . Then

$$n_{11} = \frac{(0 + n_1 + 3) \sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{(\frac{1}{2})(\frac{1}{2})}} - (0 + 3)}{\sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{(\frac{1}{2})(\frac{1}{2})}} + 1} = \frac{1}{2}n_1 .$$

Thus in this case, if  $n_{11}$  is even, one uses  $n_{11} = \frac{1}{2}n_1$  for the first stage; and if  $n_{11}$  is odd, one uses either  $n_{11} = \frac{1}{2}(n_1+1)$  or  $\frac{1}{2}(n_1-1)$ . This result does not, in general, agree with that of extensive-form analysis, but does agree (at least for all the examples considered) with the results of the stage-by-stage procedure if  $n_1$  is odd and is close to the results for  $n_1$  even.

Other results are shown in Table 2.

V. STAGE-BY-STAGE PROCEDURES ASSUMING  
NORMAL DISTRIBUTIONS

5.1 Distribution Theory

In this chapter we shall again assume that we have a stage-by-stage procedure in which at the  $k$ 'th stage ( $k = 1, 2, \dots, m$ ) we allocate  $n_{1k}$  out of  $n_k$  observations to Population 1 and  $n_{2k} = n_k - n_{1k}$  observations to Population 2. However, this time we shall assume that the random variables we generate at the  $k$ 'th stage are the means  $\bar{x}_k$  of the sample of  $n_{1k}$  and  $\bar{y}_k$  of the sample of  $n_{2k}$ , where  $\bar{x}_k \sim N(\theta_1, \sigma_1^2/n_{1k})$  and  $\bar{y}_k \sim N(\theta_2, \sigma_2^2/n_{2k})$ . (In other words  $\bar{x}_k$  and  $\bar{y}_k$  are normally distributed with means  $\theta_1$  and  $\theta_2$  and known variances  $\sigma_1^2/n_{1k}$  and  $\sigma_2^2/n_{2k}$ .) Our terminal decisions shall be the same as before.

We shall also assume that the prior distributions of  $\theta_1$  and  $\theta_2$ ,  $\rho_1(\theta_1)$  and  $\rho_2(\theta_2)$ , are  $N(\bar{x}_0, \sigma_1^2/n_{10})$  and  $N(\bar{y}_0, \sigma_2^2/n_{20})$  respectively.

In order to compute the expected values for the stage-by-stage procedure as outlined in Section 3.1 for the case of normal rather than binomial random variables, we shall need expressions for the following distributions:

$$P(\theta_1 | \bar{x}_1, \dots, \bar{x}_k), P(\theta_2 | \bar{y}_1, \dots, \bar{y}_k), P(\bar{x}_k | \bar{x}_1, \dots, \bar{x}_{k-1}),$$

$P(\bar{x}_k | \bar{x}_1, \dots, \bar{x}_{k-1})$ , and  $P(\bar{y}_k | \bar{y}_1, \dots, \bar{y}_{k-1})$ .

Let  $h_{1k} = n_{1k}/\sigma_1^2$  and  $h_{2k} = n_{2k}/\sigma_2^2$  for  $k = 0, 1, 2, \dots, m$ .

We shall also assume that all  $\bar{x}_1, \dots, \bar{x}_m$ , and  $\bar{y}_1, \dots, \bar{y}_m$  are mutually independent.

It is well known (See Raiffa and Schlaifer, pp. 294-297.) that for a one stage procedure

$$P(\theta_1 | \bar{x}_1) \text{ is } N \left( \frac{h_{10} \bar{x}_0 + h_{11} \bar{x}_1}{h_{10} + h_{11}}, \frac{1}{h_{10} + h_{11}} \right) \text{ and}$$

$$P(\bar{x}_1) \text{ is } N \left( \bar{x}_0, \frac{h_{10} + h_{11}}{h_{10} h_{11}} \right). \text{ Then, if we go to the second}$$

stage and now take  $P(\theta_1 | \bar{x}_1)$  as the prior distribution, we have that  $P(\theta_1 | \bar{x}_1, \bar{x}_2) = P(\theta_1 | \bar{x}_1) | \bar{x}_2$  is

$$N \left( \frac{(h_{10} + h_{11}) \frac{h_{10} \bar{x}_0 + h_{11} \bar{x}_1}{h_{10} + h_{11}} + h_{12} \bar{x}_2}{h_{10} + h_{11} + h_{12}}, \frac{1}{h_{10} + h_{11} + h_{12}} \right)$$

$$\text{or } N \left( \frac{h_{10} \bar{x}_0 + h_{11} \bar{x}_1 + h_{12} \bar{x}_2}{h_{10} + h_{11} + h_{12}}, \frac{1}{h_{10} + h_{11} + h_{12}} \right). \text{ Then}$$

at the  $k$ 'th stage we have that

$$P(\theta_1 | \bar{x}_1, \dots, \bar{x}_k) \text{ is } N \left( \frac{\sum_{i=0}^k h_{1i} \bar{x}_i}{\sum_{i=0}^k h_{1i}}, \frac{1}{\sum_{i=0}^k h_{1i}} \right). \text{ Also, at}$$

the second stage by assuming  $P(\theta_1 | \bar{x}_1)$  is the new prior distribution we have that

$P(\bar{x}_2|\bar{x}_1)$  is  $N\left(\frac{h_{10}\bar{x}_0 + h_{11}\bar{x}_1}{h_{10} + h_{11}}, \frac{h_{10} + h_{11} + h_{12}}{h_{12}(h_{10} + h_{11})}\right)$ . Then

at the  $k$ 'th stage we have that  $P(\bar{x}_k|\bar{x}_1, \dots, \bar{x}_{k-1})$  is

$$N\left(\frac{\sum_{i=0}^{k-1} h_{1i} \bar{x}_i}{\sum_{i=0}^{k-1} h_{1i}}, \frac{\sum_{i=0}^k h_{1i}}{h_{1k} \sum_{i=0}^{k-1} h_{1i}}\right).$$

Similar formulas for and statements about

$P(\bar{y}_k|\bar{y}_1, \dots, \bar{y}_{k-1})$  and  $P(\theta_2|\bar{y}_1, \dots, \bar{y}_k)$  also hold.

Thus we see that, if we perform the analysis for the stage-by-stage procedure for normal random variables using the loss functions defined by (2.1) and (2.7), we can again assume that we have a one stage procedure at each stage and that we are merely altering the prior distribution of  $\theta_1$  and  $\theta_2$  as we proceed from stage to stage.

## 5.2 Procedure for a Linear Loss Function

On the basis of the results of Section 5.1 we shall assume we have a one-stage procedure at each stage with  $n_{11}$  on Population 1 and  $n_{21}$  on Population 2, where  $n_{11} + n_{21} = n_1$ , a constant. We also shall assume that the sample means for Population 1 and Population 2 are  $\bar{x}_1$  and  $\bar{x}_2$ , where  $P_{\theta_1}(\bar{x}_1)$  is  $N(\theta_1, \sigma_1^2/n_{11})$  and  $P_{\theta_2}(\bar{y}_1)$  is  $N(\theta_2, \sigma_2^2/n_{21})$ , and

that the prior distributions of  $\theta_1$  and  $\theta_2$  are  $\rho_1(\theta_1)$  and  $\rho_2(\theta_2)$ , where  $\rho_1(\theta_1)$  is  $N(\bar{x}_0, \sigma_1^2/n_{10})$  and  $\rho_2(\theta_2)$  is  $N(\bar{y}_0, \sigma_2^2/n_{20})$ . Let  $\theta = (\theta_1, \theta_2)$ ,  $x_1 = (\bar{x}_1, \bar{y}_1)$ ,  $h_{11} = n_{11}/\sigma_1^2$ ,  $h_{21} = n_{21}/\sigma_2^2$ ,  $h_{10} = n_{10}/\sigma_1^2$ , and  $h_{20} = n_{20}/\sigma_2^2$ . We shall assume that  $\rho(\theta) = \rho_1(\theta_1) \rho_2(\theta_2)$  and that the decisions  $a_1$  and  $b$  are defined as in Section 2.1. Then from the previous section

$$(5.1) \quad P(\theta_1 | \bar{x}_1) \text{ is } N \left( \frac{h_{11} \bar{x}_1 + h_{10} \bar{x}_0}{h_{11} + h_{10}}, \frac{1}{h_{10} + h_{11}} \right),$$

$$P(\theta_2 | \bar{y}_1) \text{ is } N \left( \frac{h_{21} \bar{y}_1 + h_{20} \bar{y}_0}{h_{21} + h_{20}}, \frac{1}{h_{20} + h_{21}} \right),$$

$$P(\bar{x}_1) \text{ is } N \left( \bar{x}_0, \frac{h_{10} + h_{11}}{h_{10} h_{11}} \right),$$

and

$$P(\bar{y}_1) \text{ is } N \left( \bar{y}_0, \frac{h_{20} + h_{21}}{h_{20} h_{21}} \right).$$

We shall in this section let

$$(5.2) \quad L(x_1, a_1, b, \theta) = \begin{cases} k_{10} + k_{11} \theta_1 + k_{12} \theta_2 & \text{for } b=1 \\ k_{20} + k_{21} \theta_1 + k_{22} \theta_2 & \text{for } b=2 \end{cases}.$$

We would like to find the optimum  $a_1$  and, if we are at the last stage, the optimum  $b$  for given  $a_1$ .

We shall assume without loss of generality that

$$(5.3) \quad k_{10} + k_{11} \bar{x}_0 + k_{12} \bar{y}_0 \geq k_{20} + k_{21} \bar{x}_0 + k_{22} \bar{y}_0.$$



Then we shall let

$$(5.4) \quad L^*(x_1, a_1, b, \theta) = \begin{cases} k_0 + k_1 \theta_1 + k_2 \theta_2 & \text{for } b=1 \\ 0 & \text{for } b=2 \end{cases},$$

where  $k_j = k_{1j} - k_{2j}$  for  $j = 0, 1, 2$ . Then

$$(5.5) \quad E_{\theta|x_1} L^*(x_1, a_1, b, \theta) = \begin{cases} M(\bar{x}_1, \bar{y}_1) = k_0 + k_1 \left( \frac{h_{11} \bar{x}_1 + h_{10} \bar{x}_0}{h_{11} + h_{10}} \right) \\ \quad + k_2 \left( \frac{h_{21} \bar{y}_1 + h_{20} \bar{y}_0}{h_{21} + h_{20}} \right) & \text{for } b=1 \\ 0 & \text{for } b=2 \end{cases}.$$

Thus

$$\min_b E_{\theta|x_1} L^*(x_1, a_1, b, \theta) = \min \{M(\bar{x}_1, \bar{y}_1), 0\}.$$

Then

$$E_{x_1} \min_b E_{\theta|x_1} L^*(x_1, a_1, b, \theta) = \int_{-\infty}^0 M P(M) dM,$$

where  $P(M)$  is  $N(E_{x_1} \{M(\bar{x}_1, \bar{y}_1)\}, \text{Var} (M(\bar{x}_1, \bar{y}_1)))$

with

$$(5.6) \quad E_{x_1} (M(\bar{x}_1, \bar{y}_1)) = k_0 + k_1 \left( \frac{h_{11} E_{x_1}(\bar{x}_1) + h_{10} \bar{x}_0}{h_{11} + h_{10}} \right) \\ + k_2 \left( \frac{h_{21} E_{y_1}(\bar{y}_1) + h_{20} \bar{y}_0}{h_{21} + h_{20}} \right) \\ = k_0 + k_1 \bar{x}_0 + k_2 \bar{y}_0$$

and

$$\begin{aligned}
 (5.7) \quad \text{Var}(M(\bar{x}_1, \bar{y}_1)) &= k_1^2 \frac{h_{11}^2}{(h_{11}+h_{10})^2} \text{Var}(\bar{x}_1) \\
 &+ k_2^2 \frac{h_{21}^2}{(h_{21}+h_{20})^2} \text{Var}(\bar{y}_1) \\
 &= k_1^2 \frac{h_{11}^2}{(h_{11}+h_{10})^2} \left( \frac{h_{10}+h_{11}}{h_{10}h_{11}} \right) \\
 &+ k_2^2 \frac{h_{21}^2}{(h_{21}+h_{20})^2} \left( \frac{h_{20}+h_{21}}{h_{20}h_{21}} \right) \\
 &= k_1^2 \frac{h_{11}}{h_{10}(h_{10}+h_{11})} + k_2^2 \frac{h_{21}}{h_{20}(h_{20}+h_{21})}
 \end{aligned}$$

since  $M$  is a linear combination of the normally and independently distributed random variables  $\bar{x}_1$  and  $\bar{y}_1$ , whose distributions are given by (5.1).

Let  $\phi(z) = \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}$ . Then

$$\begin{aligned}
 (5.8) \quad E_{x_1} \min_b E_{\theta|x_1} L^*(x_1, a_1, b, \theta) \\
 &= - \int_0^\infty \frac{1}{\sqrt{2\pi}} \sqrt{h} e^{-\frac{1}{2}(-M-E(M))^2} dM \\
 &= - \frac{1}{\sqrt{h}} \int_{\sqrt{h}E(M)}^\infty (z - \sqrt{h}E(M)) \phi(z) dz \quad ,
 \end{aligned}$$

where  $E(M) = k_0 + k_1 \bar{x}_0 + k_2 \bar{x}_2 \geq 0$  and  $h = \frac{1}{\text{Var}(M(\bar{x}_1, \bar{y}_1))}$ .

In order to find the values of  $n_{11}$  and  $n_{21}$  subject to  $n_{11} + n_{21} = n_1$  which minimize (5.8) we must find the values which minimize  $h$ , or maximize  $\text{Var}(M(\bar{x}_1, \bar{y}_1))$ . From (5.7)

$$\begin{aligned} \text{Var}(M(\bar{x}_1, \bar{y}_1)) &= \sigma_1^2 k_1^2 \frac{n_{11}}{n_{10}(n_{10}+n_{11})} + \sigma_2^2 k_2^2 \frac{n_{21}}{n_{20}(n_{20}+n_{21})} \\ &= \frac{\sigma_1^2 k_1^2}{n_{10}} + \frac{\sigma_2^2 k_2^2}{n_{20}} - \left( \frac{\sigma_1^2 k_1^2}{n_{10}+n_{11}} + \frac{\sigma_2^2 k_2^2}{n_{20}+n_{21}} \right). \end{aligned}$$

This is maximized for  $n_{11} + n_{21} = n_1$  fixed when

$$(5.9) \quad \frac{n_{10} + n_{11}}{n_{20} + n_{21}} = \left| \frac{k_1}{k_2} \right| \frac{\sigma_1}{\sigma_2}.$$

There still remains the question of whether or not using  $L^*(x_1, a_1, b, \theta)$  is equivalent to using  $L(x_1, a_1, b, \theta)$ . From (5.2), (5.3), and (5.4)

$$L^*(x_1, a_1, b, \theta) = L(x_1, a_1, b, \theta) - (k_{20} + k_{21} \theta_1 + k_{22} \theta_2).$$

Thus

$$\begin{aligned} E_{\theta|x_1} L^*(x_1, a_1, b, \theta) &= E_{\theta|x_1} L(x_1, a_1, b, \theta) \\ &- \left[ k_{20} + k_{21} \left( \frac{h_{11} \bar{x}_1 + h_{10} \bar{x}_0}{h_{10} + h_{11}} \right) + k_{22} \left( \frac{h_{21} \bar{y}_1 + h_{20} \bar{y}_0}{h_{20} + h_{21}} \right) \right] \end{aligned}$$

Therefore,

$$\begin{aligned} (5.10) \quad E_{x_1} \min_b E_{\theta|x} L^*(x_1, a_1, b, \theta) \\ = E_{x_1} \min_b E_{\theta|x_1} L(x_1, a_1, b, \theta) - (k_{20} + k_{21} \bar{x}_0 + k_{22} \bar{y}_0). \end{aligned}$$

We see that the same values of  $n_{11}$  and  $n_{21}$  minimize  $E_{x_1} \min_b E_{\theta|x_1} L(x_1, a_1, b, \theta)$  as those which minimized  $E_{x_1} \min_b E_{\theta|x_1} L'(x_1, a_1, b, \theta)$ .

Finally, if we are at the last stage and wish to decide whether to take  $b = 1$  or  $b = 2$  as the terminal decision, we see from (5.5) that we should take  $b = 1$  if  $M(\bar{x}_1, \bar{y}_1) < 0$ , either  $b = 1$  or  $b = 2$  if  $M(\bar{x}_1, \bar{y}_1) = 0$ , and  $b = 2$  if  $M(\bar{x}_1, \bar{y}_1) > 0$ .

It should be noted that in Chapter 5 of their book Raiffa and Schlaifer (1961) derive results for a more general problem than the one considered in this section but for a somewhat more restricted linear utility function. (The utility of choosing  $b = 1$  would be linear in  $\theta_1$  and independent of  $\theta_2$  in their work.) However, with slight changes in their arguments one can obtain the results of this section as a special case of their results. It is felt, however, that the derivation of this section, which uses some of their ideas, is more straightforward for this problem.

### 5.3 Procedure for a Constant Loss Function

In this section we shall assume that the assumptions and notation of Section 5.2 hold except for the loss function. We shall let

$$L(x_1, a_1, b, \theta) = \left\{ \begin{array}{l} 1 \text{ if } b = 1 \text{ and } \theta_1 < \theta_2 \\ 1 \text{ if } b = 2 \text{ and } \theta_1 > \theta_2 \\ 0 \text{ otherwise} \end{array} \right\} .$$

Then

$$E_{\theta|x_1} L(x_1, a_1, b, \theta) = \left\{ \begin{array}{l} \int_{-\infty}^{\infty} \int_{-\infty}^{\theta_2} P(\theta_1|\bar{x}_1) P(\theta_2|\bar{y}_1) d\theta_1 d\theta_2 \text{ for } b = 1 \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\theta_1} P(\theta_1|\bar{x}_1) P(\theta_2|\bar{x}_2) d\theta_2 d\theta_1 \text{ for } b = 2 \end{array} \right\} .$$

Let  $v = \theta_1 - \theta_2$ . Then  $P(v|x_1)$  is  $N(\mu_v, \sigma_v^2)$ , where from (5.1)

$$(5.11) \quad \mu_v = \frac{h_{11} \bar{x}_1 + h_{10} \bar{x}_0}{h_{11} + h_{10}} - \frac{h_{21} \bar{y}_1 + h_{20} \bar{y}_0}{h_{21} + h_{20}}$$

and

$$\sigma_v^2 = \frac{1}{h_{10} + h_{11}} + \frac{1}{h_{20} + h_{21}} = \frac{\sigma_1^2}{n_{10} + n_{11}} + \frac{\sigma_2^2}{n_{20} + n_{21}} .$$

Then

$$E_{\theta|x_1} L(x_1, a_1, b, \theta) = \left\{ \begin{array}{l} \int_{-\infty}^0 P(v|x_1) dv \text{ for } b=1 \\ \int_0^{\infty} P(v|x_1) dv \text{ for } b=2 \end{array} \right\} .$$

Let  $z_1 = (v - \mu_v)/\sigma_v$ . Then

$$(5.12) \quad E_{\theta|x_1} L(x_1, a_1, b, \theta) = \left\{ \begin{array}{l} \int_{-\infty}^{-\frac{\mu_v}{\sigma_v}} \phi(z_1) dz_1 \text{ for } b=1 \\ \int_{-\frac{\mu_v}{\sigma_v}}^{\infty} \phi(z_1) dz_1 \text{ for } b=2 \end{array} \right\}$$

and

$$\min_b E_{\theta|x_1} L(x_1, a_1, b, \theta) = \int_{-\infty}^{-\left| \frac{\mu_v}{\sigma_v} \right|} \phi(z_1) dz_1 .$$

Thus

$$\begin{aligned} & E_{x_1} \min_b E_{\theta|x_1} L(x_1, a_1, b, \theta) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(\bar{x}_1) P(\bar{y}_1) \int_{-\infty}^{-\left| \frac{\mu_v}{\sigma_v} \right|} \phi(z_1) dz_1 d\bar{x}_1 d\bar{y}_1 . \end{aligned}$$

$$\text{Let } R = \frac{\mu_v}{\sigma_v} = \frac{1}{\sigma_v} \left( \frac{h_{11} \bar{x}_1 + h_{10} \bar{x}_0}{h_{11} + h_{10}} - \frac{h_{21} \bar{y}_1 + h_{20} \bar{y}_0}{h_{21} + h_{20}} \right) .$$

Then  $P(R)$  is  $N(\mu_R, \sigma_R^2)$  where from (5.1) and (5.11)

$$(5.13) \quad \mu_R = (\bar{x}_0 - \bar{y}_0) / \sigma_v$$

and

$$\sigma_v^2 \sigma_R^2 = \left( \frac{h_{11}}{h_{10} + h_{11}} \right)^2 \frac{h_{10} + h_{11}}{h_{10} h_{11}}$$

$$\begin{aligned}
 & + \left( \frac{h_{21}}{h_{21} + h_{20}} \right)^2 \frac{h_{20} + h_{21}}{h_{20} h_{21}} \\
 = & \frac{n_{11}}{n_{10}} \left( \frac{\sigma_1^2}{n_{10} + n_{11}} \right) + \frac{n_{21}}{n_{20}} \left( \frac{\sigma_2^2}{n_{20} + n_{21}} \right) \\
 = & \frac{\sigma_1^2}{n_{10}} + \frac{\sigma_2^2}{n_{20}} - \left( \frac{\sigma_1^2}{n_{10} + n_{11}} + \frac{\sigma_2^2}{n_{20} + n_{21}} \right) \\
 = & K - \sigma_v^2,
 \end{aligned}$$

where  $K = \frac{\sigma_1^2}{n_{10}} + \frac{\sigma_2^2}{n_{20}}$  is a constant at any given stage since we are only going to let  $n_{11}$  and  $n_{21}$  vary subject to  $n_{11} + n_{21} = n_1$ .

Thus

$$(5.14) \quad \sigma_R^2 = \frac{K}{\sigma_v^2} - 1$$

and

$$\begin{aligned}
 & E_{x_1} \min_b E_{\theta|x_1} L(x_1, a_1, b, \theta) \\
 = & \int_{-\infty}^{+\infty} P(R) \int_{-\infty}^{-|R|} \phi(z_1) dz_1 dR.
 \end{aligned}$$

In the rest of this section notation and arguments similar to those presented in the Tables of the Bivariate Normal Distribution Function (1959, pp. XXVIII - XXXII) will be used.

Let  $z_2 = (R - \mu_R)/\sigma_R$  and let  $c = \sigma_R$  and  $d = \mu_R$ . Also,

let  $\phi(z_1) \phi(z_2) = \frac{e^{-\frac{1}{2}(z_1^2+z_2^2)}}{2\pi}$  . Then

$$\begin{aligned}
 & \mathbb{E}_{x_1} \min_b \mathbb{E}_{\theta|x_1} L(x_1, a_1, b, \theta) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-|cz_2+d|} \phi(z_1) \phi(z_2) dz_1 dz_2 \\
 &= \int_{-\infty}^{-\frac{d}{c}} \int_{-\infty}^{cz_2+d} \phi(z_1) \phi(z_2) dz_1 dz_2 \\
 &+ \int_{-\frac{d}{c}}^{\infty} \int_{-\infty}^{-(cz_2+d)} \phi(z_1) \phi(z_2) dz_1 dz_2 \\
 &= \frac{1}{2} \int_{-\infty}^{-\frac{d}{c}} \int_0^{-(cz_2+d)} \phi(z_1) \phi(z_2) dz_1 dz_2 \\
 &- \int_{-\frac{d}{c}}^{\infty} \int_0^{cz_2+d} \phi(z_1) \phi(z_2) dz_1 dz_2 .
 \end{aligned}$$

In all cases  $c > 0$ , but in general  $d$  can be any real number. Without loss of generality we let  $d < 0$ . (The arguments for  $d > 0$  are similar to what follows.) Let

$$I_1 = \int_{-\infty}^{-\frac{d}{c}} \int_0^{-(cz_2+d)} \phi(z_1) \phi(z_2) dz_1 dz_2$$



and let  $I_2 = \int_{-\frac{d}{c}}^{\infty} \int_0^{cz_2+d} \phi(z_1) \phi(z_2) dz_1 dz_2$  . For

$d < 0$  the areas over which  $I_1$  and  $I_2$  extend, along with convenient auxiliary lines, are shown in Figure 2. The shaded area  $A_1$  is the area over which  $I_1$  extends and  $A_2$  is the area over which  $I_2$  extends.

In Figure 2 we let  $h = \frac{-d}{\sqrt{1+c^2}}$  and  $k = \frac{-d}{c\sqrt{1+c^2}}$  .

Let  $z_1^i, z_1^{\prime\prime}, z_2^i,$  and  $z_2^{\prime\prime}$  be the  $z_1$  and  $z_2$  axes suitably ro-

tated. Let  $V(h, k) = \int_0^h \int_0^{\frac{k}{h}z_2} \phi(z_1) \phi(z_2) dz_1 dz_2$  .

Then from Figure 2 we see that

$$\begin{aligned} & E_{x_1} \min_b E_{\theta|x_1} L(x_1, a_1, b, \theta) \\ &= \frac{1}{2} - \left( \frac{\arctan \frac{h}{k}}{2\pi} + V(h, k) + \int_0^{\infty} \int_0^h \phi(z_1^i) \phi(z_2^i) dz_1^i dz_2^i \right) \\ & \quad - \left( \frac{\arctan \frac{h}{k}}{2\pi} + V(h, k) - \int_0^{\infty} \int_0^h \phi(z_1^{\prime\prime}) \phi(z_2^{\prime\prime}) dz_1^{\prime\prime} dz_2^{\prime\prime} \right) \\ &= \frac{1}{2} - 2 \left( \frac{\arctan \frac{h}{k}}{2\pi} + V(h, k) \right) . \end{aligned}$$

For fixed  $h_{10}, h_{20},$  and  $\bar{x}_0 - \bar{y}_0,$   $h$  is a constant since from (5.13) and (5.14)

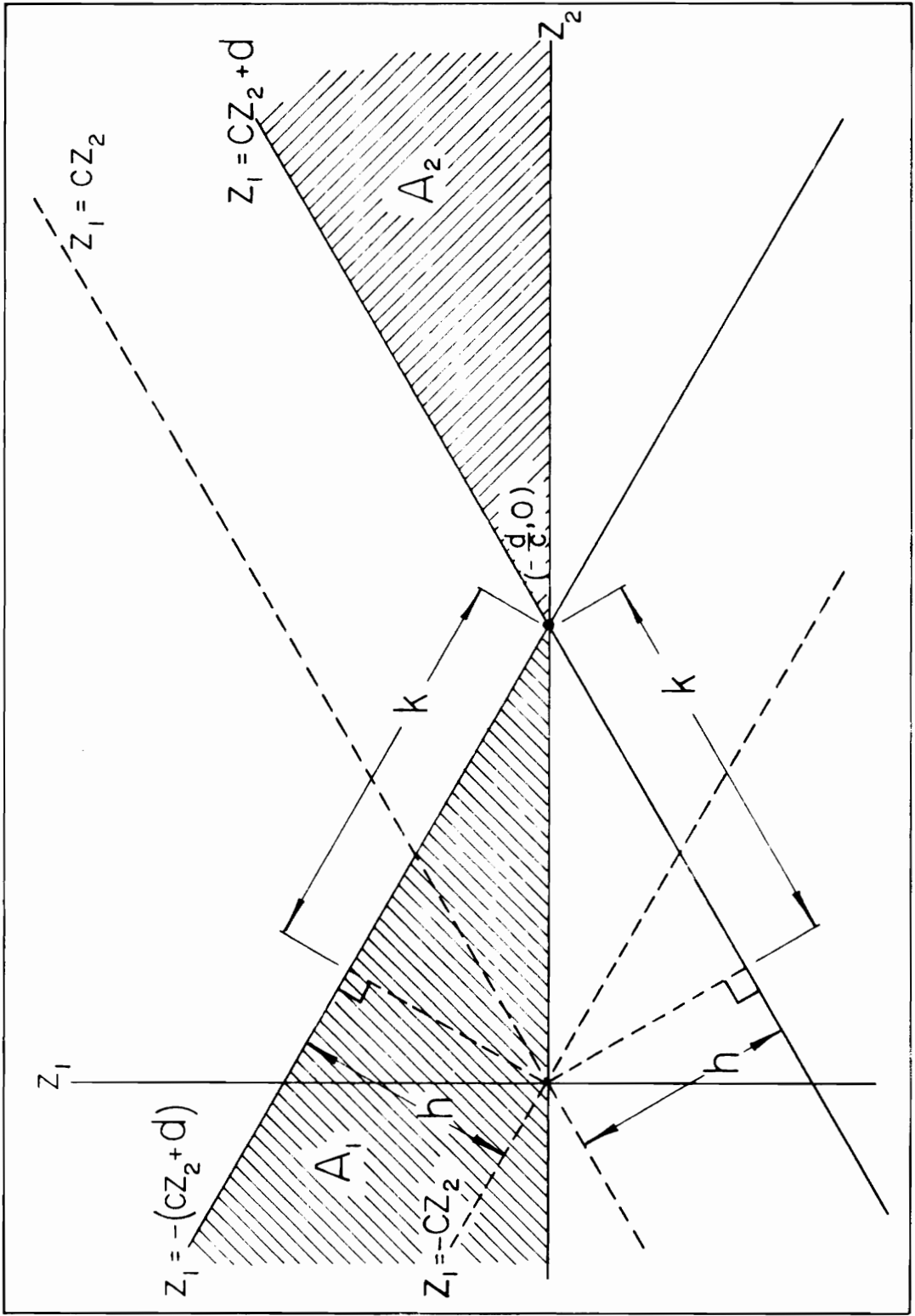


FIGURE 2 : AREAS OVER WHICH INTEGRALS ARE EVALUATED WITH AUXILIARY LINES

$$h = \frac{-d}{\sqrt{1+c^2}} = \frac{-\mu_R}{\sqrt{1+\sigma_R^2}} = \frac{\frac{\bar{x}_0 - \bar{y}_0}{\sigma_V}}{\sqrt{1 + \frac{K}{\sigma_V^2} - 1}} = \frac{\bar{x}_0 - \bar{y}_0}{\sqrt{K}} .$$

Thus we shall decrease  $E_{x_1} \min_b E_{\theta|x_1} L(x_1, a_1, b, \theta)$  by increasing  $V(h, k) + \frac{\arctan \frac{h}{k}}{2\pi}$  and this will depend only on values of  $k$ .

Consider, without loss of generality, the case in which  $h > k$  with  $h$  fixed as shown in Figure 3. In Figure 3 let  $\overline{AH} = \overline{HG} = \overline{CG} = \overline{AC} = h$  and let  $\overline{AF} = \overline{HE} = \overline{CB} = k$ . Then the volume above triangle ABC is  $V(h, k)$  and the volume above the wedge DAJ is  $\frac{\arctan \frac{h}{k}}{2\pi}$ . Decrease  $k$  by  $\Delta k$  so that  $\overline{AF}' = \overline{HE}' = \overline{CB}' = k - \Delta k$ . The volume above triangle  $ABB'$  is the net decrease in  $V(h, k)$  resulting from a decrease in  $k$  of  $\Delta k$  units. The volume above the wedge  $D'AD$  is the corresponding increase in  $\frac{\arctan h/k}{2\pi}$ . Since Angle  $BAB' =$  Angle  $D'AD$  and rotation of areas does not affect the volumes above them, it is seen that there is a net increase in  $\frac{\arctan h/k}{2\pi} + V(h, k)$  for a decrease in  $k$ .

Thus in order to maximize  $V(h, k) + \frac{\arctan h/k}{2\pi}$  one must minimize  $k = h/c = h/\sigma_R$ , which means maximizing  $\sigma_R$ . However,

$$\sigma_R^2 = \frac{K}{\sigma_V^2} - 1 \text{ so we must minimize } \sigma_V^2 . \quad \sigma_V^2 = \frac{1}{h_{10} + h_{11}} + \frac{1}{h_{20} + h_{21}}$$

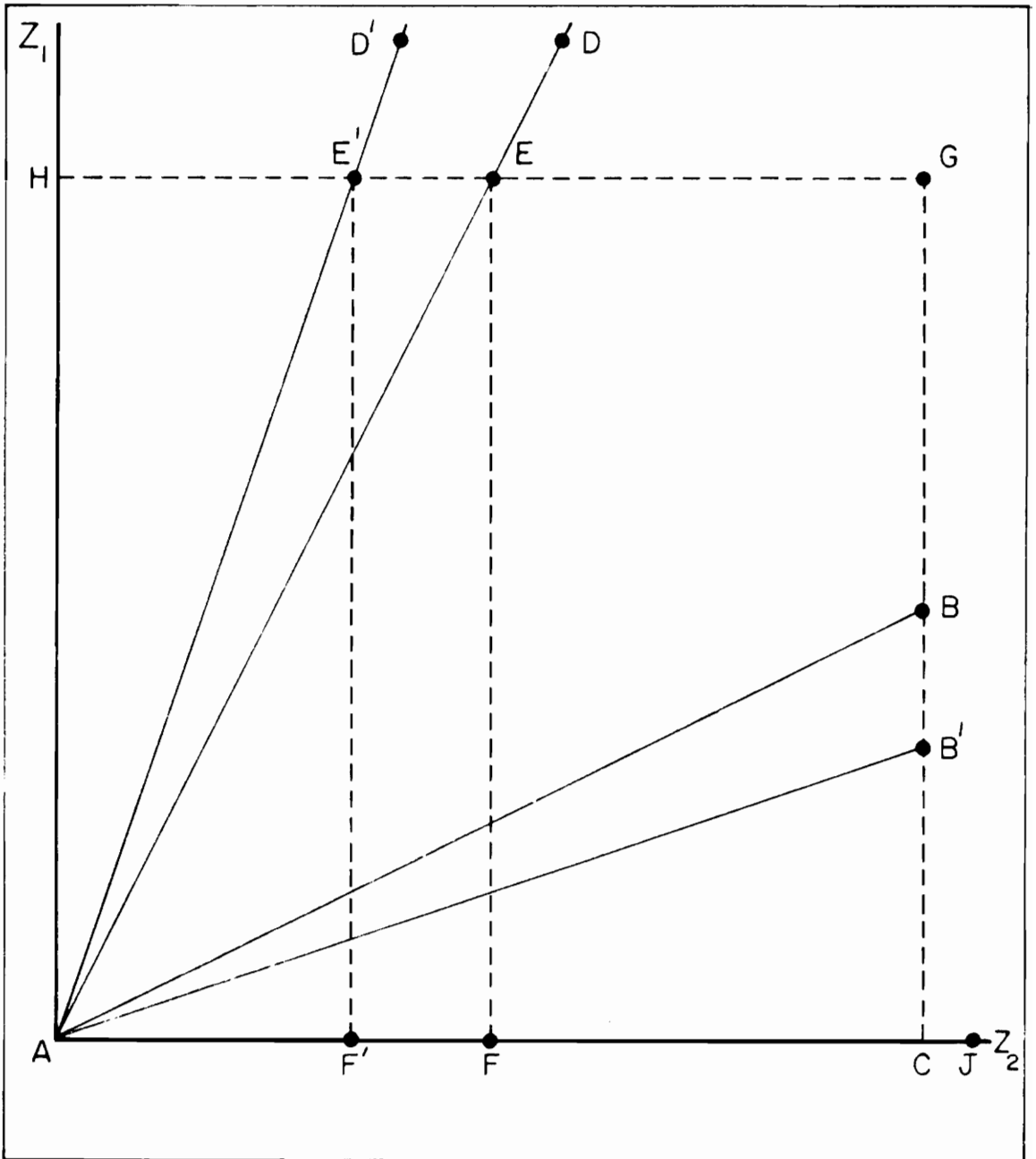


FIGURE 3

INCREASE IN  $\frac{\text{ARC TAN } \frac{h}{k}}{2\pi} + V(h,k)$  FOR  
A GIVEN DECREASE IN  $k$

$$= \frac{\sigma_1^2}{n_{10} + n_{11}} + \frac{\sigma_2^2}{n_{20} + n_{21}}, \text{ which is minimized for } n_{11} + n_{21} = n_1,$$

a constant, when

$$(5.15) \quad \frac{n_{10} + n_{11}}{n_{20} + n_{21}} = \frac{\sigma_1}{\sigma_2} .$$

It should be noted that this is the same result which one obtains from (5.9) by making  $|k_1/k_2| = 1$ .

Also, if we are at the last stage of our stage-by-stage procedure, from (5.12) we see that we take  $b=1$  when  $\mu_v > 0$ , take either  $b=1$  or  $b=1$  when  $\mu_v = 0$ , and take  $b=2$  when  $\mu_v < 0$ .

It is interesting to compare the results in (5.15) and the results of (5.9) when  $|k_1/k_2| = 1$  with the situation in which one is making the usual test of hypothesis that  $\theta_1 > \theta_2$  against  $\theta_1 < \theta_2$  when the test statistic, the difference of the two sample means, is normally distributed. In this case one computes the standardized normal variable

$$z = \frac{\bar{x}_1 - \bar{y}_1 - (\theta_1 - \theta_2)}{\frac{\sigma_1^2}{n_{11}} + \frac{\sigma_2^2}{n_{21}}} \quad (\text{taking } \theta_1 - \theta_2 = 0)$$

and rejects the hypothesis that  $\theta_1 > \theta_2$  if  $z < z_0$ , where  $z_0$  is a fixed constant. The power of this test for a given

$$\theta_1 - \theta_2 \text{ is } \Pr \left[ z < \left( z_0 - \frac{(\theta_1 - \theta_2)}{\frac{\sigma_1^2}{n_{11}} + \frac{\sigma_2^2}{n_{21}}} \right) \right]. \text{ It is obvious that the}$$

power is maximized for each  $\theta_1 - \theta_2 < 0$  if  $\frac{\sigma_1^2}{n_{11}} + \frac{\sigma_2^2}{n_{21}}$  is a minimum, or if  $n_{11}/n_{21} = \sigma_1/\sigma_2$ . (5.9) (for  $|k_1/k_2| = 1$ ) and (5.15) give the same result if  $n_{10} = n_{20} = 0$ , or if we assume the prior distributions of  $\theta_1$  and  $\theta_2$  have "infinite" variances.

VI. THEORETICAL COMPARISON OF  
EXTENSIVE-FORM ANALYSIS, THE STAGE-BY-STAGE  
PROCEDURE AND THE APPROXIMATE PROCEDURE

6.1 Asymptotic Comparison of the Stage-by-Stage Procedure  
and Extensive-Form Analysis

From (2.5) we have that

$$\begin{aligned}
 P(x_k | x_1, \dots, x_{k-1}) &= \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}} \\
 &\cdot \frac{B(\alpha_k + 1, \beta_k + 1) B(\gamma_k + 1, \delta_k + 1)}{B(\alpha_{k-1} + 1, \beta_{k-1} + 1) B(\gamma_{k-1} + 1, \delta_{k-1} + 1)} \\
 &= \frac{(\alpha_{k-1} + r_{1k}) \dots (\alpha_{k-1} + 1) (\beta_{k-1} + n_{1k} - r_{1k}) \dots (\beta_{k-1} + 1)}{(\alpha_{k-1} + \beta_{k-1} + 1 + n_{1k}) \dots (\alpha_{k-1} + \beta_{k-1} + 1)} \binom{n_{1k}}{r_{1k}} \\
 &\cdot \frac{(\gamma_{k-1} + r_{2k}) \dots (\gamma_{k-1} + 1) (\delta_{k-1} + n_{2k} - r_{2k}) \dots (\delta_{k-1} + 1)}{(\gamma_{k-1} + \delta_{k-1} + 1 + n_{2k}) \dots (\gamma_{k-1} + \delta_{k-1} + 1)} \binom{n_{2k}}{r_{2k}}
 \end{aligned}$$

since (for example)

$$(\gamma_{k-1} + r_{2k}) \dots (\gamma_{k-1} + 1) = \frac{(\gamma_{k-1} + r_{2k})!}{\gamma_{k-1}!} = \frac{\gamma_k!}{\gamma_{k-1}!}$$

Assume for  $k$  becoming large

$$\frac{\alpha_{k-1}}{\alpha_{k-1} + \beta_{k-1}} \rightarrow \theta_1, \quad \frac{\gamma_{k-1}}{\gamma_{k-1} + \beta_{k-1}} \rightarrow \theta_2,$$

$$\frac{\beta_{k-1}}{\alpha_{k-1} + \beta_{k-1}} \rightarrow 1 - \theta_1, \quad \text{and} \quad \frac{\delta_{k-1}}{\delta_{k-1} + \gamma_{k-1}} \rightarrow 1 - \theta_2.$$

Assume, further, that for  $k$  large  $\sum_{i=k}^m n_{1k}$  is small relative to  $\alpha_{k-1}$  and  $\beta_{k-1}$  and that  $\sum_{i=k}^m n_{2i}$  is small relative to  $\gamma_{k-1}$  and  $\delta_{k-1}$ . Then

$$(6.1) \quad P(x_k | x_1, \dots, x_{k-1}) \rightarrow \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}} \\ \cdot \theta_1^{r_{1k}} (1-\theta_1)^{n_{1k}-r_{1k}} \theta_2^{r_{2k}} (1-\theta_2)^{n_{2k}-r_{2k}} .$$

Also, for the loss function defined by (2.7)

$$E_{\theta} | x_1, \dots, x_m L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) \\ = \left\{ \begin{array}{l} k_{10} + k_{11} \left( \frac{\alpha_m + 1}{\alpha_m + \beta_m + 2} \right) + k_{12} \left( \frac{\gamma_m + 1}{\gamma_m + \delta_m + 2} \right) \text{ for } b = 1 \\ k_{20} + k_{21} \left( \frac{\alpha_m + 1}{\alpha_m + \beta_m + 2} \right) + k_{22} \left( \frac{\gamma_m + 1}{\gamma_m + \delta_m + 2} \right) \text{ for } b = 2 \end{array} \right\} \\ \rightarrow \left\{ \begin{array}{l} k_{10} + k_{11} \theta_1 + k_{12} \theta_2 \text{ for } b = 1 \\ k_{20} + k_{21} \theta_1 + k_{22} \theta_2 \text{ for } b = 2 \end{array} \right\}$$

for  $m$  large, which means that from Section 2.4 and (6.1)



$$\begin{aligned}
 & E_{x_m | x_1, \dots, x_{m-1}} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \\
 & \rightarrow \sum_{r_{1m}=0}^{n_{2m}} \sum_{r_{2m}=0}^{n_{2m}} \binom{n_{1m}}{r_{1m}} \binom{n_{2m}}{r_{2m}} \theta_1^{r_{1m}} (1-\theta_1)^{n_{1m}-r_{1m}} \\
 & \cdot \theta_2^{r_{2m}} (1-\theta_2)^{n_{2m}-r_{2m}} \min(k_{10}+k_{11}\theta_1+k_{12}\theta_2, k_{20}+k_{21}\theta_1+k_{22}\theta_2) \\
 & = \min(k_{10}+k_{11}\theta_1+k_{12}\theta_2, k_{20}+k_{21}\theta_1+k_{22}\theta_2) \cdot
 \end{aligned}$$

Similarly we can show that for k large

$$\begin{aligned}
 & E_{x_k | x_1, \dots, x_{k-1}} L^{(m-k+1)}(x_1, \dots, x_{k-1}, x_k, a_1, \dots, a_k) \\
 & \rightarrow \min(k_{10}+k_{11}\theta_1+k_{12}\theta_2, k_{20}+k_{21}\theta_1+k_{22}\theta_2) \cdot
 \end{aligned}$$

Also, for k large and for a linear loss function

$$\begin{aligned}
 & E_{x_k | x_1, \dots, x_{k-1}} \min_b E_{\theta | x_1, \dots, x_k} L(x_1, \dots, x_k, a_1, \dots, a_k, b, \theta) \\
 & = \sum_{r_{1k}=0}^{n_{1k}} \sum_{r_{2k}=0}^{n_{2k}} \binom{n_{1k}}{r_{1k}} \binom{n_{2k}}{r_{2k}} \frac{B(\alpha_k + 1, \beta_k + 1)}{B(\alpha_{k-1} + 1, \beta_{k-1} + 1)} \\
 & \cdot \frac{B(\gamma_k + 1, \delta_k + 1)}{B(\gamma_{k-1} + 1, \delta_{k-1} + 1)} \min \left[ k_{10} + k_{11} \left( \frac{\alpha_k + 1}{\alpha_k + \beta_k + 2} \right) \right. \\
 & \left. + k_{12} \frac{\gamma_k + 1}{\gamma_k + \delta_k + 2}, k_{20} + k_{21} \left( \frac{\alpha_k + 1}{\alpha_k + \beta_k + 2} \right) + k_{22} \left( \frac{\gamma_k + 1}{\gamma_k + \delta_k + 2} \right) \right] \\
 & \rightarrow \min(k_{10}+k_{11}\theta_1+k_{12}\theta_2, k_{20}+k_{21}\theta_1+k_{22}\theta_2) \cdot
 \end{aligned}$$

This means that asymptotically (k large with  $\frac{\alpha_{k-1}}{\alpha_{k-1} + \beta_{k-1}} \rightarrow \theta_1$ )

and  $\frac{\gamma_{k-1}}{\gamma_{k-1} + \delta_{k-1}} \rightarrow \theta_2$ ) extensive-form analysis and the stage-by-stage procedure are the same in that

$$\begin{aligned} & E_{x_k | x_1, \dots, x_{k-1}} \min_b E_{\theta | x_1, \dots, x_k} L(x_1, \dots, x_k, a_1, \dots, a_k, b, \theta) \\ &= E_{x_k | x_1, \dots, x_{k-1}} L^{(m-k+1)}(x_1, \dots, x_{k-1}, x_k, a_1, \dots, a_k) \end{aligned}$$

for a linear loss function.

### 6.2 Derivation of the Approximate Procedure as an Approximation to the Stage-by-Stage Procedure Assuming Normal Distributions

Let us assume a one stage procedure (at any given stage of a stage-by-stage procedure as described in Section 3.1). Let the prior distribution of  $\theta_1$  and  $\theta_2$ ,  $\rho(\theta)$ , be defined by (2.2). We shall then use the bivariate normal distribution

$$\rho'(\theta) = \sqrt{\frac{n_{10}}{\sigma_1^2} \cdot \frac{n_{20}}{\sigma_2^2}} \frac{e^{-\frac{1}{2} \left( \frac{(\theta_1 - \bar{x}_0)^2 n_{10}}{\sigma_1^2} + \frac{(\theta_2 - \bar{y}_0)^2 n_{20}}{\sigma_2^2} \right)}}{2\pi}$$

as an approximation for  $\rho(\theta)$  such that by equating the corresponding means and variances, we have

$$(6.2) \quad \bar{x}_0 = \frac{\alpha_0 + 1}{\alpha_0 + \beta_0 + 2} \quad , \quad \bar{y}_0 = \frac{\gamma_0 + 1}{\gamma_0 + \delta_0 + 2} \quad ,$$

$$\frac{\sigma_1^2}{n_{10}} = \frac{\frac{\alpha_0+1}{\alpha_0+\beta_0+2} \frac{\beta_0+1}{\alpha_0+\beta_0+2}}{\alpha_0+\beta_0+3} \quad , \quad \text{and}$$

$$\frac{\sigma_2^2}{n_{20}} = \frac{\frac{\gamma_0+1}{\gamma_0+\delta_0+2} \frac{\delta_0+1}{\gamma_0+\delta_0+2}}{\gamma_0+\delta_0+3}$$

so that

$$n_{10} = \alpha_0 + \beta_0 + 3 \quad , \quad n_{20} = \gamma_0 + \delta_0 + 3 \quad ,$$

$$\sigma_1^2 = \left( \frac{\alpha_0 + 1}{\alpha_0 + \beta_0 + 2} \right) \left( 1 - \frac{\alpha_0 + 1}{\alpha_0 + \beta_0 + 2} \right) \quad ,$$

and 
$$\sigma_2^2 = \left( \frac{\gamma_0 + 1}{\gamma_0 + \delta_0 + 2} \right) \left( 1 - \frac{\gamma_0 + 1}{\gamma_0 + \delta_0 + 2} \right) \quad .$$

In addition, we shall use the normal distributions  $N(\theta_1, \frac{\sigma_1^{*2}}{n_{11}})$  and  $N(\theta_2, \frac{\sigma_2^{*2}}{n_{21}})$  as normal approximations to the actual (binomial) distributions of  $r_{11}/n_{11} = \bar{x}_1$  and  $r_{21}/n_{21} = \bar{y}_1$ .

Actually, of course, if we do this,  $\sigma_1^{*2} = \theta_1(1 - \theta_1)$  and  $\sigma_2^{*2} = \theta_2(1 - \theta_2)$ . However, we shall use the approximations  $\sigma_1^{*2} = \sigma_1^2$  and  $\sigma_2^{*2} = \sigma_2^2$ , where  $\sigma_1^2$  and  $\sigma_2^2$  are given by (6.2). Note that we are assuming that  $\theta_1(1 - \theta_1)$  and  $\theta_2(1 - \theta_2)$  are "relatively constant" as  $\theta_1$  and  $\theta_2$  vary. This assumption will be appropriate when  $.2 \leq \theta_1 \leq .8$  and  $.2 \leq \theta_2 \leq .8$  but will not be good very far outside this range.

Using the above normal approximations, we can use the

results given by (5.9) and (5.15). Then, if we assume in addition that we are in the  $k$ 'th stage of a stage-by-stage

procedure, substitute  $\bar{p}_{1,k-1}$  for  $\frac{\alpha_0 + 1}{\alpha_0 + \beta_0 + 2}$ ,  $\bar{p}_{2,k-1}$  for

$\frac{\gamma_0 + 1}{\gamma_0 + \delta_0 + 2}$ ,  $\alpha_{k-1} + \beta_{k-1}$  for  $\alpha_0 + \beta_0$ ,  $\gamma_{k-1} + \delta_{k-1}$  for  $\gamma_0 + \delta_0$ ,

$n_{1k}$  for  $n_{11}$ , and  $n_{2k}$  for  $n_{21}$ , assume  $|k_1/k_2| = 1$  in (5.9),

and solve for  $n_{1k}$ , we obtain the approximate procedure as

given by (4.1) in both cases. However, if we do not assume

$|k_1/k_2| = 1$ , we obtain the following "generalized" approx-

imate procedure for the linear loss function given by (2.7):

$$(6.3) \quad n_{1k} \approx \frac{(\gamma_{k-1} + \delta_{k-1} + n_k + 3)R_{k-1}^* - \alpha_{k-1} - \beta_{k-1} - 3}{R_{k-1}^* + 1},$$

where

$$R_{k-1}^* = \left| \frac{k_1}{k_2} \right| \frac{\bar{p}_{1,k-1}(1-\bar{p}_{1,k-1})}{\bar{p}_{2,k-1}(1-\bar{p}_{2,k-1})}.$$

VII. NUMERICAL STUDIES OF EXTENSIVE-FORM  
ANALYSIS, THE STAGE-BY-STAGE PROCEDURE,  
AND THE APPROXIMATE PROCEDURE

A method of comparing the three different procedures for given loss functions and prior distributions is desired. The proposed method for this comparison is to find the exact probability of making a given decision (say  $b = 1$ ) as a function of  $\theta_1$  and  $\theta_2$ . This can be done for extensive-form analysis by computing

$$\text{Pr}(\text{choosing } b=1)$$

$$= \Sigma P_{\theta}(r_{11}, r_{21}, \dots, r_{1m}, r_{2m}, n_{11}, n_{21}, \dots, n_{1m}, n_{2m}) \\ + \frac{1}{2} \Sigma' P_{\theta}(r_{11}, r_{21}, \dots, r_{1m}, r_{2m}, n_{11}, n_{21}, \dots, n_{1m}, n_{2m}),$$

where  $\Sigma$  is the sum over all values of  $r_{11}, r_{21}, \dots, r_{1m}, r_{2m}, n_{11}, n_{21}, \dots, n_{1m}, n_{2m}$  such that

$$q_1 \sum_{j=0}^{\beta_m} (-1)^j \binom{\beta_m}{j} \frac{B(\alpha_m + \gamma_m + j + 2, \delta_m + 1)}{\alpha_m + j + 1} \\ < q_2 \sum_{j=0}^{\delta_m} (-1)^j \binom{\delta_m}{j} \frac{B(\alpha_m + \gamma_m + j + 2, \beta_m + 1)}{\gamma_m + j + 1}$$

for the constant loss function defined by (2.1) or such that

$$(k_{20} - k_{10}) + (k_{21} - k_{11}) \frac{\alpha_m + 1}{\alpha_m + \beta_m + 2} + (k_{22} - k_{12}) \frac{\gamma_m + 1}{\gamma_m + \delta_m + 2} > 0$$

for the linear loss function defined by (2.7) and where  $\Sigma'$  is

the sum over all values of  $r_{11}, r_{21}, \dots, r_{1m}, r_{2m}, n_{11}, n_{21}, \dots, n_{1m}, n_{2m}$  such that equality holds. (Either  $b = 1$  or  $b = 2$  can be taken.) We can compute

$P_{\theta}(r_{11}, r_{21}, \dots, r_{1m}, r_{2m}, n_{11}, n_{21}, \dots, n_{1m}, n_{2m})$  by computing

$P_{\theta}(r_{1m}, r_{2m} | n_{1m}, n_{2m}, r_{1,m-1}, r_{2,m-1}, n_{1,m-1}, n_{2,m-1}, \dots, r_{1k}, r_{2k}, n_{1k}, n_{2k}, \dots, r_{11}, r_{21}, n_{11}, n_{21})$

$\cdot P_{\theta}(n_{1m}, n_{2m} | r_{1,m-1}, r_{2,m-1}, n_{1,m-1}, n_{2,m-1}, \dots, r_{1k}, r_{2k}, n_{1k}, n_{2k}, \dots, r_{11}, r_{21}, n_{11}, n_{21}) \cdot$

$\dots \cdot P_{\theta}(r_{1k}, r_{2k} | n_{1k}, n_{2k}, \dots, r_{11}, r_{21}, n_{11}, n_{21})$

$\cdot P_{\theta}(n_{1k}, n_{2k} | r_{1,k-1}, r_{2,k-1}, n_{1,k-1}, n_{2,k-1}, \dots, r_{11}, r_{21}, n_{11}, n_{21})$

$\dots \cdot P_{\theta}(r_{11}, r_{21} | n_{11}, n_{21}) P_{\theta}(n_{11}, n_{21}) ,$

where  $P_{\theta}(r_{1k}, r_{2k} | n_{1k}, n_{2k}, \dots, r_{11}, r_{21}, n_{11}, n_{21})$

$$= \binom{n_{1k}}{r_{1k}} \theta_1^{r_{1k}} (1-\theta_1)^{n_{1k}-r_{1k}} \binom{n_{2k}}{r_{2k}} \theta_2^{r_{2k}} (1-\theta_2)^{n_{2k}-r_{2k}}$$

and

$P_{\theta}(n_{1k}, n_{2k} | r_{1,k-1}, r_{2,k-1}, n_{1,k-1}, n_{2,k-1}, \dots, r_{11}, r_{21}, n_{11}, n_{21})$

$= 1/n_k^*$  if the value of  $(n_{1k}, n_{2k})$  minimizes

$$E_{x_k | x_1, \dots, x_{k-1}} L^{(m-k+1)}(x_1, \dots, x_k, a_1, \dots, a_k) ,$$

$= 0$  otherwise ,

where  $n_k^*$  is the number of  $(n_{1k}, n_{2k})$  combinations which give equivalent minima to

$$E_{x_k | x_1, \dots, x_{k-1}} L^{(m-k+1)}(x_1, \dots, x_k, a_1, \dots, a_k).$$

In a similar fashion we can compute the exact probability of choosing the terminal decision  $b = 1$  for the stage-by-stage and approximate procedures.

Computer programs have been written for calculating the exact probability of  $b = 1$  for extensive-form analysis and the stage-by-stage procedure with both linear and constant loss functions in each case and also for the approximate procedure. The results of the computations with these programs are shown in Tables 4 and 5. We have, without loss of generality, taken  $\theta_1 > \theta_2$  so that  $b = 1$  is the correct decision. Under the columns labeled "Example" in each of these tables is given the assumed number of observations in each stage. Under the columns labeled "Equal Division" is given the probability of choosing  $b = 1$  if prior to the experiment it is decided that exactly half of the observations are to be allocated to Drug 1 and half to Drug 2 no matter what the results of the intermediate stages are. This probability was computed with the program for the approximate procedure by assuming that all of the observations are in one stage and the prior distribution is rectangular. Some positions in the tables are left blank either because the examples

Table 4

$$\text{Probability of } b=1 \text{ for } L(\theta, b) = \begin{cases} \theta_2 - \theta_1 & \text{for } b=1 \\ \theta_1 - \theta_2 & \text{for } b=2 \end{cases} \text{ with } \rho(\theta) = \begin{cases} 1 & \text{for } 0 \leq \theta_1 \leq 1 \\ & \text{and } 0 \leq \theta_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example	$\theta_1$	$\theta_2$	Extensive-Form Analysis	Stage-by-Stage Procedure	Approximate Procedure	Equal Division
1-1-1	.6	.4	.648000	.648000	.648000	---
1-1-1	.8	.6	.656000	.656000	.656000	---
1-1-1	.95	.8	.632749	.632749	.632750	---
4-2	.6	.4	.682561	.682560	.682560	.682560
4-2	.8	.6	.704001	.704000	.695040	.695040
4-2	.95	.8	.710841	.710841	.678357	---
3-2-3	.6	.4	.710093	.710210	.710208	.710208
3-2-3	.8	.6	.739098	.739124	.725504	.725504
3-2-3	.95	.8	.745204	.744731	.715230	.715230
5-4	.6	.4	.732768	.731440	.733431	---
5-4	.8	.6	.751542	.749378	.750673	---
5-4	.95	.8	.763867	.762961	.745676	---
1-1-1-1-1	.6	.4	.680760	.679680	.682560	---
1-1-1-1-1	.8	.6	.695240	.695360	.695040	---
1-1-1-1-1	.95	.8	.683255	.686192	.678357	---



Table 4 (continued)

Example	$\theta_1$	$\theta_2$	Extensive-Form Analysis	Stage-by-Stage Procedure	Approximate Procedure	Equal Division
8-6	.6	.4	---	.771006	.771155	.771155
8-6	.8	.6	---	.798961	.792613	.790834
8-6	.95	.8	---	.817327	.795506	.793157
12-14	.6	.4	---	---	.846028	.846233
12-14	.8	.6	---	---	.868856	.867876
12-14	.95	.8	---	---	.880464	.878446
8	.6	.4	.710208	.710208	.710208	.710208
8	.8	.6	.737280	.737280	.725504	.725504
8	.95	.8	.743770	.743770	.715230	.715230
14	.6	.4	.771157	.771157	.771155	.771155
14	.8	.6	.798908	.798908	.790834	.790834
14	.95	.8	.810577	.810577	.793157	.793157
26	.6	.4	---	---	.846233	.846233
26	.8	.6	---	---	.867876	.867876
26	.95	.8	---	---	.878466	.878466

Table 5

$$\text{Probability of } b=1 \text{ for } L(\theta, b) = \begin{cases} 1 & \text{if } b=1 \text{ and } \theta_1 < \theta_2 \\ 1 & \text{if } b=2 \text{ and } \theta_1 > \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{with } \rho(\theta) = \begin{cases} 1 & \text{for } 0 \leq \theta_1 \leq 1 \\ & \text{and } 0 \leq \theta_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example	$\theta_1$	$\theta_2$	Extensive-Form Analysis	Stage-by-Stage Procedure	Approximate Procedure	Equal Division
1-1-1	.6	.4	.648000	.648000	.648000	---
1-1-1	.8	.6	.656000	.656000	.656000	---
1-1-1	.95	.8	.632749	.632749	.632750	---
4-2	.6	.4	.682561	.682561	.682560	.682560
4-2	.8	.6	.703999	.703999	.695040	.695040
4-2	.95	.8	.710841	.710841	.678357	.678357
3-2-3	.6	.4	.707443	.710210	.710208	.710208
3-2-3	.8	.6	.737177	.739121	.725504	.725504
3-2-3	.95	.8	.761924	.744731	.715230	.715230
8	.6	.4	.710208	.710208	.710208	.710208
8	.8	.6	.737280	.737280	.725504	.725504
8	.95	.8	.743770	.743770	.715230	.715230
1-1-1-1-1	.6	.4	.671040	.671040	.682560	---
1-1-1-1-1	.8	.6	.696320	.696320	.695040	---
1-1-1-1-1	.95	.8	.709701	.709701	.678357	---

do not apply as is the case in the columns labeled "Equal Division" or because the computations take much too long to be worthwhile.

For the loss functions and prior distributions considered in Tables 4 and 5 it appears that, in general, for the examples studied there is little improvement in  $\Pr(b = 1)$  when the approximate procedure instead of just equal division is used. Also, there seems to be relatively little difference between using the stage-by-stage procedure and extensive-form analysis. The only difference in  $\Pr(b = 1)$  for the two procedures which is greater than .01 is that for the 3-2-3 example with Loss Function 3 (as defined by 2.11) when  $\theta_1 = .95$  and  $\theta_2 = .80$ . The greatest differences between the four procedures arise when  $\theta_1 = .95$ ,  $\theta_2 = .80$ , and the approximate procedure is used instead of the stage-by-stage procedure. For the 3-2-3 example, when  $\theta_1 = .95$  and  $\theta_2 = .80$  and Loss Function 3 is used, the loss in  $\Pr(b = 1)$  when the approximate rather than the stage-by-stage procedure is used is almost .03. However, for  $\theta_1$  and  $\theta_2$  symmetric about .5, in general, the approximate procedure is a bit better than the stage-by-stage procedure or extensive-form analysis.

It should be noted that in the smaller examples there is no improvement at all in using the approximate procedure rather than just equal division of the observations. We can easily derive an expression for the range of values of  $n_k$

at any given stage in the approximate procedure for which this reduces to equal division. From (4.1) (assuming  $n_{1k}$  can be any real number)

$$n_{1k} = \frac{(n_k + A_1) R_{k-1} - A_2}{R_{k-1} + 1} ,$$

where  $A_1 = \gamma_{k-1} + \delta_{k-1} + 3$

and  $A_2 = \delta_{k-1} + \delta_{k-1} + 3$  .

Equal division will be used if

$$\frac{1}{2}n_k - \frac{1}{2} \leq n_{1k} \leq \frac{1}{2}n_k + \frac{1}{2} ,$$

or

$$-\frac{1}{2} \leq \frac{(n_k + A_1) R_{k-1} - A_2}{R_{k-1} + 1} - \frac{1}{2}n_k \leq \frac{1}{2} ,$$

or

$$(7.1) \quad - \frac{R_{k-1}^{-1} - 2R_{k-1}A_1 + 2A_2}{R_{k-1}^{-1}} \leq n_k \leq \frac{R_{k-1}^{+1} - 2R_{k-1}A_1 + 2A_2}{R_{k-1}^{-1}} \text{ if } R_{k-1} > 1$$

and

$$- \frac{R_{k-1}^{-1} + 2R_{k-1}A_1 - 2A_2}{1 - R_{k-1}} \leq n_k \leq \frac{R_{k-1}^{+1} + 2R_{k-1}A_1 - 2A_2}{R_{k-1}^{-1}} \text{ if } R_{k-1} < 1.$$

Let us now consider the example in Table 2 which has  $n_1 = 4$  and  $n_2 = 2$ . In the first stage  $n_{11} = 2$  and  $n_{21} = 2$  so that for the second stage  $A_1 = 5$ ,  $A_2 = 5$ , the largest possible value of  $R_1$  is  $\sqrt{4/3}$ , and the smallest possible value of  $R_1$  is  $\sqrt{3/4}$ . In either case from (7.1) we see that we shall have equal division for  $n_2=2$ ; i.e., for any possible

$R_1$ ,  $n_{21} = 1$  and  $n_{22} = 1$ . In addition, for  $n_2$  even we shall not be able to avoid equal division unless  $n_2 \geq 4$ .

Since the differences between the probabilities in Tables 4 and 5 for the approximate procedure were so small for the relatively small examples studied and since the time necessary for computing much larger examples would be prohibitively long, a program for a simulation study of the approximate procedure was written. An example which had thirteen stages with two observations in each stage,  $\theta_1 = .95$ ,  $\theta_2 = .80$ , and  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  in the prior distribution of  $\theta_1$  and  $\theta_2$  was run with this program. From Table 4 one can see that, if there is equal division of the 26 observations before the experiment, the probability of choosing  $b = 1$  is .878466. When the example was run, the proportion of times  $b = 1$  was chosen out of 1365 repetitions was .877, which is even less than .878466. Certainly one could not conclude that the approximate procedure gave a higher probability of choosing  $b = 1$  than equal division did for this example.

It is interesting to compare equal division and the approximate procedure for very large examples. If the total

number of observations  $N = \sum_{i=1}^m n_i$  is large (e.g.,  $N \geq 100$ ) and  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ ,  $N_1 = \sum_{i=1}^m n_{1i}$  and  $N_2 = \sum_{i=1}^m n_{2i}$

will in the approximate procedure be quite close to satisfying the following relationship:

$$N_1/N_2 = \sigma_1/\sigma_2 \quad ,$$

where  $\sigma_1^2 = \theta_1(1-\theta_1)$  and  $\sigma_2^2 = \theta_2(1-\theta_2)$ . Thus the probability of choosing  $b = 1$  for the approximate procedure will be close to

$$\Pr \left( z > \frac{\theta_2 - \theta_1}{\left( \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} \right)^{\frac{1}{2}}} \right) \quad ,$$

where  $z$  is the  $N(0,1)$  variable. Equal division of the observations is covered by the special case  $N_1 = N_2 = \frac{1}{2}N$ . Table 6 gives the estimated probability of choosing  $b = 1$  for  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$  as a function of  $\theta_1$ ,  $\theta_2$ , and  $N$ .

Table 6

Approximate Probability of Choosing  $b = 1$

<u>N</u>	<u><math>\theta_1</math></u>	<u><math>\theta_2</math></u>	<u>Approximate Procedure</u>	<u>Equal Division</u>
100	.95	.90	.8328	.8298
100	.80	.75	.7257	.7257
200	.95	.91	.8691	.8669
200	.80	.76	.7530	.7528

As one can see from Table 6, there is a little improvement in using the approximate procedure over equal division for  $\theta_1$  and  $\theta_2$  close to 1 (or 0), but there is practically no difference when  $\theta_1$  and  $\theta_2$  are close to .5.

VIII. STOPPING RULES FOR THE  
STAGE-BY-STAGE AND APPROXIMATE PROCEDURES

8.1 Stopping Rules for the Stage-by-Stage Procedure

Suppose we have a stage-by-stage procedure in which it is specified there will be a maximum of  $m$  stages but the actual number of stages is not fixed. Thus we might want to stop before taking all  $m$  stages. Suppose we have run  $k - 1$  stages and want to decide whether or not to take the  $k$ 'th stage ( $k = 1, 2, \dots, m$ ). In order to define our stopping rule

we shall let  $N_{kk'} = \sum_{i=k}^{k'} n_i$  ( $k' = k, k+1, \dots, m$ ),  $N_{1kk'}$  be

the number of observations out of  $N_{kk'}$  allocated to Population 1,  $N_{2kk'}$  (where  $N_{2kk'} = N_{kk'} - N_{1kk'}$ ) be the number of observations out of  $N_{kk'}$  allocated to Population 2,  $R_{1kk'}$  and  $R_{2kk'}$  be the numbers of successes out of  $N_{1kk'}$  and  $N_{2kk'}$  respectively,  $\alpha_{k'} = \alpha_{k-1} + R_{1kk'}$ ,  $\beta_{k'} = \beta_{k-1} + N_{1kk'} - R_{1kk'}$ ,  $\gamma_{k'} = \gamma_{k-1} + R_{2kk'}$ ,  $\delta_{k'} = \delta_{k-1} + N_{2kk'} - R_{2kk'}$ ,  $x_{kk'} = (R_{1kk'}, R_{2kk'})$ ,  $C_i$  be the cost of the  $i$ 'th stage, and  $a_{kk'}$  be the decision of what  $N_{1kk'}$  to take ( $N_{1kk'} = 1, 2, \dots, N_{kk'}$ ). (Note that, if  $k' = k$ ,  $N_{kk'}$ ,  $N_{1kk'}$ ,  $N_{2kk'}$ ,  $R_{1kk'}$ ,  $R_{2kk'}$ ,  $x_{kk'}$ , and  $a_{kk'}$  reduce to  $n_k$ ,  $n_{1k}$ ,  $n_{2k}$ ,  $r_{1k}$ ,  $r_{2k}$ , and



$a_k$ , respectively.) Also, since in this chapter we shall be using only the general constant (2.1) or linear (2.7) loss functions and these depend directly only on  $b$  and  $\theta$ , we shall denote the loss function by  $L(\theta, b)$ .

Using Amster's (1962) modified Bayes stopping rule (itself modified and adapted to the problem in this thesis), we have the following rule:

If

$$(8.1) \quad \min_b E_{\theta|x_1, \dots, x_{k-1}} L(\theta, b) \leq \min_{k'} \left[ \sum_{i=1}^{k'} C_i + \min_{a_{kk'}} E_{x_{kk'}|x_1, \dots, x_{k-1}} \min_b E_{\theta|x_1, \dots, x_{k-1}, x_{kk'}} L(\theta, b) \right],$$

do not take the  $k'$ th stage.

If each  $C_i$  is relatively quite small, the stopping rule becomes approximately:

Stop sampling if

$$\begin{aligned} & \min_b E_{\theta|x_1, \dots, x_{k-1}} L(\theta, b) \\ & = \min_{a_{km}} E_{x_{km}|x_1, \dots, x_{k-1}} \min_b E_{\theta|x_1, \dots, x_{k-1}, x_{km}} L(\theta, b) . \end{aligned}$$

For the linear loss function (2.7)

$$\begin{aligned} & \min_b E_{\theta|x_1, \dots, x_{k-1}} L(\theta, b) \\ & = \min \left\{ k_{10} + k_{11} \left( \frac{\alpha_{k-1} + 1}{\alpha_{k-1} + \beta_{k-1} + 2} \right) + k_{12} \left( \frac{\gamma_{k-1} + 1}{\gamma_{k-1} + \delta_{k-1} + 2} \right) \right\} , \end{aligned}$$

$$k_{20} + k_{21} \left( \frac{\alpha_{k-1} + 1}{\alpha_{k-1} + \beta_{k-1} + 2} \right) + k_{22} \left( \frac{\gamma_{k-1} + 1}{\gamma_{k-1} + \delta_{k-1} + 2} \right) \Bigg\}$$

and

$$\min_{a_{kk}, b} E_{x_{kk}} | x_1, \dots, x_{k-1} \min_b E_{\theta} | x_1, \dots, x_{k-1}, x_{kk}, L(\theta, b)$$

$$= \min_{a_{kk}} \sum_{R_{1kk}=0}^{N_{1kk}} \sum_{R_{2kk}=0}^{N_{2kk}} \binom{N_{1kk}}{R_{1kk}} \binom{N_{2kk}}{R_{2kk}} \cdot \frac{B(\alpha_{k'} + 1, \beta_{k'} + 1)}{B(\alpha_{k-1} + 1, \beta_{k-1} + 1)} \cdot \frac{B(\gamma_{k'} + 1, \delta_{k'} + 1)}{B(\gamma_{k-1} + 1, \delta_{k-1} + 1)} \cdot \min \left\{ k_{10} + k_{11} \left( \frac{\alpha_{k'} + 1}{\alpha_{k'} + \beta_{k'} + 2} \right) + k_{12} \left( \frac{\gamma_{k'} + 1}{\gamma_{k'} + \delta_{k'} + 2} \right), k_{20} + k_{21} \left( \frac{\alpha_{k'} + 1}{\alpha_{k'} + \beta_{k'} + 2} \right) + k_{22} \left( \frac{\gamma_{k'} + 1}{\gamma_{k'} + \delta_{k'} + 2} \right) \right\} \cdot$$

For the constant loss function (2.1)

$$\min_b E_{\theta} | x_1, \dots, x_{k-1} L(\theta, b)$$

$$= \frac{1}{B(\alpha_{k-1} + 1, \beta_{k-1} + 1) B(\gamma_{k-1} + 1, \delta_{k-1} + 1)} \cdot \min \left\{ q_1 \sum_{j=0}^{\beta_{k-1}} (-1)^j \binom{\beta_{k-1}}{j} \frac{B(\alpha_{k-1} + \gamma_{k-1} + j + 2, \delta_{k-1} + 1)}{\alpha_{k-1} + j + 1}, q_2 \sum_{j=0}^{\delta_{k-1}} (-1)^j \binom{\delta_{k-1}}{j} \frac{B(\alpha_{k-1} + \gamma_{k-1} + j + 2, \beta_{k-1} + 1)}{\gamma_{k-1} + j + 1} \right\}$$

and

$$\begin{aligned}
 & \min_{a_{kk'}} E_{x_{kk'} | x_1, \dots, x_{k-1}} \min_b E_{\theta | x_1, \dots, x_{k-1}, x_{kk'}} L(\theta, b) \\
 &= \min_{a_{kk'}} \frac{\sum_{R_{1kk'}=0}^{N_{1kk'}} \sum_{R_{2kk'}=0}^{N_{2kk'}}}{B(\alpha_{k-1}+1, \beta_{k-1}+1) B(\gamma_{k-1}+1, \delta_{k-1}+1)} \frac{\binom{N_{1kk'}}{R_{1kk'}} \binom{N_{2kk'}}{R_{2kk'}}}{\cdot \min \left\{ \begin{aligned} & q_1 \sum_{j=0}^{\beta_{k'}} (-1)^j \binom{\beta_{k'}}{j} \frac{B(\alpha_{k'}+\gamma_{k'}+j+2, \delta_{k'}+1)}{\alpha_{k'}+j+1} , \\ & q_2 \sum_{j=0}^{\delta_{k'}} (-1)^j \binom{\delta_{k'}}{j} \frac{B(\alpha_{k'}+\gamma_{k'}+j+2, \beta_{k'}+1)}{\gamma_{k'}+j+1} \end{aligned} \right\} .
 \end{aligned}$$

### 8.2 A Stopping Rule for the Approximate Procedure

In this section a stopping rule for the approximate procedure will be derived by the use of normal approximations. Of course, this stopping rule will also be applicable to a stage-by-stage procedure in which normal random variables with unknown variance are generated. The general stopping rule will be that defined by (8.1), and the loss function used in this section will be the linear loss function as defined by (2.7).

It has been seen in Chapter 5 that in the stage-by-stage procedure for normal random variables one can consider his  $k$ 'th stage ( $k = 1, 2, \dots, m$ ) as the only stage of a one stage procedure with a prior distribution which is altered as one

proceeds from stage to stage. It has also been seen in Section 6.2 that the approximate procedure can be derived from the results for the stage-by-stage procedure for normal random variables. In view of the results of Chapter 5 and Section 6.2 we may suppose that we have at the  $k$ 'th stage a one-stage procedure which generates normal random variables  $\bar{x}_1$  and  $\bar{y}_1$  and has prior distributions (altered at each stage) of  $\theta_1$  and  $\theta_2$  which are  $N(\bar{x}_0, \sigma_1^2/n_{10})$  and  $N(\bar{y}_0, \sigma_2^2/n_{20})$ . Although our notation does not show it, the parameters are functions of  $k$ . In fact, we have

$$(8.2) \quad \bar{x}_0 = \bar{p}_{1,k-1}, \quad \bar{y}_0 = \bar{p}_{2,k-1}, \quad n_{10} = \alpha_{k-1} + \beta_{k-1} + 3,$$

$$n_{20} = \gamma_{k-1} + \delta_{k-1} + 3, \quad \sigma_1^2 = \bar{p}_{1,k-1}(1 - \bar{p}_{1,k-1}),$$

and  $\sigma_2^2 = \bar{p}_{2,k-1}(1 - \bar{p}_{2,k-1})$ .

Also, let

$$(8.3) \quad n_{11} = N_{1kk}, \quad n_{21} = N_{2kk}, \quad n_1 = N_{kk},$$

$$n_0 = n_{10} + n_{20}, \quad \bar{x}_1 = R_{1kk}/N_{1kk}, \quad \bar{y}_1 = R_{2kk}/N_{2kk},$$

$$x_1 = (\bar{x}_1, \bar{y}_1), \quad \text{and } C_1 = \sum_{i=k}^{k'} C_i.$$

With these substitutions and assumptions we can use the results of Sections 5.2 and 6.2. We shall assume further, as in Section 5.2 that  $k_{10} + k_{11}\bar{x}_0 + k_{12}\bar{y}_0 \geq k_{20} + k_{21}\bar{x}_0 + k_{22}\bar{y}_0$ .

Using the above assumptions and (5.1) we see that the left

hand side of the inequality of (8.1) is equivalent to  $E_{\theta} L(\theta, b)$  with the altered prior distribution of  $\theta$  and is  $k_{20} + k_{21}\bar{x}_0 + k_{22}\bar{y}_0$ . Similarly, using the above assumptions, (5.8), the arguments leading to (5.9), and (5.10) we can find the expression for the right hand side of the inequality of (8.1) and thus obtain for our stopping rule:

Do not take the  $k^{\text{th}}$  stage if

$$k_{20} + k_{21}\bar{x}_0 + k_{22}\bar{y}_0 \leq \min_{k^{\text{th}}} \left[ C_1^k - \frac{1}{\sqrt{h_{\min}}} \int_{\sqrt{h_{\min}} E(M)}^{\infty} (z - \sqrt{h_{\min}} E(M)) \phi(z) dz + (k_{20} + k_{21}\bar{x}_0 + k_{22}\bar{y}_0) \right],$$

or

Do not take the  $k^{\text{th}}$  stage if

$$(8.4) \quad \min_{k^{\text{th}}} \left[ C_1^k - \frac{1}{\sqrt{h_{\min}}} \int_{\sqrt{h_{\min}} E(M)}^{\infty} (z - \sqrt{h_{\min}} E(M)) \phi(z) dz \right] \geq 0,$$

where  $h_{\min}$  is the minimized value of  $h$  in (5.8). It will be seen from (5.8) and the arguments leading to (5.9) that

$$\begin{aligned} 1/h_{\min} &= \sigma_1^2 k_1^2 \left( \frac{1}{n_{10}} - \frac{1}{n_{10} + n_{11}} \right) \\ &+ \sigma_2^2 k_2^2 \left( \frac{1}{n_{20}} - \frac{1}{n_{20} + n_{21}} \right), \end{aligned}$$

where  $n_{11}$  and  $n_{21}$  are chosen so that (5.9) is satisfied, and that (5.8) is minimized when  $h = h_{\min}$ .

If we let  $u = \sqrt{h_{\min}} E(M)$ , then we can let

$$L_{N^*}(u) = \int_u^{\infty} (z - u) \phi(z) dz \quad ,$$

which Raiffa and Schlaifer (1961) call the "linear loss integral" and tabulate in their book. Thus by using their table one can apply this stopping rule without resorting to a digital computer. The computations are simplified if we make the following observations and assumptions:

From (5.9) it is seen that we obtain  $h_{\min}$  when

$$\frac{n_{10} + n_{11}}{n_{20} + n_{21}} = \left| \frac{k_1}{k_2} \right| \frac{\sigma_1}{\sigma_2} = R_1^* \quad .$$

Let us also assume  $n_{10}/n_{20} = R_1^*$ . (This assumption will most likely be good after the first few stages.) Then

$$n_{10} = R_1^* n_{20} \text{ and } n_{11} = R_1^* n_{20} + R_1^* n_{21} - n_{10} = R_1^* n_{21} \quad .$$

Thus  $n_{21} = \frac{n_1}{R_1^* + 1}$  and  $n_{11} = \frac{R_1^* n_1}{R_1^* + 1}$ . Also, since we let

$$n_0 = n_{10} + n_{20}, \quad n_{10} = \frac{R_1^* n_0}{R_1^* + 1} \text{ and } n_{20} = \frac{n_0}{R_1^* + 1} \quad . \text{ Then}$$

$$n_{11} + n_{10} = \frac{R_1^*}{R_1^* + 1} (n_1 + n_0) \text{ and } n_{21} + n_{20} = \frac{n_1 + n_0}{R_1^* + 1} \quad , \text{ and}$$

$$\begin{aligned}
 1/h_{\min} &= \sigma_1^2 k_1^2 \left( \frac{R_1^{i+1}}{R_1^i n_0} - \frac{R_1^{i+1}}{R_1^i (n_0 + n_1)} \right) \\
 &+ \sigma_2^2 k_2^2 \left( \frac{R_1^{i+1}}{n_0} - \frac{R_1^{i+1}}{n_0 + n_1} \right) \\
 &= \left( \frac{\sigma_1^2 k_1^2}{R_1^i} + \sigma_2^2 k_2^2 \right) (R_1^{i+1}) \left( \frac{1}{n_0} - \frac{1}{n_0 + n_1} \right) \\
 &= \left( \frac{R_1^{i+1} \sigma_1^2 k_1^2}{R_1^i} + \sigma_2^2 k_2^2 \right) (R_1^{i+1}) \left( \frac{1}{n_0} - \frac{1}{n_0 + n_1} \right) \\
 &= \sigma_2^2 k_2^2 (R_1^{i+1})^2 \left( \frac{1}{n_0} - \frac{1}{n_0 + n_1} \right) \\
 &= (|k_1| \sigma_1 + |k_2| \sigma_2)^2 \left( \frac{1}{n_0} - \frac{1}{n_0 + n_1} \right) .
 \end{aligned}$$

By adapting the above results to our stopping rule, and making the substitutions of (8.2) and (8.3) we obtain:

Do not take the  $k^{\text{th}}$  stage if

$$(8.5) \quad \min_{k^*} \left[ \sum_{i=k}^{k^*} C_i - K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk^*}}} \right. \\
 \left. \cdot L_{N^*} \left( \frac{|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|}{K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk^*}}}} \right) \right] \geq 0 ,$$

where  $K_{k-1} = \sum_{i=1}^2 |k_i| \sqrt{\bar{p}_{i,k-1} (1 - \bar{p}_{i,k-1})}$

and where  $|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|$

is used so that the case in which

$$k_{20} + k_{21} \bar{p}_{1,k-1} + k_{22} \bar{p}_{2,k-1} \geq k_{10} + k_{11} \bar{p}_{1,k-1} + k_{12} \bar{p}_{2,k-1}$$

is also included.

For the special case in which

$$L(\theta, b) = \begin{cases} k_{10} + (\theta_2 - \theta_1) & \text{for } b = 1 \\ k_{10} + (\theta_1 - \theta_2) & \text{for } b = 2 \end{cases}$$

so that  $k_0 = 0$ ,  $k_1 = -2$ , and  $k_2 = 2$  the stopping rule becomes:

Do not take the  $k^*$ th stage if

$$(8.6) \quad \min_{k^*} \left[ \sum_{i=k}^{k^*} C_i - 2K_{k-1}^0 \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk^*}}} \right. \\ \left. \cdot L_{N_{kk^*}} \left( \frac{|\bar{p}_{2,k-1} - \bar{p}_{1,k-1}|}{K_{k-1}^0 \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk^*}}}} \right) \right] \geq 0,$$

where  $K_{k-1}^0 = \sum_{i=1}^2 \sqrt{\bar{p}_{i,k-1} (1 - \bar{p}_{i,k-1})}$ .

We would like to have a better method for evaluating (8.5) than enumeration over the possible values of  $k^*$ . Let  $C$  be the cost per observation and assume that it remains the same from stage to stage. Then  $\sum_{i=k}^{k^*} C_i = C N_{kk^*}$ . Suppose

we wanted to find  $N_{kk^*}$ , where the possible values of  $N_{kk^*}$  are  $1, 2, \dots, N_{km}$ , such that the expression in brackets in (8.5)



was minimized. If we let  $n = N_{kk'}$ ,  $n' = \sum_{i=1}^{k-1} n_i + 6$ ,

$|\mu_b - m'| = |k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|$ ,  $h = 1/K_{k-1}^2$ ,  $k_t = 1$ ,  
and  $k_s = C$  in Equation 5-38a on page 115 of Raiffa and Schlaifer (1961), we have

$$v^*(e_n) = - \left[ CN_{kk'} - K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk'}}}} \right. \\ \left. \cdot L_{N^*} \left( \frac{|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|}{K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk'}}}}} \right) \right]$$

Thus our problem is equivalent to finding  $n$  such that  $v^*(e_n)$  is maximized, which is the problem solved by Raiffa and Schlaifer. Let  $n^0$  be the optimal value of  $n$ . Graphical methods are used to find  $n^0$ . Chart I in the back of their book gives  $\eta^0$  for given values of  $D_\infty$  and  $Z$ , where by substituting into their expressions for  $\eta^0$ ,  $D_\infty$ , and  $Z$  we have

$$\eta^0 = n^0 (C/K_{k-1})^{2/3},$$

$$D_\infty = \frac{|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|}{K_{k-1} / \left( \sum_{i=1}^{k-1} n_i + 6 \right)^{\frac{1}{2}}},$$

and 
$$Z = \left[ \frac{K_{k-1}}{C \left( \sum_{i=1}^{k-1} n_i + 6 \right)^{3/2}} \right]^{1/3}$$

Thus with the use of the chart our stopping rule procedure becomes:

- (1) Compute  $D_{\infty}$  and  $Z$  and refer to Figure 5.10 on page 118 of Raiffa and Schlaifer's book, which has  $Z_C$  graphed as a function of  $D'$ , where  $D' = D_{\infty}$ . If  $Z < Z_C$  for the given  $D_{\infty}$ , do not take the  $k$ 'th stage but make the terminal decision ( $b = 1$  or  $b = 2$ ) on the basis of the results for the first  $k - 1$  stages.
- (2) If  $Z > Z_C$ , refer to Chart I, find the  $\eta^0$  (and thus the  $n^0$ ) which corresponds to the given  $Z$  and  $D_{\infty}$ , and continue as follows:
- (3) If  $n^0 > N_{km}$ , compute

$$CN_{km} - K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{km}}}$$

$$\cdot L_{N^*} \left( \frac{|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|}{K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{km}}}} \right)$$

with the aid of Table II on page 356 of Raiffa and Schlaifer's book. If this expression is greater

than or equal to zero, do not take the  $k^{\text{th}}$  stage but make the terminal decision. Otherwise take the  $k^{\text{th}}$  stage.

- (4) If  $n^0 < N_{km}$  and if  $N_{kk_0^*} < n^0 < N_{k,k_0^*+1}$ ,

where  $k \leq k_0^* \leq m - 1$ , do not take the  $k^{\text{th}}$  stage if

$$\min_{k^*} \left[ C N_{kk^*} - K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk^*}}} \right. \\ \left. \cdot L_{N^*} \left( \frac{|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|}{K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + N_{kk^*}}}} \right) \right] \geq 0,$$

where in this case  $k^*$  is either  $k_0^*$  or  $k_0^* + 1$ .

- (5) If  $n^0 \leq N_{km}$  and if  $n^0 = N_{kk_0^*}$ , where  $k \leq k_0^* \leq m$ ,

take the  $k^{\text{th}}$  stage.

- (6) If  $n^0 < n_k$ , do not take the  $k^{\text{th}}$  stage if

$$C n_k - K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + n_k}} \\ \cdot L_{N^*} \left( \frac{|k_0 + k_1 \bar{p}_{1,k-1} + k_2 \bar{p}_{2,k-1}|}{K_{k-1} \sqrt{\frac{1}{\sum_{i=1}^{k-1} n_i + 6} - \frac{1}{\sum_{i=1}^{k-1} n_i + 6 + n_k}}} \right) \geq 0.$$

### 8.3 Results of Simulation Studies

A program for studying by simulation the approximate procedure with the stopping rule given by (8.6) has been written. In this program it is assumed that the cost per observation  $C$  is the same for every stage so that

$\sum_{i=k}^{k'} C_i = CN_{kk'}$ . The program generates rectangular random

variables between 0 and 1. Then at the  $k$ 'th stage, for a given  $n_{1k}$ ,  $n_{1k}$  random numbers are generated; and the number of these random numbers less than a given  $\theta_1$  is  $r_{1k}$ . Similarly, the number of  $n_{2k}$  random numbers which are less than a given  $\theta_2$  is  $r_{2k}$ . Also, ties between two equivalent values of  $n_{1k}$  or between the terminal decisions  $b = 1$  and  $b = 2$  are broken by generating a random number, taking one decision if it is greater than .5, and taking the other if it is less than .5.

The process of going through the stages and simulating the experimental outcomes of these stages until a terminal decision is made is repeated a given number of times. However, instead of going through all  $m$  possible stages before arriving at a terminal decision in a given repetition, the program has the procedure stop at the  $k$ 'th stage and make the terminal decision if (8.6) holds. After all the repetitions of the procedure have been made, the program has the

computer punch out the number of repetitions which gave  $b = 1$  as the terminal decision and for each stage (including the "0"th stage, which means that the terminal decision was made without any experimentation at all) the number of repetitions in which the terminal decision was made immediately after the stage was performed. The linear loss integral  $L_{N^*}(\mu)$ , which is needed in the computations is computed by interpolation in Table II on page 356 of Raiffa and Schlaifer's book (1961). Table II is read into the computer as data before the computations begin. Since Table II gives values of  $L_{N^*}(\mu)$  for  $0 \leq \mu \leq 4.0$  only and  $L_{N^*}(\mu)$  becomes quite small for  $\mu > 4.0$ ,  $L_{N^*}(\mu)$  is assumed to be zero for  $\mu > 4.0$ .

Tables 7 and 8 give the results of the application of the computer program to two examples. These tables show that the average number of observations decreases as  $|\theta_1 - \theta_2|$  increases for a given example. Also, as one would expect, the average number of observations decreases for a given  $\theta_1, \theta_2$  combination as the cost per observation  $C$  increases. The assumption that  $L_{N^*}(\mu) = 0$  for  $\mu > 4.0$  does not affect the results for these examples unless the cost per observation is zero or very small. In each case that the cost is zero the number of repetitions which stop before the last stage can be explained by the false assumption that  $L_{N^*}(\mu) = 0$  for  $\mu > 4.0$ .

Table 7

Results of Simulation Studies for the Approximate Procedure with the Stopping

Rule When  $m = 3$  with  $n_1 = 3$ ,  $n_2 = 2$ , and  $n_3 = 3$  and  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ 

Cost Per Observation	Number of Repetitions	$\theta_1$	$\theta_2$	Number of Repetitions Stopping After the Given Stage			Proportion of Repetitions for Which $b = 1$	Number of Repeti- tions with $u > 4.0$
				0	1	2		
0	100	.8	.6	0	0	7	.67	7
.005	100	.8	.6	0	30	11	.78	0
.005	100	.8	.2	0	53	23	.98	0
.01	100	.8	.6	0	61	18	.69	0
.01	100	.8	.2	0	79	18	.91	0



IX. GENERAL DISCUSSION OF THE RESULTS  
AND RECOMMENDATIONS

In the numerical studies done in this thesis almost all of the examples have assumed a prior distribution with  $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ . This prior distribution is the one a person would use if he knew little or nothing about the values of  $\theta_1$  and  $\theta_2$  prior to the experiment. Thus he would weight all combinations of  $\theta_1$  and  $\theta_2$  in which  $0 \leq \theta_1 \leq 1$  and  $0 \leq \theta_2 \leq 1$  equally. Also, in almost all of the examples Loss Functions 1 and 3 as defined by (2.11) have been used. These loss functions are symmetric in that they penalize the experimenter equally if he makes either of the two possible wrong terminal decisions ( $b = 1$  when  $\theta_1 < \theta_2$  or  $b = 2$  when  $\theta_1 > \theta_2$ ). However, in Loss Function 1 one has a "bonus" for making the right decision and a penalty for making the wrong decision which are proportional to  $\theta_1 - \theta_2$  with the constant of proportionality for one terminal decision the negative of the other. On the other hand, Loss Function 3 applies in the situation in which one has a loss for making the wrong decision which is independent of  $\theta_1 - \theta_2$  once it is determined that  $\theta_1 > \theta_2$  or  $\theta_1 < \theta_2$ .

It must be admitted that in the examples studied in Tables 4 and 5 the gains in using even the optimum procedure, extensive-form analysis, over merely using half of the



of the observations on Drug 1 and the other half on Drug 2 cannot be considered "tremendous." In fact, when  $\theta_1$  and  $\theta_2$  are both close to .5, equal division seems better in general. Nevertheless, if  $\theta_1$  and  $\theta_2$  are close to either 0 or 1, then the differences in the probability of choosing the correct terminal decision are large enough that they cannot be ignored -- especially if the loss for making the wrong decision is large.

However, extensive-form analysis is completely impractical for examples which are at all large even if one has access to a digital computer so that approximations must be used. The results indicate that, if one has a computer and is not certain that  $\theta_1$  and  $\theta_2$  are both close to .5, he should use the stage-by-stage procedure. In practically all of the examples studied in Tables 4 and 5 there was relatively little loss of probability of choosing the correct terminal decision when the stage-by-stage procedure was used instead of extensive-form analysis. In addition, the results of Section 6.1 give some weight to the arguments for using the stage-by-stage procedure instead of extensive-form analysis when one has a linear loss function. Finally, as was pointed out in Chapter 3, if one has a linear loss function and only a few observations in each stage, it is quite possible to do the computations for the stage-by-stage procedure by hand with the aid of a table of beta functions. (A Table of beta

functions is given by Tables of the Incomplete Beta Function (1934).)

The results for the approximate procedure for the examples studied are disappointing. There seems to be little advantage in using the approximate procedure over equal division of the observations when the prior distribution is rectangular and the loss functions are symmetric. In addition, there are sizable losses of probability of choosing the correct terminal decision when the approximate rather than the stage-by-stage procedure is used. However, the computations for the approximate procedure are quite simple and certainly do not require a computer; and there are small gains over equal division when the total number of observations and stages is large and  $\theta_1$  and  $\theta_2$  are close to 0 or 1. Thus the approximate procedure should be used if one requires only simple, unsophisticated computations to perform. The "generalized approximate procedure" as given by (6.3) can be used as an approximation for the stage-by-stage procedure (and thus extensive-form analysis) for the general linear loss function (2.7). However, the author was unable to derive a "generalized approximate procedure" for the general constant loss function (2.1) so that it appears that, when one has a constant loss function, he should use the approximate procedure only when his loss function is symmetric.

The results of Chapter 5 are not only interesting in

that they lead to the approximate procedure but also are of interest in that they are solutions to another general problem besides the principal one considered in this thesis. They give the procedure one would use in a stage-by-stage procedure if two normal random variables were generated at each stage rather than two binomial random variables. Again, results are available for the general linear loss function and for a symmetric constant loss function.

If the costs of sampling are relatively large or if the total possible number of observations is large, it is quite possible (and probable) that one will want to stop and make his terminal decision before he reaches the last possible stage. In fact, if the costs of sampling are very large relative to the possible terminal losses, one might want to make his terminal decision without any sampling at all. Thus stopping rules for the stage-by-stage procedure for both linear and constant loss functions are derived in Section 8.1. However, it is obvious that they are most unwieldy and are useful only for very small examples even if one has a digital computer available. In Section 8.2 a stopping rule for the approximate procedure when one has a linear loss function is developed. It is shown in Section 8.3 for two examples that this stopping rule behaves satisfactorily. Of course, this stopping rule can be adapted to the case in which one has a stage-by-stage procedure which generates

the normal random variables at each stage instead of binomial random variables. Also, it is recommended that it be used as an approximation when one has a stage-by-stage procedure which generates binomial random variables at each stage instead of using the stopping rule for linear loss functions developed in Section 8.1.

Thus the following final recommendations are made if one wants to decide in a multi-stage procedure which of two binomial populations has the higher probability of success and has symmetric linear or constant loss functions and rectangular prior distributions:

- (1) The approximate procedure should be used if one must have simple uncomplicated computations or if one is quite confident  $\theta_1$  and  $\theta_2$  are close to or symmetric about .5.
- (2) If one is not sure  $\theta_1$  and  $\theta_2$  are close to or symmetric about .5 and if a digital computer is available, he should use the stage-by-stage procedure. (If one has a linear loss function and only a few observations in each stage, a computer will not be necessary.)
- (3) If one feels that his costs of sampling are not very small compared to the possible losses of the terminal decision, has a linear loss function, and is using either the stage-by-stage or approximate

procedures, he should use the stopping rule developed in Section 8.2.

## X. SUGGESTIONS FOR FUTURE RESEARCH

It is obvious that many extensions to the work done in this thesis could be made. For example, the author hopes to do more computer work with non-symmetric loss functions and prior distributions which are not rectangular to determine how changing the prior distribution and the loss function affects the probability of choosing the correct terminal decision. It will also be of interest to compare extensive-form analysis, the stage-by-stage procedure and equal division for these loss functions and prior distributions.

In addition, for the cases in which one does not have symmetric loss functions or a rectangular prior distribution one might compare these four procedures with the following:

In the  $k$ 'th stage ( $k = 1, 2, \dots, m$ ) take  $n_{1k}$  and  $n_{2k}$  (subject to  $n_{1k} + n_{2k} = n_k$ ) such that

$$\frac{n_{1k} + \alpha_0 + \beta_0 + 3}{n_{2k} + \gamma_0 + \delta_0 + 3} = \left\{ \begin{array}{l} R_0 \text{ when the loss function is} \\ \text{constant and symmetric} \\ R_0^* \text{ when the loss function is} \\ \text{linear} \end{array} \right\},$$

where  $R_0$  and  $R_0^*$  are as defined in (4.1) and (6.3) when  $k = 0$ . If the loss function is linear and the prior distribution is rectangular, take  $n_{1k}$  and  $n_{2k}$  such that

$$\frac{n_{1k} + 3}{n_{2k} + 3} = \left| \frac{k_1}{k_2} \right| .$$

Of course, one might study these procedures when the random variables generated at each stage are other than normal or binomial (e.g., Poisson or exponential). Also, the case in which one has normal random variables with unknown variance generated in a stage-by-stage procedure would be of interest.

Two problems connected with this thesis for which the author has so far been unable to find a workable solution are extensive-form analysis when the random variables generated at each stage are normal with known variance and the stage-by-stage procedure with normal random variables with known variance and a constant but non-symmetric loss function ( $q_1 \neq q_2$ ).

The author hopes to do more work on the proposed stopping rule. A stopping rule for the approximate procedure with a symmetric constant loss function would not be difficult to derive. However, the computations necessary for it might be rather long if there are many stages. It would be interesting to determine if the corresponding stopping rule for the linear loss function can be used as a rough approximation.

In addition, since some of the greatest gains in using the procedures proposed in this thesis can be achieved by reducing sampling costs by the use of the stopping rule, these procedures with the stopping rule should be compared

with other procedures in which the total number of observations is a random variable such as the sequential analysis procedure proposed by Bross (1952).



## XI. SUMMARY

The general problem considered in this thesis is the following:

Suppose one desires to compare two drugs, Drug 1 and Drug 2, (or any two binomial populations) in a  $m$ -stage test procedure. The probability that Drug 1 results in an improvement in a patient in any one of the stages is  $\theta_1$ , and the corresponding probability for Drug 2 is  $\theta_2$ . We assume that it has been decided prior to the experiment that  $n_k$  observations will be taken in the  $k$ 'th stage ( $k = 1, 2, \dots, m$ ). We want to partition  $n_k$  into  $n_{1k}$  and  $n_{2k}$ , the number of observations allocated to Drug 1 and Drug 2, respectively, on the basis of the results of the observations in the previous  $k-1$  stages. After the last stage is taken, we wish to make the terminal decision that either  $\theta_1 > \theta_2$  or  $\theta_1 < \theta_2$ .

We let  $\theta = (\theta_1, \theta_2)$  and let  $x_k = (r_{1k}, r_{2k})$ , where  $r_{1k}$  and  $r_{2k}$  are the numbers of successes out of  $n_{1k}$  and  $n_{2k}$  respectively. We let  $a_k$  be the decision of what size  $n_{1k}$  (and thus  $n_{2k} = n_k - n_{1k}$ ) should be and let  $b$  be the terminal decision, where  $b = 1$  represents the decision that  $\theta_1 > \theta_2$  and  $b = 2$  represents the decision that  $\theta_1 < \theta_2$ .

We assume a prior distribution, or weighting function, for  $\theta$  of

$$(10.1) \quad \rho(\theta) = \frac{\theta_1^{\alpha_0} (1 - \theta_1)^{\beta_0} \theta_2^{\gamma_0} (1 - \theta_2)^{\delta_0}}{B(\alpha_0+1, \beta_0+1) B(\gamma_0+1, \delta_0+1)},$$

where  $\alpha_0, \beta_0, \gamma_0,$  and  $\delta_0$  are non-negative integers,  $0 \leq \theta_1 \leq 1,$  and  $0 \leq \theta_2 \leq 1.$  We wish to find strategies for finding  $a_1, a_2, \dots, a_m,$  and finally  $b,$  which are optimum relative to either the constant loss function

$$(10.2) \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) = \left\{ \begin{array}{l} q_1 \text{ if } b = 1 \text{ and } \theta_1 < \theta_2 \\ q_2 \text{ if } b = 2 \text{ and } \theta_1 > \theta_2 \\ 0 \text{ otherwise} \end{array} \right\},$$

where  $q_1$  and  $q_2$  are positive real numbers, or the linear loss function

$$(10.3) \quad L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta) = \left\{ \begin{array}{l} k_{10} + k_{11} \theta_1 + k_{12} \theta_2 \text{ if } b = 1 \\ k_{20} + k_{21} \theta_1 + k_{22} \theta_2 \text{ if } b = 2 \end{array} \right\},$$

where each  $k_{ij}$  ( $i = 1, 2, ; j = 0, 1, 2$ ) is a real number. In order to do this we use extensive-form analysis, which proceeds as follows:

- (1) For fixed values of  $x_1, \dots, x_m, a_1, \dots, a_m,$  and  $b$  compute  $E_{\theta|x_1, \dots, x_m} L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta).$
- (2) For fixed  $x_1, \dots, x_m, a_1, \dots, a_m$  compute  $L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) =$

$$\min_b E_{\theta|x_1, \dots, x_m} L(x_1, \dots, x_m, a_1, \dots, a_m, b, \theta)$$

and choose the "b" which gives the minimum.

(3) For fixed values of  $x_1, \dots, x_{m-1}$  and  $a_1, \dots, a_m$

$$\text{compute } E_{x_m|x_1, \dots, x_m} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m).$$

(4) Then for fixed values of  $x_1, \dots, x_{m-1}, a_1, \dots, a_{m-1}$

compute

$$\begin{aligned} &L^{(2)}(x_1, \dots, x_{m-1}, a_1, \dots, a_{m-1}) \\ &= \min_{a_m} E_{x_1, \dots, x_{m-1}} L^{(1)}(x_1, \dots, x_m, a_1, \dots, a_m) \end{aligned}$$

and choose the "a<sub>m</sub>" which gives the minimum.

(5) Continue in this manner with each stage until the first stage.

(6) On the first stage choose  $a_1$  such that

$$E_{x_1} L^{(m)}(x_1, a_1) \text{ is a minimum.}$$

Expressions for the expected values in Steps 1 to 6 are given in Sections 2.3 and 2.4 of this thesis. A program for extensive-form analysis was written for the I.B.M. 1620 computer, and some examples were worked out with the use of this program. It was found that, even with the computer, only small examples could be handled because of the time required.

Because of this difficulty the stage-by-stage procedure was proposed. In this procedure it is assumed that at the k'th stage ( $k = 1, 2, \dots, m$ ) there are only k stages in the

entire procedure and that one is at the terminal stage. Thus at the  $k$ 'th stage ( $l = 2, 3, \dots, m$ ) one evaluates

$$E_{x_k | x_1, \dots, x_{k-1}} \min_b E_{\theta | x_1, \dots, x_k} L(x_1, \dots, x_k, a_1, \dots, a_k, b, \theta)$$

and chooses the " $a_k$ " which minimizes the expression. At the

first stage one evaluates  $E_{x_1} \min_b E_{\theta | x_1} L(x_1, a_1, b, \theta)$  and

chooses the " $a_1$ " which gives the minimum. Expressions for these expected values are given in Section 3.1 for both the linear and constant loss functions. It is shown that, when we are at the  $k$ 'th stage, we can assume we really have only a one stage procedure with an altered prior distribution. If we have a linear loss function with only a few observations in each stage, we can do the computations for the stage-by-stage procedure without a digital computer. With a computer much larger examples can be worked when the stage-by-stage procedure is used than when extensive-form analysis is used.

Finally, the approximate procedure, which requires very simple computations and certainly does not require a computer, was proposed. It consists of minimizing  $\text{Var}(\theta_1 - \theta_2 | x_1, \dots, x_k)$  at each stage. It is shown that this is approximately equivalent to taking

$$(10.4) \quad n_{1k} \approx \frac{(\gamma_{k-1} + \delta_{k-1} + n_k + 3) R_{k-1} - \alpha_{k-1} - \beta_{k-1} - 3}{R_{k-1} + 1},$$

where  $\alpha_{k-1} = \sum_{i=1}^{k-1} r_{1i} + \alpha_0$  ,

$$\beta_{k-1} = \sum_{i=1}^{k-1} (n_{1i} - r_{1i}) + \beta_0$$
 ,

$$\gamma_{k-1} = \sum_{i=1}^{k-1} n_{2i} + \gamma_0$$
 ,

$$\delta_{k-1} = \sum_{i=1}^{k-1} (n_{2i} - r_{2i}) + \delta_0$$
 ,

and  $R_{k-1} = \left[ \frac{\bar{p}_{1,k-1} (1 - \bar{p}_{1,k-1})}{\bar{p}_{2,k-1} (1 - \bar{p}_{2,k-1})} \right]^{\frac{1}{2}}$  ,

where  $\bar{p}_{1,k-1} = \frac{\alpha_{k-1} + 1}{\alpha_{k-1} + \beta_{k-1} + 2}$

and  $\bar{p}_{2,k-1} = \frac{\gamma_{k-1} + 1}{\gamma_{k-1} + \delta_{k-1} + 2}$  .

The approximate procedure and the stage-by-stage procedure were compared in several different examples with extensive-form analysis and equal division of all the observations between Drug 1 and Drug 2 by the computation of the probability of choosing  $b = 1$  for different combinations of  $\theta_1$  and  $\theta_2$ . It was assumed for these computations that the prior distributions of  $\theta_1$  and  $\theta_2$  were rectangular ( $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = 0$ ), the constant loss function had  $q_1 = q_2 = 1$ , and the linear loss function had  $k_{10} = k_{20} = 0$ ,  $k_{11} = k_{22} = -1$ , and

$$k_{12} = k_{21} = 1.$$

It was found that for  $\theta_1 = .6$  and  $\theta_2 = .4$  the approximate procedure and equal division of the observations gave, in general, the highest probability of choosing the terminal decision  $b = 1$ . For  $\theta_1 = .95$  and  $\theta_2 = .80$  extensive-form analysis gave in practically every case the highest probability of choosing  $b = 1$ . The decrease in probability of  $b = 1$  from extensive-form analysis to the stage-by-stage procedure was, however, relatively quite small in general whereas the decrease from the stage-by-stage procedure to the approximate procedure was relatively large. In addition, in all the examples studied there was very little to no decrease when equal division rather than the approximate procedure was used. Thus it appears that, if one has a computer, he should use the stage-by-stage procedure but that, if one must have relatively simple computations, he should use the approximate procedure since there are some gains over equal division in large examples for which  $\theta_1$  and  $\theta_2$  are close to 0 or 1 even though these gains are small.

The stage-by-stage procedure for the case that normal variables with known variances rather than binomial variables are generated at each stage and the prior distributions of  $\theta_1$  and  $\theta_2$  are normal was derived for the general linear loss function and for the constant loss function with  $q_1 = q_2 = 1$ . It was shown that the optimum stage-by-stage procedure for

the linear loss function is to take  $n_{1k}$  such that

$$(10.5) \quad \frac{N_{1,k-1} + n_{1k}}{N_{2,k-1} + n_k - n_{1k}} = \left| \frac{k_1}{k_2} \right| \frac{\sigma_1}{\sigma_2} ,$$

where  $N_{1,k-1} = \sum_{i=0}^{k-1} n_{1i}$  ,  $N_{2,k-1} = \sum_{i=0}^{k-1} n_{2i}$  ,  $n_{10}$  and  $n_{20}$  are parameters of the normal prior distributions of  $\theta_1$  and  $\theta_2$ ,  $n_0 = n_{10} + n_{20}$ ,  $k_j = k_{1j} - k_{2j}$  for  $j = 1, 2$ , and  $\sigma_1$  and  $\sigma_2$  are the standard deviations of Populations 1 and 2. For the constant loss function with  $q_1 = q_2 = 1$  the optimum stage-by-stage procedure is to take  $n_{1k}$  such that

$$(10.6) \quad \frac{N_{1,k-1} + n_{1k}}{N_{2,k-1} + n_k - n_{1k}} = \frac{\sigma_1}{\sigma_2} .$$

The approximate procedure (10.4) was then derived from (10.5) with  $|k_1/k_2|=1$  and (10.6) as a normal approximation to the binomial distribution. However, if one left the term  $|k_1/k_2|$  in (10.5) in the derivation, one would obtain the "generalized" approximate procedure with  $R_{k-1}$  replaced by  $R_{k-1}^* = |k_1/k_2| R_{k-1}$  .

If the costs of sampling are large compared with the losses associated with the possible terminal decisions, we shall most likely want to stop the experimental procedure before reaching the last possible stage and make our terminal decision at that point. Suppose we have run  $k-1$  stages and want to decide whether or not to take the  $k$ 'th stage

( $k = 1, 2, \dots, m$ ). The stopping rule used for the stage-by-stage and the approximate procedures in this thesis tells us not to take the  $k$ 'th stage if

$$\begin{aligned} & \min_b E_{\theta} | x_1, \dots, x_{k-1} L(x_1, \dots, x_{k-1}, a_1, \dots, a_{k-1}, b, \theta) \\ & \leq \min_{k'} \left[ \sum_{i=k}^{k'} C_i + \min_{a_{kk'}} E_{x_{kk'}} | x_1, \dots, x_{k-1} \right. \\ & \quad \left. \cdot \min_b E_{\theta} | x_1, \dots, x_{k-1}, x_{kk'} L(x_1, \dots, x_{k-1}, a_1, \dots, a_{k-1}, a_{kk'}, b, \theta) \right] \end{aligned}$$

where  $N_{kk'} = \sum_{i=k}^{k'} n_i$ ,  $N_{1kk'}$  is the number of observations on Population 1 out of  $N_{kk'}$ ,  $N_{2kk'}$  is the number of observations on Population 2 out of  $N_{kk'}$ ,  $R_{1kk'}$  and  $R_{2kk'}$  are the numbers of successes out of  $N_{1kk'}$  and  $N_{2kk'}$ , respectively,  $C_i$  is the cost of the  $i$ 'th stage,  $a_{kk'}$  is the decision of what  $N_{1kk'}$  to take, and  $k' = k, k+1, \dots, m$ . This decision rule is applied to the stage-by-stage procedure for both the linear and constant loss functions. It is also applied to the "generalized" approximate procedure when a linear loss function is assumed. In this case one can apply the stopping rule without using a computer when he uses charts which are given in Raiffa and Schlaifer's book (1961).



### ACKNOWLEDGMENTS

The author especially wishes to express his gratitude to Dr. H. A. David for his guidance and valuable assistance in the research for this thesis.

The author is also grateful to Dr. Boyd Harshbarger for his encouragement and assistance. He is especially grateful for the fellowship which he has received during the past two years from the National Institutes of Health. In addition, he would like to express his appreciation to the other members of the Department of Statistics who have made helpful suggestions and criticisms.

Finally, the author wishes to thank Mr. Walter Johnston for preparing the figures and Mrs. Barbara Coffey for the care with which she typed the final manuscript.

BIBLIOGRAPHY

- Amster, Sigmund J. (1962). A modified Bayes stopping rule. Institute of Statistics Mimeo Series, No. 318, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.
- Bradt, R. N., Johnson, S. M., Karlin, S. (1956). On sequential designs for maximizing the sum of  $n$  observations. Ann. Math. Statist. 27 1060-1074.
- Bross, I. (1952). Sequential sampling plans. Biometrics 8 188-205.
- Chernoff, H. (1959). Sequential design of experiments. Ann. Math. Statist. 30 755-770.
- Dunnett, C. W. (1960). Statistical theory of drug screening. Proceedings of the Symposium on Quantitative Methods in Pharmacology, Biometric Society, University of Leyden. 212-231.
- National Bureau of Standards (1959). Tables of the Bivariate Normal Distribution Function and Related Functions. National Bureau of Standards Applied Mathematics Series No. 50.
- Pearson, Karl (Ed.) (1934). Tables of the Incomplete Beta-Function. The Biometrika Office, University College, London.
- Raiffa, Howard, and Schlaifer, Robert (1961). Applied Statistical Decision Theory. Graduate School of Business Administration, Harvard University, Boston.
- Robbins, H. E. (1952). Some aspects of the sequential design of experiments. Bull. Amer. Math. Soc. 58 527-535.
- Robbins, H. (1956). A sequential decision problem with a finite memory. Proc. Nat. Acad. Sci. 42 920-923.
- Taylor, R. J., and David, H. A. (1962). A multi-stage procedure for the selection of the best of several populations. J. Amer. Statist. Assoc. 57 785-796.

Vogel, Walter (1960). A sequential design for the two armed bandit. Ann. Math. Statist. 31 430-443.

Vogel, Walter (1960). An asymptotic minimax theorem for the two armed bandit problem. Ann. Math. Statist. 31 444-451.

Wald, A. (1947). Sequential Analysis. Wiley, New York.

VITA

Armand Verne Smith, Jr., was born September 20, 1935, in Chicago, Illinois. He graduated from Carmel High School in Carmel, Indiana, in 1953.

He then entered Purdue University in Lafayette, Indiana, and received the Bachelor of Science in Chemical Engineering degree in 1957. He continued his studies in the Department of Mathematics and Statistics of Purdue University and received the Master of Science degree in 1959.

After being an instructor of mathematics at the University of Dayton in Dayton, Ohio, for two years, he resumed his graduate studies in the Department of Statistics of the Virginia Polytechnic Institute in June, 1961. He has been supported in his work by a fellowship from the National Institutes of Health.

*Armand V. Smith, Jr.*

## ABSTRACT

The general problem considered in this thesis is to determine an optimum strategy for deciding how to allocate the observations in each stage of a multi-stage experimental procedure between two binomial populations (e.g., the numbers of successes for two drugs) on the basis of the results of previous stages. After all of the stages of the experiment have been performed, one must make the terminal decision of which of the two populations has the higher probability of success. The optimum strategy is to be optimum relative to a given loss function; and a prior distribution, or weighting function, for the probabilities of success for the two populations is assumed. Two general classes of loss functions are considered, and it is assumed that the total number of observations in each stage is fixed prior to the experiment.

In order to find the optimum strategy a method of analysis called extensive-form analysis is used. This is essentially a method for enumerating all the possible outcomes and corresponding strategies and choosing the optimum strategy for a given outcome. However, it is found that this method of analysis is much too long for all but small examples even when a digital computer is used.

Because of this difficulty two alternative procedures,

which are approximations to extensive-form analysis, are proposed.

In the stage-by-stage procedure one assumes that at each stage he is at the last stage of his multi-stage procedure and allocates his observations to each of the two populations accordingly. It is shown that this is equivalent to assuming at each stage one has a one stage procedure.

In the approximate procedure one (approximately) minimizes the posterior variance of the difference of the probabilities of success for the two populations at each stage. The computations for this procedure are quite simple to perform.

The stage-by-stage procedure for the case that the two populations are normal with known variance rather than binomial is considered. It is then shown that the approximate procedure can be derived as an approximation to the stage-by-stage procedure when normal approximations to binomial distributions are used.

The three procedures are compared with each other and with equal division of the observations in several examples by the computation of the probability of making the correct terminal decision for various values of the population parameters (the probabilities of success). It is assumed in these computations that the prior distributions of the population parameters are rectangular distributions and that the loss

functions are symmetric; i.e., the losses are as great for one wrong terminal decision as they are for the other.

These computations show that, for the examples studied, there is relatively little loss in using the stage-by-stage procedure rather than extensive-form analysis and relatively little gain in using the approximate procedure instead of equal division of the observations. However, there is a relatively large loss in using the approximate procedure rather than the stage-by-stage procedure when the population parameters are close to 0 or 1.

At first it is assumed there are a fixed number of stages in the experiment, but later in the thesis this restriction is weakened to the restriction that only the maximum number of stages possible in the experiment is fixed and the experiment can be stopped at any stage before the last possible stage is reached. Stopping rules for the stage-by-stage and the approximate procedures are then derived.