Optimal flexible two-stage plans

Seoung-gon Ko

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Biostatistics Commons, and the Industrial Engineering Commons

Recommended Citation
Ko, Seoung-gon, "Optimal flexible two-stage plans " (1994). Retrospective Theses and Dissertations. 10490.
https://lib.dr.iastate.edu/rtd/10490

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Optimal flexible two-stage plans

Ko, Seoung-gon, Ph.D.

Iowa State University, 1994
Optimal flexible two-stage plans

by

Seoung-gon Ko

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Approved
Signature was redacted for privacy.
In Charge of Major Work
Signature was redacted for privacy.
For the Major Department
Signature was redacted for privacy.
For the Graduate College

Iowa State University
Ames, Iowa
1994
# TABLE OF CONTENTS

## CHAPTER 1. INTRODUCTION ............................................................... 1

## CHAPTER 2. BAYES AND NEYMAN-PEARSON ASPECTS .................... 5

  2.1 Introduction ............................................................................. 5
  2.2 The Neyman-Pearson/Bayes connection ..................................... 6
    2.2.1 Lagrangian solutions of constrained optimization problems ..... 6
    2.2.2 The Bayes/Neyman-Pearson connection in one-sample case ..... 7
  2.3 Extension to two-sample case .................................................... 16
    2.3.1 One-sided case ............................................................... 16
    2.3.2 Two-sided case ............................................................... 19
  2.4 Fair comparisons with optimal standard two-stage plans .......... 22
  2.5 Remarks concluding chapter 2 .................................................. 28

## CHAPTER 3. GROUP-SEQUENTIAL ASPECTS ................................. 31

  3.1 Introduction ............................................................................. 31
  3.2 Formulation ............................................................................. 33
  3.3 Optimization ............................................................................ 35
    3.3.1 Optimization for given \( \mu_0, \mu_1 \), \( \mu_0 < \mu_1 \) .......... 35
    3.3.2 Optimal plan for linear transformations of \( \mu_0, \mu_1 \), \( \mu_0 < \mu_1 \) 44
  3.4 Alternate plans ........................................................................ 54
3.4.1 Single Sampling Plan (SS) ........................................ 55
3.4.2 Standard Two-stage Plans (STP) .............................. 55
3.4.3 Sequential Probability Ratio Test (SPRT) .................... 56

3.5 Numerical illustrations and comparisons .......................... 57

3.6 Applications to binomial (Bernoulli) case ....................... 58
  3.6.1 One-sample case ........................................... 58
  3.6.2 Illustrations: one-sample case ................................ 71
  3.6.3 Two-sample case ........................................... 77
  3.6.4 Illustrations: two-sample case ............................. 83

3.7 Remarks concluding chapter 3 .................................. 97

CHAPTER 4. CONCLUSION ............................................. 100

BIBLIOGRAPHY ....................................................... 102

ACKNOWLEDGMENTS ................................................. 105
LIST OF TABLES

Table 2.1: Relation between Bayes losses and Neyman-Pearson error rates 12
Table 2.2: The plan $\delta^*$ ................................................................. 13
Table 2.3: Second-stage sample size $M_2(d)$: $M_1 = 20$ ............... 21
Table 2.4: $\delta^o$ and the four "nearest" plans $\hat{\delta}$ for the example in section 2.2.2 28
Table 2.5: $\delta^o$ and the four "nearest" plans $\hat{\delta}$ for the example in section 2.3.1 29
Table 3.1: Approximate binomial plan corresponding Wiener plan ... 65
Table 3.2: The approximate optimal flexible binomial plan for $\Delta = 0.1$ . 72
Table 3.3: The approximate optimal flexible binomial plan for $\Delta = 0.15$ 74
Table 3.4: The approximate optimal flexible binomial plan for $\Delta = 0.2$ . 76
Table 3.5: Approximate optimal first-stage sample size $n_0^*$ for two-sample case ................................................................. 84
Table 3.6: Approximate optimal flexible binomial two-stage plans for two-sample case: $(p_0, p_1) = (0.2, 0.3)$ ......................... 85
Table 3.7: Approximate optimal flexible binomial two-stage plans for two-sample case: $(p_0, p_1) = (0.2, 0.35)$ ......................... 88
Table 3.8: Approximate optimal flexible binomial two-stage plans for two-sample case: \((p_0, p_1) = (0.2, 0.4)\) ......... 91

Table 3.9: Approximate optimal flexible binomial two-stage plans for two-sample case: \((p_0, p_1) = (0.5, 0.55)\) ......... 94
LIST OF FIGURES

Figure 2.1: ASN functions of $\delta^*$ and $\delta^o$ ........................................ 14
Figure 2.2: OC functions of $\delta^*$ and $\delta^o$ ........................................ 15
Figure 2.3: "Continuation region" and hypotheses for $\delta^*$ ............... 20
Figure 2.4: Typical continuation points "*" and "Continuation region"
boundary contours for $M_1 = 20$ ...................................................... 23
Figure 2.5: Second-stage critical regions $A_1$ for $M_2(d) = 40$ for $d = (3, 3)$ 24
Figure 2.6: Second-stage critical regions $A_1$ for $M_2(d) = 40$ for $d = (10, 15)$ 25
Figure 3.1: Contours of $L_1(T_1, k_1)$ when $(T_0, \rho_0, \rho_1) = (18.9, 193.31, 109.89)$ 39
Figure 3.2: Function $L_1(T_1, k_1^*)$ when $(T_0, \rho_0, \rho_1) = (18.9, 193.31, 109.89)$ 40
Figure 3.3: Candidates of solutions for $T_i$ for each signed equation: $r = 6$ 45
Figure 3.4: Optimal second-stage sampling efforts ............................... 59
Figure 3.5: Optimal second-stage critical values ................................. 60
Figure 3.6: $OC_2$ of $STP^*$ and $OFP^*$ for the Wiener process $w(0, 1)$ and
$w(0.5, 1)$ with $(\alpha, \beta) = (0.05, 0.10)$ ....................................... 61
Figure 3.7: $AST_2$ of $STP^*$ and $OFP^*$ for the Wiener process $w(0, 1)$ and
$w(0.5, 1)$ with $(\alpha, \beta) = (0.05, 0.10)$ ....................................... 62
Figure 3.8: $OC_2$ of $SPRT, OFP^*, STP^*$ and $SS$, for the Wiener process
$w(0.1)$ and $w(0.5, 1)$ with $(\alpha, \beta) = (0.05, 0.10)$ ................. 63
Figure 3.9: $AST_{\mu}$ of $SPRT.OFP^*, STP^*$ and $SS.$ for the Wiener process $w(0.1)\text{ and } w(0.5.1)\text{ with } (\alpha, \beta) = (0.05, 0.10)$
CHAPTER 1. INTRODUCTION

In this dissertation, we construct optimal "flexible" two-stage plans. "Flexible" means that the first-stage outcomes not only determine whether to accept the null or alternative hypothesis, or whether to proceed to the second stage, but also, in the latter case, determine second-stage sample sizes and critical values that depend on first-stage outcome. "Optimal" is meant in the sense that a suitable measure of sampling effort is minimized with respect to all flexible plans, subject to suitable constraints.

Determining or adjusting second-stage sample size and critical region based on first-stage outcome constitutes an extension of classical double sampling plans as originally proposed (Dodge and Romig [13]) and subsequently developed (Hald [16]) and Spurrier and Hewett [26]). The idea in a sense dates back to Stein's [27] seminal paper and its descendants, through the adaptive utilization of first-stage information, as detailed below, focuses on the parameter(μ) of interest, rather than on a nuisance parameter(σ) as in Stein's. Such focus on the parameter of interest is particularly reminiscent of the approach to sequential estimation indicated by Birnbaum and Healy [3]; and also of Miller and Freund's [23] more recent proposal of an ad hoc method for determining the second-stage sample size in binomial estimation problems.

The literature on optimized two-stage plans deals only with "non-flexible" plans:
plans, in other words, that call for first-stage decision or for a second stage with fixed sample size and critical region. Colton and McPherson [9] focus on first-stage rejection. They minimize alternative-hypothesis expected sample size with respect to first- and second-stage sample sizes and critical values, under fixed error rates of both kinds. Thall et al. [28], on the other hand, focus instead on first-stage acceptance and minimize the average of the expected sample sizes at the null and alternative hypotheses. Thall et al. [29], [30] treat multiple testing in similar fashion.

There have also been Bayesian approaches to two-stage plans: for example, Berry [2] reformulates McPherson's [21] approach in Bayesian terms, by also assuming a maximum total sample size and equal stage sample sizes, but allowing for the possibility of early termination in accordance with a certain Bayesian stopping rule. Cohen and Sackrowitz [7], [8] derive optimal Bayes procedures for determining the second-stage sample size in exponential family estimation, with special attention devoted to the binomial case.

In the group-sequential (Pocock [25], O'Brien and Fleming [24]) framework, McPherson [21] minimizes various measures of sampling effort, with respect to the number of "equally spaced" interim analyses, under a repeated-significance-test convention for assigning interim critical values, and fixed overall error rates of both kinds. Jennison [17] and Eales and Jennison [15], on the other hand, fix the number of interim analyses but optimize over the corresponding interim critical values. In the Bayesian framework, Cressie and Morgan [12] studied certain k-stage probability ratio tests called "variable-sample-size probability ratio tests (VPRTs)". This latter work is related to the "sequential design of experiments" initiated by Chernoff [6] (see also Bradt and Karlin [5], and Whittle [32] and Borwanker, David and Ingwell
Motivated by the above considerations and developments, we have considered certain two-stage plans through two distinct optimization formulations. While our “flexible” formulations in both cases would appear to call for a multi-dimensional optimization, these turn out to be “separable” into a family of simple two-dimensional optimizations, each corresponding to a first-stage outcome in the continuation region.

The first formulation, in Chapter 2, addressing the case of one treatment both with and without a control, is distinguished by possessing a Neyman-Pearson as well as Bayes interpretation, reminiscent of the dual interpretation of the SPRT as optimal in both a Bayes and Neyman-Pearson sense (Lehmann [20]. p.104). The Neyman-Pearson interpretation is that average expected sample size is being minimized, subject just to the two overall error rates $\alpha$ and $\beta$, respectively of first and second kind. The Bayes interpretation is that Bayes risk, involving both sampling cost and wrong decision losses, is being minimized. With $C_0$ the cost of wrongly abandoning null hypothesis, and $C_A$ the cost of wrongly staying with it, one useful outgrowth of our dual interpretation is that $(\alpha, \beta)$ and $(C_0, C_A)$ determine each other.

The second formulation, reminiscent of the group-sequential point of view, involves given first- and second-stage error rates of first and second kind (four such given error rates in all), with objective function equal to average (over the hypotheses) expected sample size. This is the formulation in chapter 3, where the single treatment case, with and without control, is considered. We choose to analyze this group-sequential formulation by means of the arcsine transformation [14] to normality of the binomial distribution. This allows translating the binomial problem to one
of testing the drift parameter of a standard Wiener process [18]. This brings to the analysis the simplifications and economies of location and scale equivariance.

The flexible plans provide an alternative to informal sample size adjustment following an interim analysis, as is sometimes done, for example, in the pharmaceutical industry, in the case of drugs already on the market being explored for new indications: equally, they in effect incorporate an initial (pilot) study into an over-all experimental plan, both when the pilot study is already done and when a pilot stage is to be incorporated into a planned-for two-stage study.

Comparisons are made with optimal non-flexible plans, which leads to conclusion that flexible plans reduce expected sample size in both formulations.
CHAPTER 2. BAYES AND NEYMAN-PEARSON ASPECTS

2.1 Introduction

Our initial motivation related to this chapter was to develop procedures in a Neyman-Pearson framework, which will (a) control errors of both kinds and (b) decrease average sampling number (ASN) by allowing second-stage sample size to depend on first-stage outcome.

As work progressed on our original Neyman-Pearson optimization formulation, we observed that the plans we were developing from an essentially Neyman-Pearson perspective had also a Bayesian interpretation, much in the way that the sequential probability ratio test simultaneously is optimal in both a Neyman-Pearson and a Bayesian sense (Lehmann [20], p.181). In particular, we observed that Neyman Pearson ASN optimization under fixed global error rates in fact is implemented by certain second-stage Bayes solutions. In other words, the plans are optimal simultaneously in both a Bayes and a Neyman-Pearson sense: the former because they are Bayes in a class \( \Gamma \) of flexible plans; the latter because they minimize a suitable measure of sampling effort with respect to all members of \( \Gamma \) satisfying certain error rate restrictions of first and second kind.

Section 2.2 treats the analysis yielding optimal flexible plans in the case of a single population where this optimization is seen to have simultaneously a Bayesian
and Neyman-Pearson character, due to the availability of the Lagrangian approach to solving constrained optimization problems. Bayesian wrong-decision losses are thereby put into correspondence with Neyman-Pearson wrong-decision error rates. Section 2.2.2 compares a particular optimal flexible with the corresponding optimal non-flexible two-stage plan. Section 2.4 describes a certain experimentwise randomization useful for this comparison.

Section 2.3 discusses extension of the ideas to the case of two populations, and Section 2.5 contains concluding remarks.

2.2 The Neyman-Pearson/Bayes connection

2.2.1 Lagrangian solutions of constrained optimization problems

Consider an arbitrary domain \( \Delta \) with elements \( \delta \) and three functions \( f(\delta) \), \( g_0(\delta) \) and \( g_1(\delta) \) defined on \( \Delta \). Suppose that we wish to find an optimum element of \( \Delta \) in the sense of minimizing \( f(\delta) \), subject to the restrictions \( g_0(\delta) \leq A \) and \( g_1(\delta) \leq B \).

If there is a \( \delta^* \) in \( \Delta \), and also two non-negative numbers \( \lambda_0 \) and \( \lambda_1 \), such that

\[
\begin{align*}
g_0(\delta^*) &= A, \\
g_1(\delta^*) &= B \\
\end{align*}
\]

(2.1)

and

\[
L(\delta^*) \leq L(\delta), \quad \delta \in \Delta
\]

(2.2)

where

\[
L(\delta) \equiv f(\delta) + \lambda_0 g_0(\delta) + \lambda_1 g_1(\delta).
\]
then $\delta^*$ is optimal in the above sense.

This assertion is verified as follows: Suppose $\delta^*$ satisfied (2.1) and (2.2), and were not optimal; in other words, suppose there were a $\delta'$ in $\Delta$ with

$$f(\delta') < f(\delta^*).$$

and also

$$g_0(\delta') \leq A,$$
$$g_1(\delta') \leq B.$$  \hfill (2.1)

Then we would have

$$L(\delta') = f(\delta') + \lambda_0(g_0(\delta') - A) + \lambda_1(g_1(\delta') - B) + \lambda_0A + \lambda_1B$$
\leq f(\delta') + \lambda_0A + \lambda_1B$$
\leq f(\delta^*) + \lambda_0A + \lambda_1B$$

$$= f(\delta^*) + \lambda_0g_0(\delta^*) + \lambda_1g_1(\delta^*)$$
$$= L(\delta^*).$$

where the middle three signs would be due respectively to (2.4), (2.3) and (2.1), and where the implied inequality

$$L(\delta') < L(\delta^*).$$

would contradict (2.2).

2.2.2 The Bayes/Neyman-Pearson connection in one-sample case

Consider an experimenter in possession of $d$ positives out of an initial sample of size $M_1$ from a Bernoulli population, for which the experimenter is willing to assume
that \( p \) equals either \( p_0 \) or \( p_A > p_0 \). Suppose as well that the experimenter is willing to restrict his/her further options either to sampling no more (\( M_0(d) = 0 \)), with an immediate decision whether \( p = p_0 \) or \( p_A \), or to conducting one further single sampling plan (\( M_2(d) > 0 \)); and that the experimenter intends to choose from among these options in Bayesian terms, involving, among other things, two cost ratios: the ratio \( C_0 \) (respectively, \( C_A \)) of the cost of wrongly deciding that \( p_A \) (respectively, \( p_0 \)) holds, to the unit cost of sampling.

With prior \( \pi_0 (0 < \pi_0 < 1) \) on \( p \), the posterior probability \( \pi_0(d) \) of \( p_0 \) is

\[
\frac{\pi_0 b(d; M_1, p_0)}{\pi_0 b(d; M_1, p_0) + \pi_A b(d; M_1, p_A)}
\]

where \( b(\cdot) \) denotes binomial probability, and correspondingly for \( \pi_A = 1 - \pi_0, p_A \) and \( \pi_A(d) = 1 - \pi_0(d) \).

The experimenter’s single sampling Bayes plan \( S^*(d) \) at \( d \) is the (possibly degenerate) single sampling plan \( S(d) \) minimizing

\[
R(d) = M_2(d) + C_0 \pi_0(d) P_A|0(d) + C_A \pi_A(d) P_0|A(d).
\]

where \( P_A|0(d) \) (respectively, \( P_0|A(d) \)) denotes the probability, under \( S(d) \), of accepting \( p_A \) (respectively, \( p_0 \)) when \( p = p_0 \) (respectively, \( p_A \)). Here, \( P_A|0(d) \) and \( P_0|A(d) \) are based, respectively, on the binomial probabilities \( b(D; M_2(d), p_0) \) and \( b(D; M_2(d), p_A) \) of the second-stage number of “positives”, \( D \). \( D = 0,1,2,..., M_2(d) \).

\( S^*(d) \) is derived by first fixing \( M_2(d) \) at, say, \( m_2 \) and minimizing \( R(d) \) over critical regions, followed by minimizing over \( m_2 \). \( m_2 = 0,1,2,... \). For \( m_2 = 0 \), minimizing \( R(d) \) amounts to accepting \( p_0 \) (respectively, \( p_A \)) if \( \frac{C_0 \pi_0(d)}{C_A \pi_A(d)} \) is greater (respectively, less) than 1. For \( m_2 > 0 \), in view of the extended Neyman-Pearson...
Lemma. the Bayes plan amounts to accepting $p_0$ (respectively, $p_A$) if the criterion

$$\frac{C_0 \pi_0(d) b(D; m_2, p_0)}{C_A \pi_A(d) b(D; m_2, p_A)}$$

is greater (respectively, less) than 1. yielding a straightforward one-sided plan, by likelihood ratio monotonicity.

$\delta^* \equiv (S^*(0), S^*(1), \ldots, S^*(M_1))$ has the following interpretations.

(i) As indicated in the preceding discussion. $S^*(d)$ is the single sampling procedure that a “one-stage” Bayesian will follow who, having obtained $d$ initial positives, uses prior $(\pi_0(d), \pi_A(d))$ on $(p_0, p_A)$.

(ii) Consider a “two-stage” Bayesian about to take, rather than already in possession of, an initial sample of size $M_1$, possibly to be followed by a second-stage single sampling plan, who uses prior $(\pi_0, \pi_A)$ on $(p_0, p_A)$. Suppose that this Bayesian has, as his options, vectors $S \equiv (S(0), S(1), \ldots, S(d), \ldots, S(M_1))$, where each $S(d)$ is an arbitrary second-stage single sampling plan. $\delta^*$ minimizes the Bayes risk $R_\delta$, for such a Bayesian, where

$$R_\delta = \pi_0 \sum_{d=0}^{M_1} b(d; M_1, p_0) \left\{ [M_2(d)]_\delta + C_0 [P_0|A(d)]_\delta \right\}$$

$$+ \pi_A \sum_{d=0}^{M_1} b(d; M_1, p_A) \left\{ [M_2(d)]_\delta + C_A [P_0|A(d)]_\delta \right\}.$$  \hspace{1cm} (2.6)

Here, the dependence of $R_\delta$ on $\delta$ comes from the fact that $M_2(d), P_0|A(d)$ and $P_0|A(d)$ are determined by $S(d)$, and hence by $\delta$.

(iii) A third interpretation of $\delta^*$ is seen by writing (2.6) as

$$R_\delta = \pi_0 \sum_{d=0}^{M_1} b(d; M_1, p_0) [M_2(d)]_\delta + \pi_A \sum_{d=0}^{M_1} b(d; M_1, p_A) [M_2(d)]_\delta$$
where $\overline{\text{ASN}}_\delta^{(2)}$ denotes the weighted average (over $p_0$ and $p_A$) of the expected second-stage sample size under $p_0$ and under $p_A$.

Applying now the Lagrangian argument in Section 2.2.1, with $\alpha_\delta = g_0(\delta)$. $\beta_\delta = g_1(\delta)$. $\overline{\text{ASN}}_\delta^{(2)} = f(\delta)$. $\pi_0 C_0 = \lambda_0$ and $\pi_A C_A = \lambda_1$, we see that $\delta^*$ minimizes $\overline{\text{ASN}}_\delta^{(2)}$, among all plans $\delta$ with error of the first (respectively, second) kind no greater than the error of the first (respectively, second) kind of $\delta^*$.

Interpretation (iii) provides the Neyman-Pearson connection, with the Bayesian economic parameterization $(C_0, C_A)$ replaced by the Neyman-Pearson parameterization $(\alpha*, \beta*)$, and the Bayes plan $\delta^*$ re-interpreted as efficient in a Neyman-Pearson sense. It may be of interest to note here that there are other statistical applications of the sorts of Lagrangian facts expounded in Section 2.2.1. For example, Cook and Wang [10] have used an analytic (as opposed to algebraic) version of Section 2.2.1 in establishing the equivalence of constrained and compound optimal design construction.

We have worked out $\delta^*$ for a particular example, in large part in order to check the validity of our original objective of decreasing $\text{ASN}$ of non-flexible two-stage plans through introduction of a flexible second stage. Thus we have examined the operating characteristic ($OC$) and $\text{ASN}$ functions of $\delta^*$ alongside those of a certain comparable optimal non-flexible two-stage plan $\delta^0$, with same first-stage sample size.
In point of fact, to insure the same error rates for $\delta^0$ and $\delta^*$ in this discrete situation, $\delta^0$ is a certain mixture of three optimal non-flexible two-stage plans, as detailed in Section 2.4.

One further point should be made, if only with regard to computation: The Lagrangian formulation in (2.1) and (2.2) presupposes that $\alpha$ and $\beta$ are to be specified first, followed by finding appropriate $\lambda_0$ and $\lambda_1$ along with $\delta^*$, a state of affairs well portrayed by the notion

$$\lambda_0(\alpha, \beta), \lambda_1(\alpha, \beta), \delta^*(\alpha, \beta).$$

But, the reverse computation, indicated by the notion

$$\delta^*(\lambda_0, \lambda_1), \alpha(\lambda_0, \lambda_1), \beta(\lambda_0, \lambda_1),$$

actually is the natural one. Here one starts out by fixing $\lambda_0$ and $\lambda_1$, followed by solving for all the second-stage Bayes plans, with wrong-decision losses $\lambda_0$ and $\lambda_1$, corresponding to all possible first-stage outcomes $d$, followed in turn by computing the error rates of first and second kind, call them $\alpha(\lambda_0, \lambda_1)$ and $\beta(\lambda_0, \lambda_1)$, of the overall procedure, call it $\delta(\lambda_0, \lambda_1)$, made up of all of these second-stage Bayes plans. The plan $\delta(\lambda_0, \lambda_1)$ is then the optimal plan $\delta^*$ corresponding to the error rate restrictions

$$g_0(\delta) \leq \alpha(\lambda_0, \lambda_1),$$

$$g_1(\delta) \leq \beta(\lambda_0, \lambda_1).$$

To meet specified $(\alpha, \beta)$ objectives, one must iterate the reverse computation till a pair $(\lambda_0, \lambda_1)$ is found such that $\alpha(\lambda_0, \lambda_1)$ and $\beta(\lambda_0, \lambda_1)$ are satisfactorily close to $\alpha$ and $\beta$. 
We illustrate the idea by following example: \( M_1 = 20, (\pi_0, \pi_A) = (\frac{1}{2}, \frac{1}{2}), (p_0, p_A) = (0.15, 0.30) \) and \((C_0, C_A)\) is chosen such \((\alpha_{S^*}, \beta_{S^*}) \approx (0.05, 0.20)\). The relation between Bayes wrong-decision losses and Neyman-Pearson error rates is illustrated in Table 2.1. Note here that \( \alpha_S \) (respectively, \( \beta_S \)) is seen relatively more sensitive to changes of \( C_0 \) (respectively, \( C_A \)) than to changes of \( C_A \) (respectively, \( C_0 \)) and the matching of wrong-decision losses \((C_0, C_A)\) with error rates \((\alpha_S, \beta_S)\) is in fact many-to-one because of discreteness. The form of \( \delta^* \) is as given in Table 2.2.

Table 2.1: Relation between Bayes losses and Neyman-Pearson error rates

<table>
<thead>
<tr>
<th>( C_0 )</th>
<th>( C_A )</th>
<th>( \alpha_S )</th>
<th>( \beta_S )</th>
<th>( \text{ASN}_S )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100</td>
<td>460</td>
<td>0.01949</td>
<td>0.19821</td>
<td>39.39920</td>
</tr>
<tr>
<td>1100</td>
<td>480</td>
<td>0.05019</td>
<td>0.18895</td>
<td>40.29906</td>
</tr>
<tr>
<td>1100</td>
<td>500</td>
<td>0.05019</td>
<td>0.18895</td>
<td>40.29906</td>
</tr>
<tr>
<td>1120</td>
<td>460</td>
<td>0.04812</td>
<td>0.19610</td>
<td>40.02428</td>
</tr>
<tr>
<td>1120</td>
<td>480</td>
<td>0.04930</td>
<td>0.19199</td>
<td>40.18055</td>
</tr>
<tr>
<td>1120</td>
<td>500</td>
<td>0.05019</td>
<td>0.18895</td>
<td>40.29906</td>
</tr>
<tr>
<td>1140</td>
<td>460</td>
<td>0.04709</td>
<td>0.19985</td>
<td>39.88343</td>
</tr>
<tr>
<td>1140</td>
<td>480</td>
<td>0.04930</td>
<td>0.19199</td>
<td>40.18055</td>
</tr>
<tr>
<td>1140</td>
<td>500</td>
<td>0.04863</td>
<td>0.18870</td>
<td>40.77308</td>
</tr>
</tbody>
</table>

\( \text{ASN} \) and \( OC \) functions for \( \delta^* \) and \( \delta^0 \) are given in Figure 2.2 and Figure 2.3.

Actual achieved error rates are \((\alpha_{S^*}, \beta_{S^*}) = (0.04949, 0.19821)\) and a pair of wrong-decision losses matched with these error rates is \((1110, 156)\). The \( OC \) functions of \( \delta^* \) and \( \delta^0 \) essentially coincide (two \( OC \) functions disagree by at most 0.002), while the \( \text{ASN} \) function of \( \delta^* \) is seen to be uniformly superior to that of \( \delta^0 \), with second-stage improvement of 8.4 \% at \( \frac{(p_0 + p_A)}{2} = 0.225 \) and near zero for extreme \( p \) since the
### Table 2.2: The plan $\delta^*$

<table>
<thead>
<tr>
<th>$d$</th>
<th>$M_2(d)$</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$A_6^a$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$A_5$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$A_6$</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>$A_6$</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>[11.43]$^b$</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>[12.53]</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>[9.44]</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>$A_4^c$</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>$A_4$</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>$A_4$</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>$A_4$</td>
</tr>
</tbody>
</table>

$^a$ Accept $p_0$

$^b$ Critical region for second sample

$^c$ Accept $p_A$
Figure 2.1: ASN functions of $\delta^*$ and $\delta^0$
Figure 2.2: OC functions of $\delta^*$ and $\delta^0$
second-stage is seldom reached for such $p$.

2.3 Extension to two-sample case

2.3.1 One-sided case

A natural extension of our formulation is to the case of a control and a treatment population, with $H_A : \mu_{control} \equiv \mu_C = \mu_0$, $\mu_{treatment} \equiv \mu_T = \mu_A > \mu_0$ and $H_0 : \mu_C = \mu_T = \mu_0$. We imagine having $d_C$ (respectively, $d_T$) positives from the control (respectively, treatment) population, out of an initial sample of size $M_1$. Analogously to the development in the previous section, we consider decision rules $\delta$, whose role is to specify how $(d_C, d_T) \equiv d$ is to determine the second-stage of sampling. Specifically, $\delta$ is to specify the values of $d$ which are to call for no additional sampling ($M_2(d) = 0$), and those values of $d$ for which an additional sample of size $M_2(d) > 0$ is to be taken from each population. In the former case (namely, $M_2(d) = 0$), $\delta$ specifies in addition whether $H_0$ or $H_A$ is to be accepted. In the latter case (namely, $M_2(d) > 0$), $\delta$ specifies in addition, as a function of first-stage outcome $d$, the numbers $(D_C, D_T) \equiv D$ of second-stage "positives" that are to lead to acceptance of $H_A$ (i.e., the critical region corresponding to $d$). Thus, a decision rule $\delta$ specifies, for every possible first-stage outcome, a second-stage (possibly zero) sample size $M_2(d)$ and second-stage (possibly degenerate) critical region for the second-stage outcomes $D$.

However, the Bayesian solution responds to the fact that the above $H_0$ and $H_A$ do not differ with respect to the control population, by ignoring all control population evidence. One possible Bayesian way, which we adopt, to restore the
relevance of control population evidence is to replace the above simple hypotheses $H_0$ and $H_A$ by composite ones $H_0'$ and $H_A'$. With prior distribution over $H_0' \cup H_A'$ conditionally uniform on $H_0'$ and $H_A'$, and prior weights $\pi_0$ and $\pi_A$ on $H_0'$ and $H_A'$, respectively. As to the form of the composite hypotheses, $H_0'$ and $H_A'$, we find some precedent (see, for example, Cox [11] and Meeker [22]) for turning to the odds ratio \( \theta(p_C, p_T) = \frac{p_T(1-p_C)}{p_C(1-p_T)} \), and we take $H_A'$ as the composite hypothesis \( \theta(p_C, p_T) = \gamma > 1 \), and $H_0'$ as the composite hypothesis $\theta(p_C, p_T) = 1$.

Thus, in analogy to (2.6), we seek a rule $\delta^*$ minimizing

\[
R_\delta = \pi_0\overline{v}_0\left[ \sum_{d_C=0}^{M_1} \sum_{d_T=0}^{M_1} b(d_C; M_1, p_C) b(d_T; M_1, p_T) \cdot \left\{ 2[M_2(d_C, d_T)]_\delta + C_0[P_A[(p_C, p_T)(d_C, d_T)]_\delta \right\} \right] + \pi_A\overline{v}_A\left[ \sum_{d_C=0}^{M_1} \sum_{d_T=0}^{M_1} b(d_C; M_1, p_C) b(d_T; M_1, p_T) \cdot \left\{ 2[M_2(d_C, d_T)]_\delta + C_A[P_0[(p_C, p_T)(d_C, d_T)]_\delta \right\} \right]
\]

where, for example, $[P_A[(p_C, p_T)(d_C, d_T)]_\delta$ is the probability of accepting $H_A'$ under $\delta$, for population parameters $(p_C, p_T)$ and first-stage outcomes $(d_C, d_T)$; also, for example, $\overline{v}_A[ b(d_C; M_1, p_C) b(d_T; M_1, p_T) ]$, incorporating area length differential, equals

\[
\left( \int_0^1 \left[ \frac{\gamma^2}{1 + (\gamma - 1)p} \right] dp \right)^{-1} \cdot \left( \int_0^1 b(d_C; M_1, p) b(d_T; M_1, \frac{\gamma p}{1 + (\gamma - 1)p}) \left[ \frac{\gamma^2}{1 + (\gamma - 1)p} \right] \frac{\gamma^2}{1 + (\gamma - 1)p} \right) dp.
\]

\( (2.7) \)
and \( av_0 \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right] \), computed from (2.7) with \( \gamma = 1 \), equals
\[
\int_0^1 b(d_C^*; M_1, P) b(d_T^*; M_1, P) \, dp.
\]

This composite formulation amounts, in Neyman-Pearson terms, to dealing with average error rates over \( H_0' \) and \( H_A' \), respectively (Wald [31], p.80), and to choosing, from among all possible decision rules \( \delta \), a rule \( \delta^* \) minimizing expected second-stage sample size averaged in the sense of the previous paragraph over all pairs \( (P_C, P_T) \) in \( H_0' \cup H_A' \), subject to the restriction that the two uniform error rate averages, respectively over \( H_0' \) and \( H_A' \), not exceed predetermined levels.

The details look as they do in section 2.2.2, except for the following. The posterior probability \( \pi_0(d) \) becomes
\[
\frac{\pi_0 av_0 \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right]}{\pi_0 av_0 \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right] + \pi_A av_A \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right]}.\]

Also, the criterion (2.5) for second-stage critical region construction now has numerator
\[
C_0 \pi_0(d) = \left( \frac{av_0 \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right]}{av_0 \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right]} \right) \quad (2.8)
\]
and corresponding denominator
\[
C_A \pi_A(d) = \left( \frac{av_A \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right]}{av_A \left[ b(d_C^*; M_1, P_C) b(d_T^*; M_1, P_T) \right]} \right) \quad (2.9)
\]

We note that the composite-hypothesis formulation avoids the cancellation of control population likelihoods and thus restores the relevance of control population evidence.
As an example, set \( M_1 = 20 \), \( \gamma = 2 \), \((\pi_0, \pi_A) = (\frac{1}{2}, \frac{1}{2})\) with \((C_0, C_A)\) chosen such \((\alpha_{\xi^*}, \beta_{\xi^*}) \approx (0.05, 0.20)\).

The "continuation region" calling for taking a second sample of size \( M_2(d) > 0 \) is shown in Figure 2.3 and \( M_2(d) \) itself as a function of \( d \) is given in Table 2.3. The observed second-stage critical regions for \( M_2(d) > 0 \) are essentially of form \( DT - DC > c \).

Actual achieved \((\alpha_{\xi^*}, \beta_{\xi^*})\) equal \((0.01999, 0.19921)\) and a pair of wrong-decision losses matched with these error rates is \((4740, 2152)\). \( \bar{A}S \bar{N}_{\xi^*}^{(2)} \) equals 129.36 from each population. \( \bar{A}S \bar{N}_{\xi_0}^{(2)} \), with \( \xi_0 \) again constructed according to section 2.4, equals 137.75 from each population, with an improvement of 6.1% at the second stage.

### 2.3.2 Two-sided case

The two-sided case differs from the previous one only in that \( H_A \) now consists both of the point set \( \theta(p_1, p_2) = \gamma \) of section 2.3.1 and the additional point set \( \tilde{\theta}(p_1, p_2) = \frac{1}{\gamma} \).

We limit ourselves here to indicating the typical shapes of first-stage continuation region and second-stage critical regions.

The possible continuation region will be collections of grid points forming essentially elliptical bands, bounded by some two of the elliptical contours, as given in Figure 2.4 for \( M_1 = 20 \) and \( \gamma = 2 \). \( \alpha \) and \( \beta \) determine the relevant band (i.e., relevant pair of elliptical contours).

Figure 2.5 and 2.6 shows the elliptical character of the possible second-stage acceptance and complementary rejection regions, given \( M_2(d) = 40 \). Figure 2.6
Figure 2.3: "Continuation region" and hypotheses for $\delta^*$
Table 2.3: Second-stage sample size $M_2(d)$: $M_1 = 20$

<table>
<thead>
<tr>
<th>$d_c \setminus d_f$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>142</td>
<td>178</td>
<td>198</td>
<td>198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>138</td>
<td>159</td>
<td>176</td>
<td>180</td>
<td>178</td>
<td>160</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>140</td>
<td>162</td>
<td>180</td>
<td>195</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>138</td>
<td>159</td>
<td>176</td>
<td>180</td>
<td>178</td>
<td>160</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>109</td>
<td>135</td>
<td>154</td>
<td>166</td>
<td>175</td>
<td>161</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>108</td>
<td>132</td>
<td>149</td>
<td>160</td>
<td>166</td>
<td>160</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>108</td>
<td>125</td>
<td>145</td>
<td>155</td>
<td>162</td>
<td>162</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>101</td>
<td>124</td>
<td>143</td>
<td>151</td>
<td>157</td>
<td>158</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>123</td>
<td>136</td>
<td>147</td>
<td>155</td>
<td>156</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>123</td>
<td>136</td>
<td>147</td>
<td>155</td>
<td>156</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>101</td>
<td>123</td>
<td>135</td>
<td>146</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>108</td>
<td>124</td>
<td>136</td>
<td>147</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>108</td>
<td>125</td>
<td>137</td>
<td>148</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>123</td>
<td>136</td>
<td>147</td>
<td>155</td>
<td>156</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>101</td>
<td>123</td>
<td>135</td>
<td>146</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>108</td>
<td>124</td>
<td>136</td>
<td>147</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>148</td>
<td>164</td>
<td>167</td>
<td>166</td>
<td>165</td>
<td>153</td>
<td>142</td>
<td>129</td>
<td>113</td>
<td>91</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>161</td>
<td>168</td>
<td>180</td>
<td>171</td>
<td>167</td>
<td>165</td>
<td>148</td>
<td>129</td>
<td>112</td>
<td>92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>179</td>
<td>198</td>
<td>196</td>
<td>187</td>
<td>179</td>
<td>164</td>
<td>148</td>
<td>129</td>
<td>111</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>176</td>
<td>206</td>
<td>209</td>
<td>199</td>
<td>188</td>
<td>168</td>
<td>149</td>
<td>128</td>
<td>102</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>176</td>
<td>204</td>
<td>206</td>
<td>199</td>
<td>179</td>
<td>150</td>
<td>121</td>
<td>92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
illustrate the fact that, when \( d \) shows some imbalance in favor of the test population versus the control population (i.e., \( d_2 > d_1 \)), then values of \( D \) most strongly pointing to \( H_0 \) are those that undo this imbalance (i.e., \( D_2 < D_1 \)). Figure 2.5 may be related to the fact that, when both sample proportions are near 0 or 1, there is no strong evidence for \( H_0 \) even when the control and test positive counts are nearly the same: if so, the relatively weak evidence for \( H_0 \) provided by a pair of very small equal counts, such as (3, 3), presumably is best bolstered by second-stage evidence both essentially maintaining the equality and also bringing both total counts \((d_i + D_i), i = \text{1.} 2\) to intermediate range.

### 2.4 Fair comparisons with optimal standard two-stage plans

The Bayes/Neyman-Pearson correspondence underlying sections 2.2 and 2.3 essentially applies as well to optimal non-flexible two-stage plans, i.e., plans with positive second-stage sample size not depending on first-stage outcome. One aspect of identifying such plans is consideration of all partitions of the set of possible first-stage outcomes \( d \) into three sets.

The Neyman-Pearson framework now calls for finding that partition, together with the assigning of immediate acceptance of either hypothesis or continuation to the three elements of the partitions, minimizing \( \text{ASV} \) averaged over both hypotheses, subject to error rate restrictions of both kinds.

For the Bayes framework, for example for the formulation of section 2.2, involving unit sampling cost and wrong-decision losses \( C_0 \) and \( C_1 \), one considers the posterior
Figure 2.4: Typical continuation points "+" and "Continuation region" boundary contours for $|M_1| = 20$
Figure 2.5: Second-stage critical regions $A_A$ for $M_2(d) = i0$ for $d = (3.3)$
Figure 2.6: Second-stage critical regions $A_A$ for $M_2(d) = 40$ for $d = (10.15)$
probability of $H_0$  
\[ \pi_0(S) = \frac{\pi_0 \sum_{d \in S} b(d; M_1, p_0)}{\pi_0 \sum_{d \in S} b(d; M_1, p_0) + \pi_A \sum_{d \in S} b(d; M_1, p_A)}. \]

conditionally on the element $S$ of a partition. Each of the three elements $S$ of a partition thereby is assigned an optimal decision: to accept either hypothesis, or to continue sampling (in which case $S$ is denoted by $T$) with additional sample size $M_2(T)$ and critical region based on the criterion with numerator
\[ C_0 \pi_0(T) \left( \frac{\sum_{d \in T} b(d; M_1, p_0) \ln D_i M_2(T), p_0)}{\sum_{d \in T} b(d; M_1, p_0)} \right). \] (2.10)

The Bayes non-flexible two-stage plan minimizing expected risk is found by searching over all partitions. We note that, for either framework, likelihood ratio monotonicity allows limiting the search over partitions to just the usual ordered intervals.

The resulting Bayes procedure has well-defined error rates $(\alpha_{\delta}, \beta_{\delta})$, and is, again, optimal for $(\alpha_{\delta}, \beta_{\delta})$, in the sense of minimizing average ASN over all labeled partitions, with error rates restricted to be no greater than $(\alpha_{\delta}, \beta_{\delta})$. As $C_0$ and $C_A$ are varied, one thus generates a collection of efficient non-flexible two-stage plans $\delta$, together with their error rates $(\alpha_{\delta}, \beta_{\delta})$. Typically, because of discreteness, no member of the above collection of error rates $(\alpha_{\delta}, \beta_{\delta})$ will coincide with $(\alpha_{\delta}, \beta_{\delta})$.

We therefore have gone, in a manner entirely analogous to the usual randomization over two procedures when only $\alpha$ is targeted in the discrete case, to randomization over three plans $\delta$, with the three error rate vectors $(\alpha_{\delta}, \beta_{\delta})$ all near $(\alpha_{\delta}, \beta_{\delta})$ and the three randomization weights chosen so that the randomization expectation of $(\alpha_{\delta}, \beta_{\delta})$ equals $(\alpha_{\delta}, \beta_{\delta})$.

The three plans $\delta$, together with their randomization weights, are determined as follows: (i) With $(\alpha_{\delta}, \beta_{\delta})$ as origin, for each of the four quadrants determine the
point \((\alpha_{\delta^*}, \beta_{\delta^*})\) nearest (in the sense of Euclidean distance) to \((\alpha_{\delta^*}, \beta_{\delta^*})\). (ii) Except for degenerate situations (e.g., when \((\alpha_{\delta^*}, \beta_{\delta^*})\) lies in the convex hull of two of these four points), \((\alpha_{\delta^*}, \beta_{\delta^*})\) will in fact lie in the convex hull of exactly two of the four possible triples that can be formed from the four points in question, corresponding to two possible experimentwise randomizations over three optimal non-flexible two-stage plans, each of the two randomization expectations of \((\alpha_{\delta^*}, \beta_{\delta^*})\) equaling \((\alpha_{\delta^*}, \beta_{\delta^*})\).

(iii) Of these two possible experimentwise randomizations, the one with the smaller randomization expectation of \(A_{\delta^*} \sqrt{2(\delta^*)}\) is chosen as representing the comparable optimal non-flexible plan \(\delta^*\).

For the example in section 2.2.2, \(\delta^*\) yields \((\alpha_{\delta^*}, \beta_{\delta^*}) = (0.04949, 0.19821)\). The four optimal non-flexible plans \(\delta^*\), call them A, B, C and D, all turn out to have the same continuation region \(T\) as does \(\delta^*\), with second-stage sample size \(m_2\) and critical values \(k_2\) as given in Table 2.4. Table 2.4 also shows the two possible three-plan randomizations mentioned in (ii), and, in accordance with (iii), identifies \(\delta^0\) as the experimentwise randomization over ABC.

For the formulation of section 2.3.1, motivated by the above-mentioned coincidence of continuation regions we have taken the flexible continuation region as applying to the non-flexible case. Regarding the construction of critical regions, we observed and used monotonicity in \(D_2\) of the conditional likelihood ratio with numerator

\[
C_{0^*} \pi_0(T) := \left( \frac{\text{av}_0[ \sum_{(d_C, d_T) \in T} \frac{b(d_C, M_1, p_{d_C}) b(d_T, M_1, p_{d_T}) b(D_C, m_2, p_{D_C}) b(D_T, m_2, p_{D_T})}{b(d_C, M_1, p_{d_C}) b(d_T, M_1, p_{d_T})} ]}{\text{av}_0[ \sum_{(d_C, d_T) \in T} b(d_C, M_1, p_{d_C}) b(d_T, M_1, p_{d_T}) ]} \right)
\]

and corresponding denominator.
Optimal flexible two-stage plans

by

Seoung-gon Ko

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department:  Statistics  
Major:  Statistics

Iowa State University
Ames, Iowa
1994
# TABLE OF CONTENTS

## CHAPTER 1. INTRODUCTION ................................................. 1

## CHAPTER 2. BAYES AND NEYMAN-PEARSON ASPECTS .................. 5

2.1 Introduction ............................................................. 5

2.2 The Neyman-Pearson/Bayes connection ................................. 6

2.2.1 Lagrangian solutions of constrained optimization problems ... 6

2.2.2 The Bayes/Neyman-Pearson connection in one-sample case ... 7

2.3 Extension to two-sample case ........................................... 16

2.3.1 One-sided case ...................................................... 16

2.3.2 Two-sided case ..................................................... 19

2.4 Fair comparisons with optimal standard two-stage plans .......... 22

2.5 Remarks concluding chapter 2 ........................................ 28

## CHAPTER 3. GROUP-SEQUENTIAL ASPECTS ................................. 31

3.1 Introduction ............................................................. 31

3.2 Formulation ............................................................. 33

3.3 Optimization ........................................................... 35

3.3.1 Optimization for given \((\mu_0, \mu_1)\), \(\mu_0 < \mu_1\) ................. 35

3.3.2 Optimal plan for linear transformations of \((\mu_0, \mu_1)\), \(\mu_0 < \mu_1\) .... 44

3.4 Alternate plans .......................................................... 51
3.4.1 Single Sampling Plan (SS) ........................................... 55
3.4.2 Standard Two-stage Plans (STP) ................................. 55
3.4.3 Sequential Probability Ratio Test (SPRT) ...................... 56
3.5 Numerical illustrations and comparisons ....................... 57
3.6 Applications to binomial (Bernoulli) case ..................... 58
  3.6.1 One-sample case .................................................. 58
  3.6.2 Illustrations: one-sample case ................................. 71
  3.6.3 Two-sample case .................................................. 77
  3.6.4 Illustrations: two-sample case ................................. 83
3.7 Remarks concluding chapter 3 .................................... 97

CHAPTER 4. CONCLUSION ................................................. 100

BIBLIOGRAPHY .......................................................... 102

ACKNOWLEDGMENTS ..................................................... 105
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Relation between Bayes losses and Neyman-Pearson error rates</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>The plan $\delta^*$</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Second-stage sample size $M_2(d)$: $M_1 = 20$</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>$\delta^g$ and the four &quot;nearest&quot; plans $\hat{\delta}$ for the example in section 2.2.2</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>$\delta^g$ and the four &quot;nearest&quot; plans $\hat{\delta}$ for the example in section 2.3.1</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>Approximate binomial plan corresponding Wiener plan</td>
<td>65</td>
</tr>
<tr>
<td>3.2</td>
<td>The approximate optimal flexible binomial plan for $\Delta = 0.1$</td>
<td>72</td>
</tr>
<tr>
<td>3.3</td>
<td>The approximate optimal flexible binomial plan for $\Delta = 0.15$</td>
<td>74</td>
</tr>
<tr>
<td>3.4</td>
<td>The approximate optimal flexible binomial plan for $\Delta = 0.2$</td>
<td>76</td>
</tr>
<tr>
<td>3.5</td>
<td>Approximate optimal first-stage sample size $n_0^*$ for two-sample case</td>
<td>84</td>
</tr>
<tr>
<td>3.6</td>
<td>Approximate optimal flexible binomial two-stage plans for two-sample case: $(p_0, p_1) = (0.2, 0.3)$</td>
<td>85</td>
</tr>
<tr>
<td>3.7</td>
<td>Approximate optimal flexible binomial two-stage plans for two-sample case: $(p_0, p_1) = (0.2, 0.35)$</td>
<td>88</td>
</tr>
</tbody>
</table>
Table 3.8: Approximate optimal flexible binomial two-stage plans for two-sample case: \((p_0, p_1) = (0.2, 0.4)\) \hspace{1cm} 91

Table 3.9: Approximate optimal flexible binomial two-stage plans for two-sample case: \((p_0, p_1) = (0.5, 0.55)\) \hspace{1cm} 94
LIST OF FIGURES

Figure 2.1: ASN functions of $\delta^*$ and $\delta^o$ ........................................ 14
Figure 2.2: OC functions of $\delta^*$ and $\delta^o$ ........................................ 15
Figure 2.3: “Continuation region” and hypotheses for $\delta^*$ .................. 20
Figure 2.4: Typical continuation points "**" and “Continuation region”
boundary contours for $M_1 = 20$ .................................................. 23
Figure 2.5: Second-stage critical regions $A_A$ for $M_2(d) = 40$ for $d = (3.3)$ 24
Figure 2.6: Second-stage critical regions $A_A$ for $M_2(d) = 40$ for $d = (10.15)$ 25
Figure 3.1: Contours of $L_1(T_1, k_1)$ when $(T_0, \hat{\rho}_0, \hat{\rho}_1) = (18.9, 193.31, 109.89)$ 39
Figure 3.2: Function $L_1(T_1, k_1^*)$ when $(T_0, \hat{\rho}_0, \hat{\rho}_1) = (18.9, 193.31, 109.89)$ 40
Figure 3.3: Candidates of solutions for $T_i$ for each signed equation : $r = 6$ 45
Figure 3.4: Optimal second-stage sampling efforts .......................... 59
Figure 3.5: Optimal second-stage critical values .............................. 60
Figure 3.6: $OC_\mu$ of $STP^*$ and $OFP^*$ for the Wiener process $w(0, 1)$ and $w(0.5, 1)$ with $(\alpha, \beta) = (0.05, 0.10)$ ................. 61
Figure 3.7: $AST_{\mu}$ of $STP^*$ and $OFP^*$ for the Wiener process $w(0, 1)$ and $w(0.5, 1)$ with $(\alpha, \beta) = (0.05, 0.10)$ ................. 62
Figure 3.8: $OC_\mu$ of $SPRT, OFP^*, STP^*$ and $SS$, for the Wiener process $w(0.1)$ and $w(0.5.1)$ with $(\alpha, \beta) = (0.05, 0.10)$ ................. 63
Figure 3.9: $AST_{\mu}$ of $SPRT.OFP^*, STP^*$ and $SS^*$ for the Wiener process $w(0.1)$ and $w(0.5, 1)$ with $(\alpha, \beta) = (0.05, 0.10)$.
CHAPTER 1. INTRODUCTION

In this dissertation, we construct optimal "flexible" two-stage plans. "Flexible" means that the first-stage outcomes not only determine whether to accept the null or alternative hypothesis, or whether to proceed to the second stage, but also, in the latter case, determine second-stage sample sizes and critical values that depend on first-stage outcome. "Optimal" is meant in the sense that a suitable measure of sampling effort is minimized with respect to all flexible plans, subject to suitable constraints.

Determining or adjusting second-stage sample size and critical region based on first-stage outcome constitutes an extension of classical double sampling plans as originally proposed (Dodge and Romig [13]) and subsequently developed (Hald [16]) and Spurrier and Hewett [26]). The idea in a sense dates back to Stein's [27] seminal paper and its descendants, through the adaptive utilization of first-stage information, as detailed below, focuses on the parameter ($\mu$) of interest, rather than on a nuisance parameter ($\sigma$) as in Stein's. Such focus on the parameter of interest is particularly reminiscent of the approach to sequential estimation indicated by Birnbaum and Healy [3]; and also of Miller and Freund’s [23] more recent proposal of an ad hoc method for determining the second-stage sample size in binomial estimation problems.

The literature on optimized two-stage plans deals only with "non-flexible" plans:
 plans, in other words, that call for first-stage decision or for a second stage with fixed sample size and critical region. Colton and McPherson [9] focus on first-stage rejection. They minimize alternative-hypothesis expected sample size with respect to first- and second-stage sample sizes and critical values, under fixed error rates of both kinds. Thall et al. [28], on the other hand, focus instead on first-stage acceptance and minimize the average of the expected sample sizes at the null and alternative hypotheses. Thall et al. [29], [30] treat multiple testing in similar fashion.

There have also been Bayesian approaches to two-stage plans: for example, Berry [2] reformulates McPherson’s [21] approach in Bayesian terms, by also assuming a maximum total sample size and equal stage sample sizes, but allowing for the possibility of early termination in accordance with a certain Bayesian stopping rule. Cohen and Sackrowitz [7], [8] derive optimal Bayes procedures for determining the second-stage sample size in exponential family estimation, with special attention devoted to the binomial case.

In the group-sequential (Pocock [25], O’Brien and Fleming [24]) framework, McPherson [21] minimizes various measures of sampling effort, with respect to the number of “equally spaced” interim analyses, under a repeated-significance-test convention for assigning interim critical values, and fixed overall error rates of both kinds. Jennison [17] and Eales and Jennison [15], on the other hand, fix the number of interim analyses but optimize over the corresponding interim critical values. In the Bayesian framework, Cressie and Morgan [12] studied certain k-stage probability ratio tests called “variable-sample-size probability ratio tests (VPRTs)”. This latter work is related to the “sequential design of experiments” initiated by Chernoff [6] (see also Bradt and Karlin [5], and Whittle [32] and Borwanker. David and Ingwell
Motivated by the above considerations and developments, we have considered certain two-stage plans through two distinct optimization formulations. While our "flexible" formulations in both cases would appear to call for a multi-dimensional optimization, these turn out to be "separable" into a family of simple two-dimensional optimizations, each corresponding to a first-stage outcome in the continuation region.

The first formulation, in Chapter 2, addressing the case of one treatment both with and without a control, is distinguished by possessing a Neyman-Pearson as well as Bayes interpretation, reminiscent of the dual interpretation of the SPRT as optimal in both a Bayes and Neyman-Pearson sense (Lehmann [20], p.104). The Neyman-Pearson interpretation is that average expected sample size is being minimized, subject just to the two overall error rates $\alpha$ and $\beta$, respectively of first and second kind. The Bayes interpretation is that Bayes risk, involving both sampling cost and wrong decision losses, is being minimized. With $C_0$ the cost of wrongly abandoning null hypothesis, and $C_A$ the cost of wrongly staying with it, one useful outgrowth of our dual interpretation is that $\{\alpha, \beta\}$ and $\{C_0, C_A\}$ determine each other.

The second formulation, reminiscent of the group-sequential point of view, involves given first- and second-stage error rates of first and second kind (four such given error rates in all), with objective function equal to average (over the hypotheses) expected sample size. This is the formulation in chapter 3, where the single treatment case, with and without control, is considered. We choose to analyze this group-sequential formulation by means of the arcsine transformation [14] to normality of the binomial distribution. This allows translating the binomial problem to one
of testing the drift parameter of a standard Wiener process [18]. This brings to the analysis the simplifications and economies of location and scale equivariance.

The flexible plans provide an alternative to informal sample size adjustment following an interim analysis, as is sometimes done, for example, in the pharmaceutical industry, in the case of drugs already on the market being explored for new indications; equally, they in effect incorporate an initial (pilot) study into an over-all experimental plan, both when the pilot study is already done and when a pilot stage is to be incorporated into a planned-for two-stage study.

Comparisons are made with optimal non-flexible plans, which leads to conclusion that flexible plans reduce expected sample size in both formulations.
CHAPTER 2. BAYES AND NEYMAN-PEARSON ASPECTS

2.1 Introduction

Our initial motivation related to this chapter was to develop procedures in a Neyman-Pearson framework, which will (a) control errors of both kinds and (b) decrease average sampling number (ASN) by allowing second-stage sample size to depend on first-stage outcome.

As work progressed on our original Neyman-Pearson optimization formulation, we observed that the plans we were developing from an essentially Neyman-Pearson perspective had also a Bayesian interpretation, much in the way that the sequential probability ratio test simultaneously is optimal in both a Neyman-Pearson and a Bayesian sense (Lehmann [20], p.191). In particular, we observed that Neyman Pearson ASN optimization under fixed global error rates in fact is implemented by certain second-stage Bayes solutions. In other words, the plans are optimal simultaneously in both a Bayes and a Neyman-Pearson sense: the former because they are Bayes in a class $\Gamma$ of flexible plans; the latter because they minimize a suitable measure of sampling effort with respect to all members of $\Gamma$ satisfying certain error rate restrictions of first and second kind.

Section 2.2 treats the analysis yielding optimal flexible plans in the case of a single population where this optimization is seen to have simultaneously a Bayesian
and Neyman-Pearson character, due to the availability of the Lagrangian approach to solving constrained optimization problems. Bayesian wrong-decision losses are thereby put into correspondence with Neyman-Pearson wrong-decision error rates. Section 2.2.2 compares a particular optimal flexible with the corresponding optimal non-flexible two-stage plan. Section 2.4 describes a certain experimentwise randomization useful for this comparison.

Section 2.3 discusses extension of the ideas to the case of two populations, and Section 2.5 contains concluding remarks.

2.2 The Neyman-Pearson/Bayes connection

2.2.1 Lagrangian solutions of constrained optimization problems

Consider an arbitrary domain $\Delta$ with elements $\delta$, and three functions $f(\delta)$, $g_0(\delta)$ and $g_1(\delta)$ defined on $\Delta$. Suppose that we wish to find an optimum element of $\Delta$, in the sense of minimizing $f(\delta)$, subject to the restrictions $g_0(\delta) \leq A$ and $g_1(\delta) \leq B$.

If there is a $\delta^*$ in $\Delta$, and also two non-negative numbers $\lambda_0$ and $\lambda_1$, such that

\[
g_0(\delta^*) = A, \quad g_1(\delta^*) = B
\]  

(2.1)

and

\[
L(\delta^*) \leq L(\delta), \quad \delta \in \Delta
\]  

(2.2)

where

\[
L(\delta) \equiv f(\delta) + \lambda_0 g_0(\delta) + \lambda_1 g_1(\delta).
\]
then $\delta^*$ is optimal in the above sense.

This assertion is verified as follows: Suppose $\delta^*$ satisfied (2.1) and (2.2), and were not optimal; in other words, suppose there were a $\delta'$ in $\Delta$ with

$$f(\delta') < f(\delta^*). \tag{2.3}$$

and also

$$g_0(\delta') \leq A. \tag{2.4}$$
$$g_1(\delta') \leq B.$$

Then we would have

$$L(\delta') = f(\delta') + \lambda_0(g_0(\delta') - A) + \lambda_1(g_1(\delta') - B) + \lambda_0 A + \lambda_1 B$$
$$\leq f(\delta') + \lambda_0 A + \lambda_1 B$$
$$< f(\delta^*) + \lambda_0 A + \lambda_1 B$$
$$= f(\delta^*) + \lambda_0 g_0(\delta^*) + \lambda_1 g_1(\delta^*)$$
$$= L(\delta^*).$$

where the middle three signs would be due respectively to (2.4), (2.3) and (2.1), and where the implied inequality

$$L(\delta') < L(\delta^*).$$

would contradict (2.2).

2.2.2 The Bayes/Neyman-Pearson connection in one-sample case

Consider an experimenter in possession of $d$ positives out of an initial sample of size $M_1$ from a Bernoulli population, for which the experimenter is willing to assume
that $p$ equals either $p_0$ or $p_A > p_0$. Suppose as well that the experimenter is willing to restrict his/her further options either to sampling no more ($M_2(d) = 0$), with an immediate decision whether $p = p_0$ or $p_A$, or to conducting one further single sampling plan ($M_2(d) > 0$); and that the experimenter intends to choose from among these options in Bayesian terms, involving, among other things, two cost ratios: the ratio $C_0$ (respectively, $C_A$) of the cost of wrongly deciding that $p_A$ (respectively, $p_0$) holds, to the unit cost of sampling.

With prior $\pi_0$ ($0 < \pi_0 < 1$) on $p_0$, the posterior probability $\pi_0(d)$ of $p_0$ is

$$
\pi_0 b(d; M_1, p_0)
\pi_0 b(d; M_1, p_0) + \pi_A b(d; M_1, p_A)
$$

where $b(\cdot)$ denotes binomial probability, and correspondingly for $\pi_A = 1 - \pi_0$, $p_A$ and $\pi_A(d) = 1 - \pi_0(d)$.

The experimenter's single sampling Bayes plan $S^*(d)$ at $d$ is the (possibly degenerate) single sampling plan $S(d)$ minimizing

$$
R(d) = M_2(d) + C_0 \pi_0(d) P_{A|0}(d) + C_A \pi_A(d) P_{0|A}(d).
$$

where $P_{A|0}(d)$ (respectively, $P_{0|A}(d)$) denotes the probability, under $S(d)$, of accepting $p_A$ (respectively, $p_0$) when $p = p_0$ (respectively, $p_A$). Here, $P_{A|0}(d)$ and $P_{0|A}(d)$ are based, respectively, on the binomial probabilities $b(D; M_2(d), p_0)$ and $b(D; M_2(d), p_A)$ of the second-stage number of "positives". $D, D = 0, 1, 2, \ldots, M_2(d)$.

$S^*(d)$ is derived by first fixing $M_2(d)$ at, say, $m_2$ and minimizing $R(d)$ over critical regions, followed by minimizing over $m_2$, $m_2 = 0, 1, 2, \ldots$. For $m_2 = 0$, minimizing $R(d)$ amounts to accepting $p_0$ (respectively, $p_A$) if $\frac{C_0 \pi_0(d)}{C_A \pi_A(d)}$ is greater (respectively, less) than 1. For $m_2 > 0$, in view of the extended Neyman-Pearson
Lemma. the Bayes plan amounts to accepting \( p_0 \) (respectively, \( p_A \)) if the criterion

\[
\frac{C_0 \pi_0(d) b(D; m_2, p_0)}{C_A \pi_A(d) b(D; m_2, p_A)}
\]

is greater (respectively, less) than 1, yielding a straightforward one-sided plan, by likelihood ratio monotonicity.

\( \delta^* \equiv (S^*(0), S^*(1), \ldots, S^*(M_1)) \) has the following interpretations.

(i) As indicated in the preceding discussion, \( S^*(d) \) is the single sampling procedure that a “one-stage” Bayesian will follow who, having obtained \( d \) initial positives, uses prior \( (\pi_0(d), \pi_A(d)) \) on \((p_0, p_A)\).

(ii) Consider a “two-stage” Bayesian about to take, rather than already in possession of, an initial sample of size \( M_1 \), possibly to be followed by a second-stage single sampling plan, who uses prior \((\pi_0, \pi_A)\) on \((p_0, p_A)\). Suppose that this Bayesian has, as his options, vectors \( \delta \equiv (S(0), S(1), \ldots, S(d), \ldots, S(M_1)) \), where each \( S(d) \) is an arbitrary second-stage single sampling plan. \( \delta^* \) minimizes the Bayes risk \( R_\delta \), for such a Bayesian, where

\[
R_\delta = \pi_0 \sum_{d=0}^{M_1} b(d; M_1, p_0) \{ [M_2(d)]_\delta + C_0 [P_A|0](d) \}_{\delta} \]

\[+ \pi_A \sum_{d=0}^{M_1} b(d; M_1, p_A) \{ [M_2(d)]_\delta + C_A [P_0|A](d) \}_{\delta} \].

\( (2.5) \)

Here, the dependence of \( R_\delta \) on \( \delta \) comes from the fact that \( M_2(d), P_A|0(d) \) and \( P_0|A(d) \) are determined by \( S(d) \), and hence by \( \delta \).

(iii) A third interpretation of \( \delta^* \) is seen by writing \( (2.6) \) as

\[
R_\delta = \pi_0 \sum_{d=0}^{M_1} b(d; M_1, p_0) [M_2(d)]_\delta + \pi_A \sum_{d=0}^{M_1} b(d; M_1, p_A) [M_2(d)]_\delta.
\]
\[ + \pi_0 C_0 \sum_{d=0}^{M_1} b(d; M_1, \rho_0)[P_A|0(d)]_\delta \\
+ \pi_A C_A \sum_{d=0}^{M_1} b(d; M_1, \rho_A)[P_0|A(d)]_\delta \\
\equiv \overline{\text{ASN}}_\delta^{(2)} + \pi_0 C_0 \alpha_\delta + \pi_A C_A \beta_\delta \]

where \( \overline{\text{ASN}}_\delta^{(2)} \) denotes the weighted average (over \( \rho_0 \) and \( \rho_A \)) of the expected second-stage sample size under \( \rho_0 \) and under \( \rho_A \).

Applying now the Lagrangian argument in Section 2.2.1. with \( \alpha_\delta = g_0(\delta), \beta_\delta = g_1(\delta), \overline{\text{ASN}}_\delta^{(2)} = f(\delta), \pi_0 C_0 = \lambda_0 \) and \( \pi_A C_A = \lambda_1 \), we see that \( \delta^* \) minimizes \( \overline{\text{ASN}}_\delta^{(2)} \) among all plans \( \delta \) with error of the first (respectively, second) kind no greater than the error of the first (respectively, second) kind of \( \delta^* \).

Interpretation (iii) provides the Neyman-Pearson connection, with the Bayesian economic parameterization \( (C_0, C_A) \) replaced by the Neyman-Pearson parameterization \( (\alpha_{\delta^*}, \beta_{\delta^*}) \), and the Bayes plan \( \delta^* \) re-interpreted as efficient in a Neyman-Pearson sense. It may be of interest to note here that there are other statistical applications of the sorts of Lagrangian facts expounded in Section 2.2.1. For example, Cook and Wang [10] have used an analytic (as opposed to algebraic) version of Section 2.2.1 in establishing the equivalence of constrained and compound optimal design construction.

We have worked out \( \delta^* \) for a particular example, in large part in order to check the validity of our original objective of decreasing \( \text{ASN} \) of non-flexible two-stage plans through introduction of a flexible second stage. Thus we have examined the operating characteristic \( (OC) \) and \( \text{ASN} \) functions of \( \delta^* \) alongside those of a certain comparable optimal non-flexible two-stage plan \( \delta^0 \), with same first-stage sample size.
In point of fact, to insure the same error rates for $\delta^0$ and $\delta^*$ in this discrete situation, $\delta^0$ is a certain mixture of three optimal non-flexible two-stage plans, as detailed in Section 2.4.

One further point should be made, if only with regard to computation: The Lagrangian formulation in (2.1) and (2.2) presupposes that $\alpha$ and $\beta$ are to be specified first, followed by finding appropriate $\lambda_0$ and $\lambda_1$ along with $\delta^*$, a state of affairs well portrayed by the notion

$$\lambda_0(\alpha, \beta), \lambda_1(\alpha, \beta), \delta^*(\alpha, \beta).$$

But, the reverse computation, indicated by the notion

$$\delta^*(\lambda_0, \lambda_1), \alpha(\lambda_0, \lambda_1), \beta(\lambda_0, \lambda_1),$$

actually is the natural one. Here one starts out by fixing $\lambda_0$ and $\lambda_1$, followed by solving for all the second-stage Bayes plans, with wrong-decision losses $\lambda_0$ and $\lambda_1$, corresponding to all possible first-stage outcomes $d$, followed in turn by computing the error rates of first and second kind, call them $\alpha(\lambda_0, \lambda_1)$ and $\beta(\lambda_0, \lambda_1)$, of the overall procedure, call it $\delta(\lambda_0, \lambda_1)$, made up of all of these second-stage Bayes plans. The plan $\delta(\lambda_0, \lambda_1)$ is then the optimal plan $\delta^*$ corresponding to the error rate restrictions

$$g_0(\delta) \leq \alpha(\lambda_0, \lambda_1),$$

$$g_1(\delta) \leq \beta(\lambda_0, \lambda_1).$$

To meet specified $(\alpha, \beta)$ objectives, one must iterate the reverse computation till a pair $(\lambda_0, \lambda_1)$ is found such that $\alpha(\lambda_0, \lambda_1)$ and $\beta(\lambda_0, \lambda_1)$ are satisfactorily close to $\alpha$ and $\beta$. 
We illustrate the idea by following example: \( M_1 = 20, (\pi_0, \pi_A) = \left( \frac{1}{2}, \frac{1}{2} \right), (p_0, p_A) = (0.15, 0.30) \) and \((C_0, C_A)\) is chosen such \((\alpha_{\delta^*}, \delta_{\delta^*}) \approx (0.05, 0.20)\). The relation between Bayes wrong-decision losses and Neyman-Pearson error rates is illustrated in Table 2.1. Note here that \(\alpha_{\delta}\) (respectively, \(\delta_{\delta}\)) is seen relatively more sensitive to changes of \(C_0\) (respectively, \(C_A\)) than to changes of \(C_A\) (respectively, \(C_0\)) and the matching of wrong-decision losses \((C_0, C_A)\) with error rates \((\alpha_{\delta}, \delta_{\delta})\) is in fact many-to-one because of discreteness. The form of \(\delta^*\) is as given in Table 2.2.

<table>
<thead>
<tr>
<th>(C_0)</th>
<th>(C_A)</th>
<th>(\alpha_{\delta})</th>
<th>(\delta_{\delta})</th>
<th>(ASN_{\delta})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100</td>
<td>460</td>
<td>0.04949</td>
<td>0.19821</td>
<td>39.39920</td>
</tr>
<tr>
<td>1100</td>
<td>480</td>
<td>0.05019</td>
<td>0.18895</td>
<td>40.29906</td>
</tr>
<tr>
<td>1100</td>
<td>500</td>
<td>0.05019</td>
<td>0.18895</td>
<td>40.29906</td>
</tr>
<tr>
<td>1120</td>
<td>460</td>
<td>0.04812</td>
<td>0.19610</td>
<td>40.02428</td>
</tr>
<tr>
<td>1120</td>
<td>480</td>
<td>0.04930</td>
<td>0.19199</td>
<td>40.18055</td>
</tr>
<tr>
<td>1120</td>
<td>500</td>
<td>0.05019</td>
<td>0.18895</td>
<td>40.29906</td>
</tr>
<tr>
<td>1140</td>
<td>460</td>
<td>0.04709</td>
<td>0.19985</td>
<td>39.88343</td>
</tr>
<tr>
<td>1140</td>
<td>480</td>
<td>0.04930</td>
<td>0.19199</td>
<td>40.18055</td>
</tr>
<tr>
<td>1140</td>
<td>500</td>
<td>0.04863</td>
<td>0.18870</td>
<td>40.77308</td>
</tr>
</tbody>
</table>

\(ASN\) and \(OC\) functions for \(\delta^*\) and \(\delta^0\) are given in Figure 2.2 and Figure 2.3.

Actual achieved error rates are \((\alpha_{\delta^*}, \delta_{\delta^*}) = (0.01949, 0.19821)\) and a pair of wrong-decision losses matched with these error rates is \((1110, 456)\). The \(OC\) functions of \(\delta^*\) and \(\delta^0\) essentially coincide (two \(OC\) functions disagree by at most 0.002), while the \(ASN\) function of \(\delta^*\) is seen to be uniformly superior to that of \(\delta^0\), with second-stage improvement of 8.4 % at \(\frac{(p_0 + p_A)}{2} = 0.225\) and near zero for extreme \(p\) since the
Table 2.2: The plan $\delta^*$

<table>
<thead>
<tr>
<th>$d$</th>
<th>$M_2(d)$</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$A_\theta^a$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$A_\theta$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$A_\theta$</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>$A_\theta$</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>$[11.43]^b$</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>$[12.53]$</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>$[9.44]$</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>$A_A^c$</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>$A_A$</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>$A_A$</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>$A_A$</td>
</tr>
</tbody>
</table>

*a Accept $p_0$

*b Critical region for second sample

*c Accept $p_A$
Figure 2.1: ASN functions of $\delta^*$ and $\delta^0$
Figure 2.2: OC functions of $\delta^*$ and $\delta^\circ$
2.3 Extension to two-sample case

2.3.1 One-sided case

A natural extension of our formulation is to the case of a control and a treatment population, with \( H_A : p_{control} = p_C = p_0 \), \( p_{treatment} = p_T = p_A > p_0 \) and \( H_0 : p_C = p_T = p_0 \). We imagine having \( d_C \) (respectively, \( d_T \)) positives from the control (respectively, treatment) population, out of an initial sample of size \( M_1 \). Analogously to the development in the previous section, we consider decision rules \( \delta \), whose role is to specify how \( (d_C, d_T) \equiv d \) is to determine the second-stage of sampling. Specifically, \( \delta \) is to specify the values of \( d \) which are to call for no additional sampling \( (M_2(d) = 0) \), and those values of \( d \) for which an additional sample of size \( M_2(d) > 0 \) is to be taken from each population. In the former case (namely, \( M_2(d) = 0 \)), \( \delta \) specifies in addition whether \( H_0 \) or \( H_A \) is to be accepted. In the latter case (namely, \( M_2(d) > 0 \)), \( \delta \) specifies in addition, as a function of first-stage outcome \( d \), the numbers \( (D_C, D_T) \equiv D \) of second-stage "positives" that are to lead to acceptance of \( H_A \) (i.e., the critical region corresponding to \( d \)). Thus, a decision rule \( \delta \) specifies, for every possible first-stage outcome, a second-stage (possibly zero) sample size \( M_2(d) \) and second-stage (possibly degenerate) critical region for the second-stage outcomes \( D \).

However, the Bayesian solution responds to the fact that the above \( H_0 \) and \( H_A \) do not differ with respect to the control population, by ignoring all control population evidence. One possible Bayesian way, which we adopt, to restore the
relevance of control population evidence is to replace the above simple hypotheses $H_0$ and $H_A$ by composite ones $H_0'$ and $H_A'$, with prior distribution over $H_0' \cup H_A'$ conditionally uniform on $H_0'$ and $H_A'$, and prior weights $\pi_0$ and $\pi_A$ on $H_0'$ and $H_A'$, respectively. As to the form of the composite hypotheses, $H_0'$ and $H_A'$, we find some precedent (see, for example, Cox [11] and Meeker [22]) for turning to the odds ratio $\theta(\hat{p}_C, \hat{p}_T) = \frac{PT(1-\hat{p}_C)}{\hat{p}_C(1-\hat{p}_T)}$, and we take $H_A'$ as the composite hypothesis $\theta(\hat{p}_C, \hat{p}_T) = \gamma > 1$, and $H_0'$ as the composite hypothesis $\theta(\hat{p}_C, \hat{p}_T) = 1$.

Thus, in analogy to (2.6), we seek a rule $\delta^*$ minimizing

$$R_\delta = \pi_0 av_0 \left[ \sum_{d_C=0}^{M_1} \sum_{d_T=0}^{M_1} b(d_C; M_1, \hat{p}_C) b(d_T; M_1, \hat{p}_T) \cdot \right. \left. \{2[M_2(d_C, d_T)]_{\delta} + C_0[PA_1(\hat{p}_C, \hat{p}_T)(d_C, d_T)]_{\delta} \} \right]
+ \pi_A av_A \left[ \sum_{d_C=0}^{M_1} \sum_{d_T=0}^{M_1} b(d_C; M_1, \hat{p}_C) b(d_T; M_1, \hat{p}_T) \cdot \right. \left. \{2[M_2(d_C, d_T)]_{\delta} + C_A[P_0(\hat{p}_C, \hat{p}_T)(d_C, d_T)]_{\delta} \} \right]$$

where, for example, $[PA_1(\hat{p}_C, \hat{p}_T)(d_C, d_T)]_{\delta}$ is the probability of accepting $H_A'$ under $\delta$, for population parameters $(\hat{p}_C, \hat{p}_T)$ and first-stage outcomes $(d_C, d_T)$; also, for example, $av_1 \left[ b(d_C; M_1, \hat{p}_C) b(d_T; M_1, \hat{p}_T) \cdot \right]$, incorporating area length differential, equals

$$\left( \int_0^1 \frac{\gamma^2}{1 + \frac{\gamma}{1 + (\gamma - 1)p}} \, dp \right)^{-1} \cdot \left( \int_0^1 \frac{\gamma p}{1 + (\gamma - 1)p} \, dp \right),$$

$$\left( \int_0^1 \frac{\gamma p}{1 + (\gamma - 1)p} \, dp \right).$$
and \( \text{av}_0 \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) \right] \), computed from (2.7) with \( \gamma = 1 \), equals

\[
\int_0^1 b(d_C^*: M_1, p) b(d_T^*: M_1, p) \, dp.
\]

This composite formulation amounts, in Neyman-Pearson terms, to dealing with average error rates over \( H_0^\prime \) and \( H_A^\prime \), respectively (Wald [31], p.80), and to choosing, from among all possible decision rules \( \delta \), a rule \( \delta^* \) minimizing expected second-stage sample size averaged in the sense of the previous paragraph over all pairs \((p_C, p_T)\) in \( H_0^\prime \cup H_A^\prime \), subject to the restriction that the two uniform error rate averages, respectively over \( H_0^\prime \) and \( H_A^\prime \), not exceed predetermined levels.

The details look as they do in section 2.2.2, except for the following. The posterior probability \( \pi_0(d) \) becomes

\[
\frac{\text{av}_0 \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) \right]}{\pi_0 \text{av}_0 \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) \right] + \pi_A \text{av}_A \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) \right]}
\]

Also, the criterion (2.5) for second-stage critical region construction now has numerator

\[
C_0 \pi_0(d) \cdot \left( \frac{\text{av}_0 \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) b(D_C^*: M_2(u), p_C^*) b(D_T^*: M_2(u), p_T^*) \right]}{\text{av}_0 \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) \right]} \right) \tag{2.8}
\]

and corresponding denominator

\[
C_A \pi_A(d) \cdot \left( \frac{\text{av}_A \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) b(D_C^*: M_2(d), p_C^*) b(D_T^*: M_2(d), p_T^*) \right]}{\text{av}_A \left[ b(d_C^*: M_1, p_C^*) b(d_T^*: M_1, p_T^*) \right]} \right) \tag{2.9}
\]

We note that the composite-hypothesis formulation avoids the cancellation of control population likelihoods and thus restores the relevance of control population evidence.
As an example, set $M_1 = 20$, $\gamma = 2$, $(\pi_0, \pi_A) = (\frac{1}{2}, \frac{1}{2})$ with $(C_0, C_A)$ chosen such that $(\alpha_{\delta^*}, \beta_{\delta^*}) \approx (0.05, 0.20)$.

The "continuation region" calling for taking a second sample of size $M_2(d) > 0$ is shown in Figure 2.3 and $M_2(d)$ itself as a function of $d$ is given in Table 2.3. The observed second-stage critical regions for $M_2(d) > 0$ are essentially of form $DT - DC > c$.

Actual achieved $(\alpha_{\delta^*}, \beta_{\delta^*})$ equal $(0.04999, 0.19924)$ and a pair of wrong-decision losses matched with these error rates is $(47.40, 21.52)$. $\text{ASV}_{\delta^*}(2)$ equals 129.36 from each population. $\text{ASV}_{\delta^0}(2)$, with $\delta^0$ again constructed according to section 2.4, equals 137.75 from each population, with an improvement of 6.1% at the second stage.

### 2.3.2 Two-sided case

The two-sided case differs from the previous one only in that $H_A$ now consists both of the point set $\theta(p_1, p_2) = \gamma$ of section 2.3.1 and the additional point set $\theta(p_1, p_2) = \frac{1}{2}$.

We limit ourselves here to indicating the typical shapes of first-stage continuation region and second-stage critical regions.

The possible continuation region will be collections of grid points forming essentially elliptical bands, bounded by some two of the elliptical contours, as given in Figure 2.1 for $M_1 = 20$ and $\gamma = 2$. $\alpha$ and $\beta$ determine the relevant band (i.e., relevant pair of elliptical contours).

Figure 2.5 and 2.6 shows the elliptical character of the possible second-stage acceptance and complementary rejection regions, given $M_2(d) = 40$. Figure 2.6
Figure 2.3: "Continuation region" and hypotheses for $\delta^*$
Table 2.3: Second-stage sample size $M_2(d)$: $M_1 = 20$

| $d_C \setminus d_T$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------------------|---|---|---|---|---|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 20                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 196 | 198 |
| 18                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 142 | 178 | 198 | 198 |
| 17                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 140 | 162 | 180 | 195 | 195 |
| 16                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 138 | 159 | 176 | 180 | 178 | 160 |
| 15                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 109 | 135 | 154 | 166 | 175 | 175 | 161 | 139 |
| 14                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 108 | 132 | 149 | 160 | 166 | 166 | 160 | 142 | 113 |
| 13                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 101 | 124 | 143 | 151 | 157 | 158 | 154 | 145 | 128 | 104 |
| 12                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 123 | 136 | 147 | 155 | 156 | 151 | 143 | 130 | 110 | 0 | 0 |
| 11                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 123 | 136 | 147 | 155 | 156 | 151 | 143 | 130 | 110 | 0 | 0 |
| 10                  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 123 | 135 | 146 | 155 | 155 | 148 | 143 | 130 | 112 | 88 | 0 | 0 |
| 9                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 124 | 136 | 147 | 155 | 155 | 147 | 143 | 132 | 113 | 91 | 0 | 0 |
| 8                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 125 | 137 | 148 | 155 | 155 | 146 | 143 | 132 | 112 | 91 | 0 | 0 |
| 7                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 125 | 137 | 148 | 155 | 155 | 146 | 143 | 132 | 112 | 91 | 0 | 0 |
| 6                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 136 | 150 | 153 | 155 | 154 | 148 | 143 | 132 | 112 | 98 | 66 | 0 | 0 |
| 5                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 138 | 152 | 164 | 164 | 159 | 152 | 124 | 126 | 112 | 98 | 66 | 0 | 0 |
| 4                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 148 | 164 | 167 | 166 | 165 | 153 | 142 | 129 | 113 | 91 | 0 | 0 | 0 |
| 3                   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 161 | 168 | 180 | 171 | 167 | 155 | 148 | 129 | 112 | 92 | 0 | 0 | 0 |
| 2                   | 0 | 179 | 198 | 196 | 187 | 179 | 164 | 148 | 129 | 111 | 87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1                   | 176 | 206 | 206 | 199 | 188 | 168 | 149 | 128 | 102 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0                   | 176 | 206 | 206 | 199 | 179 | 150 | 121 | 92 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
illustrate the fact that, when $d$ shows some imbalance in favor of the test population versus the control population (i.e., $d_2 > d_1$), then values of $D$ most strongly pointing to $H_0$ are those that undo this imbalance (i.e., $D_2 < D_1$). Figure 2.5 may be related to the fact that, when both sample proportions are near 0 or 1, there is no strong evidence for $H_0$ even when the control and test positive counts are nearly the same; if so, the relatively weak evidence for $H_0$ provided by a pair of very small equal counts, such as (3, 3), presumably is best bolstered by second-stage evidence both essentially maintaining the equality and also bringing both total counts ($d_i + D_i$), $i = 1, 2$ to intermediate range.

2.4 Fair comparisons with optimal standard two-stage plans

The Bayes/Neyman-Pearson correspondence underlying sections 2.2 and 2.3 essentially applies as well to optimal non-flexible two-stage plans, i.e., plans with positive second-stage sample size not depending on first-stage outcome. One aspect of identifying such plans is consideration of all partitions of the set of possible first-stage outcomes $d$ into three sets.

The Neyman-Pearson framework now calls for finding that partition, together with the assigning of immediate acceptance of either hypothesis or continuation to the three elements of the partitions, minimizing A.S.N averaged over both hypotheses, subject to error rate restrictions of both kinds.

For the Bayes framework, for example for the formulation of section 2.2, involving unit sampling cost and wrong-decision losses $C_0$ and $C_A$, one considers the posterior
Figure 2.1: Typical continuation points "*" and "Continuation region" boundary contours for $\lambda_1 = 20$
Figure 2.5: Second-stage critical regions $A_4$ for $M_2(d) = 40$ for $d = (3.3)$
Figure 2.6: Second-stage critical regions $A_n$ for $M_2(d) = 40$ for $d = (10, 15)$
probability of $H_0$

$$
\pi_0(S) = \frac{\pi_0 \sum_{d \in S} b(d; M_1, p_0)}{\pi_0 \sum_{d \in S} b(d; M_1, p_0) + \pi_A \sum_{d \in S} b(d; M_1, p_A)}
$$

conditionally on the element $S$ of a partition. Each of the three elements $S$ of a partition thereby is assigned an optimal decision: to accept either hypothesis, or to continue sampling (in which case $S$ is denoted by $T$) with additional sample size $M_2(T)$ and critical region based on the criterion with numerator

$$C_0 \pi_0(T) \left( \frac{\sum_{d \in T} b(d; M_1, p_0) b(D; M_2(T), p_0)}{\sum_{d \in T} b(d; M_1, p_0)} \right).$$

The Bayes non-flexible two-stage plan minimizing expected risk is found by searching over all partitions. We note that, for either framework, likelihood ratio monotonicity allows limiting the search over partitions to just the usual ordered intervals.

The resulting Bayes procedure has well-defined error rates $(\alpha^*_\delta, \beta^*_\delta)$, and is, again, optimal for $(\alpha^*_\hat{\delta}, \beta^*_\hat{\delta})$, in the sense of minimizing average ASN over all labeled partitions, with error rates restricted to be no greater than $(\alpha^*_\delta, \beta^*_\delta)$. As $C_0$ and $C_A$ are varied, one thus generates a collection of efficient non-flexible two-stage plans $\delta$, together with their error rates $(\alpha^*_\delta, \beta^*_\delta)$. Typically, because of discreteness, no member of the above collection of error rates $(\alpha^*_\delta, \beta^*_\delta)$ will coincide with $(\alpha^*_\hat{\delta}, \beta^*_\hat{\delta})$. We therefore have gone, in a manner entirely analogous to the usual randomization over two procedures when only $\alpha$ is targeted in the discrete case, to randomization over three plans $\delta$, with the three error rate vectors $(\alpha^*_\delta, \beta^*_\delta)$ all near $(\alpha^*_\hat{\delta}, \beta^*_\hat{\delta})$ and the three randomization weights chosen so that the randomization expectation of $(\alpha^*_\delta, \beta^*_\delta)$ equals $(\alpha^*_\hat{\delta}, \beta^*_\hat{\delta})$.

The three plans $\hat{\delta}$, together with their randomization weights, are determined as follows: (i) With $(\alpha^*_\hat{\delta}, \beta^*_\hat{\delta})$ as origin, for each of the four quadrants determine the
point \((a^*_\delta, b^*_\delta)\) nearest (in the sense of Euclidean distance) to \((a^*_\delta, b^*_\delta)\). (ii) Except for degenerate situations (e.g., when \((a^*_\delta, b^*_\delta)\) lies in the convex hull of two of these four points), \((a^*_\delta, b^*_\delta)\) will in fact lie in the convex hull of exactly two of the four possible triples that can be formed from the four points in question, corresponding to two possible experimentwise randomizations over three optimal non-flexible two-stage plans, each of the two randomization expectations of \((a^*_\delta, b^*_\delta)\) equaling \((a^*_\delta, b^*_\delta)\).

(iii) Of these two possible experimentwise randomizations, the one with the smaller randomization expectation of \(\frac{\text{E}(\mathcal{N}_2)}{\delta}\) is chosen as representing the comparable optimal non-flexible plan \(\delta^0\).

For the example in section 2.2.2, \(\delta^*\) yields \((a^*_\delta, b^*_\delta) = (0.04949, 0.19821)\). The four optimal non-flexible plans \(\delta\), call them A, B, C and D, all turn out to have the same continuation region \(T\) as does \(\delta^*\), with second-stage sample size \(m_2\) and critical values \(k_2\) as given in Table 2.1. Table 2.1 also shows the two possible three-plan randomizations mentioned in (ii), and, in accordance with (iii), identifies \(\delta^0\) as the experimentwise randomization over ABC.

For the formulation of section 2.3.1, motivated by the above-mentioned coincidence of continuation regions we have taken the flexible continuation region as applying to the non-flexible case. Regarding the construction of critical regions, we observed and used monotonicity in \(D_2\) of the conditional likelihood ratio with numerator

\[
C_{0^*\delta_0}(T) = \frac{\sum_{(d_C, d_T) \in T} b(d_C; M_1, p_C) b(d_T; M_1, p_T) b(D_C; m_1, p_C) b(D_T; m_2, p_T)}{\sum_{(d_C, d_T) \in T} b(d_C; M_1, p_C) b(d_T; M_1, p_T)}
\]

and corresponding denominator.
Table 2.4: \( \delta^o \) and the four “nearest” plans \( \delta \) for the example in section 2.2.2

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( \alpha^*_{\delta} )</th>
<th>( \beta^*_{\delta} )</th>
<th>( (m_2,k_2) )</th>
<th>( \text{ASN}_{\delta}^{(2)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A : (−,+)**</td>
<td>0.04186</td>
<td>0.20222</td>
<td>(55.13)</td>
<td>22.85973</td>
</tr>
<tr>
<td>B : (+,+)**</td>
<td>0.04989</td>
<td>0.21027</td>
<td>(47.11)</td>
<td>19.53468</td>
</tr>
<tr>
<td>C : (+,−)**</td>
<td>0.05921</td>
<td>0.16606</td>
<td>(55.12)</td>
<td>22.85973</td>
</tr>
<tr>
<td>D : (−,−)**</td>
<td>0.04857</td>
<td>0.19160</td>
<td>(52.12)</td>
<td>21.61279</td>
</tr>
<tr>
<td>BCD</td>
<td>0.04949</td>
<td>0.19821</td>
<td></td>
<td>21.74890(^b)</td>
</tr>
<tr>
<td>ABC \equiv \delta^o</td>
<td>0.04949</td>
<td>0.19821</td>
<td></td>
<td>21.26440(^c)</td>
</tr>
</tbody>
</table>

\( a(.+) \) denotes \( \alpha^*_{\delta} < \alpha_{\delta^*} \) and \( \beta^*_{\delta} > \beta_{\delta^*} \)

\(^b\)Randomization expectation of \( \text{ASN}_{\delta}^{(2)} \)

\(^c\)Randomization expectation of \( \text{ASN}_{\delta}^{(2)} \)

\[ C_{\alpha_{\delta}}(T) \]

\[ \left( \frac{\sum_{(d_C,d_T) \in T} b(d_C:M_1:p_C)b(d_T:M_1:p_T)b(D_C:m_2:p_C)b(D_T:m_2:p_T)}{\sum_{(d_C,d_T) \in T} b(d_C:M_1:p_C)b(d_T:M_1:p_T)} \right) \]

analogous to (2.8) and (2.9), respectively.

The resulting analysis is summarized in Table 2.5, with the experimentwise ran­domization over ACD representing the comparable optimal non-flexible plan.

2.5 Remarks concluding chapter 2

The optimization underlying Bayes flexible plans is made practical by separabil­ity, allowing separate analysis of each of a collection of two-parameter optimization problems, one each for every possible first-stage outcome.
Table 2.5: $\delta^0$ and the four “nearest” plans $\hat{\delta}$ for the example in section 2.3.1

<table>
<thead>
<tr>
<th>$\hat{\delta}$</th>
<th>$\alpha_{\hat{\delta}}$</th>
<th>$\beta_{\hat{\delta}}$</th>
<th>$m_2^a$</th>
<th>$\overline{\text{ASN}}^{(2)}_{\hat{\delta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A: (-,+)^b$</td>
<td>0.04997</td>
<td>0.19966</td>
<td>192</td>
<td>137.50033</td>
</tr>
<tr>
<td>$B: (+,+)$</td>
<td>0.05016</td>
<td>0.19931</td>
<td>192</td>
<td>137.50033</td>
</tr>
<tr>
<td>$C: (+,-)$</td>
<td>0.05029</td>
<td>0.19906</td>
<td>192</td>
<td>137.50033</td>
</tr>
<tr>
<td>$D: (-,-)$</td>
<td>0.04966</td>
<td>0.19898</td>
<td>193</td>
<td>138.32065</td>
</tr>
<tr>
<td>BCD</td>
<td>0.04999</td>
<td>0.19924</td>
<td></td>
<td>137.75244$^c$</td>
</tr>
<tr>
<td>ACD $\equiv \delta^0$</td>
<td>0.04999</td>
<td>0.19924</td>
<td></td>
<td>137.74911$^d$</td>
</tr>
</tbody>
</table>

$^a$critical region boundaries are near-linear in $(D_C, D_T)$
$^b$(-+) denotes $\alpha_{\hat{\delta}} < \alpha_{\delta^*}$ and $\beta_{\hat{\delta}} > \beta_{\delta^*}$
$^c$randomization expectation of $\overline{\text{ASN}}^{(2)}_{\hat{\delta}}$
$^d$randomization expectation of $\overline{\text{ASN}}^{(2)}_{\delta}$

By the Lagrangian approach to the solution of constrained optimization problems, Bayes flexible plans are also efficient flexible plans in a certain Neyman-Pearson sense. In the language of that Neyman-Pearson framework, flexibility of the second stage has been shown to lead to improvement in second-stage expected sample size in comparison with optimal non-flexible plans. For the parametric cases considered, this improvement appears to be uniform in $p$, and has been found to be maximal, and in the neighborhood of 7 %, for intermediate parameter values. Practitioners, no doubt inclined in any event toward adaptive treatment of the second stage, should welcome even this small improvement.

The simultaneous interpretation of $\delta^*$ in Bayesian and Neyman-Pearson terms
allows juxtaposing the pair of Bayesian wrong-decision losses, as multiples of sampling cost, and the pair of error rates of both kinds. For example, in the case of section 2.2.2, wrong-decision losses of (1110.456) are "matched" with error rates (0.04949.0.19521). Similarly, in the case of Section 2.3.1, wrong-decision losses (4740.2152) are "matched" with average error rates (0.4999.0.19924).

When the one-sample formulation of section 2.2.2 is interpreted as pertaining to the treatment population, sections 2.2.2 and 2.3.1 are seen as addressing similarly constrained statistical formulations of essentially the same subject matter question, through two alternative experimental plans. Thus it may not be inappropriate to view the two sections in conjunction with one another. Inferred wrong-decision losses for the formulation of section 2.3.1 are roughly four times greater than the inferred wrong-decision losses for the formulation of section 2.2.2; also, the optimal expected sample size \( \frac{\overline{ASN}(2) \delta^* + M_1}{\overline{ASN}(2) \delta^* + M_1} \) equals 39.40 for the formulation of section 2.2.2, and the optimal expected sample size \( 2(\overline{ASN}(2) \delta^* + M_1) \) equals 298.51 for the formulation of section 2.3.1, greater by a factor of approximately eight. Thus sampling cost is seen as considerably more (roughly twice more) sensitive to a change in experimental plan than is inferred cost of erroneous conclusion. Was this to be expected?

Note that the matching of wrong-decision losses with error rates is in fact many-to-one, in that, because of discreteness, a small \((C_0,C_1)\)-neighborhood will determine the same \( \delta^* \), and hence will be matched with a single error rate pair \((\alpha_{\delta^*},\beta_{\delta^*})\). It will of course be one-to-one in continuous, e.g., normal, cases. Note also that the above inferred wrong-decision losses are reminiscent of the shadow prices of linear programming, and, also, of the weights associated with the sequential Bayes interpretation (Lehmann [20]) of the \( S P R T \).
CHAPTER 3. GROUP-SEQUENTIAL ASPECTS

3.1 Introduction

This chapter deals with two-stage test plans with stage-specific error rates of both kinds. At the end of the first-stage an interim analysis is performed with the objective of deciding whether or not to continue the study based on results of the interim analysis. If the study is continued, the first-stage information is systematically put to work in conducting the second-stage, including its sample size and critical region, with the goal of minimizing an agree-upon measure of sampling effort, subject to agreed-upon overall, as well as stage-specific error rates of both kinds. Where this approach differs from previous work is in our casting of the design of two-stage plans in the form of stage-specific constrained optimization problems. This constraints of the optimization borrow from “group sequential” formulations the idea of “allocating” error between the two stages: indeed we go beyond the usual group sequential formulation in that both the errors of the first and of the second kind are so allocated. The objective functions for optimization problems are conditional averages of sampling effort measures.

The numerical findings in the previous chapter suggest that the focus of the previous chapter, in exact binomial computations in two-sample case, and in the treatment of composite hypotheses through average error rate control, will not produce a prac-
tically feasible sample size; this despite the fact that the plans that we developed are optimal in that framework. This point can be made clear by considering the simplest case, two-sample single-stage plan: the approach based on applying normal approximations to differences of sample proportions leads to four times smaller sample size than the composite-hypothesis approach in chapter 2.

We therefore consider two-sample plans, an approximation approach, restricted to those that condense the binomial data to differences of arcsine transformations of sample proportions (as in Thall et al. [29]). This formulation allows simple hypotheses, as well as appealing to the very convenient equivariance properties of drift-parameterized standard Wiener processes. As mentioned above, we consider not just the approximation of the formulation of chapter 2, but more flexible way to handle stage-wise error restrictions and objective functions, devoted to group-sequential formulations.

Thus, as others have considered (Lan and DeMets [19], Jennison [17]), we deal with testing the drift parameter of the standard Wiener process. This allows couching the required optimization in analytical terms, which, among other things, yields optimal second-stage critical values as linear functions of optimal second-stage sample sizes. In addition, as indicated in section 3.5, one Wiener optimization provides approximate optimizations for families of binomial (or normal) problems.

Sections 3.2 and 3.3 respectively provide the details of our formulation and optimization. Section 3.5 provides some numerical results together with numerical comparisons with other possible procedures described in section 3.4. These include the single sampling (SS) plan, the sequential probability ratio test (SPRT) and the optimal standard two-stage plan where “standard” means “without flexibility of
Section 3.6 illustrates applications to binomial testing problems and Section 3.7 contains concluding remarks.

3.2 Formulation

Let $\mu$ be the drift parameter of a standard Wiener process. We consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1 > \mu_0$, with error probabilities $P\{\text{Accept } H_1 | \mu = \mu_0\} = \alpha$ and $P\{\text{Accept } H_0 | \mu = \mu_1\} = \beta$.

In accordance with the group-sequential point of view, $\alpha$ and $\beta$ are allocated to the two stages ($\alpha_1 + \alpha_2 = \alpha$ and $\beta_1 + \beta_2 = \beta$). Such allocations which is to be distinguished from the repeated-significant-test usage (Armitage [1]) make possible independent assignment of error rates to both stages. Under the given $\alpha_1$ and $\beta_1$, the continuation region, $(l, u)$, is determined by initial sampling time $T_0$. To satisfy the four stage-specific error rate restrictions, $T_0$ must be taken from a feasible interval $(0, T_0]$. Given $(T_0, l, u)$, $\mu_1$ (respectively, $\mu_0$) is to be accepted at time $T_0$ if the process exceeds $u$ (respectively, is below $l$). Further, if it equals an intermediate value $s$, $l < s < u$, then a predetermined second stage sampling plan is implemented, with sampling time $T_s$ and critical value $k_s$ depending on $s$. Thus our plan is determined by three numbers $(T_0, l, u)$, plus the two functions $\{T_s : l < s < u\}$ and $\{k_s : l < s < u\}$.

The plan $(T_0, l, u, \{T_s : s \in (l, u)\}, \{k_s : s \in (l, u)\})$ is to be chosen in such a way as to minimize an objective function measuring sampling effort, subject to stage-specific error rate restrictions. Among the objective functions so considered are the following ($\text{AST} \equiv \text{Average Sampling Time}$):
The first objective function is the unweighted average of expected sampling times for the values of \( \mu \) between \( \mu_0 \) and \( \mu_1 \). The second is the expected sampling time for \( \mu = \bar{\mu} \). The third is the expected sampling time for \( \mu = \mu_1 \). The fourth objective function is the weighted average of expected sampling times for all values of \( \mu \) in the interval \((-\infty, \infty)\), the weights being provided by the normal distribution of \( \mu \) centered at \( \bar{\mu} \) with unit variance.

These objective functions were considered previously by several authors in deriving their optimal plans. Especially, Colton and McPherson [9] consider the third one for deriving their optimal plans which emphasize clinical efficiency when actual treatment differences exist. Also, minimizing the second one, known as the Keifer-Weiss problem, is recently considered by Jennison [17]. The first one is a generalization of a simple average of average sampling efforts at two hypothesis points as considered by Thall et al. [28]. The last one is, in fact, a common choice of many Bayesian approaches because of its conjugation.
3.3 Optimization

In this section we give an account of the manner in which we identify the optimal flexible plan, in the case of the first objective function \( OF_1 \). The other objective functions in section 3.2, i.e., \( OF_2, OF_3 \) and \( OF_4 \) can be treated in similar fashion.

3.3.1 Optimization for given \((\mu_0, \mu_1)\). \( \mu_0 < \mu_1 \)

We begin by fixing first-stage sampling time \( T_0 \) and error rates \( \alpha_1 \) and \( \beta_1 \). We then determine a "continuation interval" \((l, u)\) such that the first-stage error rates are \( \alpha_1 \) and \( \beta_1 \). Given \( T_0, l \) and \( u \), we then minimize \( \overline{AST} \cdot OF_1 \) (or, equivalently, \( \overline{AST}^{(2)} \equiv \overline{AST} - T_0 \)) subject to the restrictions that the second-stage error rates equal \( \alpha_2 \) and \( \beta_2 \), respectively. This minimization is with respect to the positive second-stage sampling time function \( T_s \), \( l < s < u \), and unrestricted second-stage critical value function \( k_s \), \( l < s < u \).

For given \( T_s \) and \( k_s \),

\[
\overline{AST}^{(2)} = (\mu_1 - \mu_0)^{-1} \int_{\mu_0}^{\mu_1} \int_{\mu_0}^{\mu_1} \phi(s \mid \mu T_0, \sqrt{T_0} \cdot T_s) \cdot ds \cdot d\mu. \tag{3.1}
\]

where \( \phi(s \mid \mu T_0, \sqrt{T_0}) \) is the normal density with mean \( \mu T_0 \) and standard deviation \( \sqrt{T_0} \). Interchanging the order of integration in (3.1), the restricted minimization of (3.1) is written

\[
\text{minimize} \{T_s\}, \{k_s\}
\]

\[
(\mu_1 - \mu_0)^{-1} \int_{l}^{u} [\Phi\left(\frac{\mu}{\sqrt{T_0}} - \frac{s}{\sqrt{T_0}}\right) - \Phi\left(\frac{\mu_0}{\sqrt{T_0}} - \frac{s}{\sqrt{T_0}}\right)] \cdot [T_s] \cdot ds
\]

subject to
\[ \int_1^u \phi(s | \mu T_0, \sqrt{T_0}) \int_{k_s}^\infty \phi(x | \mu T_s, \sqrt{T_s}) \, dx \, ds = \alpha_2 \]

and
\[ \int_1^u \phi(s | \mu T_0, \sqrt{T_0}) \int_{k_s}^\infty \phi(x | \mu T_s, \sqrt{T_s}) \, dx \, ds = \beta_2. \]

where \( \Phi(y) = \int_{-\infty}^y \phi(x | 0, 1) \, dx. \)

We implement this restricted minimization by subdividing the continuation region \((l, u)\), as determined by the first-stage sampling time \(T_0\), into a grid of \(2^r\) grid points, \( l + \frac{(u-1)}{2^r+1}, l + \frac{3(u-1)}{2^r+1}, \ldots, u - \frac{(u-1)}{2^r+1} \). This was in fact done for both \(r = 5\), \(r = 6\) and \(r = 7\), and it was found that the computation below yielded essentially the same solutions. In view of this we decided that \(r = 6\) provided a sufficiently accurate approximate formulation, and recast our problem as follows (with \(\Delta \equiv \frac{(u-l)}{2^r}, T_i \equiv T_{s_i}, k_i \equiv k_s, \) where \(s_i = l + \frac{(2i-1)(u-l)}{2^r+1}, i = 1, 2, \ldots, m-1, m; r = 6, m = 64\)).

minimize \( \{T_i\}_{i=1}^m, \{k_i\}_{i=1}^m \)
\[ (\mu_1 - \mu_0)^{-1} \sum_{i=1}^m \left[ \frac{\Phi(\frac{T_i - \Delta}{\sqrt{T_0}}) - \Phi(\frac{\mu T_0 - \Delta}{\sqrt{T_0}})}{\sqrt{T_0}} \right] \cdot [T_j] \cdot [\Delta] \] \tag{3.2}

subject to
\[ \sum_{i=1}^m \phi(s_i | \mu T_0, \sqrt{T_0}) \int_{k_i}^\infty \phi(x | \mu T_s, \sqrt{T_s}) \, dx \cdot [\Delta] - \alpha_2 = 0 \]

and \[ \sum_{i=1}^m \phi(s_i | \mu T_0, \sqrt{T_0}) \int_{-\infty}^{k_i} \phi(x | \mu T_s, \sqrt{T_s}) \, dx \cdot [\Delta] - \beta_2 = 0. \]

We abbreviate formulation (3.2) to
\[ \text{minimize} \{T_i\}_{i=1}^m, \{k_i\}_{i=1}^m \quad f(T_1, \ldots, T_m, k_1, \ldots, k_m) \] \tag{3.3}
subject to
\[ g_0(T_1, \ldots, T_m, k_1, \ldots, k_m) - \alpha_2 = 0 \]

and
\[ g_1(T_1, \ldots, T_m, k_1, \ldots, k_m) - \beta_2 = 0. \]

where \( f, g_0 \) and \( g_1 \) are of additive form:

\[
\begin{align*}
\phi(T_1, \ldots, T_m) & \equiv \sum_{i=1}^{m} \eta_i T_i, \\
g_0(T_1, \ldots, T_m, k_1, \ldots, k_m) & \equiv \sum_{i=1}^{m} g_{0i}(T_i, k_i) \\
\text{and} \\
g_1(T_1, \ldots, T_m, k_1, \ldots, k_m) & \equiv \sum_{i=1}^{m} g_{1i}(T_i, k_i).
\end{align*}
\]

We address this problem via a standard Kuhn-Tucker argument involving the Lagrangian kernel

\[
L(T_1, \ldots, T_m, k_1, \ldots, k_m; \rho_0, \rho_1) = f(T_1, \ldots, T_m, k_1, \ldots, k_m)
\]

\[
+ \rho_0 \cdot g_0(T_1, \ldots, T_m, k_1, \ldots, k_m)
\]

\[
+ \rho_1 \cdot g_1(T_1, \ldots, T_m, k_1, \ldots, k_m)
\]

\[
= \sum_{i=1}^{m} [\eta_i T_i + \rho_0 \cdot g_{0i}(T_i, k_i) + \rho_1 \cdot g_{1i}(T_i, k_i)]
\]

\[
= \sum_{i=1}^{m} h_i(T_i, k_i). \quad (3.5)
\]
The Kuhn-Tucker argument for solving (3.3) now proceeds as follows (with 
\( (T_1, \ldots, T_m) \equiv \mathbf{T} \) and \( (k_1, \ldots, k_m) \equiv \mathbf{k} \):

a) Initialize the argument by fixing positive Lagrangian multipliers \( (\rho_0, \rho_1) \).

b) For each \( i, 1 \leq i \leq m \), find \( (\mathbf{T}_i, k_i) \) by minimizing the 2-dimensional Lagrangian function \( f_i(T_i, k_i) \equiv \eta_i T_i + \rho_0 g_0 (T_i, k_i) + \rho_1 g_1 (T_i, k_i) \) with respect to \( T_i > 0 \) and unrestricted \( k_i \).

c) Now compute \( \delta_2 (\rho_0, \rho_1) \equiv g_0 (\hat{T}, \hat{k}) \) and \( \beta_2 (\rho_0, \rho_1) \equiv g_1 (\hat{T}, \hat{k}) \).

If \( (\alpha_2 (\rho_0, \rho_1), \beta_2 (\rho_0, \rho_1)) \) does not equal \( (\alpha_2, \beta_2) \) to within adequate approximation, iterate the above steps until \( (\alpha_2 (\rho_0, \rho_1), \beta_2 (\rho_0, \rho_1)) = (\alpha_2, \beta_2) \) to within adequate approximation. If so, then denote the so obtained \( (\hat{\rho}_0, \hat{\rho}_1) \) by \( (\rho_0^*, \rho_1^*) \) and corresponding \( (\hat{T}, \hat{k}) \) by \( (T^*, k^*) \).

d) In view of (3.4) and b), \( (T^*, k^*) \) in fact minimizes the 2m-dimensional Lagrangian function \( f(\mathbf{T}) + \rho_0^* \cdot g_0 (\mathbf{T}, \mathbf{k}) + \rho_1^* \cdot g_1 (\mathbf{T}, \mathbf{k}) \).

e) In view of c), d) and the section 2.2 with \( A = \alpha_2, B = \beta_2, \delta = (\mathbf{T}, \mathbf{k}) \). \( f = OF_1 \) and \( \delta^* = (T^*, k^*) \). \( (T^*, k^*) \) is seen to be optimal, in the sense of minimizing \( OF_1 \) among all \( (\mathbf{T}, \mathbf{k}) \) satisfying \( g_0 (\mathbf{T}, \mathbf{k}) \leq \alpha_2 \) and \( g_1 (\mathbf{T}, \mathbf{k}) \leq \beta_2 \).

Step b) is based on graphical evidence that \( L_i (\cdots) \) attains a minimum in \( \Omega \equiv (0, \infty) \times (0, \infty) \). This evidence is used as follows. It is easy to show that \( L_i (\cdots) \) has first-order partial derivatives of both kinds everywhere in \( \Omega \). Furthermore, it is true that \( L_i (T_i, \cdots) \) is strictly convex in \( k_i \) for all \( T_i \in (0, \infty) \), with a unique minimum in \( (0, \infty) \). Given the above graphical evidence, it can therefore be concluded that \( L_i (T_i, k_i) \) has a unique minimum in \( (0, \infty) \times (0, \infty) \), provided by setting both first-order partials equal to zero. Figure 3.1 shows contours of \( L_i (T_i, k_i) \) for the given \( T_0 \) and \( i = 1 \) (i.e., \( s = i \)). Figure 3.2 illustrates the fact that function \( L_i (T_i, k_i^*) \) is
U-shaped, where \( k_i^* \) (\( \equiv k(T_i^*) \)), further discussed below, is the optimal \( k_i \) associated with \( T_i^* \).

Figure 3.1: Contours of \( L_1(T_1, k_1) \) when \((T_0, \rho_0, \rho_1)=(18.9, 193.31, 109.89)\)

The iteration in c) is made easier by the fact that \( \alpha_2(\rho_0, \rho_1) \) is largely sensitive to \( \rho_0 \), while \( \beta_2(\rho_0, \rho_1) \) is largely sensitive to \( \rho_1 \).

We now give some further details of our solution of (3.3) in the range of \( T_0 \) where the stationary Lagrangian equations can be solved, and our determining of \( T_0 \), the largest \( T_0 \). For the simpler notation, let \( c_s = \frac{k_s}{\sqrt{T_s}} \). \( \forall s \in \{l, u\} \).

We set up the Lagrangian function for the problem (3.3) in (3.5) and differentiate with respect to each of \( 2m \) variables \( \{T_i\}_{i=1}^m \) and \( \{c_i\}_{i=1}^m \). Setting the resulting \( 2m \) expressions equal to zero produce the following \( 2m \) equations:
Figure 3.2: Function $L_1(T_1, k_1^*)$ when $(T_0, \rho_0, \rho_1) = (18.9, 193.31, 109.89)$

$$\frac{\partial}{\partial T_i} = (\mu_1 - \mu_0)^{-1} \left[ \Phi\left( \frac{\mu_1 T_0 - s_i}{\sqrt{T_0}} \right) - \Phi\left( \frac{\mu_0 T_0 - s_i}{\sqrt{T_0}} \right) \right]$$

$$- \rho_0 \cdot \phi(s_i | \mu_0 T_0 \cdot \sqrt{T_0}) \cdot \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_i - \mu_0 \cdot \sqrt{T_0})^2} \right) \cdot -\frac{\mu_0}{2\sqrt{T_i}}$$

$$+ \rho_1 \cdot \phi(s_i | \mu_1 T_0 \cdot \sqrt{T_0}) \cdot \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_i - \mu_1 \cdot \sqrt{T_0})^2} \right) \cdot -\frac{\mu_1}{2\sqrt{T_i}}$$

$$i = 1, 2, \ldots, m - 1, m$$

and

$$\frac{\partial}{\partial c_i} = - \rho_0 \cdot \phi(s_i | \mu_0 T_0 \cdot \sqrt{T_0}) \cdot \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_i - \mu_0 \cdot \sqrt{T_0})^2} \right)$$

$$+ \rho_1 \cdot \phi(s_i | \mu_1 T_0 \cdot \sqrt{T_0}) \cdot \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_i - \mu_1 \cdot \sqrt{T_0})^2} \right)$$

$$i = 1, 2, \ldots, m - 1, m$$

(3.5)
By (3.7), we have
\[ s_i + k_i - \left(\frac{T_0 + T_i}{2}\right)(\mu_0 + \mu_1) + \frac{\ln(P_i)}{(\mu_1 - \mu_0)} = 0. \]
\[ i = 1, 2, \ldots, m - 1, m \] (3.8)

Substituting (3.7) into (3.6), we find
\[
(\mu_1 - \mu_0)^{-1}\left[\Phi\left(\frac{\mu_1 T_0 - s_i}{\sqrt{T_0}}\right) - \Phi\left(\frac{\mu_0 T_0 - s_i}{\sqrt{T_0}}\right)\right]
+ \rho_1 \cdot \sigma(s_i \mid \mu_1 T_0, \sqrt{T_0}) \cdot \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_i - \mu_1 \sqrt{T_i})^2} \cdot \frac{\mu_0 - \mu_1}{2\sqrt{T_i}}\right) = 0.
\]
\[ i = 1, 2, \ldots, m - 1, m \] (3.9)

Relation (3.9) can be simplified by following steps:
\[
\rho_1 \cdot \sigma(s_i \mid \mu_1 T_0, \sqrt{T_0}) \cdot \sigma(c_i \mid \mu_1 \sqrt{T_i}, 1) \cdot \frac{\mu_0 - \mu_1}{2\sqrt{T_i}}
= - (\mu_1 - \mu_0)^{-1}\left[\Phi\left(\frac{\mu_1 T_0 - s_i}{\sqrt{T_0}}\right) - \Phi\left(\frac{\mu_0 T_0 - s_i}{\sqrt{T_0}}\right)\right],
\]
\[ i = 1, 2, \ldots, m - 1, m \]

and
\[
\sigma(k_i \mid \mu_1 T_i, \sqrt{T_i}) = \frac{2}{(\mu_1 - \mu_0)^2} \cdot \sigma(s_i \mid \mu_1 T_0, \sqrt{T_0}) \cdot \rho_1 \cdot \sigma(c_i \mid \mu_1 \sqrt{T_i}, 1).
\]
\[ i = 1, 2, \ldots, m - 1, m \]

\[ \exp \left( -\frac{1}{2} \left( \frac{k_i - \mu_1 T_i}{\sqrt{T_i}} \right)^2 \right) = \frac{2\sqrt{2\pi T_i} \cdot \left[ \Phi \left( \frac{\mu_1 T_0 - s_i}{\sqrt{T_0}} \right) - \Phi \left( \frac{\mu_0 T_0 - s_i}{\sqrt{T_0}} \right) \right]}{(\mu_1 - \mu_0)^2 \cdot \sigma(s_i | \mu_1 T_0, \sqrt{T_0}) \cdot \rho_1} \]

Taking logs and square roots, we finally obtain

\[ k_i = \mu_1 T_i \pm \sqrt{T_i} \ln \left( \frac{(\mu_1 - \mu_0)^4 \cdot \sigma(s_i | \mu_1 T_0, \sqrt{T_0})^2 \cdot \rho_1^2}{8\pi T_i \cdot \left[ \Phi \left( \frac{\mu_1 T_0 - s_i}{\sqrt{T_0}} \right) - \Phi \left( \frac{\mu_0 T_0 - s_i}{\sqrt{T_0}} \right) \right]^2} \right), \]

\[ i = 1, 2, \ldots, m - 1, m \tag{3.10} \]

Substituting (3.10) into (3.8), we have

\[ h'(T_i, k(T_i)) = s_i - \frac{T_0}{2} (\mu_0 + \mu_1) + \frac{T_i}{2} (\mu_1 - \mu_0) \pm \]

\[ \sqrt{T_i} \ln \left( \frac{(\mu_1 - \mu_0)^4 \cdot \sigma(s_i | \mu_1 T_0, \sqrt{T_0})^2 \cdot \rho_1^2}{8\pi T_i \cdot \left[ \Phi \left( \frac{\mu_1 T_0 - s_i}{\sqrt{T_0}} \right) - \Phi \left( \frac{\mu_0 T_0 - s_i}{\sqrt{T_0}} \right) \right]^2} \right), \]

\[ i = 1, 2, \ldots, m - 1, m \tag{3.11} \]

The following Lemma is useful to detail those minimizations.

**Lemma 3.1** Suppose that both \( \rho_0 \) and \( \rho_1 \) are positive. Then, for fixed \( T_i \),

\[ h_i(T_i, k(T_i)) \leq h_i(T_i, k_i). \quad -\infty < k_i < +\infty. \]
where

\[ k(T_i) = \left( \frac{\mu_0 + \mu_1}{2} \right) \cdot (T_i + T_0) - [ s + \frac{\ln(\frac{\rho_T}{\rho_0})}{(\mu_1 - \mu_0)} ]. \]

**Proof:** In the purpose of the proof, we delete the subscript \( i \).

Setting the derivative of \( h(T, k) \) with respect to \( k \) equal to zero yields the relation

\[
\rho_1 \cdot \phi(s \mid \mu_1 T_0, \sqrt{T_0}) \cdot \phi(k(T) \mid \mu_1 T, \sqrt{T})
= \rho_0 \cdot \phi(s \mid \mu_0 T_0, \sqrt{T_0}) \cdot \phi(k(T) \mid \mu_0 T, \sqrt{T}).
\] (3.12)

Further, the derivative \( h'(T, k) \) of \( h(T, k) \) with respect to \( k \), evaluated at \( k(T) + \delta \), \( \delta > 0 \), equals

\[
\rho_1 \cdot \phi(s \mid \mu_1 T_0, \sqrt{T_0}) \cdot \phi(k(T) + \delta \mid \mu_1 T, \sqrt{T})
- \rho_0 \cdot \phi(s \mid \mu_0 T_0, \sqrt{T_0}) \cdot \phi(k(T) + \delta \mid \mu_0 T, \sqrt{T})
\]

\[
= \rho_1 \cdot \phi(s \mid \mu_1 T_0, \sqrt{T_0}) \cdot \phi(k(T) \mid \mu_1 T, \sqrt{T}) \cdot \exp\left( -\left( \frac{k(T)}{\sqrt{T}} - \frac{\rho_1}{2T} \right) \right) \cdot \exp(\delta \mu_1 \sqrt{T})
- \rho_0 \cdot \phi(s \mid \mu_0 T_0, \sqrt{T_0}) \cdot \phi(k(T) \mid \mu_0 T, \sqrt{T}) \cdot \exp\left( -\left( \frac{k(T)}{\sqrt{T}} - \frac{\rho_0}{2T} \right) \right) \cdot \exp(\delta \mu_0 \sqrt{T})
\]

\[
= \rho_1 \cdot \phi(s \mid \mu_1 T_0, \sqrt{T_0}) \cdot \phi(k(T) \mid \mu_1 T, \sqrt{T}) \cdot \exp\left( -\left( \frac{k(T)}{\sqrt{T}} - \frac{\rho_1}{2T} \right) \right)
\cdot \left\{ \exp(\delta \mu_1 \sqrt{T}) - \exp(\delta \mu_0 \sqrt{T}) \right\}
> 0.
\]

where the last equality follows from (3.12), and the inequality follows from \( \mu_1 > \mu_0 \). Finally, an analogous argument shows that \( h'(T, k) < 0 \) for \( k = k(T) - \delta \).

(Q.E.D.)
When we solve each of \( m \) equations in (3.11) using Lemma 3.1, there are two candidates of optimal solutions for \((T_i, k(T_i))\). As shown in Figure 3.3, there are two possibilities to obtain candidates of optimal solutions, i.e., one is that both candidates are from negative "signed" (in (3.11)) equation and one other possibility is to obtain each candidate from each "signed" equation. Problem solutions are of two types: stationary points with \( T_i > 0 \) (for each of \( m \) equations, we choose the \((T_i, k(T_i))\) as a optimal solution if they have the smaller value of the objective function among candidates of solution with satisfying restrictions, \( T_i > 0 \), and pre-determined error restrictions) and boundary points with \( T_i = 0 \). The former always correspond to negative "signed" equation solutions.

\( T_0 \) is to be chosen from the feasible interval \((0, \overline{T_0}]\) where \( \overline{T_0} \) is the largest value of \( T_0 \) which produces the nonzero problem solutions \( T_i \). \( \forall i \). \( T_0^* \) minimizes \( OF_1 \) as optimizes in \( a) - \epsilon \), over \((0, \overline{T_0}]\). Now let \((l^*, u^*)\) be the \((l, u)\) corresponding to \( T_0^* \) and \( \{T_i^*\}_{i=1}^m \), \( \{k_i^*\}_{i=1}^m \) the solutions provided by \((l^*, u^*)\) for \( T_0 = T_0^* \). Interpreting the second-stage parameters, then gives an order-\( r \) approximation \( \{T_i^*, l^*, u^*, \{T_s^*; s \in \{l^*, u^*\}\}, \{k_s^*; s \in \{l^*, u^*\}\}\) to the optimal flexible plan \( OFP^* \) for the given \((\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_0, \mu_1)\).

### 3.3.2 Optimal plan for linear transformations of \((\mu_0, \mu_1)\), \( \mu_0 < \mu_1 \)

In this section, we consider deriving the optimal flexible plan for linear transformations of standard hypotheses with given stage-wise error restrictions, i.e., \((\alpha_i, \beta_i)\), \( i = 1, 2 \). We consider in particular \( OF_1 \) as in section 3.3.1, but, of course, the other objective functions of section 3.2, i.e., \( OF_2 \), \( OF_3 \) and \( OF_1 \), can be treated in similar fashion.
Figure 3.3: Candidates of solutions for $T_i$ for each signed equation: $r = 6$
Suppose we have an optimal flexible plan \( \delta_w(T_0^*: \mu_0, \mu_1) \equiv (T_0^*, l^*, u^*, \{T_s^*: s \in (l^*, u^*)\}, \{k_s^*: s \in (l^*, u^*)\}; \mu_0, \mu_1) \) for the given \((\alpha_1, \beta_1, \beta_2, \mu_0, \mu_1)\). Considering first the case of scale transformation, let \((\mu_0^*, \mu_1^*) = (c_1 \mu_0, c_1 \mu_1), c_1 > 0\) and define \( \delta_w(T_0^*: \mu_0^*, \mu_1^*) \) as

\[
(T_0^*, l^*, u^*, \{T_s^*: s \in (l^*, u^*)\}, \{k_s^*: s \in (l^*, u^*)\})
\]

\[
= \left( \frac{T_0^*}{c_1}, \frac{l^*}{c_1}, \frac{u^*}{c_1}, \frac{T_s^*}{c_1}, s \in (l^*, u^*) \right), \{\frac{k_s^*}{c_1}, s \in (l^*, u^*)\}\right) \quad (3.13)
\]

**Lemma 3.2** Let \((\mu_0^*, \mu_1^*) = (c_1 \mu_0, c_1 \mu_1), c_1 > 0\). Then, \( \delta_w(T_0^*: \mu_0^*, \mu_1^*) \) satisfies the stage-wise error restrictions.

**Proof:** Let \( \Phi(y) = \int_{-\infty}^{y} \phi(x \mid 0, 1) \, dx \).

Then, for the plan \( \delta_w(T_0^*: \mu_0, \mu_1) \), the first-stage error probabilities can be written as

\[
\alpha_1 = 1 - \Phi \left( \frac{u^* \, T_0^*}{\sqrt{T_0^*}} \right)
\]

and

\[
\alpha_2 = \int_{l^*}^{u^*} \phi(s \mid \mu_0 T_0^*, \sqrt{T_0^*}) \left[ 1 - \Phi \left( \frac{k_s^* - \mu_0 T_s^*}{\sqrt{T_s^*}} \right) \right] ds
\]
and

\[ \beta_2 = \int_{l^*}^{u^*} \phi(s \mid \mu_1 T_0^* \sqrt{T_0^*}) \left[ \Phi \left( \frac{k_s^* - \mu_1 T_s^*}{\sqrt{T_s^*}} \right) \right] ds. \]

Then, the first-stage error probabilities for \( \delta_{u}(T_0^*, \mu_0, \mu_1) \) are

\[ 1 - \Phi \left( \frac{u^* - c_1 \mu_0 T_0^*}{\sqrt{T_0^*}} \right) = 1 - \Phi \left( \frac{\mu^* - c_1 \mu_0 T_0^*}{\sqrt{T_0^*} c_1^2} \right) \]

\[ = 1 - \Phi \left( \frac{u^* - \mu_0 T_0^*}{\sqrt{T_0^*}} \right) \]

\[ = \alpha_1 \]

and

\[ \Phi \left( \frac{l^* - c_1 \mu_1 T_0^*}{\sqrt{T_0^*}} \right) = \Phi \left( \frac{l^* - c_1 \mu_1 T_0^*}{\sqrt{T_0^*} c_1^2} \right) \]

\[ = \Phi \left( \frac{l^* - \mu_1 T_0^*}{\sqrt{T_0^*}} \right) \]

\[ = \beta_1 \]

by (3.13). And also, for the second-stage error probabilities, consider

\[ 1 - \Phi \left( \frac{k_s^* - c_1 \mu_0 T_s^*}{\sqrt{T_s^*}} \right) = 1 - \Phi \left( \frac{k_s^* - c_1 \mu_0 T_s^*}{c_1^2} \right) \]

\[ = 1 - \Phi \left( \frac{k_s^* - \mu_0 T_s^*}{\sqrt{T_s^*}} \right), \quad \forall \ s \in (l^*, u^*) \quad (3.14) \]
and
\[
\Phi(\frac{k_s^* - c_1 \mu_1 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}}) = \Phi(\frac{k_s^* - c_1 \mu_1 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}}) = \Phi(\frac{k_s^* - \mu_1 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}}), \ \forall \ s \in (l^*, u^*) \quad (3.15)
\]
Therefore, the second-stage error probabilities are
\[
\int_{l^*}^{u^*} \phi(\tilde{s} | c_1 \mu_0 \tilde{T}_0^*, \sqrt{\tilde{T}_0^*})[1 - \Phi(\frac{k_s^* - \mu_0 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}})] \, ds
\]
\[
= \int_{l^*}^{u^*} \phi(\tilde{s} | \mu_0 \tilde{T}_0^*, \sqrt{\tilde{T}_0^*})[1 - \Phi(\frac{k_s^* - \mu_0 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}})] \, ds
\]
\[
= \alpha_2 \quad (3.16)
\]
and
\[
\int_{l^*}^{u^*} \phi(\tilde{s} | c_1 \mu_1 \tilde{T}_0^*, \sqrt{\tilde{T}_0^*})[\Phi(\frac{k_s^* - \mu_1 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}})] \, ds
\]
\[
= \int_{l^*}^{u^*} \phi(\tilde{s} | \mu_1 \tilde{T}_0^*, \sqrt{\tilde{T}_0^*})[\Phi(\frac{k_s^* - \mu_1 \tilde{T}_s^*}{\sqrt{\tilde{T}_s^*}})] \, ds
\]
\[
= \beta_2 \quad (3.17)
\]
where, in the first equalities of (3.16) and (3.17), we transform \( \tilde{s} \) to \( s \) with \( \tilde{s} = \frac{s}{c_1} \)
and use equalities of (3.14) and (3.15). \( \text{Q.E.D.} \)

Now, we have a candidate optimal flexible plan for \( (\mu_0, \mu_1) \), derived from \( \delta_w(T_0^*) \cdot (\mu_0, \mu_1) \) through (3.13), which satisfies the pre-specified stage-wise error restrictions.
Proposition 3.1 optimality under scale transformation

\( \delta_w(T_0^*: c_1 \mu_0, c_2 \mu_1) \) is optimal under the given stage-wise restrictions.

Proof: Let \((\mu_0, \mu_1) = (c_1 \mu_0, c_1 \mu_1)\).

It will suffice to show that, if \(\delta_w(T_0^*: \mu_0, \mu_1)\) satisfies stage-wise error restrictions for \((\mu_0, \mu_1)\), then

\[
OF_1[\delta_w(T_0^*: \mu_0, \mu_1)] \leq OF_1[\delta_w(T_0^*: \mu_0, \mu_1)].
\]

where

\[
OF_1[\delta_w(T_0^*: \mu_0, \mu_1)] = T_0^* + \frac{1}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{\frac{u_*}{\sqrt{T_0^*}}}^u \phi(s | \mu T_0^*, \sqrt{T_0^*}) \cdot T_s \cdot ds \cdot d\mu.
\]

By transforming with \(s = c_1 s\) and \(\mu = \frac{\mu}{c_1}\), we have

\[
OF_1[\delta_w(T_0^*: \mu_0, \mu_1)] = \frac{1}{c_1^2} \left[ T_0^* + \frac{1}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{\frac{u_*}{\sqrt{T_0^*}}}^u \phi(s | \mu T_0^*, \sqrt{T_0^*}) \cdot T_s \cdot ds \cdot d\mu \right]
\]

\[
= \frac{1}{c_1^2} OF_1[\delta_w(T_0^*: \mu_0, \mu_1)]
\]

\[
\leq \frac{1}{c_1^2} OF_1[\delta_w(T_0^*: \mu_0, \mu_1)]
\]

\[
= \frac{1}{c_1^2} \left[ c_1^2 T_0 + \frac{c_1^2}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{c_1 \frac{u_*}{\sqrt{T_0}}}^{c_1 u} \phi(c_1 s | c_1 \mu T_0^*, \sqrt{c_1^2 T_0^*}) \cdot T_s \cdot ds \cdot d\mu \right]
\]

\[
= \frac{c_1^2}{c_1^2} \left[ T_0 + \frac{1}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{\mu_0}^{\mu_1} \phi(s | \mu T_0^*, \sqrt{T_0^*}) \cdot T_s \cdot ds \cdot d\mu \right]
\]

\[
= OF_1[\delta_w(T_0^*: \mu_0, \mu_1)]. \tag{3.18}
\]
where, in the third equality of (3.18), we use the fact established essentially as in Lemma 3.2 that, if \( T_0 \) satisfies stage-wise error restrictions are satisfied for \((\mu_0, \mu_1)\), then \( T_0 = c_1^2 T_0 \) satisfies the stage-wise error restrictions for \((\mu_0, \mu_1)\). \( \text{(Q.E.D.)} \)

Next, consider the case of a location transformation. Let \((\mu_0', \mu_1') = (\mu_0 + c_2, \mu_1 + c_2)\), \( c_2 \in (-\infty, +\infty) \), and define \( \delta_w(T_0^*; \mu_0', \mu_1') \) as

\[
(T_0^*, \lambda^*, u^*, \{ T_s^*; \lambda \in (\lambda^*, u^*) \}, \{ k_s^*; \lambda \in (\lambda^*, u^*) \})
\]

\[
= (T_0^*, \lambda^* + c_2 T_0^*, u^* + c_2 T_0^*, \{ T_s^*; \lambda \in (\lambda^*, u^*) \}, \{ k_s^* + c_2 T_s^*; \lambda \in (\lambda^*, u^*) \}) \quad (3.19)
\]

**Lemma 3.3** Let \((\mu_0', \mu_1') = (\mu_0 + c_2, \mu_1 + c_2)\), \( c_2 \in (-\infty, +\infty) \). Then, \( \delta_w(T_0^*; \mu_0', \mu_1') \) satisfies the stage-wise error restrictions.

**Proof:** Let \( \Phi(y) = \int_{-\infty}^{y} \phi(x \mid 0, 1) \, dx \).

Then, for the plan \( \delta_w(T_0^*; \mu_0, \mu_1) \), the first-stage error probabilities can be written as

\[
\alpha_1 = 1 - \Phi\left( \frac{\lambda_0^* - \mu_0 T_0^*}{\sqrt{T_0^*}} \right)
\]

and

\[
\beta_1 = \Phi\left( \frac{\lambda_1^* - \mu_1 T_0^*}{\sqrt{T_0^*}} \right).
\]

Also, the second-stage error probabilities can be written as

\[
\alpha_2 = \int_{\lambda^*}^{\lambda_0^*} \phi(s \mid \mu_0 T_0^* \sqrt{T_0^*}) \left[ 1 - \Phi\left( \frac{k_s^* - \mu_0 T_s^*}{\sqrt{T_s^*}} \right) \right] ds
\]
and

$$\beta_2 = \int_{I^*} \phi(s | \mu_1 T^*_0, \sqrt{T^*_0}) \Phi\left(\frac{k^*_s - \mu_1 T^*_s}{\sqrt{T^*_s}}\right) ds.$$ 

Then, the first-stage error probabilities for $\delta_w(T^*_0; \mu'_0, \mu'_1)$ are

$$1 - \Phi\left(\frac{u^*_0 - (\mu_0 + c_2)T^*_0}{\sqrt{T^*_0}}\right) = 1 - \Phi\left(\frac{u^*_0 + c_2 T^*_0 - (\mu_0 + c_2)T^*_0}{\sqrt{T^*_0}}\right)$$

$$= 1 - \Phi\left(\frac{u^*_0 - \mu_0 T^*_0}{\sqrt{T^*_0}}\right)$$

$$= \alpha_1$$

and

$$\Phi\left(\frac{l^*_0 - (\mu_1 + c_2)T^*_0}{\sqrt{T^*_0}}\right) = \Phi\left(\frac{u^*_0 + c_2 T^*_0 - (\mu_1 + c_2)T^*_0}{\sqrt{T^*_0}}\right)$$

$$= \Phi\left(\frac{l^*_0 - \mu_1 T^*_0}{\sqrt{T^*_0}}\right)$$

$$= \beta_1$$

by (3.19). And also, for the second-stage error probabilities, consider

$$1 - \Phi\left(\frac{k^*_s - (\mu_0 + c_2)T^*_s}{\sqrt{T^*_s}}\right)$$

$$= 1 - \Phi\left(\frac{(k^*_s + c_2 T^*_s) - (\mu_0 + c_2)T^*_s}{\sqrt{T^*_s}}\right)$$

$$= 1 - \Phi\left(\frac{k^*_s - \mu_0 T^*_s}{\sqrt{T^*_s}}\right), \quad \forall \ s \in (l^*, u^*)$$

(3.20)

and

$$\Phi\left(\frac{k^*_s - (\mu_1 + c_2)T^*_s}{\sqrt{T^*_s}}\right) = \Phi\left(\frac{(k^*_s + c_2 T^*_s) - (\mu_1 + c_2)T^*_s}{\sqrt{T^*_s}}\right)$$

$$= \Phi\left(\frac{k^*_s - \mu_1 T^*_s}{\sqrt{T^*_s}}\right), \quad \forall \ s \in (l^*, u^*)$$

(3.21)
Therefore, the second-stage error probabilities are

\[
\int_{\hat{T}^*_0}^{u^*} \phi(\hat{s} | \mu_0 + c_2 \hat{T}_0^*, \sqrt{\hat{T}_0^*}) [1 - \Phi\left(\frac{k_{\hat{s}}^* - \mu_0 \hat{T}_s^*}{\sqrt{\hat{T}_s^*}}\right)] d\hat{s}
\]

\[
= \int_{\hat{T}^*_0}^{u^*} \phi(\hat{s} | \mu_0 \hat{T}_0^*, \sqrt{\hat{T}_0^*}) [1 - \Phi\left(\frac{k_{\hat{s}}^* - \mu_0 \hat{T}_s^*}{\sqrt{\hat{T}_s^*}}\right)] d\hat{s}
\]

\[
= a_2
\]

and

\[
\int_{\hat{T}^*_0}^{u^*} \phi(\hat{s} | \mu_1 + c_2 \hat{T}_0^*, \sqrt{\hat{T}_0^*}) [1 - \Phi\left(\frac{k_{\hat{s}}^* - \mu_1 \hat{T}_s^*}{\sqrt{\hat{T}_s^*}}\right)] d\hat{s}
\]

\[
= \int_{\hat{T}^*_0}^{u^*} \phi(\hat{s} | \mu_1 \hat{T}_0^*, \sqrt{\hat{T}_0^*}) [1 - \Phi\left(\frac{k_{\hat{s}}^* - \mu_1 \hat{T}_s^*}{\sqrt{\hat{T}_s^*}}\right)] d\hat{s}
\]

\[
= b_2
\]

where, in the first equalities of (3.22) and (3.23), we transform \(\hat{s}\) to \(\hat{s}\) with \(\hat{s} = \hat{s} + c_2 T_0^*\)

Proposition 3.2 optimality under location transformation

\(\delta_w(\hat{T}_0^*; \mu_0 + c_2, \mu_1 + c_2)\) is optimal under the given stage-wise restrictions.

Proof: For given \(c_2, c_2 \in (-\infty, +\infty)\), let \((\mu_0', \mu_1') = (\mu_0 + c_2, \mu_1 + c_2)\). It will suffice to show that, if \(\delta_w(\hat{T}_0^*; \mu_0', \mu_1')\) satisfies stage-wise error restrictions for \((\mu_0', \mu_1')\), then

\[
OF_1[\delta_w(\hat{T}_0^*; \mu_0', \mu_1')] \leq OF_1[\delta_w(\hat{T}_0^*; \mu_0, \mu_1')].
\]

where

\[
OF_1[\delta_w(\hat{T}_0^*; \mu_0, \mu_1')] = \hat{T}_0^* + \frac{1}{(\mu_1 - \mu_0') \int_{\hat{T}^*_0}^{u^*} \phi(\hat{s} | \mu_1 \hat{T}_0^*, \sqrt{\hat{T}_0^*}) \cdot \hat{T}_s^* d\hat{s} d\mu}
\]
By transforming with $s = \dot{s} - c_2 T_0^*$ and $\mu = \dot{\mu} - c_2$, we have

$$OF_1 [\delta_w(T_0^*; \mu_0, \mu_1)]$$

$$= T_0^* + \frac{1}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{T_0^*}^{\mu T_0^*} \phi(s | \mu T_0^* \sqrt{T_0^*}) \cdot T_s ds \, d\mu$$

$$= OF_1 [\delta_w(T_0^*; \mu_0, \mu_1)]$$

$$\leq OF_1 [\delta_w(T_0; \mu_0, \mu_1)]$$

$$= T_0 + \frac{1}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{T_0}^{\mu T_0} \phi(s + c_2 T_0 | (\mu + c_2)T_0 \sqrt{T_0}) \cdot T_s ds \, d\mu$$

$$= T_0 + \frac{1}{(\mu_1 - \mu_0)} \int_{\mu_0}^{\mu_1} \int_{\hat{T}_0} T_s ds \, d\mu$$

$$= OF_1 [\delta_w(T_0^*; \mu_0, \mu_1)].$$

(3.24)

where, in the third equality of (3.24), we used the fact, established essentially as in Lemma 3.3, that if $T_0$ satisfies stage-wise error restrictions for $(\mu_0, \mu_1)$. (Q.E.D.)

Combining Proposition 3.1 and 3.2, we have following Theorem.

**Theorem 3.1** For given $(\mu_0, \mu_1)$, let $(\theta_0, \theta_1)$ be the hypothesis points such that

$$\theta_0 = c_1 \mu_0 + c_2$$

$$\theta_1 = c_1 \mu_1 + c_2.$$

Then, if $\delta_w(T_0^*; \mu_0, \mu_1)$ is optimal for the given stage-wise restrictions, then the optimal plan $\delta_w(T_0; \mu_0, \mu_1)$, $\theta_0 < \theta_1, \theta = (\theta_0, \theta_1)$ under the same stage-wise restrictions is given by
Proof: First, apply Proposition 3.1 with \( c_1 = \frac{\theta_1 - \theta_0}{\mu_1 - \mu_0} \). In other words, apply Proposition 3.1 with

\[
(\mu_0, \mu_1) = \left( \frac{\theta_1 - \theta_0}{\mu_1 - \mu_0}, \frac{\theta_1 - \theta_0}{\mu_1 - \mu_0} \cdot \frac{\mu_0}{\mu_1} \right)
\]  

(3.26)

and then apply Proposition 3.2 to (3.26) with \( c_2 = \frac{\mu_1 \theta_0 - \mu_0 \theta_1}{\mu_1 - \mu_0} \). Then, we obtain the result. (Q.E.D.)

### 3.4 Alternate plans

The main procedures alternative to flexible two-stage plan are single sampling plans (SS), standard two-stage plans (STP), and possibly sequential probability ratio test (SPRT) plans. Through the latter typically are difficult to implement in practice, which calls for developing group sequential approaches, specially in clinical trials.
3.4.1 Single Sampling Plan (SS)

The simplest sampling plan, single sampling, with error rates $\alpha$ and $\beta$ at hypothesis points $\mu_0$ and $\mu_1$, is determined as follows for the Wiener process:

$$Pr\{\phi(x | \mu_0 T, \sqrt{T}) \geq k\} = \alpha$$
$$Pr\{\phi(x | \mu_1 T, \sqrt{T}) < k\} = \beta$$

(3.27)

The $AST_\mu$ of SS is the constant value $T^*$ satisfying (3.27) and the $OC$ of SS is given by the function $Pr\{\phi(x | \mu T, \sqrt{T}) < k^*\}$ of $\mu$, where $k^*$ satisfies (3.27).

3.4.2 Standard Two-stage Plans (STP)

We define such plans by fixing second-stage sampling time and critical values, i.e., the plans OFP without flexibility of second-stage. A unique standard two-stage plan STP, with allocation of the two error rates $\alpha$ and $\beta$ to the two-stages, exists for all first-stage sampling time $T_0$ in $(0, \hat{T}_0]$, where $\hat{T}_0$ is a value for first-stage sampling time that happens to exceed $\hat{T}_0$, upper limit of first-stage sampling time for OFP.

It's kind of obvious when we consider restrictions of two plans. The $AST_\mu$ and $OC_\mu$ function of STP are determined as follows for the Wiener process:

$$AST_\mu = T_0 + \int_{\hat{T}_0}^{T_0} \phi(s | \mu T_0, \sqrt{T_0}) \cdot T_2 \cdot ds$$

(3.28)

and

$$OC_\mu = \int_{\hat{T}_0}^{T_0} \phi(x | \mu T_0, \sqrt{T_0}) dx +$$
$$\int_{\hat{T}_0}^{T_0} \phi(s | \mu T_0, \sqrt{T_0}) \cdot \left[ \int_{\hat{T}_0}^{T_2} \phi(x | \mu T_2, \sqrt{T_2}) dx \right] ds$$
where $T_2$ and $k_2$ satisfy

\[
\alpha = \int_u^\infty \phi(x \mid \mu_0T_0, \sqrt{T_0}) \, dx + \\
\int_l^u \phi(s \mid \mu_0T_0, \sqrt{T_0}) \cdot \left[ \int_{k_2}^\infty \phi(x \mid \mu_0T_2, \sqrt{T_2}) \, dx \right] \, ds
\]

\[
\beta = \int_{-\infty}^l \phi(x \mid \mu_1T_0, \sqrt{T_0}) \, dx + \\
\int_l^u \phi(s \mid \mu_1T_0, \sqrt{T_0}) \cdot \left[ \int_{-\infty}^{k_2} \phi(x \mid \mu_1T_2, \sqrt{T_2}) \, dx \right] \, ds.
\]

AST$_\mu$ is computed according to (3.28) for all first-stage sampling times $T_0$ in $(0, T]$ and further first-stage optimization over $(0, T]$ gives minimum $T_0^{**}$, also corresponding continuation region $(l^{**}, u^{**})$ and the second-stage parameters $(T_2^{**}, k_2^{**})$ to constitute STP$^*$.  

3.4.3 Sequential Probability Ratio Test (SPRT)

The SPRT to be compared with OFP$^*$ is one with error rates $\alpha$ and $\beta$ at hypothesis point $\mu_0$ and $\mu_1$. The standard SPRT theory for Wiener process gives the following formulas for OC and AST:

\[
OC_\mu = \begin{cases} 
\frac{A - 1}{A - B} & \text{for } \mu < \bar{\mu} = \frac{\mu_0 + \mu_1}{2}, \\
\frac{\log(A)}{\log(A) - \log(B)} & \text{for } \mu = \bar{\mu}, \\
\frac{1 - B}{1 - B} & \text{for } \mu > \bar{\mu}
\end{cases}
\]

where
\[ A = \frac{1-\frac{3}{2}}{\alpha} \]
\[ B = \frac{3}{1-\alpha} \]
\[ \hat{A} = A [2\mu - (\mu_0 + \mu_1)] \]
\[ \hat{B} = B [2\mu - (\mu_0 + \mu_1)] \]

and

\[ \text{AST}_\mu = \begin{cases} 
\frac{\log(A) - \log(1 - OC\mu) + \log(B) - OC\mu}{(\mu_1 - \mu_0)(\mu - \bar{\mu})} & \text{for } \mu \neq \bar{\mu}, \\
\frac{-\log(B)\cdot\log(A)}{(\mu_1 - \mu_0)^2} & \text{for } \mu = \bar{\mu} 
\end{cases} \]

3.5 Numerical illustrations and comparisons

By following the procedures of section 3.3, as given for \( OF_1 \), we can also find the optimal plans for the \( OF_2, OF_3 \) and \( OF_4 \) of section 3.2. We have considered so for the case \( (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_0, \mu_1) = (0.025, 0.025, 0.05, 0.05, 0.50) \).

For this case, sampling effort at optimum is about equally divided between the first- and second-stages. Moreover, for the cases \( OF_1, OF_2 \) and \( OF_4 \), \( T_s \) is maximum for \( s \) roughly half-way between \( l \) and \( u \), and drops to approximately half that maximum at \( s = l \) and \( s = u \); on other hand, for the case \( OF_3 \), \( T_s \) is smallest, as would expected, at \( s = u \), and also is largest at \( s = l \). The optimal second-stage sampling time and critical values are shown in Figure 3.4 and Figure 3.5, respectively.

Furthermore, the OC functions for our four optimal flexible plans are essentially
same, while the AST functions are essentially alike, except the AST function of OF3, which is smallest near $\mu_1$, and largest near $\mu_0$. These features are pictured in Figure 3.6 and Figure 3.7. Note here that, in Figure 3.7, the AST function for OF3 also dominates the AST function of $STP^*$ derived by using the analogue of OF3.

As already indicated in the Introduction, we also present, in Figure 3.8 and 3.9, the OC and AST functions for the alternative plans $SS$, $SPRT$ and optimal standard plan, as benchmarks for our optimal flexible plans. As shown in those figures, the OC functions for all plans are essentially same and the four AST functions are strictly ordered.

### 3.6 Applications to binomial (Bernoulli) case

Wiener approximations to binomial (or normal) responses frequently occur in the clinical trial literature. (Whitehead [33], Lan and DeMets [19], Jennison [17]). In previous sections, we introduced certain optimal plans for Wiener location problems. We consider here their use in binomial models, and discuss certain advantages of approximate Wiener models for deriving optimal binomial plans. In particular, we consider obtaining approximate optimal (in the sense of $OF_1$) flexible two-stage binomial plans by Wiener approximation.

#### 3.6.1 One-sample case

Assume that we are interest in a population proportion, $p$, and have $n$ Bernoulli samples from that population. In many cases, researchers are interest in testing $H_0 : p = p_0$ against $H_1 : p = p_1, p_0 < p_1$. In the two-stage framework, let $n_0$ be first-stage sample size and $(n_F, k_F)$ be second-stage sample size and critical value when first-stage
Figure 3.4: Optimal second-stage sampling efforts
Figure 3.5: Optimal second-stage critical values
Figure 3.6: $OC_{\mu}$ of $STP^*$ and $OFP^*$ for the Wiener process $w(0, 1)$ and $w(0.5, 1)$ with $(\alpha, 3) = (0.05, 0.10)$
Figure 3.7: \( AST_\nu \) of \( STP^* \) and \( OFP^* \) for the Wiener process \( w(0.1) \) and \( w(0.5, 1) \) with \( (\alpha, \beta) = (0.05, 0.10) \)
Figure 3.8: $OC_{\mu}$ of $SPRT.OPP^*$. $STP^*$ and $S.S.$ for the Wiener process $w(0,1)$ and $w(0.5,1)$ with $(\alpha, \beta) = (0.05, 0.10)$
Figure 3.9: $AST_h$ of $\text{SPRT. OFP}^*$. $\text{STP}^*$ and $\text{SS}$, for the Wiener process $w(0.1)$ and $w(0.5.1)$ with $(\alpha, \beta) = (0.05, 0.10)$
outcome, \( r \) falls into continuation region, \( \{l_B, l_B + 1, \ldots, u_B\} \). Define the flexible binomial two-stage plan, \( \delta_B(n_0; p_0, p_1) \equiv (n_0, l_B, u_B, \{n_r: r \in \{l_B, u_B\}\}, \{k_r: r \in \{l_B, u_B\}\}; p_0, p_1) \). We can simply relate such a plan to an approximating flexible Wiener two-stage plan \( \delta_w(T_{0, \theta}; \theta_0, \theta_1) \) with hypotheses \( H_0: \theta_0 = 2\sqrt{\frac{n_0}{T_{0, \theta}}} \sin^{-1}\sqrt{\frac{p_0}{p_1}} \) and \( H_1: \theta_1 = 2\sqrt{\frac{n_0}{T_{0, \theta}}} \sin^{-1}\sqrt{p_1} \) as in Table 3.1. Note here that the approximation will be better if \( n_0 \) and \( \{n_r\} \) are large.

Table 3.1: Approximate binomial plan corresponding Wiener plan

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>( H_i: \theta = \theta_i, i = 0, 1 )</th>
<th>( H_i: p = p_i, i = 0, 1 ) where ( p_i = (\sin\frac{\theta_i}{2\sqrt{n_0 T_{0, \theta}}})^2, i = 0, 1 )</th>
</tr>
</thead>
</table>
| First-stage sampling effort | \( T_{0, \theta} \) | \( n_0 = \left[ T_{0, \theta} \right] \)
| Continuation region | \( (l, u), l, u \in R \) | \( \{l_B, l_B + 1, \ldots, u_B\} \) where \( l_B = \lfloor l \rfloor, u_B = \lceil u \rceil \) \( l_c = n_0(\sin\frac{l}{2\sqrt{n_0 T_{0, \theta}}})^2 \) \( u_c = n_0(\sin\frac{u}{2\sqrt{n_0 T_{0, \theta}}})^2 \) \( 0 \leq l_B \leq u_B \leq n_0 \)
| Second-stage functions | \( T_s \) | \( n_r = \left[ n_{sr} \right] \)
| \( k_s, s \in [l, u] \) | \( k_r = \left[ n_{sr}(\sin\frac{k_{sr}}{2\sqrt{n_{sr} T_{sr}}})^2 \right] \) \( r \in \{l_B, l_B + 1, \ldots, u_B\} \) where \( n_{sr} = \frac{n_0 T_s}{T_{0, \theta}} \)

\(^a\) Determined by degree of approximations.
\(^b\) \([\cdot]\) denotes the nearest integer.
\(^c\) Given \( r. s_r = 2\sqrt{n_0 T_{0, \theta}} \sin^{-1}\sqrt{\frac{p_0}{p_1}} \).
In transformations relating approximating Wiener plans to binomial plans, parameters (i.e., \((\theta_0, \theta_1)\) for Wiener model and \((p_0, p_1)\) for binomial model) in each plan do depend on the other's sampling effort. This fact creates additional freedom of interpretation between both types of plans: for example, suppose we have a family of Wiener plans for a certain \((\theta_0, \theta_1)\) and varying \(T_{0,0}\). Then, that family of Wiener plans corresponds to a family of binomial plans with varying \((p_0, p_1)\) and fixed \(n_0\), or a family of binomial plans with fixed \((p_0, p_1)\) and varying \(n_0\). Further, if we fix the ratio of both sampling efforts, i.e., \(\frac{n_0}{T_{0,0}}\), we have a unique \((\theta_0, \theta_1)\) for each \((p_0, p_1)\). For that \((\theta_0, \theta_1)\), we can find \(\delta_w(T_{0,0}; \theta_0, \theta_1)\) by optimization as in section 3.3.1. The ratio \(\frac{n_0}{T_{0,0}}\) also makes possible to interpreted binomial sample size relative to Wiener sampling time in both stages. Therefore, without loss of generality, we fixed \(\frac{n_0}{T_{0,0}} = 1\) to set Wiener hypotheses \((\theta_0, \theta_1)\) and interpreted binomial sample sizes corresponding to Wiener sampling time.

Now, let \(\delta_B(n_0^*, p_0, p_1)\) be the binomial plan approximated by the optimal Wiener plan \(\delta_w(T_{0,0}^*; \theta_0, \theta_1)\) for given stage-wise error restrictions, \((\alpha_i, \beta_i), i = 1, 2\) through the transformations in Table 3.1 with \(T_{0,0} = T_{0,0}^*\) and \(n_0 = n_0^*\).

**Lemma 3.4** \(\delta_B(n_0^*, p_0, p_1)\) approximately satisfies the stage-wise error restrictions.

**Proof:** It is well known that arcsine transformation of a binomial proportion multiplied by the square root of sample size achieves approximate normality with a constant variance \(\frac{1}{4}\). Therefore, if we define \(s_r \equiv 2\sqrt{n_0^* T_{0,0}^*} \sin^{-1} \sqrt{\frac{r}{n_0^*}}, 0 \leq r \leq 1\).
we have the approximate relation

\[ Pr\{s_r | p_i^*\} \approx \int_{s_r-0.5}^{s_r+0.5} \phi(s_r | 2\sqrt{n_0^* T_{0.0}^* \sin^{-1} \sqrt{p_i^*}} \cdot \sqrt{T_{0.0}^*}) ds_r. \]

(3.29)

Now, since \( n_0^* \) is the nearest integer of \( T_{0.0}^* \), the first-stage error probabilities of \( \delta_B(n_0^*: p_0, p_1) \) are

\[
\sum_{r=u_B^*+1}^{n_0^*} b(r; n_0^*, p_0) \approx 1 - \Phi(\frac{(2\sin^{-1} \sqrt{\frac{u_B^*+1}{n_0^*}} - 0.5) - 2\sin^{-1} \sqrt{p_1}}{\sqrt{T_{0.0}^*}}) \]

(3.30)

and

\[
\sum_{r=0}^{l_B^*-1} b(r; n_0^*, p_1) \approx \Phi(\frac{(2\sin^{-1} \sqrt{\frac{t_B^*-1}{n_0^*}} + 0.5) - 2\sin^{-1} \sqrt{p_1}}{\sqrt{T_{0.0}^*}}) \]

(3.31)
where, both in (3.31) and (3.30), the first approximate equality is due to (3.29), the first equality is due to Table 3.1 with $T_{0,0} = T_{0,0}^*$ and $\mu_0 = \mu_0^*$, and the second approximate equality is due to the fact that $\delta_{\nu_{(T_{0,0}; \theta_0, \theta_1)}}$ satisfies the error restrictions. Similarly applying the same transformation to at the second-stage, we can approximate the second-stage error probabilities as

$$
\sum_{r=1}^{u_B^*} b(r; \mu_0^*, \mu_0) \sum_{r_B^*} \int_{\frac{L_s}{B}}^{s_B} \phi(s_r \mid 2\sin^{-1} \sqrt{\frac{T_{0,0}}{T_{s,0}}})
\left[ 1 - \Phi\left( \frac{(k_{s,r}^* - 0.5) - 2\sin^{-1} \sqrt{\frac{\mu_0 T_{s,r}}{T_{s,0}^*}}}{\sqrt{T_{s,r}^*}} \right) \right] ds_r
\approx \int_{\frac{l^*}{\theta}}^{\mu^*} \phi(s \mid \theta_0 T_{0,0}^*, \sqrt{T_{0,0}^*})\left[ 1 - \Phi\left( \frac{k_{s,0}^* - \theta_0^* T_{s,0}^*}{\sqrt{T_{s,0}^*}} \right) \right] ds
= \alpha_2
$$

\begin{align*}
\sum_{r=1}^{u_B^*} b(r; \nu_0^*, \nu_1) \sum_{r_B^*} \int_{\frac{L_s}{B}}^{s_B} & \phi(s_r \mid 2\sin^{-1} \sqrt{\frac{T_{0,0}}{T_{s,0}}})
\left[ 1 - \Phi\left( \frac{(k_{s,r}^* + 0.5) - 2\sin^{-1} \sqrt{\frac{\mu_1 T_{s,r}}{T_{s,0}^*}}}{\sqrt{T_{s,r}^*}} \right) \right] ds_r \\
\approx & \int_{\frac{l^*}{\theta}}^{\mu^*} \phi(s \mid \theta_0 T_{0,0}^*, \sqrt{T_{0,0}^*})\left[ 1 - \Phi\left( \frac{k_{s,0}^* - \theta_1^* T_{s,0}^*}{\sqrt{T_{s,0}^*}} \right) \right] ds
= \beta_2.
\end{align*}

(Q.E.D.)
Theorem 3.2 $\delta_B(n^*_0; p_0, p_1)$ is approximately optimal for the given stage-wise restrictions.

Proof: Recall $n_0^*$ in transformations of Table 3.1 is the nearest integer of $T_{0,0}^*$. It will suffice to show that if $\delta_B(n_0^*; p_0, p_1)$ approximately satisfies the stage-wise error restrictions, then

$$OF_1[\delta_B(n_0^*; p_0, p_1)] \leq OF_1[\delta_B(n_0^*; p_0, p_1)]$$

where

$$OF_1[\delta_B(n_0^*; p_0, p_1)] = n_0^* + \frac{1}{(p_1 - p_0)} \int_{p_0}^{p_1} \sum_{r=1}^{u} b(r | n_0^*, p) n_r dp$$

By transforming with $s_r = 2\sin^{-1}\sqrt{\frac{r}{n_0^*}}$ and $\theta_i = 2\sin^{-1}\sqrt{\nu_i}, i = 1, 2$ as in Lemma 3.4, we have

$$OF_1[\delta_B(n_0^*; p_0, p_1)]$$

$$\approx n_0^* + \frac{1}{(\theta_1 - \theta_0)} \int_{\theta_0}^{\theta_1} \int_{s_{\theta}^*}^{s_{\theta}^* + 0.5} \phi(s_r | \nu, n_0^*, \sqrt{n_0^*}) \cdot n_{s_r} ds_r d\theta$$

$$\approx \frac{n_0^*}{T_{0,0}^*} [T_{0,0}^* + \frac{1}{(\theta_1 - \theta_0)} \int_{\theta_0}^{\theta_1} \int_{s_{\theta}^*}^{s_{\theta}^* + 0.5} \phi(s | \nu T_{0,0}^*, \sqrt{T_{0,0}^*}) \cdot T_{s,\theta}^* ds d\theta]$$

$$\leq \frac{n_0^*}{T_{0,0}^*} [T_{0,0} + \frac{1}{(\theta_1 - \theta_0)} \int_{\theta_0}^{\theta_1} \int_{s_{\theta}^*}^{s_{\theta}^* + 0.5} \phi(s | \nu T_{0,0}, \sqrt{T_{0,0}}) \cdot T_{s,\theta} ds d\theta]$$

$$= \frac{n_0}{T_{0,0}^*} [T_{0,0} + \frac{1}{(\nu_1 - \nu_0)} \int_{\nu_0}^{\nu_1} \int_{\theta_0}^{\theta_1} \phi(s | \nu T_{0,0}, \sqrt{T_{0,0}}) \cdot T_{s,\theta} ds d\theta]$$

$$\approx n_0 + \frac{1}{(\theta_1 - \theta_0)} \int_{\theta_0}^{\theta_1} \int_{s_{\theta}^*}^{s_{\theta}^* + 0.5} \phi(s_r | \nu, n_0, \sqrt{n_0}) \cdot n_{s_r} ds_r d\theta$$

$$\approx n_0 + \frac{1}{(p_1 - p_0)} \int_{p_0}^{p_1} \sum_{r=1}^{u} b(r | n_0, p) n_r d\theta$$

$$= OF_1[\delta_B(n_0; p_0, p_1)] \quad (3.34)$$
where in the equality of (3.34), we used the fact that, for any $n_0$, there is $T_{0, \theta} = n_0 \frac{n^*_0}{T_{0, \theta}}$ which satisfies stage-wise error restrictions. If so, so does $n_0$ by Lemma 3.4. (Q.E.D.)

Combining Theorem 3.2 and Theorem 3.1, we can derive approximate optimal flexible binomial plans for any $(p_0, p_1)$ by one optimal flexible Wiener plan.

**Theorem 3.3** Suppose we have $\delta_w(T^*_0; \mu_0, \mu_1)$ with pre-specified restrictions $(\alpha_i, \beta_i)$. 

$$i = 1, 2 \text{ and define } \delta_B(n^*_0; p_0, p_1) \text{ as }$$

$$(n^*_0, t^*_B, u^*_B, [n^*_r, r \in \{l^*_B, u^*_B\}], \{k^*_r, r \in \{l^*_B, u^*_B\} \})$$

$$= (T_{0, \theta}^*, l_c, u_c, \{\frac{\frac{n^*_0 T^*_{0, \theta}}{T^*_0}}{T^*_0} \}, s \in \{ s^*(r) \}, \{k^*_s \}, s \in \{ s^*(r) \}, \{u^*_B \}).$$

where

$$s^*(r) = 2c_1 \sqrt{\frac{n^*_0 T^*_{0, \theta}}{T^*_0}} \sin^{-1} \left( \frac{r}{\sqrt{n^*_0}} - \frac{c_2 T^*_{0, \theta}}{c_1} \right)$$

$$T^*_{0, \theta} = \frac{T^*_0}{c_1^2}$$

$$l_c = n^*_0 \left( \frac{\frac{T^*_{0, \theta}}{c_1} + \frac{c_2 T^*_{0, \theta}}{c_1^2}}{2 \sqrt{n^*_0 T^*_{0, \theta}}} \right)^2$$

$$u_c = n^*_0 \left( \frac{\frac{T^*_{0, \theta}}{c_1} + \frac{c_2 T^*_{0, \theta}}{c_1^2}}{2 \sqrt{n^*_0 T^*_{0, \theta}}} \right)^2$$

$$T^*_{sr} = \frac{T^*_{sr}}{c_1^2}$$
where \( c_1 = \frac{2 \sin^{-1} \sqrt{p_1} - \sin^{-1} \sqrt{p_0}}{\mu_1 - \mu_0} \) and \( c_2 = \frac{2 \sin^{-1} \sqrt{p_0} \mu_1 - 2 \sin^{-1} \sqrt{p_1} \mu_0}{\mu_1 - \mu_0} \).

Then, for any \((p_0, p_1)\), \(\delta_B(n_0^*, p_0, p_1)\) is approximately optimal for the given stage-wise error restrictions.

Proof: For any given \((p_0, p_1) \in (0.1) \times (0.1)\), by applying Theorem 3.1 and Theorem 3.2 consequently, after transformation as in Lemma 3.4, we obtains result. (Q.E.D.)

Illustrations for Theorem 3.3 are made in next section which contains approximated optimal flexible binomial two-stage plans for various \(p_0\) and \(\Delta\), where \(\Delta = p_1 - p_0\).

### 3.6.2 Illustrations: one-sample case

In this section, we provide approximate optimal flexible binomial plans obtained by \(\delta_B(T_0^*: \mu_0, \mu_1)\) for the given \((\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.025, 0.025, 0.05, 0.05)\). We choose \((\mu_0, \mu_1) = (0.0.5)\) which has optimal \(T_0^* = 19.1\). Table 3.2 gives the approximate optimal flexible binomial plans for selected values of \(p_0\) in \((0.1)\) and \(\Delta = 0.1\), similarly. Table 3.3 for \(\Delta = 0.15\) and Table 3.4 for \(\Delta = 0.20\).
Table 3.2: The approximate optimal flexible binomial plan for $\Delta = 0.1$

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>0.45</th>
<th>0.50</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
<th>0.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0^{**}$</td>
<td>119</td>
<td>118</td>
<td>114</td>
<td>108</td>
<td>100</td>
<td>89</td>
<td>75</td>
<td>59</td>
</tr>
<tr>
<td>$(r, u, k_r)$$^b$</td>
<td>$R_1^c$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
</tr>
<tr>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
</tr>
<tr>
<td>89</td>
<td>102</td>
<td>121</td>
<td>132</td>
<td>137</td>
<td>138</td>
<td>134</td>
<td>122</td>
<td>107</td>
</tr>
<tr>
<td>49</td>
<td>53</td>
<td>64</td>
<td>68</td>
<td>70</td>
<td>69</td>
<td>70</td>
<td>58</td>
<td>60</td>
</tr>
<tr>
<td>65</td>
<td>66</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
</tr>
<tr>
<td>85</td>
<td>64</td>
<td>115</td>
<td>126</td>
<td>132</td>
<td>132</td>
<td>127</td>
<td>126</td>
<td>114</td>
</tr>
<tr>
<td>55</td>
<td>60</td>
<td>72</td>
<td>78</td>
<td>80</td>
<td>82</td>
<td>75</td>
<td>68</td>
<td>53</td>
</tr>
<tr>
<td>67</td>
<td>68</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>73</td>
</tr>
<tr>
<td>75</td>
<td>80</td>
<td>85</td>
<td>89</td>
<td>92</td>
<td>92</td>
<td>102</td>
<td>106</td>
<td>102</td>
</tr>
</tbody>
</table>

$^a$ approximate optimal first-stage sample size

$^b$ first-stage outcome, approximate optimal second-stage sample size and second-stage critical value, respectively

$^c$ reject hypothesis $p = p_1$

$^d$ reject hypothesis $p = p_0$

$^e$ observed $\alpha$ using exact binomial probabilities

$^f$ observed $\beta$ using exact binomial probabilities
Table 3.2  (Continued)

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
<th>0.35</th>
<th>0.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0^*$</td>
<td>40</td>
<td>59</td>
<td>75</td>
<td>89</td>
<td>100</td>
<td>108</td>
<td>114</td>
<td>118</td>
</tr>
<tr>
<td>$(r, n_r, k_r)^b$</td>
<td>$R_1^c$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
</tr>
<tr>
<td>2 30 3</td>
<td>7 44 8</td>
<td>13 56 13</td>
<td>19 66 19</td>
<td>27 75 26</td>
<td>35 80 31</td>
<td>42 85 37</td>
<td>50 88 43</td>
<td></td>
</tr>
<tr>
<td>3 35 4</td>
<td>8 61 9</td>
<td>14 77 16</td>
<td>20 70 20</td>
<td>28 93 31</td>
<td>36 104 38</td>
<td>43 91 40</td>
<td>51 112 53</td>
<td></td>
</tr>
<tr>
<td>4 46 4</td>
<td>9 68 9</td>
<td>15 86 17</td>
<td>21 92 24</td>
<td>29 108 34</td>
<td>37 117 42</td>
<td>44 114 48</td>
<td>52 126 58</td>
<td></td>
</tr>
<tr>
<td>5 42 2</td>
<td>10 66 8</td>
<td>16 86 16</td>
<td>22 101 25</td>
<td>30 115 35</td>
<td>38 124 43</td>
<td>45 126 51</td>
<td>53 134 61</td>
<td></td>
</tr>
<tr>
<td>6 35 1</td>
<td>11 55 6</td>
<td>17 79 14</td>
<td>23 103 25</td>
<td>31 115 34</td>
<td>39 125 43</td>
<td>46 131 53</td>
<td>54 137 61</td>
<td></td>
</tr>
<tr>
<td>$R_0^d$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td></td>
</tr>
<tr>
<td>18 65 10</td>
<td>24 99 23</td>
<td>32 111 32</td>
<td>40 121 40</td>
<td>47 132 52</td>
<td>55 135 59</td>
<td>50 103 37</td>
<td>58 102 42</td>
<td></td>
</tr>
<tr>
<td>25 88 19</td>
<td>33 100 28</td>
<td>41 111 36</td>
<td>48 128 49</td>
<td>56 127 54</td>
<td>57 115 48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 77 16</td>
<td>34 87 23</td>
<td>42 95 29</td>
<td>49 117 44</td>
<td>57 115 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$approximate optimal first-stage sample size
$^b$first-stage outcome, approximate optimal second-stage sample size and second-stage critical value, respectively
$^c$reject hypothesis $p = p_1$
$^d$reject hypothesis $p = p_0$
$^e$observed $\alpha$ using exact binomial probabilities
$^f$observed $\beta$ using exact binomial probabilities
Table 3.3: The approximate optimal flexible binomial plan for \( \Delta = 0.15 \)

<table>
<thead>
<tr>
<th>( p_0 )</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
<th>0.35</th>
<th>0.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n^*_0 )</td>
<td>21</td>
<td>29</td>
<td>36</td>
<td>42</td>
<td>46</td>
<td>49</td>
<td>51</td>
<td>53</td>
</tr>
<tr>
<td>( (r_1, n_1, k_r) )</td>
<td>R_1</td>
<td>R_1</td>
<td>R_1</td>
<td>R_1</td>
<td>R_1</td>
<td>R_1</td>
<td>R_1</td>
<td>R_1</td>
</tr>
<tr>
<td>1 16 2</td>
<td>3 22 4</td>
<td>6 27 7</td>
<td>9 31 11</td>
<td>13 34 13</td>
<td>16 37 16</td>
<td>19 38 18</td>
<td>22 39 21</td>
<td>23 39 21</td>
</tr>
<tr>
<td>2 21 2</td>
<td>4 24 5</td>
<td>7 34 8</td>
<td>10 36 12</td>
<td>14 47 16</td>
<td>17 45 18</td>
<td>20 40 19</td>
<td>23 39 21</td>
<td>25 60 29</td>
</tr>
<tr>
<td>3 24 2</td>
<td>5 33 5</td>
<td>8 19 9</td>
<td>11 46 14</td>
<td>15 53 17</td>
<td>18 54 21</td>
<td>21 54 24</td>
<td>24 54 28</td>
<td>26 61 29</td>
</tr>
<tr>
<td>4 18 0</td>
<td>6 32 4</td>
<td>9 41 8</td>
<td>12 49 13</td>
<td>16 53 16</td>
<td>19 57 21</td>
<td>22 59 25</td>
<td>25 60 29</td>
<td>27 57 26</td>
</tr>
<tr>
<td></td>
<td>7 25 2</td>
<td>10 33 5</td>
<td>13 45 11</td>
<td>17 47 13</td>
<td>20 55 19</td>
<td>23 59 24</td>
<td>26 61 29</td>
<td>28 48 21</td>
</tr>
<tr>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
<td>R_0</td>
</tr>
<tr>
<td>( \hat{\alpha} )</td>
<td>0.05331</td>
<td>0.05170</td>
<td>0.04913</td>
<td>0.03547</td>
<td>0.04372</td>
<td>0.04921</td>
<td>0.04789</td>
<td>0.04747</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>0.07373</td>
<td>0.08484</td>
<td>0.08334</td>
<td>0.10619</td>
<td>0.09403</td>
<td>0.08953</td>
<td>0.09053</td>
<td>0.09226</td>
</tr>
</tbody>
</table>

\( a \) approximate optimal first-stage sample size
\( b \) first-stage outcome, approximate optimal second-stage sample size and second-stage critical value, respectively
\( c \) reject hypothesis \( p = p_1 \)
\( d \) reject hypothesis \( p = p_0 \)
\( e \) observed \( \alpha \) using exact binomial probabilities
\( f \) observed \( \beta \) using exact binomial probabilities
Table 3.3  (Continued)

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>0.45</th>
<th>0.50</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
<th>0.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{0}^{a}$</td>
<td>53</td>
<td>51</td>
<td>49</td>
<td>46</td>
<td>42</td>
<td>36</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>$(r, n_r, k_r)^b$</td>
<td>$R_1^c$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
<td>$R_1$</td>
</tr>
<tr>
<td>26 46 27</td>
<td>28 45 27</td>
<td>29 37 25</td>
<td>30 44 32</td>
<td>29 35 27</td>
<td>27 32 27</td>
<td>24 32 27</td>
<td>18 19 18</td>
<td></td>
</tr>
<tr>
<td>27 57 31</td>
<td>29 55 33</td>
<td>30 51 33</td>
<td>31 52 36</td>
<td>30 46 34</td>
<td>28 41 33</td>
<td>25 33 27</td>
<td>19 24 21</td>
<td></td>
</tr>
<tr>
<td>28 61 33</td>
<td>30 59 34</td>
<td>31 53 35</td>
<td>32 53 36</td>
<td>31 49 36</td>
<td>29 41 32</td>
<td>26 25 20</td>
<td>20 18 15</td>
<td></td>
</tr>
<tr>
<td>29 61 31</td>
<td>31 58 32</td>
<td>32 54 34</td>
<td>33 49 32</td>
<td>32 45 32</td>
<td>30 33 24</td>
<td>26 25 20</td>
<td>20 18 15</td>
<td></td>
</tr>
<tr>
<td>30 55 27</td>
<td>32 51 27</td>
<td>33 51 30</td>
<td>34 40 25</td>
<td>33 36 25</td>
<td>30 33 24</td>
<td>26 25 20</td>
<td>20 18 15</td>
<td></td>
</tr>
<tr>
<td>31 46 22</td>
<td>33 44 23</td>
<td>34 43 24</td>
<td>$R_0^d$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
</tr>
<tr>
<td>$\hat{\alpha}^c$</td>
<td>0.04704</td>
<td>0.04659</td>
<td>0.05226</td>
<td>0.04222</td>
<td>0.04994</td>
<td>0.04618</td>
<td>0.03982</td>
<td>0.04506</td>
</tr>
<tr>
<td>$\hat{\beta}^f$</td>
<td>0.08230</td>
<td>0.09322</td>
<td>0.08238</td>
<td>0.10068</td>
<td>0.08627</td>
<td>0.10012</td>
<td>0.11070</td>
<td>0.09796</td>
</tr>
</tbody>
</table>

*a approximate optimal first-stage sample size
*b first-stage outcome, approximate optimal second-stage sample size and second-stage critical value, respectively
*c reject hypothesis $p = p_1$
*d reject hypothesis $p = p_0$
'e observed $\alpha$ using exact binomial probabilities
*f observed $\beta$ using exact binomial probabilities
Table 3.4: The approximate optimal flexible binomial plan for $\Delta = 0.2$

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
<th>0.35</th>
<th>0.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0^*$</td>
<td>13</td>
<td>18</td>
<td>22</td>
<td>24</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>$(r, n, k, r)^b$</td>
<td>$R_t^c$</td>
<td>$R_t$</td>
<td>$R_t$</td>
<td>$R_t$</td>
<td>$R_t$</td>
<td>$R_t$</td>
<td>$R_t$</td>
<td>$R_t$</td>
</tr>
<tr>
<td>1 10 1</td>
<td>2 13 3</td>
<td>4 16 5</td>
<td>5 18 6</td>
<td>7 20 9</td>
<td>9 21 10</td>
<td>11 22 11</td>
<td>13 22 12</td>
<td></td>
</tr>
<tr>
<td>2 15 1</td>
<td>3 19 3</td>
<td>5 24 6</td>
<td>6 18 6</td>
<td>8 21 9</td>
<td>10 24 11</td>
<td>12 27 13</td>
<td>14 30 16</td>
<td></td>
</tr>
<tr>
<td>3 11 0</td>
<td>4 21 3</td>
<td>6 25 5</td>
<td>7 27 8</td>
<td>9 30 11</td>
<td>11 31 13</td>
<td>13 33 15</td>
<td>15 34 16</td>
<td></td>
</tr>
<tr>
<td>5 16 1</td>
<td>7 19 2</td>
<td>8 27 7</td>
<td>10 31 11</td>
<td>12 32 13</td>
<td>14 33 14</td>
<td>16 32 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_0^d$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td>$R_0$</td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}^e$</td>
<td>0.06155</td>
<td>0.05663</td>
<td>0.04769</td>
<td>0.05341</td>
<td>0.04263</td>
<td>0.04269</td>
<td>0.05494</td>
<td>0.04970</td>
</tr>
<tr>
<td>$\hat{\beta}^f$</td>
<td>0.07602</td>
<td>0.06703</td>
<td>0.08123</td>
<td>0.08364</td>
<td>0.09663</td>
<td>0.10441</td>
<td>0.07954</td>
<td>0.09123</td>
</tr>
</tbody>
</table>

*a* approximate optimal first-stage sample size

*b* first-stage outcome, approximate optimal second-stage sample size and second-stage critical value, respectively

*c* reject hypothesis $p = p_1$

*d* reject hypothesis $p = p_0$

*e* observed $\alpha$ using exact binomial probabilities

*f* observed $\beta$ using exact binomial probabilities
Table 3.4  (Continued)

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>0.45</th>
<th>0.50</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0^a$</td>
<td>29</td>
<td>28</td>
<td>27</td>
<td>24</td>
<td>22</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>$(r, n_r, k_r)^b$</td>
<td>$\hat{R}_1$</td>
<td>$\hat{R}_1$</td>
<td>$\hat{R}_1$</td>
<td>$\hat{R}_1$</td>
<td>$\hat{R}_1$</td>
<td>$\hat{R}_1$</td>
<td>$\hat{R}_1$</td>
</tr>
<tr>
<td>14</td>
<td>22</td>
<td>14</td>
<td>15</td>
<td>21</td>
<td>15</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>15</td>
<td>27</td>
<td>17</td>
<td>16</td>
<td>27</td>
<td>18</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>33</td>
<td>19</td>
<td>17</td>
<td>32</td>
<td>20</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td>17</td>
<td>33</td>
<td>18</td>
<td>18</td>
<td>32</td>
<td>19</td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td>18</td>
<td>28</td>
<td>14</td>
<td>19</td>
<td>26</td>
<td>14</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>$\hat{R}_0^d$</td>
<td>$\hat{R}_0$</td>
<td>$\hat{R}_0$</td>
<td>$\hat{R}_0$</td>
<td>$\hat{R}_0$</td>
<td>$\hat{R}_0$</td>
<td>$\hat{R}_0$</td>
<td>$\hat{R}_0$</td>
</tr>
<tr>
<td>0.04236</td>
<td>0.03950</td>
<td>0.04181</td>
<td>0.06427</td>
<td>0.05297</td>
<td>0.04643</td>
<td>0.05321</td>
<td></td>
</tr>
<tr>
<td>0.10521</td>
<td>0.10781</td>
<td>0.09647</td>
<td>0.07432</td>
<td>0.07952</td>
<td>0.08622</td>
<td>0.11940</td>
<td></td>
</tr>
</tbody>
</table>

*a* approximate optimal first-stage sample size

b*first-stage outcome, approximate optimal second stage sample size and second-stage critical value, respectively

*reject hypothesis $p = p_0$

*reject hypothesis $p = p_0$

*observed $r$ using exact binomial probabilities

*observed $r$ using exact binomial probabilities
For all three cases, maximum values of the optimal first-stage sample size \( n_0^* \) occur near \( p_0 = 0.5 \) and smaller values of \( n_0^* \) occur near \( p_0 = 0 \) and \( p_0 = 1 \). The upper and lower end points of the continuation regions roughly reflect the two hypothesis points, and the second-stage sample sizes are very similar to those for the first stage, no doubt reflecting the equal error rate allocation to the two stages. Finally, as expected, optimal sample sizes are reversely related to \( \Delta \). The Tables also contain \( \alpha \) and \( \beta \) rates computed by exact binomial probabilities. Here we observe, again as expected, that large sample size lead to smaller discrepancies between target error rates, \((0.05,0.10)\), and the corresponding exact error rates.

### 3.6.3 Two-sample case

Suppose we are interested in testing \( H_0 : (p_C, p_T) = (p_0, p_0) \) against \( H_1 : (p_C, p_T) = (p_0, p_1), p_0 < p_1 \). The difference of arcsine transformation between \( p_T \) and \( p_C \) is well defined because of normality. Therefore, we consider testing \( H_0 : \theta_0 = 0 \) against \( H_1 : \theta_1 = \sqrt{2}(\sin^{-1}\sqrt{\frac{n_0}{T_0}} - \sin^{-1}\sqrt{\frac{n_0}{T_0}}) \) to derive optimal flexible binomial two-stage plan. The procedures are essentially same as one-sample case in section 3.5.1 with \( \frac{n_0}{T_{0,\theta}} = 1 \).

Assume we have same numbers of Bernoulli trials for control and treatment and \( \delta_{n}(T_{0,\theta}; \theta_0, \theta_1) \) for pre-specified error restrictions \((\alpha_i, \beta_i), i = 1, 2\). Then, for the given \((p_0, p_1)\), define \( \delta_{TB}(n_0^*, l_T^*, u_T^*, \{r \in (l_T^*, u_T^*)\}) \) and

\[
(n_0^*, l_T^*, u_T^*, \{r \in (l_T^*, u_T^*)\}), \{k_r^*, r \in (l_T^*, u_T^*)\})
\]

\[
= ([T_{0,\theta}^*, l_\theta^*, u_\theta^*, \{\frac{n_0^* T_{0,\theta}^*}{T_{0,\theta}}}, s_r \in (l_\theta^*, u_\theta^*)}, \{k_r^*, s_r \in (l_\theta^*, u_\theta^*)\}) (3.35)
\]
where \( s_r = \sqrt{2n_0^* T_{0,0}^*} \left( \sin^{-1} \left( \frac{d_T}{n_0^*} \right) - \sin^{-1} \left( \frac{d_C}{n_0^*} \right) \right) \) for given \( r = (d_T, d_C). \)

**Lemma 3.5** \( \delta_{TB}(n^*_0; p_0, p_1) \) approximately satisfies stage-wise restrictions.

**Proof:** Recall \( n_0^* \) in (3.35) is the nearest integer of \( T_0^* \). By the facts that normality of the difference of normal responses and arcsine and scale transformation for both proportions, \( \frac{d_C}{n_0^*} \) and \( \frac{d_T}{n_0^*} \), as in Lemma 3.4. we have approximate probability

\[
\Pr \{ s_r \mid p_0, p_1 \} 
\approx \int_{s_r-0.5}^{s_r+0.5} \sigma(s_r) \sqrt{2(\sin^{-1} \sqrt{p_1} - \sin^{-1} \sqrt{p_0})} \sqrt{\frac{T_{0,0}^*}{n_0^*}} ds_r \quad (3.36)
\]

Therefore, the first-stage error probabilities of \( \delta_{TB}(n^*_0; p_0, p_1) \) are approximately

\[
1 - \Phi \left( \frac{u^*_{TB}}{\sqrt{n_0^*}} \right) \approx 1 - \Phi \left( \frac{u^*_0}{\sqrt{T_{0,0}^*}} \right) \]

\[
= \alpha_1 \quad (3.37)
\]

and

\[
\Phi \left( \frac{l_{TB}^* - \sqrt{2(\sin^{-1} \sqrt{p_1} - \sin^{-1} \sqrt{p_0})}}{n_0^*} \right) 
\approx \Phi \left( \frac{l_0^* - \sqrt{2(\sin^{-1} \sqrt{p_1} - \sin^{-1} \sqrt{p_0})}}{\sqrt{T_{0,0}^*}} \right) 
\approx \Phi \left( \frac{l_0^* - \theta_1 T_{0,0}^*}{\sqrt{T_{0,0}^*}} \right) 
= 3_1 \quad (3.38)
\]
and the second-stage error probabilities are approximately

\[
\int_{T_B}^{u_T B} \phi(s_r | 0, \sqrt{n_0^*}) [1 - \Phi\left(\frac{k_{sr}^*}{\sqrt{n_{sr}^*}}\right)] ds_r \\
\approx \int_{0}^{u_T} \phi(s | 0, \sqrt{T_{0,0}^*}) [1 - \Phi\left(\frac{k_{s,0}^*}{\sqrt{T_{s,0}^*}}\right)] ds \\
= \alpha_2
\]

(3.39)

and with

\[
F(s_r, p_0, p_1) = \Phi\left(\frac{k_{sr}^* - \sqrt{2} (sin^{-1} \frac{1}{\sqrt{p_1}} - sin^{-1} \frac{1}{\sqrt{p_1}}) n_{sr}^*}{\sqrt{n_{sr}^*}}\right).
\]

\[
\int_{T_B}^{u_T B} \phi(s_r | \sqrt{2} (sin^{-1} \frac{1}{\sqrt{p_1}} - sin^{-1} \frac{1}{\sqrt{p_1}}) n_{sr}^* \sqrt{n_0^*}) [F(s_r, p_0, p_1)] ds_r \\
\approx \int_{0}^{u_T} \phi(s | \theta_1 T_{0,0}^*, \sqrt{T_{0,0}^*}) [\Phi\left(\frac{k_{s,0}^* - \theta_1 T_{s,0}^*}{\sqrt{T_{s,0}^*}}\right)] ds \\
= \beta_2
\]

(3.40)

By (3.37), (3.38), (3.39) and (3.40), we can conclude that \( \delta_T (u^*; p_0, p_1) \) approximately satisfies stage-wise error restrictions. (Q.E.D.)

Since we defined \( OF_1 \) for one-sample case in section 3.2, define

\[
OF_1 [\delta_T(n_0; p_0, p_1)] \\
\equiv 2 \left[ n_0 + \frac{1}{(\theta_1 - \theta_0)} \int_{\theta_0}^{\theta_1} \int_{T_B}^{u_T B} \phi(s_r | \theta, n_0, \sqrt{n_0}) \cdot n_r ds_r d\theta \right]
\]
where $\theta_0 = 0$ and

$$\theta_1 = \sqrt{2}(\sin^{-1}\sqrt{\frac{n_0}{T_{0,\theta}}} - \sin^{-1}\sqrt{\frac{n_0}{T_{0,\theta}}} \cdot \frac{1}{T_{0,\theta}}).$$

**Theorem 3.4** $\delta_{TB}(n_0^*; p_0, p_1)$ is approximately optimal for the given stage-wise restrictions.

**Proof:** Recall $n_0^*$ in (3.35) is the nearest integer of $T_{0,\theta}$. It will suffice to show that, if $\delta_{TB}(n_0^*; p_0, p_1)$ approximately satisfies the stage-wise error restrictions, then

$$OF_1[\delta_{TB}(n_0^*; p_0, p_1)] \leq OF_1[\delta_{TB}(n_0; p_0, p_1)]$$

As mentioned in previous lemma, we can approximate

$$OF_1[\delta_{TB}(n_0^*; p_0, p_1)]$$

$$\approx 2\left(\frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s_r | \sqrt{n_0^*}, n_r) ds_r \, dv \right)$$

$$\approx \frac{2n_0^*}{T_{0,\theta}} \left[ \frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s | \sqrt{T_{0,\theta}}) \cdot T_{s,\theta} ds \, dv \right]$$

$$\leq \frac{2n_0}{T_{0,\theta}} \left[ \frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s | \sqrt{T_{0,\theta}}) \cdot T_{s,\theta} ds \, dv \right]$$

$$= \frac{2n_0}{T_{0,\theta}} \left[ \frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s | \sqrt{T_{0,\theta}}) \cdot T_{s,\theta} ds \, dv \right]$$

$$\approx 2\left(\frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s_r | \sqrt{n_0}, n_r) ds_r \, dv \right)$$

$$\approx \frac{2n_0}{T_{0,\theta}} \left[ \frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s_r | \sqrt{n_0}, n_r) ds_r \, dv \right]$$

$$\approx \frac{2n_0}{T_{0,\theta}} \left[ \frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} \int_{T_{0,\theta}}^{uT_{0,\theta}} \phi(s_r | \sqrt{n_0}, n_r) ds_r \, dv \right]$$

$$\approx OF_1[\delta_{TB}(n_0; p_0, p_1)]$$

(3.11)

where in the equality of (3.41), we used the fact that, for any $n_0$, there is $T_{0,\theta} = \sqrt{n_0}$ which satisfies the stage-wise error restrictions. Therefore, $n_0$ also satisfies...
the stage-wise error restrictions by Lemma 3.5.

(Q.E.D.)

**Theorem 3.5** Suppose we have $\delta_w(T_0^*: \mu_0, \mu_1)$ with pre-specified restrictions $(\alpha_i, \beta_i)$, $i = 1, 2$ and define $\delta_B(n_0^*; p_0, p_1)$ as

$$(n_0^*, l_T^*, u_T^*, \{n_r^*, r \in (l_T^*, u_T^*)\}, \{k_r^*, r \in (l_T^*, u_T^*)\})$$

$$= ([T_0^*, l_T^*], u_T^*, \left\{ \left[ \frac{n_0^* l_T^*}{T_0^* \theta} \right], s \in (l^*, u^*) \right\}, \left\{ k^*_s, s \in (l^*, u^*) \right\})$$

where

$$T_{0, \theta}^* = \frac{T_0^*}{c_1^2}$$

$$l_T^* = \frac{l^*}{c_1} + \frac{c_2 T_0^*}{c_1^2}$$

$$u_T^* = \frac{u^*}{c_1} + \frac{c_2 T_0^*}{c_1^2}$$

$$T_{s}^* = \frac{T_s^*}{c_1^2}$$

$$k_{s}^* = \frac{k_s^*}{c_1} + \frac{c_2 T_s^*}{c_1^2}$$

with $c_1 = \sqrt{2} \left( \sin^{-1} \sqrt{\frac{1}{\mu_1}} - \sin^{-1} \sqrt{\frac{1}{\mu_0}} \right)$ and $c_2 = -\sqrt{2} \mu_0 \left( \sin^{-1} \sqrt{\frac{1}{\mu_1}} - \sin^{-1} \sqrt{\frac{1}{\mu_0}} \right)$.

Then, for any $p_0$ and $p_1$, $\delta_B(n_0^*; p_0, p_1)$ is approximately optimal for given stage-wise error restrictions.

**Proof:** For given $p_0$ and $p_1$, apply Theorem 3.4. Then we have approximate optimal plan with $\theta_0 = 0$ and $\theta_1 = \sqrt{2} \left( \sin^{-1} \sqrt{\frac{1}{\mu_1}} - \sin^{-1} \sqrt{\frac{1}{\mu_0}} \right)$. Now, apply Theorem
3.1 with given $c_1$ and $c_2$ to $\delta_w(T^*_0; \mu_0, \mu_1)$. (Q.E.D.)

Illustrations for Theorem 3.5 are made in next section which contains approximated optimal flexible binomial two-stage plans for various $p_0$ and $p_1$.

3.6.4 Illustrations: two-sample case

In this section, we provide approximate optimal flexible binomial plans for two sample case obtained by $\delta_w(T^*_0; \mu_0, \mu_1)$ for the given $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.025, 0.025, 0.05, 0.05)$. We choose same $(\mu_0, \mu_1)$ used in section 3.6.2. Table 3.5 shows approximate optimal first-stage sample size for various $p_0$ and $p_1$ values. For fixed $p_0$, testing the closer $p_1$ has the larger approximate optimal sample sizes. If $p_1 - p_0$ is extreme, e.g., $(p_0, p_1) = (0.1, 0.8)$, the optimal first-stage sample size is too small, therefore, approximation should be poor for such cases. However, in practical point of view, we are not interest in such extreme case. For the all $p_0$ we considered, the optimal first-stage sample sizes are reasonably large when $p_1 - p_0$ is less than 0.2.

The largest sample size occurs at $p_0 = 0.5$. It is expected because we use the optimal $\delta_w(T^*_0; \mu_0, \mu_1)$ obtained by $OF_1$. The second-stage sample sizes and critical values, for selected $(p_0, p_1)$, as indicated in Table 3.5, are shown in the Table 3.6. Table 3.7. Table 3.8 and Table 3.9. For all cases, the second-stage sample sizes are similar to those of first-stage, respectively, and the large difference of first-stage evidence leads the wide rejection regions and the maximum sample size occurs in the middle of continuation region as observed in Wiener two-stage plans.
Table 3.5: Approximate optimal first-stage sample size \( n_0^* \) for two-sample case

<table>
<thead>
<tr>
<th>( p_1 \backslash p_0 )</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
<th>0.50</th>
<th>0.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15</td>
<td>414</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.20</td>
<td>119</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.25</td>
<td>69</td>
<td>665</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.30</td>
<td>36</td>
<td>178(^a)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.35</td>
<td>25</td>
<td>83(^b)</td>
<td>838</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.40</td>
<td>18</td>
<td>49(^c)</td>
<td>216</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.45</td>
<td>14</td>
<td>32</td>
<td>99</td>
<td>934</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.50</td>
<td>11</td>
<td>23</td>
<td>56</td>
<td>236</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.55</td>
<td>9</td>
<td>17</td>
<td>37</td>
<td>105</td>
<td>953(^d)</td>
<td>-</td>
</tr>
<tr>
<td>0.60</td>
<td>8</td>
<td>13</td>
<td>25</td>
<td>59</td>
<td>236</td>
<td>-</td>
</tr>
<tr>
<td>0.65</td>
<td>6</td>
<td>11</td>
<td>19</td>
<td>37</td>
<td>103</td>
<td>895</td>
</tr>
<tr>
<td>0.70</td>
<td>5</td>
<td>9</td>
<td>14</td>
<td>25</td>
<td>56</td>
<td>216</td>
</tr>
<tr>
<td>0.75</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>18</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>0.80</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>13</td>
<td>23</td>
<td>49</td>
</tr>
</tbody>
</table>

\(^a\) For the second-stage, see Table 3.6
\(^b\) For the second-stage, see Table 3.7
\(^c\) For the second-stage, see Table 3.8
\(^d\) For the second-stage, see Table 3.9
Table 3.6: Approximate optimal flexible binomial two-stage plans for two-sample case: $(p_0, p_1) = (0.2, 0.3)$

$((a_1, a_2, \hat{\beta}_1, \hat{\beta}_2) = (0.02512, 0.02506, 0.01972, 0.04957)$

<table>
<thead>
<tr>
<th>$T(d_T,dC)^a$</th>
<th>$n^* T(d_T,dC)^b$</th>
<th>$R^* T(d_T,dC)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.02870, 0.02987)</td>
<td>133</td>
<td>(0.11448, 1.57080)</td>
</tr>
<tr>
<td>(0.02987, 0.03105)</td>
<td>139</td>
<td>(0.11054, 1.57080)</td>
</tr>
<tr>
<td>(0.03105, 0.03222)</td>
<td>145</td>
<td>(0.10692, 1.57080)</td>
</tr>
<tr>
<td>(0.03222, 0.03340)</td>
<td>150</td>
<td>(0.10390, 1.57080)</td>
</tr>
<tr>
<td>(0.03340, 0.03457)</td>
<td>155</td>
<td>(0.10106, 1.57080)</td>
</tr>
<tr>
<td>(0.03457, 0.03575)</td>
<td>159</td>
<td>(0.09868, 1.57080)</td>
</tr>
<tr>
<td>(0.03575, 0.03692)</td>
<td>163</td>
<td>(0.09640, 1.57080)</td>
</tr>
<tr>
<td>(0.03692, 0.03810)</td>
<td>167</td>
<td>(0.09421, 1.57080)</td>
</tr>
<tr>
<td>(0.03810, 0.03927)</td>
<td>171</td>
<td>(0.09211, 1.57080)</td>
</tr>
<tr>
<td>(0.03927, 0.04044)</td>
<td>174</td>
<td>(0.09003, 1.57080)</td>
</tr>
<tr>
<td>(0.04044, 0.04162)</td>
<td>177</td>
<td>(0.08806, 1.57080)</td>
</tr>
<tr>
<td>(0.04162, 0.04279)</td>
<td>180</td>
<td>(0.08609, 1.57080)</td>
</tr>
<tr>
<td>(0.04279, 0.04397)</td>
<td>182</td>
<td>(0.08500, 1.57080)</td>
</tr>
<tr>
<td>(0.04397, 0.04514)</td>
<td>185</td>
<td>(0.08387, 1.57080)</td>
</tr>
<tr>
<td>(0.04514, 0.04632)</td>
<td>187</td>
<td>(0.08251, 1.57080)</td>
</tr>
<tr>
<td>(0.04632, 0.04759)</td>
<td>189</td>
<td>(0.08113, 1.57080)</td>
</tr>
<tr>
<td>(0.04759, 0.04887)</td>
<td>191</td>
<td>(0.07982, 1.57080)</td>
</tr>
<tr>
<td>(0.04887, 0.05014)</td>
<td>193</td>
<td>(0.07849, 1.57080)</td>
</tr>
<tr>
<td>(0.05014, 0.05140)</td>
<td>195</td>
<td>(0.07718, 1.57080)</td>
</tr>
</tbody>
</table>

$^a T(d_T,dC) = \sin^{-1} \left( \sqrt{\frac{d_T}{n_0^*}} \right) - \sin^{-1} \left( \sqrt{\frac{d_C}{n_0^*}} \right)$

$^b$ approximate second-stage sample size

$^c$ if $\sin^{-1} \left( \sqrt{\frac{d_T^2}{n_0^* T(d_T,dC)}} \right) - \sin^{-1} \left( \sqrt{\frac{d_C^2}{n_0^* T(d_T,dC)}} \right) \in R^* T(d_T,dC)^c$,
then reject $H_0$
### Table 3.6 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T, d_C)^a$</th>
<th>$n^* T(d_T, d_C)^b$</th>
<th>$R^* T(d_T, d_C)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.05102, 0.05219)</td>
<td>196</td>
<td>(0.07608, 1.57080)</td>
</tr>
<tr>
<td>(0.05219, 0.05337)</td>
<td>198</td>
<td>(0.07478, 1.57080)</td>
</tr>
<tr>
<td>(0.05337, 0.05454)</td>
<td>199</td>
<td>(0.07368, 1.57080)</td>
</tr>
<tr>
<td>(0.05454, 0.05572)</td>
<td>200</td>
<td>(0.07258, 1.57080)</td>
</tr>
<tr>
<td>(0.05572, 0.05689)</td>
<td>201</td>
<td>(0.07148, 1.57080)</td>
</tr>
<tr>
<td>(0.05689, 0.05806)</td>
<td>202</td>
<td>(0.07038, 1.57080)</td>
</tr>
<tr>
<td>(0.05806, 0.05924)</td>
<td>203</td>
<td>(0.06928, 1.57080)</td>
</tr>
<tr>
<td>(0.05924, 0.06041)</td>
<td>204</td>
<td>(0.06817, 1.57080)</td>
</tr>
<tr>
<td>(0.06041, 0.06159)</td>
<td>205</td>
<td>(0.06706, 1.57080)</td>
</tr>
<tr>
<td>(0.06159, 0.06276)</td>
<td>205</td>
<td>(0.06611, 1.57080)</td>
</tr>
<tr>
<td>(0.06276, 0.06394)</td>
<td>205</td>
<td>(0.06499, 1.57080)</td>
</tr>
<tr>
<td>(0.06394, 0.06511)</td>
<td>205</td>
<td>(0.06401, 1.57080)</td>
</tr>
<tr>
<td>(0.06511, 0.06629)</td>
<td>205</td>
<td>(0.06303, 1.57080)</td>
</tr>
<tr>
<td>(0.06629, 0.06746)</td>
<td>205</td>
<td>(0.06203, 1.57080)</td>
</tr>
<tr>
<td>(0.06746, 0.06864)</td>
<td>206</td>
<td>(0.06102, 1.57080)</td>
</tr>
<tr>
<td>(0.06864, 0.06981)</td>
<td>206</td>
<td>(0.06000, 1.57080)</td>
</tr>
<tr>
<td>(0.06981, 0.07099)</td>
<td>206</td>
<td>(0.05896, 1.57080)</td>
</tr>
<tr>
<td>(0.07099, 0.07216)</td>
<td>206</td>
<td>(0.05791, 1.57080)</td>
</tr>
<tr>
<td>(0.07216, 0.07334)</td>
<td>206</td>
<td>(0.05684, 1.57080)</td>
</tr>
<tr>
<td>(0.07334, 0.07451)</td>
<td>205</td>
<td>(0.05589, 1.57080)</td>
</tr>
<tr>
<td>(0.07451, 0.07569)</td>
<td>204</td>
<td>(0.05492, 1.57080)</td>
</tr>
<tr>
<td>(0.07569, 0.07686)</td>
<td>204</td>
<td>(0.05379, 1.57080)</td>
</tr>
<tr>
<td>(0.07686, 0.07803)</td>
<td>203</td>
<td>(0.05278, 1.57080)</td>
</tr>
<tr>
<td>(0.07803, 0.07921)</td>
<td>202</td>
<td>(0.05173, 1.57080)</td>
</tr>
</tbody>
</table>

$^aT(d_T, d_C) = \sin^{-1} \left( \frac{d_T}{n_0^\alpha} \right) - \sin^{-1} \left( \frac{d_C}{n_0^\alpha} \right)$

$^b$approximate second-stage sample size

$^c$if $\sin^{-1} \left( \frac{d_2T}{n_2 T(d_T, d_C)} \right) - \sin^{-1} \left( \frac{d_2C}{n_2 T(d_T, d_C)} \right) \leq R^* T(d_T, d_C)$

then reject $H_0$
Table 3.6 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T, dC)^a$</th>
<th>$n^* T(d_T, dC)^b$</th>
<th>$R^* T(d_T, dC)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.07921, 0.08038)</td>
<td>201</td>
<td>(0.05067, 1.57080)</td>
</tr>
<tr>
<td>(0.08038, 0.08156)</td>
<td>200</td>
<td>(0.04957, 1.57080)</td>
</tr>
<tr>
<td>(0.08156, 0.08273)</td>
<td>199</td>
<td>(0.04844, 1.57080)</td>
</tr>
<tr>
<td>(0.08273, 0.08391)</td>
<td>197</td>
<td>(0.04740, 1.57080)</td>
</tr>
<tr>
<td>(0.08391, 0.08508)</td>
<td>196</td>
<td>(0.04621, 1.57080)</td>
</tr>
<tr>
<td>(0.08508, 0.08626)</td>
<td>195</td>
<td>(0.04498, 1.57080)</td>
</tr>
<tr>
<td>(0.08626, 0.08743)</td>
<td>193</td>
<td>(0.04382, 1.57080)</td>
</tr>
<tr>
<td>(0.08743, 0.08861)</td>
<td>191</td>
<td>(0.04262, 1.57080)</td>
</tr>
<tr>
<td>(0.08861, 0.08978)</td>
<td>189</td>
<td>(0.04137, 1.57080)</td>
</tr>
<tr>
<td>(0.08978, 0.09096)</td>
<td>187</td>
<td>(0.04008, 1.57080)</td>
</tr>
<tr>
<td>(0.09096, 0.09213)</td>
<td>185</td>
<td>(0.03872, 1.57080)</td>
</tr>
<tr>
<td>(0.09213, 0.09331)</td>
<td>183</td>
<td>(0.03731, 1.57080)</td>
</tr>
<tr>
<td>(0.09331, 0.09448)</td>
<td>180</td>
<td>(0.03593, 1.57080)</td>
</tr>
<tr>
<td>(0.09448, 0.09566)</td>
<td>178</td>
<td>(0.03438, 1.57080)</td>
</tr>
<tr>
<td>(0.09566, 0.09683)</td>
<td>175</td>
<td>(0.03281, 1.57080)</td>
</tr>
<tr>
<td>(0.09683, 0.09800)</td>
<td>172</td>
<td>(0.03121, 1.57080)</td>
</tr>
<tr>
<td>(0.09800, 0.09918)</td>
<td>169</td>
<td>(0.02948, 1.57080)</td>
</tr>
<tr>
<td>(0.09918, 0.10035)</td>
<td>166</td>
<td>(0.02764, 1.57080)</td>
</tr>
<tr>
<td>(1.0035, 1.01153)</td>
<td>162</td>
<td>(0.02573, 1.57080)</td>
</tr>
<tr>
<td>(1.01153, 0.10270)</td>
<td>159</td>
<td>(0.02363, 1.57080)</td>
</tr>
<tr>
<td>(0.10270, 0.10388)</td>
<td>155</td>
<td>(0.02142, 1.57080)</td>
</tr>
</tbody>
</table>

\[ T(d_T, dC) = \sin^{-1} \sqrt{\frac{d_T}{n_0}} - \sin^{-1} \sqrt{\frac{d_C}{n_0}} \]

\[ b \text{ approximate second-stage sample size} \]

\[ c \text{ if } \sin^{-1} \sqrt{\frac{d_T}{T(d_T, dC)}} - \sin^{-1} \sqrt{\frac{d_C}{T(d_T, dC)}} \in R^* T(d_T, dC) \]

then reject $H_0$
Table 3.7: Approximate optimal flexible binomial two-stage plans for two-sample case: \((p_0, p_1) = (0.2, 0.35)\)

\((\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.02481, 0.02486, 0.05048, 0.05087)\)

<table>
<thead>
<tr>
<th>(T(d_T, dC)^a)</th>
<th>(n^*T(d_T, dC)^b)</th>
<th>(R^*T(d_T, dC)^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0.04203, 0.04375))</td>
<td>62</td>
<td>((0.16812, 1.57080))</td>
</tr>
<tr>
<td>((0.04375, 0.04547))</td>
<td>65</td>
<td>((0.16209, 1.57080))</td>
</tr>
<tr>
<td>((0.04547, 0.04719))</td>
<td>68</td>
<td>((0.15655, 1.57080))</td>
</tr>
<tr>
<td>((0.04719, 0.04891))</td>
<td>70</td>
<td>((0.15251, 1.57080))</td>
</tr>
<tr>
<td>((0.04891, 0.05063))</td>
<td>72</td>
<td>((0.14868, 1.57080))</td>
</tr>
<tr>
<td>((0.05063, 0.05235))</td>
<td>74</td>
<td>((0.14504, 1.57080))</td>
</tr>
<tr>
<td>((0.05235, 0.05407))</td>
<td>76</td>
<td>((0.14156, 1.57080))</td>
</tr>
<tr>
<td>((0.05407, 0.05579))</td>
<td>78</td>
<td>((0.13823, 1.57080))</td>
</tr>
<tr>
<td>((0.05579, 0.05751))</td>
<td>80</td>
<td>((0.13502, 1.57080))</td>
</tr>
<tr>
<td>((0.05751, 0.05923))</td>
<td>81</td>
<td>((0.13275, 1.57080))</td>
</tr>
<tr>
<td>((0.05923, 0.06095))</td>
<td>82</td>
<td>((0.13052, 1.57080))</td>
</tr>
<tr>
<td>((0.06095, 0.06267))</td>
<td>84</td>
<td>((0.12757, 1.57080))</td>
</tr>
<tr>
<td>((0.06267, 0.06439))</td>
<td>85</td>
<td>((0.12544, 1.57080))</td>
</tr>
<tr>
<td>((0.06439, 0.06611))</td>
<td>86</td>
<td>((0.12335, 1.57080))</td>
</tr>
<tr>
<td>((0.06611, 0.06783))</td>
<td>87</td>
<td>((0.12129, 1.57080))</td>
</tr>
<tr>
<td>((0.06783, 0.06955))</td>
<td>88</td>
<td>((0.11925, 1.57080))</td>
</tr>
<tr>
<td>((0.06955, 0.07127))</td>
<td>89</td>
<td>((0.11724, 1.57080))</td>
</tr>
<tr>
<td>((0.07127, 0.07299))</td>
<td>90</td>
<td>((0.11526, 1.57080))</td>
</tr>
<tr>
<td>((0.07299, 0.07471))</td>
<td>91</td>
<td>((0.11329, 1.57080))</td>
</tr>
</tbody>
</table>

\(T(d_T, dC) = \sin^{-1} \frac{d_T}{n_0} \cdot \sin^{-1} \frac{d_C}{n_0}\)

\(b\) approximate second-stage sample size.

\(c\) if \(\sin^{-1} \frac{d_T^2}{n_T(d_T, dC)} - \sin^{-1} \frac{d_C^2}{n_T(d_T, dC)} \notin R^*T(d_T, dC)\) then reject \(H_0\).
Table 3.7 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T, dC)^a$</th>
<th>$n^*T(d_T, dC)^b$</th>
<th>$R^*T(d_T, dC)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.07471, 0.07643)</td>
<td>92</td>
<td>(0.11134, 1.57080)</td>
</tr>
<tr>
<td>(0.07643, 0.07815)</td>
<td>92</td>
<td>(0.11001, 1.57080)</td>
</tr>
<tr>
<td>(0.07815, 0.07987)</td>
<td>93</td>
<td>(0.10808, 1.57080)</td>
</tr>
<tr>
<td>(0.07987, 0.08159)</td>
<td>93</td>
<td>(0.10673, 1.57080)</td>
</tr>
<tr>
<td>(0.08159, 0.08331)</td>
<td>94</td>
<td>(0.10481, 1.57080)</td>
</tr>
<tr>
<td>(0.08331, 0.08503)</td>
<td>94</td>
<td>(0.10345, 1.57080)</td>
</tr>
<tr>
<td>(0.08503, 0.08675)</td>
<td>95</td>
<td>(0.10154, 1.57080)</td>
</tr>
<tr>
<td>(0.08675, 0.08847)</td>
<td>95</td>
<td>(0.10017, 1.57080)</td>
</tr>
<tr>
<td>(0.08847, 0.09019)</td>
<td>95</td>
<td>(0.09878, 1.57080)</td>
</tr>
<tr>
<td>(0.09019, 0.09191)</td>
<td>96</td>
<td>(0.09687, 1.57080)</td>
</tr>
<tr>
<td>(0.09191, 0.09363)</td>
<td>96</td>
<td>(0.09545, 1.57080)</td>
</tr>
<tr>
<td>(0.09363, 0.09535)</td>
<td>96</td>
<td>(0.09403, 1.57080)</td>
</tr>
<tr>
<td>(0.09535, 0.09707)</td>
<td>96</td>
<td>(0.09258, 1.57080)</td>
</tr>
<tr>
<td>(0.09707, 0.09879)</td>
<td>96</td>
<td>(0.09112, 1.57080)</td>
</tr>
<tr>
<td>(0.09879, 0.10051)</td>
<td>96</td>
<td>(0.08963, 1.57080)</td>
</tr>
<tr>
<td>(0.10051, 0.10224)</td>
<td>96</td>
<td>(0.08813, 1.57080)</td>
</tr>
<tr>
<td>(0.10224, 0.10396)</td>
<td>96</td>
<td>(0.08661, 1.57080)</td>
</tr>
<tr>
<td>(0.10396, 0.10568)</td>
<td>96</td>
<td>(0.08506, 1.57080)</td>
</tr>
<tr>
<td>(0.10568, 0.10740)</td>
<td>96</td>
<td>(0.08349, 1.57080)</td>
</tr>
<tr>
<td>(0.10740, 0.10912)</td>
<td>96</td>
<td>(0.08190, 1.57080)</td>
</tr>
<tr>
<td>(0.10912, 0.11084)</td>
<td>95</td>
<td>(0.08070, 1.57080)</td>
</tr>
<tr>
<td>(0.11084, 0.11256)</td>
<td>95</td>
<td>(0.07904, 1.57080)</td>
</tr>
<tr>
<td>(0.11256, 0.11428)</td>
<td>95</td>
<td>(0.07736, 1.57080)</td>
</tr>
<tr>
<td>(0.11428, 0.11600)</td>
<td>94</td>
<td>(0.07604, 1.57080)</td>
</tr>
</tbody>
</table>

\[ aT(d_T, dC) = \sin^{-1} \sqrt{\frac{d_T^2}{n_0^*}} - \sin^{-1} \sqrt{\frac{d_C^2}{n_0^*}} \]

\[ b \text{ approximate second-stage sample size} \]

\[ c \text{ if } \sin^{-1} \sqrt{\frac{d_T^2}{n^*_T(d_T, dC')}} - \sin^{-1} \sqrt{\frac{d_C^2}{n^*_T(d_T, dC')}} \in R^*T(d_T, dC), \]

then reject $H_0$
Table 3.7 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T,dC)^a$</th>
<th>$n^*T(d_T,dC)^b$</th>
<th>$R^*T(d_T,dC)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.11600, 0.11772)</td>
<td>94</td>
<td>(0.07429, 1.57080)</td>
</tr>
<tr>
<td>(0.11772, 0.11944)</td>
<td>93</td>
<td>(0.07289, 1.57080)</td>
</tr>
<tr>
<td>(0.11944, 0.12116)</td>
<td>93</td>
<td>(0.07105, 1.57080)</td>
</tr>
<tr>
<td>(0.12116, 0.12288)</td>
<td>92</td>
<td>(0.06955, 1.57080)</td>
</tr>
<tr>
<td>(0.12288, 0.12460)</td>
<td>91</td>
<td>(0.06800, 1.57080)</td>
</tr>
<tr>
<td>(0.12460, 0.12632)</td>
<td>91</td>
<td>(0.06602, 1.57080)</td>
</tr>
<tr>
<td>(0.12632, 0.12804)</td>
<td>90</td>
<td>(0.06435, 1.57080)</td>
</tr>
<tr>
<td>(0.12804, 0.12976)</td>
<td>89</td>
<td>(0.06261, 1.57080)</td>
</tr>
<tr>
<td>(0.12976, 0.13148)</td>
<td>88</td>
<td>(0.06080, 1.57080)</td>
</tr>
<tr>
<td>(0.13148, 0.13320)</td>
<td>87</td>
<td>(0.05891, 1.57080)</td>
</tr>
<tr>
<td>(0.13320, 0.13492)</td>
<td>86</td>
<td>(0.05695, 1.57080)</td>
</tr>
<tr>
<td>(0.13492, 0.13664)</td>
<td>85</td>
<td>(0.05489, 1.57080)</td>
</tr>
<tr>
<td>(0.13664, 0.13836)</td>
<td>84</td>
<td>(0.05274, 1.57080)</td>
</tr>
<tr>
<td>(0.13836, 0.14008)</td>
<td>83</td>
<td>(0.05018, 1.57080)</td>
</tr>
<tr>
<td>(0.14008, 0.14180)</td>
<td>82</td>
<td>(0.04811, 1.57080)</td>
</tr>
<tr>
<td>(0.14180, 0.14352)</td>
<td>80</td>
<td>(0.04589, 1.57080)</td>
</tr>
<tr>
<td>(0.14352, 0.14524)</td>
<td>79</td>
<td>(0.04324, 1.57080)</td>
</tr>
<tr>
<td>(0.14524, 0.14696)</td>
<td>77</td>
<td>(0.04070, 1.57080)</td>
</tr>
<tr>
<td>(0.14696, 0.14868)</td>
<td>76</td>
<td>(0.03776, 1.57080)</td>
</tr>
<tr>
<td>(0.14868, 0.15040)</td>
<td>74</td>
<td>(0.03474, 1.57080)</td>
</tr>
<tr>
<td>(0.15040, 0.15212)</td>
<td>72</td>
<td>(0.03151, 1.57080)</td>
</tr>
</tbody>
</table>

\[ aT(d_T,dC) = \sin^{-1} \frac{\sqrt{\frac{d_T}{n_0^*}}} - \sin^{-1} \frac{\sqrt{\frac{d_C}{n_0^*}}} \]

\[ b \text{ approximate second-stage sample size} \]

\[ c \text{ if } \sin^{-1} \frac{\frac{d^2_T}{n_T(d_T,dC)}} - \sin^{-1} \frac{\frac{d^2_C}{n_T(d_T,dC)}} \leq R^*T(d_T,dC), \]

then reject $H_0$. 
Table 3.8: Approximate optimal flexible binomial two-stage plans for
two-sample case: \((p_0, p_1) = (0.2, 0.4)\)

\((\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.02511, 0.02510, 0.04972, 0.04938)\)

<table>
<thead>
<tr>
<th>(T(d_T, d_C)^a)</th>
<th>(n^* T(d_T, d_C)^b)</th>
<th>(R^* T(d_T, d_C)^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.05470, 0.05691)</td>
<td>37</td>
<td>(0.21705, 1.57080)</td>
</tr>
<tr>
<td>(0.05694, 0.05917)</td>
<td>38</td>
<td>(0.21142, 1.57080)</td>
</tr>
<tr>
<td>(0.05917, 0.06141)</td>
<td>40</td>
<td>(0.20357, 1.57080)</td>
</tr>
<tr>
<td>(0.06141, 0.06365)</td>
<td>41</td>
<td>(0.19874, 1.57080)</td>
</tr>
<tr>
<td>(0.06365, 0.06589)</td>
<td>43</td>
<td>(0.19187, 1.57080)</td>
</tr>
<tr>
<td>(0.06589, 0.06813)</td>
<td>44</td>
<td>(0.18759, 1.57080)</td>
</tr>
<tr>
<td>(0.06813, 0.07037)</td>
<td>45</td>
<td>(0.18347, 1.57080)</td>
</tr>
<tr>
<td>(0.07037, 0.07261)</td>
<td>46</td>
<td>(0.17951, 1.57080)</td>
</tr>
<tr>
<td>(0.07261, 0.07485)</td>
<td>47</td>
<td>(0.17569, 1.57080)</td>
</tr>
<tr>
<td>(0.07485, 0.07709)</td>
<td>48</td>
<td>(0.17199, 1.57080)</td>
</tr>
<tr>
<td>(0.07709, 0.07932)</td>
<td>49</td>
<td>(0.16840, 1.57080)</td>
</tr>
<tr>
<td>(0.07932, 0.08156)</td>
<td>49</td>
<td>(0.16658, 1.57080)</td>
</tr>
<tr>
<td>(0.08156, 0.08380)</td>
<td>50</td>
<td>(0.16312, 1.57080)</td>
</tr>
<tr>
<td>(0.08380, 0.08604)</td>
<td>51</td>
<td>(0.15975, 1.57080)</td>
</tr>
<tr>
<td>(0.08604, 0.08828)</td>
<td>52</td>
<td>(0.15646, 1.57080)</td>
</tr>
<tr>
<td>(0.08828, 0.09052)</td>
<td>52</td>
<td>(0.15472, 1.57080)</td>
</tr>
<tr>
<td>(0.09052, 0.09276)</td>
<td>53</td>
<td>(0.15152, 1.57080)</td>
</tr>
<tr>
<td>(0.09276, 0.09500)</td>
<td>53</td>
<td>(0.14979, 1.57080)</td>
</tr>
<tr>
<td>(0.09500, 0.09724)</td>
<td>54</td>
<td>(0.14667, 1.57080)</td>
</tr>
</tbody>
</table>

\(^a T(d_T, d_C) = \sin^{-1} \sqrt{\frac{d_T}{n_0^*}} - \sin^{-1} \sqrt{\frac{d_C}{n_0^*}}\)

\(^b\) approximate second-stage sample size

\(^c\) if \(\sin^{-1} \sqrt{\frac{d_{2T}}{n_T(T(d_T, d_C))}} - \sin^{-1} \sqrt{\frac{d_{2C}}{n_T(T(d_T, d_C))}} \leq R^* T(d_T, d_C)\)

then reject \(H_0\)
Table 3.8 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T, d_C)^a$</th>
<th>$n^* T(d_T, d_C)^b$</th>
<th>$R^* T(d_T, d_C)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.09724, 0.09947)</td>
<td>54</td>
<td>(0.14494, 1.57080)</td>
</tr>
<tr>
<td>(0.09947, 0.10171)</td>
<td>54</td>
<td>(0.14320, 1.57080)</td>
</tr>
<tr>
<td>(0.10171, 0.10395)</td>
<td>55</td>
<td>(0.14016, 1.57080)</td>
</tr>
<tr>
<td>(0.10395, 0.10619)</td>
<td>55</td>
<td>(0.13842, 1.57080)</td>
</tr>
<tr>
<td>(0.10619, 0.10843)</td>
<td>55</td>
<td>(0.13666, 1.57080)</td>
</tr>
<tr>
<td>(0.10843, 0.11067)</td>
<td>56</td>
<td>(0.13368, 1.57080)</td>
</tr>
<tr>
<td>(0.11067, 0.11291)</td>
<td>56</td>
<td>(0.13190, 1.57080)</td>
</tr>
<tr>
<td>(0.11291, 0.11515)</td>
<td>56</td>
<td>(0.13012, 1.57080)</td>
</tr>
<tr>
<td>(0.11515, 0.11739)</td>
<td>56</td>
<td>(0.12831, 1.57080)</td>
</tr>
<tr>
<td>(0.11739, 0.11962)</td>
<td>57</td>
<td>(0.12537, 1.57080)</td>
</tr>
<tr>
<td>(0.11962, 0.12186)</td>
<td>57</td>
<td>(0.12354, 1.57080)</td>
</tr>
<tr>
<td>(0.12186, 0.12410)</td>
<td>57</td>
<td>(0.12170, 1.57080)</td>
</tr>
<tr>
<td>(0.12410, 0.12634)</td>
<td>57</td>
<td>(0.11983, 1.57080)</td>
</tr>
<tr>
<td>(0.12634, 0.12858)</td>
<td>57</td>
<td>(0.11793, 1.57080)</td>
</tr>
<tr>
<td>(0.12858, 0.13082)</td>
<td>57</td>
<td>(0.11601, 1.57080)</td>
</tr>
<tr>
<td>(0.13082, 0.13306)</td>
<td>57</td>
<td>(0.11407, 1.57080)</td>
</tr>
<tr>
<td>(0.13306, 0.13530)</td>
<td>57</td>
<td>(0.11209, 1.57080)</td>
</tr>
<tr>
<td>(0.13530, 0.13754)</td>
<td>57</td>
<td>(0.11009, 1.57080)</td>
</tr>
<tr>
<td>(0.13754, 0.13977)</td>
<td>57</td>
<td>(0.10806, 1.57080)</td>
</tr>
<tr>
<td>(0.13977, 0.14201)</td>
<td>56</td>
<td>(0.10604, 1.57080)</td>
</tr>
<tr>
<td>(0.14201, 0.14425)</td>
<td>56</td>
<td>(0.10403, 1.57080)</td>
</tr>
<tr>
<td>(0.14425, 0.14649)</td>
<td>56</td>
<td>(0.10207, 1.57080)</td>
</tr>
<tr>
<td>(0.14649, 0.14873)</td>
<td>56</td>
<td>(0.10009, 1.57080)</td>
</tr>
<tr>
<td>(0.14873, 0.15097)</td>
<td>56</td>
<td>(0.09826, 1.57080)</td>
</tr>
</tbody>
</table>

\[ aT(d_T, d_C) = \sin^{-1} \frac{d_T}{n_0} - \sin^{-1} \frac{d_C}{n_0} \]

approximate second-stage sample size

\[ b \]

\[ c \text{if } \sin^{-1} \left( \frac{d_T}{n^* T(d_T, d_C)} \right) - \sin^{-1} \left( \frac{d_C}{n^* T(d_T, d_C)} \right) \in R^* T(d_T, d_C) \]

then reject \( H_0 \)
Table 3.8 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T, d_C)^a$</th>
<th>$n^* T(d_T, d_C)^b$</th>
<th>$R^* T(d_T, d_C)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.15097, 0.15321)</td>
<td>55</td>
<td>(0.09686, 1.57080)</td>
</tr>
<tr>
<td>(0.15321, 0.15545)</td>
<td>55</td>
<td>(0.09452, 1.57080)</td>
</tr>
<tr>
<td>(0.15545, 0.15769)</td>
<td>55</td>
<td>(0.09214, 1.57080)</td>
</tr>
<tr>
<td>(0.15769, 0.15992)</td>
<td>54</td>
<td>(0.09054, 1.57080)</td>
</tr>
<tr>
<td>(0.15992, 0.16216)</td>
<td>54</td>
<td>(0.08803, 1.57080)</td>
</tr>
<tr>
<td>(0.16216, 0.16440)</td>
<td>54</td>
<td>(0.08547, 1.57080)</td>
</tr>
<tr>
<td>(0.16440, 0.16664)</td>
<td>53</td>
<td>(0.08363, 1.57080)</td>
</tr>
<tr>
<td>(0.16664, 0.16888)</td>
<td>53</td>
<td>(0.08091, 1.57080)</td>
</tr>
<tr>
<td>(0.16888, 0.17112)</td>
<td>52</td>
<td>(0.07888, 1.57080)</td>
</tr>
<tr>
<td>(0.17112, 0.17336)</td>
<td>52</td>
<td>(0.07600, 1.57080)</td>
</tr>
<tr>
<td>(0.17336, 0.17560)</td>
<td>51</td>
<td>(0.07375, 1.57080)</td>
</tr>
<tr>
<td>(0.17560, 0.17784)</td>
<td>50</td>
<td>(0.07138, 1.57080)</td>
</tr>
<tr>
<td>(0.17784, 0.18008)</td>
<td>50</td>
<td>(0.06817, 1.57080)</td>
</tr>
<tr>
<td>(0.18008, 0.18231)</td>
<td>49</td>
<td>(0.06552, 1.57080)</td>
</tr>
<tr>
<td>(0.18231, 0.18455)</td>
<td>48</td>
<td>(0.06271, 1.57080)</td>
</tr>
<tr>
<td>(0.18455, 0.18679)</td>
<td>47</td>
<td>(0.05971, 1.57080)</td>
</tr>
<tr>
<td>(0.18679, 0.18903)</td>
<td>47</td>
<td>(0.05591, 1.57080)</td>
</tr>
<tr>
<td>(0.18903, 0.19127)</td>
<td>46</td>
<td>(0.05251, 1.57080)</td>
</tr>
<tr>
<td>(0.19127, 0.19351)</td>
<td>45</td>
<td>(0.04886, 1.57080)</td>
</tr>
<tr>
<td>(0.19351, 0.19575)</td>
<td>44</td>
<td>(0.04493, 1.57080)</td>
</tr>
<tr>
<td>(0.19575, 0.19799)</td>
<td>43</td>
<td>(0.04106, 1.57080)</td>
</tr>
</tbody>
</table>

$^a T(d_T, d_C) = \sin^{-1}\left(\sqrt{\frac{d_T}{n_0^*}} - \sin^{-1}\left(\sqrt{\frac{d_C}{n_0^*}}\right)\right)$

$^b$ approximate second-stage sample size

$^c$ if $\sin^{-1}\left(\sqrt{\frac{d_T^*}{n_T d_T, d_C}}\right) - \sin^{-1}\left(\sqrt{\frac{d_C^*}{n_T d_T, d_C}}\right) \in \{t^* T(d_T, d_C)\}$

then reject $H_0$
Table 3.9: Approximate optimal flexible binomial two-stage plans for two-sample case: \((p_0, p_1) = (0.5, 0.55)\)

\((\alpha', \beta', \beta_1', \beta_2') = (0.02501, 0.02501, 0.04997, 0.04995)\)

<table>
<thead>
<tr>
<th>(T(d_T, dC)^a)</th>
<th>(n^* T(d_T, dC)^b)</th>
<th>(R^* T(d_T, dC)^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.01240, 0.01291)</td>
<td>710</td>
<td>(0.04959, 1.57080)</td>
</tr>
<tr>
<td>(0.01291, 0.01342)</td>
<td>746</td>
<td>(0.04776, 1.57080)</td>
</tr>
<tr>
<td>(0.01342, 0.01393)</td>
<td>777</td>
<td>(0.04623, 1.57080)</td>
</tr>
<tr>
<td>(0.01393, 0.01443)</td>
<td>805</td>
<td>(0.04489, 1.57080)</td>
</tr>
<tr>
<td>(0.01443, 0.01494)</td>
<td>830</td>
<td>(0.04371, 1.57080)</td>
</tr>
<tr>
<td>(0.01494, 0.01545)</td>
<td>854</td>
<td>(0.04262, 1.57080)</td>
</tr>
<tr>
<td>(0.01545, 0.01596)</td>
<td>875</td>
<td>(0.04165, 1.57080)</td>
</tr>
<tr>
<td>(0.01596, 0.01646)</td>
<td>895</td>
<td>(0.04073, 1.57080)</td>
</tr>
<tr>
<td>(0.01646, 0.01697)</td>
<td>914</td>
<td>(0.03988, 1.57080)</td>
</tr>
<tr>
<td>(0.01697, 0.01748)</td>
<td>931</td>
<td>(0.03909, 1.57080)</td>
</tr>
<tr>
<td>(0.01748, 0.01799)</td>
<td>947</td>
<td>(0.03834, 1.57080)</td>
</tr>
<tr>
<td>(0.01799, 0.01849)</td>
<td>962</td>
<td>(0.03763, 1.57080)</td>
</tr>
<tr>
<td>(0.01849, 0.01900)</td>
<td>976</td>
<td>(0.03695, 1.57080)</td>
</tr>
<tr>
<td>(0.01900, 0.01951)</td>
<td>989</td>
<td>(0.03631, 1.57080)</td>
</tr>
<tr>
<td>(0.01951, 0.02002)</td>
<td>1002</td>
<td>(0.03568, 1.57080)</td>
</tr>
<tr>
<td>(0.02002, 0.02053)</td>
<td>1013</td>
<td>(0.03509, 1.57080)</td>
</tr>
<tr>
<td>(0.02053, 0.02103)</td>
<td>1024</td>
<td>(0.03450, 1.57080)</td>
</tr>
<tr>
<td>(0.02103, 0.02154)</td>
<td>1034</td>
<td>(0.03394, 1.57080)</td>
</tr>
<tr>
<td>(0.02154, 0.02205)</td>
<td>1043</td>
<td>(0.03340, 1.57080)</td>
</tr>
</tbody>
</table>

\(a T(d_T, dC) = \sin^{-1} \sqrt{\frac{d_T}{n_0}} - \sin^{-1} \sqrt{\frac{d_C}{n_0}}\)

\(b\) approximate second-stage sample size

\(c\) if \(\sin^{-1} \frac{d_T}{T(d_T, dC)} - \sin^{-1} \frac{d_C}{T(d_T, dC)} \in R^* T(d_T, dC)\), then reject \(H_0\)
Table 3.9 (Continued)

<table>
<thead>
<tr>
<th>( T(d_T, dC)^a )</th>
<th>( n^* T(d_T, dC)^b )</th>
<th>( R^* T(d_T, dC)^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.02205, 0.02256)</td>
<td>1051</td>
<td>(0.02288, 1.57080)</td>
</tr>
<tr>
<td>(0.02256, 0.02306)</td>
<td>1059</td>
<td>(0.02337, 1.57080)</td>
</tr>
<tr>
<td>(0.02306, 0.02357)</td>
<td>1069</td>
<td>(0.02376, 1.57080)</td>
</tr>
<tr>
<td>(0.02357, 0.02408)</td>
<td>1073</td>
<td>(0.02447, 1.57080)</td>
</tr>
<tr>
<td>(0.02408, 0.02469)</td>
<td>1079</td>
<td>(0.02508, 1.57080)</td>
</tr>
<tr>
<td>(0.02469, 0.02530)</td>
<td>1084</td>
<td>(0.02568, 1.57080)</td>
</tr>
<tr>
<td>(0.02530, 0.02591)</td>
<td>1086</td>
<td>(0.02628, 1.57080)</td>
</tr>
<tr>
<td>(0.02591, 0.02652)</td>
<td>1089</td>
<td>(0.02688, 1.57080)</td>
</tr>
<tr>
<td>(0.02652, 0.02713)</td>
<td>1092</td>
<td>(0.02748, 1.57080)</td>
</tr>
<tr>
<td>(0.02713, 0.02774)</td>
<td>1095</td>
<td>(0.02818, 1.57080)</td>
</tr>
<tr>
<td>(0.02774, 0.02835)</td>
<td>1098</td>
<td>(0.02878, 1.57080)</td>
</tr>
<tr>
<td>(0.02835, 0.02896)</td>
<td>1101</td>
<td>(0.02938, 1.57080)</td>
</tr>
<tr>
<td>(0.02896, 0.02956)</td>
<td>1104</td>
<td>(0.02998, 1.57080)</td>
</tr>
<tr>
<td>(0.02956, 0.03017)</td>
<td>1107</td>
<td>(0.03059, 1.57080)</td>
</tr>
<tr>
<td>(0.03017, 0.03078)</td>
<td>1110</td>
<td>(0.03119, 1.57080)</td>
</tr>
<tr>
<td>(0.03078, 0.03139)</td>
<td>1113</td>
<td>(0.03181, 1.57080)</td>
</tr>
<tr>
<td>(0.03139, 0.03200)</td>
<td>1116</td>
<td>(0.03242, 1.57080)</td>
</tr>
<tr>
<td>(0.03200, 0.03261)</td>
<td>1119</td>
<td>(0.03303, 1.57080)</td>
</tr>
<tr>
<td>(0.03261, 0.03322)</td>
<td>1122</td>
<td>(0.03364, 1.57080)</td>
</tr>
<tr>
<td>(0.03322, 0.03383)</td>
<td>1087</td>
<td>(0.02283, 1.57080)</td>
</tr>
<tr>
<td>(0.03383, 0.03443)</td>
<td>1082</td>
<td>(0.02343, 1.57080)</td>
</tr>
</tbody>
</table>

\[ a_T(d_T, dC) = \sin^{-1} \sqrt{\frac{d_T^2}{n_0^*}} - \sin^{-1} \sqrt{\frac{d_C^2}{n_0^*}} \]

b approximate second-stage sample size

\[ c_{\text{if}} \sin^{-1} \left( \frac{d_T^2}{n_T(d_T, dC)} \right) = \sin^{-1} \left( \frac{d_C^2}{n_T(d_T, dC)} \right) \in R^* \Rightarrow H_0 \]
Table 3.9 (Continued)

<table>
<thead>
<tr>
<th>$T(d_T, d_C)^a$</th>
<th>$n^* T(d_T, d_C)^b$</th>
<th>$R^* T(d_T, d_C)^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.03423, 0.03474)</td>
<td>1077</td>
<td>(0.02191, 1.57080)</td>
</tr>
<tr>
<td>(0.03474, 0.03525)</td>
<td>1071</td>
<td>(0.02144, 1.57080)</td>
</tr>
<tr>
<td>(0.03525, 0.03576)</td>
<td>1064</td>
<td>(0.02097, 1.57080)</td>
</tr>
<tr>
<td>(0.03576, 0.03626)</td>
<td>1057</td>
<td>(0.02048, 1.57080)</td>
</tr>
<tr>
<td>(0.03626, 0.03677)</td>
<td>1050</td>
<td>(0.01998, 1.57080)</td>
</tr>
<tr>
<td>(0.03677, 0.03728)</td>
<td>1041</td>
<td>(0.01948, 1.57080)</td>
</tr>
<tr>
<td>(0.03728, 0.03779)</td>
<td>1032</td>
<td>(0.01897, 1.57080)</td>
</tr>
<tr>
<td>(0.03779, 0.03830)</td>
<td>1023</td>
<td>(0.01843, 1.57080)</td>
</tr>
<tr>
<td>(0.03830, 0.03880)</td>
<td>1013</td>
<td>(0.01789, 1.57080)</td>
</tr>
<tr>
<td>(0.03880, 0.03931)</td>
<td>1002</td>
<td>(0.01733, 1.57080)</td>
</tr>
<tr>
<td>(0.03931, 0.03982)</td>
<td>990</td>
<td>(0.01675, 1.57080)</td>
</tr>
<tr>
<td>(0.03982, 0.04032)</td>
<td>978</td>
<td>(0.01615, 1.57080)</td>
</tr>
<tr>
<td>(0.04032, 0.04083)</td>
<td>965</td>
<td>(0.01553, 1.57080)</td>
</tr>
<tr>
<td>(0.04083, 0.04134)</td>
<td>951</td>
<td>(0.01499, 1.57080)</td>
</tr>
<tr>
<td>(0.04134, 0.04185)</td>
<td>936</td>
<td>(0.01441, 1.57080)</td>
</tr>
<tr>
<td>(0.04185, 0.04236)</td>
<td>921</td>
<td>(0.01380, 1.57080)</td>
</tr>
<tr>
<td>(0.04236, 0.04286)</td>
<td>904</td>
<td>(0.01276, 1.57080)</td>
</tr>
<tr>
<td>(0.04286, 0.04337)</td>
<td>887</td>
<td>(0.01197, 1.57080)</td>
</tr>
<tr>
<td>(0.04337, 0.04388)</td>
<td>868</td>
<td>(0.01114, 1.57080)</td>
</tr>
<tr>
<td>(0.04388, 0.04439)</td>
<td>849</td>
<td>(0.01024, 1.57080)</td>
</tr>
<tr>
<td>(0.04439, 0.04489)</td>
<td>828</td>
<td>(0.00927, 1.57080)</td>
</tr>
</tbody>
</table>

\[ T(d_T, d_C) = \sin^{-1}\left( \frac{d_T}{\sqrt{n_0}} \right) - \sin^{-1}\left( \frac{d_C}{\sqrt{n_0}} \right) \]

\[ n^* \approx \tilde{T}(d_T, d_C) \]

If 

\[ \sin^{-1}\left( \frac{d_T^*}{\sqrt{n^* T(d_T, d_C)}} \right) - \sin^{-1}\left( \frac{d_C^*}{\sqrt{n^* T(d_T, d_C)}} \right) \in n^*_T(d_T, d_C) \]

then reject $H_0$.
As expected, the approach based on differences of transformed sample proportions achieved reasonable approximate sample sizes for every $p_0$ value considered, which turned out to be roughly twice of those for one-sample case in section 3.6.2. Unfortunately, the derived plan using this approach can't recover exact binomial probabilities because of the grouping of possible outcomes. We, however, can compute error probabilities $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)$ based on derived plans. The accuracy of these observed error rates depends on normality of differences of transformed sample proportions and the size approximations of time-scaled sampling effort in Wiener model. As explained above, the interested pairs of $(p_0, p_1)$ have large enough sample sizes for both first and second stages to allow a sufficient degree of approximation. Furthermore, the discrepancies between target $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ and observed $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)$ are very small: for example, the observed error rates of the approximated plan with $p_0 = 5.0$ and $p_1 = 5.5$ in Table 3.9 disagree with the target error rates by at most $0.00005$. This discrepancy between target error rates and observed error rates is restricted to only the third decimal place even in the case of the smallest approximate sample sizes considered (see Table 3.8).

Similar results were observed in the analyses of other objective functions in section 3.2.

### 3.7 Remarks concluding chapter 3

Regarding Wiener computations per se, while the $SPRT$ holds a substantial edge in $AST$ over all three other procedures, it is often hard to implement in practice.

The four Wiener $AST$ functions shown in Figure 3.9 are strictly ordered, with the $AST$ function of $OFP^*$ uniformly about 4% smaller than that of $STP^*$. That
the domination of $STP^*$ by $OFP^*$ is uniform certainly is good news. The bad news is that it is only by about 4%, suggesting a clear limitation to what can be achieved with flexible second stages.

Recall that, in the formulation of chapter 2, one always has the greatest sampling effort in the interior of the continuation region. The approach of this chapter differs from that of chapter 2 in the fact that the greatest sampling effort can occur at the boundary of the continuation region as, for example, in the $AST_{\mu_1}$ optimization. This further flexibility may be particularly welcome when a researcher with substantial evidence for either hypothesis in the interim stage is interested in economy of observation. The free allocation of error rates of both kinds to both stages is useful for the case where the interim stage analysis is already complete and further research is required in addition to that already on hand.

We note that in contemplating our optimum sampling plans for minimizing $AST_{\mu}$ and $AST$: even optimizing expected sampling effort at $\bar{\mu} = \frac{\mu_0 + \mu_1}{2}$ cannot counteract the natural tendency of optimal sequential plans to call for large sampling effort for values of $\mu$ between $\mu_0$ and $\mu_1$. Based on this fact, it is not difficult to predict that the objective function based on the simple average at both hypotheses, as in the formulation of chapter 2, also behaves in similar fashion. To derive an approximate plan by the formulation of chapter 2, we need only to find optimal endpoint of the continuation region using the standard hypotheses under the Wiener model. Then, Transformations discussed in this chapter can be applied. Finding the endpoint of the continuation region is easily done by setting first-order derivatives of the Lagrangian kernel with respect to each endpoint equal to zero, using the empirically established fact that the Lagrangian kernel is U-shaped in the lower endpoint $l$
(and the upper endpoint $u$).

Finally, regarding Wiener computations as approximate binomial computations, the main point is that one Wiener optimization essentially solves all binomial (or normal) simple hypothesis problems, of both one-sample and two-sample type.
CHAPTER 4. CONCLUSION

Certain "flexible" two-stage plans, pertaining to single-treatment investigations, have been studied. These plans are flexible in the sense that second-stage sample size and critical region are allowed to depend on first-stage outcome and were derived such plans through two distinct optimization formulations.

The first formulation leads us to simultaneous interpretations in both Bayes and Neyman-Pearson terms. This includes the correspondence between Bayes wrong-decision losses and Neyman-Pearson wrong-decision probabilities. Moreover, on the Bayesian side, we can interpret these inferred losses in economical terms due to the fact that they are standardized by sampling cost. The resulting plans are in effect Bayes, and thus derivable by separable second-stage optimizations for every possible first-stage outcome. The second formulation, with group-sequential type error rate allocation to both stages, was investigated by standard Wiener plans. This formulation adds more flexibility in controlling stage-specific error rates of both kinds.

Optimal flexible plans obtained through the first formulation generally will show the greatest second-stage sampling effort being called for in the interior of the continuation region. The second formulation, will, in addition, through suitable choice of objective function weights, allow optimal flexible plans with greatest sampling effort somewhere on the continuation region boundary. A further sort of flexibility lies in
the fact that the first-stage can be a pilot study completed previously, or can be a specifically planned component of a two-stage study yet to be done.

Both formulations afford modest reduction in expected sample size in comparison with matched standard two-stage plans, and have the attractive feature that when using the Wiener model, they approximate the families of optimal binomial (or normal) flexible plans for both one-sample and two-sample simple hypothesis cases.

It is also possible to extend the optimal flexible plans to cases with several treatments, which may include a control.
BIBLIOGRAPHY


ACKNOWLEDGMENTS

I would like to express my sincere appreciation and gratitude to Dr. Herbert T. David for his insightful comments, advice, guidance, and constant encouragement throughout my graduate program and the writing of this dissertation.

I am indebted to all of the professors of the Department of Statistics here and at Chung Ang University for their help and guidance and I am especially grateful to the members of my dissertation committee, Dr. H. A. David, Dr. P. N. Hinz, Dr. W. Q. Meeker and Dr. S. Y. Song.

Last, but not least, special thanks to my parents, my wife Jee-Hee and my son Seokwon for their patience, understanding and encouragement, without which this dissertation would not have been possible.