AN ASYMPTOTICALLY OPTIMAL SEQUENTIAL DESIGN FOR
COMPARING SEVERAL EXPERIMENTAL CATEGORIES WITH
A STANDARD OR CONTROL

by
Charles DeWitt Roberts
December 1962

Contract No. AF 49(638)-261

Three sequential procedures are considered in order to
to decide if any of k experimental categories are
better than the standard (or control) and, if so, to
determine one of which. With a specific loss function
and a cost $c > 0$ per observation the three sequential
procedures and fixed sample size procedures are compared
in a certain asymptotic sense as $c \to 0$. In particular, one of the procedures is shown to be optimal in
this specific asymptotic sense.

This research was supported by the Mathematics Division of the
Air Force Office of Scientific Research.

Institute of Statistics
Mimeo Series No. 344
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>v</td>
</tr>
<tr>
<td>I  THE PRACTICAL SOLUTION OF THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The setup and the terminology</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Brief descriptions of Procedures I, II, and III</td>
<td>3</td>
</tr>
<tr>
<td>1.3 More detailed descriptions of Procedures I, II, and III</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Determination of $a$ and $b$</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Discussion of the merits of the three procedures in practice</td>
<td>7</td>
</tr>
<tr>
<td>1.6 Examples</td>
<td>11</td>
</tr>
<tr>
<td>(1) Normal with infinite parameter space</td>
<td>11</td>
</tr>
<tr>
<td>(2) Exponential class</td>
<td>13</td>
</tr>
<tr>
<td>(3) Normal with $k$ experimental categories</td>
<td>14</td>
</tr>
<tr>
<td>(4) Normal with two experimental categories</td>
<td>14</td>
</tr>
<tr>
<td>II ASYMPOTIC RESULTS FOR THE CASE OF TWO EXPERIMENTAL CATEGORIES</td>
<td>16</td>
</tr>
<tr>
<td>2.1 The problem</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Assumptions</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Definitions and notation</td>
<td>17</td>
</tr>
<tr>
<td>2.4 Theorem 1 (The Bayes terminal decision rule)</td>
<td>20</td>
</tr>
<tr>
<td>2.5 Theorem 2 (Optimal asymptotic sample size and risk for fixed sample size procedure)</td>
<td>22</td>
</tr>
<tr>
<td>2.6 Theorem 3 (Bounds for $\alpha_1$ and $\beta_1$)</td>
<td>27</td>
</tr>
<tr>
<td>2.7 Theorem 4 (Asymptotic lower bound for risk of any procedure)</td>
<td>37</td>
</tr>
<tr>
<td>2.8 Theorem 5 (Asymptotic risks of Procedures I, II, and III)</td>
<td>39</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>III ASYMMPTOTIC RESULTS FOR THE CASE OF $k$ EXPERIMENTAL CATEGORIES</td>
<td>47</td>
</tr>
<tr>
<td>3.1 The problem</td>
<td>47</td>
</tr>
<tr>
<td>3.2 Assumptions</td>
<td>47</td>
</tr>
<tr>
<td>3.3 Definitions and notation</td>
<td>48</td>
</tr>
<tr>
<td>3.4 Theorem 6 (The Bayes terminal decision rule)</td>
<td>49</td>
</tr>
<tr>
<td>3.5 Theorem 7 (Optimal sample size and risk for fixed sample size procedure)</td>
<td>50</td>
</tr>
<tr>
<td>3.6 Theorem 8 (Asymptotic lower bound for risk of any procedure)</td>
<td>50</td>
</tr>
<tr>
<td>3.7 Theorem 9 (Asymptotic risks of Procedures I, II, and III)</td>
<td>50</td>
</tr>
<tr>
<td>3.8 Table of prices of the procedures</td>
<td>51</td>
</tr>
<tr>
<td>3.9 Remarks concerning the relative merits of the procedures considered</td>
<td>51</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>52</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>54</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I am deeply indebted to Professor Wassily Hoeffding for proposing the problem, for his numerous suggestions and comments, and for his encouragement and kindness.

I am grateful for my National Defense Education Act fellowship which has supported me during my period of graduate study and research. Bell Telephone Labs is thanked for computer use to run my programs of the Chapter I calculations. I thank Mrs. Doris Gardner for her prompt and accurate typing of the manuscript.
INTRODUCTION

Basic problem. The basic problem here is one of practical interest. This problem is that of comparing \( k \) experimental categories with a standard (or control) in order to determine if any of the experimental categories are better than the standard (or control) and, if so, to determine one of which.

Summary and statement of results. For the above problem three sequential procedures are given with complete specification as to how the procedures are actually carried out in practice.

Justification for using these procedures is given by considering an analogous problem with three analogous procedures when there is specified a loss function and a cost \( c > 0 \) per observation. For the analogous problem it is shown in a certain asymptotic sense as \( c \) tends to zero that all three procedures are strictly better than the best fixed sample size procedure and that one of the procedures is the best procedure possible. By appealing to asymptotic results a discussion of the relative merits of the three sequential procedures as considered in practice is given.
CHAPTER I
THE PRACTICAL SOLUTION OF THE PROBLEM

1.1 The setup and the terminology. We assume there are k experimental categories and \( x^{(j)} \) is the random variable resulting from a measurement with the j-th category. We denote the probability density of \( x^{(j)} \) by \( g(x, \tau_j) \). For simplicity it is supposed here that the larger the value of \( \tau \), the more desirable the category is. We say \( \theta = 0 \) when \( \tau_1 = \tau_2 = \ldots = \tau_k = \tau_o \), and say \( \theta = j \) when \( \tau_1 = \ldots = \tau_{j-1} = \tau_{j+1} = \ldots = \tau_k = \tau_o \) and \( \tau_j = \tau_o + \Delta \) where \( \Delta > 0 \), as described in the following table (where \( \tau = \tau_o + \Delta \)):

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( x^{(1)} )</th>
<th>( x^{(2)} )</th>
<th>( x^{(3)} )</th>
<th>\ldots</th>
<th>( x^{(k)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( g(x, \tau_o) )</td>
<td>( g(x, \tau_o) )</td>
<td>( g(x, \tau_o) )</td>
<td>\ldots</td>
<td>( g(x, \tau_o) )</td>
</tr>
<tr>
<td>1</td>
<td>( g(x, \tau) )</td>
<td>( g(x, \tau_o) )</td>
<td>( g(x, \tau_o) )</td>
<td>\ldots</td>
<td>( g(x, \tau_o) )</td>
</tr>
<tr>
<td>2</td>
<td>( g(x, \tau_o) )</td>
<td>( g(x, \tau) )</td>
<td>( g(x, \tau_o) )</td>
<td>\ldots</td>
<td>( g(x, \tau_o) )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
</tr>
<tr>
<td>k</td>
<td>( g(x, \tau_o) )</td>
<td>( g(x, \tau_o) )</td>
<td>( g(x, \tau_o) )</td>
<td>\ldots</td>
<td>( g(x, \tau) )</td>
</tr>
</tbody>
</table>

(1.1.1)

The decision \( D_0 \) is preferred if \( \theta = 0 \) or if none of the experimental categories is better than the standard (or control) \( \tau \) that is, \( \tau_s \leq \tau_o \) for \( s = 1, 2, \ldots, k \) in the model (1.1.1). The
decision $D_j$ is preferred if $\Theta = j$ or if the $j$-th experimental category is better than the standard, i.e., $\tau_j > \tau_{c7}$ in the model (1.1.1). (We exclude the possibility that more than one category is better than the standard.) We will develop three sequential procedures for choosing one of the $k+1$ decisions ($D_0$, $D_1$, ..., $D_k$) so that the probability of selecting $D_0$ when $\Theta = 0$ is at least $1 - \alpha$, and the probability of selecting $D_j$ when $\Theta = j$ is at least $1 - \beta$ for each $j$, $j = 1, 2, ..., k$. This formulation is similar to that of Paulson $\tau_{27}^2$.

In all cases we will consider sequential procedures for choosing one of $D_0$, $D_1$, ..., $D_k$ which satisfy the following general assumptions:

(i) The decision to stop sampling or to continue sampling as well as the choice of the next experiment (category) may depend on the available observations.

(ii) Once the $(n+1)$-th experiment has been chosen its outcome is assumed to be independent of the $n$ previous outcomes.

For more discussion about the sequential design of experiments and general asymptotic results see Chernoff $\tau_{27}$.

The asymptotic results obtained here are similar to those found in Chernoff $\tau_{27}$. In particular, asymptotic optimality is in the same sense.

---

2 The numbers in square brackets refer to the bibliography.
We will now define three sequential procedures.

1.2 Brief descriptions of procedures I, II, and III.

Procedure I. Assign a random order for observing the k categories. If say (1, 2, ..., k) is chosen, sample on $X^{(1)}$ until a decision is made about the distribution associated with $X^{(1)}$ then if necessary start sampling on $X^{(2)}$. If sampling were begun on $X^{(2)}$, sample on $X^{(2)}$ until a decision is made about the distribution associated with $X^{(2)}$ then if necessary start sampling on $X^{(3)}$ and so on.

Procedure II. Sample in $k$ (or less) - tuples of one observation on each category beginning with a $k$-tuple. After observing each tuple decide whether or not to make a terminal decision and if sampling is continued, decide which, if any, of the categories may be eliminated from further sampling. This is the procedure suggested by Paulson [27].

Procedure III. Take one observation on each category and then on the basis of what is observed choose a category on which to sample next. After each observation select a category on which to sample next. Stop sampling from a category entirely when a decision is made about the distribution associated with that category and if necessary, continue sampling only on the remaining categories.

1.3 More detailed descriptions of procedures I, II, and III. Let $b < 0 < a$, and

$$z(j) = \log \frac{g(x(j), \tau_o + \Delta)}{g(x(j), \tau_o)} \quad \text{for} \quad j = 1, 2, \ldots, k,$$
and \( z_i^{(j)} \) be the \( i \)-th observation on \( z^{(j)} \).

**Procedure I.** Select at random one of the \( k! \) permutations of

\[ (i_1), (i_2), \ldots, (i_k) \]

We would then sample first on \( X^{(i_k)} \), then on \( X^{(i_{k-1})} \), \ldots, then on \( X^{(i_1)} \) in the following way:

(i) Stop sampling on \( X^{(j)} \) when \( b < \sum_{i=1}^{n_j} z_i^{(j)} < a \) is violated for some \( n_j \).

(ii) If for that \( n_j \) we have \( \sum_{i=1}^{n_j} z_i^{(j)} \geq a \), do no further sampling on any category and make decision \( D_j \).

(iii) If for that \( n_j \) we have \( \sum_{i=1}^{n_j} z_i^{(j)} \leq b \), begin sampling on the next category.

(iv) If for each \( j, j = 1, 2, \ldots, k \), there is an \( n_j \) for which \( \sum_{i=1}^{n_j} z_i^{(j)} \leq b \) do no further sampling and make decision \( D_0 \).

**Procedure II.** Sample in \( k \) (or less)-tuples with one observation on each of a subset of \( X^{(1)}, X^{(2)}, \ldots, X^{(k)} \) beginning with a \( k \)-tuple in the following way:

(i) Let \( m \) be an \( n \) for which \( b < \sum_{i=1}^{m} z_i^{(j)} < a \) is violated for some \( j \). If \( \sum_{i=1}^{m} z_i^{(j)} = \max \sum_{i=1}^{m} z_i^{(j)} \geq a \) stop sampling and make decision \( D_s \) where here the \( \max \) is over those \( j \) for which \( m \) observations have been taken on \( X^{(j)} \).

(ii) Stop sampling on \( X^{(j)} \) if for some \( n_j \) one of the inequalities \( b < \sum_{i=1}^{n_j} z_i^{(j)} < a \) is violated.

(iii) If \( \sum_{i=1}^{n_j} z_i^{(j)} \leq b \) for some \( n_j \) and each \( j = 1, 2, \ldots, k \) stop sampling and make decision \( D_0 \).
Procedure III. Take one observation on \( X^{(1)}, X^{(2)}, \ldots, X^{(k)} \) and then proceed as follows:

(i) Sample next on \( X^{(s)} \) if \( z_1^{(s)} = \max z_1^{(1)} \). In general after \( n_1 \) observations on \( X^{(1)}, n_2 \) on \( X^{(2)}, \ldots, n_k \) on \( X^{(k)} \) sample next on \( X^{(s)} \) if \( \frac{n}{\sum z_i^{(s)}} = \max \left\{ \frac{n}{\sum z_i^{(j)}} \right\} \).

(ii) Stop sampling on \( X^{(j)} \) when \( b < \sum_{i=1}^{n_j} z_i^{(j)} < a \) is violated for some \( n_j \).

(iii) If for some \( j \) and some \( n_j \) we have \( \sum_{i=1}^{n_j} z_i^{(j)} > a \) a stop sampling and make decision \( D_j \).

(iv) If for each \( j \) there is an \( n_j \) for which \( \sum_{i=1}^{n_j} z_i^{(j)} < b \) stop sampling and make decision \( D_0 \).

1.4 Determination of \( a \) and \( b \). Suppose \( \alpha, \beta, \) and \( \Delta \) are specified and \( \tau_0 \) is assumed known, and denote

\[
I_0 = E_\Theta \left\{ \log \frac{g(X^{(1)}_1, \tau_0)}{g(X^{(1)}_1, \tau_0 + \Delta)} \right\}
\]

and

\[
I_1 = E_\Theta \left\{ \log \frac{g(X^{(1)}_1, \tau_0 + \Delta)}{g(X^{(1)}_1, \tau_0)} \right\}
\]

where \( E_\Theta \) indicates expectation for \( \Theta \) the state of nature; also let

\[
\lambda = \min \left\{ 1, \frac{k I_0 \beta}{\alpha (k-1) I_0 \tau + (k-1) I_1 \tau} \right\}
\] (1.4.1)
For any of the three procedures, we suggest choosing

\[ a = \log \left( \frac{k}{\lambda \, \alpha} \right) \quad \text{and} \quad b = \log \beta = \frac{(k-1)\lambda \, \alpha}{k} \]  

(1.4.2)

which is what is suggested by Paulson \( I^5 \) for Procedure II.

For the three procedures we will have where \( P_\Theta \) indicates probability when \( \Theta \) is the state of nature

\[
1 - P_\Theta(D_0) = \sum_{j=1}^{k} P_\Theta(D_j)
\]

\[
\leq \sum_{j=1}^{k} P_\Theta \int \sum_{i=1}^{r} z^{(j)}_i \geq a \quad \text{for some} \quad r < \infty
\]

and

\[
1 - P_j(D_j) = 1 - P_1(D_j)
\]

\[
= P_1(D_0) + \sum_{s=2}^{k} P_1(D_s)
\]

\[
\leq P_1 \int \sum_{i=1}^{r} z^{(1)}_i \leq b \quad \text{for some} \quad r < \infty
\]

\[
+ \sum_{s=2}^{k} P_1 \int \sum_{i=1}^{r} z^{(s)}_i \geq a \quad \text{for some} \quad r < \infty
\]

Now it is well-known (see Wald \( I^6 \)) that

\[
P_\Theta \int \sum_{i=1}^{r} z^{(1)}_i \geq a \quad \text{for some} \quad r < \infty \leq e^{-a}
\]

\[
P_1 \int \sum_{i=1}^{r} z^{(1)}_i \leq b \quad \text{for some} \quad r < \infty \leq e^b \quad \text{and}
\]

\[
P_1 \int \sum_{i=1}^{r} z^{(s)}_i \geq a \quad \text{for some} \quad r < \infty \leq e^{-a}
\]

for \( s = 2, 3, \ldots, k \). Thus in order to satisfy the requirement that
\( P_0(D_0) \geq 1 - \alpha \) and \( P_j(D_j) \geq 1 - \beta \), we therefore determine \( a \) and \( b \) so that 
\[ ke^{-a} \leq \alpha \text{ and } e^b + (k-1)e^{-a} \leq \beta. \]
Let \( N_j^{(6)} \) be the sample size required for procedure \( 6 \) and category \( j \) for \( 6 = I, II, III \) and \( j = 1, 2, \ldots, k \) and 
\[ N^{(6)} = N_1^{(6)} + N_2^{(6)} + \ldots + N_k^{(6)} \]
and let \( N_j \) be the sample size required for a Wald boundary to be crossed on category \( j \). Then it follows that
\[
E_s N^{(6)} \leq E_1 N_1 + E_1 N_2 + \ldots + E_1 N_k \quad \text{for } s = 1, 2, \ldots, k.
\]
But an approximate upper bound for \( E_1 N_1 \) is \( \frac{a}{I_1} \) and an approximate upper bound for \( E_1 N_s \) is \( \frac{b}{I_0} \) for \( s = 2, 3, \ldots, k \) so that we may write (with \( \leq \) for approximately less than or equal)
\[
E_s N^{(6)} \leq \frac{a}{I_1} - \frac{(k-1)b}{I_0}
\]
Now if we minimize this approximate upper bound (that is, \( a/I_1 - (k-1)b/I_0 \)) with respect to \( a \) and \( b \) subject to 
\[ ke^{-a} \leq \alpha \text{ and } e^b + (k-1)e^{-a} \leq \beta \]
we have the values for \( a \) and \( b \) given by (1.4.1) and (1.4.2).

1.5 Discussion of the merits of the three procedures in practice.

We will assign a cost of \( c > 0 \) per observation, a loss for making a decision, and a prior distribution that assigns probability \( \xi_j > 0 \) to \( \theta = j, j = 0, 1, \ldots, k \). For our \( (k+1) \)-decision simple hypothesis problem let \( r(\theta, 6) = \text{expected loss} + \text{expected cost of observations} \) with procedure \( 6 \) if \( \theta \) is the state of nature. Let also
\[
r(6) = \sum_{j=0}^{k} \xi_j r(\theta, 6) \quad \text{the Bayes risk}
\]
associated with procedure \( 6 \) and define \( \rho_6 \) the price of procedure.
\[ \rho_0 = \lim_{c \to 0} \sup_{c} \frac{r(c)}{-c \log c} \]

Thus the price of a procedure is a type of measure of the desirability of the procedure where the more desirable procedures have small prices.

In Chapter III it is shown that for our \((k+1)\)-decision simple-hypothesis problem with a certain loss for making a decision and a certain choice of \(a\) and \(b\) and \(I_0 = -E_0 \{ z_1(1) \} \), \(I_1 = E_1 \{ z_1(1) \} \) \(t z_1(0) \), \(r(1) \) \(t z_1(1) \), we have the following table:

<table>
<thead>
<tr>
<th>Procedure 8</th>
<th>Price ( \rho_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best fixed sample size</td>
<td>( \rho_P = \frac{k}{- \log { \max(\alpha_1, \beta_1) }} ) &gt; ( \frac{k}{\min(I_0, I_1)} )</td>
</tr>
<tr>
<td>I</td>
<td>( \rho_I = \frac{k \xi_0}{I_0} + (1 - \xi_0) \left( \frac{1}{I_1} + \frac{k-1}{2I_0} \right) )</td>
</tr>
<tr>
<td>II</td>
<td>( \frac{k \xi_0}{I_0} + \frac{(1-\xi_0)}{I_1} \leq \rho_{II} \leq \frac{k \xi_0}{I_0} + (1-\xi_0) \left( \frac{1}{I_1} + \frac{k-1}{\max(I_0, I_1)} \right) )</td>
</tr>
<tr>
<td>III (asymptotically optimal)</td>
<td>( \rho_{III} = \frac{k \xi_0}{I_0} + \frac{(1-\xi_0)}{I_1} = \rho_{opt} )</td>
</tr>
</tbody>
</table>

Table (1.5.1)
By the table of prices it is seen that all three procedures asymptotically are strictly better than the fixed sample size procedure and that it is not possible to find a procedure asymptotically better than Procedure III. Now if \( 2I_0 \leq I_1 \) Procedure II is suggested over Procedure I. If \( I_1/I_0 \) or \( I_0/I_1 \) is small then Procedure II is approximately as good as (optimal) Procedure III although the possibility is left open that Procedure II is asymptotically optimal. If \( I_1/I_0 \) is small then Procedure I is approximately as good as Procedure III. Now in practice Procedure III may be difficult or troublesome to perform since one may have to change his physical apparatus after each observation and so he may perhaps prefer to choose Procedure I or II. If it is expected that a certain category will be better than the others or it is tentatively possible to rank the categories then it would appear that Procedure I could be desirable and experimentation performed by this ranking. If a ranking is not possible or it is highly suspected that more than one of the categories could be better than the standard (or control) then Procedure II could be the best to choose since it has more of a tendency to select the best of the categories rather than simply one better than the standard which is what Procedures I and III tend to do.

Monte Carlo trials were run on an IBM 7090 computer with two experimental categories for the three procedures described using 25,000 normal random deviates so that the distribution \( g(x,\tau_0) \) was normal with mean 0 and variance 1 and \( g(x,\tau) \) was normal with mean 1 and variance 1 in the table (1.1.1). The runs yielded the following
For further numerical computations in particular for a comparison of the sample size required by a fixed sample size procedure with an approximate upper bound to the expected sample size when normality is assumed see Paulson [57]. In particular from Paulson [57] with the case of the table (1.5.2) we have the following table:
### Table (1.5.3)

<table>
<thead>
<tr>
<th>Error levels</th>
<th>Sample size required for fixed sample size procedure</th>
<th>Approximate upper bound of expected sample sizes for Procedures I, II, and III</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>$\Theta = 0$</td>
</tr>
<tr>
<td>.05</td>
<td>.05</td>
<td>26</td>
</tr>
<tr>
<td>.05</td>
<td>.01</td>
<td>37</td>
</tr>
<tr>
<td>.01</td>
<td>.05</td>
<td>36</td>
</tr>
<tr>
<td>.01</td>
<td>.01</td>
<td>48</td>
</tr>
</tbody>
</table>

1.6 **Examples.** Here we assume $0 < \alpha < 1$, $0 < \beta < 1$, and $\Delta > 0$.

(1) **Normal with infinite parameter space.** Here we satisfy $\Delta, \alpha, \beta$ and assume

$$g(x, \tau) = (2\pi)^{-1/2} \exp \int -\frac{(x-\tau)^2}{2} \, d\tau.$$ 

Suppose the problem is specified by the table

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>$x^{(1)}$</th>
<th>$x^{(2)}$</th>
<th>$x^{(3)}$</th>
<th>$\ldots$</th>
<th>$x^{(k)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$g(x, w_{01})$</td>
<td>$g(x, w_{02})$</td>
<td>$g(x, w_{03})$</td>
<td>$\ldots$</td>
<td>$g(x, w_{0k})$</td>
</tr>
<tr>
<td>1</td>
<td>$g(x, w_{11})$</td>
<td>$g(x, w_{12})$</td>
<td>$g(x, w_{13})$</td>
<td>$\ldots$</td>
<td>$g(x, w_{1k})$</td>
</tr>
<tr>
<td>2</td>
<td>$g(x, w_{21})$</td>
<td>$g(x, w_{22})$</td>
<td>$g(x, w_{23})$</td>
<td>$\ldots$</td>
<td>$g(x, w_{2k})$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\ldots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$k$</td>
<td>$g(x, w_{k1})$</td>
<td>$g(x, w_{k2})$</td>
<td>$g(x, w_{k3})$</td>
<td>$\ldots$</td>
<td>$g(x, w_{kk})$</td>
</tr>
</tbody>
</table>

where

$$w_{ij} \begin{cases} 
\leq 0 & \text{if } i \neq j \\
\geq \Delta & \text{if } i = j
\end{cases}.$$
Then with the three sequential procedures with $z(j) = \Delta x(j) - \Delta^2/2$
and values of $a, b$ given by

$$\lambda = \min\{1, \frac{\beta}{\alpha(k-1)}\}$$

$$a = \log \left(\frac{k}{\lambda \alpha}\right)$$

$$b = \log \left(\beta - \frac{(k-1)\lambda \alpha}{k}\right)$$

we will have $P_o(D_o) \geq 1 - \alpha$ and $P_j(D_j) \geq 1 - \beta$ for $j = 1, 2, \ldots, k$
for all permissible values of the $w_{ij}$. This is seen since

$$P_0 \{ \Delta \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{r \Delta}{2} \right) \geq a \text{ for some } r < \infty \}$$

$$= P_0 \{ \Delta \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{\Delta + w_{01}r}{2} \right) + \frac{\Delta w_{01}r}{2} \geq a \text{ for some } r < \infty \}$$

$$\leq P_0 \{ \Delta \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{(\Delta + w_{01})r}{2} \right) \geq a \text{ for some } r < \infty \}$$

$$= P_0 \{ \left(\Delta - w_{01}\right) \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{(\Delta - w_{01})r}{2} \right) \geq \frac{\left(\Delta - w_{01}\right)a}{\Delta} \text{ for some } r < \infty \}$$

$$\leq e^{-\frac{\Delta - w_{01}}{\Delta} a}$$

$$\leq e^{-a}$$

and
\[
P_l \left\{ \Delta \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{r \Delta}{2} \right) \leq b \right\} \leq \frac{\Delta}{\Delta} \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{w_{11} r}{2} \right) + \frac{\Delta w_{11} r}{2} - \frac{r \Delta^2}{2} \leq b \text{ for some } r < \infty \}
\]
\[
\leq P_l \left\{ \frac{w_{11}}{\Delta} \left( \sum_{i=1}^{r} x_i^{(1)} - \frac{w_{11} r}{2} \right) \leq \frac{w_{11}}{\Delta} b \right\} \leq e^b
\]

Similarly for \( j = 2, 3, \ldots, k \)
\[
P_o \left\{ \Delta \left( \sum_{i=1}^{r} x_i^{(j)} - \frac{r \Delta}{2} \right) \geq a \right\} \leq e^{-a}
\]
and
\[
P_j \left\{ \Delta \left( \sum_{i=1}^{r} x_i^{(j)} - \frac{r \Delta}{2} \right) \leq b \right\} \leq e^b
\]

Therefore \( 1 - P_o(D_o) \leq k e^{-a} \leq \alpha \) and
\[
1 - P_j(D_j) \leq e^b + (k-1) e^{-a} \leq \beta \text{ for } j = 1, 2, \ldots, k .
\]

(2) \textbf{Exponential Class.} Suppose that \( \Delta, \alpha, \beta \) are specified, \( \tau_o \) is known, and we have the problem of table (1.1.1) where

\[
g(x, \tau) = A(x) \exp \int \tau s(x) - h(\tau) d\tau .
\]

Suppose \( z^{(j)} = \Delta s(x^{(j)}) - h(\tau_o + \Delta) + h(\tau_o) \) for \( j = 1, 2, \ldots, k \). Let \( a \) and \( b \) be chosen by (1.4.1) and (1.4.2) where

\[
I_o = E_o(-z_1^{(1)}) \text{ and } I_1 = E_1(z_1^{(1)}). \text{ We will have that any of the three procedures based on this } z^{(j)} \text{ will have the property}
\]
\( P_0(D_o) \geq 1 - \alpha \) and \( P_j(D_j) \geq 1 - \beta \) for \( j = 1, 2, \ldots, k \) for all \( \tau \geq \tau_o + \Delta \). This follows by an argument similar to that in example (1) by making use of the convexity of \( h \).

(3) **Normal with \( k \) experimental categories.** Suppose that \( \alpha, \beta, \) and \( \Delta \) are specified, the problem is that of table (1.1.1) with \( \tau_o \) assumed equal to 0, \( \tau \geq \Delta \), and \( g(x, \tau) = (2\pi)^{-1/2} \exp\left(-\frac{(x-\tau)^2}{2}\right) \).

Suppose \( a \) and \( b \) are chosen by (1.4.2) with
\[
\lambda = \min \left\{ \frac{1}{2}, \frac{\beta}{\alpha(k-1)} \right\}, \quad \text{and let} \quad z^{(j)} = \Delta x^{(j)} - \frac{\Delta^2}{2} \quad \text{for} \quad j = 1, 2, \ldots, k. \quad \text{By example (1) any of the three procedures based on this} \]
z\((j)\) \( \text{will have the property} \quad P_0(D_o) \geq 1 - \alpha \) and \( P_j(D_j) \geq 1 - \beta \)
for \( j = 1, 2, \ldots, k \) uniformly for \( \tau \geq \Delta \).

(4) **Normal with two experimental categories.** Suppose we have the table (1.1.1) and then for simplicity assume \( \tau_o = 0, \tau = \Delta \), and \( g(x, \tau) = (2\pi)^{-1/2} \exp\left(-\frac{(x-\tau)^2}{2}\right) \). Then \( z^{(1)}_1 = \Delta x^{(1)}_1 - \frac{\Delta^2}{2} \) and \( z^{(2)}_1 = \Delta x^{(2)}_1 - \frac{\Delta^2}{2} \) so that Procedure III would say after \( n_1 \) observations on \( x^{(1)} \) and \( n_2 \) observations on \( x^{(2)} \) sample next on \( x^{(1)} \) if \( \sum_{i=1}^{n_1} (\Delta x^{(1)}_i - \frac{\Delta^2}{2}) > \sum_{i=1}^{n_2} (\Delta x^{(2)}_i - \frac{\Delta^2}{2}) \) and otherwise sample next on \( x^{(2)} \). Since \( I_o = I_1 = \frac{\Delta^2}{2} \), we would have by (1.4.1) and (1.4.2) that
\[
a = -b = \log \frac{\beta}{\alpha} \quad \text{if} \quad \beta \leq \alpha
\]
and \( a = \frac{2}{\alpha}, \quad b = \log (\beta - \frac{\alpha}{2}) \) if \( \beta > \alpha \). In particular we have the following table:
<table>
<thead>
<tr>
<th>Error probabilities</th>
<th>Computed bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>.01</td>
<td>.01</td>
</tr>
</tbody>
</table>
CHAPTER II
ASYMPTOTIC RESULTS FOR THE CASE OF TWO
EXPERIMENTAL CATEGORIES

2.1 The problem. We wish to decide the true value of $\Theta$ in

<table>
<thead>
<tr>
<th>$\Theta$</th>
<th>$x^{(1)}$</th>
<th>$x^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\epsilon_0$</td>
<td>$\epsilon_0$</td>
</tr>
<tr>
<td>1</td>
<td>$\epsilon_1$</td>
<td>$\epsilon_0$</td>
</tr>
<tr>
<td>2</td>
<td>$\epsilon_0$</td>
<td>$\epsilon_1$</td>
</tr>
</tbody>
</table>

We will consider asymptotic results for fixed sample size procedures and Procedures I, II, and III defined in Chapter I with $-a = b = \log c$ as the cost per observation $c \to 0^+.

The choice of $a = -b = \log c$. One would prefer that as the cost tended to zero the error probabilities also tend to zero. Suppose that $\alpha$ the probability level of making an incorrect decision when $\Theta = 0$ and $\beta$ the probability level of making an incorrect decision when $\Theta \neq 0$ satisfy $\alpha = M_\alpha c$, $\beta = M_\beta c$ for some $M_\alpha > 0$, $M_\beta > 0$, and $c$ small. Then the suggested values of $a$ and $b$ given by (1.4.1) and (1.4.2) would have $a \sim -\log c$ and $b \sim \log c$. For further discussion about the choice of $-a = b = \log c$ see Chernoff [27].
2.2 Assumptions.

(1) Suppose we have the loss function

\[ L(\Theta, D) = \begin{cases} 
0 & \text{if } D = \Theta \\
1 & \text{if } D \neq \Theta 
\end{cases} \]

when \( \Theta \) is the state of nature and decision \( D \) is made.

(2) Suppose a cost \( c > 0 \) per observation is incurred.

(3) Assume (i) \( g_0 \) and \( g_1 \) are probability densities of different distributions with respect to a \( \sigma \)-finite measure \( \mu \).

(ii) \( g_0(x) = 0 \) if and only if \( g_1(x) = 0 \) almost everywhere \( (\mu) \).

(iii) The integrals

\[ I_0 = \int \left\{ \log \frac{g_0(x)}{g_1(x)} \right\} g_0(x) \, d\mu(x) \]
\[ I_1 = \int \left\{ \log \frac{g_1(x)}{g_0(x)} \right\} g_1(x) \, d\mu(x) \]

exist (finite) and are positive.

(4) The prior distribution \( \pi \) assigns probability \( \pi_j > 0 \) to \( \Theta = j \) for \( j = 0, 1, 2 \) where \( \pi_0 + \pi_1 + \pi_2 = 1 \).

2.3 Definitions and notation.

(5) For \( x_1^{(1)}, x_2^{(1)}, \ldots \) observations on \( X^{(1)} \) and \( x_1^{(2)}, x_2^{(2)}, \ldots \) observations on \( X^{(2)} \)

\( a) \ \ \ \ r_1^{(1)} = \frac{n_1}{\prod_{i=1}^{n_1} g_0(x_1^{(1)})} \), \quad r_2^{(2)} = \frac{n_2}{\prod_{i=1}^{n_2} g_0(x_1^{(2)})} \)
\[(b) \quad z_1^{(1)} = \log \frac{g_1(x_1^{(1)})}{g_0(x_1^{(1)})}, \quad z_1^{(2)} = \log \frac{g_1(x_1^{(2)})}{g_0(x_1^{(2)})}\]

\[(c) \quad n = (n_1, n_2)\]

(6) By a procedure \(\delta\) we mean a stopping rule \(\psi\), an experiment selection rule \(\chi\), and a terminal decision rule \(D\). Define \(r(\Theta, \delta)\) = expected loss + expected cost of observations to be the risk associated with procedure \(\delta\) when \(\Theta\) is the state of nature with

\[r(\delta) = \xi_0 r(0, \delta) + \xi_1 r(1, \delta) + \xi_2 r(2, \delta)\]

(7) For \(-\infty < t < \infty\) let \(t z_1^{(1)}\)

\[(a) \quad A(t) = E_0 e^{t z_1^{(1)}}\]
\[(b) \quad B(t) = E_1 e^{-t z_1^{(1)}}\]

where \(E_0, \Theta = 0,1,2\), denotes expectation when \(\Theta\) is the state of nature. (A or B may be infinite for some \(t\)-values.) Also define

\[(b) \quad \alpha_1 = \inf_t A(t)\]
\[(b) \quad \beta_1 = \inf_t B(t)\]

(8) Define for the procedure \(\delta\) fixed:

\[D = \text{the terminal decision rule where} \quad D = j \text{ when decision} \quad \Theta = j\]

is made. That is, \(D = j\) if and only if the terminal decision that \(\Theta\), the true state of nature, is \(j\) is made.
\( D_j = \) the event that decision \( \Theta = j \) is made, \( D = \{ j \} \).

\( N_1 = \) the number of observations taken on \( X^{(1)} \)

\( N_2 = \) the number of observations taken on \( X^{(2)} \)

\( N = N_1 + N_2 \)

\( M_1 = \) the least \( n_1 \) such that

\[
\sum_{i=1}^{n_1} z_i^{(1)} \geq - \log c
\]

\( M_2 = \) the least \( n_2 \) such that

\[
\sum_{i=1}^{n_2} z_i^{(2)} \geq - \log c
\]

\( M = M_1 + M_2 \)

\( U_1 = \) the least \( n_1 \) such that

\[
\sum_{i=1}^{n_1} z_i^{(1)} \leq \log c
\]

\( U_2 = \) the least \( n_2 \) such that

\[
\sum_{i=1}^{n_2} z_i^{(2)} \leq \log c
\]

\( V_1 = \) the least \( n_1 \) such that

\[
\sum_{i=1}^{n_1} z_i^{(1)} \geq - \log c
\]

\( V_2 = \) the least \( n_2 \) such that

\[
\sum_{i=1}^{n_2} z_i^{(2)} \geq - \log c
\]

\( S(n) \) will be used to denote a fixed sample size procedure based on
\( n = (n_1, n_2) \) observations with terminal decision rule \( D(n) \) and we write \( D_j(n) \) for the event \( \{ D(n) = j \} \). The expressions \( D^0(n) \) and \( D^0(n) \) will be used to denote a Bayes fixed sample size procedure and a Bayes terminal decision rule.

2.4 Theorem 1 (The Bayes terminal decision rule). Given

\[ x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)} \text{ and } x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)} \] mutually independent then

\[
D^0(n) = \begin{cases} 
0 & \text{if } r_{n_1}^{(1)} < \frac{\xi_0}{\xi_1}, \quad r_{n_2}^{(2)} < \frac{\xi_0}{\xi_2} \\
1 & \text{if } r_{n_1}^{(1)} > \frac{\xi_0}{\xi_1}, \quad r_{n_1}^{(1)} > \frac{\xi_2}{\xi_1} r_{n_2}^{(2)} \\
2 & \text{if } r_{n_2}^{(2)} > \frac{\xi_0}{\xi_2}, \quad r_{n_2}^{(2)} > \frac{\xi_1}{\xi_2} r_{n_1}^{(1)}
\end{cases}
\]

is such that \( D^0(n) \) is a Bayes terminal decision rule.

**Proof.** Given \( x = (x_1^{(1)}, \ldots, x_{n_1}^{(1)}, x_1^{(2)}, \ldots, x_{n_2}^{(2)}) \) then

\[
g_0,\bar{n}(x) = \prod_{i=1}^{n_1} g_0(x_1^{(1)}) \prod_{i=1}^{n_2} g_0(x_1^{(2)})
\]

\[
g_1,\bar{n}(x) = \prod_{i=1}^{n_1} g_1(x_1^{(1)}) \prod_{i=1}^{n_2} g_0(x_1^{(2)})
\]

\[
g_2,\bar{n}(x) = \prod_{i=1}^{n_1} g_0(x_1^{(1)}) \prod_{i=1}^{n_2} g_1(x_1^{(2)})
\]

\[
g_5,\bar{n}(x) = \xi_0 g_{0,\bar{n}}(x) + \xi_1 g_{1,\bar{n}}(x) + \xi_2 g_{2,\bar{n}}(x)
\]
\[ \xi_{j,n} = \frac{\xi_{0,n} \xi_{j}}{\xi_{0} \xi_{n}} \]

Therefore

\[ L(\xi_{n}, 0) = \xi_{0,n} L(0,0) + \xi_{1,n} L(1,0) + \xi_{2,n} L(2,0) \]

\[ = \xi_{1,n} + \xi_{2,n} \]

\[ \xi_{1,n}^{(1)} + \xi_{2,n}^{(2)} \]

\[ = \frac{\xi_{1,n}^{(1)} + \xi_{2,n}^{(2)}}{\xi_{0} + \xi_{1,n}^{(1)} + \xi_{2,n}^{(2)}} \]

Similarly

\[ L(\xi_{n}, 1) = \frac{\xi_{0} + \xi_{2,n}^{(2)}}{\xi_{0} + \xi_{1,n}^{(1)} + \xi_{2,n}^{(2)}} \]

and

\[ L(\xi_{n}, 2) = \frac{\xi_{0} + \xi_{1,n}^{(1)}}{\xi_{0} + \xi_{1,n}^{(1)} + \xi_{2,n}^{(2)}} \]

Since the Bayes terminal decision rule makes decision 0 when

\[ L(\xi_{n}, 0) < L(\xi_{n}, 1), L(\xi_{n}, 0) < L(\xi_{n}, 2) \]

decision 1 when

\[ L(\xi_{n}, 1) < L(\xi_{n}, 0), L(\xi_{n}, 1) < L(\xi_{n}, 2) \]

and decision 2 when

\[ L(\xi_{n}, 2) < L(\xi_{n}, 0), L(\xi_{n}, 2) < L(\xi_{n}, 1) \]

then the proof now follows easily.
2.5 Theorem 2 (Optimal asymptotic sample size and risk for fixed sample size procedure).

(a) If $n_0(c) = (n_{01}(c), n_{02}(c))$ where $n_{01}(c) = n_{02}(c) = \log c / \log \left\{ \max(\alpha_1, \beta_1) \right\}$, then

$$r(\delta^0(n_0)) \leq \frac{\sqrt{1 + o(1)^{1/2}} \ c \ \log c}{\log \left\{ \max(\alpha_1, \beta_1) \right\}}.$$ 

(b) If $n = n(c)$ is any sample size for which $r(\delta^0(n)) \leq 0$ ($c \log c$) then

$$r(\delta^0(n)) \geq \frac{\sqrt{1 + o(1)^{1/2}} \ 2 \ c \ \log c}{\log \left\{ \max(\alpha_1, \beta_1) \right\}}.$$ 

(Therefore $n_0$ is an optimal asymptotic (fixed) sample size.)

We first introduce some lemmas. Here we say $h(x)$ is convex if for any two real numbers $x, y$ $h(\gamma x + (1 - \gamma)y) \leq \gamma h(x) + (1 - \gamma)h(y)$ if $0 < \gamma < 1$. For definitions of $A, B, \alpha_1$, and $\beta_1$ see (7) above. The proofs of the following three lemmas follow easily.

**Lemma 1.** The functions $A(t)$ and $B(t)$ are convex.

**Lemma 2.** For $t \in [0, 1]$, $A(t) < \infty$ and $B(t) < \infty$ so that the intervals of (finite) convergence of $A(t)$ and $B(t)$ are each non-degenerate.

**Lemma 3.** (a) $0 < \alpha_1 < 1, 0 < \beta_1 < 1$ and (b) there exists $a \in (0, 1)$ and $b \in (0, 1)$ such that $\alpha_1 = A(a), \beta_1 = B(b)$.

**Lemma 4.** If we denote $\Phi_i$ as that probability measure associated with $\theta = i$, then there exists a number $\sigma > 0$ so that
(a) \( P_o(r_{n_1}^{(1)} > \frac{g_0}{\bar{g}_1}) \leq \sigma \alpha_1^{n_1}, P_o(r_{n_2}^{(2)} > \frac{g_0}{\bar{g}_2}) \leq \sigma \alpha_2^{n_2} \)

\( P_1(r_{n_1}^{(1)} < \frac{g_0}{\bar{g}_1}) \leq \sigma \beta_1^{n_1}, P_1(r_{n_2}^{(2)} > \frac{g_0}{\bar{g}_2}) \leq \sigma \alpha_1^{n_2} \)

\( P_2(r_{n_1}^{(1)} > \frac{g_0}{\bar{g}_1}) \leq \sigma \alpha_1^{n_1}, P_2(r_{n_2}^{(2)} < \frac{g_0}{\bar{g}_2}) \leq \sigma \beta_1^{n_2} \)

(b) If \( 0 < \epsilon < \min(\alpha_1, \beta_1) \), then

(i) \( P_o(r_{n_1}^{(1)} > \frac{g_0}{\bar{g}_1}) \geq (\alpha_1 - \epsilon)^{n_1} \),

\( P_1(r_{n_1}^{(1)} < \frac{g_0}{\bar{g}_1}) \geq (\beta_1 - \epsilon)^{n_1} \)

for \( n_1 \) sufficiently large, and

(ii) \( P_o(r_{n_2}^{(2)} > \frac{g_0}{\bar{g}_2}) \geq (\alpha_1 - \epsilon)^{n_2} \),

\( P_2(r_{n_2}^{(2)} < \frac{g_0}{\bar{g}_2}) \geq (\beta_1 - \epsilon)^{n_2} \)

for \( n_2 \) sufficiently large.

**Proof.** Noting that

\( P_o(r_{n_1}^{(1)} > \frac{g_0}{\bar{g}_1}) = P_o(\sum_{i=1}^{n_1} z_i^{(1)} > \log \frac{g_0}{\bar{g}_1}) \)

\[ -\alpha \log \frac{g_0}{\bar{g}_1} \leq \epsilon \alpha_1^{n_1} \]

by Theorem A.1 of the Appendix, the proof of (a) follows. Since
\[ P_0 \left( r^{(1)}_{n_1} > \frac{g_0}{\delta_1} \right) = P_0 \left( \sum_{i=1}^{n_1} z_i^{(1)} > \log \frac{g_o}{\delta_1} \right) \]

\[ \geq P_0 \left( \sum_{i=1}^{n_1} z_i^{(1)} > \nu n_1 \right) \]

for any \( \nu > 0 \)

and \( n_1 \) sufficiently large and we may choose \( \nu \) so close to 0 that

\[ \alpha^* = \inf \frac{e^{-vt} A(t)}{t} \geq \alpha_1 - \epsilon/2 \] \[ \Rightarrow \alpha^* - \epsilon/2 \geq \alpha_1 - \epsilon. \]

By applying Theorem A.1 again the proof of (b) also follows.

Now the expected loss satisfies (with \( r_1 = r^{(1)}_{n_1}, r_2 = r^{(2)}_{n_2} \))

\[ E(L(0, \delta^0(n))) = P_0(D_1^0(n) + P_0(D_2^0(n))) \]

\[ = P_0 \left( r_1 > \frac{g_0}{\delta_1}, r_1 > \frac{g_2}{\delta_1}, r_2 \right) + P_0 \left( r_2 > \frac{g_0}{\delta_2}, r_2 > \frac{g_1}{\delta_2} \right) \]

\[ \leq P_0 \left( r_1 > \frac{g_0}{\delta_1} \right) + P_0 \left( r_2 > \frac{g_0}{\delta_2} \right) \]

and

\[ E(L(0, \delta^0(n))) \geq P_0 \left( r_1 > \frac{g_0}{\delta_1}, r_2 \leq \frac{g_0}{\delta_2} \right) + P_0 \left( r_2 > \frac{g_0}{\delta_2}, r_1 \leq \frac{g_0}{\delta_1} \right) \]

\[ = \frac{1}{2} P_0 \left( r_1 > \frac{g_0}{\delta_1} \right) P_0 \left( r_2 \leq \frac{g_0}{\delta_2} \right) + P_0 \left( r_2 > \frac{g_0}{\delta_2} \right) P_0 \left( r_1 \leq \frac{g_0}{\delta_1} \right) \]

\[ \geq \frac{1}{2} P_0 \left( r_1 > \frac{g_0}{\delta_1} \right) + \frac{1}{2} P_0 \left( r_2 > \frac{g_0}{\delta_2} \right) \]

for \( n_1, n_2 \) sufficiently large by Lemma 4. Also

\[ E(L(1, \delta^0(n))) = P_1 \left( r_1 \leq \frac{g_0}{\delta_1}, r_2 \leq \frac{g_0}{\delta_2} \right) + P_1 \left( r_2 > \frac{g_0}{\delta_2}, r_2 > \frac{g_1}{\delta_2} \right) r_1 \]

\[ \leq P_1 \left( r_1 \leq \frac{g_0}{\delta_1} \right) + P_1 \left( r_2 > \frac{g_0}{\delta_2} \right) \]
and
\[ E \left(1, \delta(\mathbf{n})\right) \geq P_1 \left(r_1 \leq \frac{\delta}{\delta_1}\right) P_1 \left(r_2 \leq \frac{\delta}{\delta_2}\right) + P_1 \left(r_2 > \frac{\delta}{\delta_2}\right) P_1 \left(r_1 \leq \frac{\delta}{\delta_1}\right) \]
\[ = P_1 \left(r_1 \leq \frac{\delta}{\delta_1}\right) \]

and
\[ E \left(2, \delta(\mathbf{n})\right) = P_2 \left(r_1 \leq \frac{\delta}{\delta_1}, r_2 \leq \frac{\delta}{\delta_2}\right) + P_2 \left(r_1 > \frac{\delta}{\delta_1}, r_1 \leq \frac{\delta}{\delta_1}, r_2\right) \]
\[ \leq P_2 \left(r_2 \leq \frac{\delta}{\delta_2}\right) + P_2 \left(r_1 > \frac{\delta}{\delta_1}\right) \]

and
\[ E \left(2, \delta(\mathbf{n})\right) \geq P_2 \left(r_1 \leq \frac{\delta}{\delta_1}\right) P_2 \left(r_2 \leq \frac{\delta}{\delta_2}\right) + P_2 \left(r_1 > \frac{\delta}{\delta_1}\right) P_2 \left(r_2 \leq \frac{\delta}{\delta_2}\right) \]
\[ = P_2 \left(r_2 \leq \frac{\delta}{\delta_2}\right) \]

Therefore we have

**Lemma 5.** (a) For all \( n_1, n_2 \)
\[ E \left(0, \delta(\mathbf{n})\right) \leq P_0 \left(r_1 > \frac{\delta}{\delta_1}\right) + P_0 \left(r_2 > \frac{\delta}{\delta_2}\right) \]
\[ E \left(1, \delta(\mathbf{n})\right) \leq P_1 \left(r_1 \leq \frac{\delta}{\delta_1}\right) + P_1 \left(r_2 > \frac{\delta}{\delta_1}\right) \]
\[ E \left(2, \delta(\mathbf{n})\right) \leq P_2 \left(r_2 \leq \frac{\delta}{\delta_2}\right) + P_2 \left(r_1 > \frac{\delta}{\delta_1}\right) \]

(b) For all \( n_1 \) and \( n_2 \)
\[ E \left(1, \delta(\mathbf{n})\right) \geq P_1 \left(r_1 \leq \frac{\delta}{\delta_1}\right) \]
\[ E \left(2, \delta(\mathbf{n})\right) \geq P_2 \left(r_2 \leq \frac{\delta}{\delta_2}\right) \]
and for all sufficiently large $n_1$ and $n_2$

$$E \{0, \delta^0(n)\} \geq \frac{1}{2} P_0(r_1 > \frac{\xi_0}{\xi_1}) + \frac{1}{2} P_0(r_2 > \frac{\xi_0}{\xi_2}).$$

Proof of Theorem 2. Now we see by Lemma 5(a) and Lemma 4(a)

$$r(\delta^0(n)) \leq \xi_o (\sigma_{\alpha_1}^{n_1} + \sigma_{\alpha_2}^{n_2}) + \xi_1 (\sigma_{\beta_1}^{n_1} + \sigma_{\alpha_2}^{n_2}) + \xi_2 (\sigma_{\beta_2}^{n_2} + \sigma_{\alpha_1}^{n_1})$$

$$+ c(n_1 + n_2)$$

Now $n_{01} = n_{02} = \max \{ \log c / \log \alpha_1, \log c / \log \beta_1 \}$. Therefore

$$\frac{\log c}{\log \alpha_1} = \alpha_1 = \frac{\log c}{\log \beta_1} = \beta_1 = c = o(c \log c)$$

so that

$$r(\delta^0(n_0)) \leq \left[ 1 + o(1) \right] 2 c \log c / \log \max(\alpha_1, \beta_1)$$

which proves Theorem 2(a).

If $r(\delta^0(n)) \leq o(c \log c)$, then it follows that

$$\lim \inf_{c \to \infty} n_1 = \lim \inf_{c \to \infty} n_2 = +\infty.$$ Let $c < \epsilon < \min(\alpha_1, \beta_1)$ and for $c$ sufficiently small we have for some $M > 0$ that $r(\delta^0(n)) \leq M c \log c$

so that $P_0(r_1 > \xi_0/\xi_1) \leq -2 \xi_0^{-1} M c \log c$, $P_1(r_1 = \xi_0/\xi_1) \leq -\xi_1^{-1} M c \log c$ and also by Lemma 4(b) $P_0(r_1 > \xi_0/\xi_1) \geq (\alpha_1 - \epsilon)^{n_1}$, $P_1(r_1 < \xi_0/\xi_1) \geq (\beta_1 - \epsilon)^{n_1}$. Hence it follows that

$$n_1 \log (\alpha_1 - \epsilon) \leq \log c + \log(-2 \xi_0^{-1} M c \log c)$$

$$n_1 \log (\beta_1 - \epsilon) \leq \log c + \log(-\xi_1^{-1} M \log c)$$
\[ n_1 \geq \frac{\log c}{\log(\alpha_1 - \varepsilon)} + \frac{\log(-2 \xi_0^{-1} M \log c)}{\log (\alpha_1 - \varepsilon)} \]

\[ n_1 \geq \frac{\log c}{\log(\beta_1 - \varepsilon)} + \frac{\log(-\xi_1^{-1} M \log c)}{\log (\beta_1 - \varepsilon)} \]

Thus we have for each \( 0 < \varepsilon < \min (\alpha_1, \beta_1) \)

\[ n_1 \geq \frac{\log c}{\log(\alpha_1 - \varepsilon)} + o(\log c), \quad n_1 \geq \frac{\log c}{\log(\beta_1 - \varepsilon)} + o(\log c) \]

which implies

\[ n_1 \geq \frac{\log c}{\log \alpha_1} + o(\log c), \quad n_1 \geq \frac{\log c}{\log \beta_1} + o(\log c) \]

and so

\[ n_1 \geq \frac{\log c}{\max \{\log \alpha_1, \log \beta_1\}} + o(\log c). \]

Similarly

\[ n_2 \geq \frac{\log c}{\max \{\log \alpha_1, \log \beta_1\}} + o(\log c). \]

Therefore \( r(\delta^0(n)) \geq \lceil 1 + o(1) \rceil / 2c \log c / \max \{\log \alpha_1, \log \beta_1\} \)

which proves Theorem 2(b).

2.6 Theorem 3 (Bounds for \( \alpha_1 \) and \( \beta_1 \))

(a) \( \alpha_1 > e^{-I_0}, \beta_1 > e^{-I_1} \) and

(b) \( -\log \max(\alpha_1, \beta_1) < \min (I_0, I_1). \) \hspace{1cm} (2.6.1)

Proof. Now \( E_0 e^{z_1(1)} \geq e^{z_1(1)} - E_0 z_1(1) = e^{z_1(1)} - t I_0 \) for each \( t \)

with strict inequality holding for \( t \neq 0 \) so we have
\[ \alpha_1 = \inf_{-\infty < t < \infty} E_0 e^{-tZ_1^{(1)}} = E_0 e^{-\alpha Z_1^{(1)}} \text{ for an } \alpha \in (0,1) \text{ by Lemma} \]

3. Thus \( \alpha_1 > e^{-tZ_1^{(1)}} > \inf_{0 < t < 1} E_0 e^{-tZ_1^{(1)}} = e^{-I_0} \). Also

\[ E_1 e^{-tZ_1^{(1)}} > e^{-tI_1} = e^{-I_1} \text{ for each } t \text{ with strict inequality holding for } t \neq 0 \text{ so we have} \]

\[ \beta_1 = \inf_{-\infty < t < \infty} E_1 e^{-tZ_1^{(1)}} = E_1 e^{-\beta Z_1^{(1)}} \text{ for } a, b \in (0,1) \text{ by Lemma} \]

3. Thus \( \beta_1 > e^{-bZ_1^{(1)}} > \inf_{0 < t < 1} E_1 e^{-tZ_1^{(1)}} = e^{-I_1} \) which proves (a).

By (a) \( \log \alpha_1 > -I_0, \log \beta_1 > -I_1 \) and so \( \max\{\log \alpha_1, \log \beta_1\} > \max(-I_0, -I_1) \) which implies \( \max\{\log \alpha_1, \log \beta_1\} < \min(I_0, I_1) \) which proves (b).

**Lemma 6.** If \( S \) is an event such that \( P_0(S) > 0 \) then

(a) \[ P_0(S) E_0 \left\{ \frac{g_1(x_1^{(1)})}{g_0(x_1^{(1)})} \mid S \right\} = P_1(S) \]

(b) \[ P_0(S) E_0 \left\{ \frac{g_1(x_1^{(2)})}{g_0(x_1^{(2)})} \mid S \right\} = P_2(S) \]

(c) \[ P_1(S) E_1 \left\{ \frac{g_0(x_1^{(1)})}{g_1(x_1^{(1)})} \mid S \right\} = P_0(S) \]

(d) \[ P_2(S) E_2 \left\{ \frac{g_0(x_1^{(2)})}{g_1(x_1^{(2)})} \mid S \right\} = P_0(S) \]
if the procedure terminates with probability 1 for $\Theta = 0, 1, 2$.

Proof. Our sample space $\mathcal{X}$ consists of points

$$x = (x_1^{(1)}, x_2^{(1)}, \ldots; x_1^{(2)}, x_2^{(2)}, \ldots)$$

and so define

$$S_{jk} = \{ x : N_1 = j \text{ and } N_2 = k \}$$

$$\psi_{jk} = \begin{cases} 1 & \text{if } x \in S_{jk} \\ 0 & \text{otherwise} \end{cases}$$

$$\phi_{jk} = \begin{cases} 1 & \text{if } x \in S_{11} \cup S_{12} \\ 0 & \text{otherwise} \end{cases}$$

By assumption (3) (ii) we have that $P_0(S) > 0$ implies $P_1(S) > 0$ and $P_2(S) > 0$. Let us now prove (a).

It follows that $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_{jk} = 1$ with probability 1 for $\Theta = 0, 1, 2$. Now

$$P_1(S) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} P_1(S_{jk})$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_{jk} \phi_{jk} \frac{J}{u=1} \frac{K}{v=1} g_1(x_u^{(1)}) \frac{K}{v=1} g_0(x_v^{(2)}) \frac{J}{u=1} \frac{K}{v=1} \mu(x_u^{(1)}) \frac{K}{v=1} \mu(x_v^{(2)})$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_{jk} \phi_{jk} \frac{J}{u=1} \frac{J}{u=1} g_1(x_u^{(1)}) \frac{J}{u=1} \frac{J}{u=1} g_0(x_u^{(1)})$$

$$\cdot \frac{K}{v=1} \frac{K}{v=1} g_0(x_v^{(2)}) \frac{J}{u=1} \frac{K}{v=1} \mu(x_u^{(1)}) \frac{K}{v=1} \mu(x_v^{(2)})$$
\[ P_1(s) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} E_0 \phi_{j k} \phi'_{j k} \psi_{j k} \phi_{j k} \left[ \prod_{u=1}^{N_1} \frac{g_1(x_u^{(1)})}{g_0(x_u^{(1)})} \right] \]

\[ = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} E_0 \phi_{j k} \phi'_{j k} \psi_{j k} \left[ \prod_{u=1}^{N_1} \frac{g_1(x_u^{(1)})}{g_0(x_u^{(1)})} \right] \]

Then we have
\[ P_1(S) = E_o \phi_{N_1N_2} \frac{\sum_{u=1}^{N_1} \frac{g_1(x_u^{(1)})}{g_0(x_u^{(1)})}}{E_o \phi_{N_1N_2}} \]

and since \( E_o \phi_{N_1N_2} = P_o(S) \) we have

\[ P_1(S) = P_o(S) E_o \left( \sum_{u=1}^{N_1} \frac{g_1(x_u^{(1)})}{g_0(x_u^{(1)})} \right) \text{ which proves (a).} \]

The proofs of (b), (c), and (d) are similar.

**Lemma 7.** We have that

(a)(i) \( P_o(\sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c) = P_2(\sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c) \leq c \)

(ii) \( P_1(\sum_{i=1}^{M_1} z_i^{(1)} \leq \log c) \leq c \)

(b)(i) \( P_0(\sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c) = P_1(\sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c) \leq c \)

(ii) \( P_2(\sum_{i=1}^{M_2} z_i^{(2)} \leq \log c) \leq c \)

(c)(i) \( E_0 M_1 = E_2 M_1 = \frac{-\sqrt{1 + o(1)^2} \log c}{I_0} \)

(ii) \( E_1 M_1 = \frac{-\sqrt{1 + o(1)^2} \log c}{I_1} \)
(a)(i) \( E_{oM_2} = E_{1M_2} = \frac{-\sqrt{1 + o(1)^2 \log c}}{I_o} \)

(ii) \( E_{2M_2} = \frac{-\sqrt{1 + o(1)^2 \log c}}{I_1} \)

Proof. Let \( A_n \) be the set in the sample space on which we have \( n = M_1 \) and \( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \) and let \( B_n \) be the set in the sample space on which we have \( n = M_1 \) and \( \sum_{i=1}^{M_1} z_i^{(1)} \leq \log c \). Then on \( A_n \)

\[ \sum_{i=1}^{n} z_i^{(1)} \geq -\log c \] \( \text{That is,} \)

\[ \sum_{i=1}^{n} g_1(x_i^{(1)}) g_o(x_i^{(1)}) \geq \frac{1}{c} \]

\[ \sum_{i=1}^{n} g_o(x_i^{(1)}) \leq c \sum_{i=1}^{n} g_1(x_i^{(1)}) \]

Therefore

\[ P_o(A_n) \leq c \int_{A_n} \sum_{i=1}^{n} g_1(x_i^{(1)}) d\mu(x_1^{(1)}) \ldots d\mu(x_n^{(1)}) = c P_1(A_n) \]

so that,

\[ P_o \left( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \right) = \sum_{n=1}^{\infty} P_o(A_n) \]

\[ \leq c \sum_{n=1}^{\infty} P_1(A_n) = c P_1 \left( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \right) \leq c. \]

Similarly
\[
\begin{align*}
P_1(\sum_{i=1}^{M_1} z_i(1) \leq \log c) &= \sum_{n=1}^{\infty} P_1(B_n) \\
&\leq c \sum_{n=1}^{\infty} P_o(B_n) = c P_o(\sum_{i=1}^{M_1} z_i(1) \leq \log c) \leq c
\end{align*}
\]

which completes the proof of (a). Similarly (b) follows.

Let \( S_o \) be the set in the sample space on which \( \sum_{i=1}^{M_1} z_i(1) \leq \log c \) and \( S_1 \) the set in the sample space on which \( \sum_{i=1}^{M_1} z_i(1) \geq \log c \). We will have \( P_o(S_o) + P_o(S_1) = P_1(S_o) + P_1(S_1) = 1 \).

Now by Theorem A.2 of the Appendix we have

\[
(E_o M_1)(-I_o) = E_o(\sum_{i=1}^{M_1} z_i(1))
\]

\[
= P_o(S_o) E_o(\sum_{i=1}^{M_1} z_i(1) \mid S_o) + P_o(S_1) E_o(\sum_{i=1}^{M_1} z_i(1) \mid S)
\]

\[
\leq P_o(S_o) \log E_o(\sum_{i=1}^{M_1} \frac{g_1(x_i(1))}{g_0(x_i(1))} \mid S_o) + P_o(S_1) \log E_o(\sum_{i=1}^{M_1} \frac{g_1(x_i(1))}{g_0(x_i(1))} \mid S_1)
\]

and by applying Lemma 6 we have

\[
-I_o E_o M_1 \leq P_o(S_o) \log \frac{P_1(S_o)}{P_0(S_o)} + P_o(S_1) \log \frac{P_1(S_1)}{P_0(S_1)}.
\]

By applying Lemma 7(a) we have that \( P_o(S_1) \leq c \) and \( P_1(S_o) \leq c \) so that

\[
\lim_{c \to 0} P_o(S_1) \log \frac{P_1(S_1)}{P_o(S_1)} = 0
\]
and
\[
\lim \inf_{c \to 0} \frac{P_0(S_0) \log \frac{P_1(S_0)}{P_0(S_0)}}{\log c} = \lim \inf_{c \to 0} \frac{P_0(S_0) \log P_1(S_0)}{\log c}
\]

\[
\geq \lim_{c \to 0} \frac{P_0(S_0) \log c}{\log c} = 1.
\]

Thus we have that
\[
E_0 M_1 \geq \frac{-\sqrt{1 + o(1)/7} \log c}{1}\cdot
\]

Now \(E_0 M_1 \leq E_0 U_1\). Let us show that \(E_0 U_1 \leq -\sqrt{1 + o(1)/7} \log c/I_0\).

We shall use a type of proof which can be found in Chernoff [27].

Let \(\epsilon > 0\) be given. If \(n_1 \geq (1 + \epsilon) \log c/I_0\) we have
\[
\log c \geq -n_1 I_0/(1+\epsilon).
\]

In such case for \(t \geq 0\)
\[
P_0 \left( \sum_{i=1}^{n_1} z_i^{(1)} \geq \log c \right)
\]
\[
\leq P_0 \left( \sum_{i=1}^{n_1} z_i^{(1)} \geq -\frac{n_1 I_0}{1+\epsilon} \right) = P_0 \left( \sum_{i=1}^{n_1} (z_i^{(1)} + \frac{I_0}{1+\epsilon}) \geq 0 \right)
\]
\[
\leq \left\{ \begin{array}{l}
E_0 e^{t(z_1^{(1)} + \frac{I_0}{1+\epsilon})}
\end{array} \right\}^{n_1}
\]

But \(z_1^{(1)} + I_0/(1+\epsilon)\) has negative mean and finite moment generating function for \(0 \leq t \leq 1\). Hence the right-hand derivative of the moment generating function is negative at \(t = 0\). Thus there is a
\[ t^* = t^*_o(\varepsilon) > 0 \text{ so that } E_o e^{t^*_o(z^{(1)}_1 + I_o/(1+\varepsilon))} \leq \alpha_o \text{ for some } \alpha_o = \alpha_o(\varepsilon), \quad 0 < \alpha_o < 1. \] Thus it follows that

\[ P_o^n \left( \sum_{i=1}^{n_1} z^{(1)}_i \right) \geq \log c \leq \alpha_o^n \]

for \( n_1 \geq -(1+\varepsilon) \log c/I_o \). This is sufficient to show that \( E_o U_1 \leq -\sqrt{1 + o(1)} / \sqrt{\log c/I_o} \). Therefore we have that \( E_o M_1 = -\sqrt{1 + o(1)} / \sqrt{\log c/I_o} \) which proves (c)(i).

By Theorem A. 2 we have

\[ -(E_1 M_1) I_1 = E_1 \left( \sum_{i=1}^{M_1} z^{(1)}_i \right) \]

\[ = P_1(S_1)E_1 \left( \sum_{i=1}^{M_1} z^{(1)}_i \right) | S_1 \right) + P_1(S_1)E_1 \left( \sum_{i=1}^{M_1} z^{(1)}_i \right) | S_1 \]

\[ \leq P_1(S_1) \log E_1 \left( \prod_{i=1}^{M_1} \frac{g_0(X^{(1)}_i)}{g_1(X^{(1)}_i)} \right) | S_1 \right) + P_1(S_1) \log E_1 \left( \prod_{i=1}^{M_1} \frac{g_0(X^{(1)}_i)}{g_1(X^{(1)}_i)} \right) | S_1 \]

and by Lemma 6 we have

\[ -I_1 E_1 M_1 \leq P_1(S_o) \log \frac{P_o(S_o)}{P_1(S_1)} + P_1(S_1) \log \frac{P_o(S_1)}{P_1(S_1)}. \]

By applying Lemma 7(a) again we have that \( P_o(S_1) \leq c \) and \( P_1(S_o) \leq c \) so that

\[ \lim_{c \to 0} P_1(S_o) \log \frac{P_o(S_o)}{P_1(S_o)} = 0 \]

and
\[
\lim_{c \to 0} \inf \frac{P_1(S_1) \log \frac{P_0(S_1)}{P_1(S_1)}}{\log c} = \lim_{c \to 0} \inf \frac{P_1(S_1) \log P_0(S_1)}{\log c} \geq \lim_{c \to 0} \frac{P_1(S_1) \log c}{\log c} = 1.
\]

Thus we have \(E_1 I_1 \geq -\int 1 + o(1) \log c/I_1\). Now if \(\epsilon > 0\) and \(n_1 \geq -(1+\epsilon) \log c/I_1\) we have \(-\log c \leq n_1 I_1/(1+\epsilon)\). In such case for \(t \leq 0\)

\[P_0 \left( \sum_{i=1}^{n_1} z_i^{(1)} \leq -\log c \right) \]

\[\leq P_1 \left( \sum_{i=1}^{n_1} z_i^{(1)} \leq \frac{n_1 I_1}{1+\epsilon} \right) = P_1 \left( \sum_{i=1}^{n_1} \left( z_i^{(1)} - \frac{I_1}{1+\epsilon} \right) \leq 0 \right) \]

\[\leq E_1 e^{\epsilon \left( z_1^{(1)} - \frac{I_1}{1+\epsilon} \right)} e^{-n_1} .\]

But \(z_1^{(1)} - I_1/(1+\epsilon)\) has positive mean and finite moment generating function for \(-1 \leq t \leq 0\). Hence the left-hand derivative of the moment generating function is positive at \(t = 0\). Thus there is a \(t_1^* = t_1^*(\epsilon) < 0\) so that

\[E_1 e^{t_1^*(z_1^{(1)} - I_1/(1+\epsilon))} \leq \alpha_1 \text{ for some } \alpha_1 = \alpha_1(\epsilon), 0 < \alpha_1 < 1.\]

Thus it follows that
\[ P_1 \left( \sum_{i=1}^{n_1} z_i^{(1)} \leq -\log c \right) \leq \alpha_1 \] for \( n_1 \geq - (1+\varepsilon) \log c/I_1 \).

Therefore we have that
\[ E_{1M_1} = -\sqrt{1 + o(1)} \log c/I_1 \] which proves (c) (ii). The proof of (d) is similar to that of (c). This completes the proof of Lemma 7.

2.7 Theorem 4 (Asymptotic lower bound for risk of any procedure).

Any procedure \( \delta \) for which \( r(\delta) \leq 0 (c \log c) \) has
\[ r(\delta) \geq - (1+o(1)) \left\{ \frac{2E_0}{I_0} + \frac{(\varepsilon_1 + \varepsilon_2)}{I_1} \right\} c \log c \]

Proof. If \( r(\delta) \leq 0 (c \log c) \) then it follows that
\[ P_{i}(D_j) \leq -Mc \log c \] if \( i \neq j \) for some \( M > 0 \) and \( c \) sufficiently small. Now by Lemma 6 and Theorem A.2 we have
\[ -I_0 E_0 N_1 = E_0 \sum_{i=1}^{N_1} z_i^{(1)} \]

\[ = P_0(D_1 \text{ or } D_2) E_0 \sum_{i=1}^{N_1} z_i^{(1)} | D_1 \text{ or } D_2 \]

\[ + P_0(D_0) E_0 \sum_{i=1}^{N_1} z_i^{(1)} | D_0 \]

\[ \leq P_0(D_1 \text{ or } D_2) \log E_0 \left( \sum_{i=1}^{N_1} \frac{g_1(X_i^{(1)})}{g_0(X_i^{(1)})} | D_1 \text{ or } D_2 \right) \]

\[ + P_0(D_0) \log E_0 \left( \sum_{i=1}^{N_1} \frac{g_1(X_i^{(1)})}{g_0(X_i^{(1)})} | D_0 \right) \]

\[ = P_0(D_1 \text{ or } D_2) \log \frac{P_1(D_1 \text{ or } D_2)}{P_0(D_1 \text{ or } D_2)} + P_0(D_0) \log \frac{P_1(D_0)}{P_0(D_0)} \]
But
\[ \lim \inf_{c \to \infty} \frac{P(D_0) \log P_1(D_0)}{\log c} = \lim \inf_{c \to \infty} \frac{P(D_0) \log P_1(D_0)}{\log c} \]
\[ \geq \lim_{c \to \infty} \frac{P(D_0) \log (-M c \log c)}{\log c} = 1 \]
and
\[ \lim_{c \to \infty} P(D_1 \text{ or } D_2) \log \frac{P_1(D_1 \text{ or } D_2)}{P_0(D_1 \text{ or } D_2)} = 0 \]
so that
\[ E_0 N_1 \geq \frac{-(1+c(1))}{I_0} \log c. \]
Similarly \[ E_0 N_2 \geq \frac{-(1+c(1))}{I_0} \log c. \]
Now again by Lemma 6 and Theorem A.2
\[ -I_1 E_1 N_1 = E_1 \left( -\sum_{i=1}^{N_1} z_i^{(1)} \right) \]
\[ = P_1(D_0 \text{ or } D_2) E_1(-\sum_{i=1}^{N_1} z_i^{(1)} | D_0 \text{ or } D_2) + P_2(D_1) E_1(-\sum_{i=1}^{N_1} z_i^{(1)} | D_1) \]
\[ \leq P_1(D_0 \text{ or } D_2) \log E_1 \left( \frac{g_0(x_i^{(1)})}{g_1(x_i^{(1)})} \right) | D_0 \text{ or } D_2 \]
\[ + P_1(D_1) \log E_1 \left( \frac{g_0(x_i^{(1)})}{g_1(x_i^{(1)})} \right) | D_1 \]
\[ = P_1(D_0 \text{ or } D_2) \log \frac{P_o(D_0 \text{ or } D_2)}{P_1(D_0 \text{ or } D_2)} + P_1(D_1) \log \frac{P_o(D_1)}{P_1(D_1)} \]

But
\[ \lim_{c \to 0} P_1(D_0 \text{ or } D_2) \log \frac{P_o(D_0 \text{ or } D_2)}{P_1(D_0 \text{ or } D_2)} = 0 \]

and
\[ \liminf_{c \to 0} \frac{P_1(D_1) \log \frac{P_o(D_1)}{P_1(D_1)}}{\log c} = \liminf_{c \to 0} \frac{P_1(D_1) \log P_o(D_1)}{\log c} \]

\[ \geq \lim_{c \to 0} \frac{P_1(D_1) \log (-M c \log c)}{\log c} = 1 \]

so that
\[ E_1 N_1 \geq -(1 + o(1)) \log c/I_1. \text{ Similarly } E_2 N_2 \geq -(1 + o(1)) \log c/I_1. \]

Now \( E_1 N_2 \geq 0 \) and \( E_2 N_1 \geq 0 \) and therefore it follows that
\[ r(\delta) \geq c \xi_0 E_0 N + c \xi_1 E_1 N + c \xi_1 E_1 N + c \xi_2 E_2 N \]

\[ \geq -(1 + o(1)) \left\{ \frac{2\xi_0}{I_0} + \frac{(\xi_1 + \xi_2)}{I_1} \right\} c \log c \]

which completes the proof.

2.8 Theorem 5 (Asymptotic risks of Procedures I, II, and III). We have that

(a) \[ r(I) = -\sqrt{1 + o(1)} \left\{ \frac{2\xi_0}{I_0} + (\xi_1 + \xi_2)(\frac{1}{I_1} + \frac{1}{2I_0}) \right\} c \log c \]
(b) \(-\sqrt{1 + o(1)} \left\{ \frac{2\varepsilon_0}{\mathcal{I}_0} + \frac{(\varepsilon_1 + \varepsilon_2)}{\mathcal{I}_1} \right\} c \log c \leq r(\mathcal{I}) \leq -\sqrt{1 + o(1)} \left\{ \frac{2\varepsilon_0}{\mathcal{I}_0} + (\varepsilon_1 + \varepsilon_2)(\frac{1}{\mathcal{I}_1} + \frac{1}{\max(\mathcal{I}_0, \mathcal{I}_1)} \right\} c \log c \)

(c) \(r(\mathcal{I}) = -\sqrt{1 + o(1)} \left\{ \frac{2\varepsilon_0}{\mathcal{I}_0} + \frac{(\varepsilon_1 + \varepsilon_2)}{\mathcal{I}_1} \right\} c \log c \)

where \(r(\delta)\) denotes Bayes risk with Procedure \(\delta, \delta = I, II, III.\)

\(\text{Proof.}\) let us first consider Procedure I. Now

\[ L(0,\delta) = \mathbb{P}_0(D_1 \text{ or } D_2) \]

\[ = \frac{1}{2} \mathbb{P}_0(D_1 \text{ or } D_2 \mid \text{ started with } X^{(1)}) + \frac{1}{2} \mathbb{P}_0(D_1 \text{ or } D_2 \mid \text{ started with } X^{(2)}) \]

\[ = \frac{1}{2} \mathbb{P}_0\left( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \text{ or } \sum_{i=1}^{M_1} z_i^{(1)} \leq \log c, \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c \right) \]

\[ + \frac{1}{2} \mathbb{P}_0\left( \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c \text{ or } \sum_{i=1}^{M_2} z_i^{(2)} \leq \log c, \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \right) \]

\[ = \frac{1}{2} \left\{ \mathbb{P}_0\left( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \right) + \mathbb{P}_0\left( \sum_{i=1}^{M_1} z_i^{(1)} \leq \log c \right) \mathbb{P}_0\left( \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c \right) \right\} \]

\[ + \frac{1}{2} \left\{ \mathbb{P}_0\left( \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c \right) + \mathbb{P}_0\left( \sum_{i=1}^{M_2} z_i^{(2)} \leq \log c \right) \mathbb{P}_0\left( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \right) \right\} \]
\[
\leq \frac{1}{2} (c + c) + \frac{1}{2} (c + c) = 2c \quad \text{by Lemma 7.}
\]

\[
L(1, 5) = P_1(D_1 \text{ or } D_2)
\]

\[
= \frac{1}{2} P_1(D_1 \text{ or } D_2) \quad \text{started with } x^{(1)}
\]

\[
+ \frac{1}{2} P_1(D_1 \text{ or } D_2) \quad \text{started with } x^{(2)}
\]

\[
= \frac{1}{2} P_1(\sum_{i=1}^{M_1} z_i^{(1)} \leq \log c, \sum_{i=1}^{M_2} z_i^{(2)} \leq \log c \text{ or } \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c)
\]

\[
+ \frac{1}{2} P_1(\sum_{i=1}^{M_2} z_i^{(2)} \leq \log c, \sum_{i=1}^{M_1} z_i^{(1)} \leq \log c \text{ or } \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c)
\]

\[
= \frac{1}{2} \left[ P_1(\sum_{i=1}^{M_1} z_i^{(1)} \leq \log c) P_1(\sum_{i=1}^{M_2} z_i^{(2)} \leq \log c) + P_1(\sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c) \right]
\]

\[
+ \frac{1}{2} \left[ P_1(\sum_{i=1}^{M_2} z_i^{(2)} \leq \log c) P_1(\sum_{i=1}^{M_1} z_i^{(1)} \leq \log c) + P_1(\sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c) \right]
\]

\[
\leq \frac{1}{2} (c + c) + \frac{1}{2} (c + c) = 2c \quad \text{by Lemma 7.}
\]

Similarly \(L(2, 5) \leq 2c\).

Now \(E_0 N = E_0 N_1 + E_0 N_2\) so that

\[
E_0 N = \frac{1}{2} E_0(N_1) \quad \text{started with } x^{(1)}
\]

\[
+ \frac{1}{2} E_0(N_2) \quad \text{started with } x^{(1)}
\]

\[
+ \frac{1}{2} E_0(N_1) \quad \text{started with } x^{(2)}
\]

\[
+ \frac{1}{2} E_0(N_2) \quad \text{started with } x^{(2)}
\]
\[ E_{0N} = \frac{1}{2} \left\{ \frac{-(1+o(1)) \log c}{I_0} \right\} + \frac{1}{2} \left\{ 1 + O(c) \right\} \left\{ \frac{-(1+o(1)) \log c}{I_0} \right\} \]
\[ + \frac{1}{2} \left\{ 1 + O(c) \right\} \left\{ \frac{-(1+o(1)) \log c}{I_0} \right\} + \frac{1}{2} \left\{ \frac{-(1+o(1)) \log c}{I_0} \right\} \]
and therefore \( E_{0N} = -(1+o(1)) 2 \log c/I_0 \). Also
\[ E_{1N} = \frac{1}{2} E_{1M_1} + \frac{1}{2} P_1(\Sigma_{i=1}^{M_1} z_i^{(1)} \leq \log c) E_{1M_2} \]
\[ + \frac{1}{2} P_1(\Sigma_{i=1}^{M_2} z_i^{(2)} \leq \log c) E_{1M_1} + \frac{1}{2} E_{1M_2} \]
and by applying Lemma 7 again
\[ E_{1N} = \frac{1}{2} \left\{ \frac{-(1+o(1)) \log c}{I_1} \right\} + \frac{1}{2} \left( O(c) \right) \left( \frac{-(1+o(1)) \log c}{I_0} \right) \]
\[ + \frac{1}{2} \left\{ 1+O(c) \right\} \left\{ \frac{-(1+o(1)) \log c}{I_1} \right\} + \frac{1}{2} \left\{ \frac{-(1+o(1)) \log c}{I_0} \right\} \]
\[ = \frac{-(1+o(1)) \log c}{I_1} - \frac{(1+o(1)) \log c}{2I_0} \]

Similarly
\[ E_2N = \frac{-(1+o(1))}{I_1} \log c - \frac{(1+o(1))}{2I_0} \log c. \]

Then we have

\[ r(I) = g_o P_o(D_1 \text{ or } D_2) + g_1 P_1(D_0 \text{ or } D_2) + g_2 P_2(D_0 \text{ or } D_1) \]

\[ + g_0 c E_0 N + g_1 c E_1 N + g_2 c E_2 N \]

\[ = -(1+o(1)) \left( \frac{2g_o}{I_0} + (g_1 + g_2) \left( \frac{1}{I_1} + \frac{2}{I_0} \right) \right) c \log c. \]

which proves (a).

Now let us consider Procedure II.

\[ L(0,5) = P_0(D_1 \text{ or } D_2) \]

\[ = P_0(D_1) + P_0(D_2) \]

\[ \leq P_0 \left( \sum_{i=1}^{M_1} z_i^{(1)} \geq -\log c \right) + P_0 \left( \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c \right) \]

\[ \leq 2c \text{ by Lemma 7}. \]

\[ L(1,5) = P_1(D_0) + P_1(D_2) \]

\[ \leq P_1 \left( \sum_{i=1}^{M_1} z_i^{(1)} \leq \log c \right) + P_1 \left( \sum_{i=1}^{M_2} z_i^{(2)} \geq -\log c \right) \]

\[ \leq 2c \text{ by Lemma 7}. \]

Similarly \( L(2,5) \leq 2c \).
Now \( E_0 N = E_0 N_1 + E_0 N_2 \)
\[ \leq E_0 M_1 + E_0 M_2 \]
\[ = -(1+o(1)) \frac{2 \log c}{I_0} \quad \text{by Lemma 7} \]

We have seen in the proof of Lemma 7 that \( E_1 U_1 \leq -(1+o(1)) \frac{2 \log c}{I_1} \)
also since \( N_2 \leq U_1 \) and \( N_1 \leq U_1 \), we have that
\[ E_1 N = E_1 U_1 + E_1 N_2 \]
\[ \leq E_1 U_1 + E_1 U_1 \leq \frac{-(1+o(1))}{I_1} 2 \log c \]

and also since \( N_2 \leq M_2 \)
\[ E_1 N = E_1 N_1 + E_1 N_2 \]
\[ \leq E_1 U_1 + E_1 M_2 \leq \frac{-(1+o(1))}{I_1} \log c \quad \frac{-(1+o(1))}{I_0} \log c \]

Thus it follows that
\[ E_1 N \leq \frac{-(1+o(1))}{I_1} \log c \quad \frac{-(1+o(1))}{\max(I_0, I_1)} \log c \]

Similarly
\[ E_2 N \leq \frac{-(1+o(1))}{I_1} \log c \quad \frac{-(1+o(1))}{\max(I_0, I_1)} \log c \]

Thus we have that
\[ r(II) = \xi_0 P_0 (D_1 \text{ or } D_2) + \xi_1 P_1 (D_0 \text{ or } D_2) + \xi_2 P_2 (D_0 \text{ or } D_1) \]
\[ + \xi_0 E_0 N + \xi_1 E_1 N + \xi_2 E_2 N \]
\[-(1+o(1)) \left\{ \frac{2c}{I_o} + (\xi_1 + \xi_2) \left( \frac{1}{I_1} + \frac{1}{\max(I_0, I_1)} \right) \right\} c \log c \]

and in view of Theorem 4 this proves (b).

Now let us consider Procedure III.

\[
L(0, \delta) = P_0(D_1 \text{ or } D_2)
\leq P_0\left( \sum_{i=1}^{M_1} z_i(1) \geq - \log c \right) + P_0\left( \sum_{i=1}^{M_2} z_i(2) \geq - \log c \right)
\leq 2c \text{ by Lemma 7.}
\]

\[
L(1, \delta) = P_1(D_0 \text{ or } D_2)
\leq P_1\left( \sum_{i=1}^{M_1} z_i(1) \leq \log c \right) + P_1\left( \sum_{i=1}^{M_2} z_i(2) \geq - \log c \right)
\leq c + c = 2c \text{ by Lemma 7.}
\]

Similarly \(L(2, \delta) \leq 2c\).

Now \(E_0 N = E_0 N_1 + E_0 N_2\)

\[
\leq E_0 M_1 + E_0 M_2
\leq -(1+o(1)) \frac{\log c}{I_o}
\]

Let us now show that

\(E_1 N \leq -(1+o(1)) \frac{\log c}{I_1} \). We have that \(E_1 N = E_1 N_2\) and

\(E_1 N_1 \leq E_1 M_1 \leq -(1+o(1)) \frac{\log c}{I_1} \). Let us show \(E_1 N_2 = o(1) \log c\).

It is sufficient to show that there is \(\tau, 0 < \tau < 1\), so that after \(n_1\) observations on \(X(1)\) and \(n_2\) on \(X(2)\) we have
\[ P_1(\sum_{i=1}^{n_1} z_i^{(1)} < \sum_{i=1}^{n_2} z_i^{(2)}) \leq \tau_2. \]  Now we have that
\[ P_1(\sum_{i=1}^{n_1} z_i^{(1)} - \sum_{i=1}^{n_2} z_i^{(2)} < 0) \]
\[ \leq E_1 e^{-t \sum_{i=1}^{n_1} z_i^{(1)}} - \sum_{i=1}^{n_2} z_i^{(2)} \]
\[ \leq E_1 e^{-t \sum_{i=1}^{n_1} z_i^{(1)}} \]
\[ = \left\{ E_1 e^{-t z_1^{(1)}} \right\}^{n_1} \left\{ E_1 e^{-t z_1^{(2)}} \right\}^{n_2}. \]

But \(-z_1^{(2)}\) has a positive mean and finite moment generating function for \(-1 \leq t \leq 0\). Hence the left-hand derivative of the moment generating function is positive at \(t = 0\). Thus there is a \(t_1^* < 0\)
so that \(E_1 e^{-t_1^* z_1^{(2)}} \leq \tau\) for some \(\tau, 0 < \tau < 1\). Now since
\[ \psi_1(t) = E_1 e^{t z_1^{(1)}} \]
is convex for \(-1 \leq t \leq 0\) and \(\psi_1(0) = \psi_1(-1) = 1\) then \(\psi_1(t_1^*) = \tau_1\) where \(0 < \tau_1 < 1\) and hence \(\tau_1 \leq 1\) and we have
\[ P_1(\sum_{i=1}^{n_1} z_i^{(1)} - \sum_{i=1}^{n_2} z_i^{(2)} \leq 0) \leq \tau_2 \]
which completes the proof that \(E_1 N \leq -(1+o(1)) \log c/I_1\). Similarly
\(E_2 N \leq -(1+o(1)) \log c/I_1\). Therefore we have
\[ r(III) \leq -(1+o(1)) \left\{ \frac{2 \xi_0}{I_0} + \frac{(\xi_1 + \xi_2)}{I_1} \right\} c \log c \]
which in view of Theorem 4 completes the proof of (c). This completes the proof of the theorem.
CHAPTER III

ASYMPTOTIC RESULTS FOR THE CASE OF \( k \) EXPERIMENTAL CATEGORIES

We are now able to make straightforward generalizations to

3.1 The problem. We wish to decide the true value of \( \Theta \) in

\[
\begin{array}{cccccc}
\Theta & x^{(1)} & x^{(2)} & x^{(3)} & \cdots & x^{(k)} \\
0 & g_0 & g_0 & g_0 & \cdots & g_0 \\
1 & g_1 & g_0 & g_0 & \cdots & g_0 \\
2 & g_0 & g_1 & g_0 & \cdots & g_0 \\
3 & g_0 & g_0 & g_1 & \cdots & g_0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
k & g_0 & g_0 & g_0 & \cdots & g_1 \\
\end{array}
\]

where \( k \geq 2 \). We will consider asymptotic results for fixed sample size procedures and Procedures I, II, and III defined in Chapter I with \( -a = b = \log c \) as the cost per observation \( c \rightarrow 0^+ \).

3.2 Assumptions. The assumptions (1), (2), and (3) of Chapter II apply but now we have
(4') The prior distribution $\xi = (\xi_0, \xi_1, \ldots, \xi_k)$ assigns probability $\xi_j > 0$ to $\Theta = j$ for $j = 0, 1, \ldots, k$.

3.3 Definitions and notation.

(5') For $x_1^{(j)}, x_2^{(j)}, \ldots$ observations on $x^{(j)}, j = 1, 2, \ldots, k$

(a) $n_j^{(j)} = \frac{n_j}{\sum_{i=1}^{n} g_0(x_i^{(j)})} g_1(x_i^{(j)})$

(b) $z_i^{(j)} = \log \frac{g_1(x_i^{(j)})}{g_0(x_i^{(j)})}$

(c) $n = (n_1, n_2, \ldots, n_k)$

and (6), (7) the same as in Chapter II with

$$r(\delta) = \xi_0 r(0, \delta) + \xi_1 r(1, \delta) + \ldots + \xi_k r(k, \delta)$$

(8') Define for the Procedure $\delta$ fixed:

$D = \text{the terminal decision rule where } D = j \text{ when decision } \Theta = j \text{ is made. That is, } D = j \text{ if and only if the terminal decision that } \Theta, \text{ the true state of nature, is } j \text{ is made.}$

$D_j = \text{the event that decision } \Theta = j \text{ is made, } D = j_j, j = 0, 1, \ldots, k$

$N_j = \text{number of observations on } x^{(j)}$

$N = \sum_{j=0}^{k} N_j$

$M_j = \text{the least } n_j \text{ such that } \sum_{i=1}^{n_j} z_i^{(j)} \geq - \log c$

$M = \sum_{j=0}^{k} M_j$

$U_j = \text{the least } n_j \text{ such that } \sum_{i=1}^{n_j} z_i^{(j)} \leq \log c$
\[ v_j = \text{the least } n_j \text{ such that } \sum_{i=1}^{n_j} z_i^{(j)} > -\log c \]

\( \delta(n) \) will be used to denote a fixed sample size procedure based on \( n = (n_1, n_2, \ldots, n_k) \) observations with terminal decision rule \( D(n) \) and we write \( D_j(n) \) for the event \( D(n) = j \). The expressions \( \delta^O(n) \) and \( D^O(n) \) will be used to denote a Bayes fixed sample size procedure and a Bayes terminal decision rule.

3.4 Theorem 6 (The Bayes terminal decision rule). Given \( x_i^{(j)} \), \( j = 1, 2, \ldots, k \) and \( i = 1, 2, \ldots, n_j \) mutually independent then

\[
D^O(n) = \begin{cases} 
0 & \text{if } r_{n_1}^{(1)} < \frac{\xi_0}{\xi_1}, r_{n_2}^{(2)} < \frac{\xi_0}{\xi_2}, \ldots, r_{n_k}^{(k)} < \frac{\xi_0}{\xi_k} \\
1 & \text{if } \frac{\xi_0}{\xi_1} < r_{n_1}^{(1)} \leq \frac{\xi_2}{\xi_1}, r_{n_2}^{(2)} < \frac{\xi_0}{\xi_2}, \ldots, \frac{\xi_k}{\xi_1} r_{n_k}^{(k)} < r_{n_1}^{(1)} \\
2 & \text{if } \frac{\xi_0}{\xi_2} < r_{n_2}^{(2)} \leq \frac{\xi_2}{\xi_2}, r_{n_1}^{(1)} < r_{n_2}^{(2)} \leq \frac{\xi_3}{\xi_2}, r_{n_3}^{(3)} < r_{n_2}^{(2)}, \ldots, \frac{\xi_k}{\xi_2} r_{n_k}^{(k)} < r_{n_2}^{(2)} \\
3 & \text{if } \frac{\xi_0}{\xi_3} < r_{n_3}^{(3)} \leq \frac{\xi_3}{\xi_3}, r_{n_1}^{(1)} < r_{n_3}^{(3)} \leq \frac{\xi_2}{\xi_3}, r_{n_2}^{(2)} < r_{n_3}^{(3)}, \ldots, \frac{\xi_k}{\xi_3} r_{n_k}^{(k)} < r_{n_3}^{(3)} \\
\vdots & \\
k-1 & \text{if } \frac{\xi_0}{\xi_{k-1}} < r_{n_{k-1}}^{(k-1)} \leq \frac{\xi_{k-1}}{\xi_{k-1}}, r_{n_1}^{(1)} < r_{n_{k-1}}^{(k-1)} \leq \frac{\xi_2}{\xi_{k-1}}, r_{n_2}^{(2)} < r_{n_{k-1}}^{(k-1)}, \ldots, \frac{\xi_k}{\xi_{k-1}} r_{n_k}^{(k)} < r_{n_{k-1}}^{(k-1)} \\
k & \text{if } \frac{\xi_0}{\xi_k} < r_{n_k}^{(k)} \leq \frac{\xi_1}{\xi_k} r_{n_1}^{(1)} < r_{n_k}^{(k)} \leq \frac{\xi_2}{\xi_k} r_{n_2}^{(2)} < r_{n_k}^{(k)}, \ldots, \frac{\xi_k-1}{\xi_k} r_{n_{k-1}}^{(k-1)} < r_{n_k}^{(k)} \frac{\xi_k}{\xi_k} r_{n_k}^{(k)} < r_{n_k}^{(k)} \end{cases}
\]
is such that \( D^o(n) \) is a Bayes terminal decision rule.

**Proof.** Similar to Theorem 1.

### 3.5 Theorem 7 (Optimal sample size and risk for fixed sample size procedure).

(a) If \( n_0(c) = (n_{01}(c), n_{02}(c), \ldots, n_{0k}(c)) \) where

\[
n_{01}(c) = n_{02}(c) = \ldots = n_{0k}(c) = \log \frac{c}{\log \left( \max(\alpha_1, \beta_1) \right)}^2,
\]

then

\[
r(\delta^o(n_0)) \leq \frac{(1 + o(1)) k c \log c}{\log \left( \max(\alpha_1, \beta_1) \right)^2}.
\]

(b) If \( n = n(c) \) is any sample size for which \( r(\delta^o(n)) \leq o(c \log c) \) then \( r(\delta^o(n)) \geq \frac{(1 + o(1)) k c \log c}{\log \left( \max(\alpha_1, \beta_1) \right)} \). (Therefore \( n_0 \) is an optimal asymptotic (fixed) sample size.)

**Proof.** Similar to Theorem 2.

### 3.6 Theorem 8 (Asymptotic lower bound for risk of any procedure).

Any procedure \( \delta \) for which \( r(\delta) \leq o(c \log c) \) has

\[
r(\delta) \geq -(1 + o(1)) \left\{ \frac{k \xi_0}{I_0} + \frac{(\xi_1 + \xi_2 + \ldots + \xi_k)}{I_1} \right\} \log c.
\]

**Proof.** Similar to Theorem 4.

### 3.7 Theorem 9 (Asymptotic risks of Procedures I, II, and III).

We will have that

(a) \( r(I) = -(1 + o(1)) \left\{ \frac{k \xi_0}{I_0} + \frac{(\xi_1 + \xi_2 + \ldots + \xi_k)}{I_1} \left( \frac{1}{I_1} + \frac{k-1}{2I_0} \right) \right\} \log c \)
(b) \[-(1+o(1)) \left\{ \frac{k \xi_o}{I_o} + \frac{(\xi_1 + \xi_2 + \cdots + \xi_k)}{I_1} \right\} c \log c \leq r(\delta) \leq -(1+o(1)) \left\{ \frac{k \xi_o}{I_o} + \frac{(\xi_1 + \xi_2 + \cdots + \xi_k)}{I_1} \right\} \frac{1}{\max(I_o, I_1)} c \log c \]

(c) \[r(\text{III}) = -(1+o(1)) \left\{ \frac{k \xi_o}{I_o} + \frac{(\xi_1 + \xi_2 + \cdots + \xi_k)}{I_1} \right\} c \log c \]

where \( r(\delta) \) denotes the Bayes risk with Procedure \( \delta \), \( \delta = \text{I, II, III} \).

**Proof.** Similar to Theorem 5.

**Note.** In view of Theorems 8 and 9 we see that Procedure III is asymptotically optimal.

### 3.8 Table of prices of the procedures

We define \( \rho_\delta \), the price of Procedure \( \delta \) by

\[ \rho_\delta = \limsup_{c \to \infty} \frac{r(\delta)}{-c \log c} \]

and obtain (by use of Theorems 3, 7, and 9) the table (1.5.1) of Chapter I.

### 3.9 Remarks concerning the relative merits of the procedures considered

It is easily verified that the sequential Procedures I, II, and III all have (strictly) smaller prices than the best fixed sample size procedure. If \( 2I_o \leq I_1 \), then \( \rho_{\text{II}} \leq \rho_{\text{I}} \). (That is, if \( 2I_o \leq I_1 \), Procedure I is not better than Procedure II.) Also for \( I_1/I_o \) small \( \rho_{\text{II}} \to \rho_{\text{opt}} \) and \( \rho_{\text{I}} \to \rho_{\text{opt}} \). For \( I_o/I_1 \) small \( \rho_{\text{II}} \to \rho_{\text{opt}} \). Then Procedure II is approximately asymptotically optimal for either \( I_o/I_1 \) small or \( I_1/I_o \) small although the possibility is left open that Procedure II is asymptotically optimal.
APPENDIX

Theorem A.1 (Bounds of the sample mean). Let $X_1, X_2, \ldots$ be independent and identically distributed random variables. Define for a fixed

(1) $P_r = P \left\{ \frac{X_1 + X_2 + \ldots + X_r}{r} \geq a \right\}$

(2) $q_r = P \left\{ \frac{X_1 + X_2 + \ldots + X_r}{n} > a \right\}$

(3) $\phi(t) = E e^{tX_1}$ for all real $t$

(4) $\psi(t) = e^{-at} \phi(t)$ for all real $t$

(5) $T = \left\{ t: -\infty < t < \infty, \phi(t) < \infty \right\}$

If (i) $P(X_1 = a) \neq 1$

(ii) $T$ is a non-degenerate interval

(iii) there exists a positive $\tau$ in the interior of $T$ such that $\psi(\tau) = \inf_{t \in T} \psi(t) = \rho$ (say) then

(a) $p_n \leq \rho^n$ and for every real number $\epsilon$ such that $0 < \epsilon < \psi(\tau)$ for $n$ sufficiently large $p_n \geq (\rho - \epsilon)^n$

(b) for every real number $\epsilon$ such that $0 < \epsilon < \psi(\tau)$ for $n$ sufficiently large $q_n \geq (\rho - \epsilon)^n$. 
Proof. This is essentially due to Chernoff \( \int_{27}^{37} \) but is stated and proved in this form by Bahadur and Rao \( \int_{17}^{17} \).

(b) By an argument similar to that used by Bahadur and Rao in the proof of (a) this result follows.

**Theorem A.2 (Wald's equation).** Suppose that

(a) \( x_1, x_2, \ldots \) are identically distributed

(b) \( E x_1 < \infty \)

(c) \( N \) is a random variable whose values are the positive integers

(d) the event \( \{ N = j \} \) and the random variable \( x_k \) are independent for \( j < k \)

(e) \( E N < \infty \)

then \( E \left( \sum_{j=1}^{N} x_j \right) = (E N) (E x_1) \).

**Proof.** See Johnson \( \int_{47}^{47} \).
BIBLIOGRAPHY


