A MODIFIED BAYES STOPPING RULE

by

Sigmund J. Amster

March 1962

This research was supported by the Office of Naval Research under contract No. Nonr-655(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.
ACKNOWLEDGEMENTS

I am deeply indebted to many individuals for their inspiration and encouragement during my school years. However, it would be an injustice to the important contributions made by those not named, if any except my adviser, Dr. William J. Hall, were to be individually mentioned. Words can hardly convey my most sincere appreciation for his guidance and assistance throughout the course of this investigation.

The Office of Naval Research is to be thanked for their financial support of a portion of this research.

Finally, I wish to thank Mrs. Doris Gardner for her quick and accurate typing of this manuscript and Miss Martha Jordan for her help through the maze of administrative detail. I also wish to thank Mrs. Betty Donaghy for final preparation of the manuscript.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction of the modified Bayes rule (MBR)</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Assumptions</td>
<td>1</td>
</tr>
<tr>
<td>1.3 Notation and definition of MBR</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Calculation</td>
<td>4</td>
</tr>
<tr>
<td>1.5 Subjective probability justification</td>
<td>5</td>
</tr>
<tr>
<td>1.6 Brief literature review</td>
<td>7</td>
</tr>
<tr>
<td>1.7 Chapter summaries</td>
<td>7</td>
</tr>
<tr>
<td>II GENERAL RESULTS</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Notation</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Discussion</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Preliminary lemmas</td>
<td>13</td>
</tr>
<tr>
<td>2.4 Theorem 1: $\alpha_B \leq \alpha_0 \leq \ldots \leq \alpha_0$</td>
<td>16</td>
</tr>
<tr>
<td>2.5 Theorem 2: Convexity of $-\frac{1}{n} (d')$</td>
<td>17</td>
</tr>
<tr>
<td>2.6 Corollary 2.1: Equivalent SPRT</td>
<td>18</td>
</tr>
<tr>
<td>2.7 Theorem 3: $N_\infty \leq N_B$</td>
<td>18</td>
</tr>
<tr>
<td>III SPECIAL CASES</td>
<td></td>
</tr>
<tr>
<td>3.1 Estimation</td>
<td>20</td>
</tr>
<tr>
<td>Theorem 4: Sufficient condition to look only at $R_n(1)$</td>
<td>20</td>
</tr>
<tr>
<td>Asymptotic Minimax comparison</td>
<td>22</td>
</tr>
</tbody>
</table>
Chapter

Binomial estimation computations ........................................ 24

$\alpha_\infty$ ................................................................. 26

$\alpha_j$ ................................................................. 27

3.2 Testing hypotheses .................................................... 29

Calculation of equivalent SPRT ........................................... 29

Two composite hypotheses with linear loss ......................... 31

IV AREAS FOR FUTURE RESEARCH ..................................... 32

BIBLIOGRAPHY ................................................................ 35
CHAPTER I

INTRODUCTION

1.1 Introduction of the modified Bayes rule (MBR)

This thesis describes a stopping rule for sequential sampling which weighs the cost of additional observations against the expected gain to be derived from additional sampling. The modified Bayes rule (MBR) requires one more observation to be taken as long as the posterior risk is larger than the expected posterior risk for any additional fixed-size sample. For the present investigation, risk is defined within the Wald framework of statistical decision theory \cite{15}, using losses and costs. It is recognized however, that in some circumstances other formulations (such as Weiss \cite{16} or Lindley \cite{27}) may be more appropriate.

1.2 Assumptions

To simplify the exposition the following assumptions will be made:

A1: The experiment consists of observing, possibly sequentially, the random variables $X_1, X_2, \ldots$ (real or vector-valued) which are independent with a common probability

\footnote{This research was supported by the Office of Naval Research under contract No. N00014-85-C-0009 for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.}

\footnote{The numbers in square brackets refer to the bibliography.}
density \( f(\cdot | \theta) \) with respect to a given \( \sigma \)-finite measure \( \mu \). The same notation, \( f(\cdot | \theta) \), is used to denote the (joint) density of any number of random variables having the (possibly vector-valued) parameter \( \theta \in \mathcal{L} \).

A2: For any \( \theta \in \mathcal{L} \) and any terminal decision \( d \), the loss \( L(\theta; d) \) is non-negative.

A3: It is assumed that \( c_n^*(y^*_k) \), the "marginal" cost of \( y^*_k \), is increasing in \( k \) (i.e., for all \( x_n \) and all \( y_{k+1} = (y_k, x_{n+k+1}) \), \( c_n^*(y_{k+1}) > c_n^*(y_k) \)); where \( c_r(x_r) \) denotes the cost of observing \( x_r \), \( c_n^*(y_k) \equiv c_n(y_k, x_{n+k})c_n(x_n) \) and \( y_k = (x_{n+1}, \ldots, x_{n+k}) \). Also, for any sequence \( x_1, x_2, \ldots \), it is assumed that \( \lim_{k \to \infty} c_n^*(y_k) = \infty \).

A4: Any measurability assumptions needed to assure the existence (finitely) of the integrals used for the procedure are made.

A5: The existence of a Bayes terminal decision rule is assumed.

The Bayes terminal decision is used exclusively here and is denoted by \( d \).

The identical distribution assumption is made solely to avoid notational complexity. A2 is used primarily to justify the use of the Fubini theorem in proving properties of the MBR. A3 implies that \( \inf_k R_n(k) \) (see Section 1.3) is actually attained. A5 is used to avoid the detailed analysis (e.g., \( \frac{\text{?}}{\text{7}} \) page 297) or restrictive assumptions (e.g., \( \frac{\text{15}}{\text{7}} \) page 69) otherwise necessary to verify the existence of a Bayes terminal decision rule in each particular case.
1.3 Notation and formal definition of the MBR

The following notation will be used:

for any fixed $x_n$; $n = 0, 1, \ldots$, and $k = 0, 1, \ldots$,

\[ R_n(k) = R_n(k)(x_n) = \bar{L}_n(k) + \bar{c}_n(k), \]

where

\[ \bar{L}_n(k) = \int \int \int L(x; \theta) d(x_n, y_k) dF(y_k/\theta) d\xi_n(\theta), \]

\[ \bar{c}_n(k) = \int \int \int c_n^*(y_k) dF(y_k/\theta) d\xi_n(\theta), \]

and

\[ x^k = x \times x \times \ldots \times x \text{ (k times), (x the spectrum of a single observation)}, \]

\[ \xi_0(\theta) = \text{the prior distribution of } \theta, \]

\[ d\xi_n(\theta) = f(x_n/\theta) d\xi_0(\theta)/f_0(x_n), \]

\[ f_0(x_n) = \int f(x_n/\theta) d\xi_0(\theta), \text{ and} \]

\[ dF(y_k/\theta) = \prod_{l=1}^{k} f(y_l/\theta) d\mu(y_l). \]

Note: $R_n(0) = R_n$ is the $U_n(x)$ of $\sqrt{2\gamma}$, page 242;

$R_0(k) = R(k)$, $R_0 = R(0)$ is the $U_0$ in $\sqrt{2\gamma}$.

$R_n(k)$ may be interpreted as our present "best guess" of the posterior risk, $R_{n+k}$, if an additional sample of fixed size $k$ were observed. $R(k)$ is the average risk of a Bayes fixed-sample size (k) procedure.
The formal definition of the MBR:

at the start of sampling (if \( n = 0 \)), or after \( x_n \) is observed
(if \( n > 0 \)),

1. if \( R_n = \inf_k R_n(k) \), stop sampling;

2. if \( R_n \neq \inf_k R_n(k) \), observe \( x_{n+1} \).

(In either case, 1.2 implies that the infimum is actually achieved.)

Note: An equivalent formulation in terms of sets of distributions may sometimes be more useful (see sections 2.5 and 2.6). For each \( n \) a set of distributions \( \{ \frac{z_n}{x_n} \} \) is defined such that sampling stops if and only if \( \frac{z_n}{x_n} \) is in \( \frac{z_n}{x_n} \).

1.4 Calculation

The calculation of \( R_n(k) \) for each \( k \), is feasible whenever the Bayes fixed sample procedure can be explicitly obtained. (Part III of \( 10 \) can offer assistance in the evaluation of the integrals involved.)

On the other hand, an explicit evaluation of the stopping regions for the Bayes sequential rule (BSR) has not been generally possible. Except for certain special cases -- as testing two simple hypotheses, when Wald's sequential probability ratio test (SPRT) is such a rule, or if the BSR is truncated or fixed sample-size -- it is not usually possible to carry out the Bayes procedure. Even in these cases, the determination of the appropriate SPRT is not simple and in the truncated cases the necessary computations are exceedingly tedious. Theorem 9.3.3 in \( 37 \) gives sufficient conditions for the BSR to be truncated or non-sequential.

The MBR calculations are readily adaptable to a high-speed
computer since the analyses before and after a sample has been observed only differ in the change from a prior to a posterior distribution of \( \Theta \) (see section 2.2). The problem is further simplified if the prior and posterior distributions are of the same functional form, with only a change in parameters, since the same kind of calculation is then required at each stage (see section 3.1).

A distribution satisfying this latter property has been called a natural conjugate prior distribution in \( \mathcal{L} \) and a distribution closed under sampling in \( \mathcal{L} \). Some examples are:

<table>
<thead>
<tr>
<th>( f(x/\theta) )</th>
<th>conjugate prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>binomial</td>
<td>beta</td>
</tr>
<tr>
<td>Poisson</td>
<td>gamma</td>
</tr>
<tr>
<td>normal ((\theta, 1))</td>
<td>normal</td>
</tr>
</tbody>
</table>

For the special case of \( c_k(x_k) = kc \) the stopping rule can be conveniently expressed in terms of

\[
\gamma_n(k) = (\bar{L}_n - \bar{L}_n(k))/k.
\]

That is, stop sampling as soon as \( c \geq \gamma_n \equiv \max_k \gamma_n(k) \). In this case it is only necessary to have bounds on \( c \) to carry out the MBR.

1.5 Subjective probability justification

A subjective probability justification for the MBR may be found in interpreting the value of \( z_0 \) at any set of \( \Theta \) values as representing the original relative conviction that the true value of \( \Theta \) lies in this set. Once \( x_n \) has been observed the belief has been changed as re-
lected in the values of \( \xi_n \). In either case the distribution \( \xi \) determines which average of the risk one would like to minimize. The MBR will accept this present average risk (i.e., stop sampling) only if the cost of increasing one's convictions, through knowledge of a sample of any fixed size, is more than the expected amount to be gained. If not, one more sample will be taken and the same problem posed with the (hoped-for) better knowledge of the true state of nature.

In the context treated above, the defining property of the BSR, that of minimizing the original average risk, does not seem particularly relevant. However, the method used to determine this rule, by comparing \( R_n \) with the average risk "resulting from a continuation if at each future stage we did the best we could with the resulting observations" (\( \text{[27]} \), page 243), is really the optimal property. The MBR tries to approximate this by only considering the average risk if any future fixed-size sample were taken.

The averaging over projected observations may be alternatively performed by averaging first with respect to \( f(y_k/x_n) \), the marginal distribution obtained by integrating out \( \Theta \). However, this method is identical with the proposed computation of \( R_{n|k} \). For example if \( \mu \) is Lebesgue measure, and \( F_{\Theta}(\Theta) = f_{\Theta}(\Theta) \ d \Theta \), the following argument, in abbreviated notation, sketches the equivalence:
Let \( H = \int \varphi; d(x_n, y_k) = \int \varphi; d(x, y) \) + \( c_n(y_k) = \int \varphi; d(x, y) \) + \( c(y) \); then

\[
\int \int H g_{n+k} d\varphi dF(y/x) = \int \int H \frac{g(\varphi, y/x)}{f(y/x)} f(y/x) d\varphi d\psi,
\]

\[= \int \int H f(y/\varphi, x) g(\varphi/x) dy d\varphi,
\]

\[= \int \int H f(y/\varphi) g_n(\varphi) dy d\varphi = R_n(k).\]

1.6 Brief literature review

During the investigation of properties of the MBR, the book \( \int 107 \) by Raiffa and Schlaiffer was published. Although the authors specifically refrain from discussing the sequential decision problem, it was found that the preposterior analysis presented is exactly the same as the evaluation of an \( R_n(k) \). Previous to this, Wald \( \int 157 \) page 151, as an example of his general theory, briefly discussed the computation of \( \min_k R_n(k) \) to obtain the optimal second stage sample size. In \( \int 157 \), Chernoff introduced a sequential procedure based upon an asymptotic approach where the cost per observation goes to zero. In \( \int 177 \), Hoeffding obtained bounds for the average risk of an arbitrary sequential test. Many other writers such as Anscombe \( \int 117 \), Barnard \( \int 27 \), Bross \( \int 47 \), Lindley \( \int 27 \), and Wetherill \( \int 177 \) have contributed to the literature in this area. An excellent summary of the current state of this research is given in \( \int 57 \) by Johnson.

1.7 Chapter summaries

In Chapter II by defining a sequence of stopping rules having the MBR as a limit, the average risk for the MBR is found to be the limit of a non-increasing sequence whose initial value is that for the fixed-
sample size Bayes procedure. Since the average risk of the MBR is, of course, not less than that of the BSR, the two rules coincide if the BSR is actually a fixed-sample size rule. It is shown that the MBR is a SPRT for the problem of testing two simple hypotheses. It is also shown that the actual sample size required by the MBR is never larger than that for the BSR. Therefore, if one were to use a BSR, the computations to perform it would need to be started only after the termination of the MBR. At this point of termination, the improvement possible over the MBR is identical to the difference between the average risk of a fully sequential Bayes procedure and that of a fixed-sample size Bayes procedure in which the sample size is zero, the expectations taken with respect to the posterior distribution of $\theta$.

In Chapter III an easily satisfied condition is given for

$$\min_k R_n(k) = R_n(1)$$

in an estimation problem, where loss equals squared error, $c_k(x_k) = kc$, $f(\cdot | \theta)$ is a member of the exponential class (see section 3.1) and a conjugate prior distribution is used. When this condition holds, only a single $R_n(k)$ needs to be compared with the posterior risk to determine whether additional sampling should be performed. In addition, the asymptotic minimax rule $k_{\text{lim}}$ and MBR are compared for several estimation examples. To demonstrate the calculations needed to perform a MBR and to determine its average risk, a binomial estimation problem is presented in some detail. Although the BSR requires a fairly complicated "working-backward" method, which is apparently not feasible by hand computation, the MBR requires only the solution of several second degree equations. An example is also pre-
sented in which the particular SPRT equivalent to the MBR is found for a two simple hypotheses testing problem with $c_k(x_k) = kc$ and simple loss function.

Chapter IV contains several areas where future research may be profitable.
CHAPTER II
GENERAL RESULTS

2.1 Notation

The following additional notation will be used in this chapter:

\( \delta_B \): Bayes sequential rule (BSR).

\( \delta_j \): stopping rule which follows the MBR up to and including a sample of size \( j \); if \( R_k \neq \min_t R_k(t) \) for \( k = 0, 1, \ldots, j \) then a sample of fixed size \( m \) is taken, where \( m \) is the smallest \( t \) such that \( R_j(t) = \min_i R_j(i) \). \( \delta_0 \) is thus the fixed sample-size Bayes rule and \( \delta_\infty \) the MBR.

\( \delta_j(x_t) \): stopping rule of type \( \delta_j \) started after \( x_t \) is observed (i.e., if sampling is continued beyond \( x_{t+j} \), it is with a single fixed-size sample).

\( \alpha_B \): average risk of \( \delta_B \).

\( \alpha_B(N) \): average risk of Bayes rule truncated after \( N \) observations.

\( \alpha_j \): average risk of \( \delta_j \).

\( \alpha_j(x_t) \): posterior average risk of \( \delta_j(x_t) \), using \( \delta_t \).

\( N_B \): sample size using \( \delta_B \), a random variable.

\( N_j \): sample size using \( \delta_j \), for \( j = 0, 1, \ldots, \infty \), a random variable for \( j > 0 \).

\( E_1 \): operator taking expectation with respect to \( x_{i+1} \), holding \( x_1 \) fixed; specifically for any \( h(x_{i+1}) \),
\[ E_{i} h(x_{i+1}) = \int h(x_{i+1}) f_{i}(x_{i+1}) \, d\mu(x_{i+1}) \]

where

\[ f_{i}(x_{i+1}) = \int f(x_{i+1}/\theta) \, d\pi_{i}(\theta). \]

A_t: \{x_t: N_{\infty} = t\}, stopping region for MBR.

2.2 Discussion

Several properties follow immediately from the definitions. For any \( n, \delta_{n+1} \) will differ from \( \delta_n \) (if and only if) the sequence of observations \( x_n \) is such that \( R_n > \min_k R_n(k) \). A necessary condition for the two rules to differ is therefore that \( N_n > n \). Also, as soon as a particular \( x_n \) has been observed, the rule \( \delta_k(x_n) \) will act exactly the same as the rule \( \delta_{n+k} \). For example, using the rule \( \delta_n(n \geq 1) \), the observing of \( x_1 \) changes \( \delta_0 \) into \( \delta_1 \) and hence \( \delta_n \) into \( \delta_{n+1}(x_1) \). \( \delta_{n+1}(x_1) \) may be regarded as using a different prior distribution for a new decision problem, where the truly sequential portion has been reduced by one.

The additional properties shown in this chapter are based primarily on the recognition that:

\[ \alpha_B^{(N)} = \min \{ R_0, E_0 \min \{ R_1, E_1 \min \ldots E_{N-2} \min \{ R_{N-1}, E_{N-1} R_N \} \} \ldots \} \]

\[ \alpha_O = \min \{ R_0, E_0 R_1, E_0 E_1 R_2, \ldots \} \]

\[ \alpha_O(x_1) = \min \{ R_1, E_1 R_2, E_1 E_2 R_3, \ldots \} \]

\[ \alpha_1 = \begin{cases} R_0 & \text{if } \alpha_O = R_0 \\ E_0 \alpha_O(x_1) & \text{if } \alpha_O < R_0 \end{cases} \]

and
\[ \alpha_k = \begin{cases} R_0 & \text{if } \alpha_0 = R_0, \\ E \alpha_{k-1}(x_1) & \text{if } \alpha_0 < R_0, \end{cases} k > 1. \]

These expressions follow almost directly from the definitions, but see the lemmas, except for the first which is proved in \[37\].

To make clear the distinction between \( \delta_B \) and \( \delta_\infty \), assume that the use of \( \delta_B \) had resulted in the sequence \( x_n \). Then \( N_B = n \) if and only if

\[ (1) \quad R_n \leq \min \{ E_n, R_{n+1}, E_{n+1} \} \sqrt{\min \{ R_{n+2}, E_{n+2} \} \min(\ldots)} \ldots \].

Whereas if \( \delta_\infty \) had led to the same sequence of observations, \( N_\infty = n \) if and only if

\[ (2) \quad R_n \leq \min \{ E_n R_{n+1}, E_n E_{n+1} R_{n+2}, \ldots \}. \]

(The underscores are used for emphasis.) This, together with the fact \( E_1 \min(\ldots) \leq \min E_1(\ldots) \), is the basis for the proof of Theorem 3.

If an \((n+1)\)st observation is taken, both rules put \( R_{n+1} \) (now known) on the left-hand side and delete \( E_n \) and \( R_{n+1} \) from the right-hand side, etc. However, to actually obtain the right-hand side of \( (1) \) requires a backward induction, while in \( (2) \) a forward induction will suffice.

To indicate in more detail the computation of \( \alpha_k \), when \( x_i \) can take only a finite number of values, the case \( k = 3 \) will be examined. (See Chapter III for a numerical example.)

If \( \alpha_0 = R_0 ; \alpha_3 = R_0 \), since no sampling takes place.
If $\alpha_0 < R_0$:

1°. For each $x_1$, find $\alpha_0(x_1) = \min \sum R_1, R_1(1), R_1(2), \ldots$; then

$$\alpha_1 = E_o \alpha_0(x_1) = \sum_{x_1} \alpha_0(x_1) f_o(x_1).$$

2°. For each $x_2$, such that $x_1 \in \bar{A}_1 = \{x_1: \alpha_1 < R_1\}$,

$$\alpha_0(x_2) = \min \sum R_2, R_2(1), R_2(2), \ldots$$

then

$$\alpha_2 = \sum_{A_1} \alpha_0(x_1) f_o(x_1) + \sum_{A_1} [\sum_{x_2} \alpha_0(x_2) f_o(x_2)] f_o(x_1).$$

3°. For each $x_2$, such that $x_1 \in \bar{A}_1$,

$$\alpha_1(x_2) = \begin{cases} R_2 & \text{if } \alpha_2 = R_2 \\ E_o \alpha_0(x_3) & \text{if } \alpha_2 < R_2 \end{cases}$$

Then

$$\alpha_2(x_1) = \begin{cases} R_1 & \text{if } \alpha_1 = R_1 \\ E_o \alpha_1(x_2) & \text{if } \alpha_1 < R_1 \end{cases}$$

and

$$\alpha_3 = \begin{cases} R_o & \text{if } \alpha_o = R_o \\ E_o \alpha_2(x_1) & \text{if } \alpha_o < R_o \end{cases}$$

The repetitive nature of the above formulas imply a ready adaptability to programming for high-speed computers. It is noted that the computation of $\alpha_k$ is not necessary to perform (but only to evaluate) $a \delta_k$.

2.3 Preliminary lemmas

**Lemma 1**: For all $k \geq 1$, $R(k) = E_o E_1 \cdots E_{k-1} R_k$. 
Proof:

Let \( H = H(x_k) = L \cdot \mathcal{G}_k \cdot d(x_k) \cdot I + c_k(x_k) \). Then

\[
E_{k-1} \cdot R_k = \int_{x_k} \left( \int_{x_{k-1}} f(x_k \mid \theta) \, d \xi_{k-1} \right) \, d \mu(x_k),
\]

for \( x_k \in x_k, \omega' \in \mathcal{A}' \)

\[
= \int_{x_k} \int_{x_{k-1}} \frac{f(x_k \mid \theta)}{f_o(x_k)} \, \frac{f(x_{k-1} \mid \theta)}{f_o(x_{k-1})} \, d \mu(x_k)
\]

\[
= \int_{x_k} \int_{x_{k-1}} H \cdot f(x_k \mid \theta) \, d \xi_{k-1} \, d \mu(x_k)
\]

Analogously, by applying \( E_{k-2}, E_{k-3}, \ldots, E_0 \) successively

\[
E_0 E_1 \cdots E_{k-1} R_k = \int_{x_k} \int_{x_{k-1}} H \cdot f(x_k \mid \theta) \, d \xi_{k-1} \mid \theta \, d \mu(x_k)
\]

\[
= \int_{x_k} \int_{x_{k-1}} H \cdot d \xi_{k-1} \mid \theta \, d \mu(x_k)
\]

\[
= R(k) \quad \text{since} \quad c_k(x_k) = c_o(x_k).
\]

Lemma 2: If \( \alpha_0 < R_0 \); then for all \( k \geq 1 \), \( \alpha_k = E_0 \alpha_{k-1}(x_1) \) and \( \alpha_1 = \mathbb{E}_{0} \min \{ R_1, E_1 \cdot E_2 \cdot E_1 \cdot E_2 \cdot R_2, \ldots \} \).

Proof:

1. Given any particular \( x_1 \), the stopping rules \( \delta_k \) and \( \delta_{k-1}(x_1) \)
are exactly the same \( k \geq 1 \) since both use \( \xi_1 \) as the distribution of \( \Theta \) to determine whether or not \( R_1 = \min_k R_1(k) \).

Therefore the prior average risk, averaging with respect to \( x_1 \), is the same.

2°. If \( \alpha_0 < R_0 \), \( \delta_1 \) requires the taking of a single observation \( (x_1) \) and then either making a terminal decision

(if \( R_1 = \min_k R_1, E_1 R_2, \ldots \) or else taking a sample of

fixed size \( m \) where \( E_1 \ldots E_m R_{m+1} \leq \min_k E_1 \ldots E_k R_{k+1} \). In

either case, \( \min_k R_1, E_1 R_2, \ldots \) is the average risk conditional on \( x_1 \). The prior average risk is then obtained by
taking the (prior) expected value of this average risk.

3°. The same argument, used in the above proof, also shows that

\[ \alpha_1(x_1) = E_1 \{ \min_k R_2, E_2 R_3, E_2 E_3 R_4, \ldots \} \]

if \( x_1 \) is not in

\( A_1 \), since the only change required is that \( \xi_0 \) becomes \( \xi_1 \).

Lemma 3:

For any fixed \( m \), let \( \xi_m = \lambda \xi_1^* + (1 - \lambda) \xi_m^* \), \( \lambda \in [0, 1] \), and
define \( \alpha_0'(x_m), \alpha_0''(x_m), \ldots \), as the \( \alpha_0, \ldots \), if \( \xi_1^*, \xi_m^* \) is

used. Then, \( \alpha_0(x_m) \geq \lambda \alpha_0'(x_m) + (1 - \lambda) \alpha_0''(x_m) \).

Proof:

It is sufficient to prove the case of \( m = 0 \). Define the integer

1 by \( \alpha_0 = R(1) = \min_k R(k) \). Then
\[ \lambda \alpha_0 + (1-\lambda) \alpha''_0 = \lambda \min_k R'_k + (1-\lambda) \min_k R''_k \]
\[ \leq \lambda R'_1 + (1-\lambda) R''_1 \]
\[ = R'_1 = \alpha_0. \]

2.4 **Theorem 1**: \[ \alpha_B \leq \alpha_0 \leq \ldots \leq \alpha_n \leq \ldots \leq \alpha_1 \leq \alpha_0. \]

**Proof:**

1°. \[ \alpha_1 = \begin{cases} R_0 & \text{if } \alpha_0 = R_0 \\ \min \{E_0 \alpha(x_1), E_0 \alpha(x_2), \ldots \} & \text{if } \alpha_0 < R_0 \end{cases} \]

\[ = \min_k R_k = \alpha_0. \]

2°. \[ \alpha_2 \leq \alpha_1; \text{ if } \alpha_0 = R_0, \alpha_0 = \alpha_1 = \alpha_2 \text{ since no sampling is performed. If } \alpha_0 < R_0, \alpha_2(x_1) \leq \alpha_0(x_1) \text{ since } (1°) \text{ was proved for any } \xi_0 \text{ and is thus true for } \xi_1. \text{ Taking expectations, } E_0 \alpha_2(x_1) \leq E_0 \alpha_0(x_1), \text{ and therefore } \alpha_2 \leq \alpha_1 \text{ by Lemma 2.} \]

3°. \[ \alpha_2(x_1) \leq \alpha_1(x_1) \text{ since } (2°) \text{ was proved for any } \xi_0 \text{ and is thus true for } \xi_1. \]

4°. \[ \alpha_3 \leq \alpha_2 \text{ follows by the method used in the second part of } (2°), \text{ using } (3°) \text{ of Lemma 2.} \]

5°. \[ \alpha_m \leq \alpha_{m-1} \text{ follows by a repetition of } (3°) \text{ and } (4°) \text{, i.e., } \alpha_3(x_1) \leq \alpha_2(x_1), \text{ etc.} \]
\[ S^0. \quad \alpha_B \leq \alpha_{\infty} \text{ is the fundamental property of } \mathcal{B}. \]

2.5 Theorem 2: (see Theorem 9.4.3 of [37])

If the \( n \) is finite, then \( \mathcal{Z}_n(d') \) is convex. Where

\[ \mathcal{Z}_n(d') = \{ \xi_n : \mathcal{A}_n \text{ is in } \mathcal{A}_n \text{ and } d = d' \} \]

is such that \( \xi_n \) is in \( \mathcal{Z}_n(d') \) if and only if the MBR requires sampling to stop at stage \( n \) and a particular terminal decision \( (d') \) to be made, and

\[ \mathcal{Z}_n = \bigcup_{d'} \mathcal{Z}_n(d') \] (see section 1.3).

Proof:

Let \( \mathcal{A}_n = (\Theta_i)^m \), \( H(i, d) = \sum_{\Theta_i} d(x_n) + c_n(x_n) \) for \( i = 1, \ldots, m \), and \( \xi_n(i) = \xi_n(\Theta_i) = \lambda \xi_n'(\Theta_i) + (1-\lambda) \xi_n''(\Theta_i) \)

for \( 0 \leq \lambda \leq 1 \). Fix \( x_n \) such that both \( \xi_n' \) and \( \xi_n'' \) are in \( \mathcal{Z}_n(d') \). If \( n = 0 \), interpret \( \alpha_0(x_n) = \alpha_0 \), etc. Now

\[
\lambda \sum_{l} \xi_n'(i) H(i, d') + (1-\lambda) \sum_{l} \xi_n''(i) H(i, d')
\]

= \( \lambda \alpha_0'(x_n) + (1-\lambda) \alpha_0''(x_n) \) since both \( \xi_n' \) and \( \xi_n'' \) are in \( \mathcal{Z}_n(d') \)

\[ \leq \alpha_0(x_n) \] Lemma 3

\[ \leq \mathcal{R}_n \] since \( \alpha_0(x_n) = \min_k \mathcal{R}_n(k) \)

= \( \min_{d'} \sum_{l} \xi_n(i) H(i, d') \leq \sum_{l} \xi_n(i) H(i, d') \)

= \( \lambda \sum_{l} \xi_n'(i) H(i, d') + (1-\lambda) \sum_{l} \xi_n''(i) H(i, d') \).

Therefore, \( \alpha_0(x_n) = \mathcal{R}_n = \sum_{l} \xi_n(i) H(i, d') \); i.e., \( \xi_n \) is in \( \mathcal{Z}_n(d') \).
2.6 Corollary 2.1:

If $m = 2$, $c_n(x_n) = n$ and $L(i,a) = 1$ for $a \neq i$ and zero otherwise, then there exists an SPRT equivalent to the HBR.

Proof:

It is shown in page 267, that the convexity of $\Sigma_{-n}(d)$ is sufficient for this conclusion.

Note: It may be shown, following page 259, that if, for any $x_n$, the marginal cost $c_n(x_n)$ depends only on the values in the vector $x_n$, the sets $\Sigma_{-n}(d)$ do not depend on $n$. This result does not require the finiteness of $\Sigma_\infty$. For example, $c_n(x_n) = \Sigma c(x_i)$, where $c(x_i)$ is the cost of the observation $x_i$, is sufficient.

2.7 Theorem 3: $N_\infty \leq N_B$.

Proof:

For any $N$, at the $n$-th stage, $\delta_0(N)$ stops sampling (page 242) if and only if

$$R_n \leq E_n \min \{ R_{n+1}, E_n R_{n+2}, \ldots, E_{n-1} R_{n-n} \} \leq \min \{ \min \{ R_{n+1}, E_n R_{n+2}, \ldots, E_{n-1} R_{n-n} \} \}$$

Hence, if $\delta_B$ stops sampling at the $n$-th stage, then

$$R_n \leq \min \{ R_{n+1}, R_n(2), \ldots, R_n(n-n) \} = \min_k R_n(k)$$

i.e., $\delta_\infty$ stops sampling.
Corollary 3.1: Suppose $c(x_k) = kc$.

(i) If $\overline{L}_n \to 0$ uniformly in $x_n$, then $c_\infty$ is truncated.

(ii) If $\overline{L}_n$ is a function of $n$ only, then $\delta_\infty$ is the Bayes fixed-sample size procedure.

Proof:

(i) By $\overline{L}$, Theorem 9.3.3, $\delta_B$ is truncated. But $N_\infty \leq N_B$.

(ii) By the same theorem, $\delta_B$ is a fixed-sample size procedure.

But, using Theorem 1, $\alpha_B \leq \alpha_\infty \leq \alpha_\alpha = \alpha_B$. 
CHAPTER III

SPECIAL CASES

3.1 Estimation

For the exponential class of densities, \( f(x/\theta) = \alpha(\theta)\beta(x)\exp[\theta\gamma(x)] \), the parameters in the prior distribution combine with the sample data in a particularly simple manner when a natural conjugate prior distribution (see section 1.4) is used. In \( \int_{-\infty}^{\infty} \), page 45, the binary operation * is defined; for the case described above, the definition reduces to

\[
(\theta_0', m') \ast (\bar{x}_n, n') = (n' \frac{\bar{x}_n + m'\theta_0}{n' + m'}, \bar{x}_n + m')
\]

where \( \theta_0' \) is the prior estimate of \( \theta \) associated with \( m' \) pseudo-observations: \( (\theta_0', m') \) are prior distribution parameters, \( (\bar{x}_n, n') \) are sample sufficient statistics, and the terms on the right are posterior distribution parameters, e.g., if \( d \xi_0(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \ d \theta \) and \( f(x/\theta) = \theta^x(1-\theta)^{1-x} \), then

\[
(a/(a+b), a+b) \ast (\bar{x}_n, n) = (\bar{x}_n + a)/\left(\frac{n}{\sum_{i=1}^{n} x_i + a}\right)/\left(n+a+b\right), \ n+a+b).
\]

The following theorem gives a sufficient condition for "looking ahead" at only one potential future observation.

Theorem 4:

If: \( f(x/\theta) \) is a member of the exponential class of densities, \( L(\theta; d) = (d - \theta)^2 \), \( c_k(x_k) = kc \), and \( g_0(\theta) \) is a natural conjugate of \( f(x/\theta) \), where \( d \xi_0(\theta) = g_0(\theta) \ d \theta \),

Then: \( g \) stops sampling as soon as \( c \geq \gamma_n(1) \) (see section 1.4), if and only if
(3) \[ J_n \leq \frac{1}{2} \sqrt{\frac{3(n' + m')} {L_n}} + 1 \sqrt{\frac{1}{L_n}} , \]

where \( J_n = \int \sigma^2 \, \xi_n(\theta) \) and \( \sigma^2 = \int \frac{(x - \theta)^2}{\xi} f(x/\theta) \, dx \).

Proof:

For any fixed vector \( x_n + k = (x_n', y_k) \), it is easily shown that \( d(x_n', y_k) = \int (n' + m') \xi_n + \sum \frac{y_{i-1}}{n' + m' + k} \), where \( \xi_n = \int \xi d \xi_n(\theta) \), the mean of the posterior distribution of \( \theta \) at the \( n \)-th stage. Consequently, by definition

\[ L_n(k) = \int \int \left[ d(x_n', y_k) - \xi_n \right] f(y_i/\xi_n) \, dy_i \, g_n(\xi) \, d\theta . \]

By performing the integration, \( (n' + m' + k)^2 L_n(k) = (n' + m')^2 L_n + k J_n \), and therefore \( \gamma_n(k) = \int \frac{2(n' + m') + k}{L_n} \, L_n - J_n \} / \{ n' + m' + k \}^2 \).

Assuming that \( k \) is a continuous variable,

\[ \frac{d}{dk} \gamma_n(k) = \left[ 2 J_n - \int \frac{3(n' + m') + k}{L_n} \right] / \{ n' + m' + k \}^3 , \]

which is less than zero, for all \( k \), if and only if (3) is satisfied.

Since \( R_n \leq R_n(k) \) if and only if \( L_n \leq L_n(k) + kc \), or

\[ c \geq \gamma_n(k) , \]

and (3) implies \( \gamma_n(1) \geq \gamma_n(2) \geq \ldots \); (3) implies that \( c \) stops sampling if and only if \( c \geq \gamma_n(1) \).

In each of the examples considered, (normal, binomial and Poisson), the ratio \( J_n/L_n \) was found to be independent of \( x_n \), but no proof was obtained that this is a general property. There may be some connection with Theorem 9.3.3 of \( \int \frac{d}{dx} \). For example in the binomial case,
\[ J_n / \bar{L}_n = \int (a + \Sigma_{1} x_1^n) / \lambda \lambda + 1)^2 \int (a + \Sigma_{1} x_1^n) / \lambda \lambda + 1)^2 \]

= \lambda / (\lambda + 1), \text{ where } \lambda = a + b + n.

Whenever this is true, before sampling begins you know at what stage only \( \gamma_n(1) \) needs to be examined.

Asymptotic Minimax Comparison

Wald's \( \int_{-1}^{1} \) introduced the Asymptotic Minimax (AM) sequential estimates for \( c_k = kc \) and \( L(\theta; d\theta) = (d\theta - \vartheta)^2 \). "Asymptotic" refers to the limit as \( c \) tends to zero. By means of Wald's extension of the Cramer-Rao lower bound \( \int_{-1}^{1} \), he showed that \( T^c_0 \) and \( T^c_1 \) are both AM solutions, where

\( T^c_0: \) Take \( N_c \) observations, \( N_c \geq (c \bar{\vartheta})^{-1/2} \), and estimate \( \vartheta \) by \( \hat{\vartheta}_{N_c} \).

\( T^c_1: \) stop sampling as soon as \( c \geq \gamma_n = \int n(n+1) \bar{\vartheta}(\hat{\vartheta}_n) \lambda^{-1} \)

and estimate \( \vartheta \) by \( \hat{\vartheta}_n \).

Where \( \hat{\vartheta}_n \) is the maximum likelihood estimator of \( \vartheta \), given \( x_n, d* = \hat{\vartheta}_n, \bar{\vartheta}(\hat{\vartheta}) = E \left[ \frac{\partial \log f(y/\theta)}{\partial \theta} \right] / \theta \), and

\( \bar{\vartheta} = \inf \hat{\vartheta}(\theta) \).

In a sense explained by Wald's \( \int_{-1}^{1} \), \( T^c_1 \) is better (for small \( c \)) than \( T^c_0 \).

Since the AM procedures are among the very few general sequential estimating procedures where explicit solutions are available, a comparison with the MBR may be of interest. The similarity of the stopping criteria and estimators, as shown in Table 1, is quite striking in view of the apparent disparity between the two methods. In fact, for large \( n \), the procedures effectively coincide.
## TABLE 1

<table>
<thead>
<tr>
<th>$f(x/\theta)$</th>
<th>$g_0(\theta)$</th>
<th>$\gamma_n^1$</th>
<th>$\gamma_n(1)$</th>
<th>$\hat{\theta}_n$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(\theta,1)$</td>
<td>$N(\theta_0, \tau^{-2})$</td>
<td>$\frac{n(n+1)}{n}$</td>
<td>$\frac{((n+\tau^2)(n+\tau^2+1))^{-1}}{n}$</td>
<td>$\bar{x}_n$</td>
<td>$\frac{\bar{x}_n + (\tau^2 \theta_0/n)}{1 + (\tau^2/n)}$</td>
</tr>
<tr>
<td>$B(\theta)$</td>
<td>$\beta(a, b)$</td>
<td>$\frac{\bar{x}_n (1-\bar{x}_n)}{n(n+1)}$</td>
<td>$\frac{(\bar{x}_n + a/n)(1-\bar{x}_n + b/n)}{n}$</td>
<td>$\bar{x}_n$</td>
<td>$\frac{\bar{x}_n + (a/n)}{1 + ((a+b)/n)}$</td>
</tr>
<tr>
<td>$P(\theta)$</td>
<td>$P(\theta_o)$</td>
<td>$\frac{\bar{x}_n}{n(n+1)}$</td>
<td>$\frac{\bar{x}_n + (\theta_0/n)}{(n+1)^2 (1 + (2/n))}$</td>
<td>$\bar{x}_n$</td>
<td>$\frac{\bar{x}_n + (\theta_0/n)}{1 + (1/n)}$</td>
</tr>
</tbody>
</table>

where: $B(\theta)$ is the binomial distribution, $P(\theta)$ is the Poisson, $\beta(a, b)$ is the beta, and

$$P(\theta_o) \propto e^{-t} t^{\theta_0 - 1}$$
Binomial estimation computations

To indicate two possible forms that the stopping regions for the NBR may take, and to show the numerical computations required to obtain them, two binomial examples will be considered. The second will be given in considerable detail to include the computation of several \( \alpha_k \)'s. The form of the BSR for these examples is not known, and hand computation does not seem feasible. It is only known that they are truncated \( \sqrt{\alpha} \), at a value not less than 27 and 12 respectively (Corollary 3.1), and that the stopping boundaries are nowhere within those of the NBR (Theorem 3).

Example 1:

\[ L(\theta; d) = (d-\theta)^2, \quad c_k(x_k) = k/4000 \quad \text{and} \quad g_0(\theta) = \beta(2, 2) = \theta(1-\theta). \]

Let: \[ s = \sum_{1}^{n} x_i, \quad r' = (n+4)^2 (n+5)^2. \]

From Table 1, \( N = n \) if and only if \( 1/4000 \geq (s+2)(n-s+2)/r', \)

i.e., \[ s \geq \left( \frac{1}{2} \right) \left( n + (n+4) \sqrt{1 - ((n+5)^2/1000)^2} \right)^{1/2} = \sigma_n. \]

In the \( (n, s) \) plane, the stopping regions are indicated by the cross-hatching in Figure 1.
Example 2:

\[ L(\varnothing; d) = (\bar{a} - \varnothing)^2, \quad c_k(x_k) = k/2000 \quad \text{and} \quad g_n(\varnothing) = \mathbb{B}(9, 1). \]

Using Table 1, \( N_n = n \) if and only if

\[ s \geq \left( \frac{1}{2} \right) \left[ (n-8) + (n+10) \sqrt{1 - ((n+11)^2/500)} \right]^{1/2} = \sigma_n. \]

For this example, the stopping region is shown in Figure 2.

![Figure 2](image)

In this example the prior belief that \( \varnothing \) is large is sufficiently strong (relative to the cost per observation) that sampling cannot stop for "too few" successes.

Computation of \( \alpha_\infty (= \alpha_{12}) \)

Let \((n, \lambda_k)\) represent the coordinates in the \((n, s)\)-plane of the accessible stopping points where \( k = 1, 2, \ldots, 12 \) indexes these points. These appear below in Table 2.

Let \( t(y, z) = \text{number of distinct paths in the } (n, s) \text{ plane, from } (0, 0) \text{ to } (y, z) \text{ which do not pass through a } (n, \sigma_n). \) Several methods are available in the literature, e.g., see \( \text{pdf} \), for finding \( t(y, z) \). If
there were no boundary,

\[ t(y,z) = t(y-1,z) + t(y-1,z-1) \text{ for } y \geq z \geq 1 \]

\[ t(y,z) = \begin{cases} 
0 & y < z \\
1 & y = z 
\end{cases} \]

could be used. By removing the contributions from each \( t(n,\sigma_n) \), this recursive method is easily adapted to the present situation. Letting \( t_k = t(n,\lambda_k) \), where \( n = n(k) \), the values in Table 2 were obtained.

Let \( d_k \) be the Bayes terminal decision, if sampling stops at the point \( (n, \lambda_k) \). Since this is the mean of the posterior distribution, (see Table 1), \( d_k = (9 + \lambda_k)/(n + 10) \).

\[ \begin{array}{c|ccccccccccc}
\text{n} & 2 & 5 & 7 & 8 & 9 & 10 & 11 & 12 & 12 & 12 & 12 \\
\hline
k & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
\hline
\lambda_k & 2 & 4 & 5 & 5 & 5 & 5 & 4 & 4 & 3 & 2 & 1 & 0 \\
\hline
t_k & 1 & 2 & 7 & 23 & 53 & 103 & 292 & 156 & 65 & 12 & 1 \\
\hline
d_k & \frac{11}{12} & \frac{13}{15} & \frac{14}{17} & \frac{14}{18} & \frac{14}{19} & \frac{14}{20} & \frac{13}{21} & \frac{13}{22} & \frac{12}{22} & \frac{11}{22} & \frac{10}{22} & \frac{9}{22} \\
\end{array} \]

By definition,

\[ \alpha_n = 9 \sum_{1}^{12} \left( (d_k - \Theta)^2 + c \cdot n(k) \right) p_\Theta(k) d\Theta, \]

where

\[ p_\Theta(k) = P \int_{-\infty}^{x_n} s = \lambda_k / \Theta, \quad \lambda_k = t_k \Theta^k (1 - \Theta)^{n-\lambda_k} \]
After performing, with the aid of Table 2, the operations indicated,
\[ \alpha_\infty = .00623 + .00138 = .00761 \]

Computation of \( \alpha_j \)

For any \( x_n \),
\[ g_n(\varrho) = \frac{\varrho^{s+i}(1-\varrho)^{n-s}}{\varrho^{s+9} (n+1-s)} \quad \text{where} \quad s = \sum_{i=1}^{n} x_i \]

Also,
\[ R_n(k) = \int_{0}^{1} \int_{0}^{k} \left\{ \inf \int \varrho^{s+i}(1-\varrho)^{n-s-k} \right\} \frac{(s+i)(n-s-k-i)}{(n+k+10)^2 (n+k+11)} \varrho^{i} (1-\varrho)^{k-i} g_n(\varrho) + ck \]

where \( g_{n+k}(\varrho) = \frac{\varrho^{s+i}(1-\varrho)^{n-s-k-i}}{\varrho^{s+9} (n+s+k-i+1)} \) \varrho,

and therefore,
\[ R_n(k) = \int_{0}^{1} \int_{0}^{k} \frac{(s+i)(n-s-k-i)}{(n+k+10)^2 (n+k+11)} \varrho^{i} (1-\varrho)^{k-i} g_n(\varrho) + ck \]
\[ = \frac{(s+9)(n-s+1)}{(n+10)(n+11)(n+10+k)} + ck. \]

Assuming that \( k \) is a continuous variable, \( \frac{d}{dk} R_n(k) = 0 \) if and only if

\[ k = k' = \left\lfloor \sqrt{\varrho} \right\rfloor = \text{integer closest to } \varrho. \]

\(^1\circ\) \( \alpha_o = \min_k R(k) \)

Since \( k' = \left\lfloor \sqrt{\frac{2000(9-0)}{(10)(11)}} \right\rfloor^{1/2} - 10 \]
\[ = 3, \]
\[ \alpha_o = R(3) = .00629 + .00150 = .00779. \]
2°. \( \alpha_1 = \sum_{x_1} \alpha_0(x_1) f_0(x_1) + c \)

For \( x_1 = 0 \), \( k' = \left[ \left( \frac{2000(18)}{(11)(12)} \right)^{1/2} - 11 \right] \) = 6, and hence
\( \alpha_0(0) = R_1(6) \).

For \( x_1 = 1 \), \( k' = \left[ \left( \frac{2000(10)}{(11)(12)} \right)^{1/2} - 11 \right] \) = 1, and hence
\( \alpha_0(1) = R_1(2) \).

Also, \( f_0(0) = 1 - f_0(1) = 9 \int_0^1 (1-\theta) \theta^8 d\theta = 1/10 \).
\( \alpha_1 = 1/2000 + (9/10)\alpha_0(1) + (1/10)\alpha_0(0) = 0.00648 + 0.00125 = 0.00773 \).

3°. \( \alpha_2 \)

For \( x_2 = (0,0) \), \( k' = 7 \), for \( x_2 = (0,1) \) or \( (1,0) \), \( k' = 4 \) and
for \( x_2 = (1,1) \), \( k' = 0 \). Also \( f_0(1,1) = 9 \int_0^1 e^{10} d\theta = 9/11 \),
\( f_0(0,1) = f_0(1,0) = 9 \int_0^1 e^9(1-\theta) d\theta = 9/110 \) and therefore
\( \alpha_2 = \frac{2}{2000} + \sum_{x_2} \alpha_0(x_2) f_0(x_2) = 0.00629 + 0.00138 = 0.00767 \).

4°. \( \alpha_3 \)

For \( x_3 = (1,0,1) \) or \( (0,1,1) \), \( k' = 3 \), for \( x_3 = (1,0,0) \) or \( (0,1,0) \) or \( (0,0,1) \), \( k' = 5 \) and \( k' = 7 \) for \( x_3 = (0,0,0) \). Note that
\( x_2 = (1,1) \in A_2 \). \( f_0(1,0,1) = 3/44 \), \( f_0(0,0,1) = 3/220 \),
\( f_0(0,0,0) = 1/220 \), etc.

\( \alpha_3 = \left( \frac{9}{11} \right) \int x_3 (1,1) + 2/2000 \sum_{x_3} \int x_3 + 3/2000 \sum f_0(x_3) \)

\( = 0.00626 + 0.00141 = 0.00767 \).

By carrying more decimal places, it is found that \( \alpha_2 = 0.007675 \)
and \( \alpha_3 = 0.007671 \).
Since the stopping point (2,2) contributes .00563 to each $\alpha_j$, the reduction possible as compared with $\alpha_0$ is rather small in this example. It is noted that although the $\alpha_j$'s are monotonic, the contributions from expected cost and loss are not.

3.2 Testing hypotheses

The testing of two simple hypotheses is another problem to which the MBR may be applied. By Corollary 2.1, for $c_n(x_n) = nc$ and $L(i, d.) = 0$ if $d_i = i$ and $\parallel d$, if $d \neq i$, there exists an equivalent SPRT. From Theorem 3, the particular SPRT must have boundaries at, or within, the SPRT corresponding to the Bayes solution. Since the BSR for this problem is available (by an iterative process $\int^3_7$), an explicit comparison of the two procedures may be possible. No other analytic comparison of the SPRT’s has proven feasible. By the use of Wald's well-known formulas $\int^3_7$, approximate expressions for the OC function and ASN are obtainable for the MBR in this case.

The actual use of the MBR is not dependent on establishing the equivalent SPRT. If the sample number is expected to be small (for example, if the cost per observation is high), it may be simpler to determine whether or not sampling should continue directly from the MBR definition.

To illustrate the computations needed to determine the equivalent SPRT, the following numerical example is presented. This same binomial example is found in $\int^3_7$, page 280. The considerable decrease in the computational complexity, when compared with the BSR, is evidence for the remarks made in section 1.4.
The problem is to test $H_1: \Theta = \Theta_1 = 1/3$ against $H_2: \Theta = \Theta_2 = 2/3$, when

$$L(\theta_i, d) = \begin{cases} 0 & \text{if } d = \theta_1 \\ 1 & \text{if } d \neq \theta_1 \end{cases}, \quad c_k(x_k) = k/38.25, \quad \xi_0(\theta_1) = 1/2.$$ 

Let: $a_k = \text{posterior probability that } \Theta = \Theta_1, \text{ if } x_k \text{ is observed},$

$$s = \frac{k}{k * x_k},$$

$$a' = 1 - b' = \text{probability that } \Theta = \Theta_1 \text{ and}$$

$$h = \log(a'/b').$$

Then, for any $a'$, $a_k < 1/2$ if and only if

$$a'(\frac{1}{3})^{s} (\frac{2}{3})^{k-s} < b' (\frac{2}{3})^{s} (\frac{1}{3})^{k-s}, \quad \text{or}$$

$$s > s^* = \frac{b'}{2 \log 2} + \frac{k}{2}.$$

Then, if $a' = a_0$, using the notation of section 1.4,

$$\gamma(k) = \frac{b' P_2 - a' P_1}{k},$$

where $P_1 = \sum_{s=1}^{k} \binom{k}{s} (\frac{1}{3})^{s} (\frac{2}{3})^{k-s}$ and $P_2 = \sum_{s=s^*}^{k} \binom{k}{s} (\frac{2}{3})^{s} (\frac{1}{3})^{k-s}$, it is assumed that $a_0 > 1/2$. To establish the critical value of $a'$, that is, such that $a \geq a' > 1/2$ implies sampling stops and $H_1$ is accepted, $c$ is compared with $\gamma(k)$ for various values of $a_0$, where $c = 1/38.25 = .0262$.

If $h' = 1$, then $a' \geq .75$, and $\gamma(2) = .020$, $\gamma(3) = .013$ and $\gamma(4) = .021$ indicating that $h' = 1$ is too large. If $h' = .6$, then $a' \geq .65$, $\gamma(2) = .042$ showing $h = .6$ to be too small. By continuing this process, it is quickly found that $a' = .72$ is approximately the
critical value. To find the SPRT boundaries, the notation of \( \sqrt{27} \)
page 279 is used and

\[
\alpha_* = \left[ \log \left( \frac{.72}{.5} \right) \left( \frac{.5}{.28} \right) \right]^{-1} = 2
\]

i.e., the SPRT has boundaries (-2,2). Confirming Theorem 3, these
boundaries are not outside any of the Bayes solutions; in fact, they
coincide with the innermost Bayes solution.

For the composite hypotheses testing problem when \( c_k(x_k) = kc \)
and \( L(\theta, a) \) is a linear function of \( \theta \) when \( a \) is wrong and zero
if \( a \) is right, a detailed discussion may be found in Chapter 5 of
\( \sqrt{107} \). This chapter contains a presentation of the computations in-
volved in a preposterior analysis. Although much of the complexity in
such an analysis lies in the difficulty of finding \( \min_k R_n(k) \), which is
not needed for the MBR, their treatment will materially assist the de-
termination of the stopping regions. For example, their Chart I for
the normal distribution, variance known problem, can be used to obtain
the optimal non-sequential size \( \rho^o \) in their notation; see \( \sqrt{67} \) for
an equivalent treatment). For the MBR, at each successive stage it is
only necessary to determine whether \( \rho^o > 0 \).
CHAPTER 4

AREAS FOR FUTURE RESEARCH

Many questions regarding the MBR remain to be investigated. In addition to those peculiar to this procedure, there are those applicable to any Bayesian-type rule. The basic problem remains of establishing a common "numeraire" so that losses and costs can be quantitatively compared. Also, rules for establishing the specific \( \xi_0 \) to reflect the prior information or opinion regarding the parameter \( \Theta \) are needed.

A brief discussion of several open problems is presented below:

1. Investigation of \( \alpha_m \)

An indication of the gain achieved by using a sequential procedure would be given by a knowledge of \( \alpha_m - \alpha_\infty \). If good bounds were available for \( \alpha_\infty - \alpha_B \), a limit to the possible improvement with more sophisticated procedures (see 2.) would be known. At present only lower bounds for \( \alpha_B \left[ \int T \right] \) can be compared with \( \alpha_0 \). In fact, except for the case of testing two simple hypotheses, little is known about the advantage of using a BSR.

2. Other modifications

In the same way as the MBR discussed here has been defined in terms of "looking ahead" at fixed samples of any size, a double sequence of such rules can be defined in terms of a truncated Bayes look into the future. That is, let \( \delta_{ij} \) be such that as long as \( N \leq j \),
if \( R_n > \min \left\{ E_n \sum_{i=1}^{\min\{N_n+1, E_{n+1} \sum_{i=1}^{\ldots R_{n+1} - 1} \ldots, R_{n+i-1} \ldots \}}^{7}, R_{n+i+1} \right\} \)

another sample is taken and otherwise sampling is stopped. Since the right-hand side of the above expression is always not greater than \( \min \left\{ R_n(k) \right\} \), a general theorem of sample size inequality (like Theorem 3) can be shown. Whether the additional computational complexity implied by such rules is worthwhile remains to be investigated. The possible similarity with the sequence developed in \( \left[127\right] \) also is open to question.

3. A single critical \( R_n(k) \)

A general criterion for determining which \( R_n(k) \) is the minimum does not seem feasible in the light of the complexity shown in Chapter 5 of \( \left[107\right] \). However, some relaxation of the conditions of Theorem 4 may be possible although the minimizing \( k \) will presumably be dependent upon \( x_n \). Bounds on \( R_n \) or \( \gamma_n \) (see section 1.4) would provide sufficient conditions for either stopping or continuing.

4. Choice of alternative experiments

For problems such as the "k-armed bandit", where the distribution of the random variable depends upon the population sampled, an adaptation of the MBR may provide a fruitful approach. By separating the decision rule into one portion which decides whether to continue sampling at all, and another which evaluates the expected result of sampling from each specific population, a sequential procedure could be defined. Although not optimal in general, it is conjectured that the lost efficiency may be "slight" for at least a restricted class.
of such problems.

5. Average sample number

Theorem 1 can be used to give a, presumably poor, upper bound (in terms of $\alpha_0$) for the expected value of $N$ averaged over the prior distribution of $\Theta$. Some knowledge of $E_\Theta N_\infty$ may be possible by an adaptation of the technique used in [7].

6. Properties of the Bayes procedure

The existence of a general sequential procedure whose average risk is not greater than the fixed sample-size Bayes risk may be useful in evaluating properties of the BSR. For example, since $\alpha_\infty$ is less than $\alpha_0$ for the estimation problem found in section 3.1, the BSR cannot be a fixed sample-size procedure.
BIBLIOGRAPHY


